Explaining medical practice variation

Social organization and institutional mechanisms

Judith D. de Jong

ISBN 978-90-6905-898-6

http://www.nivel.nl
nivel@nivel.nl
Telefoon 030 2 729 700
Fax 030 2 729 729

Cover photo:	Johan Hansen
Cover design:	Richard van Kruysdijk
Word processing/lay out:	Christel van Well
Printing:	PrintPartners Ipskamp BV, Enschede

Explaining medical practice variation
Social organization and institutional mechanisms

Het verklaren van variatie in medisch handelen
Sociale organisatie en institutionele mechanismen
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Universiteit Utrecht,
op gezag van de rector magnificus, prof. dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 4 juli 2008 des middags te 2.30 uur

door

Judith Daniëlle de Jong
geboren op 21 maart 1976
te Weststellingwerf

Promotoren:
Prof. dr. P.P. Groenewegen
Prof. dr. G.P. Westert

Part of the research in this book was carried out with financial support from the Netherlands Organization for Scientific Research (NWO), under grant 510-05-0001.

Contents

1

General introduction and research questions

This chapter introduces the subject of the book and describes existing explanations for variation in medical practice. The research questions answered in this book are specified, hypotheses are derived and a demarcation of the research domain is provided. In addition, the methodological approach is introduced and the relevance of the study for both science and society are explained. The theory and methodology used for the following six chapters is discussed, thus showing the link between the six studies. This chapter ends with an outline of the structure of this book.

1.1 Introduction

Imagine the unfortunate situation that if you break your back and are admitted to hospital X, you will have to stay in bed for four to six months. But if you had been admitted to hospital Y you would have been kept in bed for six to twelve weeks (example based on a study by Öner, 1999). Why does this medical decision differ between these hospitals?

To most people medical decision-making is a black box. Patients, and other lay people, are not able to judge directly the decisions of physicians and this gives the medical profession its special social and legal status (e.g. Evans, 1990; Freidson, 1975; Starr, 1982). In general, people expect that the medical treatment they receive is provided by physicians who adhere to professional norms which are based on evidence. *"Medical practice – in contrast to most other social practice - has the image of being based on solid, scientific grounds. We have all been brought up with an understanding that medical care was established through a continual process of interaction between medical practice and medical science, with ever more sophisticated knowledge becoming available to the defenders of our precious health, the medical profession."* (Anderson and Mooney, 1990). The existence of variations in medical practice challenges that general belief. If the assumption was true that treatment by physicians is based on theoretical knowledge and the medical condition of the patient, similar patients with similar conditions would receive the same treatment, irrespective of the physician, hospital or practice they attend. However, examples from literature show something different. *"Residents of New Haven are about twice as likely to undergo a bypass operation for coronary artery disease as their counterparts in Boston, who are more likely to be treated by other means. On the other hand Bostonians are much more likely to have their hips and knees replaced by a surgical prosthesis than are New Havenites, whose physicians tend to prescribe medical treatment for these conditions. Bostonians are twice as likely to have a carotid endarterectomy…"* (Wennberg, 1990). Apparently, there are regional variations; medical treatment seems to depend on the place patients live.

Another example shows that the diagnosis does not always depend on the medical condition of the patient. Bakwin (1945) described a test in which a number of children were presented to pediatricians to see whether they needed a tonsillectomy or not. The pediatricians selected 45% of these children for a tonsillectomy. The remaining children were presented to other pediatricians, 46% were said to need a tonsillectomy. The test went on until they ran out of pediatricians, and every time the percentage of children who were said to need a tonsillectomy was about the same.

These examples also show that research attention to variation in medical practice is not new. It is a phenomenon that was already recognized in the 1930s. Back in 1938 Glover described the incidence of tonsillectomy in school children in England and found a geographical variation (Glover 1938). He showed that whether or not a child underwent tonsillectomy depended on the area in which the child happened to live, and not so much on the medical condition of that child. Still, it lasted till the 1970s before these variations received more attention in the USA and till the 1980s before interest was raised in Europe.

Today, it is widely acknowledged that variations in medical practice exist. This finding has been well documented and described (Paul-Shaheen et al., 1987). The literature shows variations in primary care with regard, for example, to diagnoses, contact frequencies, ordering of diagnostic tests, referral rates, drug prescribing, and reasons for follow-up (e.g. Taroni et al., 1990; Zaat, 1991; Martinus, 1993; Verhaak, 1993; Davis and Gribben, 1995; Peterson, 1997; O'Donnell, 2000; Guthrie, 2001; Davis et al., 2002; Davis, 2002; Congdon, 2006). Variations in secondary care have been reported for a variety of decisions such as: hospital admissions, decisions on surgery, diagnostic procedures, and length of stay (e.g. McMahon et al., 1989; Westert, 1992; Westert et al., 1993; Weinstein et al., 2004). However, variations were not random, clear patterns of variation were found on several levels of aggregation, for example in countries, regions, hospitals, and physicians (e.g. McPherson et al., 1981; Wennberg and Gittelsohn, 1982; Read et al., 1983; Paul-Shaheen et al., 1987; Westert et al., 1992; Arndt et al., 1995; Wennberg, 1999; Ashton et al., 1999; O'Connor et al., 1999).

Variation in medical treatment in itself is not remarkable. Patients differ and if differences are medically relevant then this is taken into account in the treatment they receive. Furthermore, patients can have different preferences concerning medical treatment that might influence medical decisions. However, these factors relating to the patient make up only part of the variation in medical treatment. The remaining part is caused by other, non-medical, factors. It is that part of the variation, which is not due to patients and appears to be non-random that is the subject of this book. In order to explain these variations, the social organization of the medical profession and medical practice is emphasized. The observation of variation in medical practice alone does not in itself provide an explanation. In addition insight is necessary into the behavioral mechanisms that underlie variation.

1.2 The problem of medical practice variation

What exactly is the problem of variation in medical practice? Non-random patterns of variation call into question whether there are factors other than medical science and the clinical condition of patients influencing the decision of a physician (McKinley et al., 1996; Burns and Wholey, 1991). Although medicine is widely held to be science, many clinical decisions do not rely on a strong scientific foundation, simply because such foundation does not exist (Andersen and Mooney, 1990). At the beginning of this book we gave an example of variation between hospitals in length of stay for spinal injuries. This variation occurs because of lack of consensus among the profession at large about the appropriate treatment (Öner, 1999). However, there is no variation within hospitals and therefore we conclude that there is a local consensus. While explaining the existence of variation, a lack of scientific evidence or consensus does not explain why there is no variation within the hospitals.

Research has shown that the choice of medical treatment can be influenced by for instance uncertainty about the most effective practice, the response to regulations, the remuneration system, the availability of beds and facilities, and the type of insurance coverage (Greenfield et al., 1992; Delnoij, 1994; Westert, 1992; Kroneman, 2001). Apparently, non-medical factors are needed to explain variations in medical practice (Bachman and Freeborn, 1999; Donelan et al., 1997; Eisenberg, 1985; Greenfield et al., 1992; Langley et al., 1992). Besides these factors, the influences of the (social) work-environment on medical decisions have been studied (Arndt et al., 1995; Westert et al., 1993; Westert, 1992). McClure (1982) postulates that systems of people, including medical care systems, tend to perform in the way they have been structured and rewarded to perform. *"To explain the performance of a particular system of physicians and hospitals, it is therefore necessary not only to know the structure and incentives within this system, but also to know the structure of the larger environment in which this system is placed and the incentives that this environment places on a given system"* (McClure, 1982). The medical profession is subject to changes in that larger environment, such as technological, organizational and societal changes (Freidson, 1975). These changes will probably have an effect on medical behavior.

Technological changes increase both the medical possibilities and the means of spreading medical knowledge, resulting in more medical interventions and increasing health care costs. Organizational changes such as working in groups instead of working alone affect the work environment. Societal

changes while affecting the environment also affect the profession from within. Examples are processes of rationalization that occur throughout society and have influenced the medical profession such as transparency and accountability. Other important societal changes are increased individualization and trends towards consumerism. In general, there is a societal trend towards customization of products. This influences the relationship between patients and medical professionals. Processes of rationalization that are supposed to decrease variation could contradict with a change towards patient choice in health care that might increase variation.

In short, part of variation in medical practice is due to non-medical factors. This observation is in contrast with the idea that medical practice is based on solid scientific grounds. The environment of physicians plays a role in their medical decision-making. Besides, societal changes affect variation in medical practice. Contrary to an over-professionalized theory of physician behavior, there is variation in medical practice that cannot be reduced to clinically relevant differences between patients. Medical practice variation has been found to show clear patterns by country, region, hospital and practice. This patterning of variation demands explanation.

1.3 Existing explanations for medical practice variation

Before elaborating on explanations for non-random variations, I will go into an explanation for the existence of variation itself that is lack of consensus about the appropriate treatment, in other words professional uncertainty.

Professional uncertainty in medical decision making

Variation is the result of applying theoretical knowledge to individual patients. In applying that knowledge there can be professional uncertainty in how this should be done and in what treatment should be chosen. Wennberg and Gittelsohn (1982) hypothesized that the amount of variation surrounding a certain decision is determined by the level of professional uncertainty. The larger the professional uncertainty, the more possibilities for individual beliefs, the more variation in individual styles of practice will be found. For some diagnoses it is clear what should be done, thus limiting the range of reasonable decisions. For other diagnoses there is professional uncertainty on what should be done. This offers an opportunity for different decisions that all seem appropriate and causes medical practice variation. When the level of professional uncertainty is high, there is more opportunity to take factors into account other than medically relevant characteristics.

These are so-called weak situations or low-cost situations. Professional uncertainty explains why there is more variation for some diagnoses than for others. However, it does not explain patterns of variations in medical practice for the same diagnosis.

Preferences and constraints

Several explanations for patterns of medical practice variation have been suggested in the literature. These can be divided into a preference-centered and a constraint-centered approach, summarized in table 1.1.

Table 1.1 Summary of Explanations for Variation

Type of Explanation	Content	Implications	Study Citations
Preference centered	Practice styles are caused by differences in preference for certain procedures	Practice style remains the same when the social context is changed	Wennberg and Gittelsohn (1975) and Chassin (1993)
Constraint centered	Practice styles are caused by differences in characteristics of the (social) context, providing opportunities and constraints	Similarities of practice styles within the same social context, practice style changes with changes in the social context	Westert (1992) and Westert and Groenewegen (1999)

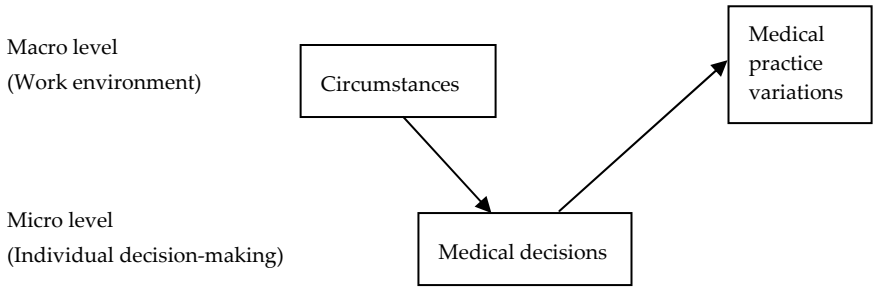
Preference centered explanations

Most of the existing explanations are based on the idea that physicians differ in their preferences concerning treatment. Wennberg and Gittelsohn (1975) developed a practice style hypothesis after concluding that differences in the populations served did not explain differences in utilization rates among hospital market areas (Wennberg, 2004). The supply of facilities and physicians did explain some, but not most of the variation. Later studies showed that variation is a worldwide phenomenon, demonstrating similar patterns of variation for common procedures, *"even though obvious differences exist in the supply of surgeons, the organization and financing of services, and in the cultural and demographic characteristics of hospital market area residents. There is one common factor: physicians in each of these areas of the world read the same medical literature, participate in the same scientific traditions, and share the same scientific uncertainties concerning the value of certain procedures"* (Wennberg, 2004). They hypothesized that the judgment and preferences of groups of physicians give rise to a unique pattern over time, which has been termed a 'surgical signature'. Wright et al. (1999) argued that through processes of

attraction and retention, practice styles tend to cluster, resulting in geographical patterns of variation that are consistent over time. Chassin (1993) suggested that one of the causes of this clustering of practice styles is a difference in the prevalence of physicians who are enthusiastic for certain procedures due to authoritative teachers in continuous medical education.
The practice style and enthusiasm hypotheses assume that physicians have a preference for certain procedures, but there is no clear theory on how these preferences come into being or change. Although we do not argue that preferences do not exist, there are two difficulties arising with the approach based on preferences (Westert and Groenewegen, 1999). The first is that it is difficult to explain why physicians change their behavior suddenly, for example when their remuneration system is changed as was found by Krasnik et al. (1990). As opposed to giving feedback on medical behavior a change in remuneration system has nothing to do with professional content or uncertainty. Physicians collectively changed their medical behavior, for non-medical reasons. The second problem concerns the methodology. It is difficult to test hypotheses on the origins of differences in preferences. In view of the fact that preferences are developed in the course of education and socialization, data collection has to go a long way back.

Constraint centered explanations
In order to solve the problems of the preference centered approach, Westert and Groenewegen (1999) offered an additional approach. They emphasized (social) conditions or local circumstances that influence the behavior of physicians by providing opportunities and constraints. Physicians are assumed to be goal-oriented individuals. The general goals of physicians are taken as given rather than their preferences for treatment. Westert (1992) applied this approach and introduced a model of local standards that predicts similarities among physicians who share a common work environment and thus a social system and similar constraints. Figure 1.1 depicts how circumstances influence the medical decisions physicians take. Different circumstances, providing different opportunities and constraints, cause different medical decisions, thus bringing about medical practice variations.

Figure 1.1 Relationship between circumstances and medical practice variations

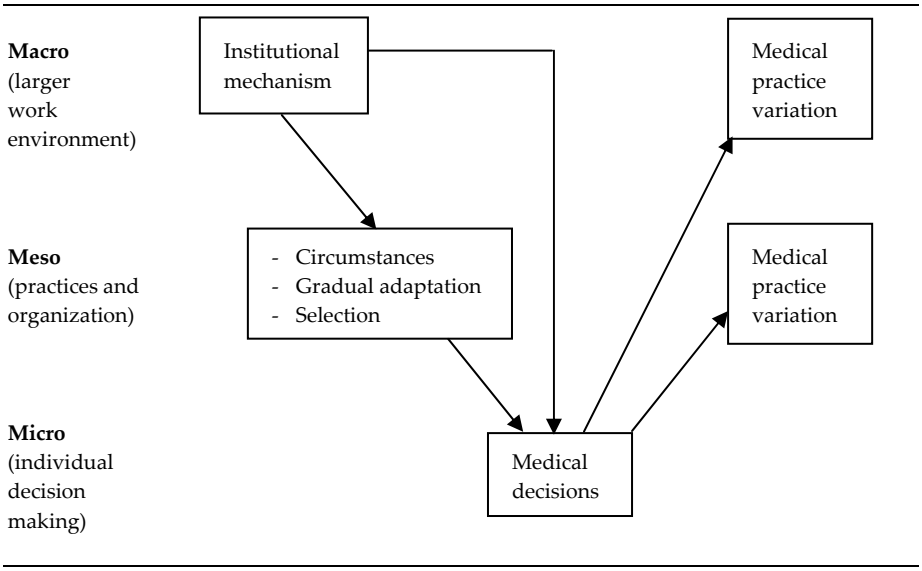


The combination of uncertainty and the importance of circumstances can be derived from the work of Freidson (1975). He theorized that for some diagnoses there are no strong expectations, or norms, on how to treat. Professional uncertainty therefore exists and there is an opportunity for different practice styles. He described the social system in which physicians take their medical decisions and expected that within groups physicians behave in similar ways, because they are mutually dependent. While professional uncertainty provides for the opportunity of variation in medical behavior, while sharing a work environment is related to similarities in behavior. Professional uncertainty does not mean that individual physicians are uncertain. Physicians find solutions to the problem of professional uncertainty, for example, by doing what colleagues are doing (Eddy, 1984). When acting like others, one does not have to explain, or legitimize, one's behavior. It becomes a norm. Thus, patterns of variation come into being. Variation between and homogeneity within shared work environments is empirically confirmed by Westert (1992).

In this book variation between and homogeneity within shared work environments will be further explained. In the first part, three mechanisms that generate homogeneity will be outlined: selection, gradual adaptation towards group norms and rapid adaptation to circumstances. Selection of new physicians in a group might be based on similarities. Gradual adaptation to each other, when physicians work together, might cause similarities. Rapid adaptation to circumstances, when there are incentives for certain behaviors, might be a cause of similarities as well.
In the second part of the book, three institutional mechanisms will be introduced which complement these ideas: the regulative mechanism, the normative mechanism and the cultural cognitive mechanism. The regulative mechanism refers to rules and regulations, the normative mechanism relates

to professional norms and the cultural cognitive mechanism concerns the framework physicians' use in their decision-making. The theoretical model is depicted in figure 1.2. The theory will be explained in the next sections.

Figure 1.2 Theoretical model for the explanation of medical practice variation



1.4 Research questions

The focus of this book is on variation in medical practice. Two types of questions are asked; a descriptive question addressing the extent of variation, and an explanatory question addressing explanations for medical practice variations. This book focuses on the explanatory question.

Before seeking explanations it is important to describe the extent to which variation in medical practice occurs. The first question asked in this book is:

1 *'To what extent does medical practice vary between (groups of) physicians?'*

Shared circumstances influencing the behavior of physicians could cause patterns of similarities in behavior. Although much has changed in medicine and much progress has been made on how to solve medical problems,

variations in medical practice remain. In this book we focus on different mechanisms influencing variation. The second question addressed is:

2 *'How can variations in medical practice between (groups of) physicians be explained, after taking medically relevant factors into account?'*

1.5 Theoretical considerations and general hypotheses

The importance of circumstances

The importance of examining physicians' behavior in relation to the group of physicians to whom they belong, has increased since medicine has become teamwork and physicians have become more dependent on hospitals (Groenewegen and Van Lindert, 2001). Working in partnerships or groups implies mutual dependency and influences, perhaps unintentionally and unconsciously, treatment decisions (Groenewegen et al., 2002b). Freidson (1975) argued that in group practice, physicians cannot make their decisions autonomously, because they depend on colleagues, and are controlled by them. Even though individual physicians can be conscientious and ethical, circumstances can cause them to change their performance. Furthermore, Freidson (1975) stated that the social environment in which physicians work is more important for their medical behavior than their formal professional education. He found evidence from several studies that with regard to the level of technical performance, client approach, 'cynicism' and ethicality there was less variation with the individual's professional training than with the social environment in which they work after education.

These examples point towards the units at which homogeneity might be expected. The appropriate unit is one of the questions arising in the debate in which explanations for variation are sought. Following hypotheses and findings from the literature, the general idea in this book is that variation in medical practice between physicians is related to differences in incentives and circumstances between their work environments.

Visibility of behavior

An important aspect of working in the same environment, the same practice, the same partnership, is that physicians use shared resources, such as personnel and equipment. These shared resources imply mutual dependency. The work environment serves as a social system in which

decisions take place and this social system may cause physicians to make similar medical decisions. Norms develop within groups of physicians. Physicians run the risk of being criticized by others working in the same group or partnership, when deviating from the partnerships' norm. Within partnerships both collective and individual goals affect behavior and can cause similarities. To protect their collective goals and to overcome free-riding and maintain solidarity, colleagues develop an informal system (Lazega, 2001). Physicians avoid being criticized by conforming to the partnership norm. The question that arises is: What are the conditions for (group) norms to develop and for local standards to emerge?
Whether local standards come into being or not, might depend on the extent to which physicians working in a partnership or hospital form a social group, and are able to develop into a normative community. It is unlikely that (group) norms develop for behavior that is unlikely to be criticized. Behavior will only be criticized by colleagues if these colleagues know about it and if it has consequences for them. This is the case when behavior, or the outcome of behavior, is visible, or shared resources are used.

Hypothesis 1: *The more visible the activities of physicians are to colleagues in the same practice, the less variation there is among them.*

A next step is to search more specifically for the causes of patterns of medical practice variations. The theoretical model used is illustrated in figure 1.2 and explained here in more detail.

Selection, gradual adaptation and circumstances
Physicians sharing a work environment are hypothesized to show similarities. Three different mechanisms are specified that might explain medical practice variation. The selection of new partners might be directed towards similarities, like seeking like. A gradual adaptation to each other within a partnership might result from processes of peer approval. Finally, the circumstances that are shared by partners might lead to similar behavior more or less directly. These are three general processes that are supposed to generate homogeneity within a group.

The process of selection affects the people who are chosen to work with. This mechanism includes both self-selection and selection by others. For physicians, careful selection might be very important in finding a new partner, because it is not so easy to get rid of a partner once she or he has entered the partnership. A bad choice therefore has huge consequences. It is expected that physicians will search for new partners among physicians they

already know, or who have characteristics that can be used in the selection process as proxies for characteristics that signal trustworthy colleagues. In the process of selection implicitly or explicitly a profile will be used that a new physician has to meet. Nowadays, it is common that vacancies for physicians are advertised. It is, however, noteworthy that advertisements only rarely stipulate specific requirements from candidates (Groenewegen and Marangos, 2004; Groenewegen and Hofland, 2003). Similarities in attitudes and in certain characteristics, such as studying at the same university or in age, would give an indication that selection processes play a role. Selection will primarily be based on attitudes and certain characteristics, and not so much on actual medical behavior. Attitudes can come up for discussion in a job interview and characteristics will be mentioned in the application form, while behavior only becomes visible when people are already working together.

Selection does not necessarily cause similarities. A physician can be selected because of a difference with the existing partners. A reason for this can be that the partnership wants to provide for omissions in treatment for their patients. For example, a partnership can choose a woman, to fulfill the need of patients to talk to a female doctor for instance when they have typical female problems (Van den Brink et al., 1994).

Although similar people are attracted to each other (Fehr, 1996), it is also often assumed that people who interact influence each other and therefore become more alike (Leenders, 1995). The classical sociological forces of socialization and normative control can generate homogeneous practices through gradual adaptation. Formal education is one way of socialization, informal education occurs when physicians start working together. When a physician starts to work with others in a partnership it is expected that sooner or later they will adapt to each other. They run the risk of losing social approval if their medical performance is criticized by their colleagues (Freidson, 1975; Westert, 1996; Westert and Groenewegen, 1999). In (al kinds of) groups sanctions are imposed against group members who deviate from norms (Wilensky and Ladinsky, 1967). For physicians peer review is increasingly part of normal routine in partnerships or local groups. Besides, when people evaluate their own personal skills or self-image, they rely on role equivalents (Burkhardt, 1994); physicians will compare themselves with colleagues. Consequently, there may be pressures towards adaptation within partnerships. The process of gradual adaptation will lead to similarities in attitudes and behavior. There will be more similarities the longer physicians work together.

Another process that can cause similarities is the (rapid) adaptation to the circumstances physicians work under. Circumstances provide constraints and opportunities for behavior, therefore change in behavior can be rapid when circumstances change. Research into the effects of a change in the system of payment, for instance, showed a rapid adaptation to the new circumstances as found in study by Krasnik et al. (1990) already mentioned.

Within the same conditions there will consequently be less variation in behavior than between conditions. This process of rapid adaptation through shared circumstances can be distinguished from selection and gradual adaptation to each other by comparing physicians working under the same circumstances. The adaptation to common circumstances will supposedly be much quicker than the adaptation through interaction, peer review and behavioral confirmation.

Hypothesis 2: *Physicians in partnerships are more similar in professional attitudes and behavior to each other than they are to physicians with whom they do not share a work environment.*

Hypothesis 3a: *Physicians select colleagues that are similar in attitudes and behavior.*
Hypothesis 3b: *Physicians adapt their attitudes and behavior to the colleagues they share a work environment with.*
Hypothesis 3c: *Physicians facing similar circumstances show similarities in behavior.*

A further test of the importance of circumstances
We will apply a test to find out whether circumstances are more important in explaining variation than preferences. The hypothesis that physicians working in two different hospitals act differently in each of those hospitals will be tested. This would imply that for physicians treating similar patients, variation within hospitals would be small compared with variation between hospitals. A second implication would be that physicians working in more than one hospital conform to the usual practice of each hospital (Westert, 1992). This implication cannot be deduced from an approach based on preference, as the preferences of an individual would not change when working in another hospital. In short, if preferences explain medical behavior physicians working in two hospitals would show the same behavior in both hospitals. If circumstances explain medical behavior, and circumstances differ between two hospitals, physicians working in both hospitals would show different behavior in each of these hospitals.

Indications were found by Westert et al. (1993) that physicians working in several different hospitals in the Netherlands have a patient length of stay close to the usual practice in the hospital where the surgery was performed. Griffiths et al. (1979), using data from the British National Health Service, also found that average post-operative stays were similar among physicians who practice in the same hospitals, while the average between hospitals was significantly different for physicians who practice in more than one hospital.

Hypothesis 4: *Variation between physicians working in the same hospital is less than variation between hospitals.*

Hypothesis 5: *Physicians working in different hospitals, in which medical behavior differs, will show medical behavior most similar to the usual behavior in the hospital in which the patient is treated.*

Institutional mechanisms

The profession of medicine is increasingly subjected to the influences of market competition, enforcing it towards standardization (Ritzer and Walczak, 1988). There is a change from professional dominance to managerial-market orientation (Scott et al., 2000). The United States may be in front, but Europe is catching up with the introduction of guidelines, protocols, diagnostic related groups or similar reimbursement systems that exert pressure to make more efficient use of health care resources. In terms of rationalization, the profession is changing from being characterized by substantive rationality to formal rationality. It is changing from being led by social values aiming to make rational choices to being controlled by rules and regulations (Ritzer and Walczak, 1988).

Institutional changes in the health care sector that increased formal rationality are expected to reduce variation in medical practice as physicians are increasingly operating in a predictable manner. Besides, they are also increasingly operating in a similar environment due to processes of rationalization that cause similarities in institutions; hospitals are becoming more alike (Scott et al., 2000). Based on a literature review Groenewegen and Westert (2004) concluded that there is indeed a downward trend in medical practice variation. How can these changes be explained?

Scott (2001) distinguishes three mechanisms explaining how institutions work: the regulative, normative and the cognitive mechanisms. These three mechanisms influence variation. The regulative mechanism does so by setting rules, monitoring and sanctioning. This mechanism fits the shift

towards a managerial orientation previously mentioned. The managed care system in the USA is an example of the regulative mechanism. Insurers here set rules on the treatment that can be given to patients. The normative mechanism refers to peers' expectations and matches professional control. An example is guidelines developed by the profession itself. These guidelines have a normative character. The cognitive mechanism relates to a common framework of meaning, similarities are caused by imitation or by the use of the same (computerized) systems of interpretation that influence treatment decisions. Computerized systems rationalize medical practice and can make it more scientific. They are meant to improve the capacity of physicians to make better decisions whereas the complex problems physicians deal with surpass their cognitive capacity (Berg, 1995). The focus of the second part of this book will be on these institutional mechanisms. The influence of each of these mechanisms on variation will be examined.

The relationship between the institutional mechanisms and circumstances is that institutional mechanisms, as used in this book, shape the working environments that influence variation. Therefore, we will focus on these institutional mechanisms and their relationship to variation in medical practice. In general, it is hypothesized that all three institutional mechanisms decrease variation in medical practice within the units to which they apply. Once rules are set, monitored and sanctioned within a hospital, variation within that hospital will decrease. The theory underlying this general hypothesis is put forward by Groenewegen and Westert (2004) in order to explain trends in medical practice variations. Figure 1.2 illustrates the relationship between institutional mechanisms and variation in medical practice. It shows how institutional mechanisms will directly affect medical decisions. Additionally, institutional mechanisms affect the working environment in which physicians take their medical decisions.

Hypothesis 6: *Regulative, normative and cognitive mechanisms decrease variation in medical practice within the units to which they apply.*

1.6 General methodological strategy

Decisions on a micro level result in patterns of variation at higher levels. To explain these patterns of variation, mechanisms on a micro level need to be specified. As already explained in this chapter a constraint-centered approach is chosen. The general idea is that decisions are the result of

considering alternatives. Which alternative is chosen depends on social circumstances and institutional organization.

In order to allow for the hypotheses to be empirically tested, data sets are constructed using secondary data that give the opportunity to test the hypotheses. Data are used from two national data collections held in the Netherlands amongst general practitioners: the First and Second Dutch National Survey of General Practice (DNSGP1 and DNSGP2). Furthermore, data were obtained from the New York Statewide Planning and Research Cooperative System (SPARCS).

Data

The First and Second Dutch National Survey of General Practice

Data were collected in 1987 and 2001 from 103 and 104 general practices in the Netherlands respectively. Practices were selected based on three stratification criteria: region, level of urbanization and practice type (Van der Velden, 1999; Westert et al., 2005). The data were collected via a questionnaire for GPs, a census among all patients on the practice lists consisting of approximately 12,000 health interviews per time point and more than one million recorded contacts of patients with their GPs in both years. DNSGP1 included 103 practices, 161 physicians and 335,000 patients. DNSGP2 included 104 practices, 195 physicians and 390,000 patients. Data on morbidity, contact details and interventions such as drug prescriptions and referrals were collected on paper forms (DNSGP1) or extracted from electronic medical records (DNSGP2).

The New York Statewide Planning and Research Cooperative System

The New York Statewide Planning and Research Cooperative System (SPARCS) is a comprehensive patient data system established as a result of cooperation between the health care industry and government. SPARCS is a major management tool assisting hospitals, agencies, and health care organizations with decision-making in relation to financial planning and monitoring of inpatient and ambulatory surgery services and costs in New York State. Initially created to collect information on discharges from hospitals, SPARCS currently collects data on patient characteristics, diagnoses and treatments, services, and charges for every hospital discharge, ambulatory surgery patient, and emergency department admission in New York State. For this study the 1999, 2000 and 2001 SPARCS-data were used. Seven Diagnosis Related Groups (DRGs) were studied consisting of two medical, one surgical and four obstetric DRGs. More than 1000 physicians and 150 hospitals were included per DRG.

Methods

The circumstances under which decisions are taken are important, when searching for explanations of medical practice variation. Data that are used for this purpose are hierarchically structured: patients are nested within physicians and physicians are nested within practices or hospitals. Individual observations are not independent and this should be taken into account in the analyses. Besides, the question is at what level variation shows up. Analyzing the data with general linear models would be problematic. Multilevel models are used to analyze hierarchical structured data (Goldstein, 1995; Snijders and Bosker, 1999; Hox, 2002).

Multilevel analysis is an extension of general linear models, like ordinary least squares or logistic regression analysis. Logistic regression analysis allows us, for example, to estimate the relationship between referral and age in a study population, assuming that there is no physician or practice effect in addition to the characteristics of the patient. With multilevel analysis, however, total variation in clinical activity is separated into different parts for instance at the patient level, the physician level and hospital level. The analyses in this book were performed using multilevel analyses.

1.7 Demarcation of the empirical research domain

This book's major purpose is to find explanations for the existence of (patterns of) variation, to get at the mechanisms that cause these variations. Most existing research focuses on specific diagnoses or specific treatment decisions. In this book we will study explanations on a more general level. We will choose certain situations and decisions in order to test general ideas. We will study medical decisions by general practitioners in the Netherlands and decisions taken by medical specialists in the USA, because these different situations best fit the hypotheses that are tested. The focus is on the social organization of medicine and work-related circumstances, not on its content. This means that no conclusions are drawn on the quality of medical practice.

We study medical practice variations because of the position of physicians. Historically, their actions are highly autonomous. They are directly related to collective costs and their clients have to trust their expertise, because they lack knowledge to judge their decisions. Everyone, sooner or later, will have to deal with physicians and literally put their lives in their hands. Besides, there is a policy interest in medical practice variations because, world-wide,

countries deal with rising health care costs and the existence of variation can be interpreted as a sign of unnecessary spending.

This book is not meant to provide an answer to the question of whether variation in medical practice is good or bad. Without a doubt, medical practice variation raises questions about the quality of care and about the use of resources. Although different treatments can be effective, and innovations are not likely to occur when there is no variation, the existence of variation will have an effect on the medical profession. Physicians will have to explain why there is variation. Policy makers and third party payers will get involved as they might be convinced that health care expenses can be limited when all physicians choose the most cost-effective treatment. Variation might give patients a choice in the treatment they prefer, at least in theory and as far as patients are able to choose. It might also help in finding better treatments but as long as there are questions about the justifiability of variation, physicians will have to deal with it, limit or explain it, or otherwise insurers or the government probably will.

The research questions addressed in this book will be answered from a social sciences perspective. Variations in medical practice will be linked to the behavior of physicians. The work of epidemiologists shows that variation exists. For understanding the existence of medical practice variation insight in the behavioral mechanisms that underlie variation is necessary.

1.8 Scientific relevance of the study

The scientific relevance of our study is twofold as two problems are addressed: existing explanations for medical practice variation fall short and, as a result, not analyzed on the right level.

Understanding medical practice variation
Research describing medical practice variation has led to the acknowledgement that non-random medical practice variations exist. Non medical factors are needed to explain those variations. Several studies aimed at identifying variables associated with medical practice variations. The findings, however, do not always point in the same direction. Apparently, the relationships are conditional. The existing literature shows that medical practice variation exists and points towards directions for where to find explanations. Still, the mechanisms underlying medical practice variation

are not fully understood. There have been attempts to give a theoretical explanation of medical practice variations (see section 1.3 of this book). However, these explanations fall short as they are contradicted by research results. There is no question about the existence of variation and that it should be addressed. The problem, however, is that "*we still don't know very much about why physicians do what they do, or how to get them to do different things. An important subsidiary problem is that most of our contemporary health services research, whether of the Wennberg epidemiological school or the more hegemonic microeconomic models, doesn't help us very much in developing that understanding.*" (Vladeck, 2004). Vladeck (2004) furthermore states that the methods and insights of sociologists and social psychologists are necessary in order to understand medical practice variation. We will examine medical practice variation from a sociological perspective as we are interested in how the social organization influences medical decisions. We will use a constraint-centered approach instead of a preference-centered approach to overcome the problems of the latter.

Because most explanations were based on the preferences of physicians, research focused on individuals. We argue that for explaining medical practice variations it is necessary to examine both the macro level of the work environment and the micro level of the medical decisions. Medical practice variation is a macro level phenomenon. Following Coleman (1990) macro level phenomena can typically be explained by micro level mechanisms. Decisions on a micro level, the decisions individual physicians take, result in patterns on a macro level which translate to patterns of variations in medical practice. Which decisions physicians make depends on circumstances, both (work-related) social circumstances and institutional organization. Thus, differences between circumstances cause differences in medical decisions resulting in medical practice variations. Therefore, circumstances are important in explaining medical practice variations.

Freidson (1975) theorized that the social system in which physicians take their decisions influences these decisions, because of mutual dependencies. He predicted similarities within groups of physicians. Westert (1992) empirically confirmed this hypothesis for decisions on post-operative lengths of stay in hospitals and introduced the theory of local standards. The study by Westert showed the level at which variation in medical practice occurs. It indicated that the major causes of variation are not primarily to be found in physician characteristics but in characteristics of the (social) environment. This theory can be further tested, and extended moreover, as

current techniques offer the opportunity for a refinement of the empirical methodology.

In this book a number of hypotheses for explaining medical practice variation are considered, data sets are constructed that allow for empirical testing of the hypotheses and rigorous analyses are applied. Most of the hypotheses tested in this book are new. Some hypotheses have been tested before, but mostly with little data. We are able to test a large variety of hypotheses with extensive data and advanced analyses, for physicians in primary and in secondary care, in the Netherlands and in the USA. So far, the theory has only been tested on physicians in hospitals, not on physicians in primary care. If the theory holds in general, it would apply to both physicians in secondary and in primary care. In addition, we will further test existing explanations in this book, and will extent the theoretical model. Insights are used from theories on selection processes, social interactions, the medical profession and institutional mechanisms. Mechanisms like selection, gradual adaptation and rapid adaptation to circumstances are distinguished in explaining medical practice variations and are empirically tested. This has never been done before. In addition, institutional mechanisms are introduced and empirically tested as an explanation. This is also a new element in explaining medical practice variation.

Progress will also be made concerning the empirical approach, as the methodology used for testing the hypotheses is partly new. Although multilevel models have been used in health services research, there are hardly any studies that focus on variation components. The shortcoming of using general linear models instead of multilevel analysis is a loss of information through aggregation. This occurs when lower level characteristics are assigned to higher levels, or misestimated precision in estimating coefficients, when higher level characteristics are assigned to lower level units. Multilevel analysis overcomes these problems. This book is meant to contribute to our understanding not only of practice variations but also of the mechanisms underlying such variation.

1.9 Societal relevance of the study

When similar people, that is, patients with the same medically relevant characteristics, are not treated in a similar manner then a number of questions are raised that are of concern in health care and society as a whole. Do patients receive the treatment that works best? Is the relationship

between costs and effects acceptable? Do all people have equal access to health care? These questions are of importance to all stakeholders in the health care system, and concern effectiveness, efficiency and equity (McPherson, 1990). Evidence on medical practice variations implies that there might be inappropriate services, wasting resources and maybe even harm to patients (Evans, 1990). The implications of variation in medical practice can be viewed from different perspectives; a health outcome perspective, a consumer or patient perspective, a health profession perspective, and a regulatory perspective.

Variation viewed from different perspectives

The health outcome perspective

The health outcome perspective relates to the appropriateness of the care provided. If similar patients receive different care, are some receiving unnecessary care? Is there under or over-use of health care resources? Although over-use gets more attention, under-use could be a problem as well (McNeil, 2001). Sources of undesirable variation need to be identified, whether they are indicators of over-use or under-use, with an emphasis on those sources that can be influenced to improve the quality of medical care. If similar patients receive different treatments, or if in some areas it is more likely to receive a certain procedure than in others, some do not get what they need, or get more than they need. Both are undesirable. The former means that some people's health could be improved; the latter means that there is a waste of resources by providing ineffective, unnecessary care. Part of variation can be legitimate, while another part is not. Anderson and Mooney (1990) define variation as a result of differences in patient characteristics such as morbidity and age, as legitimate, while variation caused by ignorance is considered illegitimate. That part of variation that is not legitimate should be reduced. Variation may raise questions about the quality of care, although the level of variation is not directly linked to the quality of care. Up to now, there is no evidence that less variation is related to a higher quality of care (Weide et al., 1992; Fertig et al., 1993).

The patients' perspective

From the point of view of patients there is the difficulty of assessing the quality of care. Between physicians and patients there is information asymmetry. Patients do not have the knowledge to judge whether their physician is giving the most appropriate care. One of the roles of physicians is to be the patients' advocate and to provide the care that is needed. Patients expect to receive the best treatment no matter to which physician or hospital they go. They have to trust physicians to provide that care but variation

undermines that trust. Trust in physicians also depends on circumstances, and probably influences how variation is interpreted. Variation can be accepted to a certain extent when patients are convinced that physicians act in their best interest. This, however, probably changes when patients are not so sure about physicians' reasons for choosing certain treatment. For instance in the managed care system in the USA, where physicians have a financial risk, patients are not sure that physicians give the treatment that is best for the patient, or that the decision is based on financial considerations (Kralewski et al., 2000). This undermines patients' trust in physicians and will probably affect how variation is interpreted.

The health profession's perspective

From the health profession's perspective there is the problem of undermining the profession's claim that treatment is based on scientific evidence and therefore physicians should enjoy privileges derived from that claim. Physicians have argued that their medical behavior can be judged only by physicians, and not by laypersons. Their unique knowledge entitles them to special status in debates about health care and plans on the organization and financing of the health care system. Variation in medical behavior raises questions on the knowledge of physicians, or the way they interpret it. Variation could be a sign of the complexity of medical problems. However, it should be justified. If they can not justify it, variation undermines the special status of physicians, their autonomy. They might more easily be portrayed as an interest group protecting their economic and social standing (Blumenthal, 1994). The existence of variation in medical behavior is a reason for policy makers and managers to challenge professional autonomy and to control the behavior of physicians more and more. If physicians want to protect their collective autonomy they will have to limit variation themselves, before someone else does.

The regulator's perspective

The last perspective we mentioned is the regulator's perspective. We have already referred to it when discussing the other perspectives. The questions that are raised by variation relate to efficiency; is the medical care provided worth the costs? Growing awareness of variation is related to increasing costs of medical care. Insurers interpret variation as a sign of physicians using unnecessary (expensive) treatments and they therefore try to set a norm for all treatments. The government aims at a health care sector that is affordable and, *"faced with lack of convincing evidence by doctors on health outcomes on the basis of which they might otherwise justify their use of scarce resources, politicians may well want to emphasize cheapness rather than efficiency"*

(Andersen and Mooney, 1990). In this context, information on the causes and explanations of variation can be used to improve the quality of health care.

The danger of medical practice variation lies in the fact that it can be misused or misinterpreted by the different actors in the health care debate. Physicians have a legitimate reason for concern, in their role as the patients' advocate. There is a shift from professional to managerial control. Power over decisions concerning patient care is shifting from physicians to third parties such as financiers. The primary interest of financiers is to maximize returns on health care investments. They might intervene in ways that may reduce costs but adversely affect the welfare of patients (Blumenthal, 1994). It is important to have a thorough understanding of medical practice variation to be able to prevent its misuse at the cost of patients' health.

Understanding medical practice variation is crucial in developing effective health care policies that aim at influencing the decision-making of physicians (Stano, 1993). Physicians should use new knowledge on better treatment modalities. However although innovations are introduced, they do not diffuse fast. Awareness of and a positive attitude towards, a certain, better way of practice is not enough to change behavior (Lomas et al., 1989). Apparently, the decision-making of physicians is influenced by more than research evidence, even if there are clear and concrete recommendations. There are external barriers, such as time limitations, insufficient staff or increased practice costs that affect the ability of physicians to execute recommendations (Cabana et al., 1999). Besides, professional uncertainty, lack of scientific foundation, and difference of opinion exist in medical practice and are causes of variation. In order to deal with professional uncertainty, physicians may become skeptical towards scientific evidence and more sensitive to peer influences (Freidson, 1975; Hulscher et al., 1999; Hulscher et al., 2001).

Policy makers, insurance companies, patient organizations and physicians generally aim at decreasing variation in medical practice. Their question is, if, and how, the behavior of physicians can be influenced. Before this question can be answered it is important to measure the extent of variation and understand its causes. Knowing how medical practice variation comes into being is the first step towards reducing it without reducing the quality of care.

1.10 Structure of the book

The results of the study are presented from chapter 2 to chapter 7. The hypotheses as derived in this chapter are further specified in these six chapters. The chapters are written as separate articles and can be read independently of each other. They can be divided in two parts; chapter 2 to chapter 4 go further into the importance of circumstances, chapter 5 to chapter 7 deals with institutional mechanisms. In table 1.2 a summary of the hypotheses derived in chapter one is given.
In chapter 2 the hypothesis that there is a clustering of variation within general practices when behavior is visible, will be tested. Chapter 3 studies why general practitioners within the same practices resemble each other. Three mechanisms are specified; selection, mutual adaptation and circumstances. In chapter 4 we test whether circumstances instead of preferences explain variation. The hypothesis that physicians working in two different hospitals act differently in those hospitals will be tested. Chapter 5 examines the behavior of physicians under restrictive (regulatory) circumstances of managed care activities such as those of insurers in the USA. In chapter 6 the behavior of physicians under restrictive (normative) circumstances, such as the professional guidelines for general practitioners in the Netherlands, will be examined. In chapter 7 cultural-cognitive processes on variation are examined. The hypothesis that physicians using decision aids show less variation in prescription will be tested. Chapter 8 provides a summary of conclusions of all six studies and an overall discussion.

Table 1.2 Summary of hypotheses and chapters in which they are tested

Hypothesis	**Chapter in which the hypothesis is tested**
Hypothesis 1: The more visible the activities of physicians are to colleagues in the same practice, the less variation there is among them	Chapter 2
Hypothesis 2: physicians in partnerships are more similar in professional attitudes and behavior to each other than they are to physicians with whom they do not share a work environment	Chapter 3
Hypothesis 3a: physicians select colleagues that are similar in attitudes and behavior Hypothesis 3b: physicians adapt their attitudes and behavior to the colleagues they share a work environment with Hypothesis 3c: physicians facing similar circumstances show similarities in behavior	Chapter 3
Hypothesis 4: variation between physicians working in the same hospital is less than variation between hospitals	Chapter 4
Hypothesis 5: physicians working in different hospitals in which medical behavior differs, will show medical behavior similar to the usual behavior in the hospital in which the patient is treated	Chapter 4
Hypothesis 6: regulative, normative, and cognitive mechanisms decrease variation in medical practice within the units to which they apply	Chapter 5, 6 and 7

2

Medical practice variation

Does it cluster within general practices?

A shorter version of this chapter was published in:
De Jong JD, Groenewegen PP, Westert GP. Medical practice variation: does it cluster within general practitioners practices? In: Westert GP, Jabaaij L, Schellevis FG (eds.). *Morbidity, performance and quality in primary care: Dutch General Practice on stage.* Oxford: Radcliff Publishing, 2006, chapter 14.

Abstract

In the ideal situation a patient receives a treatment, independent of the general practitioner (GP) or the practice where this GP is working. But this is not everyday reality. Patients are not alike. GPs are not alike. Practices are not alike. And this leads to (sometimes huge) variation in medical practice between GPs. Insurance companies, patient organizations, physicians and policy makers are all interested in this phenomenon. We have studied whether visibility of a GP's clinical behavior and the use of shared resources influence decisions on prescription, referrals, diagnostics, treatment, and advice. We have found that when behavior is visible to colleagues, or shared resources are used, GPs working in the same practice behave more alike. The other way around, when behavior is less visible or no shared resources are used, GPs are less inclined to behave in the same way.

2.1 Introduction

It is known that there is variation in medical practice between general practitioners. This variation has been found for the frequency of contacts (Groenewegen et al., 1992), registration of diagnoses (Marinus, 1993; Davis et al., 2002), diagnostic test ordering (Davis et al., 2002; Zaat, 1991; Peterson et al., 1997; Guthrie, 2001; Verstappen, 2004), referral (Groenewegen et al., 1992; O'Donnell, 2000), prescription (Davis et al., 2002), and return visits (Verhaak, 1993; Davis and Gribben, 1995; Taroni et al., 1990). Variation is due to differences in patient characteristics and morbidity, but also to GPs who differ in the treatment of similar patients (Mulder, 1996; Delnoij and Spreeuwenberg, 1997). These variations have primarily been studied between individual GPs. GPs used to work in single-handed practices, but by now, more than half of the Dutch GPs work in partnerships or groups (Boerma and Fleming, 1998). Working in partnerships or groups implies mutual dependency and influence on treatment decisions (Groenewegen et al., 2002b). In this chapter we will study to what extent GPs working in the same practice behave alike and to what extent GP practices differ between each other.

What is the relevance of understanding variations in treatment?
The observation that similar patients are not treated similarly raises a number of questions that are of central concern in health care. Do patients receive the treatment that works best? Is the relation between costs and effects acceptable? Do all people have equal access to health care? These questions are of importance to all stakeholders in the health care system, and concern effectiveness, efficiency and equity. Policy makers, insurance companies, patient organizations and GPs generally aim at decreasing variation in medical practice. Their question is if and how behavior of physicians can be influenced. Before this question can be answered it is important to measure the extent of variation and understand its causes.

Why do GPs in the same practice behave more alike?
Freidson (1975) already argued that professional behavior was more related to the (social) circumstances in which physicians work than to their professional education. This implies similarities between colleagues who share a work environment. Westert (1992) found similarities among physicians working in the same hospital in the use of hospital care, while there was variation between the hospitals. De Jong et al. (2003) (Chapter 3) found that general practitioners working in the same partnership showed similarities in attitudes and stated medical behavior, while there were

differences between general practitioners working in different partnerships. Apparently, sharing a work environment is related to similarities in behavior. Based on the literature, we expect to find less variation within than between GPs' practices for different clinical activities.

The idea is that variation in medical practice between physicians is related to differences in incentives and circumstances between their work environments, providing opportunities and constraints on behavior. An important aspect of working in the same environment, the same practice, is that GPs use shared resources, like assistants and equipment. The work environment serves as a social system in which decisions take place and this social system may cause physicians to make similar medical decisions. Visibility of behavior is an important precondition for similarities to come into being. When behavior is visible, norms may develop. Similarities in clinical behavior can be caused by more or less conscious processes. GPs run the risk of being criticized by other GPs working in the same partnership, because they see each others' behavior. This risk can be minimized by showing similar behavior as those colleagues. Colleagues develop an informal system to help protect their common interests and to overcome free-riding and maintain solidarity (Lazega, 2001).

This study
We studied how visibility of clinical behavior and the use of shared resources influence variation in medical practice between GPs working in different practices. Some of the clinical activities that are studied are more visible for colleagues than others, some use shared resources, and others do not (table 2.1). We expect more variation between practices than between GPs working in the same practice for clinical activities in the least visible situation, e.g. prescription, referrals, diagnostic tests performed in an external laboratory, and advice. When clinical behavior is visible and/or use shared resources, similarities are expected between GPs working in the same practices. Examples are diagnostic tests performed in the GPs practice, and treatment.

2.2 Data and methods

Second Dutch National Survey of General Practice
Data were used from the Second Dutch National Survey of General Practice (DNSGP2), collected in 2001. A total of 104 GP practices participated in this national data collection. Data were collected on contacts, patients, GPs and

practices, using Electronic Medical Records, and questionnaires (Westert et al., 2005).
Practices were excluded from the analyses when recording of the data was far from complete, e.g. the time of recording was only a few weeks, no referrals were recorded, or less than 100 interventions were recorded per 1,000 patients. Single-handed practices were also excluded.

