FINANZWISSENSCHAFTLICHE SCHRIFTEN

Robert Nuscheler

On Competition and Regulation in Health Care Systems

PETER LANG

FINANZWISSENSCHAFTLICHE SCHRIFTEN

Robert Nuscheler

On Competition and Regulation in Health Care Systems

Health care systems are under reform in many countries. This typically involves a shift towards more competition. But still, markets are highly regulated. This study analyzes competition and regulatory measures in four important fields using the modern tools of microeconomic theory and microeconometrics. The book demonstrates how price regulation interacts with the quality of care and shows that non-price competition amongst providers affects the social desirability of a gatekeeping system. Using data from the German Socio-Economic Panel, the conventional wisdom of risk selection by German sickness funds is challenged.

Robert Nuscheler, born in Berlin in 1969, studied from 1991 to 1998 Economics and Mathematics at the Freie Universität Berlin and the University of Limerick. From 1998 to 2001 he was a research fellow at the Institute of Public Finance and Social Policy (FU Berlin). In 2001 he joined the Social Science Research Center Berlin (WZB), where he works as a research fellow in the unit *Market Processes and Governance.*

Retrodigitization in 2018

www.peterlang.com

On Competition and Regulation in Health Care Systems

FINANZWISSENSCHAFTLICHE SCHRIFTEN

Herausgegeben von den Professoren
Konrad, Krause-Junk, Littmann, Oberhauser, Pohmer, Schmidt

Band 112

PETER LANG
Frankfurt am Main · Berlin · Bern · Bruxelles · New York · Oxford · Wien

Robert Nuscheler

On Competition and Regulation in Health Care Systems

PETER LANG
Europäischer Verlag der Wissenschaften

Bibliographic Information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at <http://dnb.ddb.de>.

Zugl.: Berlin, Freie Univ., Diss., 2003

This book is available Open Access thanks to the kind support of ZBW – Leibniz-Informationszentrum Wirtschaft.

D 188
ISSN 0170-8252
ISBN 3-631-53527-9
US-ISBN 0-8204-7682-X
ISBN 978-3-631-75167-1 (eBook)

Printed in Germany 1 2 3 4 6 7

www.peterlang.de

Preface

This study on competition and regulation in health care systems is, except for minor adjustments, identical to my doctoral thesis that was submitted to the Department of Economics at the Free University of Berlin in November 2003. I would like to thank everybody who helped me in writing this dissertation. First of all, I would like to thank Kai A. Konrad (my supervisor and the chairman of my doctoral committee) for his encouragement, inspiration and permanent support. His suggestions and his intuition greatly improved this dissertation. Additionally, he gave me the opportunity to present and discuss my work at several international conferences and strongly supported the building of a tight network with researchers from the Economics Department of the University of Bergen and its Programme for Health Economics. I am also greatly indebted to my second supervisor, Helmut Bester, for many valuable and insightful discussions and suggestions. I also benefitted from presentations in the Microeconomic Colloquium at the Free University of Berlin—many thanks to the participants. Numerous discussions with friends, colleagues and co-authors were also extremely helpful. In particular, I thank Anette Boom, Kurt Brekke, Volker Dahms, Tomaso Duso, Astrid Jung, Sebastian Kessing, Thomas Knaus, Daniel Krähmer, Johannes Münster, Julio Robledo, Lars-Hendrik Röller, Odd Rune Straume, and Roland Strausz. Last, but not least, I thank my wife, Tina, for her constant encouragement and support. I dedicate this book to her, and to Luis and Henri for giving me a smile even when I returned home late.

Berlin, November 2004.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Outline

In this thesis four different topics in the field of health economics are addressed. In Chapter 2 the behavior of a vaccine monopolist is analyzed. At the heart of the analysis is the monopolist's incentive to reduce supply in order to increase the willingness to pay for the vaccine through the increased risk of infection. Chapter 3 deals with physician competition when prices are regulated. Competition will then be in quality and location or specialization. The impact of the regulators commitment power on the market outcome is analyzed. It is supposed that giving general practitioners a gatekeeper role in the health care system increases the efficiency of care. Moreover, it is usually argued that this contributes to cost containment. This conventional wisdom is challenged in Chapter 4, where the competitive effects of gatekeeping are analyzed. Risk selection in the German Public Health Insurance System is analyzed in Chapter 5. After free choice of sickness funds was made available in 1996 a significant distortion of competition occurred due to risk separation. We test whether some strategic aspects are operating in the background (adverse selection or cream skimming). Finally, Chapter 6 offers a summary in German.

To better motivate these papers and to put them into a broader context this introduction is provided. The characteristics, or sometimes the peculiarities, of the health care market are briefly described in Section 1.2. These characteristics often result in market failures which provide a motivation for public policy interventions. Guided by the content of the above mentioned chapters, in the following Sections 1.3 to 1.6 some of these characteristics are described in detail. The contribution of the thesis to the economics literature and to the public policy debate is presented at the end of each of these sections.

At the end of the introduction the most important trade offs identified throughout are summarized. We will conclude that it will be, in general, impossible for a regulator to account for all these trade offs such that efficiency will rarely be achieved. The amount of trade offs is not only challenging for a regulator but also for researchers. The number of trade offs that can simultaneously be analyzed is limited by algebraic tractability. This motivates the partial approaches adopted throughout the thesis. The following table describes the organization of the thesis.

Topic	Introduction	Contribution	Chapter
Vaccines	1.3	1.3.4	2
Price regulation	1.4	1.4.6	3
Gatekeeping	1.5	1.5.4	4
Risk selection	1.6	1.6.6	5

Table 1.1: Organization of the thesis.

1.2 Characteristics of the health care market

1.2.1 Welfare economics and market failures

Health care systems are typically characterized by a (large) number of regulatory rules, including the organization of health care. This involves, among other things, the financing and the delivery of care. The international variety of systems is remarkable. The United Kingdom and Italy, for example, opted for a National Health Service coming along with little competition. Medical care is primarily provided by the state or state-owned companies. Financing is by general taxation. Germany and France are among those that have a social insurance. In Germany providers are private, state-owned, or operated by other institutions like the church. There is thus some competition on the providers' side. The German market for social health insurance is competitive. However, competition is subject to numerous rules so that this form of competition may well be labelled as 'regulated competition'. Finally, the United States is the example of a highly privatized health care market. The vast majority of providers and health insurers are private companies. Although competition is perhaps most pronounced in the United States there is nevertheless a large number of regulatory rules in 'managed care'.

This brief overview demonstrates that there obviously is a fundamental trade off between competition and regulation and that there is no unique solution to it. In general, the introduction of some 'rules' by a regulator will have an impact on competition in the respective market. In most cases, competition will be dampened by such rules but it may well be that there are some particular measures that foster competition (see, e.g., Chapter 4).

Economists should ask, and they do, what the reasons for regulation are or why the competitive market does not achieve an efficient outcome. An obvious starting point is the *first theorem of welfare economics*. Following this fundamental theorem, the equilibrium of an economy is efficient if there are markets for all relevant commodities and if all these markets are competitive.[1]

Thus, to argue for public policy interventions at least one of the prerequisites of the theorem must be violated. In the following we briefly discuss some violations and provide illustrative examples.

1.2.2 External effects

When there are external effects a competitive market will, in general, not arrive at an efficient outcome since there is usually no market for externalities. The prime example for positive externalities in the health care market are vaccinations. Once an individual is immunized through vaccination, he or she can no longer communicate the disease and the risk of infection for all other individuals is reduced. However, if an individual is about to decide whether or not to be immune he or she weighs the individual costs that may, for example, arise from the price of the vaccine or the (potential) side effects against the individual benefit of immunization. The benefit to other individuals is not internalized and this leads to too low immunization rates. In the theory of vaccination, which is described in detail in section 1.3, the most prominent measures to correct for this market failure are subsidies and mandatory vaccination programs.

Although Breyer et al. (2003, p. 169) argue that positive externalities are more relevant for the health care market, there are examples for negative externalities. Smoking not only damages the health of the smoker but also the health of individuals in the direct neighborhood, i.e., the own family and

[1] For a more formal definition see Gravelle and Rees (1992, p. 490).

colleagues. Moreover, there are some costs associated with smoking related fires (Santerre and Neun (1996, p. 246)). As smokers do not internalize these negative external effects, they smoke too much or, to put it differently, there are too many smokers in the society. As a response, there are usually considerable taxes on cigarettes and the like. Especially in the United States there are many rules that ban smoking in the work place and in public buildings, e.g., administrative offices, restaurants, and bars.

Besides these 'physical' external effects there may also be 'psychological' external effects, e.g. altruism (see Breyer at al. (2003, pp. 170-171)). Consider an individual with no access to medical care that is unresponsively in distress. Altruistic people would be willing to help. However, as help is a public good, i.e. an individual transfer benefits the entire (altruistic) population, and as the associated positive externality is typically not internalized, the willingness to transfer will be too low. Since a donor is usually interested in the consumption (of medical care) of the recipient and not in his actual utility this problem is likely to be more severe if transfers cannot be given in kind but in cash. Thus, to increase the willingness for redistribution the social planner may have to distort relative prices and give the transfers in kind.

1.2.3 Market transparency and the quality of care

Another prerequisite of the first theorem of welfare economics is market transparency. In particular, patients or customers must be perfectly informed about product quality. Before we address the observability of quality in health care markets note that there is no consensus about what quality actually is in such markets. Following Donabedian (1980, pp. 79-85) there are (at least) three dimensions: the structure, the process, and the outcome of care. Actual quality may be some arbitrarily weighted index of these dimensions.

Consider that quality is well defined and measurable.[2] But quality is still hardly observable for the patient.[3] As production and consumption of health care often occur simultaneously (uno actu principle) the patient has, prior to consumption, no own information about the product's quality. However patients may observe quality after consumption, and medical care then is an experience good (Nelson (1970)). However, there are cases where even this information cannot be obtained. Consider the case of appendicitis

[2]This is one of our central assumptions in Chapters 3 and 4.

[3]In Chapter 3 we will nevertheless assume that quality is observable for the patient. This assumption is relaxed in Chapter 4.

and recommended appendix surgery. Once the appendix is removed the patient can no longer verify whether the surgery was actually necessary. In such cases medical care is a credence good (Darby and Karny (1973)).[4]

Irrespective of whether medical services are credence or experience goods there are obvious moral hazard problems. Physicians are, for example, usually not only the providers of medical care but also act as experts who, at least to some extent, determine the amount care of consumed. This may result in physician induced demand. How the lack of market transparency may affect the market outcome, e.g., investment incentives and product characteristics, will be analyzed in some detail in Section 1.5. Regulatory measures to address these quality related problems are, for example, customer protection rules like (monitored) minimum quality standards or licences that are conditional on regular training. Moreover reimbursement rules may be designed to give providers the right incentives. In Section 1.4.3 we will argue that physician induced demand, for instance, may be mitigated by proper price regulation. However, as Section 1.4 will demonstrate this may be in conflict with other objectives.

1.2.4 Market transparency and health insurance

Illness occurs irregularly and its occurrence is unpredictable. Consequently the demand for medical care is also irregular and unpredictable (Arrow (1963, p. 948)). As individuals are typically risk averse, the nature of the demand for medical care creates a demand for health insurance.

Consider that there is a competitive health insurance market with no administrative costs. If insurers have perfect information, especially about the risk type of the insured, fair premiums would emerge and buyers will demand full coverage. In such a case, patients are not responsive to the prices of the medical care providers. This has two important impacts. First, there cannot be any price competition on the providers' side if the insured have free choice of provider. Additionally, prices are often regulated. In the absence of price as a strategic variable, providers will resort to other variables to increase market share. These could include, for example, the location of a medical practice or its specialization. Under the constraints of the previous subsection, there may also be considerable quality competition. Second, on the patients' side moral hazard problems arise. These can be either the ex ante or ex post type. The consumer's lack of cost consciousness (a) means that they refrain from undertaking efficient prevention[5] (*ex ante*

[4] For the role of information in the medical care market see also Pauly (1988).

[5] Investments in health like preventive activities are usually unobservable to insurers.

moral hazard) and (b) results in over-consumption, i.e. in inefficient use of medical services (*ex post moral hazard*). An optimal health insurance contract therefore would, in general, involve some deductibles or co-payments. Both moral hazard problems would be mitigated and, in the absence of price regulation, there would be some room for price competition.

Consider that there are two risk types in the society and that the individuals' information about risk type is private. Then health insurance contracts with full coverage at fair premiums would not be incentive compatible. The high risk types would imitate the low risk type in order to receive the low premium. Rothschild and Stiglitz (1976) have shown that there may be a separating equilibrium in the market where the low risks receive part coverage at fair premiums and the high risks full coverage at fair premiums. This phenomenon is known as *adverse selection* (see also Section 1.6). The asymmetry of information results in a welfare loss. This loss may be reduced by introducing a mandatory health insurance with part coverage and uniform premiums.

For equity reasons, insurers may be mandated to offer uniform premiums, i.e. premiums that are not conditional on the risk type (community rating), and a standard benefit package. This creates an incentive for insurers to select the low risks as they earn profits with them and incur losses with high risks. How these cream skimming incentives can be mitigated is at the heart of the analysis in Section 1.6.

1.2.5 Market power

The first theorem of welfare economics requires perfectly competitive markets. Once there is market power the theorem no longer applies. First, consider the providers of health care. There are several sources of market power. In the primary care market, for example, patients are often better informed about the quality of their general practitioner (GP) than about the quality of competing GPs, if we consider medical services to be experience goods. This gives the GP some (local) monopoly power lowering his incentives to provide high quality. There is thus a trade off between information and competition. The same argument applies to the secondary care market.[6]

Second, there may be market power in the health insurance market. In the most extreme case, there is a monopoly like in the United Kingdom and Italy where financing is, as already mentioned above, by general taxation. There are limited incentives to provide quality and to contain costs.

[6] In Chapters 3 and 4 we analyze monopolistic competition of health care providers.

Moreover, responsiveness to consumer preferences will be lacking. But, on the other hand, there are no cream skimming incentives and the potential welfare losses from adverse selection can be avoided. Again, there is a trade off between competition and regulation.

Third, to fuel technological progress it is, in general, necessary to allow the inventing firm to obtain monopoly profits in order to recover R&D expenditures. Without patent protection innovation incentives would be inefficiently low. Here the trade off is between competition and dynamic efficiency. How the technical progress is related to provider reimbursement and to the demand for health insurance was demonstrated by Weisbrod (1991) in his seminal paper on the "Health Care Quadrilemma".

1.3 Communicable diseases and vaccines

1.3.1 The vaccination externality and public policy

As was already mentioned above, the most important characteristic of the vaccine market is the externality that is associated with vaccination. If an individual decides to become vaccinated, there are two effects. First, the vaccinated individual is immune and can no longer catch the particular disease. Second, (with most vaccines) the individual can no longer communicate the disease. The source of the externality is, of course, the second effect. When an individual thinks about immunization he or she weighs the individual benefits of immunity against the individual costs. The individual costs could be, for example, time costs, a disutility from expected side effects from the vaccine, or simply the price of the vaccine. The individual does not take into account the positive external effect on others that can no longer be infected from him or her. This will, in general, result in too low immunization rates and call for public health intervention.

The two most important public policy measures are Pigouvian subsidies and mandatory vaccination programs. However, due to the prevalence elasticity of demand, independent of market structure, these measures have limited impact on actual demand. Consider the case of Pigouvian subsidies. In the first place demand increases as the price is reduced by the subsidy. But the increased demand reduces the prevalence of the disease and thereby the risk of infection. This, in turn, reduces the benefits of vaccination and therefore demand. The prevalence effect also weakens mandatory vaccination programs. The reduced prevalence of the disease lowers the willingness to become vaccinated of those outside the program. This is why disease eradication usually cannot be achieved at participation rates below one.

The economic literature on epidemiology, infectious diseases or vaccinations is surprisingly young. The first paper that addressed these issues was Brito et al. (1991). They set up a static model to address the appropriateness of compulsory vaccination and to derive the optimal public policy. A vaccine yielding perfect protection is considered. However, there is a disutility from vaccination, e.g., side effects. Given these costs it is not surprising that compulsory vaccination is, in general, not optimal. In fact, it is shown that the (competitive) laissez-faire outcome always dominates the compulsory (regulatory) outcome. Nevertheless, because of the vaccination externality, the laissez-faire immunization rate is too low. Subsidizing those who vaccinate and taxing the susceptible improves the allocation. If the marginal utility of income is constant, then the social optimum can be achieved.

The form and optimality of public health interventions crucially depends on the assumptions made about agent heterogeneity, market structure and dynamics. The most obvious critisism of the Brito et al. (1991) paper is its static nature. However, we believe that the externality can be studied in a static environment and that the results carry over into a dynamic context. Francis (1997) challenges this view. He shows that the vaccination externality disappears when a dynamic framework is considered. But this is a knife-edge result as he assumes identical individuals who also behave identically. The externality would not disappear if agent heterogeneity, as, for example, in Brito et al. (1991), had been considered. And indeed in the dynamic model of Geoffard and Philipson (1997) there still is an externality as the demand is prevalence elastic. At the focus of their paper is the difficulty private markets have in achieving full immunization, i.e., eradication of a disease. As already discussed above, they show that the impact of public programs aimed at increasing demand, for example Pigouvian subsidies and mandatory vaccination programs, is limited.[7]

To summarize, the (individual) willingness to pay for vaccination increases in the prevalence of the disease. Or, more generally, the willingness to pay increases in the risk of infection. Another important determinant of the willingness to pay is income. Empirical studies have revealed a positive income effect (England et al. (2001) and Philipson (1996)). Although these studies were at a national level (China and the United States, respectively), this relationship also holds in an international context. The developing countries are more likely to go without essential immunization and consequently most of the disease burdens of vaccine preventable diseases fall on those countries. Some examples will be presented next.

[7]For a literature overview see Philipson (2000).

1.3.2 Examples and empirical evidence

The World Health Organization (WHO) lists a number of vaccine preventable diseases, e.g., diphtheria, influenza, hepatitis A and B, measles, mumps, tetanus, poliomyelitis, rubella, and smallpox.[8]

Consider poliomyelitis and the polio eradication initiative. Its existence provides clear evidence that, first, private markets fail to eradicate the disease and, second, that it seems desirable to eradicate polio internationally. There are basically two reasons for this: the externality of communicable diseases and charity. That infection probabilities are internationally linked was recently demonstrated by the outbreak of the Severe Acute Respiratory Syndrome (SARS) that originated in China and spread around the world, and especially to North America. So it is in 'everybody's' interest to contribute financially to the eradication initiative. However, contributions are, like vaccination itself, a public good. Private provision therefore comes along with underprovision. And indeed, the funding gap of the initiative for 2003 to 2005 as of December 2002 was US$ 275 million (WHO (2003, p. 24)). Nevertheless, polio is close to eradication.

The only disease that has ever been eradicated is smallpox. At the 33rd World Health Assembly on May 8, 1980, smallpox was declared eradicated. The prevalence of the disease is thus zero and so, too, was, until recently, the willingness to pay for the associated vaccine. But there may be an exogenous risk of infection, i.e., a risk independent of the prevalence of the disease. For the case of smallpox, this exogenous risk is the likelihood of a terrorist attack with the polio virus or laboratory outbreaks (small stocks of the virus were allowed to be hold in laboratories for research purposes). A positive willingness to pay resulted, leading the United States to buy 209 million doses of the smallpox vaccine (USA Today, November 26, 2002). The threat posed by bio-terrorism was demonstrated by the anthrax attacks in 2001. In response, the U.S. government bought 100 million Cipro doses (USA Today, October 29, 2001).

Measles is a disease that is far from being eradicated. It is estimated that there are annually over 30 million cases (WHO (2001, p. iv)) resulting in about 770,000 deaths in 2000. More than 99 percent of the disease burden of measles falls on developing countries (Kremer (2002, p. 71)). This is mainly due to the low immunization rates in those countries.[9] In devel-

[8] See http://www.who.int/vaccines/en/vaccprevdis.shtml.

[9] "Failure to deliver at least one dose of measles vaccine to all infants remains the primary reason for high measles morbidity and mortality. Many measles deaths may be preventable by utilizing existing immunization services more efficiently. Poor management, logistical problems and missed opportunities for immunization are among the main

oped countries the vaccines are often free of charge, i.e., they are highly subsidized. Germany is one example where, however, vaccination is not mandatory. In the United States, where immunization typically is a prerequisite for school enrollment, it is mandatory (Brito et al. (1991, p. 70)). There is an initiative of the WHO and the United Nations Children's Fund (UNICEF) to reduce the mortality rates of measles (WHO (2001)).

1.3.3 Monopoly power

Currently monopoly power is a side issue in the vaccine market as the patents for most common infectious diseases have expired. However, biotechnological progress, together with the current patent policy, where living organisms can be protected, monopoly power will be of major concern in the near future. A vaccine monopolist basically has two incentives. First, he aims at keeping the disease alive, since once the disease is eradicated so is the monopolist. Second, to increase markups he reduces supply in order to increase the risk of infection and thereby the willingness to pay for vaccination (Geoffard and Philipson (1997, p. 222)).

In Chapter 2 we provide evidence that there already is monopoly power and that its importance has increased during the last few years. Here we will provide two illustrative examples for cases where currently no vaccine is available, namely, anthrax and the human immunodeficiency virus (HIV) that causes the acquired immune deficiency syndrome (AIDS).

The terrorist attacks with anthrax in the United States in 2001 not only demonstrated that there is a non-prevalence related risk of infection but also that there is monopoly power and that monopolists fight for their profits. The antibiotic Cipro is the most appropriate anti-anthrax drug and Bayer holds the patent. The original price in 2001 was US$ 1.77 a pill. After some threats of re-engineering, especially from Brazil and Canada, Bayer finally agreed to sell 100 million pills at US$ 0.95 each (The Associated Press, October 24, 2001). Thus there is (or at least there was) a considerable markup. Interestingly, there was a lawsuit in 1997. And "[...] Bayer persuaded generic drug maker Barr Laboratories to drop a legal challenge to Bayer's patent on Cipro by agreeing to pay Barr about US$ 28 million a year until the patent expires in 2003" (USA Today, October 29, 2001).

There was a similar but more pronounced battle about AIDS drugs in 2001. South Africa, with more than 20 percent infected adults, faces the highest AIDS burden worldwide. Drug treatment costs for AIDS in the developed world are about US$ 15,000 per year and person. As the gross

reasons for the underutilization of services [...]" (WHO (2001, p. 4)).

domestic product per capita in South Africa is only US$ 6,800 it is not surprising that some pressure to cut prices arose.[10] At around US$ 1,000 the price offered in South Africa by the patent holders of the AIDS drugs was already much lower than in the developed countries. However, the Indian generic drug maker, Cipla, announced it would break patent protection and make an offer of US$ 600. This threat resulted in a drop in prices (New York Times, March 8, 2001).

The UN Accelerating Access Initiative supports differential pricing for AIDS drugs. In this context, Roche was more or less forced to increase the discount on its AIDS drugs for developing countries to roughly 90 percent of the Swiss price (see Médecins Sans Frontières (2003)). This example demonstrates that, due to the enormous political pressure and the threat of re-engineering, firms cannot expect to earn good profits for AIDS drugs in the developing world. This will also be true for a future vaccine which is expected to be available in 10 to 15 years (Desmond and Greener (2003)).[11] Differential pricing improves access to vaccination and is therefore, in a static context, socially desirable. But the improved access reduces the prevalence of the disease and with it the willingness pay for vaccination in the developed world and thus monopoly profits. This may be a reason why R&D efforts to make an AIDS vaccine are relatively moderate.[12]

1.3.4 Contribution of the thesis

Chapter 2 contributes to Brito et al. (1991) and Geoffard and Philipson (1997) by considering two important issues. First, we consider agent heterogeneity with respect to income. By making individual willingness to pay increasing in income we appropriately model this empirical fact.[13] Usually agent heterogeneity is introduced, as in Brito et al. (1991), through varying disutility of vaccination if at all. There may be indeed some heterogeneity in this respect, but it is difficult to observe. As there is clear evidence that income has a positive effect on the willingness to pay for vaccination (or on the probability of being vaccinated) it seems a natural step to incorporate agent heterogeneity through income into the theoretical analysis. Moreover, as public policy measures usually affect income, a model taking income effects into account will approximate the consequences of such

[10] The numbers were taken from Reekie (2000).

[11] Desmond and Greener (2003) develop a strategy for how to use a potential HIV vaccine and whom to vaccinate.

[12] Philipson and Mechoulan (2003) analyze R&D incentives in the presence of consumption externalities.

[13] For empirical studies see England et al. (2001) and Philipson (1996).

measures better.

Second, we concentrate on monopoly power on the supply side. As already mentioned above, a monopolist has an incentive to, first, keep the disease alive and, second, to increase the willingness to pay by cutting supply. In their dynamic model, Geoffard and Philipson (1997) address the first incentive, but, as a consequence of their steady state formulas, do not consider the second incentive. Using a static model, we provide the missing part of the analysis. Additionally, our reduced form approach enables us to come up with comparative static results. The amount of vaccines offered is, of course, lower than without the externality. Supply falls further as the strength of the external effect increases. The willingness to pay becomes more responsive to the prevalence of the disease and fosters the monopolist's incentive to cut supply.

In standard monopoly a perfect price discriminating monopolist is socially efficient. We show that this is not necessarily true with this type of negative externalities. Selling vaccination to the marginal susceptible consumer not only reduces the price at that margin, but also all other prices the monopolist can demand. The reduced prevalence of the disease reduces the willingness to pay of all other consumers. If the external effect is sufficiently strong, there will not be an efficient allocation in the market, which for our model means that full immunization is not achieved. This result is new. It demonstrates that differential pricing may not be sufficient to guarantee access to vaccines for the developing world. Thus, although the UN accelerating access initiative supports differential pricing, it may not be sufficient for disease eradication. Additional measures, for example, organized vaccination programs that reduce distributional costs, may be necessary.

We also analyze the two most standard public health interventions, namely Pigouvian subsidies and mandatory vaccination programs. As already mentioned above, both these policies are limited in their impact since the prevalence elasticity of demand counters the respective positive effects of these measures. For the subsidies things may be even worse. Consider that the subsidy has to be financed by taxation, then the positive price effect of the subsidy is opposed by two negative effects, the prevalence effect and the income effect. If the latter is sufficiently large, the market brings about a lower immunization rate than without the subsidy. However, in such cases—although counter intuitive—taxes would be the optimal response. Nevertheless, this result strengthens Philipson's (2000) argument which states that "Pigouvian subsidies traditionally seen as resolving the under-provision problem of vaccines can be short-run, or out of steady state, arguments" (p. 1777), since these may fail even in static settings. If income is observable to the social planner, then mandatory vaccination programs can be much more effective than usual. With the poor covered in such a program and the rich served by the monopolist, full immunization can be achieved at participation rates that are strictly below one. This result is also new and it offers an efficiency based rationale for the vaccination programs usually supported by the United Nations, the Worldbank, or the World Health Organization.

1.4 Provider payment and incentives

Following the seminal paper of Arrow (1963), uncertainty is one of the major characteristics of the market for medical care. For individuals, illness is not a predictable occurance, it is a random deviation from the natural cause of events. With risk averse consumers this uncertainty constitutes a demand for health insurance.

Consider the case of full coverage, when the patient is not responsive to prices at all. He or she simply chooses the 'best' provider using all information available. The insurer is, as the third party payer, obliged to pay the provider's bill. In such a framework there is not much room for price competition.[14] Let us assume that health insurers are perfect agents for their members.[15] Moreover it seems reasonable to assume that insurers can offer contracts to providers as the market power of insurers compared to that of providers is usually much greater. There are simply far fewer insurers than providers. Then, as perfect agents for their members, insurers will offer optimal contracts. Thereby providers should be induced to make treatment decisions that are efficient from the patients' perspective.

Some price competition could be easily introduced by making consumers responsive to prices, e.g., by demanding proportional co-payments. As argued above (see Section 1.2), the optimal health insurance contract will, in general, involve a deductible in order to mitigate moral hazard problems. Third party payments will nevertheless remain significant and so will the insurers' objective of giving the right incentives.

[14] Health insurers may introduce some aspects of price competition by selective contracting. However, selective contracting may be forbidden as it can be used as a cream skimming device (see Section 1.6 and Chapter 5).

[15] With a competitive health insurance market, firms are induced to be responsive to consumer preferences.

1.4.1 A general payment formula

As provider incentives crucially depend on the reimbursement scheme it is convenient for the analysis to consider the following general payment function,

$$P = F + pN + \beta T + \gamma C.$$

The overall amount the provider receives is denoted by P. F is simply a transfer. N denotes the number of treated patients and, thus, p is a flat rate per case. For a single treatment the amount β is paid. This fee-for-service is paid for the total number of treatments performed T. Finally, the provider may be reimbursed for the share $\gamma \in [0, 1]$ of total costs C. Of course, both the flat rate and the fee-for-service imply some cost sharing. But as incentives differ substantially we treat these cases separately.[16]

1.4.2 Cost sharing

The question of whether providers should bear at least some of their costs received most attention in the literature on optimal provider payment. In the following two arguments are presented where optimal reimbursement involves cost sharing between providers and insurers, namely, cost reduction or cost containment objectives and selection incentives.

1.4.2.1 Cost containment

For this section we consider that the number of patients to be treated by a single provider is fixed, $N = \bar{N}$. Then a flat rate per case is simply a transfer. This is why we can set $p = 0$ without loss of generality. As the focus is on cost reduction incentives, we also set $\beta = 0$. Thus the provider receives $P = F + \gamma C$.[17]

The provider has strong incentives to reduce treatment costs when $\gamma = 0$. The provider is the residual claimant and thus cost reduction incentives are high powered. There are low powered incentives to contain costs when the health insurer bears the entire costs of treatment, $\gamma = 1$ (Laffont and Tirole (1993, p. 6)). The optimal reimbursement contract now depends on the specific environment.

[16] Although Section 1.4.4 is most relevant for the thesis, we provide Sections 1.4.2 and 1.4.3 in order to better motivate price regulation in the health care market. Even if there are no gains of price regulation in the models considered in Section 1.4.4, the price may nevertheless be regulated for one of the reasons mentioned in Sections 1.4.2 and 1.4.3.

[17] The arguments presented here can be found in Breyer et al. (2003, pp. 353-368) and, in a more general context, in Laffont and Tirole (1993, Chapter 1).

Consider a risk neutral provider and a risk neutral insurer. By exerting costly effort the provider can reduce costs. Like in Breyer et al. (2003) let us assume that actual costs are random. This ensures that the insurer cannot infer the effort level from observable cost figures. The cost function and the distribution of noise are common knowledge. At the first-best effort level, the marginal disutility of effort equals marginal cost savings. The first-best could easily be implemented if effort were contractible. However, this is likely to be infeasible. The first-best may nevertheless be implemented. Optimal reimbursement then involves $\gamma = 0$ and a transfer F that guarantees participation of the provider. Thus, given risk neutrality, making the provider the residual claimant induces efficiency.

This result relies on the assumption of risk neutrality and it no longer holds when providers are risk averse. If the provider were responsible for all costs the insurer would have to pay a large transfer in order to compensate the provider for bearing the entire risk, i.e., the insurer would have to pay a large risk premium to guarantee participation. In general, it pays for the insurer to distort cost reduction incentives and bear a fraction of actual costs. The inefficiency in effort is outweighed by the reduced risk premium. The optimal (second-best) reimbursement system with risk aversion thus involves cost sharing, i.e., $\gamma \in (0, 1)$ (Breyer et al. (2003, pp. 359-362)).

A similar result obtains when there is asymmetric information about costs. Consider that there are firms of different efficiency in the market. Efficiency is private information of the firm. The insurer only knows the distribution of the efficiency parameter.[18] For the health care market, efficiency may be reinterpreted as the 'case-mix' at a provider (Breyer et al. (2003, p. 365)). If a provider has to treat mostly high severity patients then average treatment is more costly than if mostly low severity patients were to be treated. It seems reasonable to assume that providers are better informed about the case-mix than the insurer. If there is no cost sharing, i.e. providers only receive a capitation payment F and $\gamma = 0$, the first-best effort level can be implemented by giving the most inefficient type his reservation utility.[19] As efficient providers can always mimic the most inefficient one, they receive a (large) informational rent. For the insurer, it pays to distort cost reduction incentives for the inefficient types by introducing some cost sharing. The optimal trade off between inefficiency in cost reduction effort and reduced informational rents yields the optimal contract.

[18]In such a situation one has to deal with an adverse selection problem (efficiency types) and a moral hazard problem (cost reduction effort).

[19]It is assumed that it pays to contract with all efficiency types.

1.4.2.2 Selection incentives

Consider that the patients now differ in treatment costs as well as in the benefits of treatment. From a social planner's perspective, a particular patient should be treated if the social benefits of treatment exceed the social costs.[20] It seems reasonable to assume that providers can observe the benefits of treatment and that the insurer cannot. Consider that the provider receives a flat rate per patient and a transfer, $P = F + pN$. As the treatment costs of patients differ substantially, the provider has an incentive to reject patients with high costs.[21]

If treatment costs were observable to the insurer, the fee could be adjusted to costs and efficient treatment decisions could be induced. However, it seems at bit unrealistic to assume that insurers are fully informed about actual treatment costs. Substantial selection incentives will remain. These can be mitigated if insurers bear some of the actual treatment costs.[22] For a formal derivation of this result see Breyer et al. (2003, pp. 365-368).

Ma (1994) combines cost reduction efforts and the incentives to provide quality. When providers are able to reject patients, the optimal contract implementing the first-best effort is a mixture of prospective payment and cost reimbursement. Ellis (1998) does not analyze cost reduction efforts but adds monopolistic competition to the selection story.

1.4.3 Demand inducement

In Section 1.2.3 we already discussed the problems that arise when the quality of care is not observable or when it is ex post not possible to judge whether a treatment was actually necessary. Considering credence goods, Emons (1997) showed that, in the case of idle capacity, the provider will start to treat too much as long as the mark-up for treatments is positive. Consider that $\gamma = 0$ then physician induced demand (PID) can only obtain when the fee-for-service β exceeds marginal costs (of treatment and effort). If β is below marginal costs, there is cost sharing and cost reduction incentives

[20] In this subsection we mostly follow Breyer et al. (2003, pp. 376-385) since they concentrate on selection incentives only. There is, of course, a more general literature beyond that, e.g., Ma (1994) and Ellis (1998). See the brief discussion below.

[21] This compares to the cream skimming incentives in sickness fund competition when there is community rating (see Section 1.6.2.2).

[22] Again, there is an analogue in sickness fund competition: the selection incentives stemming from community rating may be mitigated with a risk adjustment mechanism. As these are usually highly incomplete substantial selection incentives remain (see Section 1.6.4). Cost sharing with the central sponsor may improve the market outcome.

arise. This could result in another from of inducement: the provider may induce the patient not to choose treatment, i.e., there may be rationing.

There is a large debate about whether PID is a serious problem for health care markets. This not only demonstrates that fee-for-service reimbursement is widely used but also that the fee seems to be sufficiently large and thus makes demand inducement profitable. For an overview on PID and, especially, on the empirical literature testing for PID, see McGuire (2000, Section 5).