Table 2.1 Clinical activities of General Practitioners included in the analyses

Clinical activity	**Description**	**Visible/using shared resources**	**Expectation**
Prescription	Prescription of medicines	No/no	Variation primarily within practices
Referrals:			
primary care	Referral to e.g. physical therapy, mental health care	No/no	Variation primarily within practices
secondary care	Referral to medical specialists	No/no	Variation primarily within practices
Diagnostics performed in the GPs practice	Blood test, urine test, blood pressure, weight, mammae examination	Yes/yes	Similarities primarily within practices
Diagnostics performed in a laboratory	Ordered blood test, ordered urine test, X-rays, electrocardiogram, echoscopy, endoscopy	No/no	Variation primarily within practices
Treatment	Wound care, removal of earwax, surgery, bandage, catheterization, IUD-insertion	Yes/yes	Similarities primarily within practices
Advice	Advice about work, over the counter medication	No/no	Variation primarily within practices

Multilevel analyses
Multilevel models are used to analyze hierarchical structured data (Leyland and Groenewegen, 2003; Snijders and Bosker, 1999; Hox, 1995). Multilevel analysis is an extension of general linear models, like logistic regression analysis. Logistic regression analysis allows us, for example, to estimate the relationship between referral and age in our study population, assuming that there is no physician or practice effect in addition to the characteristics of the patient. With multilevel analysis, however, total variation in clinical activity is separated into three parts: a part due to differences between patients, a part due to differences between physicians and a part due to

differences between practices (Leyland and Groenewegen, 2003; Diez Roux, 2002). In the analyses a model with three levels is used: practices, GPs, and patient diagnoses. Based on the patient list of the GP, patients are attributed to GPs.
Besides the intercept and the variance components, we will present two coefficients of correlation: the intra class correlation (ICC), measured in two different ways. The ICC is a measure of homogeneity within practices, or, in other words, the degree of resemblance within the same practice. It measures the extent to which for instance referral rates for patients are similar in the same practice. A higher ICC means that there is more homogeneity within practices; patient within the same practice are treated similarly. The ICC shows us how much of the variation is due to the combined effect of practices, and GPs, the ICC2 tells us whether this variation is primarily situated between practices or between GPs. In other words: a higher ICC2 means that GPs working in the same practices are more alike and that practices differ amongst each other, without the 'disturbing' effect of different patient diagnoses.

The model
As dependent variables we used whether:
- drugs were prescribed;
- a patient was referred to primary care, or to secondary care;
- diagnostic tests were ordered;
- diagnostic tests were performed in the GPs practice;
- treatment was given;
- advice was given.

To control for differences in data collection we included a variable for the Electronic Medical Record (EMR) system. To make patient populations comparable we corrected for medically relevant characteristics of the patient. The included 'case-mix' indicators were: age, sex, subjective health status, and as an additional indicator of health status, the number of contacts with the general practitioner per year (Powell et al., 2004). For the analysis in which prescription is studied we also included a variable that indicates whether a practice is dispensing for all or part of the patient population. This is done because medication prescribed by medical specialists was also included in the data collection for GPs who are dispensing, but are not recognized as such.
In the general model we did not take into account the diagnoses of the patients. Due to too many diagnoses, no model could be fitted in which the diagnoses were taken into account. In a second analysis seven diagnoses

were selected that frequently occur in the GPs' practice: hypertension, lower back pain, insomnia, depression, cough, respiratory tract infection, and diabetes mellitus (table 2.2). For these diagnoses we performed the same analyses as described for the general model. Treatment was not included in this analysis, because treatment, as we defined it (see table 2.1), was less relevant for these diagnoses and consequently occurred only in few cases.
For the analyses the MLwiN software package was used. All variables included in the models were centered around there means (Snijders and Bosker, 1999).

Table 2.2 Description of the variables (numbers) included in the analyses

	Prescription	**Referral**	**Diagnostics**
Practices	29	31	36
General practitioners	56	63	68
Patient diagnoses	88,553	96,074	107,739
Age:	Patients:	Patients:	Patients:
0-19	11,057 (12.5%)	12,054 (12.5%)	13,414 (12.5%)
20-44	23,826 (26.9%)	26,175 (27.2%)	29,171 (27.1%)
45-64	28,262 (31.9%)	30,655 (31.9%)	34,670 (32.2%)
65-74	12,923 (14.6%)	13,974 (14.5%)	15,694 (14.6%)
>75	12,485 (14.1%)	13,216 (13.8%)	14,790 (13.7%)
Number of contacts (mean per patient and standard deviation)	16.2 (14.2)	16.2 (14.5)	15.9 (14.2)
Health status:	Patients:	Patients:	Patients:
very good/good	58,668 (66.3%)	64,308 (66.9%)	71,991 (66.8%)
moderate	25,164 (28.4%)	26,703 (27.8%)	30,086 (27.9%)
bad/very bad	4,721 (5.3%)	5,063 (5.3%)	5,662 (5.3%)

EMR:	Practices:	Contacts:	Practices:	Contacts:	Practices:	Contacts:
microhis	5	14,796	5	14,796	5	14,796
promedico	4	15,581	4	15,834	7	24,595
elias	2	4,820	3	6,576	4	8,164
arcos	3	3,683	3	5,221	4	6,537
machis	15	49,673	16	53,647	16	53,647

Pharmacy	5 practices, 16,029 cases	
ICPC diagnoses:		Number of contacts:
hypertension (K86)		3,690
lower back pain (L03)		1,006
insomnia (P06)		1,198
depression (P76)		1,055
cough (R05)		1,434
respiratory tract infection (R74)		990
diabetes mellitus (T90)		1,318

2.3 Results

First, we will present some general results from the analyses. Second, the results for the analysis with all diagnoses will be presented. Third, we will present the results for the analysis with selected diagnoses.

General results
The ICC ranges from 0.03 to 0.20, meaning that between 3% and 20% of the variation is located between practices and GPs. In other words, between 80% and 97% of the variation is due to patients.
It seems that not much of the variation is located between practices. Still, the effects are important. Take for instance prescription: the intercept is 0.15, which is equivalent to an average probability of 0.54 for receiving prescription within a 'patient diagnosis'. The average probability of receiving prescription per 'patient diagnosis' varies from 0.35 to 0.71 between practices. This means that the probability of receiving prescription is 2.0 times higher between high and low prescribing practices.

General model: analysis with all diagnoses
There is significant variation both between and within practices for prescription, and diagnostics performed in the GPs' practice (table 2.3). For referral to primary care, referral to secondary care, diagnostics performed in a laboratory, and giving advice there is significant variation between GPs only, while for treatment there is significant variation between practices. The ICC varied from 0.04 to 0.20, for giving advice, which means that 4% up to 20% of the total variation is located between practices. Based on the ICC2 we conclude that there is more variation between practices than between GPs for diagnostics performed in the GPs' practice, and treatment. Prescription, referral to primary care, referral to secondary care, diagnostics performed in a laboratory, and advice show more variation between GPs than between practices. Our expectations are confirmed for all clinical activities (table 2.3).

Analysis with selected diagnoses
For the analysis with the selected diagnoses, in which we took into account data collection variables, patient characteristics, and diagnoses, we found significant variation between and within practices for diagnostics performed in the GPs' practice (table 2.3). For prescription, referral to secondary care, and advice, the variation between GPs was significant, as there was no significant variation between practices. Referral to primary care, and diagnostics performed in a laboratory show no significant variation between practices or between GPs. The ICC varies from 0.03 to 0.19 for diagnostics

performed in the GPs' practice. Based on the ICC2 we conclude that there is clustering of variation within practices for diagnostics performed in the GPs' practice. Our expectations are confirmed for all clinical activities.

Table 2.3 Intercept and variance components for several clinical activities for all diagnoses, and for selected diagnoses (Standard error in parentheses)

	General model (all diagnoses)	**Selected diagnoses**
Prescription:		
intercept	0.15 (0.08)	1.22 (0.12) *
practice level variance	0.06 (0.03)*	0.17 (0.10)
GP level variance	0.09 (0.03)*	0.22 (0.07) *
ICC	0.02	0.05
ICC2	0.41	0.44
Referral within primary care:		
intercept	-4.44 (1.21)*	-6.20 (0.31)*
practice level variance	0.05 (0.04)	0.13 (0.18)
GP level variance	0.09 (0.04)*	0.35 (0.21)
ICC	0.04	0.11
ICC2	0.34	0.27
Referral to secondary care:		
intercept	-3.72 (0.08)*	-5.10 (0.22)*
practice level variance	0.03 (0.03)	0.00 (0.00)
GP level variance	0.11 (0.04)*	0.76 (0.26)*
ICC	0.04	0.13
ICC2	0.19	0.00
Diagnostics in practice:		
intercept	-3.06 (0.08)*	-2.87 (0.12)*
practice level variance	0.07 (0.03)*	0.24 (0.09)*
GP level variance	0.06 (0.02)*	0.12 (0.06)*
ICC	0.04	0.19
ICC2	0.54	0.66
Diagnostics in lab:		
intercept	-3.59 (0.08)*	-3.84 (0.10)*
practice level variance	0.00 (0.00)	0.00 (0.00)
GP level variance	0.15 (0.04)*	0.09 (0.05)
ICC	0.04	0.03
ICC2	0.00	0.00
Treatment:		
intercept	-4.50 (0.11)*	--
practice level variance	0.14 (0.06)*	--
GP level variance	0.05 (0.03)	--

- table 2.3 continues -

- table 2.3 continued -

	General model (all diagnoses)	Selected diagnoses
ICC	0.05	--
ICC2	0.75	--
Advice:		
intercept	-4.52 (0.18)*	-4.45 (0.14)*
practice level variance	0.00 (0.00)	0.00 (0.00)
GP level variance	0.81 (0.17)*	0.41 (0.14)*
ICC	0.20	0.11
ICC2	0.00	0.00

* $p<0.05$

ICC: a higher ICC means that there is more homogeneity within practices; patients are treated alike in a practice

ICC2: a higher ICC2 means that there is more homogeneity between GPs within the same practice; GPs act alike in a practice

2.4 Conclusions and discussion

The treatment by physicians is assumed to be based on theoretical knowledge and the medical condition of the patient only. Patients, and other lay persons, are not able to judge the decision of physicians and this gives the profession its special social and legal status (Evans, 1990). The existence of variation, even if clinically relevant variables are taken into account, undermines this position: if physicians do what is best for their patients, based on evidence, how come there is variation in treatment between similar patients? Existing explanations are based on individual preferences or (social) circumstances (Wennberg and Gittelsohn, 1975; Westert and Groenewegen, 1999).

In this chapter we found that clustering of variation within practices depends on the clinical activity studied. GPs working within the same practice differed more from colleagues working in other practices than from their colleagues working in the same practice for activities like treatment and diagnostics performed in the GPs' practice. It was the other way round for prescription, referral to primary care, referral to secondary care, diagnostics performed in the laboratory, and advice. For these activities practices looked more alike than GPs working in the same practice. These clinical activities are less visible to colleagues, and do not make use of shared resources. When only diagnoses that occur frequently are selected and these diagnoses are taken into account in the analyses, these conclusions

do not change. Our hypothesis that there is less variation within practices when shared resources are used and when behavior is visible for colleagues, was confirmed.

Visibility of behavior

In this study we assumed that behavior is visible or not. In reality this is not an on/off phenomenon: some behavior is more visible than other. We did not make such distinctions in our study, but it might be useful in explaining the difference in the amount of variation between the clinical activities. Furthermore, we used a very narrow definition of visibility: actual behavior that can be seen by colleagues. Using a more broad definition, behavior is also 'visible' when colleagues talk about it. We could think of meetings, to which a small number of GPs attend, and in which prescription is discussed. In these meetings prescribing behavior of individual GPs can be evaluated, and is therefore 'visible'. This could have an influence on when and what to prescribe. Attending GPs can become more alike in their prescribing behavior.

Limitations of this study

In the analysis in which all diagnoses (the general model) were included, we did not take into account the specific diagnoses. This was not possible, because there were too many diagnoses, and a model could not be fitted. Differences in diagnoses in a GPs practice are an important source of variation in clinical activities. This, however, would be reflected in all clinical activities. If differences in diagnoses were an explanation of the remaining variation, all clinical activities should show patterns of variation by practice (if diagnoses differ between practices) or by GP (if diagnoses differ between GPs). That is not what we found.

Guidelines have been developed for diagnoses that frequently occur, and this could result in less variation between and within practices. However, our conclusions did not differ between the analyses in which all diagnoses are included, and the diagnoses in which only seven diagnoses were included.

Our hypothesis was about variation and not about characteristics of patients, GPs or practices that might explain variation. Often research jumps to explaining variation without carefully examining variation in the first place. For instance, there is large empirical literature about the effects of the introduction of guidelines on physician behavior. However, we almost never see the obvious hypothesis tested that the introduction of guidelines reduces variation between physicians. An exception is Verstappen et al. (2003), who reported a larger decrease of variation in the experimental group after an

intervention in laboratory test ordering (but gave no statistical test on the difference).

We now have some ideas on whether medical practice variation is within or between practices. A next step is to search more specific for the causes of patterns of medical practice variations, and how these variations should be dealt with. Multilevel analyses help us identify the appropriate level at which interventions for changing GPs' behavior should be targeted.

3

Mutual influences of general practitioners in partnerships

This chapter was published as:
Jong JD de, Groenewegen PP, Westert GP. Mutual influences of general practitioners in partnerships. *Soc Sci Med*, 2003; 57:1515–24

Abstract

The aim of this study was to find out whether or not general practitioners (GPs) within the same partnership show more similarities in attitudes and behavior than GPs in different partnerships, and what the causes of these similarities might be. Knowledge of the causes of patterns of similarities within medical teams contribute to understanding medical practice variation, which is crucial in developing effective health care policies.

Data were used from the Dutch National Survey of General Practice (1987/ 1988), consisting of a stratified sample of 161 Dutch GPs, who served 335,000 patients in total.

To find out whether GPs in the same partnership are indeed more similar than GPs randomly chosen from different partnerships, we constructed two kinds of pairs: all possible pairs of GPs working in the same partnership (actual pairs), and randomly constructed pairs of GPs who are not working in the same partnership (random pairs). For each pair difference scores were computed for a variety of attitudes and behavior. Difference scores for actual and random pairs were analyzed using multi-level analysis.

Most differences in attitudes and behavior were smaller for actual pairs than for random pairs. Furthermore, in the majority of the cases differences were no longer statistically significant after explanatory variables indicating selection, gradual adaptation and rapid adaptation through shared circumstances were taken into account.

It was found that Dutch GPs working in the same partnership showed more resemblance in attitudes and behavior than GPs not working in the same partnership. Most indications point towards circumstances, and to a lesser extent towards adaptation, as an explanation of similarities within partnerships. The implication of this study is that medical practice variations are not merely individual differences in preferred practice style, but are patterned by social processes in partnerships and local circumstances.

3.1 Introduction

Medicine, including general practice, has become teamwork. Although lagging behind countries such as the UK and Denmark, by now more than half of Dutch general practitioners (GPs) work in partnerships or groups (Boerma and Fleming, 1998). Working in partnerships or groups implies mutual dependency and influence (perhaps unintentionally and unconsciously) on treatment decisions (Groenewegen et al., 2002b). Uncertainty and differences of opinion are omnipresent and causes of variation in medical practice, albeit not the only causes. The style of practice of GPs is influenced by more than the availability of medical knowledge (Westert, 1996).

Although variation is expected and observed (Evans, 1990; Wennberg and Gittelsohn, 1982), there are also similarities in treatment patterned by the work-environment of doctors. A study on variation in length of hospital stay showed that variation within a medical team is small, compared to variation between teams (Westert, 1992). Hence, similarities are patterned by hospitals (Arndt et al., 1995; Westert et al., 1993).

The issue of similarities in attitudes and behavior within partnerships is pertinent to the broader research area of medical practice variations. Understanding medical practice variation is crucial in developing effective health care policies that aim at influencing the decision-making of physicians (Stano, 1993). Physicians should use new knowledge on better treatment modalities but although best practices exist, they hardly spread. Awareness of and even a positive attitude towards a certain, better, way of practice is not enough to change behavior (Lomas, 1989). The decision-making of physicians is influenced by more than research evidence, even if there are clear and concrete recommendations. Professional uncertainty and difference of opinion exist in medical practice and are supposed to cause variation. GPs develop mechanisms for dealing with professional uncertainty and this can cause them to be skeptical towards scientific evidence and more sensitive to peer influences (Hulscher et al., 1999; Hulscher et al., 2001). Therefore, social influences are potentially important for the implementation of guidelines and for changing clinical behavior of GPs. Social influences are sometimes used in implementation projects when networks of peers and colleagues are used. The idea is that, e.g., guidelines are more readily accepted if made and implemented by the profession (Grol, 2001). The mechanisms, however, remain unclear. A review of several implementation projects leads to the recommendation that comprehensive

strategies are conditional to successful implementation (Cretin et al., 2001; Grimshaw et al., 2001; Grol, 2001; Gross and Pujat, 2001). Explaining medical practice variation can be useful in developing evidence-based implementation methods.

Against this background, we expect that there is substantial variation in attitudes and behavior between GPs, but that this variation is smaller for GPs who work together in partnerships. The explanation for this phenomenon might be sought in three directions: the selection of new partners might be directed towards similarities (like seeks like), gradual adaptation to each other within a partnership might result from processes of peer approval, and, finally, the circumstances that are shared by partners might lead to similar behavior.
A first general, descriptive question is asked:

> *'Do GPs working in partnerships show more resemblance in their attitudes and in the treatment they choose than they do to other GPs?'*

A second, explanatory question is:

> *'Are the observed similarities caused by selection, gradual adaptation and/or the shared circumstances GPs work under?'*

Background and hypotheses
Different mechanisms that might explain variation in medical practice have been described in the literature (e.g. Chassin, 1993; Wennberg, 1993; Wennberg and Gittelsohn, 1975; Wennberg and Gittelsohn, 1982). We search for explanations of medical practice variation in differences in characteristics of the social context. Westert (1992) introduced a model of local standards that predicts similarities among colleagues who share one work-environment. In this model, which is based on constraints instead of preferences, differences between groups of doctors are caused by inter-group differences in circumstances, such as bed supply. McClure (1982) also described effects of circumstances on the treatment doctors choose, adaptation was implicated by ideas about education and refresher courses (Chassin, 1993). Selection is found in literature on friendships (Fehr, 1996; Zeggelink, 1993), marriage (Kalmijn, 1998) and situations in which personnel is selected (e.g. Sessa and Taylor, 2000).
The processes concerning partner selection, gradual adaptation and rapid adaptation to shared circumstances are probably the same for other social situations. Probably there is resemblance in the way we select friends, find a

life companion, or select personnel or a partner to work with in a partnership. Of course there are certain differences, but basic mechanisms are the same for all of these examples. One may wonder, therefore, on what basis doctors already in the partnership choose a (new) doctor to work with. This may be personal attraction, which can be compared to the way friends are chosen (Zeggelink, 1993).
For GPs careful selection might be very important in finding a new partner, because it is not so easy to get rid of a partner once s/he has entered the partnership. GPs in the Netherlands are independent professionals and mobility between partnerships is virtually absent. It is expected that GPs search for new partners among GPs they already know, or who have characteristics that can be used in the selection process as proxies for characteristics that signal trustworthy colleagues. In the process of selection implicitly or explicitly a profile will be used that a new GP has to meet. It is common that vacancies for GPs are advertised. It is, however, noteworthy that advertisements only rarely stipulate specific requirements from candidates. Similarities in attitudes and similarities in certain characteristics (university, age, etc.) would give an indication that selection processes play a role. Attitudes can come up for discussion in a job interview, while behavior only becomes visible when people are already working together.

It is not necessarily so that selection causes similarities. A GP can be selected because of a difference with the existing partners, a reason for this can be that the partnership wants to provide for omissions in treatment for their patients. For example, a partnership can choose a woman, to fulfill the need of patients to talk to a female doctor for instance when they have typical female problems (Van den Brink et al., 1994).

Although similar people are attracted to each other (Fehr, 1996), it is also often assumed that friends influence each other and therefore become more alike (Leenders, 1995). Thus, similarities can also be caused by gradual adaptation. When a GP starts to work with other GPs in a partnership it is expected that sooner or later they will adapt to each other. Peer review is increasingly part of normal routine in medical partnerships. Doctors risk losing social approval if their medical performance is criticized by their colleagues (Westert, 1996; Westert and Groenewegen, 1999). When people evaluate their own personal skills or self-image, they rely on role equivalents (Burkhardt, 1994). Consequently, there may be pressures towards adaptation within partnerships. The process of gradual adaptation will lead to similarities in attitudes and behavior and this will be stronger, the longer GPs work together.

Another process that can cause similarities is the adaptation to the circumstances doctors work under. If for instance a hospital is nearby, GPs tend to send their patients to the hospital sooner than GPs at greater distances. Within the same conditions there will consequently be less variation in behavior than between conditions. This process of rapid adaptation through shared circumstances can be distinguished from selection and gradual adaptation to each other by comparing GPs working under the same circumstances. The adaptation to common circumstances will supposedly be much quicker than the adaptation through interaction, peer review and behavioral confirmation. Research into the effects of a change of payment system, for instance, showed a rapid adaptation to the new circumstances (Krasnik et al., 1990).
Three parallel ways of distinguishing between the different explanations, selection, adaptation and working under the same circumstances, can be followed (figure 3.1). In the first a distinction is made between selection on the one hand and adaptation and circumstances on the other. In the second we can distinguish between selection and circumstances as showing no time effects and gradual adaptation in being a long-term process (GPs will show more similarities if they are working in the same partnership for a longer time). In the third, circumstances are separated from selection and adaptation by looking at effects of the conditions the GPs work under (GPs show more similarities when working under the same conditions).

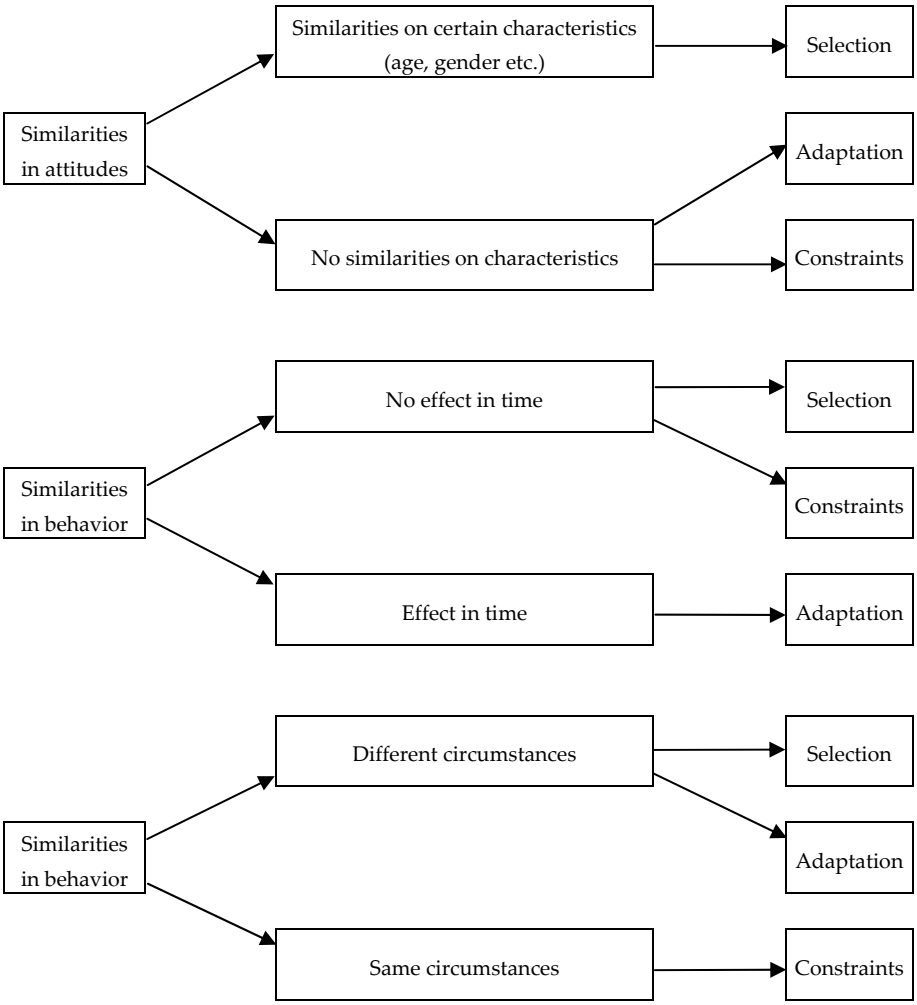
The general hypothesis is that:

Hypothesis 1: *GPs in partnerships are more similar in professional attitudes and behavior to each other than to randomly chosen GPs.*

The explanation of this phenomenon is sought in three directions:

Hypothesis 2.1: *Similarities in attitudes, combined with similarities in characteristics, such as age, university and gender, point towards selection.*
Hypothesis 2.2: *Similarities in behavior combined with an effect for the time GPs re working in the same partnership point towards adaptation.*
Hypothesis 2.3: *Similarities in behavior combined with an effect for the same circumstances point towards similarities caused by circumstances.*

Figure 3.1 Interpretation scheme to distinguish between selection, adaptation and circumstances; three parallel ways



3.2 Data and methods

To find out whether GPs in the same practice show more similarities, we compared their attitudes, self-reported behavior (both collected through a questionnaire), and data on actual behavior, collected during a three-month period in which GPs recorded basic data on each consultation. These are the

dependent variables in our analysis. The data were collected in the First Dutch National Survey of General Practice, from April 1987 till March 1988 in a stratified sample of 161 Dutch GPs in 103 practices, who served 335,000 patients in total (Van der Velden, 1999).

For the analysis of the behavior of GPs in doctor-patient contacts we ruled out as much as possible differences in the demand side as a cause of differences between GPs. This was done by analyzing the data separately for a number of chapters of the international classification of primary care (ICPC-chapters) (Lamberts et al., 1993). Those chapters were included for which enough consultations were present and that showed variation in behavior between GPs (less than 80% of GPs doing the same) on a rather general level (diagnostics, therapeutics, prescription and referral). We selected only first contacts with the GPs in which no co-morbidity was present. GPs who did not fill in all questions on independent variables in the mailed questionnaire (n=12) were left out of the analysis. Solo practices (n=51) were left out of the analysis, because they might form a different group. In all, 96 GPs in 42 practices were selected. Similarities between GPs were analyzed for a large number of dependent variables, describing attitudes, self-reported behavior and behavior reported in doctor-patient contacts (for a detailed description of the variables see Foets et al., 1991).

To find out whether GPs in the same partnership are indeed more similar in attitude and behavior than randomly chosen GPs, we could have used a standard multi-level model, with two levels: doctors at the lower level nested within practices at the higher level. If the between practice variance is significantly larger than zero, there is clustering at the practice level. In other words, doctors working in the same practice show similarities. However, with this standard multi-level model there are two problems: due to a small number of partners the estimates have wide confidence intervals. Secondly, it is not possible to explain similarities. If within practice variation becomes smaller by, for example, adding the age of the GPs as an explanatory variable, this is due to a general effect of age, but it does not necessarily mean that GPs working together show more similarities because of similar ages. Hence, both because of small numbers within practices and because of the specific hypotheses about the mechanisms we had to use a different strategy.

The alternative strategy is based on pairs of GPs. Our basic assumption is that the absolute difference for a certain dependent variable between two GPs is smaller if these GPs work in the same practice, than if they work in

different practices. We constructed two kinds of pairs: all possible pairs of GPs working in the same partnership (actual pairs), and randomly constructed pairs of GPs not working in the same partnership (random pairs). The actual pairs are made within one partnership; two GPs working in the same partnership form an actual pair. For each partnership, all possible actual pairs are made. For the random pairs the second GP of a pair is randomly chosen with replacement from all GPs in the sample, with exclusion of the GPs from the same partnership. Difference scores were computed for all dependent variables, for each pair. E.g. for practice 1 (table 3.1) we have three difference scores, one for the actual pair and two for the random pairs. The variables now measure the extent to which the two GPs differ.

Table 3.1 Pairs of GPs

Practice	**GPs**	**All actual pairs**	**Possible random pairs**
1	GP **1**.1	(**1**.1,**1**.2)	(**1.**1,**2.**1);
	GP **1**.2		(**1.**2,**n.**1);
2	GP **2**.1	(**2**.1,**2**.2); (**2**.1,**2.**n);	(**2.**1,**n.**1);
	GP **2**.2	(**2**.2,**2.**n)	(**2.**2,**1.**2);
	GP **2.**n		(**2**.n,**1**.1)
n	GP **n.**1	(**n**.1,**n**.2); (**n.**1,**n.**n);	(**n.**1,**1.**2);
	GP **n.**2	(**n.**2,**n.**n)	(**n.**2,**2.**2);
	GP **n.**n		(**n**.n,**2**.n)

Because it is plausible that GPs will show consistent similarities in attitudes and behavior for different types of attitudes and behavior, different random pairs for each type of attitude and behavior were made. By doing so, the tests for the different types of attitudes and behavior are independent. This is important, because we were interested in differences between pairs of doctors, and not in consistency of attitudes or behavior of individuals.

To take into account the dependence of pairs that belong to the same practice, multi-level analysis was used. Difference scores computed for each dependent variable for actual and random pairs were analyzed using MLwiN (Snijders and Bosker, 1999). We examined size and direction of the differences between pairs, to see if these differences are smaller within actual pairs than within random pairs. We did not develop any hypotheses about the actual dependent variables separately, except for the distinction between attitude and behavior. Hence, we will not pay attention to or try to

interpret specific differences. We only report on the number of times we found smaller differences for actual pairs than for random pairs.
Since it is assumed that actual pairs differ from random pairs, we should explicitly model this. For both the actual and the random pairs we have separate means and separate variances at the two levels.

For the equation we refer to appendix 3.A.

Similarities
To test the hypothesis whether actual pairs show more similarities, we will only examine the fixed effects (β_0 and β_1, appendix 3.A), for there are not enough pairs to estimate meaningful variances. It is expected that β_0, the mean difference for the dependent variable within actual pairs, approaches zero and β_1, the mean difference for the dependent variable within random pairs, is significantly greater than zero. Consequently, β_1 will be significantly greater than β_0.

Explanations
For testing the mechanisms selection, adaptation and circumstances, differences between members of the pairs are examined to find out whether the introduction of explanatory variables decreases differences in attitude or behavior between pairs. Because of the hypothesis that differences within actual pairs are smaller than within random pairs, we tested one-tailed at a 95% confidence interval (corresponding to chi^2=2.71). The explanatory variables are centered around the overall mean of the pairs for that variable, separately for actual and random pairs. The equation is presented in appendix 3.B.

Selection
For variables indicating selection, a difference is measured for both the actual and the random pairs. If selection causes similarities, two conditions must apply. First, the difference for variables indicating selection, for example age, within actual pairs will be smaller than within random pairs. Secondly, when adding the indicators of selection, the differences in the dependent variables between actual and random pairs will decrease (absolute β_0- β_1, appendix 3.B). The effect of a difference in an indicator will be more or less the same for actual and random pairs. We used two variables to indicate selection: age and gender of GPs.

Adaptation
We only used one variable indicating adaptation, viz. the years GPs work together. This variable was defined only for actual pairs. Random pairs have not worked together and hence the number of years they have worked together has been fixed at zero. Adding this variable will decrease the difference within actual pairs (β_0, appendix 3.B). This difference will be smaller when GPs have been working together for a longer period of time, therefore, the coefficient (β_4, appendix 3.B) will show a negative sign.

Circumstances
Differences in circumstances only exist within random pairs. Actual pairs work under the same circumstances by definition. Therefore, the difference for actual pairs is fixed at zero. Variables indicating circumstances will decrease differences within random pairs (β_1, appendix 3.B). Furthermore, we will look at the sign of the explanatory variable coefficient (β_5, appendix 3.B), which is expected to be positive.

We will only look at those parameters, that are relevant for testing the hypotheses, as indicated in table 3.2.

All variables indicating selection, adaptation and circumstances were entered in one analysis. Hence the effect of each variable was in combination with the other variables. The result of the analysis in which explanatory variables are included may be compared with the empty model. Including the explanatory variables can improve the model (smaller differences within pairs) or worsen it (larger differences within pairs).

Table 3.2 Testing explanations, expected effects

Parameter	**Selection variables**	**Adaptation variables**	**Circumstances variables**
Effect actual pairs (β_0)	decrease	decrease	-
Effect random pairs (β_1)	-	-	decrease
Effect adaptation (β_4)	-	negative	-
Effect circumstances (β_5)	-	-	positive
Difference expl vars actual pairs ($\Delta(Sap)_{ij}$)	$<(\Delta((Srp)_{ij})$	-	-
Difference between actual and random pairs (absolute β_0- β_1)	decrease	-	-

3.3 Results I: Similarities within partnerships

Table 3.3 presents the mean differences between actual and random pairs concerning the attitudes, self-reported behavior and behavior reported in doctor-patient contacts. For attitudes, differences within actual pairs are smaller than within random pairs for nine out of fourteen variables. For five variables, the differences within the random pairs are smaller. Three significant values are found. There is no convincing evidence for the hypothesis that GPs working in the same partnership (actual pairs) show more similarities in attitudes than GPs not working in the same partnership (random pairs).

In the case of self-reported behavior, GPs working in the same partnership show for all three variables more similarities. Two of these differences are statistically significant, which is in line with the hypothesis.

For behavior reported in doctor-patient contacts, it was found that the duration of a contact in general and the time a consulting hour contact takes show significant differences. Actual pairs are more alike than random pairs.

Owing to the number of separate items of behavior, analyzed for each of seven ICPC-chapters, a large number of tests have been performed. The results of those tests are reported separately in table 3.4. A total of 71 items were tested and 24 statistically significant values were observed. This is more than could be expected on the basis of chance (p=0.05), on which basis we could expect only four statistically significant values.

We may conclude that in the majority of cases differences within actual pairs are smaller than within random pairs, indicating that GPs in the same partnership are more similar than GPs not working in the same partnership. Differences indicating that actual pairs are less alike than random pairs with a magnitude approaching significance were not observed.

Table 3.3 Differences between actual and random pairs of GPs in attitudes, self-reported and behavior reported in doctor-patient contacts

Attitudes	**Mean difference between pairs**	**Chi square**
Uncertainty	0.30	0.13
Risks	**-0.16**	**2.80**
Task perception:		
general	0.06	0.00
diagnosis	-0.05	1.43
therapy	**-0.07**	**4.25**
somatic	0.07	0.51
psycho-social	-0.05	0.60
Feeling of competence:		
somatic	0.09	1.03
psycho-social	-0.09	1.77
Democratic attitude	**-0.10**	**5.77**
Social attitude	-0.08	1.78
Locus of control	-0.04	0.86
Psycho-social factors	0.00	0.00
Prescription	-0.02	0.09
Self-reported behavior:		
medical techniques used	**-0.29**	**11.90**
handling work style elements	**-0.19**	**3.92**
task profile	0.09	0.82
Behavior reported In doctor-patients contacts:		
time a contact takes	**-0.45**	**5.64**
time a consulting hour contact takes	**-0.59**	**6.75**

'bold printed' means for which variables the differences are significant
a positive difference means that random pairs are more alike than actual pairs

Table 3.4 Summary of differences between actual and random pairs of GPs in actions for certain complaints

Variable	Respira-tory	Musculo-skeletal	Blood, blood-forming	Circula-tory	Diges-tive	Eye	Ear
Diagnosis	**-0.03**	–	-0.01	0.01	0.00	-0.01	-0.01
Internal diagnosis	**-0.05**	–	-0.01	0.00	0.00	0.00	0.00
External diagnosis	–	–	0.06	–	–	–	–
Clinical diagnosis	**-0.03**	–	-0.02	0.01	0.01	**-0.02**	-0.01
Blood test	–	–	-0.03	–	–	–	–
Haematology	–	–	0.03	–	–	–	–
Treatment	-0.06	-0.04	**-0.12**	**-0.05**	-0.05	0.00	**-0.05**
Therapeutic counseling	-0.03	-0.06	-0.05	-0.01	-0.03	0.00	-0.02
Information	-0.04	-0.03	-0.04	-0.04	**-0.11**	**-0.09**	**-0.07**
Other investigation	–	–	–	–	–	-0.01	–
Medical techniques	–	–	–	–	–	–	**-0.05**
Wait and see	–	-0.02	0.00	–	-0.01	–	–
Rules	–	–	–	–	0.00	–	–
Medication	**-0.03**	**-0.03**	**-0.09**	0.01	**-0.04**	-0.03	-0.02
Sys. antibiotic	**-0.04**	–	-0.01	–	–	–	–
Analgetic	**0.05**	–	–	–	–	–	–
Cough/cold-remedy	**-0.04**	–	–	–	–	–	–
Nasal use	**-0.03**	–	–	–	–	0.00	**-0.04**
Anti-rheumatic	–	**-0.04**	–	–	–	–	–
Other medication	–	–	-0.03	0.00	-0.03	**-0.03**	-0.01
Referral within:							
primary care	–	**-0.03**	–	–	–	–	–
secondary care	–	–	–	0.00	–	–	–
Referral cutting specialism	–	–	–	–	–	**-0.03**	–

'bold printed' means for which variables the differences are significant

'–' means that it was not in the analysis

3.4 Results II: Explanations

Selection

The mean difference in age for random pairs is less than the mean difference in age for actual pairs, which is contrary to our expectation. Actual and random pairs show approximately the same distribution of exclusively male or female pairs. Adding the variables indicating selection to the model decreases differences within actual pairs (β_0) for 39 out of 90 variables (43%).

The differences between pairs decrease (absolute β_0- β_1) for 40 variables (44%). This implicates that differences in age and gender do relate to differences in attitude and behavior. However, there is no selection based on similar age or gender that makes actual pairs more alike than random pairs.

Adaptation
For 30 out of 90 variables (33%) the difference within actual pairs (β_0) decreases when explanatory variables are included. For 46 out of 90 variables (51%) the coefficient (β_4) shows a negative sign; differences are smaller when GPs are working together longer.

Circumstances
We took into account four variables for circumstances. For 90 dependent variables we added the four explanatory variables. For 46 out of 90 variables (51%) the differences within random pairs (β_1, appendix 3.B) decrease when explanatory variables are included. We found in 182 out of 360 cases (51%) a positive sign for the coefficients of the four variables. This means that differences are larger when circumstances differ.

Table 3.5 Summary of the effects for indicators on selection, adaptation and circumstances for 90 variables, based on sign, not significance

Parameter	Selection variables	Adaptation variables	Circumstances variables
Effect actual pairs (B_0)	43% decrease	33% decrease	-
Effect random pairs (B_1)	-	-	51% decrease
Effect adaptation (B_4)	-	51% negative	-
Effect circumstances (B_5, B_6, B_7, B_8)	-	-	51% positive
Difference expl vars actual pairs ($D(Sap)_{ij}$)	9.5 (sd 7.3)	-	-
Difference expl vars random pairs ($D(Srp)_{ij}$)	6.7 (sd 6.3)	-	-
Difference between actual and random pairs (absolute B_0-B_1)	44% decrease	-	-

3.5 Conclusions and discussion

In this study we found more similarities between GPs sharing a work-environment than between GPs not sharing a work-environment. Similarities mainly concerned self-reported behavior and behavior reported in doctor-patient contacts. Less similarities were found in attitudes. This gave a first clue to an explanation of similarities based on circumstances. If

selection and adaptation were the main mechanisms causing similarities, we would have expected attitudes to be more similar within partnerships.

Differences between GPs are smaller when variables indicating selection, adaptation and circumstances are included in the statistical analysis. Overall, for 51 out of 90 variables (57%) the differences between actual and random pairs (absolute β_0- β_1) decrease when all explanatory variables are included. Furthermore, it can be concluded that most indications point towards circumstances and to a lesser extent towards adaptation.
The implication of this study is that medical practice variations are not merely individual differences in preferred practice style, but patterned by social processes in partnerships and local circumstances. This knowledge is important to enhance the prospects of effective health care policies (Stano, 1993). If the mechanisms of selection, adaptation and circumstances cause GPs to be alike, this has implications for the way guidelines should be implemented. Just making GPs aware of the fact that guidelines exist will not work if they share a practice with GPs confirming each others behavior: why would they use the guidelines? Knowing what causes GPs to act the way they do provides opportunities to find effective incentives for influencing their medical behavior. Davis et al. (2000) concluded in their study on physician practice styles that each explanation of variation has different implications for the development and successful implementation of clinical guidelines. It is important to find explanations for variation and to recognize that these can be a cause of persistent behavior of GPs. Our study suggests that variation between practices is primarily caused by differences in circumstances. For the implementation of guidelines, these circumstances should be taken into account. Furthermore, adaptation seems to play a role and this knowledge can be used for implementation e.g. by starting peer groups. The limited influence of adaptation found in this study would implicate that it is not enough to send one GP from a partnership to a training course, hoping that information will also diffuse to colleagues.

In this study no evidence was found that selection based on similar age and gender forms an explanation for similarities between GPs sharing a work-environment. This finding does not necessarily imply that selection does not play a role in explaining similarities within partnerships. The results of this study point towards the importance of similarities based on circumstances and to a lesser extent on adaptation, at least for the GPs in the analyses. More research should be done to provide a broader test of the mechanism of selection.

In this study a limited number of indicators for selection, adaptation and circumstances were used, based on availability in the data set. For selection only age and gender could be used. We assumed that similarities in age and gender of partners would implicate selection and that similarities in attitudes and behavior could be explained by these variables. There is always implicit or explicit selection when GPs try to find a partner. Our analysis derives selection from characteristics being the same, but actual selection processes could also be based on opposite characteristics, e.g. when an older GP specifically looks for a younger partner. The university where GPs studied, as mentioned as an indication for selection in the hypothesis, was not taken into account, because the data on this variable were not appropriate for use in this study. Based on the variables used, we concluded that selection does not play a role, first of all because these variables do not cluster within partnerships. However, there still could be other explicit as well as implicit characteristics that are important in explaining similarities within practices.
Data used in this study are rather old; they were collected in 1987/1988. This would have been a problem in a purely descriptive study. However, mechanisms causing more similarities within partnerships than between partnerships will not have changed and could be tested with these data. There is no reason to assume that selection and adaptation processes are strongly influenced by changes over time. However, it could be argued that, dependent on specific circumstances, one is more important than another. Secondary data were used and the sample size was fixed. It should be mentioned that the sample was not very large, only 96 GPs, and to get better results more GPs should be included.

Furthermore, we tested at the level of pairs of GPs and not at the level of individual GPs. One outlying GP in a partnership of three causes two outlying relationships of a total of three relationships. There is thus an underestimation of similarities within the partnership and our hypothesis is tested more strictly than when it would have been possible to study individual GPs instead of pairs.
Processes of selection, adaptation and circumstances described in this study are general for people working and living together. Selection is important because a colleague should be someone with whom it is pleasant to work and with whom there is not much to argue about. Probably there are more occupations that are comparable to the situation of GPs. We could think of lawyers and other occupations in which client situations differ, every client is unique, but still there should be uniformity in the 'product'. Unlike GPs, lawyers and auditors are trained for a long time before they become a

partner (Lazega, 2001). The amount of time it usually takes to become a full partner could cause a difference between professions in the importance of selection and adaptation. It might be true that selection plays an important role in explaining similarities within partnerships and differences between partnerships. No matter whether these are medical partnerships, partnerships of lawyers or others. Members of the partnership select their partners in order to minimize future problems in cooperation.

For a number of reasons, the analyses presented here are not definite, it could be questioned whether the variables used really implicate the mechanism we presumed they would implicate (validity) and whether the number of different operationalizations is enough to give a good guess about the explanation of similarities. However, a first step has been taken by showing the existence of similarities between partners and by specifying three possible mechanisms to explain these similarities. A further step requires a more in-depth study of the processes of selection and mutual influence that take place in teams of health care providers. The study of these mechanisms contributes to our understanding of problems in the implementation of guidelines.

Appendix 3.A Equation 1 *Basic model*

$\Delta y_{ij} = \beta_{0ij}\Delta(y_{ap})_{0ij} + \beta_{1ij}\Delta(y_{rp})_{1ij}$

$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$

$\beta_{1ij} = \beta_1 + u_{1j} + e_{1ij}$

Δy_{ij} = absolute difference in the value of the dependent variable

i = pairs

j = practices

$\Delta(y_{ap})$ = intercept variable for actual pairs

$\Delta(y_{rp})$ = intercept variable for random pairs

β_{0ij} = mean and variance parameters for actual pairs

β_0 = mean for actual pairs

u_{0j} = level 2 (practices) variance for actual pairs

e_{0ij} = residual variance for actual pairs

β_{1ij} = mean and variance parameters for random pairs

β_1 = mean for random pairs

u_{1j} = level 2 (practices) variance for random pairs

e_{1ij} = residual variance for random pairs

Appendix 3.B Equation 2 *Model with explanatory variables*

$$\Delta y_{ij} = \beta_{0ij}\Delta(y_{ap})_{0ij} + \beta_{1ij}\Delta(y_{rp})_{1ij} + \sum_{h=1}^{a}\beta_{2ij}\Delta(S_{ap})_{hij} + \sum_{h=1}^{b}\beta_{3ij}\Delta(S_{rp})_{hij} + \sum_{h=1}^{c}\beta_{4ij}\Delta(A_{ap})_{hij} + \sum_{h=1}^{d}\beta_{5ij}\Delta(C_{rp})_{hij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = \beta_1 + u_{1j} + e_{1ij}$$

Δy_{ij} = absolute difference in the value of the dependent variable

i = pairs

j = practices

$\Delta(y_{ap})_{0ij}$ = intercept variable for actual pairs

$\Delta(y_{rp})_{1ij}$ = intercept variable for random pairs

β_{0ij} = conditional mean and variance parameters for actual pairs

β_0 = conditional mean for actual pairs

u_{0j} = level 2 (practices) variance for actual pairs

e_{0ij} = residual variance for actual pairs

β_{1ij} = conditional mean and variance parameters for random pairs

β_1 = conditional mean for random pairs

u_{1j} = level 2 (practices) variance for random pairs

e_{1ij} = residual variance for random pairs

$\Delta(S_{ap})_{hij}$ = absolute difference of variables (h = 1 till h = a) indicating selection for actual pairs

$\Delta(S_{rp})_{hij}$ = absolute difference of variables (h = 1 till h = b) indicating selection for random pairs

$\Delta(A_{ap})_{hij}$ = absolute difference of variables (h = 1 till h = c) indicating adaptation for actual pairs

$\Delta(C_{rp})_{hij}$ = absolute difference of variables (h = 1 till h = d) indicating circumstances for random pairs

$\beta_{2ij}, \beta_{3ij}, \beta_{4ij}, \beta_{5ij}$ = regression coefficient for variables

h = number of variables

a = 2

b = 2

c = 1

d = 4

4

Variation in hospital length of stay

Do physicians adapt their length of stay decisions to what is usual in the hospital where they work?

This article was published as:
De Jong JD, Westert GP, Lagoe R, Groenewegen PP. Variation in hospital length of stay: do physicians adapt their length of stay decisions to what is usual in the hospital where they work? *Health Serv Res,* 2006; 41(2):374-94

Abstract

To test the hypothesis that physicians who work in different hospitals adapt their length of stay decisions to what is usual in the hospital under consideration.

Secondary data were used, originating from the Statewide Planning and Research Cooperative System (SPARCS). SPARCS is a major management tool assisting hospitals, agencies, and health care organizations with decision-making in relation to financial planning and monitoring of inpatient and ambulatory surgery services and costs in New York State.

Data on length of stay for surgical interventions and medical conditions (a total of seven DRGs) were studied, to find out whether there is more variation between than within hospitals. 1999, 2000 and 2001 data from all hospitals in New York State were used. The study examined physicians practicing in one hospital and physicians practicing in more than one hospital, to determine whether average length of stay differs according to the hospital of practice. Multilevel models were used to determine variation between and within hospitals. A t-test was used to test whether length of stay for patients of each multihospital physician differed from the average length of stay in each of the two hospitals.

There is significantly ($p<0.05$) more variation between than within hospitals in most of the study populations. Physicians working in two hospitals had patient lengths of stay comparable to the usual practice in the hospital where the procedure was performed. The proportion of physicians working in one hospital did not have a consistent effect for all DRGs on the variation within hospitals.

Physicians adapt to their colleagues or to the managerial demands of the particular hospital in which they work. The hospital and broader work environment should be taken into account when developing effective interventions to reduce variation in medical practice.

4.1 Introduction

A persistent finding in health services research is that hospital utilization varies widely (Wennberg, 1999; Ashton et al., 1999; Paul-Shaheen et al., 1987). These variations have been observed between geographic areas, hospitals and physicians. The variation within these units of analysis has been found to be smaller than between the units, for different types of services, numbers of admissions and length of stay (Arndt et al., 1995; O'Connor et al., 1999; Read et al., 1983; Westert et al., 1993; Wennberg and Gittelsohn, 1982).

Several explanations have been sought for the variation between and similarities within units, a summary can be found in table 4.1. Wennberg and Gittelsohn (1975) suggested that an explanation lies with the physicians themselves. Firstly, they theorized that professional uncertainty explains whether the specific procedure or diagnosis will have high or low variation. Secondly, they hypothesized that the judgment and preferences of groups of physicians give rise to a unique pattern over time, which has been termed a 'surgical signature'. Chassin (1993) suggested that variation is caused by a difference in the prevalence of physicians who are enthusiastic about certain procedures. The 'surgical signature' and enthusiasm hypotheses assume that physicians have a preference for certain procedures, but the behavioral mechanisms that produce different practice styles remain unclear.