1.4.4 Non-price competition

When prices are regulated, providers will resort to other variables to increase market share. Obvious dimensions are the quality of care and practice or hospital location. The provider's specialization is also likely to be affected by strategic concerns. Location and specialization refer to horizontal product differentiation. The health care market is thus characterized by monopolistic competition. This is most easily analyzed in a Hotelling (1929) or Salop (1979) type model. Adding a vertical dimension, i.e. the quality of care, is straightforward (see, for example, Neven and Thisse (1990)).

The Hotelling as well as the Salop model have some common peculiarities that require some discussion. First, there is unit demand, that is, each single buyer buys at most one unit of the good. There is no uncertainty what makes (health) insurance redundant. Moreover, there is no difference between capitation payment and fee-for-service reimbursement so that PID plays no role. Second, there are typically exogenous constant marginal costs of production. We can therefore abstract from the moral hazard problem in cost reduction discussed above. Third, to providers, patients are typically identical. In fact they differ in the horizontal dimension, but the costs arising from a potential mismatch is usually entirely borne by consumers. Thus, there are no selection incentives.[23] Fourth, in the circular model of Salop (1979) analyzing entry is straightforward.

These nice features of the models enable us to concentrate on non-price competition and its (potentially) associated inefficiencies.[24] Horizontal and vertical differentiation are usually analyzed in a sequential game where location choices precede quality decisions. The interaction between quality and location choices has been investigated by Economides (1989) under price

[23] One exception is Calem and Rizzo (1995) who consider that providers have to bear some costs of the mismatch. However, they do not address selection incentives.

[24] Of course, these features may also be viewed as a disadvantage as important dimensions of the health care market are neglected. However, a joint analysis is simply untractable.

competition and Brekke et al. (2002) under price regulation. A general result is that quality competition can be softened by locating further apart. Brekke et al. (2002) show that price regulation can be beneficial. The idea of price regulation can already be found in Ma and Burgess (1993). However, they only consider vertical product differentiation and take locations as exogenously given. Nevertheless, price regulation may reduce inefficiencies in quality.

In models combining horizontal and vertical product differentiation like, for example, Gravelle (1999), Brekke et al. (2002), Chapter 3, and Chapter 4, the optimal (uniform) price will not result in efficiency. An increase in price will tighten quality competition as the marginal patient becomes more valuable. To dampen costly competition, providers aim at making their products less substitutable, i.e. they locate further apart or they differentiate more. The second-best price is characterized by the optimal trade off between the net benefits derived from quality provision and the mismatch costs arising from horizontal product differentiation. In Chapter 3 a circular model is applied. As a symmetric equilibrium is assumed, location is of minor importance. Instead the focus is on entry. The second-best price is then characterized by the optimal trade off between quality and the number of firms, i.e., physician density.[25]

1.4.5 Reimbursement and incentives in practice

There is some international variation in reimbursement rules. Hospital payment is nevertheless mostly prospective giving strong cost reduction incentives to providers. One exception is Portugal where costs are mostly reimbursed. If providers receive a uniform prospective payment per patient, i.e. a flat rate per case, selection incentives may arise when patients differ in treatment costs. As argued above, cost sharing would then be an appropriate response to mitigate risk selection. This, in principle, means that payments must be increasing in the severity of illness or in the costs of treatment. One widely used approach are Diagnosis Related Groups (DRGs), where payment is contingent on diagnosis. To a different extent DRGs are applied, for example, in Sweden, Australia, the United States and, from 2004 onwards, in Germany. A DRG payment system clearly reduces selection incentives but may induce manipulation of diagnoses.

[25] The analysis of location and quality in a Hotelling (1929) duopoly model and the analysis of entry and quality in a Salop (1979) model are very similar. Close locations in the Hotelling model obtain when quality competition is relatively soft. Then there would be many firms in the Salop framework and consequently firms will be closely located.

In Denmark, Portugal, Sweden, Spain, and Italy, physicians are widely paid a salary. Physicians are thus not exposed to any financial risk. Another dominant form of reimbursement is fee-for-service. Although there may be incentives to induce demand, Belgium, Austria, Switzerland, and Germany opted for such a system. The German case provides some evidence for demand inducement. There is not only a fee-for-service but also a global budget since 1993, i.e., a relative value system is applied. Every service is scored with a certain number of points. The monetary value of one point is endogenous. It is determined by dividing the budget by the total number of points submitted for reimbursement by all physicians. Consequently the global budget is always met. The monetary value of one point dropped continuously while the number of services performed increased rapidly. This observation is well in line with the so called 'target income hypothesis' and PID.[26]

It is tempting to compare national health care expenditures and identify the impact of the reimbursement system on aggregate expenditure. However, it is well known from the empirical literature on international health care expenditure comparisons (see, for example, Barros (1998)) that institutional details, including reimbursement systems, have no significant explanatory power.

To assess the impact of reimbursement rules on expenditure (or on provider behavior) one has to rely on natural experiments, like the German one mentioned above. Pauly (2000, pp. 556-557) reports some empirical observations. The switch of Medicare, the health insurer for the elderly in the United States, from cost reimbursement to fixed prices for hospitals in the mid-1980s resulted in lower growth rates of costs and in a reduced average length of stay. Hillman et al. (1989) found that capitation and salary leads to lower supply than under fee-for-service. Pauly (2000, p. 557) concludes that the incentive effects in the hospital sector are largely as expected but that results for physicians are more ambiguous.[27]

1.4.6 Contribution of the thesis

Due to the peculiarities of the health care market, markets are typically highly regulated. This mostly includes some degree of price regulation. We discussed some objectives of price regulation above. Provider reimbursement may, for example, be designed such that providers have proper cost reduction incentives. This generally requires some amount of cost sharing.

[26] In Chapter 3 we provide an alternative explanation for this phenomenon.

[27] For an overview on health care utilization studies see also Glied (2000, pp. 731-739).

To prevent providers from favorable risk selection, reimbursement should be risk adjusted. Moreover, demand inducement can be mitigated by not rewarding the amount of treatments performed or invoiced. Although the scope of price regulation is quite different in the scenarios mentioned, there is one thing they have in common: the price is not under the control of the provider, i.e., there is no price competition.

When there is no possibility of competing in prices, providers will resort to other variables to increase profits. In Section 1.4.4 we already argued that quality and location are obvious strategic variables. When a patient does not have to pay for treatment, the decision about which provider to consult will be based on other characteristics. If the patient can observe the quality of treatment, which is what we will assume in Chapter 3, providers have incentives to invest in quality. This quality may be interpreted as medical machinery since good equipment will typically be positively related to treatment quality. But quality also includes non treatment related categories such as catering quality or the size of hospital rooms. Consider that patients are indifferent after they have evaluated quality. Then a particular patient would approach the provider which is closest. Thus patients are not only responsive to quality but also to location.

The effect of price regulation is analyzed in a 3-stage game of complete information where the providers of medical care have the following choices: entry, location and quality. The sequential structure is motivated by the different degree of irreversibility in strategic decisions (see Chapter 3 for a detailed discussion). As entry is analyzed, it is convenient to use the circular model of Salop (1979). In the subgame perfect Nash equilibrium the quality of care increases in price. With an increase in price the marginal patient becomes more valuable thus stimulating quality investments. However, to escape costly quality competition firms will locate as far apart as possible. In a circular model this basically means that providers will be arranged symmetrically around the circle. Taking symmetry as given, the increase in price feeds back to the entry stage. As quality competition is intensified there will be fewer entrants. Note, that the equilibrium distance of firms increases and that quality competition relaxes accordingly.

We then analyze price regulation. And, as in most sequential games, the regulator's ability to commit to price policies is crucial for the outcome of the game. We consider two cases. First, we allow for perfect commitment. The regulator will set the second-best price and the market will end up with excess entry and too much quality. Second, partial commitment is considered. The regulator now sets the price after entry and location decisions have been made, but prior to quality decisions.[28] Then excess entry would occur and total quality provision would be efficient. As social welfare is higher in the commitment case, the regulator would like to bind herself to the second-best policy.

It is not possible to implement the first-best outcome, since there are two regulation goals (to induce the efficient number of firms and efficient quality of care) and only one regulatory variable (price). So the regulator might think of obtaining another instrument. In Section 3.5.2 we derive a number of policies that would yield efficiency. Licence fees are probably the most obvious instrument. The price could be used to induce efficient quality and the fee could be set such that the efficient entry occurs. Another option is to allow for more competition. If providers are also allowed to compete in prices, and if price and quality decisions are simultaneous, then the time consistent regulatory outcome would result. Thus, price competition does not mitigate the inefficiencies. Note, that this result may also be phrased differently: if the regulator has full commitment power, then restricting competition by regulating prices is socially beneficial.

The chapter contributes to the literature on horizontal and/or vertical product differentiation, e.g., D'Aspremont et al. (1979), Novshek (1980), Salop (1979), Gabszewicz and Thisse (1980), and Neven and Thisse (1990), by considering price regulation. In contrast, Ma and Burgess (1993) allow for price regulation. As they consider fixed locations, they cannot analyze the relationship between quality and location. Chapter 3 is closely related to Economides (1993) and Gravelle (1999). Both consider the same strategic variables and the same sequential structure as we do. In their models, location choices have no effect on quality provision thus precluding (non-price) competition in the presence of price regulation. Economides (1993) focuses on different sequential structures and their impact on equilibria. In contrast to Chapter 3, Economides (1993) does not consider price regulation. Our model builds on both Economides (1993) and Gravelle (1999) in order to consider the important effects of location choice on quality choice, thereby capturing non-price competition in the health care market. In addition, we explicitly consider, as these and the other papers mentioned do not, problems of time consistency. With regard to Gravelle (1999), who applies a similar model to the health care market, Chapter 3 contributes to the understanding of the relationship between the second-best optimum, the

[28]In Chapter 3 this variant of the model is labelled 'time consistent regulation'. This is justified as the regulator has no incentive to change the price after quality decisions have been made. Since there is no shadow price of public funds the regulator is ex post indifferent between all prices.

time consistent regulatory outcome and the equilibrium with price competition. Moreover, adding two more aspects to the Gravelle (1999) analysis, we explicitly derive a number of first-best efficient regulatory schemes and provide some empirical evidence.

1.5 Market transparency and gatekeeping

1.5.1 Credence goods and experience goods

Consider the German public health insurance system. In a recent report to the federal ministry of health the so-called concerted action committee in health care (Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen, SVRKAiG) found simultaneous occurrance of underprovision, overprovision, and inappropriate provision of health care (see SVRKAiG (2002)). If an individual suffers from a bad back, for example, which is a national disease in Germany, it is likely that there are taken too many x-ray images (overprovision). At the same time too little preventive care offered to avoid back problems (underprovision). Inappropriate provision is the rule rather than the exception. There may be inappropriate referrals by general practitioners to specialists or hospitals (including no referral in cases where a referral would have been efficient). The prescription of drugs and aid is highly inappropriate. The current reimbursement rules for pharmacists involve a payment that is proportional to product prices giving clear incentives to recommend expensive drugs.[29]

The quality of care could be improved by mitigating the under-, over-, and inappropriate provision problem. However, providers' incentives to improve quality are limited as it is usually difficult for the patients to judge about the quality of care or, more general, about their valuation of medical care prior to consumption. As was already motivated in Section 1.2.3 patients may learn their valuation through consumption. Then medical care is an experience good (Nelson (1970)). However, it may well be that this is impossible even after consumption. In this case medical care is a credence good (Darby and Karny (1973)). To better understand market forces and incentives a formal analysis is required.

In the simple example of Tirole (1988, p. 107) the market breaks down when there is a credence good monopolist.[30] When there is a monopoly

[29]The current health care reform plans include a switch from this type of payment to flat rates per prescription.

[30]Emons (1997) studies a credence good market and allows for entry and price competition. There are two services, diagnosis and repair (or treatment). In cases where excess

supplier of an experience good and a one-shot relationship the market also breaks down unless there are some a priori informed consumers (Tirole (1988, pp. 107-108)). For the case of health care these informed patients could be, for example, the chronically ill. As they are responsive to quality, incentives to invest in quality are created.[31] Higher quality benefits both the informed consumers and the uninformed ones. In Tirole's example, more information is always better as it favors efficiency. A social planner should thus increase information in the market, e.g., by allowing for consumer reports. However, the impact of consumer reports in the health care market may be limited since individual quality may well differ for two patients even though they received the same treatments and suffered from the same disease.[32] Thus, the value of the own experience exceeds the value of the experience made by somebody else. Moreover the physician-patient relationship is an important component of the quality of care or, at least, in the perception of it (Breyer et al. (2003, p. 175)).

With frequent diseases like, for example, infections, there are potentially repeat purchases. The likelihood of a repeated office visit by a patient is higher the higher the quality learned from the initial treatment. This gives providers an incentive to provide high quality. To secure future demand, building up a reputation pays. How well these reputation mechanisms actually work depends on the accuracy of quality signals. Zweifel and Breyer (1997) state that "Frequently, the quality of medical service cannot be judged correctly even after utilization, since the causality between treatment and change of health status may be blurred by other biological processes, such as the self-healing powers of the body" (p. 134) pointing to the considerable noise of these signals. Arrow (1963) is even more sceptical and states "Recovery from disease is as unpredictable as is its incidence" (p. 951).

The empirical evidence from Germany suggests that reputation mechanisms seem to be too weak to induce proper quality provision. Thus, all measures ensuring a certain minimum quality level would be desirable. Entry regulation and educational requirements are such measures. Strengthening liability rules (malpractice suits) for providers or the introduction of disease management programs may also help. As is clear from Section 1.4.4, price regulation may also be beneficial.

---

entry occurs the specialists (or physicians) have idle capacity and start treating too much and induce demand.

[31]The same mechanism is at work in Chapter 4 where the impact of gatekeeping general practitioners on specialist competition is analyzed.

[32]In their empirical study, Dranove et al. (2003) showed that report cards worsen health outcomes.

1.5.2 Market transparency and product differentiation

In Section 1.4.4 above we argued that, in the light of price regulation, providers of medical care will shift competition to other strategic variables. Basically, providers have two variables at their disposals, the quality of care and specialization (or location). From the preceding subsection it is clear that informational asymmetries will affect quality competition and, in general, quality competition will be weaker than under complete information. Here we will briefly describe the potential impacts of imperfect information about the vertical dimension on the horizontal dimension of product differentiation or on the number of varieties in the market. In Section 1.4.4 we already motivated that horizontal product differentiation is most conveniently analyzed in a spatial model like the ones of the Salop (1979) or Hotelling (1929) type.

First, consider the circular model of Salop. Riordan (1986) considers experience goods, i.e., consumers learn the quality through consumption. Higher quality thus induces more repeat purchases. The consumers' responsiveness to quality is nevertheless lower than under complete information resulting in less than optimal quality provision. As quality is costly, profits of firms are higher than in the perfect information case. With free entry the higher attractiveness of the market leads to more entrants, i.e., there are more varieties in the market than under perfect information. In the alternative location interpretation the equivalent result would be a higher physician density in the incomplete information case.[33]

In the standard Hotelling (1929) model there are two forces at work. The market capturing effect is the centripetal force in the market pushing the firms close together. As products then are close substitutes, (price) competition is intense. To relax competition firms locate apart from one another. This is the centrifugal force in the market.[34] D'Aspremont et al. (1979) have shown that the centrifugal force outweighs the centripetal force resulting in maximum product differentiation. Bester (1998) demonstrated that this might not hold when there is uncertainty about product quality. Like Riordan (1986), he considers an experience good and repeat purchases. As high prices signal high quality, prices become rigid relaxing price competition. This reduces incentives for product differentiation and may result in 'minimum differentiation'.

In Bester (1998), more information increases differentiation incentives.

[33] Wolinsky (1984) considers imperfect information about product variety and obtains ambiguous results, i.e., there may be too little or too many varieties in the market.

[34] This mechanism persists when prices are regulated and fimrs are engaged in quality competition (see, for example, Brekke et al. (2002)).

This is the other way round in Schultz (2002). There are, like in Chapter 4 below, informed and uninformed consumers. An increase in market transparency, i.e. an increase in the share of informed consumers, then leads to less product differentiation. However, as firms locate outside the unit interval in any case, there always is 'maximum differentiation'.

Although this brief overview revealed ambiguous effects of information on product differentiation it nicely demonstrated that uncertainty along one dimension, in general, feeds back into other dimensions of competition. The lack of information and its impact on provider competition is carefully modelled in Chapter 4 below.

1.5.3 The key position of gatekeepers in health care

In his article for the Handbook of Health Economics, Scott (2000) addresses the 'Economics of general practice'. He analyzes the general practitioner (GP) - patient relationship, the utilization of GP services, and GP behavior in general. Although all these issues are important, we neglect them as the gatekeeping GPs in Chapter 4 are assumed to be 'machines' acting perfectly on behalf of their patients.

Scott (2000, p. 1177) mentions three positive effects that are attributed to gatekeeping. First, gatekeeping is said to reduce health care costs by reducing unnecessary treatments. Second, the secondary care sector is used more efficiently as the patients more appropriately treated by a GP are screened out. Third, GPs have better information about the quality provided in the secondary care sector. They can simply aggregate the experiences over the patients referred. Better matches will result, increasing the efficiency of care.

For the reasons mentioned, many countries have given general practitioners a central role in the health care system. In the National Health Service (NHS) in the United Kingdom, for example, people have to enroll with a GP. If they need specialist treatment or hospital care, they need a referral from their GP. There is thus no direct access to secondary or hospital care.[35] In the United States there are two public programs (Medicare for the elderly and Medicaid for the poor), the private health insurance market, and there is Managed Care. The most important form of Managed Care is the Health Maintenance Organization (HMO). HMOs have (few) contracted providers. Patients have to give up free choice of provider upon enrollment, i.e., they can only consult contracted providers. In several HMOs, free choice is fur-

[35] For a brief overview on the NHS see Schulenburg and Greiner (2000, pp. 195-207). A detailed treatment is provided in European Observatory on Health Care Systems (1999).

ther restricted by gatekeeping.[36] Finally, the Scandinavian countries have gatekeepers and there are some aspects of gatekeeping in the current health care reform proposal in Germany.

The widespread enthusiasm for gatekeeping is a bit surprising as evidence for the positive effects is lacking. International comparisons of aggregate health care spending, for example, usually find no significant impact of institutional variables, including gatekeeping, raising doubts whether the first effect is actually important (see, e.g., Barros (1998)).[37] With the current health care reform proposal, German sickness funds will be obliged to offer 'gatekeeper programs'. If a customer enrolls in such a program the insurer is allowed to offer a premium reduction. This natural experiment may provide some evidence in favor of gatekeeping if premium reductions occur.

1.5.4 Contribution of the thesis

Taking the third positive effect of gatekeeping literally the problems associated with experience goods can be mitigated by giving GPs a gatekeeper role in the health care system. Putting it to the extreme, what we will do in Chapter 4, GPs are perfectly informed about the secondary care market. If one abstracts from agency problems and assumes that GPs truthfully convey their information, a patient can make informed choices once a GP is consulted.

The conventional wisdom is that more information is better since consumers can then make informed choices and end up in better matches. But this argument neglects the potential competitive effects. In Chapter 4 we fill that gap.[38] We consider a Hotelling (1929) model with two providers of specialist care. The vertical product characteristic is, like in Chapter 3, quality. The horizontal dimension now is the specialization of providers. There are informed patients and uninformed patients in the market. Informed patients can observe their own disease and the quality and specialization of providers, i.e., they have complete information. Uninformed consumers cannot observe specialization and quality and are, additionally, unaware of their own disease. They only know that specialist treatment is required and

[36] For more information on Managed Care see Glied (2000).

[37] Glied (2000, pp. 740-743) summarizes some studies about managed care and cost growth and finds no unambiguous results. As gatekeeping is most widely applied in managed care there seems to be no clear impact of gatekeeping.

[38] For an oligopolistic market for experience goods, Krähmer (2003) has shown that the value of information may be negative as the nature of competition is affected by the amount of information available.

therefore pick a secondary care provider randomly.

We consider two versions of the basic model. In both versions we consider that consumers have complete information once a gatekeeping GP is consulted. In the first part (Section 4.3) we consider the share of informed individuals to be exogenously given. In the second part (Section 4.4) we endogenize this share by modelling the decision to consult a gatekeeping GP.

When the share of informed patients is exogenously given (this may, for example be the chronically ill who have obtained all relevant information through repeated consumption), the regulator can only demand compulsory gatekeeping, i.e., all patients have to consult a GP before accessing the secondary care market (strict gatekeeping). We refer to this variant of the model as 'direct gatekeeping'. Consider that the regulator imposes a strict gatekeeping regime, then the increased information in the market increases the responsiveness of consumers to investments in the quality of care. Quality competition thus intensifies. As already discussed in Sections 1.4.4, 1.4.6, and 1.5.3, providers will try to escape costly quality competition by making their products less substitutable. In the framework considered in Chapter 4, this means that specialization in the market increases. If the mismatch costs to patients associated with specialization are relatively small, the quality of care is an important determinant of demand. Quality competition is then tight and thus very costly. Excessive specialization may result so that mismatch costs are higher with informed patients than when some patients are uninformed. So gatekeeping does not necessarily improve the matching of patients with providers. In these cases gatekeeping may be socially detrimental if the improved quality does not outweigh worsened matching. This result was not previously known and it holds even though gatekeeping is considered to be costless. We demonstrate that the negative effects of gatekeeping can be mitigated by proper price regulation. The increased quality competition induced by more information can be offset by cutting the price accordingly. As quality remains unchanged, there are no relocation tendencies. Matching improves and so does social welfare. Thus, with proper price regulation, imposing a strict gatekeeping regime is always socially desirable.

In the second version of the model we endogenize the decision to consult a gatekeeping general practitioner. A patient consults a GP if the individual costs of a GP visit are at most as large as the (expected) benefits. The costs may simply be time costs. The benefit is the utility derived from an expected better match. The sequence of events is as follows: the regulator sets price, the providers choose specializations and then quality, the patients decide whether to consult a GP, and, finally, the patients demand secondary care. As above, an increase in price fosters quality competition. To soften that competition, specialization increases. If specialization is already above the efficient degree, expected mismatch costs for the uninformed patients increase. Consequently more patients will go to a GP. The market becomes more competitive, resulting in higher quality and more specialization. An increase in price thus has a twofold impact.

As the amount of gatekeeping in equilibrium can be influenced by price regulation, we refer to this variant of the model as 'indirect gatekeeping'. The welfare results are similar to those of the direct gatekeeping scenario and so is the intuition. If the parameters of the model are such that information has only little impact on the strategic variables, price regulation that yields strict gatekeeping is always socially desirable. But when the competitive effects are strong, the price optimally trades off three dimensions: specialization, quality, and GP consultation. Not surprisingly, this will, in general, not result in efficiency as there are three regulation goals and only one regulatory variable. Strict gatekeeping is not optimal. As gatekeeping now involves a cost it may be not very surprising that strict gatekeeping can be socially detrimental. Nevertheless, this insight is new.

In Chapter 4 we use a Hotelling (1929) model and combine horizontal with vertical product differentiation. In Chapter 3 we did the same for the Salop (1979) model. Since the models have many similarities, Chapter 4 is related to the same set of papers as Chapter 3, e.g., D'Aspremont et al. (1979), Novshek (1980), Gabszewicz and Thisse (1980), Economides (1989), and Neven and Thisse (1990) and Economides (1993). None of these papers considered price regulation, imperfect information or gatekeeping. Gravelle (1999) also considers price regulation but not gatekeeping. Bester (1998) introduces uncertainty about quality to the Hotelling (1929) model and shows that this may result in minimum product differentiation. This result is mirrored in Chapter 4 where we obtain no product differentiation when there are only uninformed consumers in the market. However, once some patients become informed, differentiation incentives are created. Another related paper is Schultz (2002). He analyzes price competition instead of quality competition but also considers a fraction of uninformed consumers. The specialization incentives are the other way round. Increasing the share of informed customers reduces product differentiation. However, firms always locate outside the market. Chapter 4 directly builds on Brekke, Nuscheler and Straume (2002). They analyze price regulation with horizontal and vertical product differentiation in the Hotelling (1929) model. However, they consider a game of complete information. The paper complements

Chapter 4 as it is equivalent to a strict gatekeeping regime, i.e., Brekke, Nuscheler and Straume (2002) study a corner solution of Chapter 4, though an important one.

1.6 Sickness fund competition

In the 1970s many countries began to introduce some aspects of market orientation in their health care systems. These included intensified competition amongst providers of medical care, e.g., general practitioners and hospitals, as well as competition amongst health insurance companies. In this section we concentrate on the latter. Examples of countries that introduced or fostered competition between health insurance companies (or sickness funds) are the Netherlands, Belgium, Israel, Switzerland and Germany.

1.6.1 The benefits of competition

The main motivation for allowing for (more) competition are efficiency gains. According to Van de Ven and Van Vliet (1992, p. 24), there are three positive effects of competition:

1. enhancement of the quality of care,
2. improved efficiency of care,
3. and more responsiveness to consumer preferences.

These benefits can easily be obtained with selective contracting, i.e. when sickness funds can decide about the providers to contract with. Sickness funds may simply contract with providers that offer high quality at reasonable prices. The trade off between quality and its costs may be driven by the preferences of their enrollees. Competition between sickness funds for (new) members thus induces competition amongst providers. When the possibilities of selective contracting are restricted, like, for example, in Germany, a lot of the potential gains of sickness fund competition cannot be realized. Nevertheless, some gains can be realized even without selective contracting. All other things equal, inefficient sickness funds, i.e. funds with high administrative costs, have to demand higher contribution rates or premiums than efficient funds. As a result, the most inefficient firms will be driven out of the market. The third benefit of competition can, for example, be addressed by benefit packages. These can be tuned so that they meet the preferences of the target group. The quality of service and advice is another example.

1.6.2 Risk selection

Allowing for competition in the health insurance market is commendable and it promises efficiency gains. However, there may be some serious drawbacks of competition and the most important among these is risk selection. We distinguish two types of risk selection: adverse selection and cream skimming. Whether there is adverse selection or cream skimming crucially depends on the information available to insurers and consumers.

1.6.2.1 Adverse Selection

Consider there are two risk types in the population, high risk consumers (consumers with high expected costs of health care utilization) and low risk consumers (consumers with low expected costs of health care utilization). Then, at any premium level, the high risk consumers demand more health insurance coverage than the low risks consumers. If insurers can observe the risk type of consumers the market outcome will be efficient. Both types will receive full coverage and fair premiums, i.e., premiums that equal their individual expected health care costs.

However, complete information is a heroic assumption. It seems more plausible to assume that consumers are better informed about their own risk type than the insurer. When the risk type is private information of consumers, *adverse selection* may obtain.

Suppose that an insurer offers the efficient contracts without being able to observe consumers risk types. Then the high risk consumers would claim to be low risks in order to obtain full coverage at the reasonably lower premium. As the contract designed for the low risks will also attract the high risks (adverse selection), the insurer will incur a loss. The 'efficient contracts' can thus not be the equilibrium of the competitive health insurance market.

To prevent the high risks from mimicking the low risks the contract for the latter must be distorted. Incentive compatibility requires a reduction of coverage. As Rothschild and Stiglitz (1976) have shown, there may be an equilibrium where this procedure results in an equilibrium. Assume that such a separating equilibrium exists. Then the high risks would be indifferent between the contract designed for them and the one designed for the low risks.[39] The informational asymmetry thus harms the low risks only. They receive only part coverage (at fair premiums) but the high risks get their efficient, full coverage, contract.

[39] Although they are indifferent it is assumed that the high risks choose the 'high risk contract'.

Rothschild and Stiglitz (1976) have also shown that the separating equilibrium is the only possible equilibrium. In particular no pooling equilibrium can exist, i.e., there is no equilibrium where risk independent premiums obtain. Thus, there never is an equilibrium where the low risks subsidize the high risks. Nevertheless there may be a pooling contract that destroys the candidate separating equilibrium. This is the case when the share of high risk consumers is low. Then the health insurance market may break down or is inherently unstable.[40]

1.6.2.2 Cream Skimming

To a large extent health status is not considered to be the responsibility of the individual (see, e.g., Kifmann (2002, p. 719)). For equity reasons the regulator mostly requires community rating, i.e., insurers must charge the same premium to all insured. Of course, given risk aversion, this could also be justified by an efficiency argument from an ex ante perspective. Prior to realization of the risk type each individual strictly prefers community rating over risk related premiums as it enables him or her to obtain an insurance against premium risk. Community rating creates considerable incentives for risk selection as sickness funds make losses with the high risks and earn profits with the low risks. Funds will thus try to select the low risks. Hence the incentive for favorable risk selection does not necessarily arise from asymmetric information, it may be a result of regulation (Pauly (1984) and Wilson (1977)).

Barros (2003, p. 420) and Van de Ven and Ellis (2000, p. 773) define *cream skimming* as risk selection that occurs because insurers prefer low risks to high risks. As is clear from the above, community rating creates significant cream skimming incentives. Contingent on the insurers' information, Van de Ven and Ellis (2000, pp. 773-774) distinguish three forms of cream skimming:

1. If health plans can identify risk types, they can directly select the low risks ('direct cream skimming'). They could, for example, reject females and the elderly upon enrollment and thereby obtain a low risk

[40] Wilson (1977) introduced a stronger than Nash equilibrium concept. Firms anticipate the consequences of their offers, i.e. they take into account the fact that contracts that become unprofitable through their offer will be taken from the market. In such an environment an equilibrium always exists. Hence the health insurance market is stable. Whether a particular health insurance market is unstable or not is thus basically an empirical question. Note, however, that a pooling equilibrium can exist in the Wilson model. This also applies contracting costs are introduced as in Newhouse (1996).

pool consisting of young males. Such a favorable risk pool can also be attained by, e.g., selective advertising and golden handshakes.

2. Suppose insurers cannot observe the risk type but know about other non-observable but relevant, i.e. costly, risk factors. Examples are diseases like cancer or AIDS. Then insurers can prevent these high risks from enrolling by providing bad service and bad therapy to both cancer and AIDS patients. With selective contracting insurers may contract with providers with bad reputations in both fields.

3. If the insurer cannot observe the risk type and has no idea about any relevant risk factor, he can still select the low risks by offering low-option plans with (high) deductibles or by reducing offices and information centers to a minimum.

In the last two versions adverse selection is used as a cream skimming device. We refer to them as 'indirect cream skimming'. For some additional means of indirect risk selection see, for example, Breyer et al. (2003, p. 298), Van de Ven and Van Vliet (1992, pp. 28-31) and Van de Ven et al. (2003, p. 91).

1.6.2.3 The adverse effects of risk selection

Consider a separating equilibrium with full coverage at high premiums for high risks and part coverage for low risks at low premiums emerges. As low risk individuals prefer full coverage, there is an efficiency loss from adverse selection. Additionally, the high premiums may prevent non-affluent high risks from obtaining sufficient coverage. The health insurance market may be unstable coming along with real social costs (Van de Ven and Ellis (2000, p. 775)).

Moreover, and maybe more relevant, risk selection counters the positive effects of competition. Van de Van and Van Vliet (1992, p. 24) and Van de Ven and Ellis (2000, pp. 775-776) mention three adverse effects:

1. Competition in the health insurance market is supposed to increase the incentives to provide high quality. But when an insurer (selectively) contracts with high quality providers or invests in good service, e.g., by increasing the number of offices or by good access to call centers, it risks mainly attracting the high risks. This is why insurers do, or are at least tempted to do, the exact opposite: provide bad service and contract with bad providers. In the short run such a policy promises good profits. However, in the long run insurers risk their reputation.

High risks will nevertheless enroll with such an insurer when the negative effects of low quality are outweighed by lower contributions. Thus, in equilibrium, all insured may receive low quality.

2. Inefficient funds that successfully engage in risk selection could drive efficient funds out of the market or at least capture some of their market.

3. Cream skimming is itself costly. Since there are no positive welfare effects from reallocating individuals to funds this is pure waste.

1.6.2.4 Empirical evidence

There is a large empirical literature on risk selection. One of the most recent studies is Nicholson et al. (2003). For U.S. data they find that Health Maintenance Organizations (HMOs) successfully select the low risks. HMOs basically use their benefit package for selection.[41]

One the most prominent examples of adverse selection is the 'Harvard death spiral' (see Cutler and Zeckhauser (1998, pp. 11-14), Cutler and Zeckhauser (2000, pp. 616-623), and Cutler and Reber (1998)). At the beginning of the 1990s, Harvard University offered its employees two insurance types, a generous Preferred Provider Organization (PPO) and a number of HMOs. The University paid about 90 percent of the contributions. Although the PPO was more expensive than the HMOs the additional costs to be borne by the PPO enrollees were relatively low. In 1995 Harvard decided to only pay a fixed percentage of the lowest cost plan. By fostering competition, plans should have had the right incentive to contain costs. The young and healthy switched from the PPO to an HMO. To break even, the PPO had to raise its contribution leading to further switches. Once again these were the (relatively) young and healthy. In 1998 the increase in the premium would have been too large so that PPO was closed down. Cutler and Reber (1998) estimate the associated welfare loss to people who would have enrolled with the PPO to be 2 to 4 percent of baseline premiums.

1.6.3 Regulatory measures

The adverse effects of risk selection are numerous and therefore public policy interventions are justified. In the following we describe measures that are taken or that can be taken to reduce incentives for risk selection. However,

[41] Cutler and Zeckhauser (2000, pp. 616-624) provide an excellent overview of the empirical literature on risk selection.

the benefit of each measure has to be weighed against the costs that may arise, for example, from reduced competition.

The regulator usually requires open enrollment, i.e., insurers must not reject (high risk) applicants. The most straightforward form of direct cream skimming is thus not feasible. Insurers could be allowed to use their information for risk rating premiums. Then, as was pointed out by Pauly (1984), there would be no cream skimming incentives. Instead premiums would depend on observable risk characteristics. However, risk rating could make it difficult for high risks to obtain convenient coverage.[42]

Individuals can be mandated to buy some basic health plan coverage. Thereby the most extreme form of adverse selection, where the low risks escape cross subsidization and go without any coverage, can be avoided. Depending on the coverage of the basic health plan, such a policy can yield a Pareto improvement. If the coverage is too large, the low risks may be worse off than without any health insurance. However, such a mandatory pooling contract may increase risk selection incentives in the supplementary health insurance market. If providers of the basic health plan are also allowed to supply supplementary insurance, Kifmann (2003) demonstrated that supplementary contracts can well be used as a selection device. Thus separating the provision of basic and supplementary health insurance is the imperative.

Health insurers can skim the cream irrespective of the information available, i.e., for all three forms of cream skimming, by selective contracting or by the design of their benefit packages. This can be simply prevented by complete regulation of the benefit package and by forbidding selective contracting. But the costs may be substantial. Without selective contracting insurers would be prevented from contracting only with cost-effective providers. So overall efficiency may decrease. The health plans are less able to respond to consumer preferences when a uniform benefit package is dictated. This clearly yields a welfare loss (Keeler et al. (1998)).

As low risk individuals tend to choose contracts with higher deductibles, the regulator could restrict free choice of deductibles or forbid deductibles at all. Again, this yields a welfare loss from reduced responsiveness to consumer preferences. Additionally (ex post) moral hazard problems become more severe distorting health care utilization. Selective advertising or advertising in general could be forbidden. If advertising is informative this may also result in a welfare loss. Van de Ven and Ellis (2000, pp. 778-779)

[42]As already argued above, community rating is imposed for equity purposes or rather efficiency reasons from an ex ante perspective. As community rating just creates cream skimming incentives this regulation induces demand for further regulatory measures.

mention some additional measures, e.g., risk adjustment and risk sharing.

1.6.4 Risk adjustment

The most prominent measure taken to prevent risk selection, or at least to mitigate the distortions caused by risk selection, is risk adjustment. Thereby contributions to every single fund should be adjusted so that they reflect the actual risk structure. Cream skimming or adverse selection may then occur within risk premium groups. An accurate mechanism results in rather homogenous groups. With an accurate mechanism the benefits of favorable risk selection are low and, as risk selection is more difficult, the costs are likely to be high. This measure is most prominent as it does not affect competition (Van de Ven et al. (2003, p. 78)). Good risk adjustment eliminates risk selection incentives *and* ensures tight competition.[43]

In Figure 1.1 below we show how the payment flows can be organized. System A) may be called an external subsidy system, while B) is an internal subsidy system. Of course, the two systems are conceptually identical. The basic purpose of the figure is to emphasize that there is a central sponsor or solidarity fund involved in risk adjustment. If risk adjustment is incomplete, it may be complemented by *risk sharing*. With risk sharing the funds are not fully responsible for health care costs as the solidarity fund covers part of them. This obviously reduces selection incentives. Risk sharing also obviously reduces the incentives for efficiency (see also Section 1.4.2.2).

A complete overview of risk adjustment and risk sharing is beyond the scope of this section. We concentrate on variables that can be used to adjust premiums, i.e., on risk adjusters and their predictive power.[44] To assess the accuracy of risk adjustment, information is required about the maximum variance of health care expenditure that can be explained. Newhouse (1996) and Van Vliet (1992) found the predictable proportion to be around 20 percent. In the following we concentrate on prospective risk adjustment mechanisms.

Health care spending clearly increases with age. Spending patterns of males and females differ significantly. So the most obvious risk adjusters are age and gender. Pope et al. (1998) found that such a simple demographic model only explains roughly 1 percent of health care expenditures.[45] In-

[43]Cutler and Zeckhauser (2000, p. 625) and Van de Ven and Ellis (2000, pp. 780-783) mention some drawbacks. To some extent risk adjustment can be manipulated and can distort incentives.

[44]For a comprehensive treatment of risk adjustment and risk sharing see Van de Ven and Ellis (2000, pp. 779-828).

[45]Concerning the predictive power we only refer to Pope et al. (1998) which is a

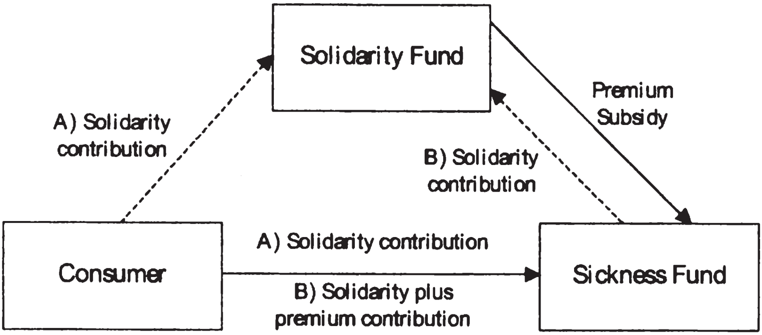

Figure 1.1: Risk adjustment systems in A) Belgium, Israel, and The Netherlands and B) in Germany and Switzerland. Source: Van de Ven et al. (2003, pp. 77-78).

cluding prior year expenditures increases R^2 to 4 percent. However, such a scheme is easily manipulated as an increase in expenditure increases subsidies in the subsequent year. The same applies to the so-called Pharmacy Cost Groups (PCGs) where chronic conditions are inferred from prescription drug consumption.

Diagnostic information may also be used to improve risk adjustment. There are several measures that could be taken, for example, Ambulatory Care Groups (ACGs), Diagnostic Cost Groups (DCGs), Major Diagnostic Categories (MDCs), Hierarchical Condition Categories (HCCs), and the Disability Payment System (DPS).[46] Pope et al. (1998) find an R^2 of 7 percent for an age, gender, DCG/HCC model. Adding diagnostic information to age and gender thus increases the predicted variance by 6 percentage points. But there are also some efficiency costs as such a model may lead the plans to screen aggressively for certain conditions in order to receive higher subsidies.

Including self-reported health measures is more a theoretical option as its application would require asking the entire population to fill out a health

study about US Medicare. International comparisons of R^2 may be misleading as the predictability strongly depends on services that are subject to risk adjustment (Van de Ven and Ellis (2000, 792)).

[46] For an explanation of these concepts see Van de Ven and Ellis (2000, pp. 798-802).

questionnaire. This is simply not feasible. However, including the self-reported health status into an age gender scheme increases R^2 from 1 to 3 percent.

1.6.5 International experience

In a recent special issue of Health Policy on 'Risk Adjustment in Europe', the experiences with sickness fund competition in 5 European countries, namely, Belgium, Germany, Israel, The Netherlands, and Switzerland, are summarized. The information presented here is taken from Van de Ven et al. (2003) which is part of that special issue.[47]

All countries introduced or fostered competition among health insurers during the 1990s. The premiums are risk adjusted in all countries. However, they use different risk adjusters and the complexity of the schemes also varies. The most incomplete risk adjustment can be found in Israel where only differences in age structures are offset. Switzerland additionally adjusts for gender and region. The risk adjusters applied in Germany are age, gender, disability, sick pay claim, and income (p. 87). All schemes have in common that they are highly incomplete, as should be clear from the previous section. As all countries except Israel have community rating, there are substantial incentives for risk selection (p. 87).

These are softened by some degree of risk sharing in Belgium and The Netherlands. Together with the low mobility of consumers there is no selection problem. There is no risk sharing in Germany and Switzerland, i.e., sickness funds completely bear the costs of health care. Together with the high mobility of consumers, especially in Germany, serious selection problems result (pp. 87, 89). There is anecdotal evidence from both countries that sickness funds try to select the low risks (p. 91). Risk selection in Switzerland may be driven by deductibles or by supplementary health insurance. Deductibles are more or less forbidden in the German public health insurance system. Moreover funds must not offer supplementary health insurance. Although there is a large contribution rate differential between funds in Germany, there is no evidence of any successful risk selection activities (see Chapter 5).

[47]The practice of risk adjustment and risk sharing in the US and for some other countries can be found in Van de Ven and Ellis (2000, pp. 829-835).

1.6.6 Contribution of the thesis

Whether the sickness funds in the German public health insurance system actively (and successfully) try to select the low risks has not been tested so far. There is quite a lot of anecdotal evidence for risk selection activities but, up to now, whether these activities actually have any significant effect has remained unclear. Chapter 5 fills this gap.

Of course, there is some literature on German sickness fund competition but none of these studies addresses risk selection. The two reports by Lauterbach and Wille (2001) and Jacobs et al. (2002), for example, use health care expenditure data and find that switchers have substantially lower health care costs. But this does not necessarily have anything to do with adverse selection or cream skimming. It may well be that the switchers in their samples are low risks and, as these typically have lower switching costs, their results may simply be explained by a transaction cost argument.

The main empirical task is to separate risk selection activities from switching costs. This requires splitting the sample into two subsamples. One sample should consist of all funds that are suspected of actively selecting the low risks and the other subsample should consist of all remaining funds. The starting point is thus a hypothesis. When looking at the variation of (additional) benefits in Germany it turns out that the company-based sickness funds (Betriebskrankenkassen, BKKs) are more likely to provide services that healthy individuals demand and less likely to provide those meeting the preferences of the sick. Combined with the anecdotal evidence, the hypothesis must be that the BKKs try to select the low risks. The sample is thus split into BKK and non-BKK members.

To identify risk selection the transitions of consumers between sickness funds must be analyzed. Obviously, there is risk selection by BKKs if the healthy switch from non-BKKs to BKKs and the sick from BKKs to non-BKKs. Nicholson et al. (2003) take that approach and find that Health Maintenance Organizations (HMOs) practice risk selection. However, as they model the switching decision with a simple probit model for both subsamples, the impact of switching costs on transition decisions remains unclear.

To assess the magnitude of switching costs one has to distinguish different switching decisions within each subsample. Consider, for instance, the non-BKK members. These have to make a choice from three alternatives: no switch, switch to a BKK, switch to a non-BKK. This requires a multinomial model. If the fact of being a low risk rather than a high risk increases the probability of switching to a BKK more than the probability of switching to a non-BKK, then the risk selection hypothesis is supported. The hypothesis is also supported when there is a stronger health effect on the probability of switching to a BKK than to a non-BKK among BKK members. Splitting the file and applying a multinomial procedure thus enables us to test the risk selection hypothesis twice. This approach is new. To be able to obtain clear cut results, the variable measuring individual risk should be accurate and one dimensional. We fit an ordered probit model for the self-assessed health status and thereby create a health index summarizing all relevant health information. The complete empirical model thus consists of two stages with an ordered probit model for health status at the first stage and a multinomial logit model for health insurance choice at the second stage. This strategy is also new.

We find no evidence of any (successful) risk selection activities in the German system. The indications that BKKs try to select the low risks may nevertheless be true, but we show that these activities are not successful. The different (additional) benefit structures may simply reflect the preferences of the enrollees and complete regulation would yield welfare losses.

Van de Ven et. al (2003, p. 89) nevertheless conclude that risk selection is a problem in Germany. With our result we can say exactly what they mean and what should be done: due to the much larger flow towards the BKKs and lower switching costs of low risk consumers there is indeed risk separation. This originates in the historically better risk structures of BKKs together with incomplete risk adjustment. Public policy should thus be directed towards mitigating comparative advantages stemming from this asymmetry, i.e. an improvement of the risk adjustment mechanism is required. Thus, good risk adjustment is important even if no successful cream skimming activities are found. This is an important policy message with international applications. A complete regulation of benefit packages, or forbidding any type of indirect selection in general, does not help. On the contrary, this would oppose the positive effects of competition. Since distortions do not originate in the behavior of sickness funds, competition should be strengthened. Obvious dimensions are deductibles and selective contracting.

1.7 Too many trade offs for efficiency

Whether there is too much or too little competition in health care can not be answered in general. It depends on the industry under study and/or on the particular health care system and the market analyzed within the system. As we have seen, a vaccine monopolist has perverse incentives. He not only aims at keeping the disease alive but also exploits the external effects to increase markups. The resulting twofold inefficiency cannot be socially optimal in a static setting like the one adopted in the following chapter. However, things are less clear in a dynamic context. The high monopoly profits stimulate R&D activities. These would be lacking with perfect competition. Ex post efficiency thus induces dynamic inefficiency.

An important determinant of monopoly profits, or profits in general, is demand. So consider there is a number of innovating firms in a R&D race. The winning firm receives a patent and enjoys monopoly power for a certain period of time. The effort of participants depends on expected monopoly profits. Calculating expected profits is far from being straightforward. It not only depends on the probability of winning the race but also on, first, the expected price, second, whether the new product becomes part of the insurers' benefit packages, and, third, on the willingness of providers to adopt new technologies.

The willingness of providers to adopt new technologies crucially depends on reimbursement rules and on treatment costs. With capitation payment (even if on a DRG basis) a very costly new technology will never be adopted although it might by efficient. So R&D effort into costly technologies will be scarce when a major part of the market is characterized by capitation payments. Of course, R&D incentives are higher with cost sharing or fee-for-service such that, in a dynamic model, quality will be higher in these systems. But these systems distort cost reduction incentives and, consequently, costs will be higher.

However, technology adoption is not only driven by provider payment but also by competition amongst providers. When considering non-price competition in terms of quality and specialization, for example, providers can avoid the adoption of costly high quality technologies by increasing specialization of their clinics. Via this channel increased specialization feeds back into the R&D race. Another channel is patient information. If patients cannot observe the quality of providers, the providers willingness to adopt high quality technologies will be low. Increasing information in the market, for example by the introduction of a gatekeeper system, fuels R&D activities.

Quality is not only an issue on the providers side but also in the health insurance market. Consider the health insurance market to be competitive. If insurers must not charge risk based premiums (community rating) and if there is open enrollment, insurers have strong incentives to skim the cream, i.e. to actively select the insured with low expected costs of health care utilization (the good risks). As quality is more valued by bad risks (they are more likely to demand health care), insurers can prevent the high risks from enrolling by offering low quality. This can be achieved by contracting with providers of low quality. The providers' incentives to adopt new technologies of high quality will be distorted and again reduced R&D expenditures will result.

Since low quality may well be in line with consumer preferences, softening competition in the health insurance market, e.g. by forbidding selective contracting, may create a distortion. Additional inefficiencies may arise as softened competition amongst insurers also reduces competition amongst providers. Cream skimming incentives can be lowered by risk adjusting premiums. But risk adjustment schemes are typically highly incomplete such that significant selection incentives remain. Risk sharing is the second most prominent measure to mitigate inefficiencies arising from favorable risk selection. The equivalence to cost sharing mentioned above implies a distortion of cost reduction incentives.

This brief summary of incentives and of conflicting regulatory goals demonstrates that it will be, in general, impossible for a regulator to design a set of regulatory rules that implement an efficient allocation. A regulator thus has to optimally trade off a number of inefficiencies against one another. Moreover, it will also be impossible to simultaneously address all mentioned trade offs—and the ones discussed are only a small selection. This is why we concentrate only on the most important aspects in each of the following chapters. Although 'partial models' have their limitations, we think that all chapters provide important insights and significantly contribute to the scientific as well as to the political discussion on the optimal design of health care systems.

Chapter 2

Monopoly Pricing in the Market for Vaccines

2.1 Motivation

Traditionally, vaccinations were regarded as one of the prime examples of positive externalities.[1] Consequently, government intervention in the form of mandatory vaccinations and Pigouvian subsidies were considered to be appropriate policy responses to the distortions caused by the externality. More recently, this traditional view has been challenged by various contributions that produced a number of somewhat conflicting results about the form and optimality of government intervention in the market for vaccines (see e.g. Brito et al. (1991), Francis (1997), and Geoffard and Philipson (1997)). These results typically depend on the specific assumptions made in the models about agent heterogeneity, market structure and dynamics. This chapter contributes to this literature by considering strategic incentives and optimal government responses in the context of two hitherto neglected dimensions. First, individuals are assumed to differ with respect to income. Second, monopoly power on the supply side is considered.

In the existing theoretical literature agent heterogeneity is usually introduced, if at all, through the assumption that the disutility of vaccinations, e.g. side effects, varies. Empirically, disutility is difficult to observe. As empirical studies of individual vaccination decisions usually find a clear positive relationship between income and the probability of being vaccinated, introducing agent heterogeneity into the theoretical analysis through in-

[1]This chapter is almost identical to Kessing and Nuscheler (2003).

come differences would seem to be a natural step. Philipson (1996, Table 2, p. 624), for example, finds a positive income effect on the probability of measles vaccination for children in the U.S. England et al. (2001, p. 19) report that, if there is a fee, as with hepatitis B in China, "poorer people are more likely to go without essential immunization". Moreover, since government action usually affects people's incomes, such an analysis promises to be a better approximation of the consequences of different policy measures.

The second key element in our treatment is its focus on monopoly power on the supply side. This assumption is motivated by recent developments in the vaccine industry. Important changes in U.S. legislation in 1986, which effectively shield manufacturers from the liability risk of new vaccines, resulted in a substantial increase in R&D efforts and these have recently lead to a dramatic increase in the availability of a number of new vaccines (BusinessWeek Online (2002)). Russell (2002) points out that two developments have also increased monopoly power significantly. First, there has been a shift from commodity vaccines to vaccines which are heavily protected by intellectual property rights. The new Hepatitis B vaccine introduced in the late 1980s has for example about thirty associated patents. Similarly, Reiss and Strauss (1998) document that between 1980 and 1995 patent applications for vaccines at the European patent office rose by a factor of seven and that this development has been fuelled by the progress made in the field of genetically engineered vaccines in particular.[2] Second, the ongoing concentration in the industry at all levels, from research and development to marketing organizations, has left only a few key players. Furthermore, as firms are increasingly specializing in specific diseases and their core fields of expertise, competitive pressure is being further attenuated.

A vaccine monopolist has two main incentives: (i) to keep the disease alive and (ii) to increase the prevalence of the disease in order to increase the willingness to pay for vaccination.[3] In their dynamic model Geoffard and Philipson (1997) address the first incentive, but remain silent about the second incentive. We provide the missing part of the analysis using a static model. On the demand side, we consider the case where the population has to pay for the vaccinations, i.e. the costs are not covered by health insurance companies or the state. Consequently, there is no bargaining either between

[2] "The four industry leaders (Merck, GlaxoSmithKline, Aventis Pasteur, and Wyeth) are estimated to spend more than US$750 million a year on vaccine R&D—as much as a fivefold jump at some companies since 1992." (BusinessWeek Online (2002))

[3] Although not in a monopoly context, the case of measles offers some insights: more than 99 percent of the disease burden of measles fall on low and middle income countries, with more than 770,000 deaths in the year 2000. Full immunization could save roughly 28 million disability adjusted life years (see Kremer (2002, pp. 70-71)).

insurance companies or state agencies.[4]

We summarize both the income dependence of the individual willingness to pay and the external effect of a reduced infection probability due to a higher number of vaccinated individuals using a simple linear aggregate demand schedule faced by the monopolist. The linearity assumption allows explicit results to be obtained, but none of the results depend on it qualitatively. The decisive element in this setting is the monopolist's second strategic incentive mentioned above. This is most easily analyzed in a static environment. But the results will also apply in a dynamic framework, since the importance of the external effect increases.[5] Although the emphasis of the analysis is on the case without price discrimination we also consider perfect price discrimination. All qualitative results, including the comparative statics, are robust. With price discrimination, vaccination discrimination is in fact reduced. But, in contrast to the standard model without external effects, the outcome may still be inefficient. The findings of the robustness of the strategic incentives are relevant for policy recommendations since multi-tier pricing is pervasive in real world vaccine markets (Russell (2002)).[6]

In the theory of public goods, the problem of under-provision can be eliminated by Pigouvian subsidies. However, although vaccinations are an example of privately provided public goods, subsidies do not work very well. At first demand increases as the individual price is reduced. However, this increase lowers the infection probability and thus reduces the willingness to pay. This counteracting effect limits the effect of these subsidies (see Geoffard and Philipson (1997, p. 225)). We show that subsidies may make things even worse. We assume that the price subsidy is financed by lump-sum taxation creating a negative income effect. If this effect is sufficiently large, the positive price effect is overcompensated and a smaller proportion of people are vaccinated. This contrasts with the classical regulation arguments for Pigouvian subsidies and strengthens Philipson's (2000) argument, that "Pigouvian subsidies traditionally seen as resolving the under-provision problem of vaccines can be short-run, or out of steady state, arguments" (p. 1777), since these may even fail in static settings. Recently, Philipson

[4]For an empirical analysis that tests whether price discrimination or bargaining is present in the U.S. vaccine market see Kauf (1999).

[5]Francis (1997) showed that in his dynamic setting the externality disappears. The allocation is efficient. But with heterogenous individuals this result does not generally hold.

[6]The UN Accelerating Access Initiative supports differential pricing for AIDS drugs (see http://www.unaids.org/acc_access/). In this context Roche was more or less forced to increase the discount on their AIDS drugs for developing countries to roughly 90 percent of the Swiss price (see Médecins Sans Frontières (2003)).

and Mechoulan (2003) have argued that subsidies are likely to distort R&D incentives.

Another public policy usually suggested is mandatory vaccination. If mandatory vaccination programs do not cover the whole population, the people vaccinated lower the probability that the susceptible will be infected. The willingness to pay is reduced, i.e., mandatory demand crowds out voluntary demand. This is a standard argument for why it is difficult to eradicate a disease by mandatory vaccination if not the entire population is included in the program (see Philipson (2000, p. 1781)). However, such a program is much more effective with income-dependant demand: as people's incomes differ, the public program can cover the poor and the monopolist the rich. Of course the willingness to pay of the rich is reduced, but it remains relatively high due to the income effect. Full vaccination can be achieved with a mandatory participation rate that is strictly smaller than one. Thus, our analysis provides an efficiency argument for public health vaccination programs that focus on the poor like those typically supported by the World Health Organization (WHO) and the Worldbank.

The approach presented here is related to Brito et al. (1991). They consider a static model with a continuum of individuals whose disutility from vaccination differs. Since vaccines are provided free of charge, price discrimination cannot be studied in their setting. The first-best outcome can be implemented by subsidizing those who decide to vaccinate, or by taxing those without immunization. But when the subsidy has to be financed through taxation, the first-best can only be attained under the strong assumption of identical marginal utility of income across individuals. In their dynamic model, Geoffard and Philipson (1997) address the question of disease eradication. Both price subsidies and mandatory vaccination programs have limited impact, since the positive effects of the respective policies are partly offset by the negative effect of the externality.

This chapter is also related to the literature on network externalities, e.g. Bensaid and Lesne (1996), Cabral et al. (1999), and Mason (2000). The main difference is the sign of the network effect. This is positive in these models but negative in ours, leading to completely different results. With a positive network effect, introductory pricing may occur to built up a certain critical network size. With vaccinations it is the other way round: in order to prevent the market shrinking or disappearing a critical mass will never be exceeded.

This chapter is organized as follows: in Section 2.2 we present the main ingredients of our model. The monopolist's price setting problem and the comparative static properties of this solution are studied in Section 2.3.

Perfect price discrimination is analyzed in Section 2.4. In Section 2.5 we discuss public policies that may be used to reduce discrimination and thus increase social welfare. Section 2.6 concludes. The Appendix provides a generalization of the reduced form applied throughout the chapter.

2.2 The model

Consider a population of mass one with individuals who differ in income but are otherwise homogenous. Income is denoted a and is continuously distributed on the interval $[a_L, a_H]$, where $0 < a_L \leq a_H$. An individual's willingness to pay for vaccination depends on her income a and the expected share of individuals who get vaccination, $\theta^e \in [0,1]$:

$$p = p(\theta^e, a). \tag{2.1}$$

The higher the expected rate of immunization θ^e, the lower the expected share of susceptible individuals $1-\theta^e$. A high θ^e is associated with a low expected risk of infection π^e, $\partial\pi^e/\partial\theta^e < 0$. Clearly, the willingness to pay for vaccination increases in the risk of infection. We thus postulate $\partial p/\partial\theta^e < 0$, which captures the external effect of vaccinations. Furthermore, in line with the empirical evidence, it is assumed that the willingness to pay increases in income $\partial p/\partial a > 0$. While the external effect of vaccinations is a general feature of the market, a positive income effect is not so obvious. In the Appendix (see Section 2.7, interpreting vaccination as an insurance decision, we derive a sufficient condition on preferences for yielding a positive income effect. Finally, we assume $p(1,a) > 0$, implying the existence of an exogenous infection risk. While made for simplicity, this can be justified by infection threats from other countries[7], accidental laboratory outbreaks, or terrorist attacks[8].

To simplify the analysis, and in order to derive explicit closed form solutions, we summarize the individual willingness to pay by the following simple linear scheme

$$p(\theta^e, a) = z_\theta(1-\theta^e) + z_a a, \tag{2.2}$$

where $z_a \in (0,1)$ measures the income effect and $z_\theta > 0$ the importance of the external effect. The upper bound on z_a is justified by normality, while

[7]The way infectious diseases can spread around the world was recently demonstrated with the Severe Acute Respiratory Syndrome (SARS) that originated in China.

[8]Although smallpox is said to be eradicated, there is a positive willingness to pay for vaccines.

the lower bound reflects our central assumption of a positive income effect. Furthermore we assume that the population is uniformly distributed on the interval $[a_L, a_H]$. None of these assumptions is necessary for the results we derive below. However, their use significantly eases the presentation of the main ideas. As will become clear, a downward sloping aggregate demand function is sufficient for most results. We discuss the conditions under which demand is downward sloping in the appendix.

There is a monopolist who provides a vaccine that yields perfect protection against the disease and that has no side effects. His price setting problem is analyzed in a two-stage game. At stage 1 the monopolist sets the price p_m. We analyze two versions of the game, in Section 2.3 we study standard monopoly pricing. Here p_m is constant and denotes the price at which the monopolist is willing to sell to all consumers actually demanding vaccination. In Section 2.4 the case of perfect price discrimination is addressed. The price may depend on income so that $p_m = p_m(a)$ is a price schedule. At stage 2 individuals observe prices, form expectations about the vaccination rate, and thus about the infection probability, and decide whether to vaccinate or not, i.e. aggregate demand is realized.

Solving the game backwards leads to a subgame perfect Nash equilibrium. Deriving aggregate demand requires first analyzing the role of consumers' expectations for vaccination decisions. As the analyses differ for the two cases studied they are relegated to the respective sections of the chapter (see Lemma 2.1 and Lemma 2.3 below). Once aggregate demand is derived, determining the monopolist's optimal policy is straightforward.

2.3 Monopoly pricing

Let us first consider the case of standard monopoly pricing where the monopolist only quotes a single price. In order to derive the stage two vaccination equilibrium we assume symmetric expectations and require expectations to be consistent, i.e. expectations must be fulfilled in equilibrium. The second useful property is the sorting of individuals by income. In particular, for a given expected infection risk, an individual who decides to be vaccinated knows that everybody richer than herself will also be vaccinated.

Before we solve for the equilibrium of the vaccination subgame, we define the critical consumer $\overline{\theta}$. Let $p_m > 0$ be some fixed price for the vaccine, then $\overline{\theta} = \overline{\theta}(p_m)$ solves

$$p_m = p\left(\overline{\theta}, a\left(\overline{\theta}\right)\right) = z_\theta \left(1 - \overline{\theta}\right) + z_a \left(a_H - \overline{\theta}\Delta\right), \tag{2.3}$$

where $\Delta := a_H - a_L$. The income of the critical consumer is $\overline{a} := a(\overline{\theta}) = a_H - \overline{\theta}\Delta$. Since the willingness to pay $p(\overline{\theta}, a(\overline{\theta}))$ strictly decreases in $\overline{\theta}$, the critical consumer is well-defined, i.e. $\overline{\theta}$ is unique.

In Lemma 2.1 we show that there are unique expectations for every given price p_m. How consistency of expectations can be used to derive the aggregate demand is demonstrated in Lemma 2.2.

Lemma 2.1 *Individuals facing a price p_m will rationally expect $\overline{\theta}(p_m)$ to be the immunization rate.*

Proof. The proof is by contradiction. So, suppose that individuals expect the immunization rate $\theta^e > \overline{\theta}$. Then, the willingness to pay for vaccination of type $\overline{\theta}$ is $p(\theta^e, a(\overline{\theta})) = z_\theta(1-\theta^e) + z_a(a_H - \overline{\theta}\Delta) < p_m$. Individual $\overline{\theta}$ will not demand vaccination and neither will all consumers with lower income than $a_H - \overline{\theta}\Delta$. Thus immunization with expectations $\theta^e > \overline{\theta}$ will actually be lower than $\overline{\theta}$ so that expectations can never be confirmed. A similar reasoning applies to all $\theta^e < \overline{\theta}$ proving inconsistency of all $\theta^e \neq \overline{\theta}$. ■

The lemma implies that we can concentrate on cases where the two arguments of the willingness to pay function are identical. To ease notation we will thus write $p(\theta) := p(\theta, a(\theta))$. Notice that we also omit the bar.

Lemma 2.2 *The aggregate demand function the monopolist is facing at the first stage of the game is given by*

$$\theta(p_m) = \frac{z_\theta + z_a a_H - p_m}{z_\theta + z_a\Delta}. \tag{2.4}$$

Proof. Since the vaccination equilibrium at price p_m is fully characterized by $\overline{\theta}(p_m)$, deriving aggregate demand simply requires solving equation (2.3) for $\overline{\theta}$. ■

Given these lemmata, the game boils down to a game of complete information. By setting the price, the monopolist can directly influence expectations about the immunization rate and thus exploit the external effect associated with vaccinations. Alternatively the monopolist's policy may be derived by optimization with respect to p or θ. For notational convenience we stick to the latter yielding the following objective function

$$\Pi(\theta) = p(\theta)\theta = (z_\theta(1-\theta) + z_a(a_H - \theta\Delta))\theta. \tag{2.5}$$

We consider constant marginal costs of zero, implying disease eradication, i.e. $\theta = 1$, as being socially optimal. The first order condition is derived by differentiation yielding[9]

$$\theta^* = \frac{z_\theta + z_a a_H}{2(z_\theta + z_a \Delta)}. \tag{2.6}$$

Without the externality, the monopolist would face the inverse demand schedule $p(\theta) = z_\theta + z_a(a_H + \theta\Delta)$ yielding an optimal supply of $\frac{z_\theta + z_a a_H}{2 z_a \Delta} > \theta^*$. With the externality, the monopolist has an incentive to reduce supply in order to increase the willingness to pay and thus profit. The externality reduces the elasticity of demand and thereby amplifies monopoly power. This interpretation demonstrates that this result is very general. It holds as long as aggregate demand is downward sloping.[10] The price corresponding to θ^* is

$$p_m^* = \frac{z_\theta}{2} + z_a a_H. \tag{2.7}$$

The price increases in all exogenous parameters except a_L. The comparative static properties of θ^* are much more informative. First, note that $\partial\theta^*/\partial z_a > 0$. With an increasing income effect, the relative importance of the external effect of vaccinations is reduced and with it the incentive to cut the supply. More interesting is the effect of a change in the external effect parameter z_θ which is clearly negative, i.e. $\partial\theta^*/\partial z_\theta < 0$. The higher the external effect of susceptible individuals on the willingness to pay, the higher the monopolists' incentive to exploit this effect, i.e. to reduce the amount of vaccines sold.

To study the effect of income inequality on equilibrium let $a_H = a + \Delta/2$ and $a_L = a - \Delta/2$. Then $\theta^* = \frac{z_\theta + z_a(a+\Delta/2)}{2(z_\theta + z_a\Delta)}$. The income inequality effect is observed by differentiation with respect to Δ yielding $\partial\theta^*/\partial\Delta < 0$. The more unequally the income is distributed among the population, the more severe the problem of vaccination discrimination. Note that the average income a is not affected by changes in Δ. Now consider that the population as a whole becomes richer, but (absolute) inequality remains unchanged: $\partial\theta^*/\partial a > 0$. The income effect becomes more important relative to the discrimination effect. Consequently, a higher share of the population decides to vaccinate. To summarize, vaccination discrimination is more likely to occur in societies that are poor or face substantial income inequality.

[9] To avoid θ^* exceeding one, it is assumed that $z_\theta + z_a(\Delta - a_L) \geq 0$.

[10] See the appendix for a condition for preferences that yield a downward sloping demand.

2.4 Perfect price discrimination

In this section we consider the case where the monopolist can observe individual income. Without externality, this induces efficiency and enables him to obtain the entire rent. Although this represents a benchmark case, it nevertheless deserves particular attention because multi-tier pricing is pervasive in vaccine markets (Russell (2002)). This observation holds for national markets but even more so at the international level, where developing countries receive vaccines at significantly lower prices than developed countries. Of course, the mechanisms of our model are also valid in such an international context, if an international link exists between the infection probabilities.[11]

Again, there is a two-stage game. Analyzing the impact of expectations on demand is little more involved with perfect price discrimination because demand has to be determined for every possible price schedule that may be offered. The following lemma demonstrates that there is a unique expectation for every relevant price schedule.

Lemma 2.3 *The monopolist offers a price schedule of the following type:*

$$p_m\left(\theta;\overline{\theta}\right)=\begin{cases} z_\theta\left(1-\overline{\theta}\right)+z_a\left(a_H-\theta\Delta\right) & for \quad \theta\leq\overline{\theta} \\ \infty & for \quad \theta>\overline{\theta} \end{cases}. \tag{2.8}$$

Given this schedule, individuals rationally expect $\overline{\theta}$ to be the immunization rate.

Proof. Before addressing expectations we have to show that it is sufficient to analyze price schedules like those mentioned above. First, note that the schedule will have a cut-off value of income $\overline{a}\in[a_L,a_H]$ such that all individuals with incomes higher than $\overline{a}$ will demand vaccination and those with lower incomes will not. If this were not the case, the monopolist would benefit from reallocating vaccinations. At $\overline{a}$ the monopolist will demand the entire willingness to pay $p\left(\theta^e,\overline{a}\right)$. For incomes exceeding $\overline{a}$ he simply adds the income effect $z_a\left(a-\overline{a}\right)$ and again obtains the entire rent. Individuals whose incomes fall short of $\overline{a}$ receive no offer. Using the relationship $a=a_H-\theta\Delta$, the relevant price schedules may be written as in equation (2.8).

Given this schedule, the (symmetric) expectation θ^e can never exceed $\overline{\theta}$ but may be lower. So, consider that $\theta^e<\overline{\theta}$. Then the willingness to pay of

[11]The recent outbreak of SARS in China and its spread to Europe and, in particular, North America dramatically demonstrates the correlation between infection risks.

type θ is $p(\theta^e, \theta) = z_\theta (1 - \theta^e) + z_a (a_H - \theta\Delta) > z_\theta (1 - \overline{\theta}) + z_a (a_H - \theta\Delta) = p_m (\theta; \overline{\theta})$. Since this holds for all $\theta \leq \overline{\theta}$, all individuals with income higher than $a_H - \overline{\theta}\Delta$ will vaccinate contradicting $\theta^e < \overline{\theta}$. Thus, $\theta^e = \overline{\theta}$ results. ■

By offering a price schedule as shown in equation (2.8), the monopolist sets not only prices but also the quantity offered in the market. As $\overline{\theta}$ is unique, optimization over $\overline{\theta}$ completes the analysis. The objective function is given by

$$\Pi(\overline{\theta}) = z_\theta(1 - \overline{\theta})\overline{\theta} + z_a \int_0^{\overline{\theta}} (a_H - \theta\Delta) d\theta. \tag{2.9}$$

Optimization with respect to $\overline{\theta}$ yields

$$\theta^{ppd} = \frac{z_\theta + z_a a_H}{2z_\theta + z_a \Delta}. \tag{2.10}$$

Qualitatively, the comparative static properties of θ^{ppd} compare with those of θ^*. Comparing equations (2.6) and (2.10) unambiguously reveals $\theta^{ppd} > \theta^*$. The monopolist's ability to demand prices that are conditional on the willingness to pay reduces the prevalence of the disease. This is well in line with standard monopoly theory. However, the following proposition contrasts with it, highlighting the peculiarity of the vaccine market, namely, the external effect.