Westert and Groenewegen (1999) offered an alternative to this preference-centered approach, emphasizing incentives and environmental conditions that influence the behavior of physicians by providing opportunities and constraints. Westert (1992) applied this approach in a model of local standards that predicts similarities among physicians who share a common work environment and thus a social system and similar constraints. We test whether variations in medical practice are indeed related to the hospital in which physicians practice. This would imply that variation within hospitals is small compared to variation between hospitals for physicians treating similar patients. A second implication would be that physicians working in more than one hospital conform to the usual practice of each hospital (Westert, 1992). This implication cannot be deduced from a preference-centered approach, since the preferences of an individual would not change when working in another hospital.

Indications were found by Westert et al. (1993), that multihospital physicians in the Netherlands have a patient length of stay close to the usual practice in the hospital where the surgery was performed. Griffiths et al. (1979), using data from the British National Health Service, also found that average postoperative stays were similar among physicians who practice in the same

hospitals, while the average between hospitals was significantly different for physicians who practice in more than one hospital.
In the present study, we tested these implications with data from the USA, where it is quite common for physicians to work in more than one hospital. As a result, the implications can be tested more reliably than was done by Westert et al. (1993) with 23 physicians in five hospitals (only four working in more than one hospital) and Griffiths (1979) with nine physicians in eight hospitals (six working in more than one hospital). Furthermore, we used a different methodology. We took length of stay, which is a well-defined and important indicator in hospital management, as the outcome variable.

Table 4.1 Summary of explanations for variation

Type of explanation	Content	Hypotheses	Study citations
Preference centered	Practice styles are caused by differences in preference for certain procedures	Practice style remains the same when social context is changed	e.g. Wennberg and Gittelsohn (1975) Chassin (1993)
Constraint centered	Practice styles are caused by differences in characteristics of the (social) context, providing opportunities and constraints	Similarities of practice styles within the same social context, practice style changes with changes in the social context	e.g. Westert (1992) Westert and Groenewegen (1999)

Background and hypotheses
In general, people show similar behavior within work environments; colleagues show similarities (Denton, 1999); attitudes and beliefs of individuals change towards the group norm (Dambrun et al., 2002). *"...people both shape and are shaped by social networks"* (Pescosolido, 1992). Similarities as a consequence of attitude and belief change is a gradual process. Other explanations for similar behavior are selection and adaptation to the circumstances under which physicians work (chapter 3) (De Jong et al., 2003). Similarities as a consequence of the latter mechanisms is much quicker than gradual adaptation. Physicians will adapt to the usual practice in the hospital in which they treat patients, in order to avoid being criticized (Westert, 1992; Eddy, 1984). Physicians expected to show similar behavior will be selected by the hospital management or by colleagues. If this strategy is followed, local standards of medical care come into being within medical teams that share the same work environment (Westert and Groenewegen, 1999). Whether local standards come into being or not, might depend on the extent to which the physicians working in a hospital form a social group,

and are able to develop into a normative community. We assume that the chances of developing local standards in a hospital are lower when more physicians in this hospital work in more than one hospital. If there is a small proportion of physicians who practice in this hospital only, average length of stay is expected to be less standardized.
In this study we first examined whether there is less variation between physicians working in the same hospital than between different hospitals. Secondly, we examined whether physicians working in more than one hospital practice differently in different hospitals. Thirdly, we tested the effect of the proportion of physicians working in only one hospital on the variation within hospitals.
The results of this study have implications for the explanation of variation between hospitals and similarities within hospitals. If physicians working in different hospitals behave according to the local standards in the hospitals where the procedures were performed and thus show intra-doctor variation, selection, based on characteristics that relate to similar behavior, is not an explanation for more similarities within hospitals than between them. This is an important test to find out whether selection holds true, because the physicians are the same. If selection based on similarities is an explanation of similar behavior, a physician would not do different things in different hospitals.
If variations between physicians are related to the hospital, as we argued:

Hypothesis 1: *The variation between physicians working in the same hospital is less than the variation between hospitals.*

Hypothesis 2: *Physicians working in different hospitals in which the average lengths of stay differ, will choose lengths of stay similar to the average in the hospital in which the patient is treated.*

Normative processes cause similarities within hospitals. The influence of colleagues or of the hospital management will be more important when the average length of stay within a hospital is more standardized; deviation from the average length of stay is more obvious when there is less variation. The distribution will show less variability for those physicians who only practice in that hospital, since they are more dependent on each other and will avoid being criticized. Therefore:

Hypothesis 3: *The higher the proportion of physicians working in only one hospital, the less variation in the length of stay there will be between physicians in that hospital.*

4.2 Data and methods

Description of the data

Data were obtained from the New York Statewide Planning and Research Cooperative System (SPARCS), which is a comprehensive patient data system established as a result of cooperation between the health care industry and government. SPARCS is a major management tool assisting hospitals, agencies, and health care organizations with decision-making in relation to financial planning and monitoring of inpatient and ambulatory surgery services and costs in New York State. Several research articles have been published that are based on SPARCS data (Pasley et al., 1995; Westert and Lagoe, 1995; Hainsworth et al., 1997; Lagoe, 1998; Murphy et al., 1999; Noetscher et al., 1999; Lagoe et al., 2001; Noetscher et al., 2001; De Jong et al., 2004).

We used 1999, 2000 and 2001 SPARCS-data to study seven DRGs: two medical (DRGs 88 and 127), one surgical (DRG 209) and four obstetrical (DRGs 358, 359, 370 and 371). Cases for which no physician was known were excluded (1.3% of all cases). Only patients above the age of twenty were included and patients with extremely long stays (defined as the average length of stay plus 1.96 times the standard deviation) were excluded. This involved a minimum of 0.86% and a maximum of 3.72% of cases per procedure. In 2001, 70% of the physicians worked in only one hospital and 30% worked in more than one hospital. About 93% of all physicians work in one or two hospitals, 7% work in more than two hospitals. This is about the same for all three years of data. The study populations for all three years are summarized in table 4.2.

Table 4.2 Study population: number of patients, physicians, hospitals and percentage of excluded cases per DRG 1999, 2000, 2001

DRG	Diagnosis/procedure	Number of discharges			Number of physicians			Number of hospitals			% cases excluded		
		1999	2000	2001	1999	2000	2001	1999	2000	2001	1999	2000	2001
88	Chronic Obstructive Pulmonary Disease	38,424	36,478	34,401	9,125	8,779	8,476	237	234	228	0.86	3.46	2.62
127	Congestive Heart Failure	62,682	62,599	59,763	11,282	11,285	11,048	233	234	229	3.21	3.72	3.00
209	Major joint procedures	28,426	29,827	32,016	1,267	1,238	1,240	210	207	204	2.58	3.04	3.13
358	Hysterectomy with complications	8,408	8,412	8,137	2,236	2,154	2,104	215	208	207	1.87	3.41	3.58
359	Hysterectomy without complications	21,962	22,926	22,604	2,630	2,608	2,520	216	215	212	2.46	2.09	1.78
370	Cesarean section with complications	11,731	12,125	11,954	2,217	2,240	2,183	164	161	159	2.61	2.54	2.42
371	Cesarean section without complications	39,844	42,980	42,770	2,602	2,654	2,571	163	161	159	1.31	1.25	1.22

Methods

Data for each of the seven diagnoses and procedures contained in the Diagnosis Related Groups were separately evaluated with respect to three hypotheses.

Hypothesis 1: Variation between and within hospitals

The initial component of the study attempted to identify the most significant sources of variation in physician practice with respect to lengths of stay. This was accomplished by comparing variations in lengths of stay between hospitals with variations in stay within hospitals, for all physicians. It can be argued that most of the variation in hospital utilization is caused by patient characteristics, like severity of illness, and case-mix (Powell et al., 2004). Some hospitals might treat patients with a higher severity of illness. It is possible that physicians can choose the hospital where to treat their patients on the basis of the severity of illness, or on other unmeasured characteristics. In our study we excluded as much as possible that the variation found is related to severity of illness, and case-mix. The comparison involved the proportion of total variation in stays that did not result from patient age, gender (male, female), race (white, black, native, Asian, other, unknown), or severity of illness (index from one to four). Hospital severity of illness was identified using the All Patients Refined System developed by the 3M Corporation of Wallingford, Connecticut, USA (Averill et al., 2002). The APR System is used to identify severity in hospitals throughout the United States and Europe. This system identifies severity of illness by assigning a level of illness to each individual ICD-9 medical records code. The levels of illness range from 1-Minor, 2-Moderate, 3-Major, 4-Extreme. For each patient, the principal diagnosis and each secondary diagnosis receives one of these levels of severity.

In order to identify a level of severity for an individual patient, the All Patients Refined System uses a formula to develop a single level based on the levels for each diagnosis. All variables were centered around their means, so the model has an interpretable meaning (Snijders and Bosker, 1999).

We also measured the extent to which physicians in the same hospital choose similar lengths of stay. This is done by calculating the intraclass correlation (ICC), defined as the variation between hospitals divided by the total variation in length of stay, and it was accomplished using multi-level analysis (Snijders and Bosker, 1999). Multi-level analysis is an extension of ordinary least squares regression analysis. Ordinary least squares regression analysis allows us, for example, to estimate the relationship between length of stay and age in our study population, assuming that there is no physician

or hospital effect in addition to the characteristics of the patient. With multilevel analysis, however, total variation in length of stay is separated into three parts: a part due to differences between patients, a part due to differences between physicians and a part due to differences between hospitals (Leyland and Groenewegen, 2003). The model is described in appendix 4.A.
We did not adjust for hospital differences, because this is part of the context and thus subject of the study: the objective was not to explain the differences in length of stay, but to explain patterns of variation.

Hypothesis 2: Variation in physicians working in two hospitals
The second component of the analysis evaluated differences in practice patterns in physicians who practiced in more than one hospital. Inclusion criteria were developed for physicians and hospitals.
Physicians were included for the analysis if they discharged patients from either one or two hospitals. These situations accounted for a substantial majority of the physicians in the data (93%). All physicians who practiced at more than two hospitals were excluded (7%). Those who practiced at only one hospital formed a comparison group; those who practiced at two hospitals were the focus of the study. In order to permit meaningful comparison of physician lengths of stay per DRG, a hospital was included if it had at least two physicians who practiced only at this hospital as well as a third physician who practiced at another hospital as well for a specific DRG.
The analysis for this component of the study involved comparison of physician lengths of stay with a 'local standard' or usual practice. The local standard for a hospital was defined as the mean length of stay for physicians working only in this hospital. In order to provide a baseline for subsequent analyses, local standards for hospitals in the study were compared, to determine which of these facilities had lengths of stay that differed by at least 0.5 day.
One half day was chosen since it is associated with savings in expenses, as well as with additional revenue if another patient can be admitted to the vacant bed. In 2001, the difference in mean lengths of stay between the two hospitals where the physician practiced was less than 0.5 days for 48% of all physicians who worked in different hospitals. The focus of the analysis was on comparison of the average length of stay for each multihospital physician with the local standard. A t-test (Hays, 1973) was used to determine whether the mean length of stay for patients of each multihospital physician differed from the local standard at each of the two hospitals. When this test was non-significant, the physician's practice pattern was assumed to be similar to the local standard at that hospital (measured as the mean length of stay

generated by the physicians who only work in that hospital). SPSS was used for this t-test. The means of the hospital and the multihospital physicians are compared in this procedure, which computes the differences between the multihospital physicians' average and the hospital's average in each case, and tests whether the average difference is other than 0.
This component of the analysis also compared the quantitative differences between local standard lengths of stay and differences between multihospital physician lengths of stay between facilities. This analysis determined whether these differences in stays pointed in the same direction. A sign test was used in this comparison (Wonnacott and Wonnacott, 1990). Although a rather weak test, it is appropriate, for we are only interested in the direction of the difference. First, local standards were subtracted and secondly physician stays were subtracted for the same pairs of hospitals. If the signs (positive or negative) of the differences were the same, it was concluded that the physician adjusted patients' lengths of stay were in the direction of the usual practice pattern at the hospital.

Hypothesis 3: Influence of the proportion of physicians practicing in only one hospital on total variation in length of stay within a hospital
The potential impact of lengths of stay generated by physicians working in only one hospital on total practice variation within a hospital was also evaluated. This was measured using a covariance estimated with multi-level analysis, measuring that part of the variation that can be attributed to the proportion of physicians practicing in only one hospital. The exact model is described in appendix 4.B. The percentage of physicians working in one hospital was centered around 99%. It was expected that the variation within hospitals would decrease when the proportion of physicians working in only one hospital increased.

4.3 Results

Results of the study for hypothesis 1 are summarized in table 4.3. This information relates to that proportion of the variation in stays that was not associated with age, gender, race, or severity of illness.

The data in table 4.3 indicate there was significantly more variation in lengths of stay between hospitals than within hospitals in most of the study populations. These populations included patients treated for major joint procedures, cesarean section with and without complications in all three years of the data, as well as those with chronic obstructive pulmonary

disease in one of the three years of the data. Variation within hospitals was significantly greater than variation between hospitals in the case of patients who received hysterectomies without complications in three years of data, and for hysterectomies with complications in two years of data. The ICC tells us how strong the resemblance is within hospitals (Leyland and Groenewegen, 2003).

On the basis of these analyses, hypothesis 1 was confirmed. The finding that there is more variation between than within hospitals was the basis for the second component of the analysis.

Results of the study for hypothesis 2 are summarized in tables 4.3 and 4.4. This information relates to lengths of stay for physicians working in two hospitals and the relationships between their practice patterns and local standards, defined as the mean length of stay for physicians working in one hospital.

For almost all of the study populations, including both medical and surgical patients, the t-tests indicated that lengths of stay for multihospital physicians (between 71 and 1,009 physicians, depending on the DRG) did not differ significantly from the local standard in the three years (table 4.4).

It should be emphasized that these comparisons involved the stays chosen by multihospital physicians and the local standards at each location of practice (hospital A and hospital B). The only significant difference between length of stay chosen by multihospital physicians and local standards relates to a single year (1999) and involves major joint procedures and hysterectomy without complications.

The length of stay for hysterectomy without complications is not related to the hospital, an indication that is based on the observation that variation is lower between hospitals than between physicians. The last component of this part of the analysis compared quantitative differences between local standard hospital lengths of stay and differences between multihospital physicians' lengths of stay for the same hospitals. Results of this sign test are summarized in table 4.4.

This information demonstrates that differences between hospital stays pointed in the same direction as differences between physician stays in a majority of the cases for all seven Diagnosis Related Groups in each of the three years of study data. The proportion of physicians for whom the stays pointed in the same direction varied from 51 to 86%.

Table 4.3 Average lengths of stay, corrected for age, sex, race and severity of illness, 95% confidence intervals for the average per hospital, Intra Class Correlation (ICC), and variance components per DRG for 1999, 2000, and 2001

		1999				2000				2001			
DRG	**Diagnosis/ procedure**	**Average (s. error)**	**Hospital 95% conf. int.**	**ICC**	**Variance components**	**Average (s. error)**	**Hospital 95% conf. int.**	**ICC**	**Variance components**	**Average (s.error)**	**Hospital 95% conf. int.**	**ICC**	**Variance components**
88	Chronic obstructive pulmonary disease	6.01 (0.08)	3.66-8.36	0.09	**1.44 (0.15)*** **1.09 (0.07)**** 13.85 (0.11)***	5.48 (0.06)	3.69-7.26	0.08	0.83 (0.09)* 0.69 (0.04)** 8.77 (0.07)***	5.37 (0.07)	3.49-7.24	0.09	0.92 (0.10)* 0.74 (0.05)** 8.84 (0.07)***
127	Congestive heart failure	5.78 (0.06)	4.04-7.53	0.06	0.79 (0.08) 0.77 (0.04) 11.91 (0.07)	5.65 (0.06)	3.91-7.39	0.06	0.79 (0.08) 0.85 (0.04) 10.82 (0.07)	5.63 (0.06)	3.89-7.38	0.06	0.79 (0.09) 0.86 (0.04) 11.36 (0.07)
209	Major joint procedures	6.42 (0.10)	3.96-8.88	0.23	**1.57 (0.18)** **0.65 (0.04)** 4.60 (0.04)	6.13 (0.08)	3.97-8.28	0.21	**1.21 (0.14)** **0.59 (0.04)** 3.92 (0.03)	5.80 (0.08)	3.75-7.86	0.22	**1.10 (0.13)** **0.53 (0.04)** 3.42 (0.03)
358	Hysterectomy with complications	4.12 (0.04)	3.32-4.91	0.04	**0.17 (0.03)** **0.34 (0.04)** 3.36 (0.06)	3.83 (0.04)	3.13-4.53	0.05	**0.13 (0.02)** **0.24 (0.03)** 2.20 (0.04)	3.74 (0.05)	2.73-4.75	0.10	0.27 (0.04) 0.23 (0.03) 2.26 (0.04)
359	Hysterectomy without complications	2.90 (0.02)	2.34-3.46	0.09	**0.08 (0.01)** **0.16 (0.01)** 0.69 (0.01)	2.82 (0.02)	2.30-3.33	0.08	**0.07 (0.01)** **0.14 (0.01)** 0.67 (0.01)	2.75 (0.02)	2.17-3.34	0.10	**0.09 (0.01)** **0.14 (0.01)** 0.67 (0.01)
370	Cesarean section with complications	4.62 (0.04)	3.87-5.38	0.04	**0.15 (0.03)** **0.06 (0.02)** 3.23 (0.05)	4.53 (0.04)	3.69-5.36	0.06	**0.18 (0.03)** **0.07 (0.02)** 2.87 (0.04)	4.43 (0.05)	3.44-5.42	0.08	**0.25 (0.04)** **0.05 (0.02)** 3.07 (0.04)
371	Cesarean section without complications	3.69 (0.03)	2.98-4.40	0.16	**0.13 (0.02)** **0.03 (0.00)** 0.64 (0.00)	3.66 (0.03)	2.94-4.38	0.17	**0.13 (0.02)** **0.03 (0.00)** 0.62 (0.00)	3.59 (0.03)	2.88-4.31	0.17	**0.13 (0.02)** **0.03 (0.00)** 0.62 (0.00)

* variance component at the hospital level
** variance component at the physician level
*** variance component at the patient level
values are bold if the difference between the hospital and physician variance is significant

Table 4.4 T-test and sign test for multihospital physicians 1999, 2000 and 2001

DRG	Diagnosis/procedure	N (physicians in 2 hospitals)			T-test pair physician1 in A and hospital A			T-test pair physician1 in B and hospital B			Sign difference between hospitals equals difference for the physicians (%)		
		1999	2000	2001	1999	2000	2001	1999	2000	2001	1999	2000	2001
88	Chronic obstructive pulmonary disease	537	516	516	nsig	nsig	nsig	nsig	nsig	nsig	56.1	54.7	54.7
127	Congestive heart failure	877	972	1,009	nsig	nsig	nsig	nsig	nsig	nsig	55.8	57.8	57.6
209	Major joint procedures	123	135	95	sig*	nsig	nsig	nsig	nsig	nsig	68.3	67.4	68.4
358	Hysterectomy with complications	120	87	103	nsig	nsig	nsig	nsig	nsig	nsig	58.3	60.9	58.3
359	Hysterectomy without complications	100	73	71	nsig	nsig	nsig	sig*	nsig	nsig	51.0	56.2	57.7
370	Cesarean section with complications	128	130	124	nsig	nsig	nsig	nsig	nsig	nsig	67.2	62.3	62.9
371	Cesarean section without complications	104	74	74	nsig	nsig	nsig	nsig	nsig	nsig	86.5	70.3	73.0

* $p<0.05$

The combination of analyses which comprised the second component of the study demonstrated that physicians who practiced at more than one hospital produced lengths of stay that were similar to the local standards where they were practicing. The t-tests demonstrated that stays for multihospital physicians did not differ significantly from the local standard for five of the DRGs in all three years, and for two of these categories in two of the three years. Finally, the sign tests demonstrated that differences between hospital stays were similar to differences between stays chosen by multihospital physicians in a majority of the cases evaluated.

These two analyses provided convincing evidence of the linkage between multihospital physician lengths of stay and local standards. On the basis of this information, hypothesis 2 was confirmed.

Results of the study of hypothesis 3 are summarized in table 4.5. This information relates to the potential impact on lengths of stay of the proportion of physicians working in only one hospital on total practice variation within a hospital.

The results of the physician variance analysis demonstrated that the relationship between the proportion of the physicians working in only one hospital and the total physician variation was non-significant for a majority of the DRGs and years of data evaluated. This means that the variation does not change when the proportion of physicians working in one hospital changes. Significant relationships were identified for chronic obstructive pulmonary disease, congestive heart failure, major joint procedures, and cesarean section with complications in two years of data and in two isolated instances, viz. hysterectomy with complications in 2000 and cesarean section without complications in 2001. Only for chronic obstructive pulmonary disease the variation decreased when the proportion of physicians working in one hospital increased, in all other significant cases the variation increased when the proportion of physicians increased.

On the basis of this analysis, hypothesis 3 was not confirmed. The conclusion might indicate a connection between length of stay choice and hospital influences, like management. Besides normative processes, regulative processes cause similarities. Regulative processes, or formal rules, will cause these similarities also when there is a small proportion of physicians working in one hospital and not in another also. In the last decades there has been a change from professional control, through normative processes, to managerial control, which is regulative (Scott et al., 2002).

Table 4.5 Variances of the physician level intercept and the proportion of physicians working in one hospital within hospitals, in 1999, 2000 and 2001

DRG	Diagnosis/procedure	1999		2000		2001	
		Variance	Significance	Variance	Significance	Variance	Significance
88	Chronic Obstructive	1.11 (0.08)		0.60 (0.05)		0.73 (0.05)	
	Pulmonary Disease	-2.22 (0.60)	p<0.01	1.18 (0.53)	p=0.03	-0.36 (0.45)	nsig
127	Congestive Heart Failure	0.79 (0.04)		0.85 (0.04)		0.88 (0.05)	
		-1.39 (0.55)	p=0.01	0.01 (0.51)	nsig	-1.58 (0.48)	p<0.01
209	Major joint procedures	0.78 (0.05)		0.71 (0.05)		0.55 (0.04)	
		-1.00 (0.10)	p<0.01	-0.73 (0.13)	p<0.01	-0.19 (0.17)	nsig
358	Hysterectomy with	0.35 (0.05)		0.28 (0.03)		0.21 (0.03)	
	complications	-0.35 (0.33)	nsig	-0.63 (0.19)	p<0.01	0.22 (0.30)	nsig
359	Hysterectomy without	0.17 (0.01)		0.15 (0.01)		0.15 (0.01)	
	complications	-0.09 (0.06)	nsig	-0.07 (0.07)	nsig	-0.07 (0.07)	nsig
370	Cesarean section with	0.08 (0.02)		0.08 (0.02)		0.09 (0.02)	
	complications	-0.33 (0.15)	p=0.03	-0.19 (0.19)	nsig	-0.55 (0.16)	p<0.01
371	Cesarean section without	0.03 (0.00)		0.03 (0.00)		0.03 (0.00)	
	complications	0.01 (0.02)	nsig	-0.02 (0.02)	nsig	-0.04 (0.01)	p=0.01

4.4 Conclusions and discussion

In the classical conception of medicine as a profession, medical practice is supposed to be uniform due to the shared body of (theoretical) knowledge. Variation originates from the necessity to apply this theoretical knowledge to individual patients. Underlying assumption is one of professional autonomy; the decision of physicians is influenced by science and the medical condition of the patient only.
Residual variation still remains, however, when clinical variables and patient characteristics are taken into account and this residual variation is not random, but shows clear patterns. The most studied ones are small area variations (Wennberg and Gittelsohn, 1982; Wennberg and Gittelsohn, 1975; Ashton et al., 1999). Researchers differ as to the attribution of this residual variation, some arguing that physicians prefer different practice styles, e.g. as a result of their education and professional socialization. The spatial phenomenon then emerges as a result of (self-)selection of physicians in certain areas or adaptation of preferences to those of others, resulting in a typical 'surgical signature'.
Other researchers have argued that the circumstances under which physicians work have a profound influence on their decisions, and that these shared circumstances make for homogeneity and differing circumstances result in variation. Freidson (1975) argued that in group practice, physicians can not make their decisions autonomously, because they depend on colleagues, and are controlled be them. Although individual physicians can be conscientious and ethical, circumstances can cause them to change their performance.

The physician influence on utilization of care has been the subject of several studies (e.g. Burns and Wholey, 1991; Geller et al., 1996). In a study by Burns et al. (1994) it was concluded that the physician is an important source of variation. Freidson (1975) stated that the social environment in which physicians work is more important for their medical behavior than their formal professional education. For instance physicians working in hospitals are subject to collegial norms, affecting their (medical) behavior. Coser (1957) demonstrated that there are different norms between wards within the same hospital. These differences, informal as opposed to formal, were related to different constraints due to differences in medical tasks. The wards, surgical and medical, were situated on two sides of the same floor.
In the debate in which explanations for variation are sought, the question arises at what level of analysis homogeneity might be expected. Reasoning from shared circumstances as a source of homogeneity, this study focused

on hospitals as more appropriate than areas (unless areas are defined on such a scale that they form the market area of a single hospital). Because we studied several DRGs separately, this is equivalent to studying wards of the same specialty. The hypothesis that variation within hospitals was small compared to variation between hospitals was confirmed in this study and persistent over three years, in most of the study populations.
This result, however, is still compatible with both the approach based on preferences for a certain practice style and the approach based on work circumstances. We therefore also analyzed the decisions of the same physicians in different hospitals. Preferences for a certain practice style are supposed to be relatively stable within the same person, but circumstances may clearly differ. Different analyses confirmed that physicians working in two hospitals with different average lengths of stay have a length of stay similar to the usual practice in the hospital where the procedure was performed.

Underlying assumption in our study is that the physician, not the hospital, decides over patients' discharges. The hospital-physician relationship has changed from physician owned in the early 20th century to one of joint control in the late 1980's (Shortell, 1991). However, the lack of common economic incentives makes it difficult for physicians and hospitals to cooperate (Shortell et al., 2000). With a change from professional dominance to managerial-market orientation (Scott et al, 2000), the question of whether the within-hospital similarities are choice or constraint is raised. Are physicians the central actors, or is it management? It might be that in the modern hospital formal management is more important than normative control by physicians. Still, it is the physician who signs the discharge note, and who runs the risk of malpractice suits. For the hospital, although there are incentives to do less, quality is important as well. Therefore, although regulative control might have become more important, physicians can still be considered important actors in the length of stay decisions. The physician is the one who discharges patients, and patients can only be discharged on the day their physician is around.
However, this might change in the future when for instance hospitalists, physicians employed by the hospital, sign discharge notes for patients of other physicians. For hospitals this has the advantage of being able to discharge a patient, who is in the condition of being discharged, when the physician is not around.

The hypothesis that variation within a hospital would be smaller when more physicians practice in that hospital alone was not confirmed. This finding

could, however, be due to the fact that there are not many hospitals in which there is a low proportion (range 0.16-0.99, mean 0.82) of physicians working only in that hospital. Results in this study are consistent with the results found in the study by Westert et al. (1993) and the study by Griffiths et al. (1979).

This study demonstrated that hospitals are important in studying variation in physicians' practice and explaining length of stay and practice variation. It is important to understand the variation phenomenon, because it will facilitate effective interventions to improve quality of care (Blumenthal, 1994). Knowing where variation originates, combined with knowledge about which variation is undesirable, is the key to successful interventions. The existence of variation is often interpreted as a sign of overuse of health care resources and resources could be saved if all physicians adapted to the lowest utilization rates (Fisher et al., 1994). Although overuse gets more attention, however, underuse could be a problem as well (McNeil, 2001). Sources of undesirable variation need to be identified, whether they are indicators of overuse or underuse, with an emphasis on those sources that can be influenced to improve the quality of medical care. If similar patients receive different treatments, or if in some areas it is more likely to receive a certain procedure than in others, some do not get what they need, or get more than they need. Both are undesirable. The former meaning that some people's health could be improved, the latter that there is a waste of resources by providing ineffective, unnecessary care. Part of variation can be legitimate, while another part is not. That part of variation that is not legitimate should be reduced. Although, up to now, there is no evidence that less variation is related to a higher quality of care (Fertig, 1993, Weide et al., 1992). Patients should receive the treatment that works best, against acceptable costs. So, besides the question whether variations matter for mortality, morbidity, and quality of life, the question whether some patients receive high quality of care against lower costs is important.

Beyond the importance of understanding the variation phenomenon, it is necessary to study physicians who practice in more than one hospital. Previous research showed that length of stay was longer for the medical patients and shorter for the surgical patients of physicians working in several hospitals, cesareans occur more frequently, and it was found that inpatient resource use is higher for physicians working in several hospitals (Burns and Wholey, 1991; Burns et al., 1995; Miller et al., 1996).
In this study we did not find that physicians who practice in two hospitals choose different lengths of stay compared to physicians who practice in one

hospital only. We found instead, that multihospital physicians had lengths of stay comparable to the usual practice in the hospital. This implies that physicians adjust to colleagues or circumstances in the hospital where they perform procedures. This implication is useful in our understanding of the variation phenomenon; there are forces within shared work environments that cause physicians to make similar decisions. In this study, we were not able to distinguish between influences exerted by colleagues and by circumstances. This could have been done if specific characteristics of the hospital had been taken into account. Those characteristics were unknown.
Patient selection could also be an explanation. Multihospital physicians may apply some criterion that is possibly related to length of stay, when choosing the hospital to which they admit patients. This, however, would reinforce the theory of the influence of the hospital on length of stay. The purpose of this article is not to describe differences between hospitals based on patient selection and therefore the possibility of composition effects should be excluded as much as possible. We did this by adjusting for case-mix including severity of illness.

An important question is whether these work environments can be used in interventions to improve the quality of medical care. Several further possible explanations remain for the patterns of variation within work environments. It may be that physicians adapt to the colleagues they work with - that they follow the pack as Eddy (1984) called it - or that there is another factor in the work environment, such as hospital management or the availability of beds and facilities (Westert and Groenewegen, 1999; Westert, 1992; Kroneman, 2001), which influences their medical choices.

There is probably not one single explanation, although one might be more important than the other. If variation is very much related to the influence of colleagues, consensus conferences are a useful instrument for effective interventions. Two strategies could be applied, viz. broad conferences trying to influence physicians from many different hospitals to adopt the same evidence-based standards and guidelines, or hospital-based conferences trying to influence all physicians in a certain group to change their 'local standards' towards more global, evidence-based standards. It is, however, questionable whether conferences are effective in changing physician behavior, or that stronger interventions should be developed. Thom (2000) showed in a study on the effects of a training course for physicians to improve their behavior and to increase patients' trust, that the intervention was not strong enough. Available evidence suggests that physician profiling can be efficient (Evans et al., 1995). Physician profiling is a technique used to

change hospital length of stay choice of physicians, by comparing their individual average to a benchmark figure, adjusted for severity of illness. The physicians are confronted with their average, without knowing the average of their direct colleagues. This individual-based, managerial approach avoids the informal standards within a hospital, which could be a barrier when trying to influence length of stay choices in a group approach. Managerial interventions might be more effective when applied to individuals, while interventions focusing on professional content might be more effective when the approach is based on teams of physicians.

Appendix 4.A The multilevel model

$$y_{ijk} \sim N(XB,\ \Omega)$$

$$y_{ijk} = \beta_{0ijk}x_{0k} + \beta_1 x_{1ijk} + \beta_2 x_{2ijk} + \beta_3 x_{3ijk} + \beta_4 x_{4ijk} + \beta_5 x_{5ijk} + \beta_6 x_{6ijk} + \beta_7 x_{7ijk} + \beta_8 x_{8ijk} + \beta_9 x_{9ijk} + \beta_{10} x_{10ijk} + \beta_{11} x_{11ijk}$$

$$\beta_{0ijk} = \beta_0 + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0,\ \Omega_v) \ :\ \Omega_v = \begin{bmatrix} \sigma^2_{v0} \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u0} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

γ_{ijk} = dependent variable (length of stay)
i = patients
j = physicians
k= hospitals

β_{0ijk} = mean and variance parameters
β_0 = mean (centered)
β_1 = age (centered)
β_2 = sex (centered)
β_3 = severity 2 (centered)
β_4 = severity 3 (centered)
β_6 = severity 4 (centered)
β_7 = white (centered)
β_8 = black (centered)
β_9 = native (centered)
β_{10} = asian (centered)
β_{11} = other (centered)

v_{0k} = level 3 (hospitals) residuals
υ_{0jk} = level 2 (physicians) residuals
e_{0ijk} = level 1 (patients) residuals

σ^2_{v0} = level 3 (hospitals) residual variance
σ^2_{u0} = level 2 (physicians) residual variance
σ^2_{e0} = level 1 (patients) residual variance

Appendix 4.B Multilevel model with co-variance estimations

$$y_{ijk} \sim N(XB,\ \Omega)$$

$$y_{ijk} = \beta_{0ijk}x_{0k} + \beta_1 x_{1ijk} + \beta_2 x_{2ijk} + \beta_3 x_{3ijk} + \beta_4 x_{4ijk} + \beta_5 x_{5ijk} + \beta_6 x_{6ijk} + \beta_7 x_{7ijk} + \beta_8 x_{8ijk} + \beta_9 x_{9ijk} + \beta_{10} x_{10ijk} + \beta_{11} x_{11ijk} + \beta_{12jk} x_{12k}$$

$$\beta_{0ijk} = \beta_0 + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\beta_{12jk} = \beta_{12} + v_{12k} + u_{12jk}$$

$$\begin{bmatrix} v_{0k} \\ v_{12k} \end{bmatrix} \sim N(0,\ \Omega_v) \ :\ \Omega_v = \begin{bmatrix} \sigma^2_{v0} & \\ \sigma_{v120} & \sigma^2_{v12} \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \\ u_{12jk} \end{bmatrix} \sim N(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ 0 & \sigma^2_{u12} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

γ_{ijk} = dependent variable (length of stay)
i = patients
j = physicians
k= hospitals

β_{0ijk} = mean and variance parameters
β_0 = mean
β_1 = age (centered)
β_2 = sex (centered)
β_3 = severity 2 (centered)
β_4 = severity 3 (centered)
β_6 = severity 4 (centered)
β_7 = white (centered)
β_8 = black (centered)
β_9 = native (centered)
β_{10} = asian (centered)
β_{11} = other (centered)
β_{12} = proportion of physicians working in one hospital (centered)

ν_{0k} = level 3 (hospitals) residuals
υ_{0jk} = level 2 (physicians) residuals
e_{0ijk} = level 1 (patients) residuals

v_{12k} = level 3 (hospitals) residuals for the proportion of 1-hospital physicians
υ_{12jk} = level 2 (physicians) residuals for the proportion of 1-hospital physicians

σ^2_{v0} = level 3 (hospitals) variance without that part of the variance that can be attributed to the proportion of 1-hospital physicians
σ^2_{v12} = part of the level 3 variance that can be attributed to the proportion of 1-hospital physicians
σ_{v120} = covariance between the proportion of 1-hospital physicians and the hospital variance
σ^2_{u0} = level 2 (physicians) variance without that part of the variance that can be attributed to the proportion of 1 hospital physicians
σ^2_{u12} = part of the level 2 variance that can be attributed to the proportion of 1 hospital physicians
σ^2_{e0} = level 1 (patients) variance

5

Does managed care make a difference?

Physicians' length of stay decisions under managed and non-managed care

This article was published as:
De Jong JD, Westert GP, Noetscher CM, Groenewegen PP. Does managed care make a difference?: physicians' length of stay decisions under managed and non-managed care. *BMC Health Serv Res*, 2004; 4:3

Abstract

In this study we examined the influence of type of insurance and the influence of managed care in particular, on the length of stay decisions physicians make and on variation in medical practice.

We studied lengths of stay for comparable patients who are insured under managed or non-managed care plans. Seven Diagnosis Related Groups were chosen, two medical (COPD and CHF), one surgical (hip replacement) and four obstetrical (hysterectomy with and without complications and cesarean section with and without complications). The 1999, 2000 and 2001- data from hospitals in New York State were used and analyzed with multilevel analysis.

Average length of stay does not differ between managed and non-managed care patients. Less variation was found for managed care patients. In both groups, the variation was smaller for DRGs that are easy to standardize than for other DRGs.

Type of insurance does not affect length of stay. An explanation might be that hospitals have a general policy concerning length of stay, independent of the type of insurance of the patient.

5.1 Introduction

There is concern that factors other than the medical needs of a patient influence decision–making by physicians (Bachman and Freeborn, 1999; Donelan et al., 1997). Non-medical factors play a role in explaining medical practice variation (Eisenberg, 1985; Greenfield et al., 1992; Langley et al., 1992). Among the factors that influence medical treatment are uncertainty of the most effective practice, response to regulations, method of patients' payment to the physicians, and type of insurance coverage (Greenfield et al., 1992).

In this study we examined the influence of type of insurance, and the influence of managed care in particular, on the decisions physicians take and on variation in medical practice. Managed care plans have evolved in the USA, where they are widely used to control costs by combining the financing and delivery of health care. Providers are at financial risk in capitated plans and the insured have less choice where treatment and health care providers are concerned. The potential of managed care is attractive to policy makers in Europe.

Different types of insurance coverage are linked to different premiums and different ways in which providers of care are paid. As a consequence, the type of insurance cover that patients have might influence the treatment given to patients. In traditional insurance plans, providers are paid fee-for-service. The insured can choose the physicians they want and consult them whenever they want, resulting in maximum freedom for both the insured and the provider. These plans were fully unmanaged in the past, but even these plans use managed care to some extent, nowadays (Weiner and De Lissovoy, 1993).

The Health Maintenance Organization (HMO) is the best known type of managed care in the US (Bachman and Freeborn, 1999). HMO insured are obliged to choose a primary care physician, and treatment by specialists is only compensated after referral by the primary care physician. Physicians within the HMO network are often at financial risk; they are capitated or face a risk-sharing withdrawal (Scott et al., 2000; Schut, 1986; HIAA, 1999; AMA, 1999; Mechanic, 2000; Simon et al., 1998; Kerr et al., 1995; Arnould and DeBrock, 1990).

It should be emphasized that the terms managed and non-managed care were developed decades ago, when Health Maintenance Organizations were

well defined organizations that used specific techniques to manage hospital utilization. These definitions have become less clear, however, particularly during the last few years. Managed care organizations have adopted more features of traditional health plans, such as the ability of patients to access specialty physicians directly, while at the same time, traditional insurance plans have made greater use of utilization controls once practiced largely by health maintenance organizations.

Managed care insurers use various utilization management strategies to reduce health care costs, primarily by avoiding unnecessary hospital admissions. This is done by using the primary care physician as a gatekeeper, reducing length of stay and negotiating reduced payments to providers for services (Mechanic, 2000; Fairfield et al., 1997a; Fairfield et al., 1997b). Moreover, preauthorization for specialty care is required (Simon et al., 1998). Restrictions on the treatment a physician can provide are greatest in fully managed delivery systems. Concurrent treatment and retrospective utilization review are common (Weiner et al., 1993).

The performance of physicians is judged on length of stay, among other things. This makes length of stay a valuable outcome variable, although lengths of stay in the USA are already shorter than in Europe.

Different types of insurance provide different constraints and incentives that influence the length of stay decision. Using ordinary regression analysis, it was found that the way patients are insured (Shi, 1996; Bradbury et al., 1991) and physicians are paid (Lave and Frank, 1990) significantly influences length of stay. A review carried out by Miller and Luft (1994) reported that length of stay was shorter for patients in Health Maintenance Organizations (HMOs; a managed care organization), in fifteen out of sixteen observations from thirteen studies using data from 1980 onwards. Our study is different from these studies, as a result of our methodology and the focus on managed and non-managed care alone. HMOs (managed care) will be compared to traditional plans (non-managed care).

The question we seek to answer is the following:

> *'Do physicians choose different lengths of stay for comparable patients who are differently insured (managed/non-managed)?'*

Hypotheses

HMOs try to control hospital costs, which means they have to influence a variety of decisions made by the insured and their physicians. It is most effective to prevent the insured from being admitted to the hospital, but once in hospital, length of stay should be influenced. In this study, we focus on this influence on length of stay.

Physicians decide on treatment strategies and timing of discharge, thereby determining the length of stay. On the other hand, physicians are confronted with constraints that influence their decisions. Our assumption is that variation between the decisions made by physicians is related to systematic differences in the constraints they face. In this study, we focus on one important set of constraints, viz. those set by the insurer of the patients.

Constraints for the HMO insured will be far more restrictive than for the traditionally insured, implying that variation between physicians treating patients with an HMO insurance will be less than the variation between physicians treating other patients. Physicians treating patients with an HMO insurance face similar constraints and incentives, leading to similar length of stay decisions. Inpatient days are very expensive for insurers, who will therefore try to limit these expenses, by using incentives for physicians to discharge patients as soon as possible. Physicians sometimes receive a bonus from the HMO-insurer for example, if they reach a certain utilization target and physicians will try to earn this bonus in order to increase their income.
On the basis of the differences in constraints between managed and non-managed care plans, we hypothesize that:

Hypothesis 1: *Length of stay will be shorter for managed care patients.*

Hypothesis 2: *There will be less variation in length of stay for managed care patients.*

Physicians decide on treatment strategies and timing of discharge, thereby determining length of stay. Neither the hospital nor the insurer signs the discharge note, and hospital and insurer can only try to influence that decision. Physicians deal with different insurers within one hospital, and variation within a hospital is therefore to be expected.

Hypothesis 3: *The influence of managed care, i.e. less variation in length of stay, will primarily be found at physician level.*

Apart from the difference between managed care and non-managed care in the substance of the constraints they apply to physicians, the restrictiveness of the constraints may vary with market conditions. Physicians can be induced to follow rules set by the insurer if the physicians are dependent on the insurer. In the case of managed care plans, physicians have a contract and this offers opportunities for influencing behavior.
Managed care plans set rules for hospitals and physicians to follow and the importance of following those rules will be higher when a physician has a lot of managed care patients. The physician will avoid losing these patients by following the rules as best as possible.

Hypothesis 4: *The higher the proportion of managed care patients the physician has, the shorter the length of stay and the less the variation in length of stay at physician level.*

The same applies to hospitals. When there are a lot of managed care patients, a hospital will try to influence the decision physicians make, and thus to ensure the criteria are met.

Hypothesis 5: *The higher the proportion of managed care patients the hospital has, the shorter the length of stay and the less the variation at hospital level.*

The effectiveness of an insurer in influencing physicians is conditional to the dependency of the physicians on that insurer. If a physician deals with one insurer, it will be possible for that insurer to control that physician's behavior. If the physician has an alternative, an insurer will have less power over medical decisions.

Hypothesis 6: *Physicians who deal with fewer insurers will have less variation in length of stay for managed care patients.*

Again, the same applies to hospitals. Hospitals dealing with fewer insurers will be more dependent on these insurers and will therefore be more easily controlled. These hospitals will try to be more effective in controlling the physicians practicing in the hospital, in order to keep the insurer satisfied. Stringent credentialing and utilization reviews will be carried out (Burns and Wholey, 1991).

Hypothesis 7: *Hospitals dealing with fewer insurers will experience less variation in length of stay for managed care patients*.

There are interdependencies between hospitals and physicians. Hospitals need good physicians to attract patients, physicians need hospitals to care for their patients and to provide equipment. Whether one is able to influence the behavior of the other in cases of divergent incentives, depends on the existence of an alternative. The importance of the relationship between physicians and hospitals will be greater when physicians practice in fewer different hospitals (Shortell, 1991). Physicians will be more dependent, and are thus more easily controlled when they practice in fewer hospitals. As a consequence, physicians will show less variation in their length of stay choice when they work in fewer different hospitals.

Hypothesis 8: *Physicians practicing in fewer different hospitals will have less variation in length of stay for managed care patients.*

Insurers pay hospitals on a DRG basis and DRGs consist of conditions requiring similar lengths of stay in the hospital (Burns and Wholey, 1991). Rules on length of stay made by insurers will not be as restrictive for all DRGs and there will probably be a difference between those DRGs that are easy to standardize and those that are not. Surgical DRGs, for example, can be more easily standardized than medical DRGs like Chronic Obstructive Pulmonary Disease (COPD).

Hypothesis 9: *The easier it is to standardize treatment for a specific DRG, the less variation in length of stay there will be for patients under managed care.*

5.2 Data and methods

Description of the data

Data were obtained from the New York Statewide Planning and Research Cooperative System (SPARCS), which is a comprehensive patient data system established as a result of cooperation between the health care industry and government. SPARCS is a major management tool assisting hospitals, agencies, and health care organizations with decision-making regarding financial planning and monitoring of inpatient and ambulatory surgery services and costs in New York State. It is important to recognize the fact that there are huge inter-state differences in insurance programs. Medicaid in one state, for instance, is different to Medicaid in another state.

Managed care penetration in New York State is below the average for the USA; an average of 10% of inpatient contacts is under a managed care

program. There are 62 counties (58 in the analysis) and insurance plans differ per county. The number of physicians per 10,000 civilian population in New York State is 35.3, which is higher than the US average of 25.5 (1995 data; US Department of Health and Human Services, 2000).

We used 1999, 2000 and 2001 SPARCS-data and seven DRGs were studied: two medical (DRGs 88 and 127: Chronic Obstructive Pulmonary Disease and Congestive Heart Failure), one surgical (DRG 209: hip replacement) and four obstetrical (DRGs 358, 359, 370 and 371: hysterectomy with and without complications, cesarean section with and without complications). Cases for which no physician was known were omitted (1.3% of all cases) and only patients above the age of twenty were included. Patients with extremely long stays (defined as the average length of stay plus 1.96 times the standard deviation) were excluded, which involved a minimum of 0.86% and a maximum of 3.72% of cases per procedure. The study populations for all three years are summarized in table 5.1.

Analyses

Each DRG was analyzed separately, with three groups of patients being created within each DRG: one for managed care (HMO, Medicaid HMO and Medicare HMO), one for non-managed care (Blue Cross/Blue Shield, commercial insurance, Medicaid, Medicare) and a group containing all patients not insured or otherwise insured. Mean age for the managed care patients is 47.9 years, for the non-managed care patients 61.1 years, and for all other patients 49.5 years. The percentages of managed care patients for each DRG are fourteen for COPD, fourteen for CHF, twenty-one for total hip, thirty-eight for hysterectomy with complications, thirty-eight for hysterectomy without complications and thirty-seven and forty for Cesarean section with and without complications respectively.

Multi-level analysis was used to examine the differences in length of stay between the managed and the non-managed care groups, thus acknowledging the fact that patients are hierarchically nested within physicians and physicians within hospitals (Snijders and Bosker, 1999). We controlled for age, sex, race and co-morbidities. Characteristics of the physician included in the analysis were the proportion of patients insured under a managed care system (range 0-1), the number of hospitals in which the physician practices (range 1-7) and the number of insurers (range 1-10) that was dealt with. Hospital characteristics that were included were the proportion of patients insured under a managed care system (range 0-0.6) and the number of insurers (range 1-12). All variables are centered.

Table 5.1 Study population: number of patients, physicians, hospitals and percentage of excluded cases per DRG 1999, 2000, 2001

DRG	Diagnosis/procedure	Number of discharges			Number of physicians			Number of hospitals			% cases excluded		
		1999	2000	2001	1999	2000	2001	1999	2000	2001	1999	2000	2001
88	Chronic Obstructive Pulmonary Disease	38,424	36,478	34,400	9,125	8,779	8,476	237	234	228	0.86	3.46	2.62
127	Congestive Heart Failure	62,682	62,599	59,763	11,282	11,285	11,048	233	234	229	3.21	3.72	3.00
209	Hip replacement	28,426	29,827	32,016	1,267	1,238	1,240	210	207	204	2.58	3.04	3.13
358	Hysterectomy with complications	8,408	8,412	8,137	2,236	2,154	2,104	215	208	207	1.87	3.41	3.58
359	Hysterectomy without complications	21,962	22,926	22,604	2,630	2,608	2,520	216	215	212	2.46	2.09	1.78
370	Cesarean section with complications	11,731	12,125	11,955	2,217	2,240	2,183	164	161	159	2.61	2.54	2.42
371	Cesarean section without complications	39,844	42,980	42,769	2,602	2,654	2,571	163	161	159	1.31	1.25	1.22

Furthermore, we controlled for differences in insurance programs between counties by adding county as a level to our model. The model consists of four levels, viz. the level of the patient, the physician, the hospital and the (hospital) county. Separate models were fit for each year. Table 5.2 shows how the different hypotheses were tested. The variance is used as a measure of variation.

Table 5.2 Description of the hypothesis testing

Hypothesis	Description	Method of testing
1	Shorter length of stay managed care	Mean length of stay for the managed care and the non-managed care group are compared
2	Less variation length of stay managed care	Variation in length of stay for the managed care and the non-managed care group are compared
3	Influence managed care at physician level	The variation for the managed care group and the non-managed care group at physician level is compared to the variation for both groups at hospital level
4	Shorter length of stay, less variation when more managed care patients per physician	The regression coefficient for the proportion of managed care patients per physicians is examined as well as the covariance between this proportion and the variation in length of stay; both are expected to be negative
5	Shorter length of stay, less variation when more managed care patients per hospital	The regression coefficient for the proportion of managed care patients per hospital is examined as well as the covariance between this proportion and the variation in length of stay; both are expected to be negative.
6	Fewer insurers per physician, less variation in length of stay	The covariance between the number of insurers per physician and the variation in length of stay is examined and expected to be positive.
7	Fewer insurers per hospital, less variation in length of stay	The covariance between the number of insurers per hospital and the variation in length of stay is examined and expected to be positive.
8	Fewer different hospitals per physician, less variation in length of stay	The covariance between the number of hospitals per physician and the variation in length of stay is examined and expected to be positive.
9	DRGs that can be standardized show less variation	Variation for all DRGs is compared, most variation is expected in medical DRGs and DRGs with complications

5.3 Results

Table 5.3 shows the mean length of stay for the managed and the non-managed care groups for each DRG. Length of stay for the managed care group is slightly shorter in two out of seven cases. The differences are very small, however, being even less than the half-day that is the minimum to save costs. The hypothesis (1) is not confirmed.