Proposition 2.1 *For a sufficiently strong external effect, $z_\theta > z_a a_L$, a perfect price discriminating monopolist is socially inefficient, i.e. $\theta^{ppd} < 1$.*

As the proof is straightforward, we only provide the intuition and a graphical illustration: with negative network effects increasing demand not only reduces the price for the marginal consumer, it also reduces the willingness to pay of all other consumers. Consider Figure 2.1. Instead of serving θ^{ppd} suppose that the monopolist covers the entire market. The additional rent he gains is given by area A. But, as the higher share of vaccinated individuals reduces the willingness to pay, all individual prices are reduced. He loses rent that amounts to area B. With full immunization the monopolist is clearly worse off.

The results of this section may simply be summarized by $\theta^* < \theta^{ppd} \leq 1$. Both inequalities require some further discussion. First, allowing the monopolist to perfectly price discriminate improves access to vaccines and is thus socially desirable. In an international context price discrimination is the rule rather than the exception. However, when providing vaccines to developing countries at lower rates than to developed countries

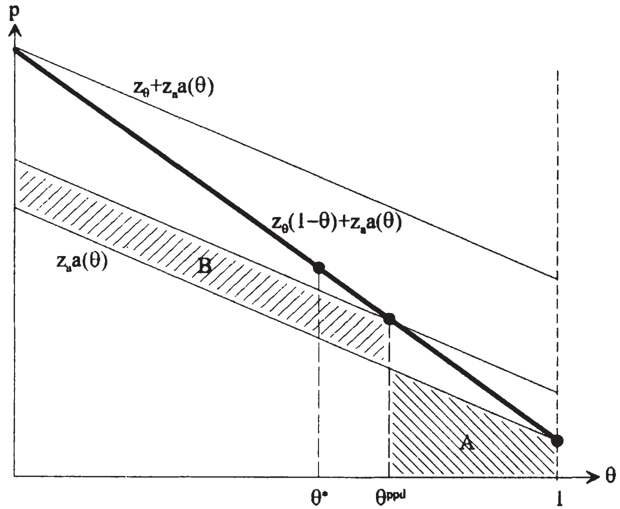

Figure 2.1: The case of perfect price discrimination with vaccination externality.

the monopolist risks undermining prices in developed countries, e.g. by re-imports. This may force the monopolist back to the uniform monopoly price. But developed countries have an incentive to prohibit re-imports since higher monopoly profits in the case of perfect price discrimination facilitate stronger R&D incentives in a dynamic framework (see Kremer (2002, pp. 76-77)). Of course, the same argument holds for arbitrage prevention if price discrimination is employed at the national level. Second, although perfect price discrimination yields higher welfare than standard monopoly, public health intervention may still be necessary to correct for the externality.

2.5 Public policy

Obviously, within our setting of zero marginal cost, the socially optimal policy would be to have the monopolist cover the whole market. We now discuss the consequences of two standard public health interventions, namely price subsidies and mandatory vaccination programs. To evaluate their benefits, we analyze their potential to increase the degree of immunization in society.

2.5.1 Price subsidies

Consider a policy of paying the monopolist a per unit subsidy of size $s > 0$. This is usually a standard tool for alleviating the inefficiency caused by monopoly and a Pigouvian cure for the vaccination externality. Unfortunately subsidies do not work very well in the market for vaccines since the positive effect of the subsidy is opposed by the prevalence effect: the increased demand due to the subsidy reduces the prevalence of the disease and thereby the willingness to pay for vaccination. Things get worse when the subsidy is to be financed by taxation. Consider, for example, a head tax of size T. The income distribution in this case shifts to $[a_L - T, a_H - T]$ creating a negative income effect. The government budget constraint is given by $T = \theta s$. Thus the monopolist now faces a willingness to pay of

$$p(\theta; s) = z_\theta(1-\theta) + z_a\left(a_H - \theta s - \theta\Delta\right). \tag{2.11}$$

He actually receives $p(\theta; s) + s$. If there is a positive subsidy, the monopolist chooses the price such that

$$\theta^s = \frac{z_\theta + z_a a_H + s}{2(z_\theta + z_a s + z_a\Delta)}. \tag{2.12}$$

Proposition 2.2 *There exists a critical income effect* $z_a^{crit} \in (0,1)$ *such that* $\theta^s < (=,>)\ \theta^*$ *if* $z_a > (=,<)\ z_a^{crit}$.

Proof. In principle $\theta^s = \theta^*$ could be solved for z_a. Since the actual size of z_a^{crit} is of minor interest, we study two benchmark cases and apply a continuity argument.[12] To decide on the effectiveness of the subsidy we have to compare θ^s with the laissez-faire share θ^* of equation (2.6):

$$\theta^s - \theta^* = \frac{s}{2}\frac{z_\theta\left(1-z_a\right) + z_a\Delta - z_a^2 a_H}{\left(z_\theta + z_a s + z_a\Delta\right)\left(z_\theta + z_a\Delta\right)}. \tag{2.13}$$

Since the denominator of the right hand side of equation (2.13) is always positive, the sign of $\theta^s - \theta^*$ is determined by the numerator: if $z_a = 0$ the numerator turns out to be $z_\theta > 0$. The problem of discrimination is reduced by Pigouvian subsidies, since there is no income effect to offset the positive effect of the subsidy. This coincides with the result when a public budget

[12] $z_a^{crit} = \frac{1}{2a_H}\left(\Delta - z_\theta + \sqrt{z_\theta^2 + 2z_\theta\left(a_H + a_L\right) + \Delta^2}\right)$.

constraint is not considered. More interesting, if $z_a = 1$, then $\theta^s < \theta^*$. Discrimination is further increased by subsidizing vaccines. This is due to the income effect caused by financing the subsidy. By continuity, there exists some value $z_a^{crit} \in (0,1)$ such that the subsidies have no effect. In this case, the positive price effect of the subsidy on demand is exactly offset by the two negative effects, the prevalence effect and the financing effect. If the income effect is sufficiently large, i.e. $z_a > z_a^{crit}$, a subsidy makes things even worse. ■

This is in contrast with the classical regulatory arguments, where Pigouvian subsidies are applied to correct for the inefficiencies due to the externality. This extreme effect was not previously known in the theory of vaccinations. Philipson (2000, p. 1777), for example, states that subsidies are limited in their impact in dynamic settings, but that they may have an effect in the short-run or out of steady state. In our static setting, we have shown—strengthening this result—that subsidies may have no effect or may even have a negative effect. In a related paper, Philipson and Mechoulan (2003) point to another pitfall of Pigouvian subsidies, namely, the distortion of R&D incentives.

2.5.2 Mandatory vaccination

Another public policy usually suggested is mandatory vaccination. If mandatory vaccination programs do not cover the whole population, the individuals vaccinated reduce the infection probability of the susceptible. The willingness to pay is reduced, i.e. mandatory demand crowds out voluntary demand. This is a standard argument for why it is difficult to eradicate a disease by mandatory vaccination if not the entire population is included in the program (see e.g. Philipson (2000, p. 1781)). This argument also applies to the model presented here if the social planner has no information about individual income levels. But consider that income is observable, then a program is much more effective than usual. Since the individuals differ in income the public program may only cover the poor and the monopolist the rich.[13]

Let $m \in [0,1]$ be the share of mandatory vaccinated individuals. Consider that these are the 100 times m percent poorest in the society. The

[13] We assume that the social planner can observe income, while the monopolist cannot, or does not, use this information. As mentioned above, perfect price discrimination yields the same results qualitatively. Since vaccination discrimination is lower, the optimal program with perfect price discrimination will be smaller in size, but may still be necessary.

willingness to pay (of the rich) is now given by

$$p(\theta; m) = z_\theta(1 - \theta - m) + z_a(a_H - \theta\Delta). \tag{2.14}$$

The optimum is obtained by differentiation with respect to θ and is attained if

$$\theta^m = \frac{(1-m)\, z_\theta + z_a a_H}{2(z_\theta + z_a\Delta)}. \tag{2.15}$$

The overall share of vaccinated individuals is given by $\min\{1, m + \theta^m\}$. If $m+\theta^m < 1$, then the effect of extending the mandatory vaccination program on the share of vaccinated individuals is clearly positive: $d(m + \theta^m)/dm > 0$.

Proposition 2.3 *With mandatory vaccination programs full immunization is achieved at participation rates strictly smaller than 1.*

Proof. Solving $m + \theta^m = 1$ for m yields

$$\overline{m} = \frac{z_\theta + z_a\Delta - z_a a_L}{z_\theta + 2z_a\Delta}. \tag{2.16}$$

The equivalence $\overline{m} < 1 \Leftrightarrow z_a a_H > 0$ proves the assertion. ■

Mandatory vaccination is more effective than in other models, e.g. Geoffard and Philipson (1997), since the negative effect of the externality is reduced by the still high income effect. As long as the income effect is positive, a residual demand $\theta^m > 0$ served by the monopolist remains. Note that the government's information about individual income heterogeneity enables it to counter the strategic pricing behavior of the monopolist originating precisely from such differences among individuals. The proposition provides an efficiency based argument for distribution-oriented public health vaccination programs like those typically supported by the WHO or the Worldbank.

In line with our comparative static results on income inequality, the share of the population to be included in the program is higher, the higher income inequality. If a society has a high amount of inequality, it is accompanied by a serious amount of vaccination discrimination. Thus, the mandatory vaccination program must be relatively large for full immunization. Of course, if income inequality is relatively low, there is no need for a public vaccination program, since the monopolist already serves the entire market.

2.6 Conclusion

We presented a simple static model to study the effects of monopoly power on the supply side in the market for vaccines. We highlighted the importance of income inequality when analyzing the monopolist's incentive to exploit the external effect of vaccinations to maximize profits.

In the monopoly solution the poor are discriminated against, i.e., remain susceptible, in order to increase the willingness to pay of the rich. Interpreting individuals as countries, the developing countries are strategically left without immunization. Discrimination was found to be more severe if the prevalence elasticity of demand is high, i.e., when the income effect is low or the impact of the external effect is high. Societies with low average wealth or high income inequality are left with a high share of susceptible individuals. With perfect price discrimination the prevalence of the disease will be lower. But, in contrast to the standard monopoly model without external effects, the outcome may still be inefficient.

If the social planner is not informed about the individual income levels, he is left with two policy alternatives—Pigouvian subsidies and mandatory vaccination. Unfortunately, subsidies are of limited use since the positive price effect is opposed by two negative effects, the prevalence effect and the income effect. In some cases subsidies make things even worse, raising doubts about whether Pigouvian subsidies are appropriate at all. As usual, mandatory vaccination programs fail to eradicate the disease if not the entire population is included in the program. Things change dramatically when the social planner is informed about individual income. The public health interventions may then be conditional on income. A mandatory program covering the poor only yields full vaccination at participation rates that are strictly smaller than one. Thus, when income is observable, mandatory vaccination programs strictly dominate Pigouvian subsidies. This provides an efficiency-based argument for public health vaccination programs directed towards the poor or, within an international context, towards poor countries, like the ones advocated and carried out by the Worldbank or the WHO.

2.7 Appendix

Here we provide a micro foundation for our reduced form approach. We argue that a vaccination can be seen as an insurance against the disease. Using this, we will derive a sufficient condition for the willingness to pay being increasing in income. Furthermore we show that under this condition a unique equilibrium always exists.

Vaccination as insurance. Consider an individual with an original income of $a > 0$, which reflects individual productivity or wage-earning abilities, and preferences which obey the von Neumann-Morgenstern axioms. The individual is exposed to the threat of becoming infected with a transmittable disease. The probability of infection is given by $\pi \in (0, 1]$, which is taken to be exogenous to the individual.[14] The monetary loss from infection depends on income, $\beta = \beta(a)$. It is sensible to assume that $\beta > 0$ and $\beta' \in [0, 1]$, since illness will lead to absence from work for a certain time. Hence, a high income individual will lose (weakly) more than a low income individual. A vaccine is available that yields perfect protection against the disease and has no side effects. The price for being vaccinated is denoted p. Then, the utility for a vaccinated individual with income a is given by

$$u = u(a - p). \tag{2.17}$$

The individual decides to vaccinate, if, and only if, the utility u exceeds the expected utility Eu of remaining without protection, where

$$Eu = \pi u(a - \beta(a)) + (1 - \pi)u(a). \tag{2.18}$$

The decision to vaccinate amounts to the choice between the certain outcome and the original risky outcome. Thus, the willingness to pay for vaccinations $p(\pi, a)$ equals the sum of two components, the increase in expected income, $\pi\beta$, and the risk premium. Applying the approximation formula for the risk premium derived by Arrow and Pratt (see Pratt (1964)), we have $p(\pi, a) \approx \pi\beta - \frac{u''(EX)}{u'(EX)}\frac{Var(X)}{2}$, where $EX = a - \pi\beta$ and $Var(X) = \pi(1 - \pi)\beta^2$. Then, for a given infection probability, the willingness to pay globally increases in income if for all $a > 0$ the following condition holds:

$$\frac{u'''}{u''} - \frac{u''}{u'} > 2\beta' \frac{u'/(u''(1 - \pi)) - \beta}{(1 - \pi\beta')\beta^2}. \tag{2.19}$$

The numerator of the right hand side is always negative while the denominator is positive. In the case of constant absolute risk aversion, the left hand side is zero, implying a strictly increasing willingness to pay for

[14] Of course, in equilibrium, this probability will depend on the number of susceptible individuals. This issue is addressed below. Like above, π can be interpreted as the (symmetric) expectation about the infection risk.

vaccinations if $\beta' > 0$. If the utility function exhibits constant relative risk aversion, β' must be sufficiently large for the willingness to pay to be non-decreasing in income.

In the following we will assume that condition (2.19) holds. Furthermore we assume sorting of individuals, $\partial\theta/\partial a < 0$, and, reflecting the vaccination externality, $\partial\pi/\partial\theta < 0$. The following two paragraphs compare to the treatment in Section 2.3 (see e.g. Lemma 2.1).

Expectations. Consider that the price p is exogenous. Moreover assume that $(\overline{\pi}, \overline{a})$ solves $u(\overline{a} - p) - Eu(\overline{\pi}, \overline{a}) = 0$, where $\overline{\pi} = \pi(\theta(\overline{a}))$. Then expectations will be such that $\pi^e = \overline{\pi}$, i.e. $(\overline{\pi}, \overline{a})$ is an equilibrium of the second stage game.

Proof: We have $u(\overline{a} - p) - Eu(\overline{\pi}, \overline{a}) = 0$, where $\overline{\pi} = \pi(\theta(\overline{a}))$. Suppose that $\pi^e > \overline{\pi}$. Then $u(\overline{a} - p) - Eu(\pi^e, \overline{a}) > 0$. Given the price p, the former indifferent individual with income $\overline{a}$ now obtains a positive rent. Since the willingness to pay is increasing in income there exists an income level $\widehat{a} < \overline{a}$ with $u(\widehat{a} - p) - Eu(\pi^e, \widehat{a}) = 0$. It follows that $\widehat{\theta} > \overline{\theta}$ and with it $\widehat{\pi} < \overline{\pi}$. Thus expectations with $\pi^e > \overline{\pi}$ can never be confirmed. A similar argument applies when $\pi^e < \overline{\pi}$ proving $\pi^e = \overline{\pi}$.

Uniqueness. Consider that the price p is exogenous. Assume that $u(\overline{a} - p) = Eu(\overline{a}, \pi(\theta(\overline{a})))$. Then $\overline{a}$ is unique.

Proof: Consider an income $\widehat{a} > \overline{a}$. We know from above that $u(\widehat{a} - p) - Eu(\widehat{a}, \pi(\theta(\overline{a}))) > 0$. Since $\widehat{a} > \overline{a} \Leftrightarrow \widehat{\theta} < \overline{\theta} \Leftrightarrow \widehat{\pi} > \overline{\pi} \Leftrightarrow Eu(\widehat{a}, \pi(\theta(\widehat{a}))) < Eu(\widehat{a}, \pi(\theta(\overline{a})))$: $u(\widehat{a} - p) - Eu(\widehat{a}, \pi(\theta(\widehat{a}))) > u(\widehat{a} - p) - Eu(\widehat{a}, \pi(\theta(\overline{a}))) > 0$. Thus no income $\widehat{a} > \overline{a}$ can be a solution of our problem. As the same applies to every $\widehat{a} < \overline{a}$, $\overline{a}$ is unique.

Consequences. From our analysis above we know that those who demand vaccination are, as long as condition (2.19) holds, the rich. Of course, it is also possible to construct the theoretical case where the willingness to pay decreases with income, e.g. if $\beta' = 0$ and constant relative risk aversion is assumed. However, this is contradicted by empirical evidence. A direct implication of the increasing willingness to pay is uniqueness. This is different from models with positive externalities where multiple equilibria may occur. Having these results it is straightforward to generalize our reduced form to some function $p(\theta, a)$ with $\partial p/\partial\theta < 0$ and $\partial p/\partial a > 0$ implying a downward sloping aggregate demand schedule.

Chapter 3

Price Regulation, Physician Density and the Quality of Care

3.1 Motivation

In many countries the suppliers of health care services are compensated for medical treatments according to a fee schedule chosen by some regulating authority.[1] When there is no possibility of competing in prices, suppliers of health care resort to other variables to increase market share. Consider, for instance, competition between physicians. If a patient needs medical assistance two major factors influence his decision for a private medical practice (or private clinic): the quality provided by the physicians and the distance to the medical practices. Hence, the physicians aiming at higher profits will choose quality levels and locations strategically.

There is a mutual dependency between location and quality decisions. Consider two orthopedists A and B, both with an X-ray unit. Assuming that they are closely located, quality competition will then be intensive. This can be interpreted in two ways: first, physician B, for example, may view these locations as a handicap, because they have to compete in a small market. To increase his market share, B can improve his quality (e.g. by buying a computer tomograph). Second, assume that B has the quality advantage described above. Since A cannot react with price cuts to B's advantage he will lose market share. To avoid this, A will also buy a computer tomograph. To reduce quality competition, physicians will locate as far apart as possible. However, quality levels may still remain high

[1] This chapter is a revised version of Nuscheler (2002).

because of their deterrent effect on other physicians and potential entrants. The strategic importance of quality may lead to inefficiently high quality provision.

We study the incentives and consequences of price regulation in a 3-stage non-cooperative game where the providers of health care have the following successive choices: entry, location and quality. The players' decisions are simultaneous at all stages. The sequential structure reflects the different degrees of irreversibility in the strategic decisions. Quality, interpreted as medical machinery or effort spent on treatments, is variable in the short run. Quality choices are strategic given the locations. These cannot be changed in the short run because of institutional barriers and high transaction costs of relocating. For example, in Germany the licences are only valid for small districts. If a physician wants to relocate he has to apply for a new licence. Finally, the entry decision is variable only in the long run because educational requirements are high. Alternatively, the model may be interpreted as competition among private hospitals. Quality in terms of medical machinery and size of the hospital rooms or beds per room are far more flexible than location.

The game is solved by backward induction leading to a subgame perfect Nash equilibrium. The first finding is that the regulator is not able to implement the first-best efficient allocation when price is the only regulatory variable. This is not very surprising since it is in general not possible to achieve two goals (optimal quality level and optimal number of clinics) when only one policy variable is available (price). We consider two versions of price regulation with and without commitment. If the regulator cannot commit herself to a price, the set of possible prices is reduced to time consistent policies. The symmetric non-cooperative equilibrium with price commitment is a second-best optimum with excess capacity in both quality and entry. The second-best price turns out to be not time consistent and thus cannot be an equilibrium in the game without commitment. The time consistent price policy leads to the first-best level of total quality provision, but also to excess entry. Since welfare is lower in the game without commitment, the regulator would like to commit herself to the second-best price instead of applying the time consistent (quality-optimal) price.

Of course, this problem can be solved by introducing a second regulation variable. The number of suppliers can be fixed by licences to the first-best optimal level. But, from the increases in physician density in all of the 15 EU member states from 1989 to 1999 (see EUROSTAT (2001, p. 6)), it would appear that entry is not very much restricted. A "natural" suggestion for overcoming the inefficiencies and the commitment problem is to allow physicians to compete in quality *and* prices at stage 3 of the game. Like the social planner, the physicians trade off the benefits of quality against its costs and use the price to reduce quality competition. Consequently, the (price-) competitive equilibrium coincides with the time consistent regulatory outcome. Introducing price competition has no effect. If the regulator is to be (re-) elected by patients the time consistency problem is more severe. Once physicians have entered the market, patients care only about the net gain they derive from quality provision, leading to lower welfare than in the time consistent regulatory equilibrium.

Our model is related to the wide range of quality competition (or vertical differentiation) and spatial competition (or horizontal differentiation) literature, for example, D'Aspremont et al. (1979), Novshek (1980) and Salop (1979) for horizontal differentiation, Gabszewicz and Thisse (1980) for vertical differentiation and Neven and Thisse (1990) for both. But, due to the peculiarities of most of the health care systems, there is a major difference between these models and ours: the absence of the price as a strategic variable. The idea of price regulation is found in Ma and Burgess (1993). They show, in a vertically differentiated setting, that the introduction of price regulation reduces inefficiencies in quality. Since locations are exogenous, their analysis stops where ours begins. The current chapter is closely related to Economides (1993) and Gravelle (1999). Both consider the same strategic variables and the same sequential structure as those in the model presented here. In their models, location choices have no effect on quality provision thus precluding (non-price) competition in the presence of price regulation. Furthermore, Gravelle (1999) assumes marginal costs to be increasing in quality thereby introducing an adverse effect of quality provision. In contrast, we emphasize the public good property of the quality of medical care by considering zero marginal costs. Economides (1993) focuses on different sequential structures and their impact on equilibria. In contrast to this chapter, Economides (1993) does not consider price regulation. As we do, Economides (1993) concentrates on competitive equilibria where firms are direct competitors. Gravelle (1999) follows Salop (1979) and also analyzes monopoly and kink equilibria. Our model builds on both Economides (1993) and Gravelle (1999) in order to consider the important effects of location choice on quality choice, thereby capturing non-price competition in the health care market. In addition, we explicitly consider, as these and the other papers mentioned do not, problems of time consistency. With regard to Gravelle (1999), who applies a similar model to the health care market, this chapter contributes to the understanding of the relationship between the second-best optimum, the time consistent regulatory outcome and the equilibrium with price competition. Moreover, adding two more aspects to the Gravelle (1999) analysis, we explicitly derive a number of first-best efficient regulatory schemes and provide some empirical evidence.

The chapter is organized as follows: in Section 3.2 we introduce the model. The non-cooperative equilibrium is determined in Section 3.3 followed by the welfare analysis in Section 3.4. Regulation and time consistency is studied in Section 3.5. Price competition is introduced in Section 3.6. In Section 3.7 empirical evidence found in the German health care system is presented. Section 3.8 concludes.

3.2 The model

Consider a population of a city which is uniformly distributed on a circle with circumference 1. Each inhabitant of this city has constant (inelastic) demand for one identical medical treatment. Location of a patient on the circle is denoted by $x \in [0,1]$ and the location of physician i by $l_i \in [0,1]$ for $i = 1, ..., n$. The physicians are indexed such that $0 \leq l_1 \leq l_2 \leq ... \leq l_n \leq 1$. A patient's utility from one medical treatment by physician i is

$$u(x, q_i, l_i; p) = q_i - p - c(x - l_i)^2, \tag{3.1}$$

where $q_i \geq 0$ is the quality of medical practice i. Quality is the medical equipment of an institution and is assumed to be measurable and observable. Alternatively, quality may be interpreted as an index capturing different dimensions of quality. In the spirit of Donabedian (1980, pp. 79-85) these dimensions could be the structure, process, and outcome of care. The regulated fee for the treatment is denoted by p and is identical for all physicians. Notice that p may also be interpreted as the premium of a health insurance with full coverage. Since we have assumed that the individuals have constant demand for one medical treatment, a health insurance is redundant. The third term measures transaction costs. These costs are quadratic in the arc-length distance from x to l_i with $c > 0$.[2] The utility from not consulting a physician is assumed to be $-\infty$. By this assumption, physicians are never local monopolists, their neighbors will always be direct competitors. The monopoly and kink equilibria studied in detail by Salop (1979) and Gravelle (1999) do not exist in the model presented here.[3] For

[2]Both Economides (1993) and Gravelle (1999) consider linear transaction costs. In consequence, marginal benefits of an increase in quality are independent of locations. This precludes an important issue of physician non-price competition.

[3]This seems reasonable when analyzing the health care market since, at least in the developed countries, everybody has access to medical care. This is in contrast to Gravelle

simplicity, we do not allow physicians to compete with physicians who are not their neighbors. This is without loss of generality, because a situation of competition between non-neighboring physicians will never be an equilibrium in an entry game. In the language of Salop (1979), there is always a competitive, but never a super-competitive, region.

The patient who is indifferent between the two neighboring medical practices i and $i+1$ is called the critical patient, x_i^{crit}, and is implicitly given by $u\left(x_i^{crit}, q_i, l_i; p\right) = u\left(x_i^{crit}, q_{i+1}, l_{i+1}; p\right)$:

$$x_i^{crit} = \frac{q_i - q_{i+1}}{2c\left(l_{i+1} - l_i\right)} + \frac{l_i + l_{i+1}}{2}. \tag{3.2}$$

Note that the critical patient is given by this equation if, and only if, physicians' locations do not coincide. It can be shown that identical locations cannot be an equilibrium in pure strategies of the location subgame (see proof of Lemma 3.1 in the appendix). Given the critical patients, the market share of physician i is $M_i = x_i^{crit} - x_{i-1}^{crit}$. The regulator's reimbursement policy is designed as a fee-for-service $p \geq 0$ generating benefits of pM_i. We assume symmetry of the physicians' costs structures and that the cost function $K\left(q_i\right)$ for the provision of quality q_i is quadratic, $K\left(q_i\right) = kq_i^2$, $k > 0$. Notice that these costs are fixed with respect to the number of treated patients. Without loss of generality, other fixed costs are assumed to be zero. Setting marginal costs to zero stresses the local public good character of medical institutions.[4] Physician's i profit function is then given by

$$\Pi_i = pM_i - kq_i^2. \tag{3.3}$$

We do not address the problem of physician induced demand or quality elastic (total) demand in this chapter. Thus, the only incentive to provide high quality is an increase in market share. Since quality levels are only due to quality competition, they can be viewed as a lower bound for more realistic, but more complicated, elastic demand or supplier induced demand models. Regarding equations (3.2) and (3.3), the physicians' profits depend on their distance to their neighbors and their comparative quality advantages. We assume the medical practices to be for-profit institutions. Their objective is to maximize profits with respect to location and quality. This

(1999), where utility of not consulting a physician is normalized to zero. Consider, for instance, influenza or appendicitis for the hospital interpretation. In Gravelle's model the "patient" may prefer to not take pharmaceuticals or not have an appendectomy, i.e. the he may prefer to die.

[4] Economides (1993) assumes the same for private goods. Of course this is equivalent to the assumption of constant marginal costs $MC > 0$. Then p is the net-price: $p = \tilde{p} - MC$.

problem is analyzed in a three-stage non-cooperative game of complete information consisting of:

- Stage 1: On the basis of the expected (or announced) price $p^e \in [0, \infty)$ the potential entrants decide simultaneously whether to enter the industry, and n physicians actually enter.
- Stage 2: The n physicians simultaneously choose their locations $l_i \in [0, 1]$, $i = 1, ..., n$.
- Stage 3: The regulator chooses $p \in [0, \infty)$ and then the n physicians decide simultaneously on their quality levels $q_i \in [0, \infty)$, $i = 1, ..., n$.

The sequential structure of the game is argued by the differences in irreversibility of the strategic decisions. The entry decision is variable only in the long run, because of demanding educational requirements. Institutional barriers and transaction costs prevent locations being variable in the short run. Assuming that the medical practices have plenty of capacity, quality, interpreted as medical machinery, is variable in the short run. This also applies to the price set by the regulator. A theoretical, rather than real life, reasoning for the sequential structure is as follows: if the physicians choose quality and location simultaneously, then, by the same argument as in Gabszewicz and Thisse (1992, pp. 291-292), no equilibrium in pure strategies exists. Novshek (1980) solved this problem considering conjectural variations. With quality decision at stage 2 and location at stage 3, a symmetric equilibrium in pure strategies exists. Since the equilibrium quality provision is zero, quality competition cannot be studied in this setting.

Since the regulator's price decision is at stage three and the entry decisions on the basis of the expected price are at stage one, the question of the regulator's ability to commit herself is crucial for the outcome of the game. We consider two games, one in which the regulator can commit on p at stage 1 and a time consistent choice of p without commitment. For clarification, put it differently: there are two four-stage games. The commitment game involves the following choices: price, entry, location, and quality. The game without commitment involves: entry, location, price, and quality. The second game may also be labelled as 'partial commitment game' since prices are set prior to quality. We will nevertheless refer to the resulting policy as the time consistent regulatory policy as the regulator has ex post no incentive to deviate from that policy. Actually she is ex post indifferent between all prices. However, before addressing the regulator's price setting problem we have to determine the non-cooperative outcome and the first-best efficient allocation of the game.

3.3 The non-cooperative equilibrium

3.3.1 Quality

We solve the game by backward induction leading to a subgame perfect Nash equilibrium. Note that we concentrate on pure strategy equilibria. At the third stage of the game each physician maximizes his profits with respect to quality, taking quality levels of the others and all location choices as given yielding

$$q_i^* = \frac{p}{4ck}\left(\frac{1}{l_{i+1}-l_i}+\frac{1}{l_i-l_{i-1}}\right), \qquad i = 1, ..., n. \tag{3.4}$$

We do not observe reaction functions connecting the quality levels of the medical practices directly.[5] Since symmetry of the cost functions is common knowledge and locations are observable, each physician is able to calculate his rivals' equilibrium quality levels. Thus, the direct quality reaction can be viewed as being hidden in the locations. Besides this indirect reaction to others' quality levels, physician i directly reacts with his quality choice to the location choices of the neighboring practices. Locational disadvantages can be reduced by providing high quality.[6] In this sense, quality and location are aggressive strategic variables. Minimum quality is chosen when physician i locates in the middle of the practices $i-1$ and $i+1$.

The zero slope of the reaction functions hinges on the assumption of separability of quality in the utility function. This is a standard assumption in these kinds of models to keep the analysis tractable. As long as the transaction costs are convex, the quality level depends on the own and on the neighbors' locations. That quality increases in price is fairly robust. When the utility function is given by $f(q) - p - c(x-l)^2$ and the costs of quality provision by $K(q)$, then it is sufficient to assume the following plausible properties: $f' > 0$, $f'' \leq 0$, and $K'' \geq 0$. More quality benefits the patient but at a decreasing rate.[7] Improving quality gets more expensive the higher quality is.

[5]This result may be surprising at a first sight. But it is theoretically appealing since the zero conjectural variation assumption of the Nash equilibrium is consistent with the slope of the reaction functions (see Bresnahan (1981)).

[6]Ma and Burgess (1993) found a similar strategic structure. In their model disadvantages in quality can be reduced by price cuts one stage later. Note that the effect of location choices on quality choices disappears when considering linear transaction costs as in Economides (1993) and Gravelle (1999).

[7]The fact that more quality may harm patients can be precluded as physicians will never invest along such dimensions. Recall that quality was supposed to be observable.

Note that equation (3.4) is not an equilibrium for all possible locations as quality goes to infinity when practices' locations become close. Quality competition may be called 'ruinous'. In Chapter 4 we explicitly deal with that problem by defining an 'existence set', i.e., a set where locations imply qualities that yield at least zero profits (and a pure strategy equilibrium). This approach is also taken in Economides (1984, 1986, 1989), Hinloopen and Marrewijk (1999), and Lambertini (2001). As we study an entry game and let the number of firms be determined by a zero profit condition, the existence set might be a singleton for every provider. This is somewhat inconvenient as a relocation in equilibrium then immediately destroys the pure strategy equilibrium.[8] However, this problem almost disappears when there is a sufficiently large cost associated with entry, e.g., (time consuming) education. As these costs are sunk at the quality stage, firms will make positive profits resulting in an existence set with positive mass. To simplify the analysis we abstract from such fixed costs which is without loss of generality. For the moment, we assume that firms are sufficiently located apart guaranteeing existence of a pure strategy equilibrium. We are more precise on that in the proof of Lemma 3.1.

3.3.2 Location

To obtain the optimal location choices of the practices we have to study the relation between quality levels and locations in more detail. We have to bear in mind that, besides the own quality level, the optimal quality levels of the practices $i-1$ and $i+1$ also depend on the location choice of physician i according to equation (3.4). From the respective first order conditions it is straightforward to derive the following reaction coefficients

$$\frac{\partial q_{i-1}^*}{\partial l_i} = -\frac{p}{4ck}\frac{1}{(l_i - l_{i-1})^2}, \tag{3.5}$$

$$\frac{\partial q_i^*}{\partial l_i} = \frac{p}{4ck}\left(\frac{1}{(l_{i+1} - l_i)^2} - \frac{1}{(l_i - l_{i-1})^2}\right), \tag{3.6}$$

$$\frac{\partial q_{i+1}^*}{\partial l_i} = \frac{p}{4ck}\frac{1}{(l_{i+1} - l_i)^2}. \tag{3.7}$$

Consider the situation shown in Figure 3.1. Suppose practice i decides to locate closer to practice $i+1$ ($dl_i > 0$) starting from a symmetric situation.

[8] Mostly this will not happen in an ϵ-environment since the number of firms in the market is an integer while zero profits may require a non-integer number of firms.

The competition in this market area becomes more intensive. The intensity is further increased by the enhancement of quality of practice i according to equation (3.6). Practice i captures some of the market of practice $i+1$. Since the location choices are simultaneous, and the optimization is with respect to given locations of the others, there is no reaction in location of practice $i+1$. Physician's $i+1$ response to the approaching (and quality improving) practice i is to increase his own practice's quality in the extent of (3.7). Competition between physicians i and $i-1$ is weakened because of the increased distance. Practice $i-1$ can lower its quality following (3.5) without losing much of the market, if any at all. We have $dq^*_{i+1} > dq^*_i > 0 > dq^*_{i-1}$ for $dl_i > 0$ and $dq_{i+j} = 0$ for all $j \notin \{-1, 0, 1\}$. The relocation's impact on quality provision leaves the market share of physician i unchanged. Since the quality level has increased, i's profit is reduced.

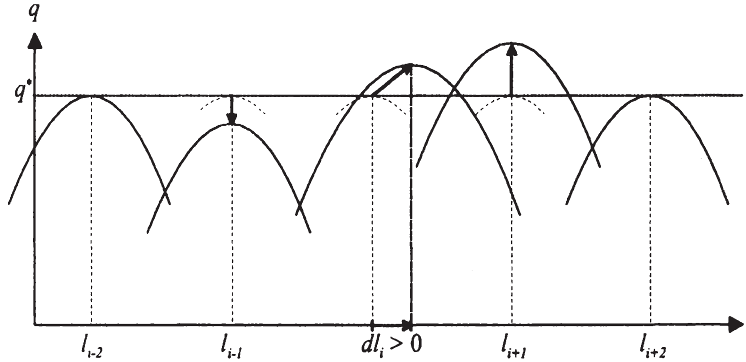

Figure 3.1: Starting from the perfect symmetric arrangement a relocation of physician i closer to $i+1$ ($dl_i > 0$) alters equilibrium quality provision.

Lemma 3.1 *The perfectly symmetric arrangement of the practices is a Nash equilibrium of the location subgame.*

The proof is presented in the appendix. Assume that asymmetric equilibria do exist. If the number of entrants in an asymmetric situation equals the number of physicians in the symmetric case, total quality provision will be higher and some physicians will incur a loss. Therefore we will have a (weakly) smaller number of practices in asymmetric equilibria. Since we want to study inefficiency in quality due to quality competition and not due to asymmetry, we limit ourselves to the symmetric case.

In the symmetric equilibrium the optimal quality provision of equation (3.4) reduces to

$$q_i^* = \frac{np}{2ck}, \qquad i = 1, ..., n. \tag{3.8}$$

In equilibrium the quality increases linearly in the number of entrants and also linearly in price.[9] Quality competition sharpens, the more physicians enter the market and the better paid the medical treatments get. Regarding the first mentioned effect, we see that this model tends to produce strong inefficiencies in total quality. An increase in the transportation cost parameter c leads to a decrease in equilibrium quality. The reason for this is that providing a high quality is less effective than before as patients' responsiveness to quality is reduced by higher transaction costs. The marginal benefits of quality are decreasing when c increases (this is like closing the umbrellas in Figure 3.1 a little). Increases in k directly drive the costs of quality provision up. All other things equal, the quality level has to be reduced to equalize marginal benefits and marginal costs.

3.3.3 Entry

Suppose that the profits earned outside the health care market are zero. Then the number of entrants is implicitly given by the zero profit condition $\frac{p}{n} - k(\frac{np}{2ck})^2 = 0$. The only real solution is

$$n^* = \left(\frac{4c^2k}{p}\right)^{\frac{1}{3}}. \tag{3.9}$$

For simplicity we ignore the fact that n^* may not be integer valued. With a first brief look at equation (3.9), the result seems counter intuitive. Why should the equilibrium number of entrants be lowered by an increase in price? Initially, one would expect the opposite, since the benefits per patient are increasing. But we have seen in equation (3.8) that a higher price increases quality competition among the entrants. This higher intensity in competition is anticipated at the entry stage. Because of the quadratic

[9]When neglecting non-price competition as, for example, in Economides (1993) and Gravelle (1999), quality levels are independent of the number of firms. This seems unreasonable as the number of firms is a measure of competition intensity. Consequently, first-best efficient quality is in terms of individual quality levels in Gravelle (1999) where it is in total quality in our model. Thereby we stress the substitutability between quality and the number of practices, i.e., access to medical care.

quality costs, the higher price has a deterrent effect.[10] The argument for the parameters c and k works in the opposite direction. Quality competition becomes weaker with an increase in transportation costs or quality costs making the health market more attractive. The quality cost effect is weaker since the positive effect is accompanied by a negative cost effect of quality provision. Before studying welfare we summarize the results of this section.

Proposition 3.1 *The symmetric non-cooperative equilibrium of the sequential 3-stage entry-location-quality game is given by:* $q^* = \frac{pn^*}{2ck} = (\frac{p^2}{2ck^2})^{\frac{1}{3}}$, $l_i^* - l_{i-1}^* = \frac{1}{n^*}$, *for all* $i = 1, ..., n^*$ *and* $n^* = (\frac{4c^2k}{p})^{\frac{1}{3}}$.

3.4 The first-best optimum

In the presence of convex costs of quality provision and convex transportation costs, the first-best optimum is symmetric, i.e. the quality levels are equal, $q_i = q$ for all $i = 1, ..., n$, and the practices are arranged symmetrically on the circle. Then the welfare function is given by

$$W = q - knq^2 - \frac{c}{12n^2}. \tag{3.10}$$

The first two terms measure the net welfare gain from quality provision. This is reduced by the expected (or average) transportation costs. When equal welfare weights are considered, the price does not appear in the welfare function, i.e. p is a welfare neutral transfer from patients to physicians.[11] Welfare is indirectly, not directly, affected by changes in p. The first-best efficient outcome is obtained by differentiation with respect to q and n.

Proposition 3.2 *The welfare optimum of the sequential 3-stage entry-location-quality game is attained iff (1) the number of entrants is* $n^{fb} = \frac{2ck}{3}$, *and (2) the entrants are arranged symmetrically, i.e. with distance* $l_i^{fb} - l_{i-1}^{fb} = \frac{1}{n^{fb}}$ *for all* $i = 1, ..., n^{fb}$, *and (3) the quality provided is* $q_i^{fb} = \frac{1}{2n^{fb}k} = \frac{3}{4ck^2}$ *for all* $i = 1, ..., n^{fb}$. *The associated welfare is* $W^{fb} = \frac{3}{16ck^2}$.

[10] This reasoning holds, for instance, for cost functions of the form $K(q) = kq^\tau$, where $k > 0$ and $\tau > 1$. For more complex cost functions the reverse may obtain. If marginal costs of quality provision are very low, then the costs incurred by the tightened competition are low, making the health care market more attractive. This also applies when marginal costs increase rapidly as the prohibitive costs prevent intensive quality competition occurring.

[11] This is why the optimal policy of the 'partial commitment game' is time consistent.

Note that total quality provision is independent of n and c and is equal to $\frac{1}{2k}$. Obviously, the higher quality costs are, the lower the welfare maximizing level of total quality. Increasing transportation costs drives the welfare maximizing number of firms up without changing total quality provision, i.e. physicians' quality and physician density are substitutes.

3.5 Price regulation and time consistency

In Section 3.3 we solved the 3-stage game with the flat-rate p seen as exogenous or as the expected (equilibrium) fee. We now address the regulator's price setting problem. Her goal is to set the "welfare maximizing" fee such that the non-cooperative solution stated in Proposition 3.1 approaches the first-best optimum stated in Proposition 3.2.

3.5.1 Two benchmarks and the first-best optimum

Given that the fee influences the quality decisions of the physicians as well as the number of entries, we cannot expect to be able to implement the first-best optimum by just setting the welfare maximizing price. We have two regulation goals (the first is to fix the number of firms to $n^{fb} = \frac{2ck}{3}$ and the second is to induce quality choice $q_i^{fb} = \frac{3}{4ck^2}$ for all $i = 1, ..., n^{fb}$) and only one regulatory variable. To illustrate this, we study two benchmarks. The first benchmark is created by setting a fee such that the total quality provision is optimal (quality optimal price p_q) and the second by setting a fee inducing optimal entry (entry optimal price p_n).

Lemma 3.2 *If the regulation authority chooses the fee* $p_q = \frac{c}{n_q^2} = \frac{1}{16ck^2}$ *total quality is fixed to the first-best level. The number of entrants is* $n_q = 4ck > \frac{2ck}{3} = n^{fb}$ *and thus identified as excess capacity. Welfare at* p_q *is* $W_q = \frac{11}{192ck^2} \approx \frac{0.057}{ck^2}$.

Individual quality levels are $q_q = \frac{1}{8ck^2}$ and are thus not first-best efficient, while total quality provision is. The quality optimal benchmark differs from the first-best solution only by number of physicians who provide the first-best efficient amount of total quality.

Lemma 3.3 *To induce the optimal number of practices, a fee of* $p_n = \frac{27}{2ck^2}$ *is required. When faced with this price, physicians provide a quality level of* $q_n = \frac{9}{2ck^2} > \frac{3}{4ck^2} = q^{fb}$ *which is identified as excess capacity in quality. At this price welfare is* $W_n = -\frac{147}{16ck^2}$.

It is worth noting that we observe excess capacity in both benchmarks. This is because the quality level varies in price linearly, and the number of entrants varies inversely. Since $p_q < p_n$, it is impossible to implement the first-best optimum by just choosing the correct price.

3.5.2 First-best efficient regulation

The above mentioned problem can be solved by applying a second regulatory variable or by non-constant compensation. Suppose the regulator could compensate on the basis of any non-constant reimbursement scheme. If marginal compensation increases with market share, optimal quality provision could be induced. Average compensation could be set to induce proper entry.[12] Let M be the market share of the physician. If, for instance, individual compensation is based on $p_{ind}(M) = -\frac{3}{2ck^2} + \frac{5}{4k}M$, the symmetric non-cooperative equilibrium will be first-best efficient (see appendix for a derivation). Licence fees are somewhat simpler. If the price induces the optimal quality level, the regulator has to charge an amount of $\frac{45}{16c^2k^3}$ to implement the first-best optimum. Introducing lower bounds for quality ($q \geq q^{fb}$), like the minimum standard in Economides (1993), and controlling entry by a suitable fee ($p = \frac{3}{8ck^2}$) will also be first-best efficient. Disease management programs may be seen as such a minimum standard.[13]

Licence fees can be seen as a tax imposed through demanding (and time consuming) educational requirements.[14] Looking at the increases in physician densities in the 15 EU member states from 1989 to 1999 (EUROSTAT (2001, p. 6)), it would seem that the effect of the "tax" on entry is limited. Furthermore, as far as we know, neither non-constant compensation nor explicit licence fees have ever been considered in health care regulation. One reason may be that the assumption of a benevolent regulator is not always appropriate. Health care regulation may be viewed as the outcome of a bargaining game or as the outcome of a rent seeking contest where interest groups strategically interact. Another reason may be that prestige is a (major) factor in investments in health care. This especially applies to the financing of hospitals and to high-tech medicine.

[12] We are indebted to Amihai Glazer for pointing this out.

[13] In Germany the first disease management programs started on July 1, 2002. However, there is a large political debate about whether these guidelines introduce upper or lower bounds on the quality of care.

[14] We thank an anonymous referee for providing this interpretation.

3.5.3 The second-best optimum

Suppose that the regulator can commit herself to price policies. Welfare is maximized with respect to the price. Substituting the non-cooperative solutions for q and n of Proposition 3.1, (3.10) simplifies to

$$W = \frac{23}{24}\left(\frac{p^2}{2ck^2}\right)^{\frac{1}{3}} - p.$$

Proposition 3.3 *The second-best optimum of the sequential 3-stage entry-location-quality game is attained iff the practices are arranged symmetrically and the price is set to* $p^{sb} = \left(\frac{23}{36}\right)^3 \frac{1}{2ck^2}$. *The resulting equilibrium values are:* $n^{sb} = \frac{72}{23}ck$, $q_i^{sb} = \left(\frac{23}{36}\right)^2 \frac{1}{2ck^2}$ *for all* $i = 1, ..., n^{sb}$ *and* $W^{sb} = \left(\frac{23}{36}\right)^3 \frac{1}{4ck^2} \approx \frac{0.065}{ck^2}$.

These values are between the values of the benchmarks given by Lemma 3.2 and Lemma 3.3, implying excess capacity in both entry and total quality provision. The deviation of these values from those induced by p_q is small and consequently welfare is close to W_q as well.

3.5.4 Time consistent regulation

When the social planner cannot commit herself to a certain price, the time consistency issue arises. Whether the second-best outcome can be implemented or not depends on the credibility of the announcement of $p = p^{sb}$ prior to stage 1 of the game. If the regulator has an incentive to reoptimize after the entry and location decisions based on p^{sb} have been made, then p^{sb} is not credible. Differentiating (3.10) with respect to q we observe $p^{reop} = \frac{c}{n^2}$ to be the welfare maximizing fee. Welfare increases from $\frac{0.065}{ck^2}$ to $\frac{0.071}{ck^2}$. But, with rational physicians, it is impossible to have different fees at the different stages of the game. Physicians anticipate the regulator's incentive to reoptimize between stages two and three and base their decisions on p^{reop}. Since the functional form corresponds to the quality optimal fee, the decisions are based on p_q resulting in welfare $\frac{0.057}{ck^2}$ as calculated above. Since $W^{sb} > W_q$, the regulator would like to bind herself to p^{sb}. Of course, this argument, and with it, the time inconsistency result do not rely on the second-best price.

Proposition 3.4 *When the regulator sets the price between stages 2 and 3 the optimal price is given by* p_q. *As the regulator has no incentive to change price after quality decisions have been made,* p_q *is the optimal time consistent policy.*

3.5.5 The median voter equilibrium

In a democracy it may be more appropriate to consider an elected regulator. Suppose the regulator wants to be re-elected, then the price will be set in favor of the electorate. If there is majority voting, the median voter decides on p. Assuming that the share of physicians relative to the share of patients in the population is small, the fee will be chosen to maximize the patients' utility function. Utility is obtained by replacing the quality cost term in the welfare function (3.10) by p.

Suppose the voters take the effect of changes in p on firms profits and the equilibrium number of firms into account. If they can decide on the number of licences and the fee, they can implement the first-best optimum. The second-best optimum is the outcome if they can only decide on the fee.

But the sequential structure of the game implies myopic patients. Hence, the voting behavior described above is not time consistent. Given a certain number of practices, patients care only about their net gain from quality provision $q - p$. Differentiating this expression with respect to p yields a first order condition of $n = 2ck$. If this equation holds, the voters have no incentive to change the price after entry and location decisions have been made. The price corresponding to $n = 2ck$ characterizes the voting equilibrium.

Proposition 3.5 *The symmetric voting equilibrium of the sequential 3-stage entry-location-quality game is attained iff the price is given by* $p^v = \frac{1}{2ck^2}$. *The resulting equilibrium values are:* $q^v = \frac{1}{2ck^2}$, $n^v = 2ck$, *and* $W^v = -\frac{1}{48ck^2}$.

Excess capacity in both entry and total quality is observed. The price exceeds the second-best price implying higher quality provision but fewer entrants than in the second-best equilibrium. Compared to the time consistent regulatory equilibrium, price and quality are higher and the number of firms is smaller. Total quality provision doubled while costs are 8 times as high. Relative to the costs, the quality of care is poor.[15] The welfare derived from the values of Proposition 3.5 is lower than in the time consistent regulatory case, i.e $W^v < W_q$ (see Lemma 3.2). Thus, the commitment problem is more severe.[16]

---

[15] This is exactly what may be concluded from the World Health Report 2000 for the German health care system (see WHO (2000, p. 153)).

[16] When precluding physician non-price competition as in Gravelle (1999), completely different results are obtained, since the voters no longer have the incentive to promote quality competition by increasing the fee. The voting equilibrium then has lower price,

3.5.6 A numerical example

To illustrate the results of the welfare and time consistency analysis we provide a numerical example for $c = 48$ and $k = \frac{1}{8}$ (see Table 3.5.6 below). The first-best optimum is attained if the total quality of 4 quality units is provided by 4 physicians. This can be done by setting $p = 3$ and demanding $\frac{5}{8}$ as licence fee. In this case, expected welfare is 0.25. The benchmarks shown in the following two rows show excess entry in the quality optimal case and excess quality provision in the entry optimal case. In the commitment game (second-best) we observe excess capacity in terms of both entry and total quality. But this policy is not time consistent. The regulator's incentive to reoptimize is shown in the second last row where a price reduction after the entry decisions have been made (based on the second-best price) increases welfare from 0.09 to 0.10. The unique time consistent equilibrium is given by the quality optimal price policy. The commitment problem is more severe in the median voter equilibrium as indicated by the last row.

	p	n	q	$n \cdot q$	W
first-best	–	4	1	4	0.25
quality optimal	0.08	24	$0.1\bar{6}$	4	0.08
entry optimal	18	4	6	24	−12.25
second-best	0.17	18.78	0.27	$5.1\bar{1}$	0.09
reoptimized	0.14	18.78	0.21	4	0.10
median voter	$0.6\bar{6}$	12	$0.6\bar{6}$	8	$-0.4\bar{4}$

Table 3.1: A numerical example for $c = 48$ and $k = \frac{1}{8}$.

3.6 Price Competition

In the presence of the commitment problem discussed in the previous section, a natural suggestion would be to allow physicians to also compete in prices. Suppose that at stage 3 of the game physicians decide simultaneously on quality *and* prices. Then the critical patient, i.e. the patient who

lower quality, and fewer firms. This changes implications for the movement of the health care market, especially for the quality of care, when exposing the health care market to price competition.

is indifferent between the practices i and $i+1$, is given by

$$x_i^{crit} = \frac{q_i - q_{i+1}}{2c(l_{i+1} - l_i)} - \frac{p_i - p_{i+1}}{2c(l_{i+1} - l_i)} + \frac{l_i + l_{i+1}}{2}. \qquad (3.11)$$

The optimal mix of prices and quality is obtained by differentiating the profit function with respect to p_i and q_i.

Proposition 3.6 *The symmetric equilibrium of the sequential 3-stage entry-location-quality and price game is given by the quality optimal benchmark, i.e. by the optimal time consistent regulatory outcome.*

The proof is presented in the appendix. Since the competitive outcome coincides with optimal time consistent regulatory outcome, the commitment problem cannot be solved by introducing price competition into the health care market. This is not very surprising since both the regulator and the physicians trade off the benefits of quality provision against its costs, where price is used to reduce quality competition.

3.7 Reimbursement of physicians in Germany

In recent decades, we have seen many health care reforms in Germany. But reimbursement of physicians has always remained on a fee-for-service basis. In 1993 there was a switch from cost reimbursement to sectoral budgets in the German health care market. The (regional) statutory sickness funds[17] bargain with the (regional) physicians' associations about a (regional) budget for panel doctors. After the budgets are agreed, the physicians' associations reimburse the physicians on a fee-for-service basis subject to the constraints given by the fixed budgets. More precisely: a relative value system is applied. Every service is scored with a certain number of points. The monetary value of one point is endogenous. It is determined by dividing the budget by the total number of points submitted for reimbursement by all physicians. Consequently the budgets in Germany are always met.[18]

Our model can be applied to the German reimbursement system. The mass of patients was assumed to be equal to one. Then, total benefits of the market are p. This is why p can alternatively be interpreted as a fee-for-service or as a budget for physician services. The point value in our model

[17] In Germany health insurance companies are called sickness funds.

[18] For a more detailed description of the German physician reimbursement system see OECD (1992, pp. 57-72) and more recent European Observatory on Health Care Systems (2000, pp. 102-106).

is observed very easily. Every patient receives the same service independent of the physician he visits, and it is not possible for the physicians to treat a patient more than once. Hence, changes in p directly translate into identical changes in the point value. If one treatment is scored with one point the point value is p.

From equation (3.9) we know that the equilibrium number of entrants is a decreasing function of p. This counter intuitive result was explained by equation (3.8): the quality increases linearly in price, and, in the presence of convex costs of quality provision, this has a deterrent effect. Or, the other way round, a drop in prices makes the health care market more attractive. The reduced quality competition encourages further entries. This inverse relationship is observed in Germany (see Figure 3.2). The point value was decreasing from 1993 to 1996. The improvement in 1997 is due to the introduction of clinic budgets in 1997 (see SVRKAiG (1998, pp. 369-372)). Since the clinic budgets were introduced in July 1997, the 1998 value is higher. Without these changes the point value would have decreased further.[19] In the political debate it is argued that the increasing number of physicians caused the decline in the point value. This effect is called the treadmill effect in a fixed budget system. Applying the target income hypothesis, an expected decrease in the point value implies an increase in the number of (invoiced) treatments, i.e. supplier induced demand (see Benstetter and Wambach (2001)). This is very intuitive and is supported by Figure 3.2. Our model shows that, due to spatial and quality competition, this relation persists in the absence of supplier induced demand.

Suppose that supplier induced demand is absent from the German health care market. Furthermore, assume that the point value in Germany is between the benchmarks studied in Section 3.5 and that the regulator cannot commit herself. Excess capacity in total quality and entry was identified in this price range.[20] To attain the first-best solution price cuts should be accompanied by a drop in physician licences. Since this is not observed in Germany, given that our model applies, the regulation of the health care market is not welfare maximizing. It seems as if the market converges to the time consistent equilibrium instead of to the first-best efficient outcome. Entry regulation is too weak.

One may argue that applying our model to this question is not appropri-

[19] We are indebted to Dr. Dahlhausen from the Kassenärztliche Bundesvereinigung (KBV) for providing the pointvalue data. For some of that data see KBV (2000, p. C6).

[20] The Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (SVRKAiG) reports, depending on the disease, under-provision and over-provision of health care services in Germany (see SVRKAiG (2002)). Our model only explains over-provision due to quality competition.

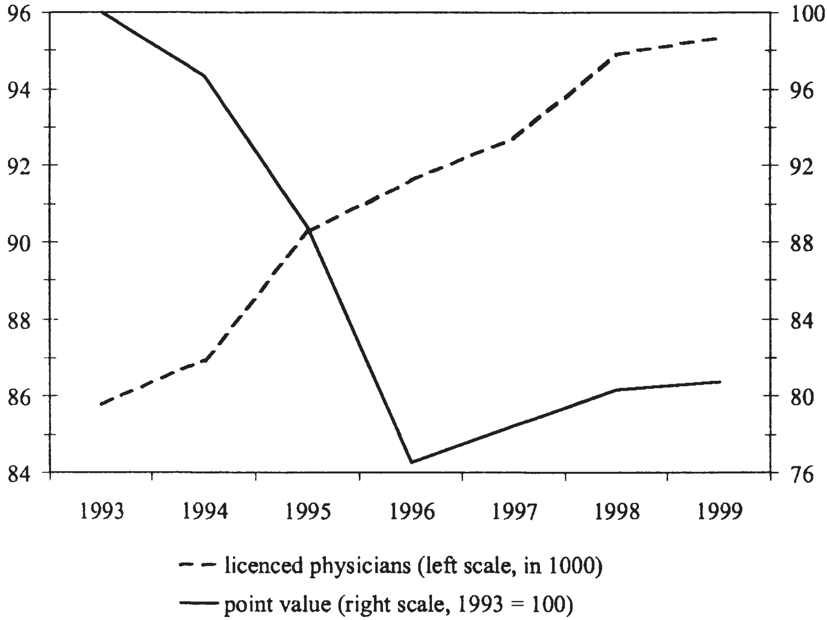

Figure 3.2: For the Western Länder a reciprocal relationship between the point value and licenced physicians is observed. *Source:* KBV (2001, p. A20) and BMG (1999, p. 260).

ate since the point value was decreasing but the size of the budget was not. As a proxy for the budget's size we use the expenditure for treatments by panel doctors (see Figure 3.3). From 1993 to 1999 the expenditure increased by roughly 9 percent in 1991 prices.[21] Technical progress is one of the main drivers of that growth. Making this progress available to the insured improves the quality of the medical treatments they receive.[22] This can be interpreted as an increase in voter influence on health care regulation. The increase in expenditure, i.e. in price, drives quality up from the time consistent quality level to (or in direction of) the voting equilibrium value. Our model cannot explain the increase in active physicians without changing the specification of the quality cost function.[23] However, approaching the

[21] We deflated the numbers for expenditure on treatments found in Bundesminiterium für Gesundheit, BMG (1999) and KBV (2001) by the CPI for all households.

[22] Alternatively, the expenditure for outpatient care per capita can be used for the same argument. The increase from 1993 to 1998 is about 11 percent.

[23] As argued in Section 3.3, if marginal costs of quality provision are low or prohibitive, then the number of practices increases in the budget. Since the quality provided in

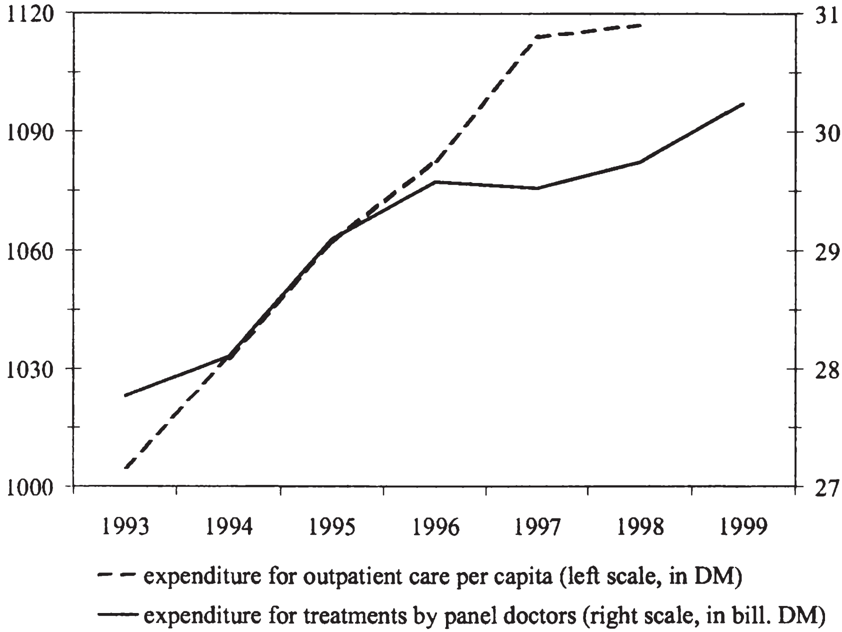

Figure 3.3: Expenditure for outpatient care per capita and for treatments by licenced physicians, both for the Western Länder and both in 1991 prices. *Source:* Statistisches Bundesamt (2001, p. 129), BMG (1999, p. 402), and KBV (2001, p. G2).

first-best efficient outcome starting from the time consistent outcome, price increases should be accompanied by a drop in physician licences. We again conclude that licences are allocated too generously.

3.8 Conclusion

We presented a model of physician competition in a price regulated environment. In the absence of price competition physicians resort to other variables to increase profits: location and quality. The problem was analyzed in a 3-stage entry-location-quality game. We restricted ourselves to inelastic demand and symmetric equilibria to obtain a lower benchmark case with respect to the level of quality provision. We found that quality is an im-

Germany is likely to be high, it may be reasonable to assume prohibitive or at least very high marginal costs.

portant strategic variable used aggressively to reduce locational handicaps. The closer two physicians are located, the tighter competition, and, consequently, the higher the quality provided. An increase in the fee reduces the equilibrium number of entrants. The higher price sharpens quality competition to an extent that it has a deterrent effect. The inverse relationship of the point value and the number of licenced physicians observed in Germany is mainly attributed to the treadmill effect. This may indeed be the more reasonable explanation, but our model shows that the relation also persists in the absence of supplier induced demand due to quality competition.

The regulator's goal is to set a welfare maximizing fee. We studied two prices resulting in two benchmarks, one in which the number of firms is first-best efficient, and the other in which total quality is. Since these two prices do not coincide, the first-best optimum cannot be implemented by just setting the correct fee. If the regulator can commit herself, she will apply the second-best fee which lies between the benchmark prices. But only the second of the two benchmarks was identified as being a credible price policy. Hence, in the game without commitment, the second-best policy is ruled out as time inconsistent. The regulator is limited to the quality optimal price. Since expected welfare is lower in the game without commitment, the regulator would like to bind herself to the second-best price. The commitment problem cannot be solved either by introducing price competition into the health care market or by letting voter patients decide on the price. The outcome of the price competition game corresponds to the time consistent regulatory outcome. In the voter equilibrium, the commitment problem is more severe.

Theoretically, the first-best efficient outcome can be implemented by simultaneously limiting entry by means of licences. Studying the relation between the price and the number of physicians makes it possible to judge whether such a policy is being applied. In both the scenarios discussed in the previous section it was indicated that, although there is entry regulation in Germany, like need-related planning in the allocation of physician licences, entry regulation seems too weak.

3.9 Appendix

Proof. (*Lemma 3.1*) Since we want to prove that the symmetric arrangement of practices is a Nash equilibrium we take symmetric locations of all practices $j \neq i$ as given and show that the best physcian i can do is to locate in the middle between practices $i-1$ and $i+1$.

We restrict the analysis to the existence set Q, where i's location decision yields a pure strategy equilibrium in the quality subgame. Q is non-empty if the distance between all frims, $1/n$, is sufficiently large. In other words, n must be sufficiently low. Q is non-empty if $n \leq n^*$ (see equation (3.9)).

To show that i's best response is locate in the middle between $i-1$ and $i+1$ we differentiate the objective function with respect to l_i. Substituting the expressions following the equations (3.5) to (3.7) into the first order condition and rearranging yields

$$\begin{aligned}\frac{8c^2k}{p^2}\frac{\partial \Pi_i^{II}}{\partial l_i} &= \frac{1}{\left(l_i - l_{i-1}^*\right)^3} - \frac{1}{\left(l_{i+1}^* - l_i\right)^3} \\ &+ \frac{1}{\left(l_{i+1}^* - l_i\right)^2\left(l_i - l_{i-1}^*\right)} - \frac{1}{\left(l_{i+1}^* - l_i\right)^2\left(l_{i+2}^* - l_{i+1}^*\right)} \\ &+ \frac{1}{\left(l_i - l_{i-1}^*\right)^2\left(l_{i-1}^* - l_{i-2}^*\right)} - \frac{1}{\left(l_i - l_{i-1}^*\right)^2\left(l_{i+1}^* - l_i\right)}.\end{aligned} \tag{3.12}$$

It is easy to see that, in the symmetric solution, the right-hand side of (3.12) is zero. To identify this decision as a global maximum on $\left(l_{i-1}^*, l_{i+1}^*\right)$ we have to show that i is worse off when he deviates from the symmetric location or that he is better off when he moves to the center, starting from any asymmetric location. Although (3.12) is hard to interpret, the sign is determined for all $l_i \in \left(l_{i-1}^*, l_{i+1}^*\right)$. Assume that the practices have symmetric locations except i which is closer to $i-1$ than to $i+1$, then: $l_{i-1}^* - l_{i-2}^* = l_{i+2}^* - l_{i+1}^* = \frac{1}{n}$, $l_{i+1}^* - l_i > \frac{1}{n}$ and $l_i - l_{i-1}^* < \frac{1}{n}$. By making a small step to the midpoint, i can increase profits since the right hand side of (3.12) is positive. By the same argument, this is true when i is closer to $i+1$. Notice that, in this case, a movement to the center requires dl_i to be negative. Hence, given the symmetric location of the others, i cannot be better off when he deviates from the perfect symmetric solution, and he is actually worse off. ■

Proof. (*The reimbursement $p(M) = -\frac{3}{2ck^2} + \frac{5}{4k}M$ yields the first-best solution.*) The average individual compensation is set such that the physicians exactly break even when providing the first-best quality level in a market with n^{fb} competitors:

$$\frac{p_{ind}(n^{fb})}{n^{fb}} = k(q^{fb})^2.$$

We obtain $p_{ind}(n^{fb}) = \frac{3}{8ck^2}$. Using the definition of the critical patient from equation (3.2), the market share of physician i is given by $M_i =$

$\frac{n}{2c}(2q_i - q_{i+1} - q_{i-1}) + \frac{1}{n}$. When the individual reimbursable price depends on market share, the objective function is given by $\Pi_i = p_{ind}(M_i)M_i - kq_i^2$. Assuming symmetry, the first order condition for profit maximizing quality provision is $q_i^* = \frac{1}{2ck}(p'_{ind} + p_{ind}n)$. This equation must hold at n^{fb} yielding:

$$\frac{p'_{ind}(n^{fb})}{2ck} + \frac{p_{ind}(n^{fb})}{3} = q^{fb}.$$

Inserting q^{fb} and $p_{ind}(n^{fb})$ gives the solution for p': $p'_{ind}(n^{fb}) = \frac{5}{4k}$. There is an infinite number of reimbursement schemes satisfying the conditions $p_{ind}(n^{fb}) = \frac{3}{8ck^2}$ and $p'_{ind}(n^{fb}) = \frac{5}{4k}$. The easiest is the one mentioned in the text, i.e. $p(M) = -\frac{3}{2ck^2} + \frac{5}{4k}M$, where M is physician market share. ■

Proof. (*Proposition 3.6*) Substituting the expressions for the critical patients into the profit function yields the objective function for the third stage of the game:

$$\Pi_i = \frac{p_i}{2c}\left[\frac{q_i - q_{i+1}}{l_{i+1} - l_i} - \frac{p_i - p_{i+1}}{l_{i+1} - l_i} - \frac{q_{i-1} - q_i}{l_i - l_{i-1}} + \frac{p_{i-1} - p_i}{l_i - l_{i-1}}\right] + \frac{p_i}{2}(l_{i+1} - l_{i-1}) - kq_i^2.$$

When solving for the symmetric equilibrium, the first order conditions are given by

$$q = \frac{np}{2ck} \quad \text{and} \quad p = \frac{c}{n^2}$$

proving the assertion. ■

Chapter 4

Gatekeeping and Secondary Care Competition

4.1 Motivation

The UK and the Scandinavian countries are examples of countries where general practitioners (GPs) have a gatekeeping role in the health care system.[1] Patients do not have direct access to secondary care. They need a referral from their (primary care) GP to get access to a hospital or a specialist.[2] Restricting direct access to secondary care by giving GPs a gatekeeper role is currently on the political agenda in Germany, while in Sweden there has been some debate about whether patients should be able to approach a specialist or a hospital directly. The current chapter contributes to the discussion on gatekeeping by analyzing the competition effects that arise when GPs are equipped with a gatekeeping role.

In general, there are three main arguments for introducing gatekeeping in health care markets (see Scott (2000)). Firstly, it is usually claimed that gatekeepers contribute to cost control by reducing 'unnecessary' interventions.[3] Second, efficiency in the health care system increases as the patients more appropriately treated by a GP are screened out. Third, it is argued

[1] This chapter closely follows Brekke, Nuscheler and Straume (2003).

[2] In the US, several Health Maintenance Organizations also practice gatekeeping. Recently, however, some HMOs have relaxed the restrictions on access to specialists (see, e.g., Ferris et al., 2001).

[3] Although this is a common argument for restricting access to secondary care, the empirical evidence that gatekeeping actually contributes to lower health care expenditures seems to be scarce (see, e.g., Barros (1998)).

that secondary care is used more efficiently since "GPs usually have better information than patients about the quality of care available from secondary care providers" (Scott (2000, p. 1177)). In the present model we focus on the third argument, highlighting the fact that making this information available to patients changes the nature of competition between secondary care providers, which in turn affects the social desirability of gatekeeping.

As pointed out in a seminal paper by Arrow (1963), uncertainty and asymmetric information make health care markets different from other markets. Uncertainty generates demand for health insurance, implying that non-price strategies are important in attracting patients as the consumption of medical care is paid for by a third party (see also Section 1.4). Asymmetric information is present in the sense that consumers (patients) are typically less informed about their health conditions, and thus the appropriate treatment, than the providers of medical care. In this chapter we stress both these features of health care markets.

Building on the familiar model of Hotelling (1929), we consider a secondary care market with two providers (hospitals). In order to attract patients, and thus obtain third party payments, the hospitals have two strategic variables at their disposal - location and quality of care. We refer to location as the specialization or service mix at a hospital, though it may also be interpreted in geographical terms. Thus, hospitals engage in non-price competition in terms of both horizontal and vertical differentiation of services.