Table 5.4 shows the variation in length of stay for the managed and non-managed care groups for each DRG. The variation within the managed care group is significantly smaller for five out of seven DRGs, which is consistent with the hypothesis (2). Most of the variation between managed and non-managed care groups can be found at patient level.
The difference in the variation for the managed and non-managed care groups at hospital level and at physician level is measured as a ratio (the variation of the managed care group divided by the variation of the non-managed care group). The ratio is one if the variation for both groups is the same, less than one if the variation of the managed care group is smaller, and greater than one if the variation of the managed care group is greater than the variation of the non-managed care group. If the difference in variation between the managed and the non-managed care groups is insignificant, the ratio is set at 1. A difference in variation between the two groups is found only at physician level for DRG 88, Chronic Obstructive Pulmonary Disease, (1.43, $p<0.1$), DRG 127, Congestive Heart Failure, (0.71, $p<0.1$), DRG 209, hip replacement, (0.73, $p<0.05$), and DRG 370, Cesarean section with complications, (3.44, $p<0.1$). These results do not provide unequivocal evidence indicating that the variation within the managed care group is smaller than the variation within the non-managed care group at physician level, and the hypothesis is not confirmed (3). There are no differences between the variations for both groups at hospital level for all DRGs.

Table 5.5 summarizes the regression coefficients and the covariances for the different variables. All four significant regression coefficients for the proportion of HMO patients per physician show that the higher the proportion of managed care patients that physicians have, the shorter the length of stay. The covariance shows the relation between the proportion of managed care patients and the variation in length of stay for managed care patients. Two significant covariances have a negative sign, which means that variation between physicians is lower when the proportion of managed care

patients is higher; two other significant covariances show the opposite. The hypothesis is not confirmed (4).

The significant regression coefficients for the proportion of managed care patients per hospital show that length of stay is higher when the proportion of managed care patients is higher. Our hypothesis is not confirmed. The significant covariances show opposite effects. The hypothesis is not confirmed (5).

As expected, variation in length of stay is higher when the number of insurers per physician is higher (hypothesis 6). We found two significant covariances for the influence of the number of insurers per hospital, indicating a higher variation when the number of insurers is higher. This is consistent with the hypothesis (7). Covariances for the number of hospitals in which a physician practices show that the variation in length of stay is higher for three DRGs and is lower for another DRG when a physician practices in more hospitals. The hypothesis (8) is not confirmed.

We compared variation for the seven DRGs to test the last hypothesis (9) on whether variation in length of stay will be less when treatment is easy to standardize. The comparison shows (table 5.4) that variation is smallest for DRGs 359 and 371, which are the obstetrical DRGs without complications and are DRGs that can be standardized. Variation is greatest in the medical DRGs (88 and 127), which are less easy to standardize. This is true for both the managed care patients and the non-managed care patients. The hypothesis (9) is confirmed.

Table 5.3 Mean length of stay (LOS) for managed and non-managed care groups for each DRG

DRG	Diagnosis/procedure	Mean LOS in days (s. error) managed care group			Mean LOS days (s. error) non-managed care group			Difference between managed and non-managed care groups (days)		
		1999	2000	2001	1999	2000	2001	1999	2000	2001
88	Chronic Obstructive Pulmonary Disease	3.92 (0.31)	3.93 (0.24)	3.47 (0.24)	3.77 (0.29)	3.79 (0.20)	3.35 (0.22)	0.15	0.14	0.12
127	Congestive Heart Failure	1.77 (0.25)	1.77 (0.25)	1.74 (0.25)	1.71 (0.22)	1.58 (0.23)	1.67 (0.22)	0.06	0.19	0.07
209	Hip replacement	5.21 (1.16)	4.35 (0.74)	5.02 (0.40)	5.25 (1.15)	4.62 (0.74)	5.01 (0.39)	-0.04	-0.27	0.01
358	Hysterectomy with complications	3.08 (0.75)	1.93 (0.45)	2.06 (0.45)	3.18 (0.74)	1.92 (0.45)	2.10 (0.46)	-0.10	0.01	-0.04
359	Hysterectomy without complications	2.38 (0.31)	2.35 (0.17)	2.68 (0.13)	2.42 (0.31)	2.28 (0.17)	2.71 (0.12)	-0.04	0.07	-0.03
370	Cesarean section with complications	2.95 (0.14)	2.40 (0.80)	1.89 (1.76)	2.82 (0.13)	2.37 (0.80)	1.88 (1.77)	0.13	0.03	0.01
371	Cesarean section without complications	2.88 (0.30)	2.69 (0.21)	3.28 (0.22)	2.84 (0.30)	2.70 (0.21)	3.26 (0.21)	0.04	-0.01	0.02

Table 5.4 Variation in length of stay for managed and non-managed care groups for each DRG

DRG	Diagnosis/procedure	Variation managed care group			Variation non-managed care group			Difference between managed and non-managed care groups (days)		
		1999	2000	2001	1999	2000	2001	1999	2000	2001
88	Chronic Obstructive Pulmonary Disease	14.4	9.52	9.27	17.1	10.5	10.9	-2.7**	-0.98**	-1.63**
127	Congestive Heart Failure	11.9	11.2	11.2	14.1	13.0	13.8	-2.2**	-1.80**	-2.65**
209	Hip replacement	4.91	4.59	4.04	6.67	5.72	4.95	-1.8**	-1.13**	-0.91**
358	Hysterectomy with complications	3.27	2.28	2.32	4.18	2.71	2.72	-0.9**	-0.43**	-0.40**
359	Hysterectomy without complications	0.89	0.83	0.84	0.90	0.88	0.86	-0.01	-0.05	-0.02
370	Cesarean section with complications	3.38	2.93	3.39	3.84	3.61	3.55	-0.46*	-0.68**	-0.16
371	Cesarean section without complications	0.80	0.77	0.74	0.83	0.82	0.81	-0.03	-0.05**	-0.07**

* $p<0.05$

** $p<0.001$

Table 5.5 Effects on length of stay; relevant regression coefficients (RC) and covariance with the managed care group (COV, whether this coefficient is positive or negative) for each variable per DRG

DRG	Diagnosis/Procedure	HMO patients per physician		HMO patients per hospital		Number of insurers per physician	Number of insurers per hospital	Number of hospitals per physician
		RC	COV	RC	COV	COV	COV	COV
88	Chronic obstructive pulmonary disease	-0.22	pos	0.48	neg**	pos	neg	pos*
127	Congestive heart failure	0.03	pos**	0.46	neg	pos	pos	pos
209	Hip replacement	-0.97**	neg**	-1.60**	neg**	pos**	neg	pos**
358	Hysterectomy with complications	-0.25**	neg	-0.25	pos	pos*	pos	pos*
359	Hysterectomy without complications	-0.08	neg**	0.49*	neg	pos**	pos	pos
370	Cesarean section with complications	-0.29*	pos*	0.78	pos*	pos	pos**	neg**
371	Cesarean section without complications	-0.08**	neg	0.59*	pos	pos**	pos*	neg

* $p<0.1$

** $p<0.05$

5.4 Conclusions and discussion

In this study, we found no difference in length of stay between managed and non-managed care patients. Furthermore, it appeared that there was less variation in length of stay for managed care patients. All results are summarized in table 5.6. Contrary to our expectations, this difference is not primarily found at physician level, nor is it found at hospital level. It is found at patient level, however, which means that patients with managed care insurance plans differ from patients with non-managed care insurance plans. This implies the existence of some sort of selection; patients insured under a managed care system are more similar than other patients.

Table 5.6 Results of the hypothesis testing

Hypothesis	Description	Test result
1	Shorter length of stay managed care	Not confirmed
2	Less variation length of stay managed care	Confirmed
3	Influence managed care at physician level	Not confirmed
4	Shorter length of stay, less variation when more managed care patients per physician	Confirmed, not confirmed
5	Shorter length of stay, less variation when more managed care patients per hospital	Not confirmed, not confirmed
6	Fewer insurers per physician, less variation in length of stay	Confirmed
7	Fewer insurers per hospital, less variation in length of stay	Confirmed
8	Fewer different hospitals per physician, less variation in length of stay	Not confirmed
9	DRGs that can be standardized show less variation	Confirmed

Selection by managed care insurance plans has been found in some other studies, which conclude that the managed care insured are younger and healthier (Miller and Luft, 1994; Block, 1997; Hellinger, 1995). The mean age of the managed care patients in this study is 47.9 years, while this is 61.1 years for the non-managed care patients. Fourteen percent of the CHF-patients are insured under a managed care plan, whereas this applies to 40% of the patients with a Cesarean section. There is less variation for procedures that are easy to standardize, such as those where no complications occur, irrespective of the type of insurance.

The question that comes to the fore is how unmanaged is non-managed care? New York State has lower managed care penetration than the US–average and approximately 10% of patients are covered by a managed care program, which suggests two possibilities. One is that utilization controls for these plans are more aggressive and that there are greater differences in utilization between them, because managed care penetration is so low. This was obviously not the case, however. A more plausible explanation is that the need to compete and limit managed care penetration has caused traditional insurance plans to adopt many of the techniques used in managed care.

Yet another possibility, is that the management of care in United States hospitals is increasingly provider-driven. In this context, hospitals will apply utilization controls to all payors to reduce expenses, rather than to individual groups of patients. This point has a major application to lengths of stay in New York State, where since 1986 all payors have reimbursed hospitals on a per discharge basis, rather than by the day. This means that hospitals have an incentive to reduce expenses for all payors, rather than simply those regarded as managed care plans.

Some hospitals employ case managers who do the discharge planning for all patients in the hospital, using benchmarks such as clinical pathways, criteria to determine necessity of admission or length of stay parameters. A hospital may compare its length of stay to similar hospitals in California, which generally have a shorter length of stay for the same diagnosis. This case management does not take into account whether a patient is under a managed care or a non-managed care program. Where an insurer is concerned, it can be highly effective to put pressure on a hospital and to let the hospital, in its turn, put pressure on the physicians. Length of stay or other requirements can be included in contracts between insurer and hospital, obliging the hospital to report to the insurer on the clinical care rendered by physicians or others. In order to avoid unnecessary utilization or losing a contract, it can be worth the effort for the hospital to employ case managers. Hospital norms can be communicated to physicians, patients and others involved in care, to ensure that hospital expectations regarding length of stay can be met. If the hospitals set their length of stay norm below that of all insurers and manage discharge effectively, there will be no difference in length of stay related to type of insurance. Since some insurers are more effective in length of stay management than hospital case management, some insurers will employ nurses in the hospital to manage for them. In these cases, insurance-related differences are possible.

In this study, there was a difference in the effect of the proportion of managed care patients on the physicians and on the hospitals. Where physicians are concerned, length of stay is shorter when this proportion is higher, but this effect was not found in hospitals. This means that physicians' length of stay choice is influenced by the insurer, while the hospital does not change its policy. The proportion of managed care patients does not have an unequivocal effect on the variation in both physicians and hospitals, indicating that there is no insurer effect.

The hypothesis that there is more variation when physicians practice in many hospitals is not confirmed. The option of treating patients in another hospital does not influence variation in length of stay for managed care patients, which would seem to be an effect of the insurer in combination with less variation when the number of insurers is lower.

Contrary to our study, insurance and payment were found to have a significant influence on length of stay in the studies mentioned in the introduction (Shi, 1996; Lave and Frank, 1990; Miller and Luft, 1994), which compared several types of insurance. Ordinary least squares regression was used to measure effects on length of stay, thus neglecting the fact that data are on different levels of aggregation. Hospital characteristics were also assigned to the patient level and when regression coefficients of hospital characteristics are assigned to the patient level, the units of analysis are considered to be independent observations. Patients are hierarchically nested within hospitals, however, and so the assumption of independent observations is not correct. As a consequence, different levels of analysis should be taken into account by using multi-level analysis (Snijders and Bosker, 1999). Furthermore, it is important to recognize the fact that there are considerable inter-state differences between insurance programs with the same name. Medicaid in one state is different to Medicaid in another state for instance, and these must therefore be considered as different programs, or analyses have to be made for individual states.

Bradbury et al. (1991) compared a specific type of HMO with traditional insurance programs by hospital, thus keeping hospital characteristics constant. Due to the fact that there had to be enough admissions of both types of insurance to a hospital for ordinary regression analysis to be possible, only ten (of the initial 78) hospitals were included in the analyses, a problem that could have been overcome with multi-level analysis. Results showed that for this specific type of HMO (the independent practice association or IPA), length of stay is shorter than for patients insured under

a traditional program. In addition to using a different methodology that might lead to different outcomes, all studies were carried out in the eighties and early nineties. The potential impact of the evolution of traditional insurance plans to include managed care techniques should not be discounted, and the terms traditional insurance and managed care plans may have become anachronisms in this context. This study may suggest the need for a more sophisticated approach to the subject, focusing on the impact of specific utilization management techniques.

Managed care was introduced in the US to keep health care costs from rising. Health care costs are also rising rapidly in Europe and a solution is being pursued as a consequence. While limiting the supply side and setting budgets were seen as solutions at first, there is a shift towards managed care nowadays. Although health care costs in the United States are the highest worldwide, the health care system there serves as an example to European countries as the introduction of managed care techniques is examined (Fairfield et al., 1997b; Erdmann and Wilson, 2001; Robinson, 1998).

Policy makers believe that managed care reduces costs without affecting the quality of health care. Nevertheless, it is open to question whether a cost reduction that provides less care and involves shorter lengths of stay truly does not affect the quality of care. Evidence of cost reduction is found in the short length of stay that is experienced in the United States, although it remains unclear whether this short length of stay is an effect of managed care or possibly of something else in the system. Furthermore, health care costs in the United States have continued to rise, despite the increasing number of managed care insured. This might be caused by the high costs that come with managed care systems, such as the costs of monitoring, or it may be due to increasing costs of medication, supplies, and various treatments. Some medications and treatments are still considered experimental, causing research and development costs to increase. Additionally, competing companies may develop similar drugs and treatments, each vying for use. The increases in United States health care costs could also be a pricing issue. In order to survive, hospitals may be raising prices and reducing expenses to order to keep up with payors (Anderson et al., 2003). It is also easy to lose sight of some (negative) effects of managed care that will come to the fore when elements from the health care system in the United States are transferred to European countries, since analyses of international health care policy have demonstrated that elements from one system cannot simply be transferred to a different system (Kimberley and De Pouvourville, 1993; Ross et al., 2000).

Managed care fits into the role of sickness funds in European social health insurance systems. Dutch sickness funds have a lot in common with HMOs for example; there is a contract with providers, providers receive a budget, a primary care physician is obligatory for the insured and monitoring of providers is common (Groenewegen et al., 2002a). Experiments with HMOs are taking place in Switzerland (Colombo, 2001; European Observatory on Health Care Systems, 2000) and there is interest in Germany in what is called 'Integrierte Versorgung', which is networks of health care providers that receive a budget from sickness funds (Groenewegen et al., 2002a). These can also be compared to HMOs.

In this study, we did not find evidence that it is managed care that has an effect on length of stay and thus on costs related to inpatient days. There seems to be something else that is causing the short length of hospital stay in the United States, independently of the patient's insurance.

What we found is that it is not restrictions imposed by the insurer that result in patterns of variation, since there are hardly any differences in length of stay of managed care and non-managed care patients. It might be the case that hospitals respond to the way they are paid; payment per DRG means that it is always (cost) efficient to keep length of stay short. Or it may be that the knowledge that managed care is on the increase is causing hospitals to react in advance by developing strategies to make sure that they will have (managed care) patients in the future (Zhang et al., 1999).

6

Do guidelines create uniformity in medical practice?

This article was submitted as:
De Jong JD, Groenewegen PP, Spreeuwenberg P, Schellevis F, Westert GP.
Do guidelines create uniformity in medical practice?

Abstract

This article aimed to test the general hypothesis that guidelines create uniformity, or reduce variation, in medical practice. Medical practice variation has policy interest and is one of the reasons for developing guidelines. The development and implementation of guidelines was considered in the broader context of processes of rationalization. We focused on the influence of guidelines introduced for family physicians in the Netherlands on variation in drug prescription. These voluntary guidelines are developed by the profession. It showed that although there was an overall increase in variation in prescription the increase was less in the cases of diagnoses for which guidelines were introduced. Guidelines, primarily, had an effect on variations in single-handed practices.

6.1 Introduction

In the classical conception of medicine as a profession, medical practice is largely uniform through the shared body of (theoretical) knowledge. Variation originates from the necessity to apply this theoretical knowledge to individual patients. However, when clinical variables and patient characteristics are taken into account, there is variation left. This variation could stem from the invention, and differential adaptation of this invention by physicians, like the invention of new drugs, medical treatments, and technical apparatus. Whatever the origin of this part of variation, it is striking that this variation has been found to show clear patterns by country, region, hospital and practice. Explanations for variation are sought in differences in opinions or enthusiasm for certain procedures between individual physicians, and in differences in constraints and social influences for groups of physicians (Wennberg and Gittelsohn 1975; Chassin 1993; Westert and Groenewegen 1999). Variation in medical practice is not a bad thing by definition; without variation there probably will be no progress. However, it is the downside of variation that attracts attention from third parties. Evidence of variations in medical practice suggests the possibility of inappropriate servicing, wasting of resources and even actual harm to patients (Evans, 1990). The existence of variation has policy interest and is one of the reasons, besides rising health care costs, for developing guidelines. The use of clinical guidelines that give recommendations about appropriate health care is a way of reducing variation and maintaining, or improving, the quality of health care (Grilli et al., 2000; Langley et al., 1998; Lomas et al., 1989). A wide variety of guidelines has been developed in the last decades for hospitals and physicians (Hibble et al., 1998). In this article guidelines for family physicians in the Netherlands will be studied.

In the Netherlands guidelines are developed for family physicians by the Dutch College of General Practitioners. The first appeared in 1989 and over 80 guidelines for different diagnoses have appeared ever since (http://www.nhg.artsennet.nl/guidelines). While several studies examined the adherence to guidelines (e.g. Schers et al., 2000; Hermens et al., 2001; Grol, 2001; Tiemeier et al., 2002), the impact on variation among physicians is hardly ever studied (e.g. Verstappen et al., 2003).

The development and implementation of clinical guidelines can be seen in the light of broader processes of rationalization, which occur everywhere in modern society. Processes of rationalization lead to more uniformity; guidelines introduced and followed by physicians create uniformity. The more physicians who follow the guidelines, the greater the uniformity. Still, processes of rationalization do not lead to more uniformity in all respects.

Although standards may specify for instance the therapeutic substances of drugs that are preferred for certain conditions, many different brands of drugs, containing the same therapeutic substances, can coexist.
This article focuses on the influence of guidelines on variation in drug prescription. It will not look at the contents of the guidelines, nor test whether guidelines are being followed, nor whether the quality of medical treatment is increased by the introduction of guidelines. It will test the general hypothesis that guidelines create uniformity. The general question addressed is: '*Is variation reduced after guidelines are introduced?'*. In other words: *'Do guidelines indeed create uniformity?'*.

Background and hypotheses
To answer this question we must first mention the broader processes of rationalization that occur throughout modern society. Secondly, we will be more specific about rationalization in medicine. Thirdly, the subject of guidelines will be discussed. Fourthly, more information about family physicians and guidelines in the Netherlands will be given. Finally, hypotheses will be formulated on when to expect a decrease in variation after guidelines are introduced. More specific expectations can then follow.

Rationalization in society
According to Weber processes of rationalization occur throughout society not only in technology, science and industry. Unexpectedly, arts and religion experience rationalization too even though they are assumed to be irrational (Ultee et al., 1996). These processes of rationalization, also referred to using terms such as technical progress, modernization, bureaucratization, professionalization, and secularization, are causing uniformity. Ritzer (2000) chose a more timely term for the processes of rationalization: 'McDonaldization'. With this term Ritzer refers to the most well-known example of a sector in which rationalization took place: the fast-food industry. Rationalization does not only apply to production, it also applies to consumption of goods and services like the use of chip cards. Common for rationalizing, in both production and consumption, is the attention given to time management.
Rationalization has to do with agreements and protocols. It also has to do with efficiency, predictability, calculability, quality - which does not necessarily mean the best quality but simply creating a stable product - and control. Processes and products are standardized and what one gets becomes highly predictable.

Rationalization in medicine

In medicine rationalization processes also take place. Worldwide, the profession of medicine is increasingly subject to the influences of market competition, forcing it towards standardization (Ritzer and Walczak, 1988). The production and diffusion of medical knowledge and technology are increasingly international. There is a change from professional dominance to managerial market orientation (Scott et al., 2000). The United States is in front but Europe is on its heels with the introduction of guidelines, protocols, diagnostic related groups or similar reimbursement systems that exert pressure to make more efficient use of health care resources. In terms of rationalization: the profession is changing from being characterized by substantive rationality to formal rationality[1], from being led by social values when making rational choices, to being controlled by rules and regulations (Ritzer and Walczak, 1988). The institutional changes in the health care sector that lead to increased formal rationality are expected to reduce variation in medical practice as physicians are increasingly operating in a predictable manner. They are also increasingly operating in a similar environment as rationalization processes cause similarities in institutions such as hospitals (Scott et al., 2000). Uniformity in the work environment will probably also have effects on variation in medical practice. Based on a literature review, Groenewegen and Westert (2004) concluded that there is indeed a downward trend in medical practice variation. How can we explain these changes?

Scott (2001) distinguishes three mechanisms by which institutions work: the regulative, normative, and the cognitive mechanism. Guidelines can contribute to more homogeneity through all three mechanisms. Guidelines with a regulative character consist of formal rules in which the behavior of physicians is monitored and sanctioned in order to influence behavior. Guidelines with a normative character are followed voluntarily. In a broad sense the medical profession can be seen as a normative community. However, within any social community, peers can exert pressure. There can be a social obligation to follow the guidelines. The cognitive mechanism relates to a common framework of meaning, similarities are caused by imitation or by the use of the same (computerized) systems of interpretation that influence treatment decisions. Computerized systems rationalize medical practice and can make it more scientific. They are meant to improve the capacity of physicians to make better medical decisions, whereas the

---

1 According to Kalberg (1980) four different types of rationality can be identified; practical, theoretical, substantive and formal rationality (see box 6.1).

complex problems physicians deal with surpasses their cognitive capacity (Berg, 1995).
Guidelines can be developed by different stakeholders such as insurance companies and organizations of medical professionals. They are supposed to increase the quality of care, or reduce costs, depending on which body is producing the guideline. The source of the guidelines is important as this is related to the acceptance of the guidelines by physicians. It determines too whether they are normative, meaning that there are no formal sanctions when the guidelines are not followed, or regulative, including formal sanctions (box 6.2). For instance insurance companies can develop guidelines in order to reduce costs, this goal in itself limits the acceptance amongst physicians. These guidelines, however, may still be followed because insurance companies can exert regulative pressure using formal sanctions such as through the authorization and rules on reimbursement for hospitals, physicians and patients.
In this article guidelines developed by the professional organization for family physicians in the Netherlands will be studied. These guidelines are normative, or voluntary rules, thus in essence it is up to the individual physician whether they are followed.

Box 6.1 Four types of rationality (Kalberg, 1980)

Practical rationality
Based on the capacity for means-end rational action, shaped by individuals' purely pragmatic and egoistic interests.

Theoretical rationality
Rooted in abstract cognitive processes instead of action, rational action may follow indirectly from theoretical rational thinking.

Substantive rationality
Derived from value-rational action, shaped by a coherent set of social values.

Formal rationality
Based on the capacity for means-end rational action, shaped by (universally applied) rules, laws and regulations.

Box 6.2 Sources of guidelines, underlying motivations and acceptance (Onion and Wally, 1995)

The profession
Guidelines developed by professionals (are seen to) aim at improving medical practice. Guidelines developed by professionals are generally seen as credible and accepted by the profession at large. An example of guidelines developed by the profession itself is the NHG-guidelines in the Netherlands. There are in essence only normative sanctions for those who do not work according to the guidelines.

Insurers, or the pharmaceutical industry
Guidelines developed by insurance companies (are seen to) aim at limiting use of medical resources. They are seen by professionals to be influenced by financial goals, even if they were the same as those developed by the profession (Tunis et al., 1994), and less easily accepted. Clinical guidelines in the USA are an example of such guidelines. Guidelines developed by the pharmaceutical industry are also approached with suspicion by professionals, irrespective of their quality. Regulative sanctions are used to have physicians follow the guidelines.

The Netherlands
The role of family physicians in the Netherlands is described in box 6.3. The Netherlands are a precursor in the development of clinical guidelines compared to other European countries. Guidelines are developed by the Dutch College of General Practitioners (NHG), they have developed and published guidelines since 1989. The guidelines were developed in order to improve the quality of family physicians practice. The guidelines can be used to support family physicians in their daily practice (Tielens, 1987), protect them from mistakes and legitimize medical behavior. The guidelines relate to diagnostics, treatment, referral, and prescribing. The NHG aims to achieve evidence-based practical guidelines that are widely accepted. In order to increase acceptance, the target group is involved in their development (see box 6.4). The idea is that guidelines are more readily accepted and acted upon if made and implemented by the profession itself (Grol, 2001).
Guidelines are both a reflection of what is common in the profession and what should become common (Burgers, 1994; Meulenberg and Goudswaard, 1997). Only part of the NHG-guidelines published between 1989 and 2000 are evidence-based. They are mostly based on a consensus between members of the study group that is involved in the development of the guideline (Tasche et al., 2001). Of 130 recommendations in 28 guidelines published between 1993 and 1997, 44% is evidence-based (Burgers et al., 2002).

Box 6.3 Family physicians in the Netherlands

Family physicians in the Netherlands have a gate-keeping role. All publicly insured patients (all patients with an income below 30,700 Euro for the year 2002, 60% of the total population) are on the patient list of a family physician or practice. Publicly insured patients have no direct access to specialist care; they have to consult their family physician first. Most of the privately insured patients also visit their family physician before consulting a medical specialist. Family physicians treat over 90% of all complaints themselves (Cardol et al., 2004; Committee of the Health Council, 2004), family physicians prescribe 75% of all drugs, and 65% of all consultations end with a prescription (Foets and Stokx, 1993).
Family physicians are working in single-handed practices (42%), in duo practices (33%), or in group practices (25%) (Van den Berg et al., 2004). The number of family physicians working in single-handed practices is decreasing over time.

Box 6.4 Guideline development procedure

The procedure to develop guidelines is as follows: the NHG selects a topic for which a guideline should be formulated, they mostly concern medical practice or organization; a working group of family physicians develops the guideline and this is sent to another 50 family physicians for comment. After the guideline is revised, it is sent to an independent scientific committee for authorization (Geijer et al., 1999). Then, the guideline is published in the official journal of the Dutch College of Family Physicians. Both members and non-members of the NHG react positively to the procedure and consider the NHG capable of formulating widely accepted guidelines to be used in family practice (Grol, 1989).

Why and when would variation be reduced by the introduction of guidelines?

Variation is expected to decrease when guidelines are followed. It is not certain that people will follow guidelines, for being different can be valued more than being similar (Brunsson and Jacobsson, 2000). Being different is important when people need to distinguish oneself from the other, for instance when there are many physicians it will be more important to have specific skills in order to be able to offer different treatments from the others. However, people will try not to be different if there is a degree of uncertainty. In situations where one is unsure about what to do, it is safer to demonstrate similar behavior to the others. Following guidelines means belonging to a group and it gives people approval and status, or protection when behavior needs to be explained. By following guidelines, it is easier to explain what has been done, and to achieve acceptance. This applies equally to family physicians as it does to all people. We will discuss two important mechanisms that influence the reduction of variation when guidelines are

introduced: the uncertainty of the outcomes of, or the appropriateness of, medical behavior and social integration in the profession. With these mechanisms differences between physicians working in single-handed practices and physicians working in group practices can be expected.

Guidelines are developed to support family physicians in their daily practice, and in this way ensure the quality of care (Lomas et al., 1989). Support is mostly appreciated in cases where one is uncertain about what should be done. Therefore, it could be argued that guidelines are used in order to reduce uncertainty. However, are individual family physicians uncertain about what should be done? Variation in medical practice might be interpreted as a sign of collective uncertainty, but it does not necessarily implicate individual uncertainty (Evans, 1990). Besides, there are other solutions to the problem of individual uncertainty; showing similar behavior as colleagues in general (of which following guidelines would be an example) or as colleagues in local practice situations (Eddy, 1984). When acting like others, one does not have to explain, or legitimize, one's behavior. It becomes a norm.

In general, people are most influenced by peers when there is a high degree of uncertainty, and objective, unambiguous information is not available (Bandura, 1986). Hence, the ones most likely to use the guidelines to reduce uncertainty, if they are, are those family physicians who are independent actors, who do not have colleagues around them to whom they might turn; physicians in single-handed practices. Physicians surrounded by colleagues are expected to show similar behavior to those colleagues. This behavior could be based on professional guidelines, but also on local standards.

In addition to the reduction of uncertainty, which is an *individual* interest, there is also the *collective* interest for following guidelines. Variation in medical behavior undermines the scientific basis of medicine and trust. Due to an asymmetric division of information between patients and family physicians who have specific knowledge, patients are unable to judge the quality of the services provided and therefore do not know if physicians are doing the right thing. For that reason, to reduce uncertainty among the clients, the public, about the quality of services, trust in the profession is important (Lulofs, 1981).

Trust in the profession is a collective good therefore all individuals benefit equally from it. The individual professionals, the family physicians, can contribute to the collective good by working according to the guidelines themselves, and by disciplining colleagues who do not. As Evans (1990, page 131) wrote: *"…curb your individual autonomy to protect your collective autonomy"*. The individual family physician will contribute to the collective if it is rational to follow the collective interests, that is if the cost of deviating

from the collective interest is high, for instance when deviating means not being able to join locum groups (Lulofs, 1981). Disciplining colleagues is inherent to a profession for autonomy is the test of professional status and self-regulation is the test of autonomy. A profession is granted autonomy from society based on the expectation that the profession itself will regulate the performance of its own members (Freidson, 1975). If professional guidelines are followed, an outside party, like an insurer, will have less reason to set rules in order to reduce variation, for example to the lowest level of performance in order to reduce costs. Assuming that all physicians attach importance to professional autonomy, and do not want to be controlled by third parties, the collective interest of working according to guidelines is the same for all family physicians. Yet there are differences in costs.

Family physicians in single-handed practice rely more on patients for social approval, while family physicians sharing their work environment with colleagues rely more on these role equivalents for their social approval (Freidson, 1975). Family physicians working in single-handed practices gain more social approval if their patients are satisfied. These physicians too may be less likely to spend more time in activities other than those directly related to patient care. Family physicians working in partnerships increase social approval by engaging in professional activities. This is related to following guidelines because the ones most likely to work according to the guidelines are the ones most integrated in the profession, those who are most likely to spend time keeping up with professional developments, and those who risk being sanctioned by colleagues (Coleman et al., 1966). Family physicians who do not share their work environment with colleagues are less likely to be disciplined by colleagues because their behavior is less visible to colleagues (Lulofs, 1981). The literature supports this. Family physicians spending more time in direct patient care were more likely to deviate from indicators derived from guidelines (Hutten, 1998). These guidelines from which indicators were derived, however, were not published at the time of data collection, but were to be published in the near future. This is an indication that guidelines reflect usual practice, as performed by family physicians working in group practices.

Besides disciplining each other, other physicians are an important source of information. Grol (2001) studied the successes and failures of the implementation of guidelines and found as sources of information: scientific journals; discussion of the guideline with a local group of physicians; and contact with other physicians and course attendance. Discussion and contact with colleagues goes without saying for physicians sharing their work environment, but it is less easy for family physicians working in single-

handed practices. Of course the work environment is not the only place where colleagues meet, and it is also possible for single-handed practitioners to consult a colleague. However, there is a difference between uniformity in talk and uniformity in behavior; people can discuss the way one should behave and agree on that, but actually show different behavior in each others absence. The crux is whether behavior is, or becomes, visible or not. Only when behavior is visible does one run the risk of being criticized.
Although behavior is not always visible in a shared work environment, it is more visible there than in a single-handed practice, as colleagues in shared practices see each others' patients. To avoid the risk of being criticized they will imitate those close colleagues with status and success, and behavior within a group will show similarities. Thus it is expected that family physicians working in groups developed their own, local standards, even in the absence of guidelines. Therefore, when guidelines are introduced little change in variation between physicians in group practices is expected. A change in behavior is, however, perfectly conceivable when guidelines deviate from local standards. So differences can be expected in the reduction of variation between family physicians in single-handed and those in group practices. In this article we will study the reduction in variation for drug prescription. It is hypothesized that:

Hypothesis 1: *Variation in drug prescription has decreased with the introduction of guidelines.*

Hypothesis 2: *Variation in drug prescription has decreased more for family physicians working in single-handed practices than for family physicians working in group practices.*

6.2 Data and methods

First and Second Dutch National Survey of General Practice
Data were used from the First and Second Dutch National Survey of General Practice (DNSGP1 and DNSGP2), collected in 1987 and 2001 respectively. In 1987 there were no guidelines, while in 2001 75 guidelines were available (http://www.nhg.artsennet.nl/guidelines). Data were collected on contacts, patients, family physicians and practices. For a description of the data-collection and comparability we refer to Westert et al. (2005). A summary of differences between both surveys can be found in table 6.1. We will elaborate on two differences between the surveys that influenced the design of the analyses: the registration period and the method of data collection.

Table 6.1 Summary of differences between the First and Second Dutch National Survey on General Practice

	DNSGP1	DNSGP2
Year	1987-1988	2000-2002
Number of practices	103	104
Number of FPs	161	195
Number of patients	335,000	390,000
Registration period	3 months	1 year
Data collection	On registration forms	FP information systems

Registration period
From all family physicians participating in DNSGP2, data for one year were retrieved from the electronic medical records. Family physicians participating in DNSGP1 were divided into four groups. Each recorded information on sheets during one of four consecutive three month periods. This difference in registration period complicates direct comparison of variation in both studies. Therefore we chose to compare differences, which will later be explained more fully, within DNSGP1 to differences within DNSGP2.

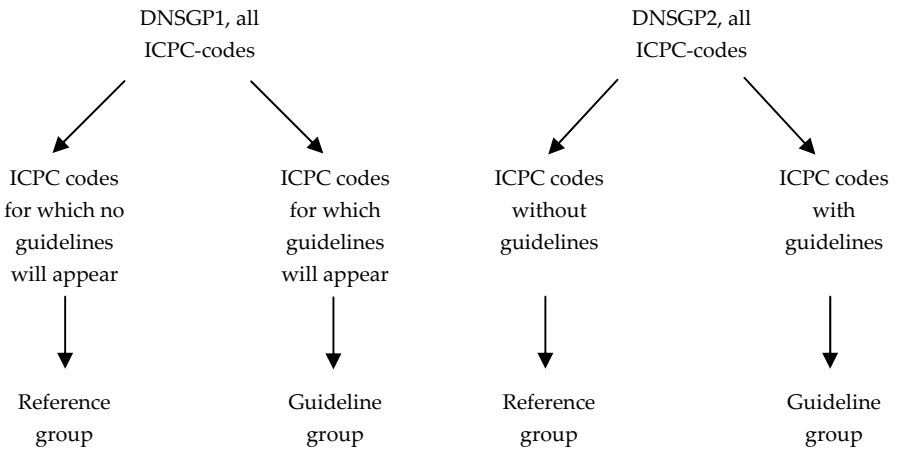
Data collection
During DNSGP1 the data were collected on paper registration forms, while the data for DNSGP2 were retrieved from Electronic Medical Records (EMR)[2], including diagnoses, prescription, and referral. The problem of differences in data collection between DNSGP1 and DNSGP2 was solved by comparing differences within one study to differences within the other study, as already mentioned in the previous paragraph.

Groups of diagnoses with and without guidelines
In order to make a comparison possible, it was decided to examine differences within studies between groups of diagnoses. For DNSGP1 two groups were created, one containing all diagnoses for which guidelines were about to appear, and one for which no guidelines were developed until

[2] In the Netherlands different software packages are used, all working slightly different. Software differences can cause differences in registration. We tested whether this was important in our analysis and decided not to take the different EMRs into account, because it did not have a significant effect but did complicate the analysis.

DNSGP2.[3] For DNSGP2 we also made two groups of diagnoses, one for which guidelines have appeared and one for which there were no guidelines in 2001 (figure 6.1). So the diagnoses in both studies were divided into two groups, a guideline group and a reference group. The first group was made by selecting codes of the international classification of primary care (ICPC - codes) (Lamberts et al., 1993) related to 75 NHG guidelines (appendix 6.A). In the reference group we have all ICPC-codes for which no guideline existed in 2001.

Figure 6.1 Creating the guideline and the reference group



Exclusions

Practices were excluded from the analyses if the data recorded were far from complete, for example if the period of recording data was only a few weeks, or ICPC-codes were scarcely recorded. Cases were excluded too for which no drugs could be identified, based on the ATC-5 code (Anatomical Therapeutic Chemical, 1993). Diagnoses that occurred less frequently, defined as less than 100 patients per diagnosis per study, were excluded. Furthermore, in DNSGP1 and DNSGP2 the same diagnoses were included.

In total 36 practices were thus excluded from the analyses, 32 due to incomplete recording. And 11,735 cases were excluded because no drugs could be identified based on the ATC-5 code. The actual numbers of diagnoses in the study population, practices, family physicians and 'family

[3] Between DNSGP1 and DNSGP2 ICPC codes differ, we adjusted ICPC-codes used during DNSGP1 to the codes used during DNSGP2.

physician diagnoses' [4] in our study can be found in table 6.2. The exclusions did not have an important negative effect on how representative this study was (table 6.3).

Table 6.2 Number of diagnoses, practices, family physicians and FPs diagnoses in our study

	DNSGP1 Guideline group	DNSGP1 Reference group	DNSGP2 Guideline group	DNSGP2 Reference group
ICPC-codes	77	57	77	57
Practices	102	102	69	69
FPs	160	160	108	108
FPs diagnoses	7,858	5,115	7,093	4,460

Table 6.3 A comparison between our population, the study population and the total population of Dutch FPs at the time of data gathering. The distribution of sex, age and types of practice

	Our population DNSGP1 (%)	Study population DNSGP1* (%)	Total population in 1987* (%)	Our population DNSGP2 (%)	Study population DNSGP2** (%)	Total population in 2001** (%)
Sex:						
male	85	85	92	78	73	74
female	15	15	8	22	27	26
Age:						
<35	13	32	9	1	4	6
35-39	31	30	27	10	15	15
40-44	26	12	23	13	18	20
45-49	13	12	11	35	30	25
50-54	7	7	8	24	25	22
55-59	8	6	9	16	8	10
60+	3	1	12	1	1	2
Type of practice:						
solo	32	32	56	34	31	43
shared	36	37	30	18	28	33
group/health centre	32	31	14	48	42	25

* Foets et al., 1992

** Schellevis et al., 2004

[4] Diagnoses are counted per family physician, 'family physician diagnoses' is the sum of all these different diagnoses counted per physician

Dependent variable

In this article variation will be represented as a concentration; the higher the concentration the less variation in drug prescription.

As a dependent variable a measure of concentration was used, the Herfindahl-Hirschman Index (HHI) (Den Hartog and Janssen, 2000; Zwanziger et al., 2000). This measure was based on the 'market share' of the drugs prescribed per diagnosis. The kind of drugs was identified based on the ATC-5 code. The HHI was measured as $\sum(a/b)^2$; where *a* is the number of times a specific drug was prescribed per diagnosis, and *b* is the total number of times any drug is prescribed for this diagnosis. This was measured for each drug prescribed per diagnosis and these values were added together for all drugs prescribed per diagnosis. The range of this index goes from a low point of 1 divided by the number of drugs prescribed per diagnosis to a maximum of 1. A low index means that all drugs are equally often prescribed while 1 means that there is only one drug prescribed. The HHI was multiplied by 100 for computational reasons.

Analyses

Because our data are hierarchically structured, multi-level models are the appropriate statistical approach (Leyland and Groenewegen, 2003; Snijders and Bosker, 1999; Hox, 1995). In our case the hierarchy is as follows: family physicians are nested within practices and patients are nested within family physicians. Ideally the data would be analyzed with these three levels. A multi-response model with the proportion of patients receiving different types of drugs as an outcome would be appropriate. Data would be analyzed by diagnosis and the total variation would be calculated by adding the variation for all types of drugs. Although in theory this is the ideal approach, in practice it turned out to be impossible. The reason is that there are not enough data because of too few patients for each type of drug. Therefore, a different method was chosen.

As a dependent variable we used a measure of concentration, the HHI. The concentration index is not a characteristic of a single patient but of the aggregate of patients with the same diagnosis. It is measured per diagnosis per physician. Therefore the level of the patients is not relevant. The diagnoses per physician are populations, consisting of patients, with certain characteristics, including diagnosis, which are important in our analyses. These populations are nested within family physicians and therefore modeled as a level (see appendix 6B for the full model). The total variation in prescription is separated into three parts: a part due to differences between populations, a part due to differences between family physicians,

and a part due to differences between practices (Leyland and Groenewegen, 2003; Diez Roux, 2002).
Since different family physicians have populations with the same diagnoses the different HHIs are dependent between family physicians. What family physicians can prescribe does also depend on the diagnosis, and if they have populations with the same diagnoses, it is more likely that they will act the same. Therefore we produced a model, which allows for dependence between observations. The estimated means of the HHI will be presented.
To account for differences in patient population between family physicians, we included the mean age and sex of the patient populations of a family physician. Furthermore the number of patients per diagnosis as well as the number of different drugs available per diagnosis, measured as the total number of different drugs prescribed by all family physicians per diagnosis, the number of months from 1987 until the guidelines were introduced, and the number of family practitioners working in a practice were centered around a meaningful value, so that the estimated HHI has an interpretable meaning, and included in the model.

Testing hypotheses
The hypothesis that variation in prescription has decreased with the introduction of guidelines was tested by comparing the difference in HHIs between the reference and the guideline group for DNSGP1 and DNSGP2. If guidelines reduce variation:

- The HHI for the guideline group would have increased between DNSGP1 and DNSGP2 while the HHI for the reference group remained the same, increased less, or decreased (results are in table 6.5, A and B).

The hypothesis that variation in prescription for the guideline group has decreased more, for family physicians in single-handed practices than for family physicians in group practices, was tested by examining differences between family physicians in single-handed and group practices, for the guideline and reference group separately. And also if guidelines work differently for family physicians in group practices and for family physicians in single-handed practices:

- There would be a difference in HHI between single-handed practices and group practices for both the guideline and the reference group during DNSGP1. The HHI would be higher for the group practices (results are in table 6.7, E and F).
- There would be a difference in HHI between single-handed practices and group practices for the reference group during DNSGP2. The HHI would be higher for the group practices (results are in table 6.7, G).

- The difference in HHI for the guideline group between DNSGP1 and DNSGP2 for single-handed practices would be large compared to this difference for group practices (results are in table 6.7, L and K).
- The difference in HHI for the guideline group between DNSGP1 and DNSGP2 for single-handed practices would be large compared to the difference for the reference group (results are in table 6.7, L and J).
- The difference in HHI for the guideline group between DNSGP1 and DNSGP2 for group practices would be comparable to the difference for the reference group (results are in table 6.7, K, I and M).

6.3 Results

A summary of hypotheses, tests, and results can be found in table 6.8.

Is variation reduced after guidelines are introduced?
Table 6.4 and table 6.5 show that the HHI is higher during DNSGP1 than during DNSGP2. More different drugs were prescribed per diagnosis during DNSGP2. This is true for both the guideline and the reference group (table 6.5, A and B). The difference in HHI between the reference and the guideline group decreased, because the HHI decreased more for the reference than for the guideline group (table 6.5, C, D). The number of different drugs prescribed per diagnosis increased more for the reference than for the guideline group. The hypothesis that variation decreased after the introduction of guidelines was not confirmed. The results indicate that guidelines did not reduce variation in drug prescription, but tempered the increase of variation in prescription.

Table 6.4 Mean HHI for the reference and guideline group during DNSGP1 and DNSGP2

Description of the group	Mean HHI (standard error)
DNSGP1, reference group	82.7 (0.8)
DNSGP1, guideline group	74.3 (0.7)
DNSGP2, reference group	73.6 (1.0)
DNSGP2, guideline group	68.4 (0.9)

Table 6.5 Differences between the HHI for the guideline and the reference group during DNSGP1 and DNSGP2

Description of the difference	Difference	p-value
A: DNSGP1, reference group AND DNSGP2, reference group	9.1	<0.001
B: DNSGP1, guideline group AND DNSGP2, guideline group	5.9	<0.001
C: DNSGP1, reference group AND DNSGP1, guideline group	8.3	<0.001
D: DNSGP2, reference group AND DNSGP2, guideline group	5.2	<0.001

Is there a difference between group and single-handed practices?

For all groups the HHI is higher during DNSGP1 than during DNSGP2 (table 6.6).

First the assumption was tested that group practices have local standards. This implies that there is a difference in HHI between single-handed practices and group practices for the guideline group as well as for the reference group. Table 6.7 shows that there is no difference for the guideline group (F), but there is a difference for the reference group (E). In line with the assumption, the HHI is higher for the group practices, indicating that less different drugs are prescribed in group practices. Table 6.7 (G) shows that there is no difference in HHI between family physicians in single-handed practices and family physicians in group practices for the reference group during DNSGP2.

The difference in HHI for the guideline group between DNSGP1 and DNSGP2 for single-handed practices was not found to be large compared to this difference for group practices (table 6.7, K and L). Table 6.7 (L and J) also shows that the difference in HHI for the guideline group between DNSGP1 and DNSGP2 for single-handed practices is not large compared to the difference for the reference group. Finally, it was found that the difference for the guideline group between DNSGP1 and DNSGP2 for group practices is comparable to the difference for the reference group (table 6.7, K, I and M).

The results in table 6.7 (I,J,K,L) generally give indications of an increase in the total number of drugs prescribed for the reference as well as the guideline group, for both single-handed and group practices. It also shows that the number of drugs increased more for the reference group compared to the guideline group and that this effect applies most to single-handed practices.

Between DNSGP1 and DNSGP2 the difference between single-handed and group practices reversed. During DNSGP1 there were less different drugs prescribed in group practices, while during DNSGP2 there were less different drugs prescribed in single-handed practices. Based on these results,

we conclude that guidelines tempered the increase in the number of different drugs prescribed primarily for family physicians working in single-handed practices. The hypothesis that variation in drug prescription has decreased more for family physicians working in single-handed practices than for family physicians working in group practices, was not confirmed.

Table 6.6 Mean HHI for the reference and guideline group, single-handed and group practices during DNSGP1 and DNSGP2

Description of the group	Mean HHI (standard error)
DNSGP1 reference group, group practice	83.8 (1.0)
DNSGP1 reference group, single handed practice	79.3 (1.1)
DNSGP1 guideline group, group practice	75.1 (0.9)
DNSGP1 guideline group, single handed practice	72.5 (1.1)
DNSGP2 reference group, group practice	71.4 (1.4)
DNSGP2 reference group, single handed practice	73.9 (1.2)
DNSGP2 guideline group, group practice	65.8 (1.2)
DNSGP2 guideline group, single handed practice	70.1 (1.2)

Table 6.7 Differences between the mean HHI for the guideline and the reference group, single-handed and group practices during DNSGP1 and DNSGP2

Description of the difference	Difference	p-value
E: DNSGP1 reference group, group practice AND DNSGP1 reference group, single handed practice	4.5	=0.002
F: DNSGP1 guideline group, group practice AND DNSGP1 guideline group, single handed practice	2.5	=0.06
G: DNSGP2 reference group, group practice AND DNSGP2 reference group, single handed practice	-2.6	=0.16
H: DNSGP2 guideline group, group practice AND DNSGP2 guideline group, single handed practice	-4.3	=0.01
I: DNSGP1 reference group, group practice AND DNSGP2 reference group, group practice	12.4	<0.001
J: DNSGP1 reference group, single handed practice AND DNSGP2 reference group, single handed practice	5.3	<0.001
K: DNSGP1 guideline group, group practice AND DNSGP2 guideline group, group practice	9.2	<0.001
L: DNSGP1 guideline group, single-handed practice AND DNSGP2 guideline group, single-handed practice	2.4	=0.06
M: I and K	3.2	=0.13

Table 6.8 Description of hypotheses, testing and results

General hypothesis	Assumption	Test	Result in table	Test result in line with expectation?
Variation in prescription has decreased with the introduction of guidelines		The HHI for the guideline group would have increased between DNSGP1 and DNSGP2	Table 6,.5, B	No
		The HHI for the reference group remained the same, increased less, or decreased	Table 6.5, A	Yes
Variation in prescription has decreased more for family physicians working in single-handed practices than for family physicians working in group practices	Group practices have local standards	A difference in HHI between single-handed practices and group practices for both the guideline (1) and the reference group during DNSGP1 (2). The HHI would be higher for the group practices (3)	Table 6.7, E and F	No (1) Yes (2) Yes (3)
		A difference in HHI between single-handed practices and group practices for the reference group during DNSGP2 (1). The HHI would be higher for the group practices (2)	Table 6.7, G	No (1) No (2)
	Standards (guidelines) in single-handed practices (group practices already used local standards)	The difference in HHI for the guideline group between DNSGP1 and DNSGP2 for single-handed practices is large compared to this difference for group practices	Table 6.7, L and K	No

- table 6.8 continues -

- table 6.8 continued -

General hypothesis	Assumption	Test	Result in table	Test result in line with expectation?
		The difference in HHI for the guideline group between DNSGP1 and DNSGP2 for single-handed practices is large compared to the difference for the reference group	Table 6.7, L and J	No
		The difference in HHI for the guideline group between DNSGP1 and DNSGP2 for group practices is comparable to the difference for the reference group	Table 6.7, K, I and M	Yes

6.4 Conclusions and discussion

Due to processes of global rationalization, more uniformity occurs throughout the world. Since those rationalization processes also take place in medicine, for instance guidelines are introduced for family physicians in The Netherlands, more uniformity is to be expected.
It was found in this study that although more different drugs are prescribed (lower HHI) during DNSGP2 for both the reference and the guideline group, the change is higher for the reference group (table 6.4). The overall increase in the number of different drugs prescribed can be explained by the fact that there are far more different drugs available during DNSGP2 than during DNSGP1. Moreover, the pharmaceutical industry influences prescription by introducing more different drugs onto the market, and patients might ask for certain drugs, resulting in more different drugs prescribed during DNSGP2. This would lead to less uniformity in prescribing. Guidelines themselves could also be a cause of more variation in prescription for in some of the guidelines it is recommended to start with one drug and then follow it with another. That the increase in the number of different drugs prescribed is lower for the guideline group, indicates that guidelines temper the effect of the availability of more drugs.