The major aim of our model is to highlight the informational role of gatekeepers in such secondary care markets. Without gatekeepers, we assume that at least some of the patients are uninformed about both their own specific diagnosis and the exact characteristics of the secondary care market. Thus, with direct access to secondary care, patients' choices may be subject to substantial errors. First, a patient may end up in a poor match, i.e., he may choose the hospital that is less able to cure his disease. Second, he may decide to go to the specialist who provides the lower quality of care. By introducing GP gatekeeping we assume that all relevant information is transmitted to the patients, thereby enabling them to make informed choices.[4] Thus, GPs observe the actual disease of a patient with certainty and know which specialist is more able to cure a particular disease. Additionally, we assume that GPs obtain perfect quality signals. Both features are in line with the above mentioned second argument for introducing gatekeeping.

[4] We abstract from agency problems by assuming that the GPs truthfully convey their information to the patients.

We analyze the informational role of gatekeepers by applying two different variants of the basic model. In the first part of the chapter we consider an exogenously given number (fraction) of patients that are *a priori* fully informed about their disease and the most appropriate treatment for this condition. One possible interpretation is to think of these patients as the chronically ill who have obtained all relevant information through repeated consumption. Introducing GP gatekeeping is then simply equivalent to making the uninformed fraction of patients fully informed. Since gatekeeping can only be regulated directly, we refer to this variant as 'direct gatekeeping'. Although we consider gatekeeping to be costless, we find that introducing strict gatekeeping, i.e., making it compulsory to get a referral to secondary care, is not necessarily socially desirable. The reason is that more informed patients lead to more intense quality competition between hospitals, amplifying the hospitals' incentives to differentiate their services. Consequently, gatekeeping may induce too much quality and differentiation from the viewpoint of social welfare.[5] However, we show that, under second-best price regulation, gatekeeping is always socially beneficial. Thus, gatekeeping should be accompanied by proper price regulation.

One might ask, however, if an uninformed patient would not *voluntarily* consult a GP before seeking secondary care. We argue that consulting a GP may in itself involve costs for the patients, such as out-of-pocket payments, travelling and/or time costs. A patient would then have to compare the benefits of consulting a GP - in terms of reduced risk of choosing the less suitable hospital - against such costs. This situation is analyzed in the second part of the chapter, where we let the fraction of informed patients be endogenously determined by this trade off. A crucial feature of this variant of the model is that GP consultation can be indirectly influenced through price regulation. Consequently, we will refer to this mechanism as 'indirect gatekeeping'. The endogeneity of the consultation decision alters hospitals' incentives. Although differentiation relaxes quality competition, it also increases the fraction of informed patients, since the (expected) benefits of consulting a GP now are higher. Thus, the incentives for differentiation of services are weakened. Moreover, there is now a real cost of introducing a strict gatekeeping regime, which is reflected in the individual costs of GP consultation. In this case, individual consultation incentives coincide with

[5] This result is related to Dranove et al. (2003), who empirically analyze whether public disclosure of patient health outcomes at the level of the individual physician or hospital ('report cards') is beneficial to patients and social welfare. They find that report cards led to both selection behavior by providers and improved matching of patients with hospitals. However, on net this led to higher levels of resource use and to worse health outcomes (for sicker patients).

the social incentives, implying that there is no need to regulate gatekeeping directly. However, we show that second-best price regulation implies a *de facto* strict gatekeeping regime - in which every patient finds it beneficial to consult a GP before accessing the secondary care market - if quality costs or GP consulting costs are sufficiently low, or if mismatch costs are sufficiently high.

The chapter relates to both the general literature on spatial competition and the literature on (imperfect) competition in health care markets. The interaction between quality and location choices has been investigated by Economides (1989) under price competition and Brekke et al. (2002) under price regulation. This chapter contributes to this literature by introducing imperfect information into the framework. As previously mentioned, we find that the hospitals' incentives to differentiate services are significantly altered by the presence of uninformed consumers. In particular, we find that uninformed consumers tend to soften the incentives for horizontal differentiation. In this respect our findings are in the spirit of Bester (1998), who shows that quality competition may induce minimum differentiation - i.e., agglomeration at the market center - when consumers are uncertain about product quality and use observed prices to ascertain the quality of goods.

In a related paper, Calem and Rizzo (1995) analyze hospitals' choices of quality and speciality mix (location) under exogenous prices. An incentive for closer locations is introduced by assuming that the hospitals cover a fraction of their patients' mismatch costs. Besides this particular assumption, their paper differs from ours in two important ways. Firstly, they are not concerned with imperfect information and the issue of gatekeeping and how this affects the nature of competition in the market for secondary care. Second, they do not consider the implications of optimal price regulation on the hospitals' incentives with respect to quality and location choice.[6]

The model also relates to the more general literature on transparency in imperfectly competitive markets.[7] Increased transparency on the consumer side of the market typically leads to intensified price competition and thus

[6] Two other related papers applied to the primary care market are Gravelle (1999) and Nuscheler (2002), where the latter is Chapter 3 of this thesis. Both papers address the issue of competition between physicians by investigating the interaction between quality and location choices when prices are regulated. Building on the seminal contribution of Salop (1979), they apply a circular model with attention directed towards entry of physicians into the market, so the focus of these papers is clearly quite different from ours.

[7] See, e.g., Varian (1980), Burdett and Judd (1983), Schultz (2002, 2003), Lommerud and Sørgard (2003).

to a more socially desirable market outcome. This chapter contributes to this literature by analyzing the effects of improved transparency in markets that are characterized by *non-price competition*. In this case, more intense competition between firms does not necessarily improve social welfare. Improved market transparency consequently has ambiguous welfare effects.[8]

The remainder of the chapter is organized as follows. The basic framework is presented in Section 4.2. In Section 4.3, we consider the case of 'direct gatekeeping', where the fraction of informed patients is exogenously given. In Section 4.4, we analyze the case of 'indirect gatekeeping', where the fraction of informed patients are endogenously determined by individual GP consultation decisions. In both sections we derive the quality and specialization equilibria and analyze the social desirability of gatekeeping. We also discuss the issue of optimal price regulation. Section 4.5 provides concluding remarks.

4.2 The model

There is a continuum of patients with mass 1 distributed uniformly along the Hotelling line $S = [0, 1]$. The location of a patient is denoted $z \in S$ and is associated with the disease he suffers from. A disease z can be seen as a realization of a random variable Z which is uniformly distributed on S. All patients need one medical treatment to be cured. There are two health care providers - henceforth called hospitals - both able to cure all diseases. However, they are differentiated with respect to the disease they are best able to cure. Specialization of a hospital is denoted $x_i \in \mathbf{R}$, $i = 1, 2$. Without loss of generality, we will assume throughout the chapter that $x_1 \leq x_2$. Note that the degree of specialization is not restricted to the disease space S. Thus hospitals may locate outside S.[9]

In addition to specialization, there is a second strategic variable used by the hospitals to attract patients, namely the quality of care $q_i \geq 0$, $i = 1, 2$. Quality costs are assumed to be symmetric and quadratic, kq_i^2, where $k > 0$. These costs are considered to be fixed, i.e., they are independent of how many patients are actually treated. This implies that quality has the characteristics of a public good at each hospital. Examples of such

[8]Another related paper in this strand of the literature is Baye and Morgan (2001), who analyze the competition effects of information gatekeepers on the Internet, where such gatekeepers create a market for price information by charging fees to firms that advertise prices and to consumers who access the list of advertised prices.

[9]This assumption is made for convenience, but does not qualitatively affect any of the results presented here.

quality investments are the cost of searching for and hiring more qualified medical staff, additional training of existing medical staff, and investments in improved hospital facilities, which can be related to both medical machinery and non-medical facilities such as room standard or catering quality. Without loss of generality, other fixed costs are set to zero. Marginal production costs are assumed to be constant and equal to zero. This cost structure stresses the importance of fixed costs which seems reasonable for the hospital market.[10] The price for one treatment is denoted $p \geq 0$ and is set by some regulatory authority.[11] As the price is independent of which hospital is actually attended it may alternatively be interpreted as a premium for a health insurance with full coverage. The profit function of hospital i is given by

$$\Pi_i = pD_i - kq_i^2, \tag{4.1}$$

where D_i is the demand for hospital i treatment.

A patient's (ex-post) utility when going to hospital i is given by

$$u_i\left(z;p\right) := u\left(q_i, x_i, z; p\right) = v + q_i - p - t(z - x_i)^2. \tag{4.2}$$

The maximum gross willingness to pay for hospital treatment, v, is assumed to be sufficiently large for the entire market to be covered. Thereby, we preclude monopoly and kink equilibria and concentrate on competitive ones.[12] Notice that this assumption essentially means that all patients have access to hospital or specialist care, which seems reasonable, at least for developed countries (without waiting lists). The last term measures the mismatch costs incurred when treated by hospital $i = 1, 2$. The parameter $t > 0$ determines the importance of mismatch costs relative to the quality of care. Of course, mismatch costs would be zero if the patient suffers exactly from the disease for which the hospital he goes to is specialized. Mismatch costs are assumed to be quadratically increasing in distance.

To evaluate the effects of gatekeeping we consider two different patient types. The fraction $\lambda \in [0,1]$ of the population is fully informed when accessing the hospital market, i.e., these patients know their own location (diagnosis) and the specialization and quality provision of each hospital.

[10] The assumption of production-independent quality costs is widely used in the literature on quality competition in health care markets (see, e.g., Calem and Rizzo (1995), Lyon (1999), Gravelle and Masiero (2000), Barros and Martinez-Giralt (2002)).

[11] All results we derive also hold for constant marginal costs $MC > 0$. Let $\widetilde{p}$ denote the mill price, then the mark-up is given by $p = \widetilde{p} - MC$.

[12] In a circular model, Economides (1993) makes similar assumptions (see also Chapter 3), whereas Salop (1979) and Gravelle (1999) study monopoly and kink equilibria in detail.

In the first part of the chapter (direct gatekeeping, Section 4.3) we will assume that the number of fully informed patients is exogenously given. In this case we can think of these patients as the chronically ill, who know exactly what disease they are suffering from and have obtained sufficient information about the hospital market through repeated consumption. In the second part of the chapter (indirect gatekeeping, Section 4.4) we will endogenize λ by explicitly modelling patients' decisions about consulting a GP before accessing the hospital market. We will assume that the GP has a gatekeeper role in the system and that he or she obtains all significant information. This information is then truthfully conveyed to those patients consulting the GP, making them fully informed about all relevant variables.

To simplify the analysis we assume that the fully informed patients are uniformly distributed on S. Members of the remaining part of the population, $1-\lambda$, only know v, the distribution of Z, and that hospital treatment is required. They cannot observe x_i, q_i, and z. For these patients secondary care is an experience good. Their ex-ante utility of attending hospital i is given by

$$Eu_i\left(Z;p\right) := Eu\left(q_i^e, x_i^e, Z; p\right) = v + q_i^e - p - tE(Z - x_i^e)^2, \qquad (4.3)$$

where the superscript e denotes the expected value of the respective variable. Patients learn their ex-post utility given by (4.2) only through actual consumption.

For the direct gatekeeping scenario we will consider that the regulator can only influence λ through direct regulation. In theory, it is possible to imagine that the regulator can influence the amount of information available to patients in the market through several different means. We will, however, focus on what is probably the most realistic regulatory instrument, namely introducing a strict gatekeeping regime, where all patients are required to consult a GP before accessing the hospital market. Thus, the scope for regulating λ is restricted to setting $\lambda = 1$. In the indirect gatekeeping scenario, the regulator can indirectly influence the endogenously determined value of λ through price regulation.

The impact of introducing gatekeeping to the market for hospital or secondary care is analyzed in a 5-stage game:

- Stage 1: the regulator sets her available regulatory variables. These are one or both of p and λ. Regulation on the latter variable is restricted to setting $\lambda = 1$.

- Stage 2: the hospitals simultaneously decide on their specializations, x_1 and x_2, where $x_1, x_2 \in \mathbf{R}$, and $x_1 \leq x_2$.

- Stage 3: the hospitals simultaneously set their quality levels $q_1 \geq 0$ and $q_2 \geq 0$.

- Stage 4: patient information about x_i, q_i, and z can be obtained by consulting a gatekeeping general practitioner who truthfully conveys information about the relevant variables. The choice of consulting a GP is reserved for the second version of the model (Section 4.4). In the first version (Section 4.3) the share of fully informed patients, λ, is exogenously given.

- Stage 5: the patients demand secondary care treatment.

The sequential structure of the game is argued by the different degree of irreversibility of strategic decisions. Clearly, the decision of whether to consult a gatekeeping GP and/or which hospital to go to is the most flexible decision to be taken in the entire game. Changing quality or specialization requires more effort and investment. In both cases it may be necessary to replace some medical machinery and/or have the current staff undergo significant training, or even hire new staff. Although it may sometimes be hard to distinguish between quality investments and a change of specialization, it seems logically consistent to assert that hospitals first decide what to produce (their service or speciality mix), and then determine the quality of services.[13] This sequential structure is common in models that combine horizontal and vertical differentiation (see, e.g., Economides (1989), Calem and Rizzo (1995), Bester (1998), Gravelle (1999)).

That the regulator can determine λ and p at the beginning of the game essentially means that we consider commitment power on her side. This assumption is, of course, crucial as in most sequential games. With respect to λ, this may be justified since introducing a strict gatekeeping system ($\lambda = 1$) must be regarded as a major reform of the health care system. This may be less clear with the price. As in Brekke et al. (2002) and Chapter 3 there will be an incentive to reoptimize after specializations have been chosen. Nevertheless, since commitment is valuable for the regulator, one could argue that she should be able to obtain such commitment power, either through reputation or by creating institutional mechanisms that makes it costly, or otherwise difficult, to change the regulated price. In any case, as price regulation is not the major focus of the present chapter we will concentrate on the commitment case.

[13] Calem and Rizzo (1995) discuss this in some more detail.

4.3 Direct gatekeeping

In this section we will consider that $\lambda \in [0, 1]$ is exogenous and can only be regulated directly by setting $\lambda = 1$. Hence, the regulator determines whether or not to introduce a strict gatekeeping system, and thereby make all patients fully informed. The game is solved by backward induction.

4.3.1 The specialization-quality game

4.3.1.1 The demand for secondary care

The share $1 - \lambda$ of the population is uninformed about the actual quality levels and about specializations. Moreover, these people do not know the disease they suffer from. To make a decision about which hospital to consult, patients have to evaluate their expected utility, given by equation (4.3), for both hospitals. Imposing symmetry, these patients are indifferent between hospitals in expected terms. Both hospitals receive one half of these patients, $(1 - \lambda)/2$.

This assumption is not necessary, but it eases the presentation of the main ideas dramatically. Actual demand depends on the patients' beliefs which influence expected utilities. Since these beliefs do not change the optimization problem of the hospitals (see below), they can be neglected at the earlier stages of the game. Nevertheless, beliefs have to be confirmed in equilibrium and, as we concentrate on symmetric equilibria, beliefs will also be symmetric. Our assumption that each hospital gets half of the uninformed patients is thus the outcome of a more general treatment.[14] Given the symmetry assumption, the decision about which hospital to attend reduces to flipping a fair coin - which seems not unrealistic.

In contrast, the informed fraction of the population, λ, is responsive to quality investments and specialization decisions as both strategic variables and the own disease are observable. The informed patient who is indifferent between hospital 1 and hospital 2 suffers from disease $\overline{z}$, which is obtained by solving $u_1(z; p) = u_2(z; p)$ for z, where u_i is given by equation (4.2), yielding

$$\overline{z} = \frac{q_1 - q_2}{2t(x_2 - x_1)} + \frac{x_1 + x_2}{2}. \tag{4.4}$$

[14] Although there is a game of incomplete information (the fraction $1 - \lambda$ of patients do not know their disease type) and imperfect information (the fraction $1 - \lambda$ of patients cannot observe qualities and specializations), beliefs are irrelevant for the outcome. Subgame perfection is thus sufficient to obtain a unique symmetric (perfect Bayesian) equilibrium. This changes when λ is endogenized.

The demand for hospital 1 is thus $D_1 = \lambda\overline{z} + (1-\lambda)/2$. Hospital 2 receives the residual demand $D_2 = 1 - D_1 = \lambda(1-\overline{z}) + (1-\lambda)/2$.

4.3.1.2 Quality competition

We look for an equilibrium in pure strategies in the quality subgame.[15] If a pure strategy equilibrium exists, it is found by inserting demand from equation (4.4) into the profit function (4.1) and optimizing with respect to q_1, which yields the optimal quality provision for both hospitals for given specializations:

$$q^*(\Delta;\lambda,p) = \frac{p\lambda}{4tk\Delta}, \tag{4.5}$$

where $\Delta := x_2 - x_1$. The equilibrium quality levels depend only on the distance between hospitals' locations and not on their actual locations. An immediate implication is that optimal specializations will be characterized by some certain distance and not by absolute locations.

From (4.5) we see that $\lim_{\Delta\to 0} q^*(\Delta;\lambda,p) \to \infty$. Since quality investments are costly, this means that (4.5) yields negative profits if Δ is sufficiently small. In other words, there exists a (small) range of $\Delta \in \left[0,\overline{\Delta}\right]$ where investment incentives are so strong that hospitals are led into 'ruinous competition'. Thus, in order to secure positive profits - and thus pure strategy equilibrium existence - we have to impose a restriction that the hospitals are not located too closely. Let Q be the set of all location pairs (x_1,x_2) such that a Nash equilibrium exists in the quality subgame. Using (4.5), it is easily shown that $(x_1,x_2) \in Q$ if

$$k > \frac{p\lambda^2}{8t^2\Delta^2(1-\lambda+\lambda(x_1+x_2))}. \tag{4.6}$$

Assuming that (4.6) is satisfied, the comparative static results are straightforward: the smaller product differentiation, i.e., the smaller Δ, the more intense is quality competition. Patients are more responsive to quality improvements when mismatch costs are small. Thus t is a measure of competition intensity. Not very surprisingly, an increase in the quality cost parameter k has an adverse effect on quality provision. The better medical treatments are paid, the higher the benefits of capturing additional market shares from the competitor. At this stage of the game the only means of

[15]The concept of mixed strategies does not seem to make much sense in the context of hospital quality investments, so we disregard this possibility by assumption.

competition is the quality of care and thus hospitals will improve their quality. The comparative statics with respect to λ are the same as with respect to p: more informed patients lead to higher quality provision.

Because of its exogeneity, the fraction $1 - \lambda$ of the population cannot have any effect on competition, thus, λ can be interpreted as the density of patients that are distributed along the Hotelling line. When defining $\widehat{p} := \lambda p$ we obtain the same results as Brekke et al. (2002).

4.3.1.3 Specialization

At this stage of the game hospitals decide on their specialization, taking the effects on quality competition and demand into account. In order to obtain a perfect pure strategy equilibrium of the specialization-quality game, we follow the approach taken in similar location models[16] and restrict the strategy space of the specialization game to the set Q, for which a pure strategy equilibrium of the quality game obtains. Intuitively, it seems highly plausible to assume that the hospitals will not consider locations which trigger incentives for 'ruinous competition'. Following Economides (1986), we define the direction in which $\partial \Pi_i / \partial x_i$ is positive as the 'relocation tendency' of firm i. An equilibrium of the specialization game must then be at the zero relocation locus, $\partial \Pi_1 / \partial x_1 = \partial \Pi_2 / \partial x_2 = 0$, and a perfect equilibrium of the specialization-quality game is defined as the intersection between the zero relocation locus and the existence set Q. Formally, a specialization equilibrium (x_1^*, x_2^*) exists if

$$\frac{\partial \Pi_i \left(x_1^*, x_2^*\right)}{\partial x_i} = 0; \quad \frac{\partial^2 \Pi_i \left(x_1^*, x_2^*\right)}{\partial x_i^2} < 0; \quad \left(x_1^*, x_2^*\right) \in Q; \quad i = 1, 2.$$

Inserting the optimal quality levels into hospital 1's profit function, we obtain the following partial derivative with respect to x_1:

$$\frac{\partial \Pi_1}{\partial x_1} = \frac{\lambda p}{2} - \frac{p^2 \lambda^2}{8 t^2 k \Delta^3}. \tag{4.7}$$

As already mentioned, setting $\partial \Pi_1 / \partial x_1 = 0$ only yields Δ^*. There exists a continuum of locations fulfilling $x_2 - x_1 = \Delta^*$. Imposing symmetry leads to a unique equilibrium of the game, provided that $(x_1^*, x_2^*) \in Q$, where

$$x_1^* (\lambda, p) = \frac{1}{2} \left(1 - \Delta^*\right) \text{ and } x_2^* (\lambda, p) = \frac{1}{2} \left(1 + \Delta^*\right), \tag{4.8}$$

[16]See, e.g., Economides (1984, 1986, 1989), Hinloopen and Marrewijk (1999), Lambertini (2001).

and

$$\Delta^*(\lambda, p) = \left(\frac{p\lambda}{4t^2k}\right)^{\frac{1}{3}}. \tag{4.9}$$

It is easily shown that the second order conditions are met. However, it remains to identify the exact condition for equilibrium existence. According to the specification of the game, two requirements must be met. First, we need to have that $(x_1^*, x_2^*) \in Q$. Second, it must not be a profitable strategy for either firm to deviate in the quality subgame by offering zero quality and only serve the uninformed consumers arriving in equilibrium. Using (4.9) and (4.5), and imposing symmetry in the profit function, it is straightforward to show that both requirements are met, thus guaranteeing the existence of a unique symmetric equilibrium, if

$$k > \frac{p\lambda^4}{32t^2}. \tag{4.10}$$

For the remainder of the analysis, we will assume that this condition is met.

The hospitals' location incentives are governed by two opposing forces. *Ceteris paribus*, each hospital can obtain a larger share of the market by moving closer to its rival. On the other hand, closer locations imply that quality competition is intensified, as can be seen from equation (4.5).

Consider an increase in the price p. This will strengthen the market share effect, since hospitals now receive a higher mark-up on each treatment. However, a price increase also means that quality competition is amplified. From (4.9) we see that the latter effect always dominates: a higher price implies that hospitals aim at dampening the resulting increase in quality competition by locating further apart. Indeed, as long as the fee for secondary care treatments exceeds marginal costs (and λ is strictly positive), quality competition among providers induces product differentiation.

An identical mechanism determines the relationship between patient information and locations. More informed patients will result in stronger quality competition and hospitals will respond by differentiating more.[17] A social planner thus faces a trade off when setting the price or taking measures to improve information in the market. The improved quality has to be weighed against the change in aggregate mismatch costs.

We have already identified the mismatch cost parameter t as a measure of competition intensity. A low t boosts quality provision and - to dampen

[17] This result is clearly dependent on the mode of competition. If we allow the firms (hospitals) to compete on prices, and not qualities, the opposite result would apply (see Schultz (2002)).

this effect - hospitals locate further apart. Finally, an increase in the quality cost parameter k reduces quality competition, resulting in less product differentiation. When inserting (4.9) into (4.5) we obtain the equilibrium quality levels of the game:

$$q^*(\lambda, p) = \left(\frac{p^2\lambda^2}{16tk^2}\right)^{\frac{1}{3}}. \tag{4.11}$$

4.3.2 Social Welfare

Consider a social planner who aims at maximizing social welfare. Assuming symmetry in qualities and locations the social welfare function is given by

$$W = v + q - 2kq^2 - \frac{t}{12} + \frac{t}{4}\Delta(\lambda - \Delta). \tag{4.12}$$

Note that we consider that acquiring information about the market is costless, i.e., gatekeeping involves no costs.[18] We will relax this assumption when endogenizing λ in Section 4.4.

4.3.2.1 The second-best optimum

Let us first consider the case where λ cannot be regulated by the social planner at all. In this sense the solution derived here may be called a 'constrained first-best', or simply the second-best. Quality provision is second-best efficient when

$$q^{sb} = \frac{1}{4k}. \tag{4.13}$$

Maximizing the last term of equation (4.12) yields $\Delta^{sb} = \lambda/2$, which determines the second-best efficient specializations

$$x_1^{sb} = \frac{1}{2} - \frac{\lambda}{4} \text{ and } x_2^{sb} = \frac{1}{2} + \frac{\lambda}{4}. \tag{4.14}$$

The regulator faces the following fundamental trade off: on the one hand, the mismatch costs incurred by the informed patients are minimized when hospitals locate at $\frac{1}{4}$ and $\frac{3}{4}$, respectively. These locations would obtain when the entire population is informed, $\lambda = 1$. On the other hand, as

[18] If we interpret p as a per patient (or per treatment) reimbursement from a government agency, this particular specification of social welfare also relies on the assumption that the third party (i.e., the regulator) is able to raise the necessary funds in a non-distortionary manner.

the uninformed patients choose a provider randomly, their mismatch costs are at a minimum when hospitals do not specialize and agglomerate at the market center, i.e., at $\frac{1}{2}$. Minimum differentiation would obtain for $\lambda = 0$. Balancing these opposing effects leads to locations $x_1^{sb} \in \left[\frac{1}{4}, \frac{1}{2}\right]$ and $x_2^{sb} \in \left[\frac{1}{2}, \frac{3}{4}\right]$.

4.3.2.2 The first-best optimum

Now consider that the regulator has the available option of introducing a strict gatekeeping regime, which amounts to setting $\lambda = 1$. From equation (4.12) it is easily seen that $\partial W / \partial \lambda \geq 0$ for all feasible values. The social planner would thus implement a strict gatekeeping regime whenever $\Delta > 0$. The first-best solution is consequently given by[19]

$$\lambda^{fb} = 1,\ q^{fb} = \frac{1}{4k},\ x_1^{fb} = \frac{1}{4} \text{ and } x_2^{fb} = \frac{3}{4}. \tag{4.15}$$

4.3.3 Gatekeeping

The aim of this subsection is to show that introducing strict gatekeeping, i.e., setting $\lambda = 1$, is not necessarily socially beneficial when the price is exogenously given. This may be surprising at first sight since strict gatekeeping implies that additional information is acquired. Taking the competitive effects into account it may turn out that - although gatekeeping is costless - strict gatekeeping is harmful from a social welfare perspective. The relationship between social welfare and the share of informed patients is given by the following proposition:

Proposition 4.1 *For an exogenously given price, social welfare is maximized at (i) $\lambda = 1$ if mismatch costs are sufficiently high, (ii) $\lambda \in (0, 1)$ if mismatch costs are sufficiently low and quality costs are sufficiently high.*

Proof. Inserting (4.9) and (4.11) into (4.12) yields a welfare function $W(p, \lambda)$. We can easily calculate

$$\frac{\partial W}{\partial \lambda} = (2p)^{\frac{1}{3}} \left(\frac{1}{8} \left(\frac{2p}{\lambda t k^2} \right)^{\frac{1}{3}} - \frac{1}{6} \left(\frac{\lambda}{t^2 k} \right)^{\frac{1}{3}} (2p - t) \right) \tag{4.16}$$

[19]Notice that the solution $\Delta = 0$ and $\lambda = 0$ is always dominated by $\Delta = 1/2$ and $\lambda = 1$.

and

$$\frac{\partial^2 W}{\partial \lambda^2} = -(2p)^{\frac{1}{3}} \left(\frac{1}{24} \left(\frac{2p}{\lambda^4 t k^2} \right)^{\frac{1}{3}} + \frac{1}{18} \left(\frac{1}{t^2 \lambda^2 k} \right)^{\frac{1}{3}} (2p - t) \right) \qquad (4.17)$$

(i) We have that $\frac{\partial W}{\partial \lambda} > 0$ for all permissible values of λ if $t > 2p$. (ii) Assume that $t < 2p$. In this case we have that $\frac{\partial W}{\partial \lambda} > (<) 0$ if $k < (>) \overline{k} := \frac{27pt}{32\lambda^2(2p-t)^3}$. Since $\lim_{\lambda \to 0} \overline{k} \to \infty$ and $\frac{\partial^2 W}{\partial \lambda^2} < 0$ it follows that social welfare is maximized for a unique value of λ that lies strictly between 0 and 1 if $k > \frac{27pt}{32(2p-t)^3}$. ∎

Ceteris paribus, more informed patients lead to more intense competition between the hospitals, which implies a higher provision of quality and more differentiation. If mismatch costs are high, the degree of competition between hospitals is low, which further implies that the incentives for horizontal differentiation are also low. In this case from a welfare point of view there is underprovision of quality and an insufficient degree of differentiation. A larger share of informed patients would thus increase efficiency with respect to both quality provision and horizontal differentiation.

However, more informed patients could lead to *excessive competition* if mismatch costs are sufficiently low. If, in addition, quality costs are sufficiently high, so that first-best quality provision is relatively low, a fully informed market would lead to both excessive differentiation *and* overprovision of quality. This could be sufficient to outweigh the benefits of increased patient information on aggregate mismatch costs, implying that social welfare is maximized in a situation where not all patients are fully informed.

The welfare implications of introducing a strict gatekeeping regime follows immediately:

Proposition 4.2 *For an exogenously given price, introducing a strict gatekeeping regime is detrimental to social welfare if (i) mismatch costs are sufficiently low, (ii) quality investments are sufficiently costly, or (iii) the fraction of a priori informed patients is sufficiently high.*

In other words, costless gatekeeping can reduce social welfare due to excessive competition between health care providers. In order better to illustrate the main mechanisms behind this result we provide a numerical example.

4.3.3.1 A numerical example

Let $p = 1$ and $k = 1$. Then the remaining parameters of the model are t and λ.[20] We illustrate the case of fairly intense competition with $t = 1/2$ (Case 1) and moderate competition with $t = 3/2$ (Case 2). In Table 4.1 we present the outcome of the location-quality game, with the associated level of social welfare, for different values of λ.

	Case 1, $t = 1/2$			Case 2, $t = 3/2$		
λ	q^*	Δ^*	$W^* - v$	q^*	Δ^*	$W^* - v$
0.1	0.11	0.46	0.022	0.07	0.22	−0.072
0.2	0.17	0.58	0.043	0.12	0.28	−0.043
0.3	0.22	0.67	0.051	0.16	0.32	−0.021
0.4	0.27	0.74	0.051	0.19	0.35	−0.002
0.5	0.31	0.79	0.046	0.22	0.38	0.015
0.6	0.36	0.84	0.035	0.25	0.41	0.030
0.7	0.39	0.89	0.021	0.27	0.43	0.043
0.8	0.43	0.93	0.003	0.30	0.45	0.054
0.9	0.47	0.97	−0.018	0.32	0.46	0.065
1	0.5	1	−0.042	0.35	0.48	0.075

Table 4.1: Equilibrium outcomes for $p = 1$ and $k = 1$.

As can be seen from equation (4.5), quality competition is intense for low values of the mismatch cost parameter t. Thus, hospitals provide higher quality in Case 1 than in Case 2. To mitigate costly quality competition, hospitals aim at making their products less substitutable. This incentive is clearly higher in Case 1, partially offsetting the competition effect. In Case 1, increasing the share of informed patients is beneficial for low values of λ. Besides the net benefits derived from higher quality provision, patients may also gain from reduced mismatch costs. As λ increases, though, the centrifugal force drives hospitals further away from the market center, combined with an increase in quality provision. At $\lambda = 0.4$ we see that there are both overprovision of quality *and* too much differentiation, compared with the first-best solution, implying that a further increase in λ unambiguously reduces welfare.[21] In fact, since $\lim_{\lambda \to 0} (W^* - v) = 0$ we see that imple-

[20] Of course v is another not yet specified parameter. The actual size is irrelevant for the model (as long as v is sufficiently large), so we will keep this general.

[21] As hospitals still specialize within the disease space, our example shows that Propo-

menting a strict gatekeeping system would be socially detrimental even if there are no informed patients to begin with. This changes when Case 2 is considered, where moderate specialization incentives are at work. In this case it pays to generally demand a GP referral.

4.3.4 Price regulation

The results of the previous section hinge on the assumption that the price is exogenous. We will now consider the case where the regulator is able also to use the price as a regulatory instrument. Assuming second-best price regulation, the following result obtains:

Proposition 4.3 *With second-best price regulation, introducing a strict gatekeeping system always improves social welfare.*

Proof. Again, inserting (4.9) and (4.11) into (4.12) yields the welfare function $W(p,\lambda)$. By defining $\widehat{p} := p\lambda$ we can define a new welfare function $\widehat{W}(\widehat{p},\lambda)$. Maximizing $W(p,\lambda)$ with respect to p and λ is then equivalent to maximizing $\widehat{W}(\widehat{p},\lambda)$ with respect to $\widehat{p}$ and λ. Taking the partial derivative with respect to λ yields $\frac{\partial \widehat{W}(\widehat{p},\lambda)}{\partial \lambda} = \frac{t}{4}\left(\frac{\widehat{p}}{4t^2k}\right)^{\frac{1}{3}} > 0$. Thus, social welfare is maximized by setting $\lambda = 1$. ■

From (4.9) and (4.11) we know that p and λ have identical effects on equilibrium differentiation and quality provision. Thus, by using the price instrument properly, the regulator can induce exactly the same location-quality outcome for any given value of λ. Consider an increase in the share of informed patients in the market. The resulting effects - stronger quality competition and larger differentiation - can be exactly offset by reducing the price accordingly. This would, however, have an unambiguously positive effect on social welfare - even though differentiation and quality provision remain unchanged - since expected aggregate mismatch costs are reduced when fewer patients run the risk of attending the 'wrong' hospital. Thus, the regulator can maximize social welfare by introducing a strict gatekeeping system in order to make all patients fully informed, and then use price regulation to correct for the potential negative effects of increased information.[22]

---

sition 4.1 does not rely on the assumption that hospitals are allowed to locate outside the disease space.

[22] A more detailed discussion of the optimal second-best price in the case of a strict gatekeeping system is presented in Brekke et al. (2002).

4.4 Indirect gatekeeping

In this section we endogenize the share of informed patients, λ. We assume that patients have the choice of consulting a gatekeeping GP, thereby obtaining all relevant information, before accessing the hospital market. To obtain an interior solution for λ we consider cost heterogeneity with respect to GP consultation. Let $y \in [0,1]$ denote the cost type of a patient. The associated costs are assumed to be ay, where $a > 0$. This heterogeneity can simply be justified by an opportunity cost argument, e.g., by varying time costs due to different wage earning abilities. To simplify the analysis we assume that patient types are uniformly distributed on the disease space S.

The GP consultation decision is based on the expected benefits of gatekeeping relative to a patient's cost type. Benefits are in expected terms as prior to consultation none of the patients can observe specialization, quality and disease. So, in a game-theoretic sense, the consultation decision is simultaneous to specialization and quality decisions. Since hospitals cannot observe patients' consultation decisions, and since they do not know a patient's cost type, they have to form beliefs about the actual consultation rate. We are solving for the perfect Bayesian equilibrium where expectations will be confirmed. Additionally, we require that beliefs have to be consistent out of equilibrium. This restriction is discussed in some detail below.

4.4.1 The specialization-quality game

4.4.1.1 The demand for secondary care and GP consultation

The demand for hospital 1 is exactly the same as in the previous section for a given share λ of informed patients. So for this subsection it remains to determine the consultation decision.

When deciding whether to approach a (randomly chosen) hospital directly or to consult a gatekeeping GP first, a patient has to weigh the costs of going to a GP against the benefits. Imposing the same symmetry assumption as previously, the quality of hospital care is unimportant for this problem. The quality received is independent of whether a GP was consulted or not. Determining the benefits of gatekeeping simply requires ascertaining the expected reduction in mismatch costs. This requires to forming expectations Δ^e about the degree of product differentiation Δ in the market. We will assume that these expectations are symmetric, which seems plausible as patients are (except for consultation costs) ex-ante identical. As before we will assume that patients know that the equilibrium will be symmetric, i.e., that hospitals locate equidistantly from the market centre, but on opposite sides.

The expected mismatch costs when directly approaching a hospital are

$$MMC_{DA}^{e} = \frac{t}{2}\int_{0}^{1}\left(z-\frac{1}{2}\left(1-\Delta^{e}\right)\right)^{2}dz + \frac{t}{2}\int_{0}^{1}\left(z-\frac{1}{2}\left(1+\Delta^{e}\right)\right)^{2}dz. \tag{4.18}$$

When consulting a GP first, expected mismatch costs are reduced to

$$MMC_{GP}^{e} = t\int_{0}^{\frac{1}{2}}\left(z-\frac{1}{2}\left(1-\Delta^{e}\right)\right)^{2}dz + t\int_{\frac{1}{2}}^{1}\left(z-\frac{1}{2}\left(1+\Delta^{e}\right)\right)^{2}dz. \tag{4.19}$$

The expected benefit of gatekeeping, $B^e := MMC_{DA}^e - MMC_{GP}^e$, is thus

$$B^{e} = \frac{t\Delta^{e}}{4}. \tag{4.20}$$

The equilibrium value of λ is obtained by equating the expected benefits of gatekeeping to its actual costs, $t\Delta^e/4 = a\lambda$, and solving for λ, yielding

$$\lambda = \frac{t\Delta^{e}}{4a}. \tag{4.21}$$

The comparative static results are intuitive: the higher the costs of consulting a GP, a, the lower the share of patients actually going to a GP. The benefits of gatekeeping are increasing in mismatch costs, since some costs may be avoided by obtaining information. Aggregate expected mismatch costs are determined by two different factors. For any given positive distance between the hospitals, these costs are obviously increasing in the mismatch cost parameter t. In addition, expected aggregate mismatch costs are increasing in the degree of horizontal differentiation. The further apart the hospitals are located, the more costly, in terms of mismatch costs, to attend the 'wrong' hospital.

4.4.1.2 Quality competition and specialization

We now assume that for *any* patient consultation strategy hospitals' expectations about the fraction of informed consumers is equal to the actual fraction induced by that strategy. In other words, we require beliefs to be consistent out of equilibrium. With this a hospital's best response against patients strategies can be written in terms of the fraction of informed patients. Notice that the restriction on strategies that the lowest cost types demand GP referral, as used in equation (4.21), is irrelevant from a hospital perspective. Only the expected share of informed patients matters and not their actual composition.

The equilibrium of the quality subgame is thus simply obtained by substituting λ by λ^e in equation (4.5), yielding

$$q^{**}\left(\Delta, \lambda^e; p\right) = \frac{p\lambda^e}{4tk\Delta}. \tag{4.22}$$

The comparative static properties are comparable to those with direct gatekeeping. The same applies to product differentiation, which is obtained in a similar fashion:

$$\Delta^{**}\left(\lambda^e; p\right) = \left(\frac{p\lambda^e}{4t^2k}\right)^{\frac{1}{3}}. \tag{4.23}$$

4.4.1.3 The solution of the game

Imposing rational expectations, we obtain the solution of the game. One requirement is that hospitals' expectations are confirmed, $\lambda^e = \lambda$, and the other that patients' expectations are confirmed, $\Delta^e = \Delta$. By inserting equation (4.21) into equation (4.23), and solving for Δ, we obtain equilibrium product differentiation

$$\Delta_u^{**} = 0 \text{ and } \Delta_s^{**} = \frac{1}{4}\left(\frac{p}{tka}\right)^{\frac{1}{2}}, \tag{4.24}$$

where the subscript 'u' indicates that the equilibrium is unstable, and 's' that it is stable. This will be shown in the proposition below.

The corresponding quality levels are obtained by substituting equation (4.23) into (4.22), taking the relationship in equation (4.21) into account, yielding

$$q_u^{**} = 0 \text{ and } q_s^{**} = \frac{p}{16ak}. \tag{4.25}$$

Using the equilibrium product differentiation displayed in equation (4.24), we can solve for the share of GP patients by substituting the respective values into equation (4.21):

$$\lambda_u^{**} = 0 \text{ and } \lambda_s^{**} = \frac{1}{16}\left(\frac{pt}{ka^3}\right)^{\frac{1}{2}}. \tag{4.26}$$

Proposition 4.4 *With endogenous GP consultation there are two symmetric equilibria of the specialization-quality-consultation game. (i)* $S_u^{**} =$

$(\Delta_u^{**}, q_u^{**}, \lambda_u^{**})$, and (ii) $S_s^{**} = (\Delta_s^{**}, q_s^{**}, \lambda_s^{**})$, where $\Delta_i^{**}, q_i^{**}, \lambda_i^{**}$, $i = u, s$, are given by equations (4.24), (4.25), and (4.26). The equilibrium given in (i) is unstable and the equilibrium shown in (ii) is stable.

Proof. (i) That S_u^{**} is an equilibrium of the game is straightforward. Consider that all patients expect that the hospitals will not differentiate, $\Delta^e = 0$. Given these expectations, the benefits of gatekeeping are zero, $B^e = 0$. Since there are positive costs of consulting a GP, nobody will actually go to a gatekeeper, i.e. $\lambda = 0$. Hospitals correctly anticipate these expectations: they know that patients are completely uninformed and thus not responsive to quality investments. Consequently, hospitals set $q_u^{**} = 0$. Hospitals do not incur any costs in this scenario so they cannot do better than confirm expectations on specialization and agglomerate at the same point. However, this equilibrium is unstable when expectations have to be consistent out of equilibrium. Then, wrong expectations on λ on the hospitals' side directly translate into wrong expectations on Δ when taking equation (4.21) into account. Specialization is then

$$\Delta^{**} = \left(\frac{p\Delta^e}{16atk}\right)^{\frac{1}{3}}. \tag{4.27}$$

Now consider that $\Delta^e \in (0, \Delta_s^{**})$. Since $\Delta^{**}(\Delta_s^{**}) = \Delta_s^{**}$ and Δ^{**} is increasing and concave in Δ^e, this implies $\Delta^{**}(\Delta^e) > \Delta^e$. As expectations are not confirmed, Δ^e cannot be an equilibrium of the game. Moreover, since actual differentiation exceeds expectations, there is no force driving expectations back to zero, proving instability.

(ii) Here it remains to show that the equilibrium is stable. First, assume that $\Delta^e \in (0, \Delta_s^{**})$, then - as above - we have that $\Delta^{**}(\Delta^e) > \Delta^e$, driving expectations back to Δ_s^{**}. Second, consider that $\Delta^e > \Delta_s^{**}$; then we have $\Delta^{**}(\Delta^e) < \Delta^e$. This not only proves that $\Delta^e > \Delta_s^{**}$ can never be an equilibrium, but also that expectations will be driven back towards Δ_s^{**}. ■

Requiring consistent beliefs out of equilibrium enables us to prove instability of S_u^{**} and stability of S_s^{**}. Although it seems plausible to concentrate on S_s^{**} note that S_u^{**} can easily be supported as a stable equilibrium when consistency is not imposed. Consider that hospitals expect that nobody will become informed, $\lambda^e = 0$, independent of what patients' strategies would actually suggest. Then hospitals do not differentiate and there are zero benefits of gatekeeping. As patients correctly anticipate missing specialization incentives, $\lambda^e = 0$ will always be confirmed.

As we concentrate on the stable equilibrium, we ease notation by suppressing the index 's' in the remainder of the chapter. In order to analyze the comparative static properties of the equilibrium it is convenient to neglect the restriction $\lambda^{**} \in [0,1]$. We will be more precise about that later.

The share of the population going to a gatekeeping GP increases in the mismatch cost parameter, t, as this drives up the benefits of gatekeeping. It also increases in price. This is an indirect effect stemming from specialization. Price increases boost quality competition and, to dampen this effect, hospitals aim at reducing the substitutability of their products, increasing the benefits of gatekeeping. Obviously, λ^{**} is a decreasing function of a. The higher the disutility incurred by consulting a GP the lower the share of patients who actually consult one. Finally, an increase in the quality cost parameter, k, reduces quality competition and thereby differentiation incentives. This, in turn, reduces the benefits of gatekeeping.

Compared to the previous section, there are two major differences in the comparative static properties of quality. Firstly, the mismatch costs parameter has no effect. With direct gatekeeping, patients were more responsive to quality investments at lower values of t, amplifying quality competition. As can be seen from equation (4.22), this is also true here, but this effect is opposed by the consultation effect. A lower t reduces the benefits of gatekeeping, resulting in a less competitive market. With linear costs of GP consulting these two effects exactly offset. Secondly, an increase in consulting costs, a, lowers λ^{**} and thereby softens quality competition. The latter effect is also present with respect to the specialization decisions: high consulting costs weaken the competition effect, making product differentiation less desirable.

4.4.2 Social welfare

Subtracting the consulting cost term $a\int_0^\lambda x dx$, the social welfare function can be rewritten from equation (4.12),

$$W = v + q - 2kq^2 - \frac{t}{12} + \frac{t}{4}\Delta(\lambda - \Delta) - \frac{1}{2}a\lambda^2. \tag{4.28}$$

The first order conditions for the first-best solution are obtained by differentiation, yielding

$$\Delta^{fb} = \frac{\lambda}{2},\ q^{fb} = \frac{1}{4k},\ \text{and}\ \lambda^{fb} = \frac{t\Delta}{4a}. \tag{4.29}$$

First-best quality is identical to the previous analysis, and depends only on the costs of quality investments. Specialization is the second-best from Section 4.3.2.1, and is thus conditional on the share of GP patients. Most importantly, although not surprising, the first-best λ coincides with individual decisions, so there is no need to regulate gatekeeping directly.

Proposition 4.5 *The first-best efficient solution of the game with endogenous gatekeeping has quality* $q^{fb} = \frac{1}{4k}$ *and*
(i) $\Delta^{fb} = 0$ *and* $\lambda^{fb} = 0$ *for* $t < 8a$,
(ii) $\Delta^{fb} = \lambda^{fb}/2$ *and* $\lambda^{fb} \in [0, 1]$ *for* $t = 8a$, *and*
(iii) $\Delta^{fb} = \frac{1}{2}$ *and* $\lambda^{fb} = 1$ *for* $t > 8a$.

Proof. The first-best solution in (i) is an interior solution where both first order conditions, $\Delta^{fb} = \frac{\lambda}{2}$ and $\lambda^{fb} = \frac{t\Delta}{4a}$, are satisfied. As λ is restricted to the unit interval there are situations where $\lambda^{fb} = \frac{t\Delta}{4a}$ does not hold, i.e. when parameters are such that λ exceeds one. Considering the unrestricted Hotelling model the first order condition for Δ can always be met. Inserting $\Delta = \frac{\lambda}{2}$ into (4.28) and differentiating yields $\frac{\partial W}{\partial \lambda} = \lambda \left(\frac{t}{8} - a\right)$. For $t > 8a$ the regulator sets λ to its maximum, $\lambda^{fb} = 1$, and $\Delta^{fb} = 1/2$. Obviously, the regulator is indifferent between all feasible values of λ when $t = 8a$. ■

The mismatch cost parameter has two opposing effects on the benefits of gatekeeping. On the one hand, a higher value of t increases expected aggregate mismatch costs for every given pair of hospital locations. On the other hand, an increase in t reduces quality competition and leads to less differentiation, which reduces the benefits of gatekeeping. Using (4.20) and (4.24) it is straightforward to show that the former (direct) effect always dominates, i.e., $\partial B^e/\partial t > 0$, implying that the benefits of gatekeeping are increasing in the mismatch cost parameter t. Consequently, when t is sufficiently low relative to GP consulting costs, $t < 8a$, the benefits of increased information, in terms of reduced expected mismatch costs, are outweighed by the costs of going to a GP.

4.4.3 Price regulation

When setting the optimal price, the regulator trades off inefficiencies along three different dimensions: quality provision, horizontal differentiation and GP consultation. In general, this will not result in a strict gatekeeping regime. The objective function is obtained by inserting the equilibrium values of the specialization-quality-consultation game into the welfare function. The candidate second-best price is then found to be

$$p^{sb} = \frac{t}{8} + 3a. \tag{4.30}$$

The following equilibrium obtains:

$$\Delta^{sb} = \frac{\left[2tka\left(t+24a\right)\right]^{\frac{1}{2}}}{16tka},\ q^{sb} = \frac{24a+t}{128ak},\ \text{and}\ \lambda^{sb} = \frac{\left[\frac{2t(24a+t)}{ka^3}\right]^{\frac{1}{2}}}{64}. \quad (4.31)$$

The price equilibrium given in (4.30) exists if $\left(x_1^{**}\left(p^{sb}\right), x_2^{**}\left(p^{sb}\right)\right) \in Q$. Using (4.31), the existence condition (provided that the solution is interior) is given by

$$k > \frac{24a+t}{1024a^2}. \quad (4.32)$$

It is straightforward to show that the second-best price given by (4.30) is an interior solution for a subset of the parameter values, defined by

$$k > \overline{\overline{k}} := \frac{t\left(24a+t\right)}{2048a^3}.$$

Thus, if $k < \overline{\overline{k}}$ we have a corner solution with $\lambda^{sb} = 1$.[23] Given that $\partial\overline{\overline{k}}/\partial t > 0$ and $\partial\overline{\overline{k}}/\partial a < 0$, the implications for indirect gatekeeping follow immediately:

Proposition 4.6 *Second-best price regulation implies a de facto strict gatekeeping regime if quality costs or GP consulting costs are sufficiently small, or if mismatch costs are sufficiently high.*

The intuition is basically the same as before: gatekeeping is a way of reducing expected aggregate mismatch costs, and the social benefits of gatekeeping are consequently linked to these costs that are increasing in t and decreasing in k. Of course, the role of GP consulting costs is straightforward: the smaller a, the larger share of the population consults a GP to obtain information.

In the following we will only consider interior solutions.[24] Thus, GP consultation will generally be inefficient. The efficiency properties of the second-best interior solution are summarized as follows:

Proposition 4.7 *The second-best (interior) solution of the specialization-quality-consultation game has the following properties: (i) for $t < 8a$ there is too much differentiation given λ^{sb} and inefficiently low quality provision, (ii)*

[23] From (4.32) it follows that an interior solution always meets the existence condition if $a < t/2$.

[24] For a discussion of optimal price regulation in the corner solution (strict gatekeeping), see Brekke et al. (2002).

for $t = 8a$ the first-best optimum is implemented, $\Delta^{sb} = \frac{1}{8}\left(\frac{2}{ak}\right)^{\frac{1}{2}}$, $q^{sb} = \frac{1}{4k}$ and $\lambda^{sb} = \frac{1}{4}\left(\frac{2}{ak}\right)^{\frac{1}{2}}$, (iii) for $t > 8a$ there is insufficient differentiation given λ^{sb} and inefficiently high quality provision.

Proof. First-best specialization, conditional on the share of GP-patients, requires $\Delta^{sb} = \lambda^{sb}/2$. From (4.31) we find that $\Delta^{sb} - \lambda^{sb}/2 = \frac{8a-t}{128}\left[\frac{2(24a+t)}{tka^3}\right] > (<)\, 0$ if $t < (>)\, 8a$. First-best quality is given by $q^{fb} = \frac{1}{4k}$. From (4.31) we have that $q^{sb} - q^{fb} = -\frac{(8a-t)}{128ka} < (>)\, 0$ if $t < (>)\, 8a$. ∎

We have an interior solution if GP consultation costs are sufficiently high. From (4.25) we know that a high value of a implies that quality provision will be relatively low in equilibrium. We also know that a price above marginal costs is in any case a necessary condition to prevent under-provision of quality. If, in addition, mismatch costs are relatively low, the value of obtaining information will be limited and, consequently, GP consulting will be low in equilibrium. Since the first-best efficient level of quality provision is independent of mismatch costs, this implies that social welfare is maximized at a low degree of differentiation. In this case, $t < 8a$, the price that yields first-best differentiation is not high enough to generate efficient quality provision. Thus, higher quality can only be obtained at the expense of excessive differentiation, and these considerations are optimally traded off at a price which yields under-provision of quality and too much differentiation.

On the other hand, if mismatch costs are high, the first-best level of differentiation will be higher - closer to $\frac{1}{2}$ - due to higher GP consultation. In this case, $t > 8a$, the optimal degree of differentiation is obtained at a price that yields over-provision of quality. Consequently, optimal regulation implies accepting a less than optimal degree of differentiation in order to avoid too much over-investment in quality.

4.5 Conclusion

Equipping GPs with a gatekeeper role in the health care system is a major issue in the debate on health care reforms. Among politicians, the conventional wisdom is that gatekeeping contributes to cost control. This is somewhat surprising since evidence is lacking, as was demonstrated in an empirical study by Barros (1998). Economists are more concerned about efficiency arguments rather than fiscal ones. As GPs are usually better informed than patients about the characteristics of the secondary health care market, e.g. about quality and specialization of hospitals, matching of patients to hospitals may indeed be improved by gatekeeping. However, this argument neglects the potential competitive effects in the hospital market. We presented a model that analyzes the competitive effects of gatekeeping in the presence of non-price competition.

While prices were regulated, we allowed for competition in specialization and quality. We considered two versions of the basic model, one in which the share of ex ante informed patients is exogenously given (direct gatekeeping), and another where the share of informed patients is endogenously determined (indirect gatekeeping).

In the direct gatekeeping scenario we assumed gatekeeping to be costless. We found that when the price is exogenously given strict gatekeeping does not necessarily improve social welfare. This is the case when the additional information acquired by GPs boosts competition to such an extend that excessive specialization of hospitals occurs. In these cases, due to the endogeneity of specializations, mismatch costs are higher with gatekeeping than without. This raises doubts about whether gatekeeping improves efficiency. Things change dramatically when allowing for second-best price regulation. We demonstrated that strict gatekeeping, i.e. GP consultation is compulsory before accessing the secondary care market, is always socially desirable. Thus, direct gatekeeping should always be accompanied by proper price regulation.

Gatekeeping was endogenized by introducing cost heterogeneity with respect to GP consultation. Since consultation decisions of patients are the same to what a social planner would implement, there is no need for direct regulation of gatekeeping. GP consultation can be indirectly influenced by price regulation. With second-best price regulation, a strict gatekeeping regime obtains if the benefits of gatekeeping are sufficiently high (improved matching outweighs the potentially negative competitive effects) compared to its costs. When the share of GP patients is below one, the second-best outcome will, in general, be inefficient. Depending on the parameters, there may be too much differentiation and too low quality or vice versa. Direct implementation of a strict gatekeeping regime may again reduce social welfare.

The analysis demonstrates that efficiency gains that are usually attributed to GP gatekeeping cannot be taken for granted when non-price competition is incorporated into the analysis. In the short run, efficiency gains may be obtained by better matches. But quality provision may still be inefficient. In the long run, hospitals will adjust their specialization so that differentiation increases and this counteracts the positive short run effect.

Chapter 5

Sickness Fund Competition in the German Public Health Insurance System: Evidence for Risk Selection?

5.1 Motivation

There have been many health care reforms in Germany during the last decades.[1] Most of these reforms were aimed at cost containment. The movement towards more competition in the health care market began with the Health Care Reform Act 1989. The Health Care Structure Act 1993 introduced more competition among the statutory health insurance companies.

About 90 percent of the German population are insured with statutory sickness funds (in Germany the health insurance companies are called sickness funds), most of them compulsorily. Essentially free choice of sickness funds has been available since 1996. Competition among sickness funds fosters quality competition and thereby improves the quality of care. In addition, patients may benefit from the funds becoming more responsive to their preferences and more cost conscious (Van de Ven and Van Vliet (1992, p. 24)). For equity reasons, sickness funds must not charge risk related premiums, i.e., there is community rating.[2] With community rating, sickness

[1]This chapter is a revised version of Nuscheler and Knaus (2003).

[2]To a large extent health status is not considered to be the responsibility of the indi-

funds have strong incentives to select the low risks as they make profits with low risks and losses with high risks. Note that these incentives stem from regulation rather than competition (Pauly (1984)).

About 400 sickness funds are active in the market. These can be grouped into four categories: regional funds, substitute funds, company-based sickness funds (Betriebskrankenkassen, BKKs), and other funds. The BKKs historically have had (and still have) a much better risk structure than all the other funds resulting in significantly lower contribution rates even after (incomplete) risk adjustment.[3] They gained great many new members after free choice of funds was made available to the insured. A (major) part of that movement is likely to be due to transaction costs, healthy people simply have lower switching costs.[4] We will refer to this form of selection as *passive* risk selection. However, there may be an additional effect over and above these transaction costs. BKKs have a substantially different benefit structure on top of regulated (standard) benefits. They are more likely to provide benefits that healthy individuals demand (e.g. early cancer diagnosis) and less likely to provide those that sick individuals demand (e.g. cancer therapy). There is thus some room for *active* risk selection. We give more details on that below.

Given the sharp increase in BKK members and the different (additional) benefit structures of the BKKs, the main research question is whether the BKKs *actively* engaged in risk selection. Although there are no health care cost figures, the data of the German Socio-Economic Panel (GSOEP) is well suited to answering that question. To disentangle active and passive risk selection we estimate a recursive two-equation model focussing on five transition years from 1995 to 2000. At the first stage we estimate the self-assessed health status using an ordered probit model. In this way we obtain a continuous health index that is used as an explanatory variable at the second stage, where a multinomial logit model is fitted for switching behavior. As the BKKs seem to be the main beneficiaries of competition, we separately analyze BKK members and non-BKK members. The latter have to make a choice from three alternatives: stay with their fund, i.e. not switching at all, switch to a BKK fund, or switch to a non-BKK fund. It turns out that better health significantly reduces the probability of not switching, or, in other words, the sick are more likely to stay with their fund.

---

vidual (see, e.g., Kifmann (2002)). Community rating can thus be seen as an insurance against premium risk, increasing efficiency from an ex ante perspective.

[3] In order to emphasize that health insurance premiums are payroll taxes we refer to them as contribution rates throughout the chapter.

[4] Strombom et al. (2002) found that the price sensitivity is much higher for younger and healthier individuals than for elderly or sick individuals.

Health status has no significant effect on the probability of switching to a non-BKK fund. Thus transitions within the non-BKKs can be attributed to transaction costs.[5] Although the probability of switching to a BKK is positively and significantly affected by health, this provides no evidence for active risk selection by BKKs as both health effects are not significantly different. Switching behavior is thus only driven by switching costs and not by funds actively trying to select the low risks.

To complete the analysis we also investigate the switching behavior of the BKK members. Again, the less healthy are less likely to switch and there is no significant difference in the positive health effects on switching probabilities. We thus find clear evidence that transitions are a result of switching costs. Active risk selection seems to be a negligible problem. The differences in the (additional) benefit packages are too small to have any significant effect. As the flow towards the BKKs is much larger than the flow towards the non-BKKs there is nevertheless a (passive) risk selection problem. Associated distortions are best reduced by improving the risk adjustment scheme, currently on the political agenda.

With the two-stage estimation procedure we reduce the number of health status or health utilization variables from 10 to 1. This enables us to compare the health effects of the different switching types what would otherwise be impossible. Introducing an ordered probit model at the first stage we summarize all health related variables in a single continuous index. Thus, although dimensions are dramatically reduced, we do use all available information. Using health indexes rather than the discrete variables obtained from surveys is common in recent research. Van Doorslaer and Jones (2003), for example, also obtain a continuous health index from an ordered probit regression and use this index to measure inequality in self-assessed health. For a similar approach see Wagstaff and van Doorslaer (1994).

There is some literature on sickness fund competition in Germany. In their recent report, Lauterbach and Wille (2001) analyze whether those individuals who change their sickness fund entail lower health care costs than individuals who stay with their fund. They use micro data provided by several sickness funds and find that changers have a positive effect on profits after risk compensation is carried out. Using a different data set, Jacobs et al. (2002) also find that changers cause lower health care costs than non-changers.

Andersen and Schwarze (1999) analyze the determinants of switching a

[5]This interpretation requires homogeneity of non-BKK funds. Homogeneity among these and among the BKKs is assumed throughout the chapter. We discuss this in some detail in Section 5.6.3.

fund using, as we do, data from the German Socio-Economic Panel (GSOEP). They concentrate on 1997 and 1998. In their single equation probit model, individual health satisfaction affects the probability of changing positively and significantly. At the same time, they obtain a significant negative effect of self-assessed health on the probability of transition. As both variables are measures of the actual health status, the overall effect of health remains unclear. Moreover, the different signs raise doubts about the appropriateness of the specification. In the more recent work of Schwarze and Andersen (2001), the effect of the contribution rate on the probability of switching is analyzed. Using 1999 and 2000 GSOEP data and matched contribution rates, they find that a higher contribution rate in 1999 significantly increases the probability of changing the sickness fund. Since contribution rate data is not available before 1999 we cannot include these in our analysis. As a proxy we use dummies for the different types of health insurance companies. For health status, the effects in Schwarze and Andersen (2001) are ambiguous and insignificant.

All papers cited above have in common that they analyze all changers and do not distinguish between different types of switchers. The reports by Lauterbach and Wille (2001) and Jacobs et al. (2002) can thus only come up with lower average expenditures of switchers. They do not analyze switching behavior at all. This is what Andersen and Schwarze do in both of their papers applying a simple probit model. As they analyze all switchers they are unable to identify risk selection. This would require analyzing subsets of the data separately, i.e. BKKs and non-BKKs. Moreover, to identify active risk selection one needs a reference *switching* group where a multinomial procedure with at least three outcomes is needed instead of an ordinary probit model. Schut et al. (2003) consider, as we do, a multinomial logit model of health plan choice. Analyzing a different data set, they focus on price effects on switching behavior rather than risk selection and find an increasing price elasticity for the German market. This result is mirrored in our year dummies.

Nicholson et al. (2003) apply the switcher methodology to analyze risk selection in the United States. They find that people who switched to a Health Maintenance Organization (HMO) use substantially fewer medical services than in the period prior to switching and less than people who stayed with a non-HMO. This is the other way round for switchers from HMOs to non-HMOs. They thus conclude that HMOs actively select the favorable risks. In our analysis the BKKs play the role of the HMOs. However, we find no evidence for active risk selection.

Our approach contributes to the large literature on adverse selection in health care markets, e.g., Ellis (1989), Cutler and Reber (1998) and Cutler and Zeckhauser (1998).[6] Buchner and Wasem (2003, p. 30) state that "[...] large-scale empirical evidence is missing for 'anti-selection' against bad risks [...]". This is especially true for Germany, as, up to our knowledge, tests whether funds take measures to select the low risks have not been carried out so far. We fill this gap by modelling all relevant transitions between funds.

Since the self-assessed health status plays a central role in our analysis, this chapter also relates to the literature using or explaining self-assessed health. Bound (1991), for instance, uses the self-assessed health status as an explanatory variable in a retirement model. Pohlmeier and Ulrich (1992) derive a health index using self-assessed health. In the recent work of Crossley and Kennedy (2002) the reliability of self-assessed health is addressed. For Australian data, they find a substantial error in rating and that this error is correlated with observable variables like age, gender, and income. If these findings also apply to the GSOEP data, the measurement error may bias the health index obtained from our ordered probit regression.

There are few applications of recursive models in the field of health economics. One exception is the literature on (ex post) moral hazard. Holly et al. (1998), for example, study the relation of health insurance coverage and health care utilization in the Swiss health care market. Their econometric model differs from ours in that they apply a bivariate probit model with an endogenous dummy variable while we consider a recursive model with an ordered probit model at the first stage and a multinomial logit model at the second stage. To ascertain the impact of deductibles on health care use, contract choice must be endogenized since low risks tend to choose contracts with higher deductibles than high risks (see, e.g., Schellhorn (2001)). Another respect in which these papers differ from ours is that they use endogenous regressors, while our health index is considered exogenous.

The chapter is organized as follows: Section 5.2 discusses the different forms of risk selection. Public policy measures taken to prevent risk selection in Germany are described. Some more institutional background is provided in Section 5.3. In Section 5.4 the data set is introduced and descriptive statistics are shown. The empirical model, explained in Section 5.5, is followed by the results in in Section 5.6. Section 5.7 concludes.

[6] For an excellent overview see Cutler and Zeckhauser (2000).

5.2 Risk selection and regulation

5.2.1 Active and passive risk selection

We distinguish two different types of risk selection: active and passive risk selection. We will say that a fund *actively* engages in risk selection when it takes measures in order to attract the low risks, to prevent the high risks from enrolling, or to 'encourage' the high risks to disenroll.

Active risk selection can take two forms: adverse selection and cream skimming. Consider that consumers know their risk type. Then, at any premium level, the high risk consumers demand more health insurance coverage than the low risks. If insurers can only imperfectly observe risk types *adverse selection* obtains. As Wilson pointed out, adverse selection may also arise when insurers are not allowed to use their information (Wilson (1977, pp. 167-168)). Thus, adverse selection may be a result of regulation.

Barros (2003, p. 420) and Van de Ven and Ellis (2000, p. 773) define *cream skimming* as risk selection that occurs because insurers prefer low risks to high risks. Consider that a single insurer has to charge the same premium to all enrollees (community rating) and that these premiums are not risk adjusted.[7] In such an environment there are strong cream skimming incentives. Contingent on the insurers' information Van de Ven and Ellis (2000, pp. 773-774) distinguish three forms of cream skimming. First, if health plans can identify risks, they can directly select the low risks ('direct cream skimming'). They could, for example, reject females and the elderly upon enrollment and thereby obtain a low risk pool consisting of young males. Such a favorable risk pool can also be attained by, e.g., selective advertising and golden handshakes.[8] Second, suppose that insurers cannot observe the risk type but know about other non-observable but relevant, i.e. costly, risk factors. Examples are diseases like cancer or AIDS. Insurers can prevent these high risks from enrolling by providing bad service and bad therapy to both cancer and AIDS patients. With selective contracting insurers will contract with providers with bad reputation in both fields. Third, if the insurer cannot observe the risk type and has no idea about any relevant risk factor, he can still select the low risks by offering low-option plans with (high) deductibles or by reducing offices and information centers to a minimum. In the last two versions adverse selection is used as a cream

[7] When premiums are risk adjusted, risk selection must be interpreted in terms of premium risk groups. Risk selection may then occur within these groups.

[8] If insurers were allowed to use their information for risk rating premiums, there would be no, or at least weakened, (direct) cream skimming incentives (Pauly (1984)).

skimming device. We refer to them as 'indirect cream skimming'.[9]

However, risk separation may obtain even if no fund actively selects the low risks, i.e., there may be *passive* risk selection. Consider two funds L and H with identical benefits. Moreover assume that L has a better risk structure and can therefore offer a lower contribution rate. Take two members of fund H who are identical except that one is healthy and the other is sick. Both consider switching to L and they would identically benefit from lower contributions. Whether they actually switch to L depends on switching costs. These are likely to be higher for sick people. Sick people have better things to do than changing their fund, e.g. undergo treatments. Additionally, they are more likely to be bad risks. Through repeated consumption of medical services they know their fund quite well and, for example, know how to get costs reimbursed. This information would be lost if they switched (health care as well as health insurance are experience goods). Consequently, high risks are less likely to switch. If the (average) risk of the switchers is below the average of L members, the risk differential between the funds increases without any risk selection activities by L.

Our empirical model disentangles active and passive risk selection but is unable to distinguish between the different forms of active risk selection. Whether adverse selection or cream skimming is present depends on the information available to insurers and consumers. As this is hardly observable for researchers, the distinction, although important, is mainly theoretical.

5.2.2 Regulation of sickness fund competition in Germany

Risk selection has adverse effects which counter the positive effects of competition. This is why there are usually a number of measures taken to prevent risk selection, to reduce distortions of risk selection, and to remove the incentives for risk selection. The adverse effects of risk selection include (see, e.g., Van de Ven and Van Vliet (1992, p. 24) and Van de Ven and Ellis (2000, pp. 774-776)): First, the chronically ill may get bad service. Sickness funds will not contract with providers with good reputation in treating chronic diseases in order to prevent attracting these bad risks. They will nevertheless enroll when this negative effect is outweighed by lower contributions.[10] Second, inefficient funds who successfully engage in

[9]Breyer et al. (2003, p. 298), Van de Ven and Van Vliet (1992, pp. 28-31) and Van de Ven et al. (2003, p. 91) report some additional means for indirect risk selection.

[10]This is of little relevance for Germany as there is essentially no selective contracting between insurers and providers. However, a similar argument applies when these funds

risk selection could drive efficient funds out of the market, or at least capture some of their market. Third, cream skimming is itself costly. Since there are no positive welfare effects from reallocating individuals to funds this is pure waste. Fourth, in a Rothschild-Stiglitz world (Rothschild and Stiglitz (1976)), the health insurance market may be unstable and, fifth, the equilibrium, if any exists, will have inefficient coverage.

Annual open enrollment prevents sickness funds from rejecting females and the elderly. The most straightforward version of direct cream skimming is thus not feasible. By contracting with bad gynecologists and shabby sanatoriums favorable risk selection may still be possible. However, selective contracting is pretty much restricted. Funds may engage in activities to indirectly select the low risks. They could, for example, discourage high risks from joining by reducing the number of offices and information centers. Additional means are, e.g., selective advertising, informing high risks about the possibility to switch, and the design of benefit packages.

To reduce the sickness funds' means for selection, there is a standard benefit package that all funds have to provide (Social Code Book V). This amounts to roughly 95 percent of the services (Buchner and Wasem (2003)). It is therefore usually claimed that sickness fund competition is in terms of contribution rates rather than benefits (see e.g. Lauterbach and Wille (2001, p. 29)).

The major measure taken to prevent risk selection is the introduction of risk adjustment schemes. Contributions to every single fund should be adjusted to make it reflect its actual risk structure. Such a scheme was introduced in Germany in 1994, i.e. two years prior to competition. This scheme reduces the incentives for risk selection and, additionally, should prevent sickness funds having comparative advantages from (historically) better risk structures. Moreover, stability of the health insurance market increases. Income, age, gender, sick pay claims, and incapacity to work are used as risk adjusters. As is well known from the literature on risk adjustment, a scheme relying only on simple demographic variables is highly incomplete.[11] Thus substantial incentives remain for risk selection. Not surprisingly, the fear of risk selection as a consequence of sickness fund competition has been expressed quite often (see Lauterbach and Wille (2001, p. 209)). Van de Ven et al. (2003, p. 89) conclude that risk selection is a problem in Germany.

---

cut down their benefit packages. As there is some room to vary the (additional) benefits in Germany this may be more relevant.