We found no significant difference between the reference group and the guideline group for family physicians in group practices between DNSGP1 and DNSGP2. This suggests that guidelines did not have an effect on the variation for family physicians in group practices. In single-handed practices a significant increase in variation was found for the reference group, while no significant increase was found for the guideline group. In group practices a significant increase in variation was found for both groups. Therefore it was concluded that in line with the hypothesis, guidelines primarily had an effect on variation in single-handed practices.

Our study demonstrates that although guidelines seem to temper an increase in variation, the number of different drugs did not decrease. The importance of reducing the number of different drugs prescribed lies in the idea that family physicians become more familiar with the drugs, which results in better quality for patients and lower health care expenses (Denig et al., 1988; Hill-Smith, 1996).

In this study we tested whether more uniformity in prescription, in other words less variation in drugs prescribed, is created after the introduction of guidelines. In order to do so, we analyzed data from two national medical record studies, one held before the introduction of NHG-guidelines, and the other held after introduction of those guidelines. Although this sounds straightforward, it was complicated to analyze. Two developments go together in the analysis.

First, the number of available drugs has increased by almost 10% in the period between 1995 and 2001 (Medicines Evaluation Board 1999, 2005), which would imply an increase in the variation in prescription. The increasing number of available drugs is in contradiction to the tendency to more uniformity. However, it is rational for a pharmaceutical company to introduce new drugs, and maybe even more rational to replicate drugs developed by a competitor in a slightly different way. In that way there is uniformity in what they produce, although there are many different appearances. It would be interesting to examine whether guidelines somehow influence the pharmaceutical industry. For example would there be less innovation for diagnoses with guidelines for prescription drugs?

Second, guidelines were introduced which would imply a decrease in the variation in prescription. All guidelines are not introduced at the same time, but over a 15 year period. With respect to variation the effect of the introduction of a guideline would be a certain amount of variation at the beginning, more variation at a later stage when more physicians are beginning to work according to the guideline, and less variation when the introduction is completed and all physicians are working according to the guidelines (Coleman et al., 1966). Variation is thus influenced by the stage at

which guidelines are introduced in this process. In order to control for the different stage in which each guideline might be found, the number of months it existed since 1987 until the guidelines were introduced, were included in the analyses. In spite of these complications affecting the analysis, a measure had to be chosen that could be used to compute variation in prescription.

One possibility would be to count the number, or range of different drugs prescribed, but this measure is insensitive to the number of times a drug is prescribed. For example, the range is two if both drugs are each prescribed ten times, but also when one drug is prescribed ten times and the other is prescribed a hundred times. Measures of concentration are sensitive to the number of times a drug is prescribed. Therefore the Herfindahl-Hirschman Index (HHI), which is a measure of concentration, was used; the higher the HHI, the more concentrated the 'market', or the less different drugs are often prescribed.

As already stated in the introduction, the guidelines in this study are followed voluntarily. They are followed to reduce uncertainty or because peers can exert pressure. The reduction of uncertainty would apply most to family physicians in single-handed practices. Peer pressure would be exerted in group practices. Nevertheless there can be other reasons for following guidelines. Guidelines, including those developed by the profession itself, can be used by outside parties to judge treatment.

Third parties can hold the profession accountable to its own guidelines. This makes the guidelines, which are in essence voluntarily, regulative, with sanctions for those who do not comply. Non-compliance to guidelines can be used in malpractice suits (Timmermans and Berg, 2003). It has been reported that guidelines were used as evidence in malpractice suits in the USA, mostly against an individual physician (Hyams et al., 1995). In the Netherlands, examples exist of family physicians being sued for medical behavior, which was not inline with the guideline (Crul and Legemaate, 2005). As yet they were never found guilty and punished because they did not work according to the guidelines. The opinion of the family physicians has always been valued more than the guideline. Still, the family physicians had to explain why they did not follow the guidelines. To our knowledge there are no cases where a family physician had to explain why the guideline was followed. This makes following guidelines less voluntary because there is a risk when they are not followed.

In the introduction we argued that guidelines can be seen as instances of an increasing rationalization. We mentioned other terms that are used for these processes and probably the best term for processes of rationalization in medicine at this stage is 'de-professionalization' (Ritzer and Walczak, 1988;

Flap, 2001). Freidson (1975) stated that autonomy is the test of medicine as a profession. Physicians have the privilege to be free from control of non-professionals. Three claims form the basis of this privilege. Non-professionals are unable to evaluate or regulate work of professionals, due to the degree of skills and knowledge involved in professional work. Furthermore, it is claimed that professionals act in the best interest of their patients. They are the patients' advocates, even without supervision. The third claim is that the profession is self-regulatory; the physicians are the only ones to recognize and control deviant behavior. It is argued that: *"Its autonomy is justified and tested by its self-regulation"* (Freidson, 1975). If guidelines are developed within the profession they can be used as a tool by non-professionals to evaluate treatment provided by physicians. This undermines autonomy which was the basis for having the privileges of a profession; the profession is deprofessionalized. This brings us to a paradox. Guidelines have to be developed by the profession to keep their collective autonomy, to avoid non-professionals setting the rules. However, at the same time these professional guidelines can be used by others to evaluate treatment and thus limit autonomy.

Following Ritzer (2000), there is a danger of irrationality resulting from rationalization. Guidelines have been developed as a tool to reduce uncertainty and increase the quality of care. In order to guarantee quality, guidelines should be up-to-date. Their development and introduction takes time and therefore they will always be a little or even far behind. This would imply that family physicians working according to the guidelines do not necessarily provide the best care available. Guidelines recommend certain treatment, but it is always the family physician who has the responsibility for it.

It can be argued that variation is necessary in order to make improvement possible for without variation it will be impossible to know what treatment works best. There is a danger of a performance paradox when guidelines are followed. Following guidelines could lead to undesirable outcomes, for instance when they are not up-to-date, or even make improvements in medicine impossible.

To provide good quality care, guidelines should not be followed blindly, but should enhance clinical judgment. Variation could be reduced with the introduction of guidelines, but it should certainly not be expected to vanish. Less variation means more uniformity, it does not necessarily mean better quality.

Appendix 6.A ICPC codes: their occurrence in the (initial) study population and whether they belong to a guideline or not

Chapter	ICPC-code	Frequency of occurrence	Guideline yes or no
General and unspecified (A)	A97	1,309	n
	A04	1,278	n
	A85	406	n
	A03	380	y
	A77	346	n
	A06	225	n
	A12	214	n
Blood, blood forming organs, lymphatics, spleen (B)	B80	253	y
	B81	198	y
	B02	128	n
	B82	114	n
	B70	67	n
	B85	39	y
Digestive (D)	D06	1,004	n
	D12	625	n
	D73	514	y
	D87	494	n
	D01	457	n
	D93	436	y
	D02	410	y
	D11	376	y
	D84	369	n
	D03	272	n
Eye (F)	F70	596	n
	F72	279	n
	F13	197	n
	F05	195	n
	F71	187	n
Ear (H)	H81	1,199	n
	H71	825	y
	H70	645	y
	H72	320	y
	H02	225	y

- appendix 6.A continues -

- appendix 6.A continued -

Chapter	ICPC-code	Frequency of occurrence	Guideline yes or no
Circulatory (K)	K86	4,660	y
	K77	620	y
	K87	515	y
	K85	510	y
	K74	446	y
	K07	363	n
	K96	294	n
	K78	240	n
	K25	232	n
Musculoskeletal (L)	L03	1,814	y
	L92	1,021	y
	L02	1,012	n
	L01	949	n
	L15	918	y
	L99	858	y
	L17	826	n
	L86	773	y
	L04	760	n
	L08	702	y
	L18	627	n
Neurological (N)	N01	608	y
	N17	545	y
	N89	394	y
	N02	321	n
	N94	143	n
	N88	138	n
	N06	108	n
Psychological (P)	P76	1,560	y
	P06	1,018	y
	P74	557	y
	P03	484	y
	P78	412	n
	P02	349	n
	P29	152	n
	P70	147	y
	P20	146	n
	P00	124	n
	P99	112	n

- appendix 6.A continues -

- appendix 6.A. continued-

Chapter	ICPC-code	Frequency of occurrence	Guideline yes or no
Respiratory (R)	R74	2,756	n
	R05	2,181	n
	R78	1,306	y
	R96	1,128	y
	R75	1,118	y
	R97	987	y
	R95	776	y
	R21	564	y
	R81	507	n
	R76	482	y
	R02	390	n
	R09	202	n
Skin (S)	S88	1,630	n
	S74	1,522	y
	S03	1,063	n
	S87	565	y
	S18	552	n
	S11	433	y
	S93	420	n
	S79	418	n
	S02	381	n
	S96	366	y
	S82	353	n
	S86	350	n
	S76	327	y
	S97	304	y
Endocrine, metabolic and nutritional (T)	T90	2,720	y
	T93	529	y
	T86	271	y
	T92	184	y
	T82	111	n
	T08	96	n
	T83	76	n
	T85	67	y
	T27	58	n
Urology (U)	U71	1,558	y
	U04	223	y
	U01	154	y
	U02	140	y

- appendix 6.A continues -

- appendix 6.A. continued-

Chapter	ICPC-code	Frequency of occurrence	Guideline yes or no
	U95	124	y
	U05	93	y
	U27	90	n
	U99	88	n
	U00	69	n
	U70	49	y
	U07	46	y
	U76	39	n
	U75	34	n
Pregnancy, childbirth, family planning (W)	W11	1,502	y
	W78	324	y
	W12	166	y
	W14	159	n
	W90	95	y
	W15	88	n
	W01	56	n
	W02	54	n
Female genital system and breast (X)	X37	785	n
	X11	433	y
	X72	316	n
	X87	226	n
	X06	198	n
	X14	193	y
	X84	187	n
	X76	180	n
	X07	165	n
	X18	126	n
	X19	125	n
Male genital system (Y)	Y85	125	n
	Y06	107	n
	Y77	92	n
	Y75	84	n
	Y13	81	n
	Y07	69	n
	Y73	62	n

Appendix 6.B The multilevel model

$$\gamma_{ijk} \sim N(XB, \Omega)$$

$$\gamma_{ijk} = \beta_{0jk}x_{0jk} + \beta_{1jk}x_{1jk} + \beta_{2jk}x_{2jk} + \beta_{3jk}x_{3jk} + \beta_4 x_{4ijk} + \beta_5 x_{5ijk} + \beta_6 x_{6jk} + \beta_7 x_{7jk} + \beta_8 x_{8k} + \beta_9 x_{9ijk} + e_{10ijk}x_{10ijk} + ... + e_{143ijk}x_{143ijk}$$

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$
$$\beta_{1jk} = \beta_1 + v_{1k} + u_{1jk}$$
$$\beta_{2jk} = \beta_2 + v_{2k} + u_{2jk}$$
$$\beta_{3jk} = \beta_3 + v_{3k} + u_{3jk}$$

$$\begin{bmatrix} v_{0k} \\ v_{1k} \\ v_{2k} \\ v_{3k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma^2_{v0} & & & \\ 0 & \sigma^2_{v1} & & \\ 0 & 0 & \sigma^2_{v2} & \\ 0 & 0 & 0 & \sigma^2_{v3} \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{uo} & & & \\ 0 & \sigma^2_{u1} & & \\ 0 & 0 & \sigma^2_{u2} & \\ 0 & 0 & 0 & \sigma^2_{u3} \end{bmatrix}$$

$$\begin{bmatrix} e_{10ijk} \\ ..... \\ e_{143ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e10} & & \\ 0 & ..... & \\ 0 & ..... & \sigma^2_{e143} \end{bmatrix}$$

γ_{ijk} = dependent variable (HHI, per diagnosis per physician)
i = 'patient populations' (an identifier for each diagnosis per physician)
j = physicians
k = practices

$\beta 0_{jk}$ = mean HHI and variance parameters (DNSGP1, reference group)
$\beta 1_{jk}$ = mean HHI and variance parameters (DNSGP1, guideline group)
$\beta 2_{jk}$ = mean HHI and variance parameters (DNSGP2, reference group)
$\beta 3_{jk}$ = mean HHI and variance parameters (DNSGP2, guideline group)
x_4 = number of different drugs available per diagnosis (centred)
x_5 = number of patients per diagnosis (centred)
x_6 = mean age of the patient population per physician (centred)
x_7 = mean gender of the patient population per physician (centred)
x_8 = number of physicians per practice (centred)
x_9 = number of month from 1987 until guidelines were introduced per diagnosis (centred)
x_{10}.....x_{143} = diagnosis

v = level - three variance (practice level)
u = level - two variance (physician level)
e = measurement error per diagnosis

7

Do decision support systems influence (variation in) prescription?

This article was submitted as:
De Jong JD, Groenewegen PP, Spreeuwenberg P, Westert GP, De Bakker DH. Do decision support systems influence (variation in) prescription?

Abstract

Translating scientific evidence into daily practice is problematic. All kinds of intervention strategies, using educational and/or directive strategies, aimed at modifying behavior, have evolved, but have been found only partially successful. In this article the focus is on (computerized) decision support systems (DSSs). DSSs intervene in physicians' daily routine, as opposed to interventions that aim at influencing knowledge in order to change behavior. We examined whether general practitioners (GPs) are prescribing conform the advice given by the DSS and whether there is less variation in prescription when the DSS is used.

Data were used from the Second Dutch National Survey of General Practice (DNSGP2), collected in 2001. A total of 82 diagnoses, 74,9811 contacts, 133 physicians, and 85 practices was included in the analyses. GPs using the DSS daily were compared to GPs who do not use the DSS. Multilevel analyses were used to analyze the data. Two outcome measures were chosen: whether prescription was conform the advice of the DSS or not, and a measure of concentration, the Herfindahl-Hirschman Index (HHI).

GPs who use the DSS daily prescribe more according to the advice given in the DSS than GPs who do not use the DSS. Contradictory to our expectation there was no significant difference between the HHIs for both groups: variation in prescription was comparable.

We studied the use of a DSS for drug prescribing in general practice in the Netherlands. The DSS is based on guidelines developed by the Dutch College of General Practitioners and implemented in the Electronic Medical Systems of the GPs. GPs using the DSS more often prescribe conform the advice given in the DSS compared to GPs not using the DSS. This, however, did not mean that variation is lower; variation is the same for GPs using and for GPs not using a DSS. Implications of the study are that DSSs can be used to implement guidelines, but that it should not be expected that variation is limited.

7.1 Introduction

According to Grol and Grimshaw (2003): *"One of the most consistent findings in research of health services is the gap between evidence and practice. Results of studies in the USA and the Netherlands suggest that about 30-40% of patients do not receive care according to present scientific evidence, and about 20-25% of care provided is not needed or is potentially harmful"*. These are causes of unwanted variations in health service provision. Introducing evidence and clinical guidelines into daily practice is problematic. All kinds of intervention strategies, using educational and/or directive strategies, aimed at modifying behavior, have been developed, but found only partially successful (e.g. Denig et al., 1988; Grimshaw et al., 2001; Grol and Grimshaw, 2003).

This article focuses on computerized decision support systems (DSSs). The use of DSSs is a way to improve physicians performance (Garg et al., 2005). Computerized systems rationalize medical practice, by implementing evidence into medical practice, and in doing so decrease variation in medical practice (Berg, 1995; Kawamoto et al., 2005). They are meant to improve the capacity of physicians to make better decisions whilst the complex problems physicians deal with surpass their cognitive capacity (Berg, 1995). Computerized systems intervene in physicians' daily routine, as opposed to interventions that aim at influencing knowledge in order to change behavior. This was found to be effective in changing behavior for instance in blood test ordering (Van Wijk, 2000).

In the Netherlands a DSS for prescribing drugs was introduced for general practitioners (GPs) in 1998. This was done to implement professional guidelines regarding the prescription of drugs. Prescription of drugs has great importance in modern health care; the development of new and effective drugs has contributed to the increase of the health status in the OECD countries (OECD, 2005). In modern health care, appropriate prescribing of new and expensive drugs is a big challenge as the number of available drugs is increasing, making choice more complex.

Prescription of drugs is influenced by the pharmaceutical industry (e.g. Wazana, 2000), the professional environment (e.g. Denig et al., 1988), physicians' and patients' habits (e.g. Coscelli, 2000; Lundin, 2000); all factors that are not directly related to treatment outcome. Denig (1994) showed that physicians choose from a limited evoked set of drugs which comes up in their minds, given the health problem of the patient. This evoked set is influenced by advertising, and (continuing) education. More than 70% of

drugs prescribed are drugs from the evoked set (Wierenga et al., 1989). This 70% is 'top of the mind' drugs, meaning that, given the diagnosis, this drugs comes to the mind directly, based on habits. Physicians do not consider all possible treatment options, but chose from approximately two to five different options (Denig, 1994). In order to change the drugs physicians prescribe, the evoked set of physicians has to be changed, or its role in decision making should be changed. the evoked set should become less important or should even be avoided. That is where DSSs come in. Instead of relying on the evoked set of drugs for physicians to chose one in a specific situation, the DSS proposes one or more drugs of preference, based on characteristics of the patient and professional guidelines. The introduction of DSSs is expected to result in more rational prescribing and less unexplained and unwanted variation in prescribing between physicians (Denig, 1994; Kamps, 1999; Carthy et al., 2000; Jaye and Tilyard, 2002).

Educational interventions and giving feedback as methods to enhance evidence based practice rely on changing the evoked set in the minds of physicians. DSSs, however, directly affect physicians' routines. The strategy of changing physicians' routines has shown to be effective in for instance blood test ordering. Zaat et al. (1992 found that by simply changing the application form, the amount of unnecessary test ordering was drastically decreased. After changing the form back into its original format the physicians showed their old test ordering behavior.

DSSs can be useful tools in medical practice; they can assist physicians in making evidence based decisions and can therefore improve the quality of care. Quality of care will only improve with the use of DSSs if the recommendations are evidence based and the DSS is actually used as intended. DSSs do therefore not always improve clinical practice. Hunt et al. (1998) showed in their systematic review that most systems significantly improved clinical practice, but some (34%) did not. Lobach and Hammond (1997) found in their study on clinical practice in the USA that DSSs improved compliance with care standards. Shea et al. (1996) found evidence from a meta-analysis of randomized controlled studies that supports the effectiveness of data-driven computer-based reminder systems to improve prevention services in the ambulatory care setting. Johnston et al. (1994) concluded from their review that recommendations from DSSs can improve compliance with guidelines for preventive and acute care. Improvement of guideline adherence was also shown by Shiffman et al. (1999) in their systematic review on the effectiveness of computer-based guideline implementation systems. In a systematic review on the effects of DSSs on

physician performance and patient outcomes it was concluded that DSSs can enhance clinical performance for drug dosing, preventive care, and other aspects of medical care, but not convincingly for making the diagnosis (Hunt et al., 1998). Furthermore, they concluded that the effects on patient outcomes have been insufficiently studied. A study by Ramnarayan et al. (2006) suggests a promising role for DSSs in the reduction of diagnostic error. DSSs result in improved clinical practice if support is provided automatically as part of physicians' workflow, support is delivered at the time and location of decision making, actionable recommendations are provided and support is computer based (Kawamoto et al., 2005).

DSSs can influence physicians' behavior and at the same time influence variation in prescribing between physicians. Generally, variation is reduced by shared normative systems, rules and common frameworks of meaning (Scott, 2001). That is, variation decreases when people adhere to the same norms and rules, or make use of the same frameworks of meaning. DSSs can be considered frameworks of meaning, people do not have to be aware of what a DSS does when they use it and DSSs do not work via behavioral confirmation or sanctioning as norms and rules do. As such they can decrease variation amongst those who make use of the DSS.

In this article we will study a DSS which is used by GPs in the Netherlands as a tool to give advice on prescription when the diagnosis is given. The DSS proposes a prescription, given the diagnosis of the patient, taking into account age, sex and co-morbidity. As mentioned before, four features are associated with the ability to improve clinical practice, it is part of the workflow, support is delivered at the time and location of decision making, actionable recommendations are provided and it is computer based (Kawamoto et al., 2005). The DSS for GPs in the Netherlands meets these four features (Wolters et al., 2003). In the advice patient characteristics, like age, sex, co-morbidities and other drugs prescribed are taken into account. The DSS is integrated in the Electronic Medical Systems (EMS) of the GPs. The advice given is derived from professional-guidelines. These guidelines are developed by the Dutch College of General Practitioners (NHG) and are widely accepted (Grol, 1989). Wolters et al. (2002) studied the use of the DSS by GPs in the Netherlands. They found that having access to the DSS increased from 20% in 1999 to 71% of GPs in 2001, and daily use from 11% to 40%. More information on DSSs in the Netherlands can be found in box 7.1.

Box 7.1 Decision Support Systems (DSS) in the Netherlands

There are two preconditions for the DSS to have an effect: GPs had to have a computer with an EMS, and they had to use the Electronic Medical Record (EMR). In 2001 97% of all GPs had a computer with an EMS and 94% of these GPs used the EMR (Wolters et al., 2002). In order to give advice a diagnosis coded with the International Classification of Primary Care (ICPC) is needed. In the EMR a thesaurus with a search option is available to search for a possible ICPC-code. Besides the ICPC-code, information on patient characteristics is needed. This has to be available in the EMR in the appropriate place (Wolters et al., 2003). GPs are not required to use the DSS and they vary in how frequent they use it.

In this article we will examine the influence of computerized decision support on prescribing by GPs in the Netherlands. Two questions will be addressed.

1 *'Do GPs who use the DSS on a regular basis prescribe more often conform the advice given by the DSS?'*

2 *'Is there less variation in prescription among GPs who use the DSS on a regular basis compared to those who do not?'*

Hypotheses

An effective method for changing physicians behavior, or implementing new techniques, is to make them save time if they change or comply (Payne, 2000). An effective DSS minimizes the effort required by physicians to receive and act on the recommendations (Kawamoto et al., 2005). DSSs in the Netherlands give GPs information at the time a decision has to be made. Thus, GPs do not have to recall information from the mind, and can more easily prescribe according to professional guidelines, even without having to make a conscious choice. *It is expected that those who use the DSS, prescribe according to the guidelines that are incorporated in the DSS.*

The more GPs use the DSS, the more their prescriptions will be conform the advice given in the DSS. If GPs use the DSS there will be less variation in prescription. *It is expected that there is less variation in prescribing between GPs using the DSS compared to GPs not using the DSS.*

7.2 Data and methods

Data

Data were used from the Second Dutch National Survey of General Practice (DNSGP2), collected in 2001. Data were collected on contacts, patients, GPs and practices. For a description of the data-collection we refer to Westert et al. (2005). DNSGP2 is used for many different research questions and analyses. The basis of DNSGP2 was an extraction of the electronic medical records from 103 general practices during one year. This contained the ATC-coded prescription data. DSS use was assessed through a questionnaire filled in by 191 GPs working in the 103 practices in March 2001, 188 questionnaires could be used. The response was 96%. In the questionnaire it was asked if the GPs had a DSS and how many times this was used. The questionnaire was not designed for this specific article and consequently the GPs were not aware of this study. The current study is a secondary analysis of the DNSGP2 data base. The total DNSGP2 population can be found in table 7.1.

Table 7.1 Selection of GPs used in this article, the study population of DNSGP2 and the total population of Dutch GPs at the time of data collection. Distribution by sex, age and type of practice

	GPs included in this article from DNSGP2 (%)	**Study population DNSGP2* (%)**	**Total GP population in the Netherlands in 2001* (%)**
Sex:			
male	78	73	74
female	22	27	26
Age:			
<35	1	4	6
35-39	15	15	15
40-44	15	18	20
45-49	32	30	25
50-54	25	25	22
55-59	12	8	10
60+	1	1	2
Type of practice:			
single handed	51	31	43
shared	24	28	33
group/health centre	26	42	25

* Schellevis et al., 2004

Table 7.2 Description of included data (absolute numbers)

	Included data*	DSS daily users	DSS non users/ owners
Number of practices	85	29	33
Number of physicians	133	44	43
Number of patients	749,811	251,587	242,786
Number of diagnoses	82	82	82
Type of practice:			
single-handed	43	14	14
duo	20	5	10
group	22	10	9
TOTAL	85	29	33
Pharmacy:			
pharmacy included	9	3	1
with pharmacy	1	0	0
no pharmacy	73	26	32
missing	2	0	0
TOTAL	85	29	33
EMS:**			
microhis	23	5	16
promedico	29	14	5
elias	22	7	10
arcos	11	3	2
TOTAL	85	29	33
Number of physicians per practice (mean (st.dev.) [range])	1.6 (1.1 [1-5]	1.5 (1.1) [1-5]	1.3 (0.7) [1-4]
Number of different drugs per diagnosis in DSS (mean (st.dev.) [range])	4.0 (2.4) [1-11]	4.0 (2.4) [1-11]	4.0 (2.4) [1-11]
Number of different drugs prescribed per physician per diagnosis (mean (st.dev.) {range])	7.0 (6.7) [1-156]	6.7 (6.11) [1-66]	7.2 (6.9) [1-77]

* selected based on whether diagnoses are in the DSS

** the type of formulary used was related to the type of EMS. Type of EMS was included in the analyses

Method

The DSS includes an advice for 172 diagnoses. These diagnoses were included in our analyses if in at least 1,250 contacts a prescription was given for this diagnosis. A minimum was set to exclude diagnoses that hardly appear in GPs practices. The minimum was set at 1,250 contacts for computational restrictions. A total of 82 diagnoses were included in the analyses. Contacts were included only if type of medicine (ATC-5) and diagnosis (ICPC coded) were known. All participating GPs have had a

coding training at inclusion in DNSGP2. Because of missing diagnoses sixteen practices were excluded. A description of the GPs in our analysis compared to the total DNSGP2 study group as well as to the total population in the Netherlands can be found in table 7.1. There is an overrepresentation of single handed practices in our population. A total of 749,811 contacts, 133 physicians, and 85 practices was thus included in the analyses (table 7.2).

For this study two extremes were created: those GPs who use the DSS daily and those GPs who do not use the DSS at all. The latter group both includes GPs who do not have a DSS and GPs who do have a DSS but do not use it. An important difference between both groups is in the Electronic Medical System they use. We have controlled for this in the analyses.

Specifics in the methods differ between the two questions addressed in this article. Therefore, the specific methods will be described for each question separately.

Are GPs prescribing conform the advice given by the information system?
Dependent variable
As dependent variable a measure was used that indicated whether a prescription was conform DSS or not. For each diagnosis (ICPC coded) a list of prescriptions (ATC-5 coded) advised in the DSS was used (Wolters et al., 2003). These proposed prescriptions were compared with the actual prescriptions for patients. If prescriptions for patients were similar to the prescription in the list it was conform the DSS, if it was different it was not conform the DSS. This was done for both GPs using the DSS and for GPs not using the DSS.

Model
Multilevel analyses were used to take into account the structure of the data: contacts are nested within GPs and GPs are nested within practices (Leyland and Groenewegen, 2003; Snijders and Bosker, 1999; Hox, 1995). The model therefore consists of three levels; the contact, the GP and the practice level. Because the dependent variable was dichotomous, a multilevel logistic regression was performed (see appendix 7C for the full model). The percentage of contacts in which patients receive prescriptions conform the DSS was computed, taking the diagnosis into account and correcting for the specific EMS, practice type, and the number of GPs in a practice. Besides, a variable was included indicating whether or not a practice was dispensing for all or part of the population. This was done because medication

prescribed by medical specialists was also included in the data collection for GPs in a dispensing practice, but are not recognized as such. Hence, in these practices part of the prescribing decisions might have been made by medical specialists. For the model to have interpretable meaning, variables were centered around their means. Diagnoses were not only included in the fixed part of the model, but also in the random part of the model. The latter was done because the observations are not independent; different physicians face patients with the same diagnoses and hence the same advice of the DSS which has an effect on what is prescribed.

Is there less variation in prescription when the information system is used?

Dependent variable

As dependent variable to measure variation in prescribing a measure of concentration was used, the Herfindahl-Hirschman Index (HHI) (Den Hartog and Janssen, 2000; Zwanziger et al., 2000). This measure was based on whether the prescriptions for a given diagnosis were distributed over a large number of different drugs or only one or a few drugs. The kind of drug was identified based on the ATC-5 code. The HHI was measured as $\sum(a/b)^2$; where *a* is the number of times a specific drug was prescribed per diagnosis per GP and *b* is the total number of times any drug is prescribed for this diagnosis per GP. This was measured for each drug prescribed per diagnosis per GP and these values were summed for all drugs prescribed per diagnosis per GP. The range of this index goes from a low point of 1 divided by the number of drugs prescribed per diagnosis to a maximum of 1. A low index means that all drugs are equally often prescribed while 1 means that there is only one drug prescribed. The higher the concentration the less variation in drug prescription there is. The HHI was multiplied by 100 for ease of interpretation of the coefficients.

In contrast to the previous model, the HHI, the concentration index, is not a characteristic of a single patient contact but of the aggregate of patient contacts with the same diagnosis. It is measured per diagnosis per GP.

Model

The data are hierarchically structured and therefore multilevel models are the appropriate statistical approach (Leyland and Groenewegen, 2003; Snijders and Bosker, 1999; Hox, 1995). The model consists of three levels; diagnoses, GPs and practices. Since different GPs face patients with similar diagnoses, again a model, which allows for dependence between observations, was chosen (see appendix 7D for the full model).

To account for differences in patient population between GPs the mean age and sex of the patient population of a GP were initially included. These variables did only have a small effect and were excluded from the final analyses. Corrections were made for the different diagnoses by including all diagnoses in the random part of the model. They were, as opposed to the previous analysis, not included in the fixed model because the HHI was measured per diagnosis. As mentioned before, the HHI was measured per GP per diagnosis and therefore each diagnosis occurred only once for each GP. Furthermore, the specific EMS, practice type, the number of GPs in a practice, and the number of different drugs suggested in the DSS per diagnosis were included. A variable was included indicating whether or not a practice was dispensing for all or part of the population. All variables were centered to give the model interpretable meaning.

7.3 Results

Are GPs prescribing conform the advice given by the information system?
Two extremes were compared: GPs using the DSS daily were compared to GPs who do not use the DSS. The latter group involved GPs who do not have the DSS as well as GPs who do have the DSS but say they never use it. In line with our expectation GPs who use the DSS daily prescribe more according to the advice given in the DSS than GPs who do not use the DSS (table 7.3). Still, prescription by GPs who do not use the DSS is in line with advices given in the DSS in 75% of all prescriptions, against 89% for daily users. In appendix 7.A the total model is presented. The fixed effects show that there are differences between diagnoses in whether GPs prescribe conform the advice given in the DSS.

Table 7.3 Proportion of prescriptions conform DSS, corrected for EMS, type of practice, dispensing practice, number of GPs in a practice (full model in appendix 7.A)

	Conform DSS	**Difference**
DSS daily users	0.89 (0.06)	
DSS non users/havers	0.75 (0.05)	0.14 (p=0.04)

Is there less variation in prescription when the information system is used?
Contradictory to our expectation there is no significant difference between the HHIs, the indicator we used to measure variation, for GPs using the DSS

daily and GPs who do not own or use the DSS (table 7.4). Apparently the variation in prescription is comparable for both groups. In appendix 7.B the total model is presented. The fixed part of the model shows a relatively strong effect of the number of different drugs advised in the DSS. If this number is larger there is more variation in prescribing. From the random part we learn that the HHI shows differences between diagnoses. Comparing groups of diagnoses, variation in HHI is relatively low in diagnoses related to the skin and relatively high in the diagnoses from the chapter 'general' of the ICPC, the first five diagnoses of the list.

Table 7.4 Variation in prescribing, as measured by HHI, corrected for EMS, type of practice, dispensing practice, number of GPs in a practice, number of different drugs advised in the DSS (full model in appendix 7.B)

	HHI (st. error)	**Difference**
DSS daily users	40.3 (1.2)	
DSS non users/havers	41.4 (1.3)	1.1 (p=0.3)

7.4 Conclusions and discussion

GPs using a DSS are prescribing conform the advice given in the DSS more than GPs not using a DSS. This however, did not mean that variation is lower; variation is the same for GPs using a DSS and for GPs not using a DSS.
The DSS for prescribing by GPs in the Netherlands incorporates the professional guidelines of the Dutch College of General Practitioners. Therefore, the DSS can be seen as a way to implement guidelines. Guidelines and their incorporation in a DSS are part of a general trend towards rationalization of medicine in the Western world and one of the general implications of rationalization processes is a decrease in variation. The use of DSSs was hypothesized to decrease variation between physicians, because physicians who use a DSS that is based on the same professional guidelines make use of the same cognitive framework.

Although GPs are using the DSS and thus the same cognitive framework, variation did not decrease. How can this be explained? DSSs give recommendations for prescribing certain drugs. The DSS for prescribing by GPs that we studied, advises several different drugs or recommends a stepwise treatment starting with one type of drug and changing that type of

drug later on when necessary. We were not able to take stepwise use of different drugs into account in our definition of conformity to the advice of the DSS. As a consequence, variation can be generated. GPs not using the DSS prescribe from their evoked set and probably do not use a stepwise treatment strategy.

Although variation does not necessarily mean that some patients receive bad quality of care, it does raise questions related to effectiveness, efficiency and equity (McPherson, 1990). Evidence on medical practice variations implicates that there might be inappropriate servicing, waste of resources and maybe even harm to patients (Evans, 1990).

In general, it can be questioned whether the use of a DSS would actually improve the quality of prescribing for four reasons. First, there is a danger of 'cookbook medicine' where physicians blindly follow the advice from the DSS. In order to provide good quality care, guidelines and recommendations through DSSs should not be followed blindly, but should enhance clinical judgment. Second, the recommendations in the DSS are only partly evidence based. Third, the advice given in the DSS can lead to undesirable outcomes when it is not up-to-date. The fourth reason relates to who is developing and promoting the DSS and for what reason. A DSS might be developed and promoted by third party payers who would be interested in cost containment, or by the pharmaceutical industry that might also be interested in the DSS as a marketing tool for their own products.

In the Netherlands the advice given in the DSS is derived from professional guidelines, developed by the Dutch College of General Practice. The formularies correspond to the NHG-guidelines for 65-70% (Kamps, 1999). The NHG-guidelines both reflect what is common in the profession and what should become common (Burgers, 1994; Meulenberg and Goudswaard, 1997). Only part of the NHG-guidelines published between 1989 and 2000 are evidence-based. They are mostly based on consensus between members of the study groups that are involved in the development of the guidelines (Tasche et al., 2001). Of 130 recommendations in 28 NHG-guidelines published between 1993 and 1997, 44% was evidence-based (Burgers et al., 2002). Scientific evidence is developing continuously and guidelines should therefore be changed according to the latest evidence on a regular basis. GPs can be ahead of changes in the guideline and already use new evidence before the guideline is updated (Tacken, 2005). Guidelines reflecting common practice, and GPs being ahead of guidelines can explain that GPs not using the DSS prescribe conform the advice. The fact that we did not

find a difference in variation between daily users and non-users of the DSS indicates that the introduction of the DSS did not lead to simple cookbook medicine.
This article does not question whether or not the professional guidelines, incorporated in the DSS, are adequate and whether or not prescribing quality is better for the users of the DSS. These are different questions from the one discussed in this paper and would require a guideline by guideline analysis of evidence and data. In our article we only test whether or not over a large domain of diagnoses, there is indeed more conformity to the advice of the DSS among daily users than among non-users.

One of the limitations of this study is that we do not know whether there has been a change in the behavior of GPs when they started using the DSS. GPs using the DSS were compared to GPs not using the DSS in the same time period. We could not perform a longitudinal analysis which would make it possible to detect a change in behavior. From the analysis performed, we conclude that GPs using the DSS prescribe according to it and thus comply with professional guidelines more than GPs not using the DSS. It is assumed that GPs change their prescribing behavior when they start using the DSS, but it is possible that they already did prescribe conform the advice. Another limitation is that coding the diagnosis is a condition to use the DSS. We used only data from practices coding the diagnoses in the ICPC-system. We needed these codes to be able to construct our dependent variable that indicates conformity to the advice, given by the DSS. It is possible that the non-users of the DSS that nevertheless use the ICPC-coding in their EMR are a positive selection.
We assume that the coded diagnosis reflects what GPs think is the diagnosis at the time they enter its code in the system. It is this diagnosis that steers their prescription decision, irrespective of whether or not this diagnosis was correct or incorrect.
Although the questions answered in this article are fairly simple, the analyses were complicated. For both questions it was necessary to take the diagnoses into account. Furthermore, variation in prescription is not measured easily. To overcome the first difficulty, we could have chosen to examine specific diagnoses. Analyzing every diagnosis separately and aggregating the results over all diagnoses would ignore the fact that physicians' prescribing decisions in one diagnosis are correlated to the same physicians' prescribing decisions in another diagnosis. Moreover, our model structure takes into account the fact that prescribing decisions of different physicians for the same diagnosis will be correlated as well. Moreover, we were testing a general hypothesis and having separate outcomes for each

(group of) diagnoses would ignore this. To take the specific diagnoses into account and the fact that the same diagnosis is encountered by different GPs, we used a model, taking into account that the same diagnosis can 'belong' to different GPs. The statistical model we developed has the additional advantage that it is able to identify (groups of) diagnoses with relatively low and high conformity to the DSS' advice, taking the data structure into account.
The second difficulty, how to measure variation in prescription, had several possible solutions. Counting the range, the number of different drugs prescribed, was one possibility (De Bakker et al., 2007). Disadvantage of the range, however, is its insensitivity to the number of times a drug is prescribed. For example, the range is two if both drugs are each ten times prescribed, but also when one drug is ten times prescribed and the other is prescribed a hundred times. Measures of concentration are sensitive for the number of times a drug is prescribed. Therefore, the Herfindahl-Hirschman Index (HHI), which is a measure of concentration, was used; the higher the HHI the less different drugs are prescribed.

This study implicates that physicians decision making regarding prescription can be influenced with computerized decision aids. Although we only studied the effect of the DSS for drug prescribing in general practice in the Netherlands, the result we found is in line with the results of international reviews. To improve the quality of prescribing these systems should be up-to-date and the advice should be evidence based. However, the regular use of a DSS did not result in lower practice variation.

This study demonstrates that GPs using a DSS are prescribing conform the advice given in the DSS more than GPs not using a DSS. This, however, did not mean that variation is lower; variation is the same for GPs using and for GPs not using a DSS. Implications of the study are that DSSs can be used to implement guidelines, but that it should not be expected that variation is limited. As DSSs can influence prescription, it is important to make explicit who is developing the DSS and for what reason.

Appendix 7.A Proportion of prescriptions conform DSS

Fixed effects	Estimate (st. error)
DSS use:	
DSS daily users	0.89 (0.06)
DSS non users/havers	0.75 (0.05)
EMR:	
EMR1 (centered)	0.42 (0.13)
EMR2 (centered)	0.28 (0.12)
EMR3 (centered)	0.14 (0.13)
Type of practice:	
Duo practice (centered)	-0.13 (0.10)
Group practice (centered)	0.02 (0.17)
Health centre (centered)	-0.06 (0.22)
Dispensing practice (centered)	-0.27 (0.16)
Number of GPs (centered)	0.05 (0.05)
Diagnoses:	
Allergy/allergic reaction nos (centered)	-0.21 (0.06)
No disease (centered)	-5.51 (0.36)
Iron deficiency anemia (centered)	1.21 (0.07)
Pernicious/folate deficiency anemia (centered)	-2.54 (0.10)
Stomach ache/stomach pain (centered)	1.37 (0.06)
Heartburn (centered)	1.40 (0.06)
Nausea (centered)	0.20 (0.07)
Diarrhea (centered)	-0.39 (0.07)
Constipation (centered)	0.60 (0.05)
Other presumed infections of digestive system (centered)	-2.07 (0.10)
Disease of esophagus (centered)	1.76 (0.06)
Disorders of stomach function/gastritis (centered)	1.43 (0.06)
Irritable bowel syndrome (centered)	-0.91 (0.05)
Anal fissure/perianal abscess (centered)	-2.24 (0.10)
Abnormal sensations of eye (centered)	0.27 (0.06)
Infectious conjunctivitis (centered)	1.07 (0.06)
Allergic conjunctivitis (centered)	0.05 (0.06)
Otitis externa (centered)	0.61 (0.05)
Acute otitis media/myringitis (centered)	0.35 (0.05)
Vertiginous syndrome/labyrinthitis/vestibulitis (centered)	0.82 (0.07)
Angina pectoris (centered)	0.55 (0.05)
Heart failure (centered)	-0.11 (0.05)
Uncomplicated hypertension (centered)	0.20 (0.04)
Hypertension with involvement target organs (centered)	-0.31 (0.05)
Transient cerebral ischemia (centered)	1.10 (0.07)
Hemorrhoids (centered)	-0.50 (0.06)

Rheumatoid arthritis and allied conditions (centered)	-0.43 (0.05)
Osteoarthritis of hip (centered)	-0.12 (0.07)
Osteoarthritis of knee (centered)	-0.19 (0.06)
Other osteoarthritis and allied conditions (centered)	-0.25 (0.07)
Osteoporosis (centered)	-0.48 (0.06)
Headache (centered)	-0.10 (0.05)
Tension headache (centered)	0.09 (0.07)
Restless legs syndrome (centered)	-1.90 (0.08)
Vertigo/dizziness (centered)	0.23 (0.06)
Parkinsonism/paralysis agitans (centered)	0.11 (0.08)
Epilepsy, all types (centered)	-3.38 (0.10)
Migraine (centered)	1.81 (0.06)
Disturbances of sleep/insomnia (centered)	1.53 (0.04)
Affective psychosis (centered)	-2.27 (0.11)
Anxiety disorder/anxiety state (centered)	0.81 (0.05)
Depressive disorder/anxiety/depression (centered)	0.42 (0.04)
Cough (centered)	-0.84 (0.04)
Sympt/complt sinus (incl. Pain) (centered)	-2.77 (0.12)
U.R.I. (head cold)/rhinitis nos (centered)	-1.80 (0.05)
Sinusitis acute/chronic (centered)	-0.38 (0.04)
Tonsillitis acute (centered)	0.52 (0.06)
Acute laryngitis/tracheitis/croup (centered)	-2.48 (0.11)
Acute bronchitis/bronchiolitis (centered)	-4.06 (0.07)
Pneumonia (centered)	-0.88 (0.06)
Chronic bronchitis/bronchiectasis (centered)	0.52 (0.06)
Emphysema/chronic obstructive pulmonary disease (centered)	0.76 (0.04)
Asthma (centered)	0.89 (0.04)
Hayfever, allergic rhinitis (centered)	0.66 (0.04)
Localised redness/erythema/rash of skin (centered)	-3.72 (0.17)
Herpes zoster (centert)	-0.13 (0.08)
Dermatophytosis (centered)	1.29 (0.05)
Moniliasis/monilia infection/candidiasis (centered)	-2.22 (0.07)
Other infectious skin dis.nec/erysipelas (centered)	-0.26 (0.06)
Impetigo (centered)	0.59 (0.07)
Seborrhoic dermatitis/other erythematous dermatoses (centered)	-0.64 (0.05)
Atopic dermatitis/eczema (centered)	0.85 (0.05)
Contact dermatitis/skin allergy (centered)	0.28 (0.04)
Psoriasis w/wo arthropathy (centered)	0.39 (0.06)
Acne (centered)	1.95 (0.07)
Urticaria (centered)	-1.03 (0.06)
Other disease skin/subcutaneous tissue (centered)	-1.77 (0.07)
Hyperthyroidism/thyrotoxicosis (centered)	0.76 (0.08)
Hypothyroidism/myxedema (centered)	2.06 (0.07)
Diabetes mellitus (centered)	0.40 (0.04)
Gout (centered)	1.21 (0.07)
Lipid metabolism disorder (centered)	1.35 (0.05)
Cystitis/other urinary infect. Non-venereal (centered)	0.68 (0.04)
Family planning/oral contraceptive (centered)	0.33 (0.04)

Menstrual pain (centered)	1.25 (0.09)
Menstruation excessive (centered)	0.78 (0.08)
Menstruation irregular/frequent (centered)	0.74 (0.09)
Menopausal sympt./complt. (centered)	-0.33 (0.05)
Urogenital candidiasis, thrush (proven/unproven) (centered)	-0.30 (0.05)
Vaginitis/vulvitis, non venereal nos (centered)	-0.77 (0.07)
Benign prostatic hypertrophy (centered)	0.48 (0.08)

Random effects	**Variance (st. error)**
Practice level:	
DSS daily users	0.00 (0.00)
DSS non users/havers	0.03 (0.02)
GP level:	
DSS daily users	0.11 (0.02)
DSS non users/havers	0.04 (0.02)
Diagnosis level:	
Pain: generalised/unspecified	0.99 (0.03)
Allergy/allergic reaction nos	1.03 (0.03)
No disease	1.00 (0.05)
Iron deficiency anemia	0.95 (0.03)
Pernicious/folate deficiency anemia	0.92 (0.04)
Stomach ache/stomach pain	0.96 (0.02)
Heartburn	0.95 (0.02)
Nausea	1.11 (0.04)
Diarrhea	0.98 (0.04)
Constipation	0.99 (0.01)
Other presumed infections of digestive system	1.02 (0.05)
Disease of esophagus	0.91 (0.02)
Disorders of stomach function/gastritis	0.97 (0.02)
Irritable bowel syndrome	1.02 (0.02)
Anal fissure/perianal abscess	1.03 (0.05)
Abnormal sensations of eye	0.92 (0.03)
Infectious conjunctivitis	1.03 (0.02)
Allergic conjunctivitis	0.97 (0.03)
Otitis externa	0.97 (0.02)
Acute otitis media/myringitis	1.03 (0.02)
Vertiginous syndrome/labyrinthitis/vestibulitis	0.94 (0.03)
Angina pectoris	1.00 (0.01)
Heart failure	1.03 (0.02)
Uncomplicated hypertension	0.98 (0.01)
Hypertension with involvement target organs	0.99 (0.02)
Transient cerebral ischemia	0.90 (0.03)
Hemorrhoids	1.02 (0.03)
Rheumatoid arthritis and allied conditions	1.03 (0.03)

Osteoarthritis of hip	0.96 (0.04)
Osteoarthritis of knee	1.00 (0.03)
Other osteoarthritis and allied conditions	1.01 (0.04)
Osteoporosis	1.04 (0.03)
Headache	1.03 (0.02)
Tension headache	1.06 (0.04)
Restless legs syndrome	1.03 (0.04)
Vertigo/dizziness	0.97 (0.03)
Parkinsonism/paralysis agitans	1.02 (0.05)
Epilepsy, all types	1.06 (0.03)
Migraine	1.05 (0.02)
Disturbances of sleep/insomnia	0.97 (0.01)
Affective psychosis	1.12 (0.06)
Anxiety disorder/anxiety state	0.99 (0.02)
Depressive disorder/anxiety/depression	1.01 (0.01)
Cough	1.00 (0.01)
Sympt/complt sinus (incl. Pain)	1.08 (0.05)
U.R.I. (head cold)/rhinitis nos	1.00 (0.01)
Sinusitis acute/chronic	1.00 (0.01)
Tonsillitis acute	0.98 (0.03)
Acute laryngitis/tracheitis/croup	1.12 (0.05)
Acute bronchitis/bronchiolitis	1.18 (0.02)
Pneumonia	1.05 (0.03)
Chronic bronchitis/bronchiectasis	0.96 (0.02)
Emphysema/chronic obstructive pulmonary disease	1.01 (0.01)
Asthma	1.04 (0.01)
Hayfever, allergic rhinitis	1.01 (0.01)
Localised redness/erythema/rash of skin	1.11 (0.05)
Herpes zoster	1.03 (0.05)
Dermatophytosis	1.02 (0.01)
Moniliasis/monilia infection/candidiasis	1.05 (0.03)
Other infectious skin dis.nec/erysipelas	0.98 (0.03)
Impetigo	1.04 (0.03)
Seborrhoic dermatitis/other erythematous dermatoses	1.05 (0.02)
Atopic dermatitis/eczema	1.00 (0.02)
Contact dermatitis/skin allergy	0.99 (0.01)
Psoriasis w/wo arthropathy	1.04 (0.03)
Acne	1.05 (0.02)
Urticaria	1.05 (0.03)
Other disease skin/subcutaneous tissue	0.99 (0.04)
Hyperthyroidism/thyrotoxicosis	0.96 (0.04)
Hypothyroidism/myxedema	0.85 (0.02)
Diabetes mellitus	1.00 (0.01)
Gout	0.94 (0.03)
Lipid metabolism disorder	0.96 (0.01)
Cystitis/other urinary infect. Non-venereal	0.98 (0.01)
Family planning/oral contraceptive	1.03 (0.01)
Menstrual pain	0.98 (0.04)

Menstruation excessive	1.03 (0.04)
Menstruation irregular/frequent	1.05 (0.05)
Menopausal sympt./complt.	1.01 (0.02)
Urogenital candidiasis, thrush (proven/unproven)	1.03 (0.02)
Vaginitis/vulvitis, non venereal nos	1.03 (0.04)
Benign prostatic hypertrophy	1.00 (0.04)

Appendix 7.B Total model used to measure the HHI

Fixed effects	**Estimate (st. error)**
DSS use:	
DSS daily users	40.3 (1.17)
DSS non users/havers	41.4 (1.27)
EMR:	
EMR1 (centered)	-0.37 (2.99)
EMR2 (centered)	-2.06 (3.15)
EMR3 (centered)	-3.53 (3.09)
Type of practice:	
Single handed practice (centered)	10.9 (5.14)
Duo practice (centered)	9.23 (4.37)
Group practice (centered)	5.43 (2.90)
Dispensing practice (centered)	2.04 (3.16)
Number of GPs (centered)	2.07 (1.24)
Number of different drugs advised in the DSS (centered)	-3.37 (0.10)

Random effects	**Variance (st. error)**
Practice level:	
DSS daily users	27.5 (9.35)
DSS non users/havers	0.00 (0.00)
GP level:	
DSS daily users	3.48 (3.09)
DSS non users/havers	44.9 (10.9)
Diagnosis level:	
Pain: generalized/unspecified	912.9 (153.8)
Allergy/allergic reaction nos	350.1 (55.69)
No disease	1,005 (206.3)
Iron deficiency anemia	1,500 (232.6)
Pernicious/folate deficiency anemia	1,302 (235.6)
Stomach ache/stomach pain	418.2 (65.43)
Heartburn	399.2 (62.86)
Nausea	979.8 (154.3)
Diarrhea	640 (102.4)
Constipation	233.3 (36.55)
Other presumed infections of digestive system	676.2 (112.6)
Disease of esophagus	804.6 (126)
Disorders of stomach function/gastritis	484.1 (76.6)
Irritable bowel syndrome	473.6 (74.17)

Anal fissure/perianal abscess	603.7 (102.1)
Abnormal sensations of eye	716.1 (114.9)
Infectious conjunctivitis	918.8 (140)
Allergic conjunctivitis	471.7 (73.22)
Otitis externa	470.5 (72.37)
Acute otitis media/myringitis	269.5 (42.23)
Vertiginous syndrome/labyrinthitis/vestibulitis	1125 (179.1)
Angina pectoris	402.9 (64.12)
Heart failure	273.9 (43.42)
Uncomplicated hypertension	398.7 (61.52)
Hypertension with involvement target organs	619.7 (100.6)
Transient cerebral ischemia	1,626 (269)
Hemorrhoids	438.1 (67.8)
Rheumatoid arthritis and allied conditions	543.6 (88.25)
Osteoarthritis of hip	942.2 (161)
Osteoarthritis of knee	315.7 (53.24)
Other osteoarthritis and allied conditions	851.8 (146.7)
Osteoporosis	556.7 (91.72)
Headache	233.8 (36.95)
Tension headache	438.8 (71.34)
Restless legs syndrome	1,141 (194.2)
Vertigo/dizziness	818.8 (128.2)
Parkinsonism/paralysis agitans	1,484 (278.3)
Epilepsy, all types	702.5 (116.1)
Migraine	767.8 (118)
Disturbances of sleep/insomnia	304.8 (47.83)
Affective psychosis	1,131 (233.7)
Anxiety disorder/anxiety state	297.3 (47.08)
Depressive disorder/anxiety/depression	194.7 (30.77)
Cough	621.1 (95.94)
Sympt/complt sinus (incl. Pain)	1465 (282.6)
U.R.I. (head cold)/rhinitis nos	837.8 (128.9)
Sinusitis acute/chronic	99.38 (15.89)
Tonsillitis acute	572.7 (89.45)
Acute laryngitis/tracheitis/croup	1,103 (188.4)
Acute bronchitis/bronchiolitis	1,058 (162.9)
Pneumonia	495.7 (77.23)
Chronic bronchitis/bronchiectasis	527.5 (86.3)
Emphysema/chronic obstructive pulmonary disease	57.73 (9.49)
Asthma	69.96 (11.34)
Hayfever, allergic rhinitis	300 (46.59)
Localised redness/erythema/rash of skin	676.7 (107.1)
Herpes zoster	736.8 (118.6)
Dermatophytosis	197.5 (30.83)
Moniliasis/monilia infection/candidiasis	592.3 (94.62)
Other infectious skin dis.nec/erysipelas	302.6 (48.2)
Impetigo	682.8 (106.5)
Seborrhoic dermatitis/other erythematous dermatoses	488.4 (75.41)

Atopic dermatitis/eczema	188.8 (29.58)
Contact dermatitis/skin allergy	352 (54.49)
Psoriasis w/wo arthropathy	276.2 (43.5)
Acne	180.6 (28.32)
Urticaria	482.5 (74.82)
Other disease skin/subcutaneous tissue	633 (99.3)
Hyperthyroidism/thyrotoxicosis	1,067 (190.4)
Hypothyroidism/myxedema	1,731 (269.9)
Diabetes mellitus	304.3 (47.23)
Gout	384.1 (60.24)
Lipid metabolism disorder	1,051 (161.1)
Cystitis/other urinary infect. Non-venereal	130.7 (20.6)
Family planning/oral contraceptive	354.2 (54.62)
Menstrual pain	669 (111.2)
Menstruation excessive	864.6 (135.5)
Menstruation irregular/frequent	1,708 (272.2)
Menopausal sympt./complt.	243.3 (37.94)
Urogenital candidiasis, thrush (proven/unproven)	193.9 (30.39)
Vaginitis/vulvitis, non venereal nos	659.5 (106.3)
Benign prostatic hypertrophy	1,616 (267.3)

Appendix 7C The multilevel logistic model

$$\gamma_{ijk} \sim \text{Binomial}(n_{ijk}, \pi_{ijk})$$

$$\gamma_{ijk} = \pi_{ijk} + e_{0ijk} x_0^*$$

$$\text{logit}(\pi_{ijk}) = \beta_{0jk} x_{0jk} + \beta_{1jk} x_{1jk} + \beta_2 x_{2k} + \beta_3 x_{3k} + \beta_4 x_{4k} + \beta_5 x_{5k} + \beta_6 x_{6k} + \beta_7 x_{7k} + \beta_8 x_{8k} + \beta_9 x_{9k} + \beta_{10} x_{10ijk} + \ldots\ldots + \beta_{90} x_{90ijk} + e_{91i} x_{91ijk} + \ldots\ldots + e_{172i} x_{172ijk}$$

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$

$$\beta_{1jk} = \beta_1 + v_{1k} + u_{1jk}$$

$$\begin{bmatrix} v_{0k} \\ v_{1k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma^2_{v0} & \\ 0 & \sigma^2_{v1} \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \\ u_{1jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{uo} & \\ 0 & \sigma^2_{u1} \end{bmatrix}$$

$$\begin{bmatrix} e_{91ijk} \\ \ldots\ldots \\ e_{172ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e91} & & \\ 0 & \ldots\ldots\ldots\ldots & \\ 0 & \ldots\ldots\ldots\ldots & \sigma^2_{e172} \end{bmatrix}$$

γ_{ijk} = dependent variable (conform DSS or not, dichotomous)
i = ' patient populations' (an identifier for each diagnosis per physician)
j = physicians
k = practices

β_{0jk} = mean propotion conform DSS, per physician (daily users)
β_{1jk} = mean proportion conform DSS, per physician (non users/ havers)
x_2 = EMR1, per practice (centred)
x_3 = EMR2, per practice (centred)
x_4 = EMR3, per practice (centred)
x_5 = Duo practice, per practice (centred)
x_6 = Group practice, per practice (centred)
x_7 = Health centre, per practice (centred)
x_8 = Dispensing practice, per practice (centred)
x_9 = number of GPs, per practice (centred)
x_{10}.....x_{90} = diagnosis (centred)
x_{91}.....x_{172} = diagnosis

v = level - three variance (practice level)
u = level - two variance (physician level)
e = measurement error per diagnosis

Appendix 7D The multilevel model

$$\gamma_{ijk} \sim N(XB, \Omega)$$

$$\gamma_{jk} = \beta_{0jk}x_{0jk} + \beta_{1jk}x_{1jk} + \beta_2 x_{2k} + \beta_3 x_{3k} + \beta_4 x_{4k} + \beta_5 x_{5k} + \beta_6 x_{6k} + \beta_7 x_{7k} + \beta_8 x_{8k} + \beta_9 x_{9k} + e_{10i}x_{10ijk} + \ldots\ldots + e_{91i}x_{91ijk}$$

$$\beta 0jk = \beta 0 + v_{0k} + u_{0jk}$$

$$\beta 1jk = \beta 1 + v_{1k} + u_{1jk}$$

$$\begin{bmatrix} v_{0k} \\ v_{1k} \end{bmatrix} \sim N(0, \Omega v) : \Omega v = \begin{bmatrix} \sigma^2_{v0} & \\ 0 & \sigma^2_{v1} \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \\ u_{1jk} \end{bmatrix} \sim N(0, \Omega u) : \Omega u = \begin{bmatrix} \sigma^2_{uo} & \\ 0 & \sigma^2_{u1} \end{bmatrix}$$

$$\begin{bmatrix} e_{10ijk} \\ \ldots\ldots \\ e_{91ijk} \end{bmatrix} \sim N(0, \Omega e) : \Omega e = \begin{bmatrix} \sigma^2_{e10} & & \\ 0 & \ldots\ldots\ldots\ldots & \\ 0 & \ldots\ldots\ldots\ldots & \sigma^2_{e91} \end{bmatrix}$$

γijk = dependent variable (HHI, per diagnosis per physician)
i = 'patient populations' (an identifier for each diagnosis per physician)
j = physicians
k = practices

$\beta 0jk$ = mean HHI and variance parameters, per physician (daily users)
$\beta 1jk$ = mean HHI and variance parameters, per physician (non users/ havers)
x_2 = EMR1, per practice (centred)
x_3 = EMR2, per practice (centred)
x_4 = EMR3, per practice (centred)
x_5 = Duo practice, per practice (centred)
x_6 = Group practice, per practice (centred)
x_7 = Health centre, per practice (centred)
x_8 = Dispensing practice, per practice (centred)
x_9 = number of GPs, per practice (centred)
x_{10}.....x_{91} = diagnosis

v = level - three variance (practice level)
u = level - two variance (physician level)
e = measurement error per diagnosis

8

Summary of conclusions and discussion

In this chapter a summary of the study will be given. The main findings will be described and discussed. This chapter ends with a proposal for future research into variations in medical practice.

8.1 Introduction

In this book several hypotheses on the causes of medical practice variation were empirically tested. Following Freidson (1975) and Westert and Groenewegen (1999) an approach was applied that emphasized social conditions that, by providing opportunities and constraints, influence physicians' medical behavior. This approach predicts variation between physicians working in different work environments and similarities between physicians sharing the same work environment, such as a practice or a hospital. In this book it was confirmed that circumstances are especially important in explaining medical practice variation.

Three mechanisms that generate variation between and homogeneity within groups were tested: selection, gradual adaptation towards group norms and rapid adaptation to circumstances. The results pointed towards the importance of similarities based on shared circumstances. It also showed that the most common explanation of medical practice variations based on individual preferences was unsatisfactory. The implication is that medical practice variations are not merely individual differences in preferred practice styles, but are patterned by social processes in partnerships and local circumstances. Knowing that, we moved on with testing institutional mechanisms and their effect on variation.

The relationship between the institutional mechanisms and circumstances is that institutional mechanisms, as used in this book, have an effect on the circumstances that influence variation. Scott (2001) distinguishes three mechanisms of how institutions work: the regulative, normative, and the cultural-cognitive mechanism. These three mechanisms were hypothesized to influence variation. Testing the institutional mechanisms was important for understanding medical practice variation. Besides an explanation, the institutional mechanisms provide instruments that can be used in influencing variation. We showed that institutional mechanisms are effective in influencing physicians' behavior, and therefore could be used in limiting variation. However, we found no empirical evidence proving that institutional mechanisms reduce variation in this study. This book contributed to understanding medical practice variation and the mechanisms underlying it without a normative opinion on the existence of variation. With this book we made progress in explaining medical practice variations and in the methodology that can be used in analyzing medical practice variation.

8.2 Research questions and main hypotheses

In the first chapter 2 different types of questions were formulated; a descriptive question addressing the extent of variation, and an explanatory question addressing explanations for medical practice variation. The second question was the most important research question in this study.

1 *'To what extent does medical practice vary between (groups of) physicians?'*

2 *'How can variations in medical practice between (groups of) physicians be explained, after taking medically relevant factors into account?'*

In chapter 2 the hypothesis was tested that there is a clustering of variation within general practices in the Netherlands when medical behavior or the consequences of medical behavior are visible. Visibility was supposed to be an important pre-condition for similarities to arise. Chapter 3 looked at what the mechanisms are that cause general practitioners within practices in the Netherlands to resemble each other. Three mechanisms were specified; selection, mutual adaptation towards group norms and rapid adaptation to circumstances. In chapter 4 a test was applied to check whether circumstances instead of preferences explain variation. The hypothesis that physicians working in two different hospitals act differently in those hospitals was tested with data from the USA. Chapter 5 examined the behavior of physicians under restrictive (regulative) circumstances, namely managed care activities by insurers in the USA. In chapter 6 the behavior of physicians under restrictive (normative) circumstances, in this case professional guidelines for general practitioners in the Netherlands, was examined. In chapter 7 the influence of cultural-cognitive processes on variation was examined for general practitioners in the Netherlands. The hypothesis that physicians using decision aids show less variation when prescribing was tested.

8.3 Data and methods

In order to allow for the hypotheses to be empirically tested, data sets were constructed using secondary data that gave an opportunity to test the hypotheses. They were used from two national data collections held in the Netherlands amongst general practitioners: the First and Second Dutch

National Survey of General Practice (DNSGP1 and DNSGP2). These data were collected in 1987 and 2001 respectively. Furthermore, data were obtained from the New York Statewide Planning and Research Cooperative System (SPARCS). The 1999, 2000 and 2001 SPARCS-data were used. The analyses in this book were performed using multilevel analyses (table 8.1).

Table 8.1 Data used for testing hypotheses

Chapter	Hypothesis cluster	Data used
2	Circumstances	Data were used from the Second Dutch National Survey of General Practice (DNSGP2), collected in 2001. A total of 195 GPs in 104 practices serving 390,000 patients participated in this survey. Single-handed practices were excluded from the analyses. Depending on the analyses between 29 and 36 practices, 56 and 68 GPs, 88553 and 107739 contacts were included.
3	Circumstances	Data were used from the First Dutch National Survey of General Practice, collected in 1987. A total of 161 Dutch GPs in 103 practices, who served 335,000 patients participated in this survey. Single-handed practices were not included in the analysis. In all, 96 GPs in 42 practices were included in the analyses.
4 and 5	Circumstances. The regulative mechanism	Data were obtained for the years 1999, 2000 and 2001 from the New York Statewide Planning and Research Cooperative System (SPARCS). We used 1999, 2000 and 2001 SPARCS-data to study seven DRGs: two medical, one surgical and four obstetrical. More than 1000 physicians and 150 hospitals were included per DRG.
6	The normative mechanism	Data were used from the First and Second Dutch National Survey of General Practice (DNSGP1 and DNSGP2), collected in 1987 and 2001 respectively. In 1987 there were no guidelines, while in 2001 75 guidelines were available (www.nhg.nl). A total of 160 GPs in 102 practices from DNSGP1 and 108 GPs in 69 practices from DNSGP2 were included in the analyses.
7	The cultural-cognitive mechanism	Data were used from the Second Dutch National Survey of General Practice (DNSGP2), collected in 2001. A total of 104 GP practices participated in this survey. Data were collected on contacts, patients, GPs and practices, using electronic medical records, and questionnaires. A total of 82 diagnoses, 749,811 cases, 133 physicians, and 85 practices were included in the analyses.

8.4 Main findings

The importance of circumstances

In chapter 2 we studied whether the visibility of clinical behavior and the use of shared resources influence variation in medical practice between GPs working in different practices. The idea was that variation in medical practice between physicians is related to differences in incentives and circumstances between their work environments, providing opportunities and constraints on behavior. An important aspect of working in the same environment, the same practice, is that GPs use shared resources, like personnel and equipment. The work environment serves as a social, economical and technological system in which decisions take place and this system may cause different physicians to make similar medical decisions. The visibility of behavior is an important pre-condition for similarities to arise based on group norms. When behavior, or the outcomes of behavior, is, or are, visible and have consequences for others, norms may develop. Physicians run the risk of being criticized by other physicians working in the same partnership, because they see each others' behavior. This risk can be minimized by showing similar behavior to those colleagues. For testing the hypothesis several clinical activities, such as prescription, referral, diagnostics, treatment, and giving advice, were studied. Data were used from the Second Dutch National Survey of General Practice. In table 8.2 the hypotheses, how they were tested and the test results are summarized.

We found that clustering of variation within practices depends on the clinical activity studied. GPs differed more from colleagues working in other practices than from their colleagues working in the same practice in relation to activities such as treatment and diagnostics performed in the GPs' practice. Both are rather well visible to colleagues in the same practice. It was the other way round for prescribing, referral to primary care, referral to secondary care, diagnostics performed in an external laboratory, and for giving advice. For these activities there was less variation between practices than within them, GPs working in the same practice did not show more similarities than GPs working in different practices. These clinical activities are less visible to colleagues and so do not make use of shared resources. This confirmed our hypothesis that there is less variation within practices when shared resources are used and when behavior is visible to colleagues, was confirmed.

These findings gave us some ideas on whether medical practice variation is within, or between practices. The findings indicated that partnerships

indeed develop into a normative community for those behaviors that are visible to physicians working in the same partnership. A next step was to search more specifically for the causes of patterns of medical practice variations.

In chapter 3 we elaborated upon the question of why physicians sharing a work environment show similarities. We achieved this by testing the hypothesis that physicians in partnerships are more similar in professional attitudes and behavior to each other than to randomly chosen physicians.

Three different mechanisms were specified that might explain patterns of medical practice variation: the selection of new partners might be directed towards similarities in as much as like seeks like, gradual adaptation to each other within a partnership might result from processes of peer approval, and, finally, the circumstances that are shared by partners might lead to similar behavior. These are three general processes that are supposed to generate homogeneity within groups or variation between groups. Data were used from the First Dutch National Survey of General Practice. A variety of attitudes and behavior were examined. Table 8.2 summarizes the hypotheses, how they were tested and the test results themselves.

It was found that Dutch GPs working in the same partnership showed more resemblance in attitudes and behavior than GPs not working in the same partnership. Most indications pointed towards circumstances, and to a lesser extent towards adaptation, as an explanation of similarities within partnerships. This gave a first clue to an explanation of similarities based on circumstances. No evidence was found that selection based on similar age and gender forms an explanation for similarities between GPs sharing a work environment. This finding does not necessarily imply that selection does not play a role in explaining similarities in behavior within partnerships but at least selection does not seem to be made on the basis of equal age or gender. The implication of these findings is that medical practice variations are not merely individual differences in preferred practice style, but patterned by social processes in partnerships and local circumstances.

In chapter 4 we tested whether circumstances instead of preferences explain variation. The hypothesis that physicians working in two different hospitals act differently in those hospitals was tested. This would imply that variation within hospitals was small compared with variation between hospitals for physicians treating similar patients. A second implication would be that

physicians working in more than one hospital conform to the usual practice of each hospital (Westert, 1992). This implication cannot be deduced from a preference-centered approach, as the preferences of an individual would not change when working in another hospital.

Table 8.2 The importance of circumstances: description of hypothesis testing

Hypothesis description	Chapter	Data source	Method of testing	Test result
Most variation occurs in the least visible situation	2	DNSGP2	Variation for different medical decisions, more or less visible to colleagues was compared	Confirmed
Physicians in partnerships are more similar in professional attitudes and behavior towards each other than to randomly chosen physicians	3	DNSGP1	Similarities in attitudes and behavior between physicians working in the same partnership were compared to similarities between physicians working in different partnerships	Confirmed
Variation between physicians working in the same hospital is less than variation between hospitals	4	SPARCS	Variation between and within hospitals was compared	Confirmed
Physicians working in different hospitals in which the average lengths of stay differ, will choose lengths of stay similar to the average in the hospital in which the patient is treated	4	SPARCS	Length of stay chosen by physicians working in two hospitals was compared to the average length of stay chosen in those hospitals	Confirmed
The higher the proportion of physicians working in only one hospital, the less variation in the length of stay there will be between physicians in that hospital	4	SPARCS	The potential impact of lengths of stay generated by physicians working in only one hospital on total practice variation within a hospital was measured using a covariance estimated with multi-level analysis, measuring that part of the variation that can be attributed to the proportion of physicians practicing in only one hospital.	Not confirmed

To test the hypotheses secondary data were used, originating from the New York Statewide Planning and Research Co-operative System (SPARCS). The study examined physicians practicing in one hospital, and physicians practicing in more than one hospital in order to determine whether average lengths of stay differs according to the hospital of practice. In table 8.2 the hypotheses, how they were tested and the test results, are summarized.

The hypothesis that variation within hospitals was small compared to variation between hospitals was confirmed. This result, however, is still compatible with both the approach based on preferences for a certain practice style and the approach based on work circumstances. We therefore also analyzed the decisions of the same physicians in different hospitals. Preferences for a certain practice style are supposed to be relatively stable within the same person, but circumstances may clearly differ. Several different analyses confirmed that physicians working in two hospitals with different average lengths of stay have a length of stay similar to the usual practice in the hospital where the procedure was performed.

This study demonstrated that organizational units, such as hospitals, are important in studying medical practice variations. It was confirmed that circumstances are important in understanding medical practice variation.

Institutional mechanisms

Scott (2001) distinguishes three mechanisms of how institutions work: the regulative, the normative, and the cultural-cognitive mechanism. These three mechanisms might influence variation. We tested all three institutional mechanisms, by choosing an example of each, in three separate studies. First an example of the regulative mechanism was described, then an example of the normative mechanism was given and finally an example of the cultural-cognitive mechanism was provided.

The managed care system in the USA is an example of the regulative mechanism; insurers set rules on the treatment that can be given to patients. The normative mechanism refers to peer expectations, and matches professional control. An example is guidelines developed by the profession itself. These are guidelines with a normative character. The cultural-cognitive mechanism relates to a common framework of meaning.

In chapter 5 we examined the influence of managed care on the decisions physicians take and on variation in medical practice. The managed care system in the USA is an example of a restrictive circumstance in which

regulative mechanisms are used to influence medical behavior. The Health Maintenance Organization (HMO) is the best known type of managed care in the USA (Bachman and Freeborn, 1999).

We focused on one important set of constraints, namely those set by the insurer of the patients. Constraints on those patients insured through the HMO will be far more restrictive than for those traditionally insured, implying that variation between physicians treating patients with an HMO insurance will be less than the variation between physicians treating other patients. Physicians treating patients with HMO insurance face similar constraints and incentives, leading to similar decisions regarding length of stay.

On the basis of the differences in constraints between managed and non-managed care plans, we hypothesized that the length of stay would be shorter for managed care patients, and that there would be less variation in the length of stay for managed care patients. We studied the lengths of stay for comparable patients who are insured under managed or non managed care plans. The 1999, 2000 and 2001, data from hospitals in New York State were used and analyzed with multilevel analysis. In table 8.3 the hypotheses, how they were tested and what the test results are summarized.

The results showed no difference in the length of stay between managed and non-managed care patients. Furthermore, it appeared that there was less variation in the length of stay for managed care patients. Contrary to our expectations, this difference was not primarily found at the level of the physicians, nor was it found at the level of the hospitals.

We found a difference in the effect of the proportion of managed care patients on the physicians and on the hospitals. Where physicians are concerned, the length of stay is shorter when this proportion is higher, but this effect was not found in hospitals. This means that the physicians' choice about the length of stay is influenced by the insurer, while the hospital does not change its policy. The proportion of managed care patients did not have an unequivocal effect on the variation in both physicians and hospitals, indicating that there was no insurer effect. The option of treating patients in another hospital did not influence variation in the length of stay for managed care patients, which would seem to be an effect of the insurer in combination with less variation when the number of insurers was lower.

We did not find evidence that it is managed care that has an effect on the length of stay and thus on the costs related to inpatient days. There seems to be something else that is causing the short length of hospital stay in the United States, independent of the patient's insurance. What we found is that it is not restrictions imposed by the insurer that result in patterns of variation, since there are hardly any differences in the length of stay of managed care and non-managed care patients. It might be that hospitals respond to the way they are paid; payment per DRG means that it is always (cost) efficient to keep the length of stay short. Or it may be that the knowledge that managed care is on the increase that is causing hospitals to react in advance by developing strategies to make sure that they will have (managed care) patients in the future (Zhang et al., 1999). Actually, it seems that hospitals are managing care more than the insurer, most probably because of financial incentives.

Table 8.3 The regulative mechanism: description of hypothesis testing

Hypothesis description	Chapter	Method of testing	Test result
Length of stay will be shorter for managed care patients	5	Mean length of stay for the managed care and the non-managed care group were compared	Not confirmed
There will be less variation in length of stay for managed care patients	5	Variation in length of stay for the managed care and the non-managed care group were compared	Confirmed
The influence of managed care, i.e. less variation in length of stay, will primarily be found at physician level	5	Variation for the managed care group and the non-managed care group at physician level was compared to the variation for both groups at hospital level	Not confirmed
The higher the proportion of managed care patients the physician has, the shorter the length of stay (a) and the less the variation in length of stay at physician level (b)	5	The regression coefficient for the proportion of managed care patients per physicians was examined as well as the covariance between this proportion and the variation in length of stay; both are expected to be negative	Confirmed (a)/ not confirmed (b)
The higher the proportion of managed care patients the hospital has, the shorter the length of stay(a) and the less the variation at hospital level (b)	5	The regression coefficient for the proportion of managed care patients per hospital was examined as well as the covariance between this proportion and the variation in length of stay; both were expected to be negative	Not confirmed (a)/not confirmed (b)

- table 8.3 continues -

- table 8.3 continued -

Hypothesis description	Chapter	Method of testing	Test result
Physicians who deal with fewer insurers will have less variation in length of stay for managed care patients	5	The covariance between the number of insurers per physician and the variation in length of stay was examined and expected to be positive	Confirmed
Hospitals dealing with fewer insurers will experience less variation in length of stay for managed care patients	5	The covariance between the number of insurers per hospital and the variation in length of stay was examined and expected to be positive	Confirmed
Physicians practicing in fewer different hospitals will have less variation in length of stay for managed care patients	5	The covariance between the number of hospitals per physician and the variation in length of stay was examined and expected to be positive	Not confirmed
The easier it is to standardize treatment for a specific DRG, the less variation in length of stay there will be for patients under managed care	5	Variation for all DRGs was compared, most variation was expected in medical DRGs and DRGs with complications	Confirmed

In chapter 6 the behavior of physicians under restrictive normative circumstances, in this case the professional guidelines for general practitioners in the Netherlands, was examined. The variation was expected to decrease when guidelines are followed. Following guidelines means belonging to a group and it gives people status or protection when behavior is to be explained. By following guidelines, it is easier to explain what is done, and to achieve acceptance. We discussed two important mechanisms influencing the reduction of variation by the introduction of guidelines: the uncertainty of medical behavior and the integration in the profession. With these mechanisms differences between physicians working in single-handed practices and physicians working in group practices can be expected.

We tested whether more uniformity in prescribing, in other words less variation in drugs prescribed, was created after the introduction of guidelines, irrespective of whether they are followed or not. In order to do so, we analyzed data from two large studies (DNSGP1 and DNSGP2), one held before the introduction of guidelines developed by the Dutch College of General Practitioners (NHG) and the other held after the introduction of those guidelines. For both studies two groups of diagnoses were created and

compared, one for which guidelines were (going to be) introduced, the guideline group, and one for which there were no guidelines, the reference group. In table 8.4 the hypotheses, how they were tested and what the test results were are summarized.

Table 8.4 The normative mechanism: description of hypothesis testing

Hypothesis description	Chapter	Method of testing	Test result
Variation in prescription has decreased with the introduction of guidelines	6	Differences in a measure of concentration between two groups of diagnoses (a guideline and a reference group) were compared before and after the introduction of guidelines	Not confirmed
Variation in prescription has decreased more for GPs working in single-handed practices than for GPs working in group practices	6	Differences in a measure of concentration between GPs in single-handed and group practices for two groups of diagnoses (a guideline and reference group) were examined separately	Confirmed

We found that although more different drugs were prescribed during DNSGP2 for both the reference and the guideline group, the increase in the number of different drugs prescribed was higher for the reference group. The overall increase in the number of different drugs prescribed can be explained by the fact that there was a far greater variety of drugs available during DNSGP2 than during DNSGP1. That the increase was lower for the guideline group might indicate that the guidelines temper the effect on the availability of more drugs.

We found no significant difference between the reference group and the guideline group for GPs in group practices between DNSGP1 and DNSGP2. This suggests that guidelines did not have an effect on the variation for GPs in group practices. However, in line with our hypothesis, we found that the guidelines primarily had an effect on the variation in single-handed practices; variation increased less for the guideline group.

From this study we concluded that the introduction of guidelines, although it probably tempered the increase in variation, did not reduce variation. The introduction of guidelines alone is not enough to change behavior and reduce variation. A step further is to intervene in physicians' daily routines, instead of using rules and regulations. This can be done by using

computerized decisions aids, an example of the cultural-cognitive mechanism.

In chapter 7 the influence of cultural-cognitive processes on variation was examined. The cultural-cognitive mechanism is supposed to have an effect on variation as physicians using a similar framework to make decisions are likely to reach the same decision (with the pre-condition that not much room is left for their own interpretation). We studied a DSS which is used by GPs in the Netherlands as a tool to give advice on prescription when the, ICPC-coded, diagnosis is given. The DSS proposes a prescription, given the diagnosis of the patient, taking into account age, sex and co-morbidity.

The use of DSSs is hypothesized to decrease the variation between physicians, because physicians make use of the same cognitive framework. DSSs intervene in physicians' daily routine. They are used to facilitate for instance the use of those guidelines. Data were used from the Second Dutch National Survey of General Practice (DNSGP2), collected in 2001. In table 8.5 the hypotheses, how they were tested and what the test results were are summarized.

Table 8.5 The cultural-cognitive mechanism: description of hypothesis testing

Hypothesis description	Chapter	Method of testing	Test result
Physicians using the DSS, prescribe according to the guidelines that are incorporated in the DSS	7	Prescriptions by physicians using the DSS and those not using the DSS are compared to prescription suggested in the guidelines	Confirmed
Physicians using decision aids show less variation in prescription	7	A measure of concentration for drug prescription was compared between physicians using the DSS and those not using the DSS	Not confirmed

This study demonstrated that physicians using a DSS are prescribing conform to the advice given in the DSS more than physicians not using a DSS. This, however, did not mean that variation was lower; variation was the same for physicians using and for physicians not using a DSS. This is probably related to the fact that the DSS we studied advises several different drugs or recommends a step by step treatment starting with one type of drug and changing that type of drug later on when necessary. The implications of the study are that DSSs can be used to implement guidelines, but that it should not be expected that variation is limited. Variation is

probably only limited when DSSs do not give an opportunity for variation. When they give for instance only one advice on which medication to prescribe.

8.5 Scientific implications of the results

The scientific implications are in the theoretical model for explaining medical practice variations. We mentioned the appropriate level of analysis as one of the questions arising in the debate in which explanations for medical practice variation are sought. We argued that the occurrence of non-random variation points towards explanatory theories that go beyond the individual. On the face of it individual decisions, appear not to be as individual as they seem. People in general do not take decisions in a vacuum; they are influenced by the circumstances under which they take decisions. Circumstances provide for opportunities and constraints. In line with Coleman (1990) and Freidson (1975), Westert (1992) argued that the medical behavior of individual physicians is influenced by restrictions in the direct work environment, thus indicating that the appropriate level of analysis is the work environment. Still, research in the area of medical practice variation is most often regionally oriented. We extended the viewpoint of Westert (1992) by adding institutional mechanisms to the model. These institutional mechanisms in relation to medical practice variations had never been studied systematically before.

With our theoretical model we predicted that:
- *variation between physicians sharing a work environment is less than variation between physicians working in different work environments, because of selection, gradual adaptation and circumstances;*
- *regulative, normative, and cognitive mechanisms decrease variation in medical practice within the units to which they apply.*

The research results confirmed that circumstances are important in explaining medical practice variations. It indicated that partnerships form a normative community for those behaviors that are visible to physicians working in the same partnership. Implications were found that physicians indeed do not take their decisions autonomously. It showed that the level at which variations should be examined is the level of the environment in which physicians practice. It was found that within the same work environment physicians show similarities in behavior.
Our expectations concerning the influence of institutional mechanisms on medical practice variation were not empirically confirmed. Our results show

that these mechanisms influence medical behavior, but they do not necessarily decrease variation in medical practice. The results, however, suggest that institutional mechanisms can be used as instruments to influence medical practice variation, because behavior can be influenced. Whether variation is reduced or not probably depends on the room for variation given within the rules and regulations.

We extended an existing theoretical model, the model of local standards developed by Westert (1992), and derived new hypotheses. These hypotheses were tested using international data from both hospitals and general practice. Multilevel analyses were used to analyze the data. We showed that the theoretical model holds for different countries and is not only applicable to hospital care, but applies to general practice as well.

8.6 Methodological approach

This book is a combination of six separate studies. Each of these studies has its own methodological approach. What is common in all six studies is that secondary data are analyzed using multilevel analysis. We will start with discussing the general methodological approach, followed by a description of the limitations and strengths of the study.

General methodological approach

For testing the hypotheses secondary data from three different databases were used. The advantage of using secondary data is that it saves time; no data have to be gathered. The disadvantage is that data were not gathered for testing our specific hypotheses. Data were used from two national surveys held amongst general practitioners in the Netherlands, and from hospitals in the United States. These different datasets were used because they gave the best opportunity to test the hypotheses. For example, to test the hypothesis that physicians working in different places act differently in those places, we needed data on physicians working in more than one place. In the Netherlands this situation seldom occurs, while in the United States it is quite common for physicians to work in more than one hospital.

Because we were interested in universal mechanisms, and we did not perform a purely descriptive study, it is of less importance whether or not data are used from the same dataset, gathered at the same time and place. The mechanisms will be the same for different datasets, gathered at different times and places. Variation occurs throughout the world, and the pattern of

variation is found to be similar for common procedures among fee-for-service hospital markets even though obvious differences exist in the supply of surgeons, the organization and financing of services, and in the cultural and demographic characteristics of the people living within the hospital's market area (Wennberg, 2004).

An important analogy between the datasets that influenced the analysis is that the data are about physicians working in a certain environment. As social circumstances cause similarities between physicians working in the same environment, their clinical activities concerning their patients are not independent and this should be taken into account in the analyses. The data are hierarchically structured. Therefore, we used multilevel models to analyze the data (Snijders and Bosker, 1999; Hox, 1995).

Limitations of the methodology

Although we searched for methods that best applied to the hypotheses that needed to be tested, the methodology has several limitations. First of all we address the possibility of composition effects. We were interested in variation between physicians and partnerships or hospitals. The possibility of composition effects should be excluded as much as possible, as the purpose of the book was not to describe the differences between physicians, partnerships or hospitals based on patient selection. We excluded this possibility by adjusting for case mix, but there is always a possibility that composition effects influenced the results. In the study on hospital length of stay (chapter 4), we were not able to distinguish between influences exerted by colleagues and by other circumstances. This could have been done if the specific characteristics of the hospitals had been taken into account. Those characteristics were unavailable. Patient selection could also be an explanation of variation between hospitals. Physicians working in more than one hospital may apply some criterion that is possibly related to the length of stay, when choosing the hospital to which they admit patients. This, however, would reinforce the theory of the influence of the hospital on the length of stay.

In other studies (for example in chapter 2) we did not take into account the specific diagnoses. This was not possible, because there were too many diagnoses, and a model could not be fitted. Differences in diagnoses within a GP's practice are an important source of variation in clinical activities. This, however, would be reflected in all clinical activities. If differences in diagnoses were an explanation for the remaining variation, all clinical activities should show patterns of variation by practice (if diagnoses differ

between practices) or by GP (if diagnoses differ between GPs). That is not what we found and therefore we conclude that differences in diagnoses were not of main importance in explaining variation between practices.

Strengths of the methodology

An important advantage in performing the study as was done in this book is that it offered the opportunity to do a lot of research, to test all kinds of hypotheses, given the fact that no data needed to be collected. Data were used from different situations, instead of having to wait for changes to occur.

Multilevel analyses were used to analyze the data. Using this statistical technique solves the problem of the loss of information through aggregation, as well as the problem of misestimated precision in estimating coefficients when higher level characteristics are assigned to lower level units. With multilevel analysis, total variation in clinical activity is separated into several parts. For instance: a part due to differences between patients, a part due to differences between physicians, and a part due to differences between practices or hospitals. This is the basic methodology used in each of the studies. In all studies we were primarily interested in variation and how variation was divided between different levels.

The methodology used to test the hypotheses was state of the art at some points and challenging at other points. Statistical techniques are constantly under development and advanced techniques provide sophisticated analyses for testing hypotheses. However, we ran into the limits of what, to our knowledge, was possible with present techniques. This required solutions that approached the way one would ideally analyze the data. It can be expected that in the future it will be possible to test the hypotheses with even more sophisticated analyses. Although we would certainly apply new techniques when available, we do think that the methods used were the best we could have thought of, and however not perfect, the results contribute to our understanding of medical practice variation.

8.7 Implications of the study results for health policy

The purpose of this book was to identify the mechanisms underlying medical practice variation. The results of the study have implications for health policy. In the introduction we described the importance of understanding medical practice variations from different perspectives; a

health outcomes perspective, a consumer or patient perspective, a health profession perspective, and a regulator perspective. An important question, from all these perspectives, in the context of policy implications of this study is: (how) can physicians' behavior be influenced? Actually, for health policy implications two questions are important. Firstly, do the results explain why existing health policy aimed at influencing physicians' behavior is (not) effective? Secondly, are there consequences of the results that can be valued positively or negatively and should policy be aimed at stimulating or changing these consequences? This relates to the question of whether medical practice variation is good or bad. We will pay attention to this issue in section 8.8.

Results from the first three studies (chapters 2, 3 and 4), showed that physicians adjust to colleagues or circumstances in the practice or hospital where they perform procedures. This causes uniformity in medical behavior. In addition, we found that if behavior is visible or uses shared resources, then physicians are mutually dependent and behavior becomes more uniform. From the last three studies (chapters 5, 6 and 7) we learned that regulatory, normative and cognitive-cultural pressures influence behavior but do not reduce medical practice variation. These results have implications for health policy.

For instance the first result, the importance of circumstances, has implications for policy that aims at getting physicians to work according to evidence based standards. Not all physicians work according to them. Thus, working according to these standards, involves changes in medical behavior. The policy implication is that activities that introduce standards should be directed at physicians in the relevant units, instead of individual physicians. For instance, individual courses will be less effective than local consensus meetings, for example within practices.

The results relating to institutional mechanisms and the visibility of behavior have implications for current policies introducing pay-for-performance and performance indicators. In general, pay-for-performance connects quality measures to financial incentives. Performance indicators are primarily meant to monitor quality. We assume that both pay-for-performance and performance indicators increase visibility and mutual dependencies. They are part of a change from professional control towards managerial control. *"Society expects the health care sector to be more open, more accountable, at least when it comes to using the resources they claim from the public sector. Each institution is expected to defend its budgetary needs, year-by-year, and*

to justify the growth in needed resources, by linking the expected expenses to the expected activities." (Van Kemenade, 2007). Third parties, such as insurance companies, monitor physicians' behavior and try to limit the use of resources. The implication of our study results for policies introducing pay-for-performance and performance indicators is that it will result in more uniformity within local units (practices). Over time, when all practices are governed by the same performance indicators, these policies will result in more uniformity between practices. Another assumption is that introducing pay-for-performance increases mutual dependencies more than performance indicators in general, because of the financial incentives. The implication is that introducing pay-for-performance increases homogeneity more than performance indicators without direct financial consequences.

The result that the institutional mechanisms influence behavior but do not reduce medical practice variation has implications for long-term changes in health care that transform our thinking about health care from a professional service into a product and changes patients into consumers of health care products. We assume that cognitive-cultural pressure leads to more uniformity only if they restrict the number of alternatives for physicians effectively. Furthermore, we assume that health care products that can easily be standardized will be provided in separate units or clinics, because high volume production will be financially attractive. The implication of the study result concerning this policy context is that there will be more homogeneity for treatments that can be standardized in high volume units (clinics).

The same result has implications in the context of patients behaving more as consumers and having increasing influence on what physicians do. Patients' choice becomes more important. Increasing attention for giving patients a choice is an instrument that matches market control (Saltman and Figueras, 1998). An assumption is that normative pressure from colleagues is increasingly counteracted by patient preferences. The relative importance of colleagues' opinions is shifting as patients become more important in decision-making. Patients differ in the amount of pressure they are able to exert on physicians, for example their ability to take part in discussions over treatment. Some groups like those speaking a minority language or elderly people, are less able to exert pressure on physicians. The implication is that in the case of normative pressure by colleagues, for example guidelines, physicians adhere more closely according to those guidelines when treating patients that are not able to discuss their treatment, for instance minority groups, than for other patients who might ask the physician for treatment

that is different from the treatment recommended in the guidelines. More homogeneity within the first group can be expected due to normative pressure from colleagues than with other patients.

We now gave some examples of the policy implications of the study results and showed that our model can be used to predict changes in medical practice variation when new developments take place. However, we did not empirically test these predictions. In this light, it is interesting to monitor trends in medical practice variation and test hypotheses derived from our theoretical model. These trends, as well as other issues, could be subject of future research.

8.8 Proposal for future research

Background

Research on variation in medical practice can be traced back to the 1930s. However, it was not until the 1970s, with a publication by Wennberg and Gittelsohn (1973) in Science, that these variations received more attention in the USA. In the 1980s interest in medical practice variation was raised in Europe. Medical practice variations involving all kinds of medical treatment have been well documented ever since. It was found that these variations could not be explained by differences in patient populations. Although much progress was made describing medical practice variation, less progress was made regarding the explanation for medical practice variations. It was generally assumed that physician behavior is important in explaining medical practice variations. Still, the theory for the relation between physicians' behavior and medical practice variation was unsatisfactory. Explanations were mostly based on differences in preferences between physicians, thus ignoring the influence of circumstances.

Westert and Groenewegen (1999) proposed an additional approach to solve the problem that sudden changes in behavior could not be explained and that hypotheses derived from the preference-centered approach were difficult to test. In their constraint-centered approach, they emphasized (social) conditions or local circumstances that influence the behavior of physicians by providing opportunities and constraints. The advantage of this approach is that there is an explicit relationship between the work-related circumstances, physicians' medical behavior and medical practice variation. Using this approach Westert (1992), developed a theoretical model, a model of local standards, for explaining geographical differences in

the hospital length of stay for surgical operations in the Netherlands. The next question was whether this model also applied to other medical decisions, and whether the model should be elaborated. Both were done this book. The model was tested for general practice in the Netherlands and for hospitals in the USA and extended with institutional mechanisms. It proved to be valuable in explaining medical practice variations. However, there are still uncultivated areas. For outlining future research, three lines of reasoning can be followed. The first is to examine implications from the hypotheses and/or study results that are not studied yet. This would include research concerning time trends and the interaction between institutional mechanisms. The second is to examine hypotheses that are not confirmed in this study and to give explanations for them. This line of reasoning would lead to further testing the hypotheses or to extending the theoretical model. The third line of reasoning is of a different nature and concerns the policy implications of medical practice research, instead of concerning the theory that explains medical practice variation. It concerns the relationship between medical practice variations and the quality of care. We will now give a general outline of the proposed studies.

Time trends in medical practice variation

What we described in the section on health policy implications (section 8.7 in this book) hints at changes in medical practice variation over time. All kinds of developments in society and the health care system influence medical practice variation. Furthermore, knowledge on medical practice increases and it might be that there is (evidence-based) consensus on the most appropriate treatment for more diagnoses now than there was in the past. This would decrease professional uncertainty and therefore limit the options for individual beliefs. Subsequently, variation for these diagnoses decreases. Knowledge on trends and the causes of changes in variation adds to understanding medical practice variation.

Research on trends in medical practice variations is rare (Groenewegen and Westert, 2004). From our model hypotheses can be derived concerning the circumstances under which variation in medical practice would increase or decrease. If circumstances are more alike, physicians' medical behavior is expected to show less variation. Circumstances can become more alike through institutional change. An example of institutional change is the trend from professional control to managerial control (Scott et al., 2000). Instead of norms, rules become important. The implication is that when rules are set, monitored and sanctioned at the level of the hospital, then variation within hospitals would decrease. If they are set at a national level, for instance

performance indicators, variation is expected to decrease at a national level for medical decisions included in the performance indicators. The questions to be answered in this study are:

1 *'Are there changes in medical practice variation over time?'*
   a *'What are the conditions under which medical practice variation decreases?'*
   b *'At which level does a change in medical practice variation occur?'*

Institutional mechanisms

In this book only one example for each of the three institutional mechanisms, the regulative, the normative and the cultural-cognitive example, was examined. For each example it was found that there was a relationship with medical behavior. However, the effects of medical practice variation were not unambiguous. It should be examined whether other examples of the institutional mechanism, such as pay-for-performance initiatives and accreditation by physicians, show similar results. Furthermore, institutional mechanisms could interact. How they interact, and what their relative importance is, was not examined in this book. They are separately included in our model, each is assumed to have an effect on the circumstances under which physicians take their decisions. The questions to be answered in this study are:

2 *'Does our model apply for different applications of the institutional mechanisms?'*
   a *'What are the applications of institutional mechanisms for which the model does (not) apply?'*
   b *'What are the conditions under which changes in medical practice variation do (not) occur?'*
   c *'How do the institutional mechanisms interact?'*
   d *'What does this mean for the theoretical model?'*

Relevant actors

Our basic assumption is that physicians decide over treatment. With the changing position of physicians, the relative importance of physicians in taking medical decisions could change. The balance of power between physicians, financiers and patients could change. Financiers are becoming more involved in the medical decisions because of agreements with health care providers (Saltman and Figueras, 1998). Shared decision-making increases the influence of patients on medical decisions (O'Connor et al.,

2004). It should be examined whether our model is able to deal with these changes; can these changes be viewed as changes in circumstances under which physicians take their decisions? In the first place, we should examine if patients are actively involved in medical decision-making and should therefore be modeled as actors. Third parties mostly provide incentives instead of being actively included in the decision-making process. Because of this, they can be modeled as circumstances. If the model falls short, the influence of these actors should be incorporated in the theoretical model. The questions to be answered in this study are:

3 *'Does our basic assumption, with physicians as central actors, hold for different circumstances?'*
   a *'What are the circumstances under which physicians are (not) the central actors?'*
   b *'To what extent should the influence of other actors like patients and third parties be modeled?'*

The study proposals described above will elaborate on an existing theoretical (behavioral) model that explains medical practice variations. They concern a relevant societal problem and can provide important suggestions for health care policy. This book, as well as the study proposals described above, mainly concerned the theory underlying medical practice variations. Therefore, based on the results of this book no conclusions can be drawn concerning the quality of care. Knowledge of the relationship between medical practice variation and quality of care is necessary for developing policy that is aimed at limiting medical practice variation while increasing quality.

Medical practice variation and quality of care

The existence of variation raises all kinds of questions, and therefore it is important to understand it. A step further is to reduce it. However, it is unknown whether less variation is related to better quality of care (Fertig, 1993; Weide et al., 1992). It is unknown whether the existence of variation has advantages, for instance in the development of new medical techniques. In order to gain insight into the relationship between quality and variation a multidisciplinary approach of medical practice variation would be appropriate. A combination of medical knowledge and knowledge from the social sciences, that is knowledge on behavior, might add to our understanding of both medical practice variation and its relationship to aspects of quality. The social sciences search for the mechanisms underlying medical practice variation. The medical sciences can indicate when variation

is undesirable. From these sciences the relationship between variation and quality can be determined. They can give an in-depth consideration of the quality of medical decisions.

A further step is to examine possible interventions. When the mechanisms are known, it becomes clear where interventions could be directed. Whether this actually works in practice should be tested. Interventions could have both desirable and undesirable, both expected and unexpected, effects. These should be evaluated. The questions to be answered in this study are:

4 *'What is the relationship between medical practice variations and quality of care?'*
   a *'How can knowledge of medical practice variations be used to improve the quality of care?'*
   b *'What are the consequences of interventions aiming at decreasing medical practice variation and increasing the quality of care?'*

This research proposal thus aims at refining the theory underlying medical practice variation, to understand the causes of medical practice variation. Furthermore, it aims at using this knowledge in order to improve medical practice.

Samenvatting (Summary in Dutch)

Inleiding

In dit boek zijn verschillende hypotheses met betrekking tot de oorzaken van variaties in medisch handelen empirisch getoetst. In navolging van Freidson (1975) en Westert en Groenewegen (1999) is gekozen voor een benadering waarbij de nadruk ligt op de sociale organisatie. De sociale organisatie is van invloed op het medisch handelen van artsen door het bieden van mogelijkheden en beperkingen. Deze benadering voorspelt variatie tussen artsen die onder verschillende omstandigheden werken en overeenkomsten tussen artsen die een werkomgeving delen, zoals een praktijk of een ziekenhuis. In dit boek is bevestigd dat vooral omstandigheden belangrijk zijn bij het verklaren van variatie in medisch handelen.

Drie mechanismen die variatie tussen en overeenkomsten binnen groepen veroorzaken zijn getoetst: selectie, geleidelijke aanpassing aan de groepsnorm en snelle aanpassing aan omstandigheden. De resultaten wezen in de richting van gedeelde omstandigheden als oorzaak van overeenkomsten. De resultaten lieten ook zien dat de meest gebruikelijke verklaring van variatie in medisch handelen die is gebaseerd op individuele preferenties tekort schoot. De implicatie is dat variatie in medisch handelen niet zozeer is gebaseerd op individuele verschillen in preferenties voor een bepaalde manier van handelen, maar dat patronen van variatie worden veroorzaakt door sociale processen in samenwerkingsverbanden en lokale omstandigheden. Dit wetende, zijn we verder gegaan met het toetsen van institutionele mechanismen en hun effect op variatie.

De relatie tussen institutionele mechanismen en omstandigheden is dat institutionele mechanismen, zoals we ze gebruiken in dit boek, een effect hebben op de omstandigheden die van invloed zijn op variatie. Scott (2001) onderscheidt drie mechanismen wat betreft de werking van instituties: een regulatief, een normatief en een cultureel-cognitief mechanisme. De hypothese was dat deze drie mechanismen van invloed zijn op variatie. Het toetsen van de institutionele mechanismen was belangrijk voor het begrijpen van variatie in medisch handelen. Behalve een verklaring bieden de institutionele mechanismen ook instrumenten die kunnen worden gebruikt om variatie te beïnvloeden. We hebben laten zien dat institutionele mechanismen effectief zijn in het beïnvloeden van het medische gedrag van artsen. Daarom zouden ze gebruikt kunnen worden in het beperken van variatie. Hier hebben we in dit boek echter geen empirisch bewijs voor gevonden. Dit boek heeft een bijdrage geleverd aan het begrijpen van

variatie in medisch handelen en de mechanismen die daaraan ten grondslag liggen, zonder daarbij een normatieve opinie met betrekking tot het bestaan van variatie te betrekken. Met dit boek hebben we vooruitgang geboekt in het verklaren van variatie in medisch handelen en in de methodologie die gebruikt kan worden om variatie in medisch handelen te analyseren.

Onderzoeksvragen en hypothesen

In het eerste hoofdstuk zijn twee verschillende soorten onderzoeksvragen geformuleerd: een beschrijvende vraag met betrekking tot de mate van variatie, en een vraag die betrekking heeft op verklaringen voor variatie in medisch handelen. De tweede vraag was de belangrijkste onderzoeksvraag.

1 *'In welke mate is er variatie in medisch handelen tussen (groepen van) artsen?'*

2 *'Hoe kunnen variaties in medisch handelen tussen (groepen van) artsen worden verklaard, nadat gecorrigeerd is voor medisch relevante factoren?'*

In hoofdstuk 2 is de hypothese getoetst dat variatie clustert binnen maatschappen van huisartsen in Nederland als medisch handelen of de consequenties daarvan zichtbaar zijn. Aangenomen werd dat zichtbaarheid een belangrijke voorwaarde was voor het ontstaan van overeenkomsten. In hoofdstuk 3 is onderzocht wat de mechanismen zijn die ervoor zorgen dat huisartsen binnen maatschappen in Nederland op elkaar lijken. Drie mechanismen zijn gespecificeerd: selectie, geleidelijke aanpassing aan de groepsnorm en snelle aanpassing aan omstandigheden. In hoofdstuk 4 is een toets uitgevoerd om te bepalen of omstandigheden of preferenties variaties in medisch handelen verklaren. De hypothese dat artsen die in twee ziekenhuizen werken verschillend handelen in deze twee ziekenhuizen is getoetst met data uit de Verenigde Staten. Het medisch handelen van artsen onder restrictieve (regulatieve) omstandigheden is onderzocht in hoofdstuk 5. Het ging hier om managed care activiteiten door verzekeraars in de Verenigde Staten. In hoofdstuk 6 is het medisch handelen van artsen onder restrictieve (normatieve) omstandigheden onderzocht. Hierbij ging het om professionele richtlijnen voor huisartsen in Nederland. In hoofdstuk 7 is de invloed van cultureel-cognitieve processen op het medisch handelen van huisartsen in Nederland bestudeerd. De hypothese dat artsen die gebruik

maken van een elektronisch beslissysteem minder variëren in voorschrijfgedrag is getoetst.

Data en methode

Om de hypotheses empirisch te toetsen zijn datasets geconstrueerd. Hiervoor is gebruik gemaakt van secundaire data. Data zijn gebruikt van twee nationale dataverzamelingen onder huisartsen in Nederland: de eerste en tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk (NS1 en NS2). Deze data zijn verzameld in respectievelijk 1987 en 2001. Daarnaast zijn data gebruikt van het "New York Statewide Planning and Research Cooperative System" (SPARCS). Hiervan zijn data van de jaren 1999, 2000 en 2001 gebruikt. De analyses in dit boek zijn gedaan met multilevel analyse (tabel 1).

Tabel 1 Gebruikte data voor het toetsen van de hypotheses

Hoofdstuk	**Hypothese cluster**	**Gebruikte data**
2	Omstandigheden	Data zijn gebruikt van de tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk (NS2), verzameld in 2001. In totaal namen 195 Nederlandse huisartsen werkzaam in 104 praktijken met 390.000 patiënten deel aan deze studie. Solopraktijken zijn uitgesloten van de analyses. Afhankelijk van de analyses zijn tussen de 29 en 36 praktijken, 56 en 68 huisartsen, 88,553 en 107,739 contacten meegenomen in de analyses.
3	Omstandigheden	Data zijn gebruikt van de eerste nationale studie naar ziekten en verrichtingen in de huisartspraktijk (NS1), verzameld in 1987. In totaal namen 161 Nederlandse huisartsen werkzaam in 103 praktijken met 335,000 patiënten deel aan deze studie. Solopraktijken zijn uitgesloten van de analyses. In totaal zijn 96 huisartsen werkzaam in 42 praktijken meegenomen in de analyses.
4 and 5	Omstandigheden. Het regulatieve mechanisme	Data zijn verkregen voor de jaren 1999, 2000 en 2001 van het New York Statewide Planning and Research Cooperative System (SPARCS). We hebben de SPARCS-data van de jaren 1999, 2000 en 2001 gebruikt om zeven DRGs te bestuderen: twee medische, een chirurgische en vier obstetrische. Meer dan 1000 artsen en 150 ziekenhuizen zijn per DRG meegenomen in de analyses.

- table 1 continues-

- table 1 continued -

Hoofdstuk	Hypothese cluster	Gebruikte data
6	Het normatieve mechanisme	Data zijn gebruikt van de eerste en tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk (NS1 en NS2), verzameld in respectievelijk 1987 and 2001. In 1987 waren er nog geen richtlijnen, terwijl er in 2001 75 richtlijnen beschikbaar waren (www.nhg.nl). In totaal zijn 160 huisartsen werkzaam in 102 praktijken uit NS1 en 108 huisartsen werkzaam in 69 praktijken uit NS2 meegenomen in de analyses.
7	Het cultureel-cognitieve mechanisme	Data zijn gebruikt van de tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk (NS2), verzameld in 2001. In totaal namen 195 Nederlandse huisartsen werkzaam in 104 praktijken met 390.000 patiënten deel aan deze studie. Data warden verzameld met betrekking tot contacten, patiënten, huisartsen en praktijken. Hierbij werd gebruik gemaakt van elektronische medische dossiers en vragenlijsten. In totaal zijn 82 diagnoses, 749,811 cases, 133 huisartsen, and 85 praktijken meegenomen in de analyses.

Belangrijkste resultaten

Het belang van omstandigheden

In hoofdstuk 2 hebben we onderzocht of zichtbaarheid van klinische activiteiten en het gebruik van gedeelde bronnen van invloed is op variatie in medisch handelen tussen huisartsen. Idee was dat variatie in medisch handelen tussen artsen gerelateerd is aan verschillen in prikkels en beperkingen tussen hun werkomgevingen. De werkomgeving biedt mogelijkheden en beperkingen voor medisch handelen. Een belangrijk aspect van het werken in dezelfde omgeving, dezelfde praktijk, is dat huisartsen dezelfde bronnen gebruiken, zoals personeel en materiaal. De werkomgeving werkt als een sociaal, economisch en technologisch systeem waarin beslissingen worden genomen en dit systeem kan er de oorzaak van zijn dat verschillende artsen dezelfde medische beslissingen nemen. Zichtbaarheid van medisch handelen is een belangrijke voorwaarde voor het ontstaan van overeenkomsten die gebaseerd zijn op groepsnormen. Als handelen, of de uitkomst van handelen, zichtbaar is en consequenties heeft voor anderen, kunnen normen worden ontwikkeld. Artsen lopen het risico bekritiseerd te worden door andere artsen die in dezelfde maatschap werken, omdat ze het handelen van elkaar zien. Dit risico kan beperkt

worden door dezelfde handelingen te laten zien als collega's. Om de hypothese te toetsen zijn verschillende klinische activiteiten bestudeerd, te weten voorschrijven, verwijzen, diagnostiek, behandelen en het geven van advies. Data afkomstig uit de tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk zijn gebruikt. In tabel 2 wordt een samenvatting gegeven van de hypotheses, hoe ze zijn getoetst en wat de resultaten waren.

We hebben gevonden dat variatie tussen praktijken afhangt van de klinische activiteit. Huisartsen verschilden meer van collega's die in anderen praktijken werken dan van collega's die in dezelfde praktijk werken voor activiteiten als behandeling en diagnostiek in de praktijk. Beide activiteiten zijn zichtbaar voor collega's in dezelfde praktijk. Het omgekeerde hebben we gevonden voor voorschrijven, verwijzen naar de eerste lijn, verwijzen naar de tweede lijn, diagnostiek in een extern laboratorium en voor het geven van advies. Voor deze activiteiten was er minder variatie tussen praktijken dan binnen praktijken; huisartsen die in dezelfde praktijk werken toonden niet meer overeenkomsten dan huisartsen die in verschillende praktijken werken. Deze klinische activiteiten zijn minder zichtbaar voor collega's en er wordt geen gebruik gemaakt van gedeelde bronnen. Onze hypothese, dat er minder variatie tussen praktijken is als er gebruik wordt gemaakt van gedeelde bronnen en als medisch handelen zichtbaar is voor collega's, werd bevestigd.

Deze bevindingen gaven ons ideeën met betrekking tot de vraag of er variatie in medisch handelen binnen of tussen praktijken is. De bevindingen duiden erop dat maatschappen zich inderdaad ontwikkelen in een normatieve gemeenschap voor die medische handelingen die zichtbaar zijn voor collega's die in dezelfde praktijk werken. Een volgende stap was het meer specifiek zoeken naar oorzaken van patronen van variatie in medisch handelen.

In hoofdstuk 3 zijn we verder ingegaan op de vraag waarom artsen die een werkomgeving delen overeenkomsten vertonen door de hypothese te toetsen dat artsen in een maatschap meer op elkaar lijken wat betreft professionele attitude en gedrag dan op willekeurig gekozen artsen.

Drie verschillende mechanismen zijn gespecificeerd die patronen van variatie in medisch handelen kunnen verklaren: selectie van nieuwe collega's kan gericht zijn op overeenkomsten (soort zoekt soort), geleidelijke aanpassing aan elkaar in een maatschap kan het resultaat zijn van

bevestigingsprocessen en, tot slot, de omstandigheden die gedeeld worden door collega's kunnen leiden tot overeenkomsten in medisch handelen. Dit zijn drie algemene processen waarvan kan worden verwacht dat ze leiden tot overeenkomsten binnen en verschillen tussen groepen. Data afkomstig uit de eerste nationale studie naar ziekten en verrichtingen in de huisartspraktijk zijn gebruikt. Verschillende attitudes en handelingen zijn onderzocht. In tabel 2 wordt een samenvatting gegeven van de hypotheses, hoe ze zijn getoetst en wat het resultaat was.

We hebben gevonden dat Nederlandse huisartsen die in dezelfde praktijk werken meer overeenkomsten vertonen in attitudes en medisch handelen dan huisartsen die niet in dezelfde praktijk werken. De meeste indicaties wijzen op het belang van omstandigheden, en in mindere mate op geleidelijke aanpassing, voor de verklaring van overeenkomsten binnen maatschappen. Dit gaf ons een eerste aanwijzing voor een verklaring van overeenkomsten op basis van omstandigheden. Er is geen bewijs gevonden dat selectie gebaseerd op overeenkomsten in leeftijd en geslacht een verklaring vormt voor overeenkomsten tussen huisartsen die een werkomgeving delen. Deze bevinding houdt niet zonder meer in dat selectie geen rol speelt in het verklaren van overeenkomsten in medisch handelen binnen maatschappen, maar alleen dat selectie niet gebaseerd lijkt te zijn op overeenkomsten in leeftijd en geslacht. De implicatie van deze bevindingen is dat variatie in medisch handelen niet veroorzaakt wordt door individuele verschillen in voorkeuren voor een bepaalde manier van handelen, maar dat patronen van variatie in medisch handelen ontstaan door sociale processen in een samenwerkingsverband en lokale omstandigheden.

In hoofdstuk 4 hebben we getoetst of omstandigheden in plaats van voorkeuren variatie verklaren. De hypothese dat artsen die in verschillende ziekenhuizen werken in die ziekenhuizen anders handelen is getoetst. Dit zou betekenen dat er weinig variatie binnen ziekenhuizen was vergeleken met variatie tussen ziekenhuizen voor artsen die vergelijkbare patiënten behandelen. Een tweede implicatie zou zijn dat artsen die in meer dan een ziekenhuis werken zich conformeren aan de gebruikelijke gang van zaken in elk ziekenhuis (Westert, 1992). Deze implicatie kan niet worden afgeleid van een aanpak gebaseerd op voorkeuren, omdat de voorkeuren van een individu niet zouden veranderen als er in een ander ziekenhuis wordt gewerkt.

Tabel 2 Het belang van omstandigheden: beschrijving van de manier waarop de hypotheses zijn getoetst

Beschrijving van de hypothese	**Hoofdstuk**	**Data bron**	**Methode**	**Resultaat**
De meeste variatie komt voor in de minst zichtbare situatie	2	NS2	Variatie voor verschillende medische beslissingen, meer of minder zichtbaar voor collega's, is vergeleken	Bevestigd
Artsen in een maatschap lijken meer op elkaar wat betreft professionele attitude en gedrag dan random gekozen artsen	3	NS1	Overeenkomsten in attitudes en gedrag tussen artsen die in dezelfde maatschap werken zijn vergeleken met overeenkomsten tussen artsen die in verschillende maatschappen werken	Bevestigd
Er is minder variatie tussen artsen die in hetzelfde ziekenhuis werken dan tussen ziekenhuizen	4	SPARCS	Variatie tussen ziekenhuizen is vergeleken met variatie binnen ziekenhuizen	Bevestigd
Artsen die in verschillende ziekenhuizen werken waarbij de gemiddelde ligduur tussen de ziekenhuizen verschilt, kiezen een ligduur die vergelijkbaar is met de gemiddelde ligduur van het ziekenhuis waar de patiënt wordt behandeld	4	SPARCS	De ligduur gekozen door artsen die in twee ziekenhuizen werken is vergeleken met de gemiddelde ligduur in die ziekenhuizen	Bevestigd
Hoe groter het aandeel artsen dat in één ziekenhuis werkt, hoe minder variatie in de gekozen ligduur er is tussen artsen in dat ziekenhuis	4	SPARCS	De potentiële invloed van de gekozen ligduur door artsen die in één ziekenhuis werken op de totale variatie in een ziekenhuis is gemeten met behulp van een covariantiemaat met multilevel analyse. Op deze manier is berekend welk deel van de variatie kan worden toegeschreven aan het aandeel artsen dat in één ziekenhuis werkt.	Niet bevestigd

Voor het toetsen van de hypothese is gebruik gemaakt van secundaire data, afkomstig uit de Verenigde Staten (New York Statewide Planning and Research Cooperative System (SPARCS). We hebben artsen die in een ziekenhuis en artsen die in meer dan een ziekenhuis werkten onderzocht, om te bepalen of de gemiddelde ziekenhuisligduur verschilt afhankelijk van het ziekenhuis waar de arts zijn/haar patiënten behandeld. In tabel 2 wordt een samenvatting gegeven van de hypotheses, hoe ze zijn getoetst en wat het resultaat was.

De hypothese dat er weinig variatie tussen ziekenhuizen is vergeleken met variatie tussen ziekenhuizen is bevestigd. Dit resultaat is echter nog steeds verenigbaar met zowel een aanpak die uitgaat van voorkeuren voor een bepaalde wijze van medisch handelen als met een aanpak die gebaseerd is op werkomstandigheden. Daarom hebben we ook de medische beslissingen van dezelfde artsen in verschillende ziekenhuizen geanalyseerd. Voorkeuren voor een bepaalde manier van handelen worden verondersteld betrekkelijk stabiel te zijn voor dezelfde persoon, maar omstandigheden kunnen verschillen. Verschillende analyses bevestigen dat artsen die in twee ziekenhuizen werken, waarbij de gemiddelde ziekenhuisligduur verschilt tussen de ziekenhuizen, een ziekenhuisligduur hebben die vergelijkbaar is met de gebruikelijke praktijk in het ziekenhuis waar de procedure is uitgevoerd.

Dit onderzoek liet zien dat eenheden van organisaties, zoals ziekenhuizen, belangrijk zijn in het onderzoeken van variatie in medisch handelen. Het is bevestigd dat omstandigheden belangrijk zijn in het begrijpen van variaties in medisch handelen.

Institutionele mechanismen
Scott (2001) onderscheidt drie mechanismen met betrekking tot de werking van organisaties: een regulatief, een normatief en een cultureel-cognitief mechanisme. Deze drie mechanismen kunnen variatie beïnvloeden. We hebben elk van de drie institutionele mechanismen getoetst, door een voorbeeld van elk mechanisme te kiezen, in drie aparte studies. Als eerste hebben we een voorbeeld van het regulatieve mechanisme beschreven, daarna een voorbeeld van het normatieve mechanisme en tot slot hebben we een voorbeeld van het cultureel-cognitieve mechanisme gegeven.

Het managed care systeem in de Verenigde Staten is een voorbeeld van het regulatieve mechanisme; verzekeraars maken regels voor de behandeling van patiënten. Het normatieve mechanisme heeft betrekking op

verwachtingen van collega's, en past bij professionele controle. Richtlijnen die door de professie zelf ontwikkeld zijn, zijn hiervan een voorbeeld. Dit zijn richtlijnen met een normatief karakter. Het cultureel-cognitieve mechanisme heeft betrekking op een gemeenschappelijk denkkader.

In hoofdstuk 5 hebben we de invloed van managed care op variatie in medisch handelen onderzocht. Het managed care systeem in de Verenigde Staten is een voorbeeld van een restrictieve omstandigheid waarbij het regulatieve mechanisme wordt gebruikt om medisch handelen te beïnvloeden. De Health Maintenance Organisatie (HMO) is het meest bekende type van een managed care organisatie in de Verenigde Staten (Bachman en Freeborn, 1999).

We hebben ons gericht op een belangrijke set van beperkingen, namelijk de beperkingen die worden bepaald door de verzekeraar van de patiënt. Beperkingen voor mensen met een HMO verzekering zullen veel restrictiever zijn dan die voor mensen met een traditionele verzekering. Dit impliceert dat er minder variatie is tussen artsen die mensen met een HMO verzekering behandelen dan tussen artsen die mensen met een andere verzekering behandelen. Artsen die mensen met een HMO verzekering behandelen worden geconfronteerd met dezelfde beperkingen en prikkels, wat zal leiden tot een vergelijkbare beslissing wat betreft ziekenhuisligduur.

Op basis van de verschillen in beperkingen tussen managed en niet-managed care verzekeringen hebben we de hypothese afgeleid dat de ligduur voor mensen met een managed care verzekering korter is en dat er minder variatie in ligduur is voor mensen met een managed care verzekering. We hebben ligduur voor vergelijkbare patiënten die een managed care verzekering of een niet-managed care verzekering hebben bestudeerd. Data uit de jaren 1999, 2000 en 2001 van ziekenhuizen uit de staat New York zijn gebruikt en geanalyseerd met multilevel analyse. In tabel 3 wordt een samenvatting gegeven van de hypotheses, hoe ze zijn getoetst en wat het resultaat was.

De resultaten lieten geen verschil in ligduur tussen mensen met een managed care en mensen met een niet-managed care verzekering zien. Verder bleek dat er minder variatie in ligduur was voor mensen met een managed care verzekering. In tegenstelling tot onze verwachtingen hebben we dit niet vooral gevonden op het niveau van de arts of op ziekenhuisniveau.

Tabel 3 Het regulatieve mechanisme: beschrijving van de manier waarop de hypotheses zijn getoetst

Beschrijving van de hypothese	**Hoofdstuk**	**Data bron**	**Methode**
De ligduur voor managed care patiënten is korter	5	De gemiddelde ligduur voor de managed care en de niet managed care groepen zijn vergeleken	Niet bevestigd
Er is minder variatie in ligduur voor managed care patiënten	5	Variatie in ligduur voor de managed care groep en de niet managed care groep zijn vergeleken	Bevestigd
De invloed van managed care, namelijk minder variatie in ligduur, zal voornamelijk gevonden worden op artsniveau	5	Variatie voor de managed care groep en de niet managed care groep op artsniveau is vergeleken met de variatie voor beide groepen op ziekenhuisniveau	Niet bevestigd
Hoe groter het aandeel van managed care patiënten dat de arts behandelt, hoe korter de ligduur (a) en hoe minder variatie in ligduur op artsniveau (b)	5	De regressiecoëfficiënt voor het aandeel managed care patiënten per arts is bestudeerd. Tevens is de covariantie tussen dit aandeel en de variatie in ligduur bestudeerd. De verwachting was dat beide negatief zouden zijn	Bevestigd (a)/ Niet bevestigd (b)
Hoe groter het aandeel managed care patiënten dat behandeld wordt in het ziekenhuis, hoe korter de ligduur in dat ziekenhuis (a) en hoe minder variatie in ligduur op het ziekenhuisniveau (b)	5	De regressiecoëfficiënt voor het aandeel managed care patiënten per ziekenhuis is bestudeerd. Teven is de covariantie van dit aandeel en de variatie in ligduur bestudeerd. De verwachting was dat beide negatief zouden zijn	Niet bevestigd (a)/ niet bevestigd (b)
Artsen die met minder verschillende zorgverzekeraars te maken hebben, hebben minder variatie in ligduur voor managed care patiënten	5	De covariantie tussen het aantal verzekeraars per arts en de variatie in ligduur is bestudeerd. De verwachting was dat deze positief zou zijn	Bevestigd
Ziekenhuizen die met minder verschillende zorgverzekeraars te maken hebben, hebben minder variatie in ligduur voor managed care patiënten	5	De covariatie tussen het aantal verzekeraars per ziekenhuis en de variatie in ligduur is bestudeerd. De verwachting was dat deze positief zou zijn	Bevestigd

- vervolg tabel 3 -

- vervolg tabel 3 -

Beschrijving van de hypothese	Hoofdstuk	Data bron	Methode
Artsen die in minder verschillende ziekenhuizen werken hebben minder variatie in ligduur voor managed care patiënten	5	De covariantie tussen het aantal ziekenhuizen per arts en de variatie in ligduur is bestudeerd. De verwachting was dat deze positief zou zijn	Niet bevestigd
Hoe eenvoudiger het is om een behandeling van een bepaalde DRG te standaardiseren, hoe minder variatie in ligduur er is voor managed care patiënten	5	Variatie voor alle DRGs is vergeleken. Voor medische DRGs en DRGs met complicaties werd de meeste variatie verwacht	Bevestigd

We vonden een verschil in het effect van het aandeel managed care patiënten op artsen en op ziekenhuizen. Wat artsen betreft bleek dat de ligduur korter was naarmate het aandeel groter was. Dit effect is voor ziekenhuizen niet gevonden. Dit betekent dat de ligduurkeuze van artsen wordt beïnvloed door de verzekeraar, maar dat het ziekenhuis haar beleid niet aanpast. Het aandeel managed care patiënten had geen eenduidig effect op variatie voor artsen en ziekenhuizen. Dit indiceert dat er geen effect van de verzekeraar is. De mogelijkheid om patiënten in een ander ziekenhuis te behandelen had geen invloed op variatie in ligduur voor managed care patiënten, wat een effect zou zijn van de verzekeraar in combinatie met minder variatie als het aantal verzekeraars lager lag.

We hebben geen bewijs gevonden dat het managed care is dat een effect heeft op ligduur en dus op de kosten die samenhangen met opnamedagen. Het lijkt erop dat er iets anders is dat ervoor zorgt dat de ligduur in de Verenigde Staten kort is, onafhankelijk van de verzekering van de patiënt. Wat we hebben gevonden is dat het niet de beperkingen zijn die door de verzekeraar worden gesteld die resulteren in patronen van variatie, omdat er nauwelijks verschillen in ligduur zijn tussen patiënten met een managed care verzekering en patiënten zonder managed care verzekering. Het zou kunnen zijn dat ziekenhuizen reageren op de manier waarop ze worden gefinancierd; betaling per DRG betekent dat het altijd (kosten)efficiënt is om de ligduur zo kort mogelijk te houden. Ook kan het zijn dat de wetenschap dat managed care aan het toenemen is ertoe leidt dat ziekenhuizen alvast maatregelen toepassen om er zeker van te zijn dat ze in de toekomst (managed care) patiënten hebben (Zhang et al., 1999). Het lijkt erop dat

ziekenhuizen zorg meer managen dan de verzekeraar, zeer waarschijnlijk vanwege financiële prikkels.

In hoofdstuk 6 hebben we het handelen van artsen onderzocht onder restrictieve normatieve omstandigheden, namelijk professionele richtlijnen voor huisartsen in Nederland. We verwachtten dat variatie zou afnemen als richtlijnen worden gevolgd. Het volgen van richtlijnen houdt in dat je bij een groep hoort en het geeft artsen status en bescherming als medisch handelen moet worden uitgelegd. Door het volgen van richtlijnen wordt het eenvoudiger om uit te leggen wat je doet en om geaccepteerd te worden. We hebben twee belangrijke mechanismen bediscussieerd die van invloed zijn op de afname van variatie door de introductie van richtlijnen: onzekerheid over medisch handelen en integratie in de professie. Op basis van deze twee mechanismen kunnen verschillen tussen artsen in solopraktijken en artsen in groepspraktijken worden verwacht.

We hebben getoetst of er meer uniformiteit in het voorschrijven van geneesmiddelen is gekomen, met andere woorden of er minder variatie in voorschrijven is gekomen nadat er richtlijnen zijn geïntroduceerd. Of ze nu gevolgd worden of niet. Om dit te onderzoeken hebben we gebruik gemaakt van de data van twee grote studies (NS 1 en NS2). De ene is gehouden voor de introductie van richtlijnen, de andere is gehouden nadat richtlijnen zijn geïntroduceerd. De richtlijnen zijn ontwikkeld door het Nederlands Huisartsen Genootschap (NHG). Voor beide studies zijn groepen van diagnoses gemaakt en vergeleken. Een groep bestond uit alle diagnoses waarvoor richtlijnen zijn geïntroduceerd, de richtlijnengroep. De andere groep bestond uit alle diagnoses waarvoor geen richtlijnen zijn geïntroduceerd, de referentiegroep. In tabel 4 wordt een samenvatting gegeven van de hypotheses, hoe ze zijn getoetst en wat het resultaat was.

We hebben gevonden dat hoewel er meer verschillende geneesmiddelen werden voorgeschreven tijdens NS2 voor zowel de referentie als de richtlijnengroep, de toename in het aantal verschillende middelen hoger was voor de referentiegroep. De totale toename van het aantal verschillende geneesmiddelen dat werd voorgeschreven kan worden verklaard door het feit dat er veel meer verschillende geneesmiddelen beschikbaar waren tijdens NS2 dan tijdens NS1. Dat de toename kleiner was voor de richtlijnengroep kan indiceren dat richtlijnen het effect van de beschikbaarheid van geneesmiddelen matigen.

Tabel 4 Het normatieve mechanisme: beschrijving van de manier waarop de hypotheses zijn getoetst

Beschrijving van de hypothese	Hoofdstuk	Databron	Methode
Door de introductie van richtlijnen is variatie in voorschrijven afgenomen	6	Verschillen in een concentratiemaat tussen twee groepen van diagnoses (een richtlijnengroep en een referentiegroep) zijn vergeleken voor en na de introductie van richtlijnen	Niet bevestigd
Variatie in voorschrijven is meer afgenomen voor artsen die in solo praktijken werken dan voor arsten die in groepspraktijken werken	6	Verschillen in een concentratiemaat tussen artsen die in een solopraktijk werken en artsen die in een groepspraktijk werken voor twee groepen van diagnoses (een richtlijnen en een referentie groep) zijn apart bestudeerd	Bevestigd

We hebben geen significant verschil gevonden tussen de referentiegroep en de richtlijnengroep voor huisartsen in groepspraktijken tussen NS1 en NS2. Dit suggereert dat richtlijnen geen effect hebben op de variatie voor huisartsen in groepspraktijken. In overeenkomst met onze hypothese hebben we wel gevonden dat richtlijnen een effect hebben op de variatie in solopraktijken; variatie is minder toegenomen in de richtlijnengroep.

Op basis van deze studie concluderen we dat de introductie van richtlijnen, hoewel ze waarschijnlijk de toename van variatie dempen, variatie niet verminderen. De introductie van richtlijnen alleen is niet genoeg om medisch handelen te veranderen en variatie te verminderen. Het is een stap verder om te interveniëren in de dagelijkse routines van artsen, in plaats van het gebruiken van regels en voorschriften. Dit kan worden gedaan door gebruik te maken van gecomputeriseerde beslissystemen, een voorbeeld van het cultureel-cognitieve mechanisme.

In hoofdstuk 7 hebben we de invloed van cultureel-cognitieve processen op variatie onderzocht. Van het cultureel-cognitieve mechanisme wordt verwacht dat er een effect is op variatie omdat als artsen eenzelfde raamwerk gebruiken bij het maken van hun keuzes het waarschijnlijk is dat ze tot eenzelfde besluit komen (onder de voorwaarde dat er niet veel ruimte is voor eigen interpretatie). We hebben een beslissysteem bestudeerd dat

door huisartsen in Nederland wordt gebruikt voor advies omtrent het voorschrijven, gegeven de diagnose. Het beslissysteem stelt een geneesmiddel voor, bij de gegeven diagnose, waarbij rekening wordt gehouden met leeftijd, geslacht en comorbiditeit.

We verwachtten dat het gebruik van een beslissysteem leidt tot minder variatie tussen artsen, omdat ze gebruik maken van hetzelfde cognitieve raamwerk. Beslissystemen grijpen op de dagelijkse routine van artsen aan. Ze worden gebruikt om bij voorbeeld het gebruik van richtlijnen te ondersteunen. Data zijn gebruikt van de tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk (NS2). Deze zijn verzameld in 2001. In tabel 5 wordt een samenvatting gegeven van de hypotheses, hoe ze zijn getoetst en wat het resultaat was.

Tabel 5 Het cultureel-cognitieve mechanisme: beschrijving van de manier waarop de hypotheses zijn getoetst

Beschrijving van de hypothese	**Hoofdstuk**	**Data bron**	**Methode**
Artsen die een beslissysteem gebruiken schrijven geneesmiddelen voor volgens de richtlijnen waarop het beslissysteem is gebaseerd	7	Voorschriften door artsen die een beslissysteem gebruiken en artsen die dat niet doen zijn vergeleken met de voorschriften die geadviseerd worden in de richtlijnen	Bevestigd
Er is minder variatie in voorschrijven tussen artsen die een beslissysteem gebruiken	7	Een maat voor de concentratie van het voorschrijven van geneesmiddelen is vergeleken tussen artsen die een beslissysteem gebruiken en artsen die dat niet doen	Niet bevestigd

Deze studie liet zien dat artsen die een DSS gebruiken meer volgens het advies dat wordt gegeven voorschrijven dan artsen die geen DSS gebruiken. Dit betekende niet dat er minder variatie was; variatie was hetzelfde voor artsen die een DSS gebruiken en voor artsen die dat niet doen. Dit is waarschijnlijk gerelateerd aan het feit dat het DSS dat we bestudeerd hebben verschillende geneesmiddelen of een stapsgewijze behandeling adviseert waarbij begonnen wordt met een bepaald type geneesmiddel dat later vervangen kan worden als dat nodig is. Implicaties van de studie zijn dat een DSS kan worden gebruikt bij de implementatie van richtlijnen, maar dat niet moet worden verwacht dat variatie wordt beperkt. Variatie wordt waarschijnlijk alleen beperkt als een DSS geen ruimte geeft voor variatie, als

er bij voorbeeld slechts één advies wordt gegeven met betrekking tot het geneesmiddel dat kan worden voorgeschreven.

Wetenschappelijke implicaties van de resultaten

De wetenschappelijke implicaties bestaan uit het verder toetsen en uitbreiden van het theoretisch model voor het verklaren van variaties in medisch handelen. Het geschikte analyseniveau is, zoals we eerder bespraken, één van de kwesties die naar voren komen in de discussie over verklaringen voor variatie van medisch handelen. We argumenteerden dat het optreden van niet-willekeurige variatie wijst op verklarende theorieën die het individu overstijgen. Beslissingen die op het eerste gezicht individueel lijken, zijn waarschijnlijk niet zo individueel. Mensen nemen over het algemeen geen beslissingen in een vacuüm: ze worden beïnvloed door de omstandigheden waaronder zij beslissingen nemen. De omstandigheden creëren kansen en belemmeringen. In navolging van Coleman (1990) en Freidson (1975), redeneerde Westert (1992) dat medisch gedrag van individuele huisartsen beïnvloed wordt door beperkingen in de directe werkomgeving, daarmee aangevend dat het geschikte analyseniveau de werkomgeving is. Toch is onderzoek op het gebied van variatie in medisch handelen meestal regionaal georiënteerd. We hebben het model van Westert (1992) uitgebreid door institutionele mechanismen aan het model toe te voegen. Deze institutionele mechanismen in relatie tot variaties in medisch handelen zijn nooit eerder systematisch onderzocht.

Met ons theoretisch model voorspelden we dat:
- *variatie tussen huisartsen die een werkomgeving delen geringer is dan variatie tussen huisartsen die in verschillende werkomgevingen werken, vanwege selectie, geleidelijke aanpassing en omstandigheden;*
- *regulatieve, normatieve, en cultureel-cognitieve mechanismen variatie in medisch handelen verminderen binnen de eenheden waarop zij van toepassing zijn.*

De onderzoeksresultaten bevestigden dat de omstandigheden belangrijk zijn om variaties in medisch handelen te kunnen verklaren. Ze lieten zien dat maatschappen een normatieve gemeenschap vormen voor gedrag dat zichtbaar is voor artsen die in eenzelfde maatschap werken. Er zijn inderdaad aanwijzingen gevonden dat huisartsen geen autonome beslissingen nemen. De onderzoeksresultaten lieten zien dat het niveau waarop variaties onderzocht zouden moeten worden, het niveau is van de

omgeving waarin artsen werken. We hebben gevonden dat artsen die een werkomgeving delen gelijkenissen in gedrag vertonen.
Onze verwachtingen ten aanzien van de invloed van institutionele mechanismen op variaties in medisch handelen werden niet empirisch bevestigd. Onze resultaten laten zien dat deze mechanismen medisch gedrag wel beïnvloeden, maar dat zij variatie in medisch handelen niet noodzakelijkerwijs verminderen. De resultaten impliceren echter wel dat institutionele mechanismen kunnen worden gebruikt als instrumenten om variatie in medisch handelen te beïnvloeden, omdat gedrag kan worden beïnvloed. Of variatie nu wel of niet wordt verminderd, hangt waarschijnlijk af van hoeveel ruimte er is voor variatie binnen de regels en voorschriften.

We hebben een bestaand theoretisch model uitgebreid, het model van lokale standaarden dat ontwikkeld werd door Westert (1992), en ontleenden daar nieuwe hypothesen aan. Deze hypothesen hebben we getoetst met gebruikmaking van internationale data van zowel ziekenhuizen als huisartsenpraktijken. Multilevel analyse is gebruikt om de data te analyseren. We toonden aan dat het theoretisch model opgaat voor verschillende landen en niet alleen van toepassing is op ziekenhuiszorg, maar ook op de huisartsenpraktijk.

Methodologische benadering

Dit boek is een combinatie van zes afzonderlijke onderzoeken. Ieder onderzoek heeft zijn eigen methodologische benadering. Wat de zes onderzoeken gemeen hebben is dat secundaire data geanalyseerd worden met behulp van multilevel analyse. Allereerst bespreken we de algemene methodologische benadering; dan volgt een beschrijving van de beperkingen en de sterke punten van het onderzoek.

Algemene methodologische benadering
Voor het toetsen van de hypotheses werden secundaire data van drie verschillende databases gebruikt. Het voordeel van het gebruik van secundaire data is dat het tijd bespaart: de data zijn er al. Het nadeel is dat de data niet verzameld werden om onze specifieke hypothesen te toetsen. We gebruikten data van twee nationale studies die werden gehouden onder huisartsen in Nederland, en van ziekenhuizen in de Verenigde Staten. Deze verschillende datasets werden gebruikt omdat zij de beste mogelijkheid boden om de hypotheses te toetsen. Om een voorbeeld te noemen: om de hypothese te toetsen dat artsen die op verschillende plaatsen werken ook

verschillend handelen op die plaatsen, hadden we data nodig over artsen die op meer dan één plaats werken. In Nederland doet die situatie zich zelden voor, terwijl in de Verenigde Staten het vrij algemeen is dat artsen in meer dan één ziekenhuis werken.

Omdat we geïnteresseerd waren in universele mechanismen en we niet een louter beschrijvend onderzoek deden, is het minder belangrijk dat data van dezelfde dataset werden gebruikt, verzameld op dezelfde tijd en dezelfde plaats. De mechanismen zullen in verschillende datasets hetzelfde zijn, verzameld op verschillende tijden en plaatsen. Variatie komt overal ter wereld voor. Het patroon van variatie blijkt overeen te komen bij algemene procedures in "per verrichting betaalde" ziekenhuismarkten, ondanks de duidelijke verschillen in aantallen artsen, de organisatie en financiering van diensten en de culturele en demografische kenmerken van de inwoners van het gebied waar de ziekenhuismarkt zich bevindt (Wennberg, 2004).

Een belangrijke overeenkomst tussen de datasets is het feit dat de data gaan over artsen die in een bepaalde omgeving werken. Dit was van invloed op de analyses. Omdat sociale omstandigheden overeenkomsten veroorzaken tussen artsen die in dezelfde omgeving werken, staan hun klinische activiteiten betreffende hun patiënten niet los van elkaar. Hiermee moet ook in de analyses rekening worden gehouden. De data zijn hiërarchisch geordend. Daarom gebruikten we multilevel modellen om de data te analyseren (Hox, 1995; Snijders en Bosker, 1999).

Beperkingen van de methodologie

Hoewel we methoden zochten die het best pasten bij de te toetsen hypotheses, heeft de gekozen methodologie een aantal beperkingen. Allereerst noemen we de mogelijkheid van compositie-effecten. We waren geïnteresseerd in variatie tussen artsen en maatschappen of ziekenhuizen. De mogelijkheid van compositie-effecten moet zoveel mogelijk worden uitgesloten; immers, het doel van dit boek is niet een beschrijving te geven van de verschillen tussen artsen, maatschappen of ziekenhuizen, gebaseerd op patiëntenselectie. We hebben deze mogelijkheid uitgesloten door te corrigeren voor case-mix, maar de mogelijkheid blijft bestaan dat compositie-effecten de resultaten beïnvloeden. In het onderzoek naar de duur van ziekenhuisopname (hoofdstuk 4), konden we geen onderscheid maken tussen invloeden van collega's en die van andere omstandigheden. Dit hadden we wel kunnen doen als er rekening was gehouden met de specifieke kenmerken van de ziekenhuizen. Die kenmerken waren niet beschikbaar. De patiëntenselectie kon ook een verklaring zijn voor variatie

tussen ziekenhuizen. Artsen die in meer ziekenhuizen werken passen wellicht een criterium toe dat te maken heeft met duur van verblijf, wanneer zij een ziekenhuis uitzoeken voor hun patiënten. Dit zou de theorie echter weer bevestigen dat het ziekenhuis de duur van de opname beïnvloedt.

In andere onderzoeken (bijvoorbeeld hoofdstuk 2) hielden we geen rekening met specifieke diagnoses. Dat was niet mogelijk omdat er teveel diagnoses waren en er geen model kon worden geschat. Verschillen in diagnoses in de huisartsenpraktijk zijn een belangrijke bron van variatie in klinische activiteiten. Dit zou echter weerspiegeld worden in alle klinische activiteiten. Als verschillen in diagnoses een verklaring waren van de overige variatie, zouden alle klinische activiteiten patronen moeten vertonen van variatie per praktijk (als de diagnoses tussen de praktijken verschillen) of per arts (als de diagnoses tussen artsen verschillen). Dat hebben we niet gevonden en daarom concluderen we dat verschillen in diagnoses niet het belangrijkst waren voor de verklaring van variatie tussen praktijken.

Sterke punten van de methodologie

Een belangrijk voordeel van de manier waarop het onderzoek in dit boek is gedaan, is dat het de mogelijkheid bood om veel onderzoek te doen. We hebben een groot aantal verschillende hypotheses kunnen toetsen, omdat het niet nodig was om data te verzamelen. Er werden data gebruikt uit verschillende situaties; we hoefden dus niet te wachten op veranderingen in de situaties.

We gebruikten multilevel analyses om de data te analyseren. Door deze statistische techniek te gebruiken, ontstaan er geen problemen van informatieverlies door aggregatie of het onterecht verkleinen van de standaardfouten wanneer kenmerken van een hoger niveau aan eenheden op lager niveau worden toegekend. Met multilevel analyse wordt de totale variatie in klinische activiteit in verschillende delen gesplitst. Bijvoorbeeld: een gedeelte vanwege verschillen tussen patiënten, een gedeelte vanwege verschillen tussen huisartsen en een gedeelte vanwege verschillen tussen praktijken en ziekenhuizen. Dit is de basismethodologie die in elk van de onderzoeken werd gehanteerd. We waren bij alle onderzoeken in de eerste plaats geïnteresseerd in variatie en hoe variatie over de verschillende niveaus verdeeld was.

De methodologie die we gebruikten om de hypotheses te toetsen was aan de ene kant de best beschikbare en, aan de andere kant, een uitdaging. Hoewel statistische technieken voortdurend verder worden ontwikkeld en

vergevorderde technieken geavanceerde analyses mogelijk maken bij het toetsen van hypotheses, bereikten we de grenzen van wat, voor zover wij weten, mogelijk was met de huidige technieken. We moesten daarom naar oplossingen zoeken in de richting van hoe men idealiter de data zou analyseren. Het ligt in de lijn der verwachtingen dat het in de toekomst mogelijk zal zijn de hypotheses met nog verfijndere analyses te toetsen. We zouden zeker nieuwe technieken toepassen wanneer die beschikbaar zijn, maar we denken dat de gebruikte methoden de beste waren die we konden bedenken. De resultaten leveren een bijdrage aan het inzicht in variatie in medisch handelen.

Implicaties van de onderzoeksresultaten voor het gezondheids-beleid

Het doel van dit boek was om alle mechanismen te ontdekken die ten grondslag liggen aan variatie in medisch handelen. De resultaten van dit onderzoek hebben implicaties voor het gezondheidsbeleid. In de inleiding vermeldden we hoe belangrijk het is om variaties in medisch handelen te doorgronden vanuit verschillende perspectieven: het perspectief van de resultaten voor gezondheidsuitkomsten, dat van de consument of patiënt, het perspectief van de professies in de gezondheidszorg, en het perspectief van derden, zoals zorgverzekeraars. Een belangrijke vraag, vanuit al deze perspectieven, in de context van beleidsimplicaties van dit onderzoek is: (hoe) kan het gedrag van artsen worden beïnvloed? Voor de implicaties voor het gezondheidsbeleid zijn in feite twee vragen belangrijk. Allereerst, verklaren de uitkomsten waarom het bestaande gezondheidsbeleid dat zich richt op het beïnvloeden van het gedrag van artsen (niet) effectief is? Op de tweede plaats, hebben de uitkomsten consequenties die positief of negatief kunnen worden gewaardeerd en moet het beleid zich erop richten deze consequenties te bevorderen of te veranderen? Dit heeft te maken met de vraag of variatie in medisch handelen goed of slecht is. In de laatste paragraaf besteden we hier aandacht aan.

De uitkomsten van de eerste drie onderzoeken (hoofdstukken 2, 3 en 4) lieten zien dat artsen zich aanpassen aan hun collega's of aan de omstandigheden in de praktijk of het ziekenhuis waar zij procedures volgen. Zo ontstaat er uniformiteit in medisch gedrag. Bovendien ontdekten we dat wanneer gedrag zichtbaar is of gemeenschappelijke bronnen gebruikt, artsen wederzijds afhankelijk zijn en het gedrag uniformer wordt. We leerden uit de laatste drie onderzoeken (hoofdstukken 5, 6 en 7) dat regulatieve,

normatieve en cognitief-culturele druk het gedrag beïnvloeden, maar variatie in medisch handelen niet verminderen. Deze bevindingen hebben implicaties voor het gezondheidsbeleid.

Bijvoorbeeld: de eerste uitkomst, het belang van de omstandigheden, heeft implicaties voor beleid dat zich erop richt ervoor te zorgen dat artsen werken volgens op wetenschappelijk bewijs gebaseerde ("evidence-based") standaarden. Niet alle artsen houden zich aan deze standaarden. Om in overeenstemming met deze standaarden te werken, moeten zij hun medisch gedrag wijzigen. Voor het beleid betekent dat, dat activiteiten om de standaarden te introduceren gericht moeten worden op artsen op de relevantie eenheden, en niet op de individuele artsen. Individuele cursussen zullen bijvoorbeeld minder effectief zijn dan plaatselijke consensusbijeenkomsten, zoals binnen praktijken.

De uitkomsten betreffende institutionele mechanismen en zichtbaarheid van het gedrag hebben consequenties voor het huidige beleid dat "pay-for-performance" en prestatie-indicatoren introduceert. "Pay-for-performance" verbindt over het algemeen kwaliteitsmaatstaven aan financiële prikkels. Prestatie-indicatoren zijn in de eerste plaats bedoeld om kwaliteit te monitoren. We veronderstellen dat zowel "pay-for-performance" en prestatie-indicatoren de zichtbaarheid en wederzijdse afhankelijkheid verhogen. Zij maken deel uit van een verandering van beheer vanuit de beroepsgroep naar beheer door het management. *"De maatschappij verwacht dat de sector gezondheidszorg opener, verantwoordelijker is, althans, wat betreft het gebruiken van bronnen die zij claimen van de publieke sector. Elk instituut moet zijn begrotingsbehoeften verdedigen, ieder jaar weer, en de toename in benodigde bronnen rechtvaardigen, door een verband te leggen tussen de verwachte uitgaven en de verwachte activiteiten."* (Van Kemenade, 2007). Derden, zoals zorgverzekeraars, monitoren het gedrag van artsen en proberen het bronnengebruik te beperken. Onze onderzoeksresultaten betekenen voor beleid dat "pay-for-performance" en prestatie-indicatoren introduceert dat het zal resulteren in meer uniformiteit binnen de lokale eenheden (praktijken). Na verloop van tijd, als alle praktijken door dezelfde prestatie-indicatoren worden geleid, zal dit beleid resulteren in meer uniformiteit tussen de praktijken. Ook wordt verondersteld dat het introduceren van "pay-for-performance" de wederzijdse afhankelijkheid zal verhogen, meer dan prestatie-indicatoren in het algemeen doen, vanwege de financiële prikkels. Dit betekent dat de introductie van "pay-for-performance" homogeniteit versterkt, meer dan prestatie-indicatoren zonder directe financiële consequenties.

De uitkomst dat de institutionele mechanismen gedrag wel beïnvloeden maar variatie in medisch handelen niet verminderen, heeft implicaties voor veranderingen op de lange termijn in de gezondheidszorg die ons denken over de gezondheidszorg veranderen van een professionele dienstverlening naar een product, en die de patiënten veranderen in consumenten van gezondheidszorgproducten. We nemen aan dat cognitief-culturele druk alleen leidt tot meer uniformiteit als daarmee het aantal alternatieven voor artsen op effectieve wijze wordt beperkt. Ook nemen we aan dat gezondheidszorgproducten die gemakkelijk kunnen worden gestandaardiseerd, in afzonderlijke klinieken zullen worden geleverd, want productie in een hoog volume zal financieel aantrekkelijk zijn. De implicatie van de onderzoeksresultaten is dat er meer homogeniteit zal zijn voor behandelingen die gestandaardiseerd kunnen worden en in grote volumes kunnen worden uitgevoerd in klinieken.

Deze zelfde uitkomst heeft implicaties voor patiënten: zij gedragen zich meer als consumenten en hebben steeds meer invloed op wat artsen doen. De keuze van patiënten wordt belangrijker. Meer aandacht voor de keuze van patiënten is een instrument dat past bij marktcontrole (Saltman en Figueras, 1998). Men veronderstelt dat normatieve druk van collega's steeds meer wordt tegengewerkt door de voorkeuren van patiënten. Het relatieve belang van de mening van collega's verschuift naarmate patiënten belangrijker worden bij het nemen van beslissingen. Hoeveel druk patiënten uit kunnen oefenen op artsen verschilt per patiënt, zo ook hun vermogen om over de behandeling te discussiëren. Er zijn groepen patiënten, zoals taalminderheden of ouderen, die minder goed in staat zijn druk uit te oefenen op artsen. Dat betekent dat bij normatieve druk door collega's, bijvoorbeeld door middel van richtlijnen, artsen meer volgens deze richtlijnen handelen met patiënten die niet in staat zijn over de behandeling te discussiëren, dan met patiënten die de arts kunnen vragen om een andere behandeling dan de behandeling die door de richtlijnen wordt aanbevolen. Binnen de eerste groep verwachten we meer homogeniteit door normatieve druk van collega's dan bij andere patiënten.

We hebben een paar voorbeelden gegeven van de beleidsimplicaties van de onderzoeksresultaten en we hebben laten zien dat ons model gebruikt kan worden om veranderingen in variatie in medisch handelen te voorspellen als er nieuwe ontwikkelingen plaatsvinden. We hebben deze voorspellingen echter niet empirisch getoetst. Het is daarom interessant trends te monitoren in variaties in medisch handelen en hypotheses te toetsen die aan ons

theoretisch model zijn ontleend. Deze trends, en andere onderwerpen, zouden onderwerp voor toekomstig onderzoek kunnen zijn.

Voorstel voor toekomstig onderzoek

Achtergrond
Vanaf de jaren dertig in de vorige eeuw heeft er onderzoek plaatsgevonden naar variatie in medisch handelen. Het was echter pas in de jaren zeventig– met een publicatie van Wennberg en Gittelsohn (1973) in Science - dat deze variaties meer aandacht kregen in de Verenigde Staten. In de jaren tachtig kreeg men in Europa meer belangstelling voor dit onderwerp. Sinds die tijd zijn variaties in medisch handelen voor allerlei soorten medische behandelingen goed gedocumenteerd. Men ontdekte dat deze variaties niet konden worden verklaard aan de hand van verschillen in patiënten-populaties. Hoewel men veel vooruitgang boekte in het beschrijven van variatie in medisch handelen, werd er minder vooruitgang geboekt in de verklaring voor variaties in medisch handelen. Men nam over het algemeen aan dat het gedrag van de arts een belangrijke rol speelt in de verklaring van variaties in medisch handelen. Toch schoot de theorie over de relatie tussen het gedrag van artsen en variatie in medisch handelen tekort. De verklaringen waren vooral gebaseerd op verschillen in voorkeuren tussen artsen. De invloed van de omstandigheden werd over het hoofd gezien.

Westert en Groenewegen (1996) stelden een aanvullende benadering voor om het probleem op te lossen dat plotselinge veranderingen in gedrag niet verklaard konden worden en dat hypotheses ontleend aan de benadering waarin voorkeuren centraal staan, moeilijk te toetsen waren. In hun benadering waarin beperkingen centraal staan, benadrukten zij (sociale) condities of lokale omstandigheden die het gedrag van artsen beïnvloeden door kansen of beperkingen te bieden. Het voordeel van deze benadering is dat er een expliciete relatie bestaat tussen de werkgerelateerde omstandigheden, het medische gedrag van artsen en variatie in medisch handelen. Aan de hand van deze benadering ontwikkelde Westert (1992) een theoretisch model, een model van lokale standaarden, om geografische verschillen in verblijfsduur in het ziekenhuis voor operaties in Nederland te verklaren. De volgende vraag was of dit model ook van toepassing was op andere medische beslissingen, en of het model verder zou moeten worden uitgewerkt. In dit boek wordt het allebei gedaan. Het model werd getoetst voor huisartsenpraktijken in Nederland en voor ziekenhuizen in de Verenigde Staten. Daarnaast is het model uitgebreid met institutionele

mechanismen. Het bleek waardevol voor de verklaring van variaties in medisch handelen. Er is echter nog steeds onontgonnen terrein. We kunnen op drie manieren toekomstig onderzoek uitstippelen. Allereerst kunnen de implicaties van de hypotheses en/of onderzoeksresultaten worden getoetst die nog niet zijn onderzocht. Dit zou onderzoek naar trends in tijd en de interactie tussen institutionele mechanismen inhouden. In de tweede plaats kunnen hypotheses worden onderzocht die in dit onderzoek niet worden bevestigd en wat de redenen daarvoor zijn. Deze manier van denken zou leiden tot verdere toetsing van de hypotheses of tot een uitbreiding van het theoretische model. De derde mogelijkheid is van een heel andere orde en betreft veeleer de implicaties van onderzoek naar medisch handelen voor het beleid, dan de theorie die variatie in medisch handelen verklaart. Het gaat hier om de relatie tussen variaties in medisch handelen en de kwaliteit van de zorg. We geven nu een algemene schets van de voorgestelde onderzoeken.

Tijdtrends in variaties in medisch handelen

Wat we in het onderdeel over implicaties voor gezondheidsbeleid schreven wijst in de richting van veranderingen in variaties in medisch handelen in de loop van de tijd. Variaties in medisch handelen worden beïnvloed door allerlei ontwikkelingen in de maatschappij en het gezondheidszorgsysteem. Bovendien is er steeds meer kennis met betrekking tot medisch handelen en misschien is er voor meer diagnoses dan in het verleden het geval was ("evidence based") consensus over de meest geschikte behandeling. Onzekerheid binnen de professie zou daarmee minder worden en dus ook de mogelijkheden voor individuele overtuigingen beperken. Dientengevolge vermindert de variatie voor deze diagnoses. Kennis van trends en de oorzaak van veranderingen in variatie dragen bij aan het inzicht in variatie in medisch handelen.

Onderzoek naar trends in variaties in medisch handelen is schaars (Groenewegen en Westert, 2004). Aan ons model kunnen hypotheses worden ontleend over de omstandigheden waaronder variatie in medisch handelen zou toenemen of verminderen. Het medische gedrag van artsen toont naar verwachting minder variatie naarmate de omstandigheden meer gelijkenis vertonen. Door institutionele veranderingen kunnen de omstandigheden meer gelijkenis vertonen. Een voorbeeld van institutionele verandering is de trend van professionele controle naar controle door het management (Scott et al., 2000). In plaats van normen worden regels belangrijk. De implicatie is dat wanneer de regels zijn vastgesteld, gemonitord en gesanctioneerd op ziekenhuisniveau, variatie binnen

ziekenhuizen minder zou worden. Als zij op landelijk niveau zijn vastgesteld, bijvoorbeeld in de vorm van prestatie-indicatoren, vermindert naar verwachting de variatie op landelijk niveau voor medische beslissingen die in de prestatie-indicatoren besloten zitten. In dit onderzoek moeten de volgende vragen worden beantwoord:

1 *'Zijn er in de loop van de tijd veranderingen in variatie in medisch handelen?'*
   a *'Onder welke condities vermindert variatie in medisch handelen?'*
   b *'Op welk niveau doet zich verandering in variatie in medisch handelen voor?'*

Institutionele mechanismen

In dit boek werd slechts één voorbeeld voor elk van de drie institutionele mechanismen, het regulatieve, het normatieve en het cultureel-cognitieve, onderzocht. Bij ieder voorbeeld werd aangetoond dat er een relatie was met medisch gedrag. Echter, de effecten voor variatie in medische praktijk waren niet eenduidig. Er is nader onderzoek nodig om te zien of andere voorbeelden van institutionele mechanismen, zoals "pay-for-performance" initiatieven en accreditatie door artsen gelijksoortige resultaten laten zien. Bovendien zou er een wisselwerking kunnen zijn tussen de institutionele mechanismen. Hoe die wisselwerking eruit ziet en wat het relatieve belang van de verschillende mechanismen is, werd in dit boek niet onderzocht. Ze zijn afzonderlijk in ons model opgenomen, in de veronderstelling dat elk een effect heeft op de omstandigheden waaronder artsen hun beslissingen nemen. In dit onderzoek dienen de volgende vragen te worden beantwoord:

2 *'Is ons model van toepassing voor verschillende toepassingen van de institutionele mechanismen?'*
   a *'Voor welke toepassingen van institutionele mechanismen is ons model (niet) van toepassing?'*
   b *'Wat zijn de condities waaronder veranderingen in variatie in medisch handelen (niet) optreden?'*
   c *'Is er wisselwerking tussen de institutionele mechanismen?'*
   d *'Wat betekent dit voor het theoretische model?'*

Relevante actoren

Onze basisassumptie is dat artsen beslissen over de medische behandeling. Met de veranderende positie van artsen kan het relatieve belang van artsen bij het nemen van medische beslissingen veranderen. Het machtsevenwicht

tussen artsen, financiers en patiënten kan veranderen. Financiers raken meer betrokken bij de medische beslissingen door overeenkomsten met zorgaanbieders (Saltman en Figueras, 1998). Gedeelde besluitvorming vergroot de invloed van patiënten op medische beslissingen (O'Connor et al., 2004). Er moet onderzocht worden of ons model stand houdt bij deze veranderingen; kunnen we deze veranderingen zien als veranderingen in de omstandigheden waaronder artsen hun beslissingen nemen? Ten eerste moeten we onderzoeken of patiënten actief betrokken zijn bij medische besluitvorming en derhalve als actoren moeten worden opgenomen. Derden geven meestal prikkels in plaats van dat zij actief betrokken zijn bij het besluitvormingsproces. Daarom kunnen ze als omstandigheden worden gemodelleerd. Als het model tekort schiet, moet de invloed van deze actoren in het theoretische model worden geïncorporeerd. In dit onderzoek dienen de volgende vragen te worden beantwoord:

3 *'Gaat onze basisveronderstelling, met artsen als centrale actoren, op onder verschillende omstandigheden?'*
   a *'Onder welke omstandigheden zijn artsen (geen) centrale actoren?'*
   b *'In hoeverre moet de invloed van andere actoren zoals patiënten en derden worden gemodelleerd?'*

De hierboven beschreven onderzoeksvoorstellen bouwen voort op een bestaand theoretisch (gedrags)model dat variaties in medisch handelen verklaart. Ze betreffen een relevant maatschappelijk probleem en kunnen belangrijke suggesties geven voor het gezondheidszorgbeleid. Dit boek en de hierboven beschreven onderzoeksvoorstellen gaan hoofdzakelijk over de theorie die ten grondslag ligt aan variaties in medisch handelen. Daarom kunnen er op basis van de resultaten van dit boek geen conclusies worden getrokken ten aanzien van de kwaliteit van de zorg. Kennis van de relatie tussen variatie in medisch handelen en de kwaliteit van de zorg is onontbeerlijk voor het ontwikkelen van beleid dat erop gericht is om variatie in medisch handelen te beperken en tegelijkertijd de kwaliteit te verhogen.

Variatie in medisch handelen en kwaliteit van de zorg

Het feit dat er variatie is roept allerlei vragen op. Het is daarom belangrijk het te begrijpen. Het is een volgende stap om deze variatie te verminderen, hoewel we niet weten of minder variatie ook een betere kwaliteit van de zorg betekent (Fertig, 1993; Weide et al., 1999). Het is niet bekend of het bestaan van variatie voordelen heeft, bijvoorbeeld bij de ontwikkeling van nieuwe medische technieken. Om meer inzicht in de relatie tussen kwaliteit

en variatie te krijgen, zou een multidisciplinaire benadering van variatie in medisch handelen op zijn plaats zijn. Een combinatie van medische kennis en kennis van de sociale wetenschappen, dus inzicht in gedrag, kan ons inzicht in zowel variatie in medisch handelen als de relatie met kwaliteitsaspecten vergroten. Vanuit de sociale wetenschappen wordt gezocht naar de mechanismen die ten grondslag liggen aan variatie in medisch handelen; vanuit de medische wetenschappen kan worden aangeven wanneer variatie niet gewenst is. Op basis van kennis vanuit de medische wetenschappen kan de relatie tussen variatie en kwaliteit worden bepaald. Met behulp van medische kennis kan de kwaliteit van medische beslissingen beschouwd worden.

De volgende stap is het onderzoek naar mogelijke interventies. Als de mechanismen bekend zijn, wordt duidelijk waar interventies zich op kunnen richten. Of het ook werkt in de praktijk, dient getoetst te worden. Interventies kunnen gewenste of ongewenste, verwachte of onverwachte effecten hebben. Deze moeten worden geëvalueerd. In dit onderzoek dienen de volgende vragen te worden beantwoord:

4 *'Wat is de relatie tussen variaties in medisch handelen en de kwaliteit van de zorg?'*
   - a *'Hoe kunnen we kennis van variaties in medisch handelen gebruiken om de kwaliteit van de zorg te verbeteren?'*
   - b *'Wat zijn de consequenties van interventies die erop gericht zijn variatie in medisch handelen te verminderen en de kwaliteit van de zorg te verbeteren?'*

Dit onderzoeksvoorstel is dus gericht op de theorie die ten grondslag ligt aan variatie in medisch handelen, om de oorzaken van variatie in medisch handelen te achterhalen. Het voorstel is er bovendien op gericht om deze kennis te gebruiken om medisch handelen te verbeteren.

References

American Medical Association. *Principles of Managed Care.* Chicago: AMA, Council on Medical Service, 1999

Anatomical Therapeutic Chemical (ATC) classification index. Oslo: WHO Collaborating Centre for Drug Statistics Methodology, 1993

Anderson GF, Reinhardt UE, Hussey PS, Petrosyan V. It's prices, stupid: why the United States is so different from other countries. *Health Aff* ,2003; 22(3):89-103

Anderson TF, Mooney G. Medical practice variations: where are we? In: Anderson TF, Mooney G (eds). *The challenges of medical practice variations.* London: The Macmillan Press, 1990:1-15

Arndt M, Bradbury RC, Golec J. Indications for hysterectomy: variation within and across hospitals. *Med Care Res Reviews*, 1995; 52:342-63

Arnould R, DeBrock L. *Utilization Control in HMOs: faculty working paper 90-1698.* University of Illinois, 1990

Ashton CM, Petersen NJ, Souchek J, Menke TJ, Yu H et al. Geographic variations in utilization rates in Veterans Affairs hospitals and clinics. *New Engl J Med*, 1999; 340:32-9

Averill, Goldfield N, Steinbeck BA, Grant T. *All Patients Refined Diagnosis Related Groups (APRDRGs).* Wallingford, Connecticut, USA: 3M Health Information Systems, 2002

Bachman KH, Freeborn DK. HMO physicians' use of referrals. *Soc Sci Med,* 1999; 48(4):547-57

Bakwin H. Pseudodoxia Pediatrica. *New Engl Med,* 1945; (232):691-7

Bandura A. *Social foundations of thought and action: a social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986

Berg M. *Rationalizing medical work. Decision support techniques and medical practices*. [dissertation] Maastricht: Rijksuniversiteit Limburg, 1995

Block LE. Evolution, growth and status of managed care in the United States. *Public Health Rev,* 1997; 25:193-244

Blumenthal D. The variation phenomenon in 1994. *New Engl J Med,* 1994; 331:1017-8

Boerma WGW, Fleming DM. *The role of general practice in Primary Health Care*. London: World Health Organization, The Stationary Office, 1998

Bradbury RC, Golec JH, Stearns FE. Comparing hospital length of stay in independent practice association HMOs and traditional insurance programs. *Inquiry*, 1991; 28: 87-93

Brunsson N, Jacobsson B (eds). *A world of standards*. Oxford: Oxford University Press, 2000

Burgers J. De wetenschappelijke onderbouwing van de NHG-standaarden. *Huisarts en Wetenschap,* 1994, 37:188-93

Burgers JS, Zaat JOM, Spies TH, Van der Bij AK, Mokkink HGA, Grol RPTM. De kwaliteit van de NHG-standaarden: beoordeling van 130 kernaanbevelingen uit 28 standaarden. *Huisarts en Wetenschap,* 2002; 45(7): 349-53

Burkhardt ME. Social interaction effects following a technological change. *Acad Manage J*, 1994; 37(4):869-98

Burns LR, Wholey DR. The effects of patient, hospital, and physician characteristics on length of stay and mortality. *Med Care*, 1991; 29:251-71

Burns LR, Chilingerian JA, Wholey DR. The effect of physician practice organization on efficient utilization of hospital resources. *Health Serv Res,* 1994; 29(5):583-603

Burns LR, Geller SE, Wholey DR. The effect of physician factors on the cesarean section decision. *Med Care*, 1995; 33:365-82

Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH et al. Why don't physicians follow clinical practice guidelines? *JAMA,* 1999; 282(15):1458-65

Cardol M, Van Dijk L, De Jong JD, De Bakker DH, Westert GP. *Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk: huisartsenzorg: wat doet de poortwachter?* Utrecht: NIVEL, 2004

Carthy P, Harvey I, Brawn R, Watkins C. A study of factors associated with cost and variation in prescribing among GPs. *Fam Pract,* 2000; 17(1):36-41

Chassin MR. Explaining geographic variations: the enthusiasm hypothesis. *Med Care,* 1993; 31(5) (supplement):YS37-YS44

Coleman JS. *Foundations of Social Theory*. Cambridge: Harvard University Press, 1990

Coleman JS, Katz E, Menzel H. *Medical innovation: a diffusion study.* New York: The Bobbs-Merrill Company, Inc., 1966

Colombo F. *Towards more choice in social protection?: individual choice of insurer in basic mandatory health insurance in Switzerland. Labour market and social policy.* Occational paper number 53. Paris: OECD, 2001

Committee of the Health Council. *Gepaste zorg*. Zoetermeer: RVZ, 2004

Congdon P. Modeling multiple hospital outcomes: the impact of small area and primary care practice variation. *Int J Health Geographics,* 2006, 5: 50

Coscelli A. The importance of doctors' and patients' preferences in the prescription decision. *J Ind Econ*, 2000; 48:349-69

Coser RL. Authority and decision-making in a hospital: a comparative analysis. *Am Sociol Review*, 1957; 23(1):56-63

Cretin S, Farley DO, Dolter KJ, Nicholas W. Evaluating an integrated approach to clinical quality improvement: clinical guidelines, quality measurement, and supportive system design. *Med Care*, 2001; (39)8 Suppl 2:II-70-II-84

Crul BVM, Legemaate J. Uitspraak Tuchtcollege: protocol niet heilig. *Medisch Contact*, 2005; 39:1556-58

Dambrun M, Guimond S, Duarte S. The impact of hierarchy-enhancing vs. attenuating academic major on stereotyping: the mediating role of perceived social norm. *Crisp*, 2002; 7(8):114-35

Davis P, Gribben B, Lay-Yee R, Scott A. How much variation in clinical activity is there between general practitioners?: a multi-level analysis of decision-making in primary care. *J Health Serv Res Policy*, 2002; 7(4):202-8

Davis P, Gribben B. Rational prescribing and interpractitioner variation. *Int J Technol Assess Health Care*, 1995; 11:428-42

Davis P, Gribben B, Scott A, Lay-Yee R. Do physician practice styles persist over time?: continuities in patterns of clinical decision-making among general practitioners. *J Health Serv Res Policy*, 2000; (5)4:200-7

De Bakker DH, Coffie DSV, Heerdink ER, Van Dijk L, Groenewegen PP. Determinants of the range of drugs prescribed in general practice: a cross-sectional analysis. *BMC Health Serv Res*, 2007; 7:132

De Jong JD, Westert GP, Noetscher CM, Groenewegen PP. Does managed care make a difference?: physicians' length of stay decisions under managed and non-managed care. *BMC Health Serv Res*, 2004; 4:3

De Jong JD, Groenewegen PP, Westert GP. Mutual influences of general practitioners in partnerships. *Soc Science Med*, 2003; 57(8):1515-24

Delnoij DMJ. *Physician payment systems and cost control.* [dissertation] Utrecht: NIVEL, 1994

Delnoij DMJ, Spreeuwenberg PMM. Variations in GPs' referral rates to specialists in internal medicine. *Eur J Public Health*, 1997; 7(4):427-35

Den Hartog M, Janssen RTJM. De ontwikkeling van de markt-structuur van ziekenhuizen van 1979 tot 2000 in Nederland. *Acta Hospitalia*, 2000; 40(3):45-62

Denig P. *Drug choice medical practice rationales, routines, and remedies.* [dissertation] Groningen: Rijksuniversiteit Groningen, 1994

Denig P, Haaijer-Ruskamp FM, Zijsking DH. How physicians choose drugs? *Soc Sci Med*, 1988; 27:1381-6

Denton DW. The attraction-selection-attrition model of organizational behavior and the homogeneity of managerial personality. *Crisp*, 1999; 4(8):146-59

Diez Roux AV. A glossary for multilevel analysis. *J Epidemiol Community Health,* 2002; 56:588-94

Donelan K, Blendon RJ, Lundberg GD, Calkins DR, Newhouse JP et al. The new medical marketplace: physicians' views. *Health Aff,* 1997; 16(5):139-48

Eddy DM. Variations in physician practice: the role of uncertainty. *Health Aff,* 1984; 3(2):74-89

Eisenberg JM. The state of research about physicians' practice patterns. *Med Care*, 1985; 23(5):461-83

Erdmann Y, Wilson R. Managed care: a view from Europe. *Annu Rev Public Health*, 2001; 22:273-91

European Observatory on Health Care Systems. *Health Care Systems in transition.* Switzerland: WHO, 2000

Evans JH 3rd, Hwang Y, Nagarajan N. Physicians' response to length-of-stay profiling. *Med Care,* 1995; 33:1106-19

Evans RG. The dog in the night-time: medical practice variations and health policy. In: Andersen TF, Mooney G (eds.). *The challenges of medical practice variations*. London: Macmillan Press, 1990; 117-52

Fairfield G, Hunter DJ, Mechanic D, Rosleff F. Managed care: origins, principles and evolution. *BMJ*, 1997a; 314: 1823-6

Fairfield G, Hunter DJ, Mechanic D, Rosleff F. Managed care: implications of managed care for health systems, clinicians and patients. *BMJ,* 1997b; 314: 1895-8

Fehr B. *Friendship processes*. London: Sage Publications, 1996

Fertig A, Roland M, King H, Moore T. Understanding variation in rates of referral among general practitioners: are inappropriate referrals important and would guidelines help to reduce rates? *BMJ*, 1993; 307:1467-70

Fisher ES, Wennberg JE, Stukel ThA, Sharp SM. Hospital readmission rates for cohorts of medicare beneficiaries in Boston and New Haven. *N Engl J Med*, 1994; 331:989-95

Flap H. Professies als probleem: sociaal-wetenschappelijke vraagstellingen op het terrein van vrije beroepen. In: Arts W, Batenburg R, Groenewegen P (eds). *Een kwestie van vertrouwen*. Amsterdam: Amsterdam University Press, 2001

Foets M, Stokx L. Het voorschrijven van geneesmiddelen in de huisartspraktijk: de meest voorgeschreven werkzame stoffen. *Medisch Contact,* 1993; 48(16):489-92

Foets M, Van der Velden J, De Bakker D. *Dutch National Survey of General Practice: a summary of the survey design*. Utrecht: NIVEL, 1992

Foets M, Stokx L, Hutten J, Sixma H. *Basisrapport. De huisartsenquête: datareductie door schaalconstructie.* Utrecht: NIVEL, 1991

Freidson E. *Profession of medicine: a study of the sociology of applied knowledge.* New York: Dodd, Mead & Company, 1975, eight printing

Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux P et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA,* 2005, 293:1223-8

Geijer RMM, Burgers JS, Van der Laan JR, Wiersma T, Rosmalen CFH, Thomas S. *NHG-standaarden voor de huisarts I*. 2 ed. Maarssen: Elsevier/Bunge, 1999

Geller SE, Burns LR, Brailer DJ. The impact of nonclinical factors on practice variations: the case of hysterectomies. *Health Serv Res,* 1996; 30(6):729-50

Glover JA. The incidence of tonsillectomy in school children. *Proc R Soc Med*, 1938; 31:1219–36

Goldstein H. *Multilevel statistical models*. London: Hodder Headline, 1995

Greenfield S, Nelson EC, Zubkoff M, Manning W, Rogers W et al. Variations in resource utilization among medical specialties and systems of care: results from the medical outcomes study. *JAMA*, 1992; 267(12):1624-30

Griffiths M, Waters WE, Acheson ED. Variation in hospital stay after inguinal herniorrhaphy. *BMJ,* 1979; 6166 (1):787-9

Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet,* 2000; 355:103-6

Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care,* 2001; (39)8 Suppl 2: II-1-II-45.

Groenewegen PP, Hofland S. Maatschappen van specialisten: ontwikkeling in omvang en samenstelling. *Tijdschr Gezondheidswet*, 2003; 81(1):27-33

Groenewegen P, Marangos A. Maat gevraagd: eisen aan nieuwe leden van een maatschap. *Medisch Contact,* 2004; 59(8):278-81

Groenewegen PP, Van Lindert H. Vrij beroep in afhankelijkheid: de veranderende positie van medisch specialisten in de Nederlandse algemene ziekenhuizen. In: Arts W, Batenburg R, Groenewegen P (eds). *Een kwestie van vertrouwen.* Amsterdam: University Press, 2001

Groenewegen PP, Westert GP. Is there a time trend in medical practice variations?: a review of the literature and an critical analysis of theoretical approaches. *J Public Health*, 2004; 12(3):229-36

Groenewegen PP, De Bakker DH, Van der Velden J. *Basisrapport verrichtingen in de huisartspraktijk: Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk.* Utrecht: NIVEL, 1992

Groenewegen PP, De Jong JD, Kerssens JJ. *Voorkeursaanbieders: selectief contracteren in het nieuwe verzekeringsstelsel.* Utrecht: NIVEL, 2002a

Groenewegen PP, Dixon J, Boerma WGW. The regulatory environment of general practice: an international perspective. In: Saltman RB, Busse R, Mossialos E (eds.). *Regulating entrepreneurial behavior in European health care systems*. Buckingham: Open University Press, 2002b; 200-214

Grol R. Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care,* 2001; 39(8) Suppl 2: II-46-II-54

Grol R. De verspreiding van NHG-standaarden onder huisartsen. *Huisarts en Wetenschap*, 1989; 32(13):494-7

Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet,* 2003; 362(9391):1225-30

Grol R, Dalhuijsen J, Thomas S, In 't Veld C, Rutten G, Mokkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ*, 1998; 317:858-61

Gross PA, Pujat D. Implementing practice guidelines for appropriate antimicrobial usage: a systematic review. *Med Care*, 2001; 39(8) Suppl 2: II-55-II-69

Guthrie B. Why do general practitioners take blood? *Eur J Gen Pract,* 2001; 7:138-42,160

Hainsworth DS, Lockwood-Cook E, Pond M, Lagoe RJ. Development and implementation of clinical pathways for stroke on a multihospital basis. *J Neuroscience Nurs,* 1997; 29(3):156-62

Hays WL. *Statistics for the social sciences*. London: Holt, Rinehart and Winston 2th edition, 1973

Hellinger FJ. Selection bias in HMOs and PPOs: a review of the evidence. *Inquiry*, 1995; 32:135-42

Hermens RPMG, Hak E, Hulscher MEJL, Braspenning JCC, Grol RPTM. Adherence to guidelines on cervical cancer screening in general practice: programme elements of successful implementation. *Br J Gen Practice,* 2001; 51(472):897-903

Health Insurance Association of America. *Source book of health insurance data 1999-2000.* Washington: HIAA, 1999

Hibble A, Kanka D, Pencheon D, Pool F. Guidelines in general practice: the new Tower of Babel? *BMJ* , 1998; 317:862-3

Hill-Smith I. Sharing resources to create a district drug formulary: a county controlled trial. *Br J Gen Practice,* 1996; 46: 271-5

Hox JJ. *Multilevel analysis: techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates, 2002

Hox JJ. *Applied multilevel analysis*. Amsterdam: TT-Publications, 1995

Hulscher ME, Wensing M, Grol RP, Van der Weijden T, Van Weel C. Interventions to improve the delivery of preventive services in primary care. *Am J Public Health,* 1999; 89(5):737-46

Hulscher ME, Wensing M, Van der Weijden T, Grol R. Interventions to implement prevention in primary care. *Cochrane Database Systematic Rev* (1), 2001; CD000362

Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes. A systematic review. *JAMA* 1998, 1339-46

Hutten JBF. *Workload and provision of care in general practice*. [dissertation] Utrecht: NIVEL, 1998

Hyams AL, Brandenburg JA, Lipsitz SR, Shapiro DW, Brennan TA. Practice guidelines and malpractice litigation: a two way street. *Annals Int Med,* 1995; 122:450-5

Jaye C, Tilyard M. A qualitative comparative investigation of variation in general practitioners' prescribing patterns. *Br J Gen Pract,* 2002; 52(478):381-6

Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinical performance and patient outcome. *Annals Int Med,* 1994; 135-42

Kalberg S. Max Weber's types of rationality: cornerstones for the analysis of rationalization processes. *Am J Sociol*, 1980; 85(5):1145-79

Kalmijn M. Intermarriage and homogamy: causes, patterns, trends. *Annual Rev Sociol,* 1998; 24:395-421

Kamps GB. *Formularia en het voorschrijven door huisartsen*. [dissertation] Amsterdam: Meditekst, 1999

Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*, 2005: 330:765-8

Kerr EA, Mittman BS, Has RD, Siu AL, Leake B, Brook RH. Managed care and capitation in California: how do physicians at financial risk control their own utilization? *Ann Intern Med*, 1995; 123:500-4

Kimberley JR, De Pouvourville G. *The migration of managerial innovation.* San Francisco: Jossey Bass, 1993

Kralewski JE, Rich EC, Feldman R, Dowd BE, Bernhardt T, Johnson C, Gold W. The effects of medical group practice and physician payment methods on costs of care. *Health Serv Res,* 2000; 35(3):591-613

Krasnik A, Groenewegen PP, Pedersen PA, Von-Scholten P, Mooney G, Gottschau A, Flierman HA, Damsgaard MT. Changing remuneration systems: effects on activity in general practice. *BMJ*, 1990; 300:1698-701

Kroneman M. 2001. *Healthcare systems and hospital bed use*. [dissertation] Utrecht: NIVEL, 2001

Lagoe RJ. Benchmarking hospital admission rates by ages. *Healthcare Benchmarks*, 1998; 5(7):104-12

Lagoe RJ. Benchmarking hospital lengths of stay using histograms. *Nursing Economics*., 1999; 17(2):75-84, 102

Lagoe RJ, Noetscher CM, Murphy ME. Hospital readmission: predicting the risk. *J Nurs Care Qual*, 2001; 15(4):69-83

Lamberts H, Wood M, Hofmans-Okkes I. *The international classification of primary care in the European community*. New York: Oxford University Press, 1993

Langley GR, MacLellan AM, Sutherland HJ, Till JE. Effect of nonmedical factors on family physicians' decisions about referral for consultation. *CMAJ*, 1992; 147(5):659-66

Langley C, Faulkner A, Watkins Ch, Gray S, Harvey I. Use of guidelines in primary care-physicians' perspectives. *Fam Pract*, 1998; 15(2):105-11

Lave JR, Frank RG. Effect of the structure of hospital payment on length of stay. *Health Serv Res,* 1990; 25(2):327-47

Lazega E. *The collegial phenomenon; the social mechanisms of cooperation among peers in a corporate law partnership*. Oxford: University Press, 2001

Leenders RThAJ. *Structure and influence: Statistical models for the dynamics of actor attributes, network structure and their interdependence*. [dissertation] Amsterdam: Thesis Publishers, 1995

Leyland AH, Groenewegen PP. Multilevel modelling and public health policy. *Scand J Public Health*, 2003; 31:267-74

Lobach DF, Hammond WE. Computerized decision support based on a clinical practice guideline improves compliance with care standards. *Am J Med,* 1997; 102(1):89-98

Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice?: the effect of a consensus statement on the practice of physicians. *New Engl J Med,* 1989; 321(19):1306-11

Lulofs JG. Een markttheoretische benadering van professies. *Mens en Maatschappij,* 1981; 4(56):349-77

Lundin D. Moral hazard in physician prescription behavior. *J Health Econ,* 2000; 19(5):639-62

Marinus AMF. *Inter-dokter variatie in de huisartspraktijk*. [dissertation] Amsterdam: University of Amsterdam, 1993

McClure W. Toward development and application of a qualitative theory of hospital utilisation. *Inquiry,* 1982; 19:117-35

McKinley JB, Potter DA, Feldman HA. Non-medical influences on medical decision-making. *Soc Sci Med*, 1996; 42:769-76

McMahon LF Jr, Wolfe RA, Tedeschi PJ. Variation in hospital admissions among small areas: a comparison of Maine and Michigan. *Med Care,* 1989; 27(6):623-31

McNeil BJ. Hidden barriers to improvement in the quality of care. *New Engl J Med,* 2001; 345:1612-20

McPherson K, Strong PM, Epstein A, Jones L. Regional variations in the use of common surgical procedures: within and between England and Wales, Canada and the United States of America. *Soc Sci Med (A)*, 1981; 15(3 pt.1):273-88

McPherson K. Why do variations occur. In: Anderson TF, Mooney G (eds). *The challenges of medical practice variations*. London: The Macmillan Press, 1990; 16-35

Mechanic D. Managed care and the imperative for a new professional ethic. *Health Aff,* 2000; 19(5):100-11

Medicines Evaluation Board. *Annual report 1998.* Den Haag: Medicines Evaluation Board, 1999

Medicines Evaluation Board. *Annual report 2004.* Den Haag: Medicines Evaluation Board, 2005

Meulenberg F, Goudswaard L. Richtlijnontwikkeling in de huisartsgeneeskunde: een gevolg van professionalisering en kwaliteitsdenken. *Issue*, 1997; 4:32-4

Miller ME, Welch WP, Welch HG. The impact of practicing in multiple hospitals on physician profiles. *Medical Care,* 1996:. 34:455-62

Miller RH, Luft HS. Managed care plan performance since 1980: a literature analysis. *JAMA,* 1994; 271(19):1512-9

Mulder HC. *Het medisch kunnen: technieken, keuze en zeggenschap in de moderne geneeskunde.* [dissertation] Assen: Van Gorcum; 1996

Murphy M, Noetscher CM, Lagoe RJ. A multihospital effort to reduce inpatient lengths of stay for pneumonia. *J Nurs Care Qual,* 1999; 13(5):11-23

http://www.nhg.artsennet.nl/guidelines

Noetscher CM, Morreale GF, Lagoe RJ. Length of stay reduction: two innovative hospital approaches. *J Nurs Care Qual,* 2001; 16(1):1-14

Noetscher CM, Lagoe RJ, Hohner VK, Schmidt GM. Analyzing hospital readmissions using statewide data bases. *J Nurs Care Qual,* 1999; 13(6):57-67

O'Connor AM, Llewellyn-Thomas HA, Barry Flood A. Modifying unwarranted variations in health care: shared decision making using patient decision aids: a review of the evidence base for shared decision making. *Health Affairs*, 2004; Web exclusive VAR-63-VAR-72

O'Connor GT, Quinton HB, Traven ND, Ramunno LD, Dodds TA, Marciniak TA, Wennberg JE. Geographic variation in the treatment of acute myocardial infarction: the Cooperative Cardiovascular Project. *J Am Med Assoc,* 1999; 281:627-33

O'Donnell CA. Variation in GP referral rates: what can we learn from the literature? *Fam Pract,* 2000; 17:462-71

OECD. *Health at a glance: OECD indicators - 2005 edition.* Paris: OECD, 2005

Öner FC. *Thoracolumbar spine fractures: diagnostic and prognostic parameters.* [dissertation] Utrecht: Brouwer Uithof, 1999

Onion CWR, Wally T. Clinical guidelines: development, implementation, and effectiveness. *Postgraduate Med J*, 1995; 71:3-9

Pasley BH, Lagoe RJ, Marshall NO. Excess acute care bed capacity and its causes: the experience of New York State. *Health Serv Res,* 1995; 30(1):115-31

Paul-Shaheen P, Clark JD, Williams D. Small area analysis: a review and analysis of North American literature. *J Health Polit Policy Law,* 1987; 12:741-809

Payne TH. Computer decision support systems. *Chest,* 2000, 118(2 Suppl):47S-52S

Pescosolido BA. Beyond rational choice: the social dynamics of how people seek help. *Am J Sociol*, 1992; 97(4):1096-138

Peterson S, Eriksson M, Tibblin G. Practice variation in Swedish primary care. *Scand J Prim Health Care,* 1997; 15:68-75

Powell A, Davies H, Thomson R. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. In: Grol R, Baker R, Moss F (eds.) *Quality improvement research: understanding the science of change in health care*. London: BMJ Publishing Group, 2004; 29-50

Ramnarayan P, Roberts GC, Coren M, Nanduri V, Tomlinson A, Taylor PM, Wyatt JC, Britto JF. Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: a quasi-experimental study. *BMC Med Inform Decis Mak,* 2006; 6(1):22

Read JL, Stern RS, Thibodeau LA, Greer DE Jr, Klapholz H. Variation in antenatal testing over time and between clinic settings. *JAMA,* 1983; 249:1605-9

Ritzer G. *The McDonaldization of society*. Thousand Oaks, California: Pine Forge Press, 2000

Ritzer G, Walczak D. Rationalization and the deprofessionalization of physicians. *Social Forces,* 1988; 67:1-22

Robinson R. Managed care: US evidence and lessons for Europe. *Eurohealth,* 1998; 4(2):33-4

Ros CC, Groenewegen PP, Delnoij DMJ. All rights reserved, or can we just copy?: cost sharing arrangements and characteristics of health care systems. *Health Policy*, 2000; 52:1-13

Saltman RB, Figueras J. Analyzing the evidence on European health care reforms. *Health Affairs,* 1998; 17(2):85-108

Schellevis FG, Westert GP, De Bakker DH, Groenewegen PP. *Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk: vraagstellingen en methoden.* Utrecht: NIVEL, 2004

Schers H, Braspenning J, Drijver R, Wensing M, Grol R. Low back pain in general practice: reported management and reasons for not adhering to the guidelines in The Netherlands. *Br J Gen Pract,* 2000; 50(457):640-4

Schut FT. *Health Maintenance Organizations.* Lochem-Gent: De Tijdstroom, 1986

Scott WR. *Institutions and Organizations*. Thousands Oaks: Sage, 2001

Scott WR, Ruef M, Mendel PJ, Caronna CA. *Institutional change and healthcare organizations: from professional dominance to managed care*. London: The University of Chicago Press, 2000

Sessa VI, Taylor JJ. *Executive selection*. San Francisco: Jossey-Bass Inc., 2000

Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. *J Am Med Inform Assoc,* 1996, 399-409

Shi L. Patient and hospital characteristics associated with average length of stay. *Health Care Manage Rev,* 1996; 21(2):46-61

Shiffman RN, Liaw Y, Brandt CA, Corb GJ. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. *J Am Med Inform Assoc,* 1999; 6(2): 104-14

Shortell SM. *Effective hospital-physician relationships.* Ann Arbor: Health administration press perspectives, 1991

Shortell SM, Gillies RR, Anderson DA, Erickson KM, Mitchell JB. *Remaking health care in America.* San Francisco: Jossey-Bass, 2000

Simon CJ, Dranove D, White WD. The effect of managed care on the incomes of primary care and specialty physicians. *Health Serv Res*, 1998; 33(3):549

Snijders TAB, Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modeling.* London: Sage Publications, 1999

Stano M. Evaluating the policy role of the small area variations and physician practice style hypotheses. *Health Policy,* 1993; 24:9-17

Starr P. *The social transformation of medicine*. New York: Basic Books, 1982

Tacken MAJB. *Quality of preventive performance in general practice: the use of routinely collected data.* [dissertation] Wageningen: Ponsen and Looijen BV, 2005

Taroni F, Stiassi R, Traversa G, Raschetti R, Menniti-Ippolito F, Maggini M, Spila-Alegiani S. The nature content and interpractice variation of general practice: a regional study in Italy. *Eur J Epidemiol,* 1990; 6:313-8

Tasche M, Oosterberg E, Kolnaar B, Rosmalen R. Inventarisatie van lacunes in huisartsgeneeskundige kennis. *Huisarts en Wetenschap,* 2001; 44(3):91-4

Thom DH. Training physicians to increase patient trust. *J Eval in Clin Pract,* 2000; 6(3):245-53

Tielens VCLMG. *Consulten van huisartsen*. [dissertation] Nijmegen: Katholieke universiteit Nijmegen, 1987

Tiemeier H, De Vries WJ, Van het Loo M, Kahan JP, Klazinga N, Grol R, Rigter H. Guideline adherence rates and interprofessional variation in a vignette study of depression. *Qual Safety Health Care*, 2002; 11(3):214-8

Timmermans S, Berg M. *The gold standard: an exploration of evidence-based medicine and standardization in health care*. Philadelphia, PA: Temple University Press, 2003

Tunis SR, Hayward RS, Wilson MC, Rubin HR, Bass EB, Johnston M, Steinberg EP. 1994. Internists' attitudes about clinical practice guidelines. *Annals Int Med,* 1994; 120(11):956-63

Ultee W, Arts W, Flap H, *Sociologie - vragen, uitspraken, bevindingen*. Wolters-Noordhoff, Groningen 1992 tweede herziene druk, 1996

US Department of Health and Human Services, Centers for Disease Control. *Health, United States, 2000.* Washington, D.C.: U.S. Government Printing Office, 2000

Van den Berg MJ, Kolthof ED , De Bakker DH, Van der Zee J. *Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk: de werkbelasting van huisartsen.* Utrecht: NIVEL, 2004

Van den Brink-Muinen A, De Bakker DH, Bensing JM. Consultations for women's health problems: factors influencing women's choice of sex of general practitioner. *Br J Gen Pract,* 1994; 382:205-10

Van der Velden J. *General Practice at work*. [dissertation] Utrecht: NIVEL, 1999

Van Kemenade YW. *Health care in Europe 2007: the finance and reimbursement systems of 11 European countries.* Maarssen : Elsevier Gezondheidszorg, 2007

Van Wijk MAM. *Bloodlink: computer based blood test ordering*. [dissertation] Rotterdam: Erasmus University, 2000

Verhaak PFM. Analysis of referrals of mental health problems by general practitioners. *Br J Gen Pract,* 1993; 43:203-8

Verstappen WHJM. *Towards optimal test ordering in primary care.* [dissertation] Maastricht: Universiteit Maastricht; 2004

Verstappen WHJM, Van der Weijden T, Sijbrandij J, Smeele I, Hermsen J, Grimshaw J, Grol RPTM. Effect of a practice-based strategy on test ordering performance of primary care physicians. *J Am Med Assoc,* 2003; 289:2407-12

Vladeck BC. Everything new is old again. *Health Affairs,* 2004; October Supplement. Var :108-27

Wazana A. Physicians and the pharmaceutical industry: is a gift ever just a gift? *JAMA,* 2000; 283(3):373-80

Weide MG, Westert GP, Schouten JSAG, Groenewegen PP. *Praktijkvariatie en gepaste zorg.* Utrecht: NIVEL/RIVM, 1992

Weiner JP, De Lissovoy G. Razing a Tower of Babel: a taxonomy for managed care and health insurance plans. *J Health Polit Policy Law*, 1993; 18(1):75-103

Weinstein JN, Bronner KK, Shawver Morgan T, Wennberg JE. Trends and geographic variations in major surgery for degenerative diseases of hip, knee and spine. *Health Affairs,* 2004; October Supplement. Var:81-9

Wennberg JE. On the need for outcomes research and the prospects for the evaluative clinical sciences. In: Anderson TF and Mooney G (eds). *The challenges of medical practice variations.* London: The Macmillan Press, 1990, 158-73

Wennberg JE. Future directions for small area variations. *Med Care,* 1993; (31)5:YS75-YS80

Wennberg JE. Understanding geographic variations in health care delivery. *New Engl J Med,* 1999; 340:52-3

Wennberg JE. Practice variations and health care reform: connecting the dots. *Health Affairs,* 2004; October Suppl, Var:140-4

Wennberg J, Gittelsohn A. Small area variations in health care delivery: a population-based health information system can guide planning and regulatory decision-making. *Science,* 1973; 14(182):1102-8

Wennberg JE, Gittelsohn A. Health care delivery in Maine I: patterns of use of common surgical procedures. *J Maine Med Assoc,* 1975; (66)5:123-49

Wennberg J, Gittelsohn A. Variations in medical care among small areas. *Sci Am,* 1982; 246(4):120-34

Westert GP, Groenewegen PP. Medical practice variations: changing the theoretical approach. *Scand J Public Health,* 1999; 27:173-80

Westert GP, Schellevis FG, De Bakker DH, Groenewegen PP, Bensing JM, Van der Zee J. Monitoring health inequalities through General Practice: the Second Dutch National Survey of General Practice. *Eur J Public Health,* 2005; 15(1):59-65

Westert GP. *Variation in use of hospital care*. [dissertation] Assen: Van Gorcum, 1992

Westert GP. Artsen en de 'couleur locale': over het verklaren van lokale verschillen in medisch handelen. *TSG, 1996;* 74:52-6

Westert GP, Nieboer AP, Groenewegen PP. Variation in duration in hospital stay between hospitals and between doctors within hospitals. *Soc Science Med,* 1993; 37(6):833-9

Westert GP, Lagoe RJ. The evaluation of stays for total hip replacement. *Qual Manage Health Care,* 1995; 3(3):62-71

Wierenga B, Jong SJT, Mantel AF. Het beslissingsproces van de huisarts bij de keuze van een geneesmiddel. *Ned Tijdschr Geneeskunde,* 1989; 133(3):115-22

Wilensky HL, Ladinsky J. From religious community to occupational group: structural assimilation among professors, lawyers, and engineers. *Am Sociol Rev*, 1967; 32(4): 541-61

Wolters I, Van den Hoogen H, De Bakker D. *Evaluatie invoering elektronisch voorschrijf systeem: eindrapport.* Utrecht: NIVEL, 2003

Wolters I, Van den Hoogen H, De Bakker D: *Evaluatie invoering Elektronisch Voorschrijf Systeem: monitoringfase: de situatie in 2001.*Utrecht: NIVEL, 2002

Wonnacott TH, Wonnacott RJ. *Introductory Statistics*. New York: WILEY 5th edition, 1990

Wright JG, Hawker GA, Bombardier C, Croxford R, Dittus RS, Freund DA, Coyte PC. Physician enthusiasm as an explanation for area variation in the utilization of knee replacement surgery. *Med Care,* 1999; 37(9):946-56

Zaat JO, Van Eijk JT, Bonte HA. Laboratory test form design influences test ordering by general practitioners in The Netherlands. *Med Care,* 1992; 30(3):189-98

Zaat JOM. *De macht der gewoonte: over de huisarts en zijn laboratoriumonderzoek.* [dissertation] Amsterdam: University of Amsterdam, 1991

Zeggelink E. *Strangers into friends*. [dissertation] Amsterdam: Thesis Publishers, 1993

Zhang N, Kohn L, McGarrah R, Anderson G. The effect of managed care on hospital staffing and technological diffusion. *Health Policy,* 1999; 48:189-205

Zwanziger J, Melnick GA, Bamezai A. The effects of selective contracting on hospital costs and revenues. *Health Serv Res,* 2000; 35(4):849-67

Curriculum Vitae

Judith de Jong was born on March 21, 1976 in Weststellingwerf, the Netherlands.
From 1995 to 2000, she studied Science and Policy at Utrecht University.
After graduating she started working at NIVEL (Netherlands Institute for Health Services Research).
Her research topics included medical practice variations, comparative health systems research, the organization of medical specialists, and health system reform.
Currently she is involved in evaluating the Dutch health insurance system and in health consumer research.

Dankwoord

Het schrijven van dit proefschrift was een leuke uitdaging. Ik heb hierbij op allerlei manieren steun en hulp gekregen van veel mensen. Allereerst wil ik mijn promotoren, Peter Groenewegen en Gert Westert, bedanken.
Peter, je bleek steeds in staat om hoofdstukken, waarvan ik vond dat ze af waren, weer met een frisse blik te bekijken. Je kritische houding en enthousiasme hebben me gestimuleerd en uitgedaagd bij het schrijven van dit proefschrift. Het is bijzonder prettig om met je te werken.
Gert, dankzij jou had ik mooi werk om op voort te bouwen. Je hebt een heldere kijk op mijn werk gehad, stond altijd klaar en was erg behulpzaam. Zo heb je ervoor gezorgd dat ik naar de Verenigde Staten kon om een paar weken bij Ron te werken.
Ron, without your effort I would not have learned so much in only a few weeks in the USA. You arranged an interesting and complete program and gave me the opportunity to get a glance at the health care system in the USA. Eating nacho's at the end of the day was always good to digest the enormous amount of information I got during the day.

Ik heb dit proefschrift bij het NIVEL geschreven. Een prettige plek waar ik zowel op steun als kritisch commentaar kan rekenen. Een aantal collega's wil ik bij naam noemen. Peter Spreeuwenberg, veel van wat ik nu kan met multilevel analyse heb ik van jou geleerd. Het was plezierig om er samen over te kunnen discussiëren en oplossingen te bedenken.
Christel, je hebt van mijn tekst een mooi boek weten te maken. Zonder jouw inspanningen en nauwkeurigheid had het er niet zo mooi uitgezien.

Mieke en Liset, jullie hebben nuttig commentaar op het manuscript gegeven, bedankt daarvoor. Het is weer tijd voor een borrel om het leven te bespreken!
Majda en Ellen, het was een gezellige tijd. Ik vond het leuk om veel dingen samen met jullie te organiseren. Majda, je functie als klankbord waardeer ik erg. Daarnaast zullen weinig mensen mijn boek zo goed lezen als jij hebt gedaan. Merci!
Klim-, fiets-, borrel- en andere vrienden, jullie hebben het jaren volgehouden om te vragen hoe het met mijn proefschrift ging. Dit is het geworden. Een mooie gelegenheid voor een feest!

Alex, het is een gevaar om bij een dankwoord alleen aan de mensen te denken die in de laatste fase een bijdrage hebben geleverd. Je hebt in voorgaande jaren een luisterend oor geboden en vaak voor een plek gezorgd waar ik ongestoord kon werken. Dat, en dat niet alleen, is waardevol.

Mijn familie en in het bijzonder mijn ouders, jullie wil ik bedanken voor de liefdevolle aandacht en steun. Soms blijkt Utrecht dichterbij dan verwacht!

Lieve Johan, eigenlijk wil ik je gewoon bedanken voor wie je bent. Je liefde en steun zijn belangrijk voor me bij veel meer dan alleen het afronden van dit proefschrift. Het is fijn dat je naast me staat.

Utrecht, mei 2008
Judith de Jong