[11]See Van de Ven and Ellis (2000) and, for the German case see, Breyer and Kifmann (2001)

5.3 Institutional background

In 1977 the rapid growth of health care expenditure was stopped by the Health Insurance Cost Containment Act. With this reform, an advisory body to the government, the so-called Concerted Action Committee in Health Care, was created. Its major task is to keep contribution rates constant. This essentially means that the increase in health care expenditure is limited to the increase in contributory income (contributions are payroll taxes). Nevertheless, in the 1980s the contribution rates increased from 11.4 percent of gross income to 13 percent.[12] This pressure led to several additional health care reforms. 1989 can be seen as a starting point for introducing more competition into the health care market. Blue collar workers were put on par with white collar workers. They were now allowed to opt out of the statutory health insurance if a certain threshold income is exceeded and to buy private health insurance.[13] At that time, there was no free choice of sickness funds within the public health insurance system. Depending on their profession, members of the regional based funds were allowed to change to substitute funds and other funds (including guild funds, farmers' funds, the miners' fund, and the sailors' fund).[14] They were only allowed to change to a company-based fund (BKK) if they were actually employed in the company the fund was designed for. As a result of this limited competition, the contribution rates of the regional funds were significantly higher than those of the BKKs and the substitute funds (see Figure 5.1).

Sickness funds are not for profit organizations. They have to hold reserves of at least 2 percent and at most 8.5 percent of annual expenditures (Van de Ven et al. (2003, p. 89)). Within that frame, funds are free to set their contribution rate, however, these are subject to approval by the German Federal (Social) Insurance Authority (Bundesversicherungsamt). Note, that employers and employees equally share contributions.

With the Health Care Structure Act of 1993, competition in the health insurance market was intensified. All insured people were allowed to choose their sickness funds freely from 1996 onwards (annual open enrollment). To offset comparative advantages due to risk structure and to reduce selection incentives, a risk compensation scheme was introduced in 1994. Competition and risk adjustment led to an adaptation of the contribution rates of

[12]See Bundesministerium für Arbeit und Sozialordnung, Federal Ministry of Labour and Social Affairs (2001, Table 7.7).

[13]The information presented here and in the following is mainly taken from European Observatory on Health Care Systems, EOHCS (2000, pp. 21-37, 107-116).

[14]The Techniker Krankenkasse (technicians sickness fund), for example, was designed for technicians and engineers only.

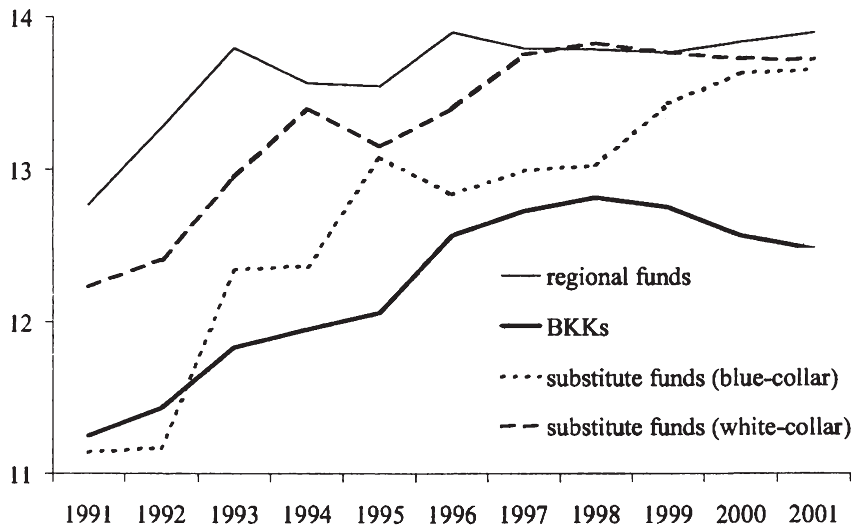

Figure 5.1: Percentage contribution rate averages for the different types of sickness funds. Source: BMG (2001, p. 396).

the non-BKK funds. Unfortunately, we are not able to separate the competition effect from the compensation effect. In 2001 the BKKs departed substantially from the average contribution rate of all other funds. Since 1993 the BKKs have, on average, the lowest contribution rates. As is shown in Table 5.1, competition between the sickness funds has led to a consolidation of the health insurance market. Due to mergers and market exits, the number of funds fell from 1,209 in 1991 by more than two thirds to 396 in 2001.

More interesting for the question addressed here are the transitions from one sickness fund to another. For the sake of presentation, we only show the trend for the members of the three main types of sickness funds, namely, regional funds, substitute funds, and BKKs (see Figure 5.2). The regional funds continuously lost members from 1991 onwards. This trend actually started much earlier. In 1970, 52.4 percent of the statutory insured population in Western Germany were insured by regional funds. This number dropped to 42.8 percent in 1991 and 37.0 percent 2001. This was due to the higher average contribution rates (see Figure 5.1). Members facing high

year	overall	regional funds	company-based funds	substitute funds
1991	1,209	276	721	15
1992	1,223	271	741	15
1993	1,221	269	744	15
1994	1,152	235	719	15
1995	960	92	690	15
1996	642	20	532	15
1997	554	18	457	14
1998	482	18	386	13
1999	455	17	361	13
2000	420	17	337	12
2001	396	17	318	12

Table 5.1: Number of active sickness funds in the German statutory health insurance market, other funds are omitted. Source: BMG (2001, p. 342).

contributions changed to substitute funds when they were allowed to do so. In 1970, 22.9 percent of West Germany's statutory insured population were insured in substitute funds. The share increased to 34.0 percent in 1991, peaked in 1997 (37.1 percent), and then dropped back to 33.9 percent in 2001 (all numbers were taken from BMG, 2001, p. 345).

The increase in members in BKKs, together with the drop of the average BKK contribution rate from 1998 onwards, may be interpreted as an indication that risk selection favors the BKKs. From Figures 5.1 and 5.2 one may conclude that the risk compensation scheme does not fully control for the different risk structures.[15] However, even if the conclusion of risk separation in favor of the BKKs is correct, this is not necessarily due to active risk selection (see Section 5.2). Once again, the main empirical task is to divide up risk selection into active and passive parts.

One of the most straightforward selection instruments is the benefit package. On top of the regulated standard benefits, sickness funds may provide additional benefits. Comparing these for the three main types of

[15] In fact, it is well known that such a simple scheme as that applied in Germany is very incomplete (van de Ven and Ellis (2000)). The federal government plans to improve the scheme by including some morbidity measures in 2007. Risk adjustment is beyond the scope of this chapter. The only thing we need for our interpretation is incompleteness. For more details about risk adjustment in Germany see Buchner and Wasem (2003).

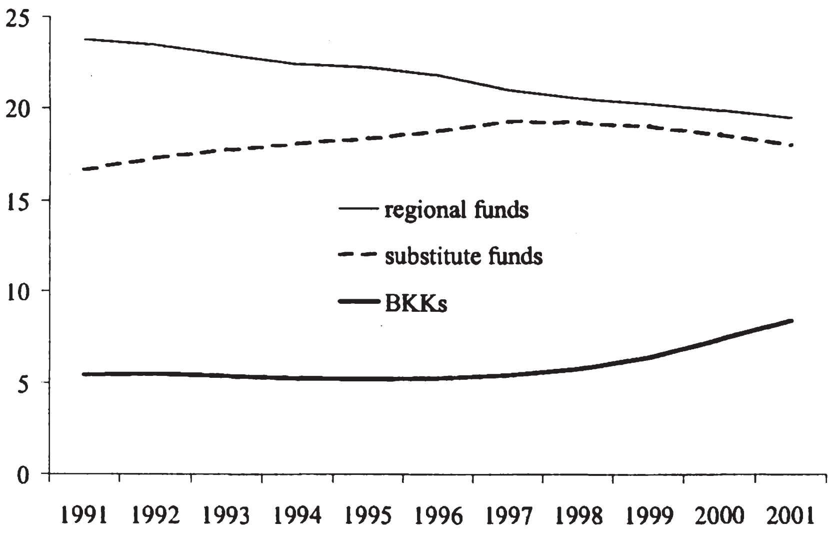

Figure 5.2: Members in 1,000,000 (without their dependants) of the different types of sickness funds. Source: BMG (2001, p. 344).

sickness funds reveals substantial differences (see Table 5.2). Take, for example, health checkups and early cancer diagnosis. Roughly 25 percent of the BKKs provide additional benefits in these fields, fields that are of more interest when actually healthy. Here less than 10 percent of the substitute funds and regional funds provide better than standard care. Things change when looking at benefits that are more likely to meet the preferences of the sick, e.g. chiro therapy, cancer therapy, and homeopathic medicine. The share of BKKs providing these benefits is well below the ones of the regional and substitute funds. These observations are in line with the hypothesis of active risk selection of BKKs by benefit packages. There is anecdotal evidence that funds also use means other than the benefit package for selection, e.g., selective advertising and delaying payments. By explicitly modelling all relevant switching decisions we are able to test whether there are any (successful) selection activities.

Benefit	Percentage of funds providing the benefit		
	regional funds	substitute funds	BKKs
Acupuncture	18.18	16.67	20.79
Anthroposophic medicine	72.73	50.00	12.50
Chiro therapy	90.91	83.33	33.33
Autohemotherapy	9.09	33.33	3.16
Homeopathic medicine	70.00	83.33	25.51
Cancer therapy	63.64	50.00	13.98
Phytotherapy	81.82	63.64	9.68
Oxygen therapy	20.00	25.00	2.11
Naturopathy	75.00	41.67	22.58
Health checkups	9.09	8.33	25.49
Early cancer diagnosis	0.00	8.33	24.51
Health seminars	72.73	36.36	47.37
Logopedia	100.00	63.64	32.35
Nutrition consultancy	90.91	41.67	12.24
Yoga/meditation	30.00	33.33	12.24
N	11 (of 17)	12 (of 12)	103 (of 318)

Table 5.2: Additional benefits (or better than standard benefits) provided by sickness funds in percent of funds that generally provide it. Source: AFW Dienstleitungsgesellschaft, April 2002, own calculations.

5.4 Data

To explore health and health insurance choice in Germany after the natural experiment in 1996 we use the German Socio-Economic Panel (GSOEP). The GSOEP is a representative longitudinal study of private households in Germany. The same private households, persons and families have been surveyed annually since 1984. This micro data panel provides extensive information on the individual characteristics needed to analyze health and health insurance choice in Germany.[16]

The empirical results presented in this study are based on the waves from 1995 to 2000. In 1998 the GSOEP was extended by the Supplementary Sample E. The Sample F was a major extension of the GSOEP in 2000. Both samples are included in our analysis. The different waves are pooled

[16]For more information on the GSOEP see Wagner, Burkhauser and Behringer (1993) and also Projektgruppe Sozio-oekonomisches Panel (1995).

into one sample.

The two dependent variables to be explained by our empirical model are (self-assessed) health status and the switching decision. Health status, $y_1 \in \{0, 1, 2, 3, 4\}$, is of ordinal scale with 5 outcomes, where 0 is bad health and 4 is very good health. To save space we reduce the analysis of the switching behavior of BKK members to a minimum. This is justified as non-BKK members account for the vast majority of switches. For now, the second dependent variable is the switching behavior of the latter. They have to make a choice from three alternatives, $y_2 \in \{0, 1, 2\}$, where 0 indicates no switch, 1 a switch to a BKK and 2 a switch within the non-BKKs. y_2 equals 1 if an individual was enrolled in a non-BKK in one year and in a BKK in the following year. All other switchers in this group must have switched within the non-BKKs. We can therefore use the variable 'change of health insurance' from the GSOEP. Thus y_2 equals 2 if this variable indicates a switch and the individual was not enrolled with a BKK before the switch and thereafter. The switching variable for BKK members is constructed accordingly.

As a sub-sample of these six waves of data we selected only individuals who were not privately insured. Only individuals who were members of the statutory health insurance are included, since they were the only ones who had new incentives to change to a BKK after the 1993 reform. Since family insured members have only limited freedom to choose their health insurance, they were excluded. Finally, we restricted our sample to the working population aged between 25 and 54. This is done in order to exclude special incentives for individuals in the education system and for those close to retirement. For non-BKK members the sample sizes before and after selection are shown in Table 5.3. In this table we also show the percentage of switchers in every year, calculated from the selected sample. From 1995 to 1996, 6.5 percent of the insured switched. This number continuously increased to 10.1 percent in 1999 revealing the remarkable dynamics in the German system. The increase may reflect the fact that the information about the possibility of switching has spread over time. This pattern is mainly due to the increasing flow towards the BKKs. The incentives to enroll with them increased over time, i.e., the contribution rate differential increased (see Figure 5.1). In the transition equation, taking 1995 as the reference year, positive coefficients for the year dummies that increase over time would be expected for the switchers towards the BKKs. Explanations of the variables (Table 5.8) and the entire sample statistics based on the pooled selected sample (Table 5.9) are shown in the Appendix .

In Table 5.10 (see Appendix) we provide some more details about the

year	N full sample	N after selection	Percentage of switchers		
			to BKK	to non-BKK	overall
1995	13,768	1,713	1.23	5.31	6.54
1996	13,511	3,571	1.96	5.15	7.11
1997	13,283	3,431	2.80	4.93	7.72
1998	14,670	3,463	3.03	6.32	9.36
1999	14,085	3,550	4.17	5.94	10.11
overall	69,317	15,728	2.80	5.56	8.35

Table 5.3: Sample selection and the percentage of changers, non-BKK members. Source: GSOEP 1995-2000, own calculations.

relationship between health and switching behavior. Take, for example, the non-BKK members with health status 3 (good). 7,975 out of 15,728 individuals (or 50.71 percent) rate their health as being good. 7,225 people, or 90.6 percent, with good health status did not switch. 50.12 percent of the 14,414 stayers rate their health as being good.

The stayers at the non-BKK funds have substantially worse average health (2.57) than the switchers, where health of out switchers, i.e. switchers to BKKs, have better health (2.76) than within switchers (2.71). There seems to be a stronger health effect for out switchers, which would be in line with our hypothesis of (active) risk selection by BKKs. Individuals with good or very good health are more likely to switch since their proportion of stayers is below the average of all stayers. The likelihood of switching clearly increases with health.[17] Essentially the same is true for the BKK members. Switching out here means that an insured switched to a non-BKK. In contrast to our hypothesis, the health effect is stronger for out switchers. We fitted the same empirical model for BKK members as for non-BKK members. The results show no significant difference in the respective positive health effects on switching probabilities, rejecting not only the (active) risk selection hypothesis for BKKs but also for non-BKKs. As Table 5.4 shows, risk selection may nevertheless obtain.

Finally, we want to shed some more light on the relationship between age, health and switching behavior. Consider, for example, the non-BKK members shown in Table 5.5 below. As already argued above, healthy people

[17]We neglect the numbers obtained for bad health status in our interpretation. Due to the few observations, numbers are very noisy.

Member type	Non-BKKs Members	Non-BKKs Health	BKKs Members	BKKs Health
Non switchers	14,414	2.57	1,878	2.58
Within switchers	874	2.71	78	2.73
New members	156	2.81	440	2.76
	15,444 (−284)	2.58 (−.00)	2,396 (+284)	2.62 (+.02)

Table 5.4: The market after 5 years of transition activity. Source: GSOEP 1995-2000, own calculations.

are more likely to switch. This may (to are large extent) be due to lower switching costs. A similar argument applies for age. Medical consumption increases with age. So old people are likely to be much better informed about their fund than young people. If they need advice they might know whom to ask and how to get costs reimbursed. This information would be lost when changing the fund, creating higher switching costs for the old. Furthermore, young people may have better access to information. As age is negatively correlated with the probability of switching and as it is one of the main determinants of health, it is not clear whether there is still an effect of health on switching behavior after controlling for age.

Age	Average health	Percentage of Stayers	Percentage of Withins	Percentage of Outs	Average health of Stayers	Average health of Withins	Average health of Outs
25–29	2.90	88.77	7.24	3.99	2.89	2.97	2.92
30–34	2.72	90.38	6.12	3.50	2.71	2.77	2.85
35–39	2.62	90.68	6.78	2.54	2.62	2.70	2.74
40–44	2.48	92.84	4.84	2.32	2.47	2.55	2.62
45–49	2.37	93.76	4.24	2.00	2.35	2.47	2.71
50–54	2.27	94.98	2.95	2.07	2.27	2.44	2.34
overall	2.58	91.65	5.56	2.80	2.57	2.71	2.76

Table 5.5: Age, health status and switching decisions of non-BKK members. Source: GSOEP 1995-2000, own calculations.

5.5 The empirical model

We estimate a recursive two equation system with an ordered probit model for health status at the first stage and a multinomial logit model for switching behavior at the second stage.[18] The (self-assessed) health status can take 5 values, $y_1 \in \{0,1,2,3,4\}$, where $y_1 = 0$ is bad health and $y_1 = 4$ is very good health. As the order of different health outcomes can be interpreted but the distance cannot, we fit an ordered probit model,

$$y_1^* = \alpha' x_1 + \epsilon_1. \tag{5.1}$$

The latent variable, y_1^*, is unobservable. Instead we observe $y_1 = j$ if $\mu_{j-1} < y_1^* \leq \mu_j$, $j = 0, ..., 4$, where $\mu_0 = 0$ is a convenient normalization. For this simple notation to be correct we have to set $\mu_{-1} = -\infty$ and $\mu_4 = \infty$. The probabilities are given by $\text{Prob}(y_1 = j|x_1) = \Phi(\mu_j - \alpha' x_1) - \Phi(\mu_{j-1} - \alpha' x_1)$, $j = 0, ..., 4$, where Φ denotes the standard normal cumulative distribution function. Estimates for the parameter vector α as well as for the threshold values μ_1, μ_2 and μ_3 are obtained by maximum likelihood. This regression yields a continuous (fitted) health index that will be used as an explanatory variable at the second stage.

The switching behavior of the non-BKK members is described by

$$y_2 = \begin{cases} 0 & \text{no switch} \\ 1 & \text{switch to BKK} \\ 2 & \text{switch to non-BKK} \end{cases} . \tag{5.2}$$

We fit a multinomial logit model, where the probabilities are given by (Greene (1997, p. 915))

$$P_{ij} := \text{Prob}(y_2 = j|x_{2i}, y_{1i}^*) = \frac{\exp(\beta_j' x_{2i} + \gamma_j y_{1i}^*)}{\sum_{k=0}^{2} \exp(\beta_k' x_{2i} + \gamma_k y_{1i}^*)}. \tag{5.3}$$

Note that we consider the error terms to be uncorrelated. Unique parameterization requires a normalization and we set $\beta_0 = \gamma_0 = 0$. For identification of the model, there must be at least one variable in x_1 that is not in x_2 in order to obtain some variation for the estimation of γ_1 and γ_2. Our identifying assumption is that all objective measures of health, i.e. IMPAIR, DISAB, DOCTOR, VISITS, HOSPITAL, SICK6, SPORTS, only affect health status and do not directly affect switching behavior (see Table

[18] For an excellent introduction to both models see, e.g., Wooldridge (2002, pp. 497-509)

5.8 for the explanation of variables). x_2 may contain a number of additional variables not included in x_1.

Due to the arbitrarily chosen reference category, the absolute values of the parameter estimates are meaningless. Only the differences can be interpreted. The log-odds ratios are given by

$$\ln\left(\frac{P_{ij}}{P_{ik}}\right) = x'_{2i}(\beta_j - \beta_k) + \widehat{y}_{1i}(\gamma_j - \gamma_k), \tag{5.4}$$

where $\widehat{y}_{1i}$ denotes the fitted value of y^*_{1i}. Nevertheless interpretation remains difficult. Things get easier when looking at the marginal effects. Define $\widetilde{\beta}' := (\beta' \quad \gamma)$ and $\widetilde{x}_2 := (x_2 \quad \widehat{y}_1)$, then the marginal effect of the attributes are given by (Ronning (1991, p. 42))[19]

$$\frac{\partial P_{ij}}{\partial \widetilde{x}_{2li}} = P_{ij}\left(\widetilde{\beta}_{jl} - \sum_{k=0}^{2} P_k \widetilde{\beta}_{kl}\right). \tag{5.5}$$

Note that we use the 'standard' marginal effects of the multinomial logit model. We thus consider $\partial\widehat{y}_1/\partial x_2 = 0$. Estimation is by maximum likelihood yielding consistency.

Let δ_j denote the marginal health effect on the probability that a non-BKK member chooses action $j = 0, 1, 2$. The marginal effects for BKK members are η_j. Note that for BKK members $y_2 = 1$ also denotes an "out-switch", i.e. a switch to a non-BKK. Then risk selection activities of BKKs are detected if $\delta_1 > \delta_2$ or $\eta_2 > \eta_1$.

The main advantage of our model is the following: there is only one health variable to interpret in the second equation. One could estimate, and we do, a model with all health care utilization variables in the second equation as well as some health dummies. Then there would be 10 'health effects' and interpretation would require that the health effects of one choice dominate all the other. As can be seen from our estimates this is clearly not the case (see Tables 5.11 and 5.12 in the Appendix). There are simply too many dimensions. We reduce dimensions to 1 by using the health index obtained from an ordered probit regression. Note that we nevertheless use all the information available although the number of dimensions is dramatically reduced.

In principle, our first stage estimates can suffer from a simultaneity bias if health status and the objective measures of health or health care

[19] Indexes: $i = 1, \ldots,$ N is an index for individuals, $j = 0, 1, 2$ is an index for the choice at stage 2, and $l = 1, \ldots,$ K is the l-th variable of $\widetilde{x}_2$.

utilization are determined at the same time. We are optimistic that this is not the case. Self-assessed health status measures health at present. At present in absolute terms means in February/March of the respective year (interviewers are in the field in that period of time). As can be seen from Table 5.8, the health care utilization measures, i.e., VISITS, DOCTOR, HOSPITAL, and SICK6, measure utilization within the last three months or, even further in the past, during the last year. So health status may be seen as the outcome of these variables.[20] It may indeed be the case that an individual is referred to a hospital due to his poor health status and that individual health is improved by hospital treatment. However, this is not the focus of the chapter as transitory health effects are measured. When analyzing risk selection the permanent component of health status is of particular interest. An individual with a hospital stay in the year prior to the interview may still have, and, as our analysis shows, does have a lower health status. We do not see any reason for why simultaneity should be a problem with IMPAIR and DISAB. We admit that this is less clear with SPORTS as poor health may prevent doing sports at all. We nevertheless included it in our analysis as active sport can also be interpreted from the health production perspective (Grossman (1972)).

The health status variable itself requires some more discussion. On the one hand, we consider health to be exogenous in the transition equation. This seems plausible since health status is measured more than 6 months prior to the switching decision. Endogeneity would require some unobservable variables that influence health and are somehow related to switching behavior. We do not see any variable that is linked to both.[21] On the other hand, and maybe more seriously, there is the potential measurement error of health. As nothing is known about the reliability of self-assessed health in the GSOEP we neglect that problem.

5.6 Results

Before interpreting the outcome of the recursive model, let us briefly examine the single equation results for non-BKK members where a multinomial logit model is fitted with 10 health related explanatory variables and a number of controls. Parameter estimates are shown in Table 5.11. The marginal

[20] Analyzing GSOEP data, Pohlmeier and Ulrich (1992) also use health care utilization measures as explanatory variables for individual health.

[21] Foster (1997) demonstrated for a binary logistic regression that the results of the 'Pseudo Instrumental Variable' estimator that we use and the Generalized Methods of Moments estimator only differ little even if there is a substantial endogeneity problem.

effects, shown in Table 5.12, reveal that it is impossible to figure out a clear health effect (both tables are in the Appendix). At first glance, it seems as if the effect of health on within switchers ($y_2 = 2$) is larger than for the out switchers ($y_1 = 2$): the marginal effects are higher for all health dummies.[22] However, it is the other way round for the health care utilization measures DOCTOR, VISITS, and HOSPITAL as well as for the more permanent health characteristics IMPAIR and DISAB. SICK6 goes, like the health dummies, in the opposite direction.[23] The mixed picture motivates our recursive approach as it enables us to reduce health dimensions from 10 to 1.

5.6.1 Health status

As we can see from Table 5.6, all measures of health care utilization, i.e. DOCTOR, VISITS, HOSPITAL, and SICK6, have the expected negative sign. They all significantly reduce health. This also applies to the variables approximating the sickness history of an individual, IMPAIR and DISAB.

Doing active sport (SPORTS) and the log of net income (LNNET) are means of health production.[24] Actively doing sport significantly increases health. As already mentioned above, there could be the reverse causality. The effect of net income on health is positive measuring, although insignificant, the impact of improved access to economic resources of health production.

The positive sign of education (EDU), measured in years spent in the education system, points to the complementarity between education and health. Its insignificance may stem from the positive correlation with income: education is a major determinant of wage earning ability.

Insured of non-German nationality (FOREIGN) significantly rate their health better than Germans, while gender (FEMALE) has no significant effect. Of course, one of the most important determinants of individual health is AGE. The elderly report a significantly lower health status. This points to the difficulty of disentangling health from age effects in the transition equation (see also Table 5.5).

---

[22] We use HEALTH0 and HEALTH1 as reference category. As there is only 1 percent with health status 0, significance is lost when using HEALTH0 only.

[23] As DISAB and SICK6 are risk adjusters they are of little importance for risk selection distortions. Individuals are eligible for sick pay after 6 weeks of sickness. Before that the employer is obliged to pay the salary.

[24] We have excluded LNNET from the single equation multinomial logit specification because of its strong correlation with LNGROSS.

	non-BKK members		BKK members	
Variable	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**
Constant	4.2034***	.1732	3.9306***	.5423
IMPAIR	−1.3377***	.0236	−1.4568***	.0677
DISAB	−.1732***	.0437	−.2435**	.1053
DOCTOR	−.1913***	.0207	−.2775***	.0564
VISITS	−.0551***	.0021	−.3528***	.0047
HOSPITAL	−.1551***	.0336	−.3408***	.0981
SICK6	−.2522***	.0482	−.4514***	.1292
SPORTS	.1910***	.0198	.2359***	.0537
EDU	.0069	.0043	.0198	.0121
FOREIGN	.1020***	.0270	.1482**	.0690
FEMALE	.0122	.0228	.0430	.0646
AGE	−.0254***	.0011	−.1384***	.0032
UNEMPL	−.0846**	.0365	−.0879	.1531
WHITEC	.0600***	.0225	.0121	.0613
LNNET	.0264	.0212	.0003	.0667
YEAR96	.0863***	.0327	.0570	.1000
YEAR97	.0956***	.0337	.0453	.0990
YEAR98	.1057***	.0336	.1371	.0982
YEAR99	.0958***	.0334	.1306	.0961
μ_1	1.3560***	.0347	1.3973***	.1011
μ_2	2.8515***	.0370	2.9073***	.1115
μ_3	4.7530***	.0399	4.7345***	.1182
Observations	15,728		2,112	
Log likelihood	−15,337.62		−2,085.42	
Pseudo R^2	0.1771		0.1847	
LR chisq(18)	6,602.96		944.78	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 5.6: First stage estimation results: ordered probit for health status.

Being unemployed (UMEMPL) has a negative impact on health status. The link may be indirect: all other things equal, the life satisfaction of unemployed people is much lower than that of the employed (Frey and Stutzer (2002)). This may feedback into actual health or at least in the self-rating of it. Unemployment plays a minor role among BKK members (see Table 5.9), so, not surprisingly, this effect is insignificant. Like with active sports, there may be the reverse causality. However, we do not follow this route and simply include this variable as a control. White collar employees (WHITEC) are found to be significantly healthier on average. This seems plausible as their jobs are typically physically less demanding. Again, significance is lacking for BKK members.

Finally, we included YEAR dummies in the regression in order to control for variation over time. For non-BKK members average health in all subsequent years is significantly higher than in 1995. No significant effects obtain for BKK members.

5.6.2 Switching behavior of non-BKK members

Using the fitted health index from the ordered probit regression we obtain the parameter estimates of the multinomial logit model as shown in Table 5.13 in the Appendix. There is only one health variable left, namely, the fitted health index Y1F. Comparing the impact of health on switching behavior is straightforward. It is tempting to simply compare the size of the estimated coefficient, but this would be misleading as the health effect does not only depend on this single coefficient (see equation (5.5)). For interpretation we will thus focus on the marginal effects shown in Table 5.7 below.

A marginal increase in the health index significantly reduces the probability of not switching, i.e., the sick are less likely to switch. As already argued above, sick people (bad risks) simply have higher switching costs. There is no significant health effect on the probability of switching within the non-BKK funds ($y_2 = 2$). In contrast, better health significantly increases the probability of switching to a BKK ($y_2 = 1$). Comparing these types of switchers thus reveals a difference: there is a health effect on top of transaction costs for out switchers.[25] We find $\delta_1 > \delta_2$ and thus some evidence for active risk selection by BKKs. Before drafting policy recommendations we have to test whether this difference is actually significant. To address this, we analyze the change in the log-odds ratio given in equation (5.5) for a (marginal) change in health status Y1F. From our parameter

[25] The p-values of the marginal health effects are .055, .086, and .278, respectively.

	Prob($y_2 = 0$)		Prob($y_2 = 1$)		Prob($y_2 = 2$)	
Variable	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**
Constant	2.6579***	.4313	−1.6482***	.2877	−1.0097***	.3587
Y1F	−.0574*	.0299	.0302*	.0176	.0272	.0251
WHITEC	−.1166**	.0547	.0742**	.0318	.0424	.0460
FULLTIME	−.1546**	.0725	.0205	.0418	.1341**	.0613
UNEMPL	.0051	.0890	−.0824	.0603	.0773	.0683
SUBST	−.0111	.0530	.0635**	.0315	−.0523	.0441
OTHER	−.0983	.0705	.1684***	.0388	−.0701	.0620
NOSINGLE	.0104	.0445	−.0322	.0248	.0217	.0377
FOREIGN	.4923***	.0824	−.0193	.0392	−.4730***	.0849
FEMALE	−.0215	.0492	.0115	.0278	.0100	.0413
AGE	.0213***	.0030	−.0068***	.0018	−.0145***	.0027
EDU	−.0316***	.0090	−.0030	.0051	.0346***	.0079
LNGROSS	−.0510	.0549	.0971***	.0333	−.0461	.0457
YEAR96	−.1025	.0839	.1112*	.0589	−.0087	.0639
YEAR97	−.1524*	.0832	.1864***	.0590	−.0340	.0649
YEAR98	−.3024***	.0811	.2044***	.0592	.0980	.0628
YEAR99	−.3534***	.0804	.2778***	.0605	.0756	.0631

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, all entries were multiplied by 10.

Table 5.7: Second stage marginal effects for non-BKK members.

estimates (see Table 5.13) we see that $\gamma_1 - \gamma_2 > 0$. Thus the probability of switching to a BKK increases relatively to the probability of switching to a non-BKK. However, this difference is not significantly different from 0, which can easily be seen when rerunning estimates with switchers to non-BKKs as a reference category. For non-BKK members we thus find no evidence for active risk selection by BKKs.

White collar workers, WHITEC, are more likely to switch and, if they switch, it is more likely for them than it is for others to enroll with a BKK. Potentially there are some financial incentives operating in the background. White collar workers are more likely to be employed by large companies. As employers pay half of the contribution there might be some employer pressure to switch to a 'cheaper' fund. Unfortunately the data on firm size is very incomplete preventing us from including it as an explanatory variable.[26]

[26] Comparing the parameter estimates of the single equation approach with the outcome

All other things equal (especially income), FULLTIME employed are significantly more likely to switch. As there are no special financial incentives this might reflect better access to information. If they switch, non-BKK funds are the main beneficiaries.

As unemployment insurance pays the contribution rate, there are no financial incentives of switching for the unemployed. Consequently, a lower propensity to switch results. However, the model yields no significant effects. Insured enrolled with a substitute fund (SUBST), or with other funds (OTHER), are more likely to switch than regional fund members. As the contribution rates of the regional funds are, on average, the highest the negative signs, although not significant, are surprising. Members of the regional funds should have stronger financial incentives to switch. This may indeed be the case, but it appears as if our dummies are improper approximations of actual contribution rates. That may be due to the large variation of contribution rates within sickness fund types. However, if a switch occurs, regional fund members are less likely to switch to a BKK. Switching activity within the non-BKK funds follows no pattern.

FOREIGN controls for the systematic difference between individuals of German nationality and non-German nationality. Being German increases the probability of switching. This can mainly be attributed to the much lower tendency of non-Germans to switch within the non-BKK funds. Household size has no effect on switching behavior as single households do not significantly differ from NOSINGLE households.

The following two variables, FEMALE and AGE, are of little interest from the risk selection perspective as they are risk adjusters in the German scheme. However, they offer some valuable insights. First, gender has no significant effect on switching behavior. Second, the probability of switching significantly decreases with AGE. A ten years increase in age yields a .02 lower probability of switching. As already argued above, transaction costs are higher for the elderly. The age effect is about twice as high for within switchers than for out switchers, pointing to the relative appeal of BKKs. Note that we obtain a significant health effect although we control for age which is one of the major determinants of health.

Higher education, measured in years spent in the education system, EDU, significantly increases the probability to switch. A three years increase in education increases the probability of switching by .01.[27] A marginal in-

of the recursive model reveals only small differences, i.e. results are robust.

[27] In Germany compulsory education is 10 years. In general 13 years are required for a school leaving examination (Abitur). So this simulation approximates the impact of different school leaving certificates.

crease in gross income (LNGROSS) significantly increases the probability of switching to a BKK, pointing to the financial incentives of doing so (contributions are payroll taxes).

Finally, as already anticipated from Table 5.3, the switching dynamics significantly increased over time. This pattern is mainly due the increasing flow towards the BKKs. Recall that the contribution rate differential increased from 1998 onwards, i.e., BKKs became more and more financially attractive, while there was an adaptation of contribution rates of all other types of funds (see Figure 5.1).[28]

5.6.3 Switching behavior of BKK members

Although there is no evidence for risk selection activities when considering non-BKK members it is theoretically possible that there is evidence for it when looking at the BKK members. A complete analysis thus requires also analyzing these. Parameter estimates and marginal effects are given in Tables 5.14 and 5.15 in the Appendix.

Our hypothesis of active risk selection of BKKs can clearly be rejected as $\eta_1 > \eta_2$. Again, the difference is not significantly different from 0. Together with our result from the previous subsection we can conclude that there is no indication of any risk selection activities in the German public health insurance system. As already mentioned above, this interpretation relies on our homogeneity assumption. It may well be that there are some BKKs as well as some non-BKKs actively trying to select the low risks. If this were true, then our conclusion would be that BKKs and non-BKKs are equally successful in risk selection. However, given the development of members and contribution rates it is obvious directing attention to BKKs and non-BKKs rather than to other subgroups of funds.

Only the following three variables require some further discussion: UNEMPL, AGE, and LNGROSS. People becoming unemployed may somehow be forced to leave the BKK of their employer. Being unemployed thus reduces the probability of staying. The flow is mainly directed towards the non-BKKs. Age is a rather unimportant determinant of switching behavior. The income effects are much more pronounced than for non-BKK members. BKK members may be more aware of the financial benefits of being enrolled with a BKK. It could also be argued that BKK members are better informed according to their higher propensity to switch (see Table 5.9). Finally, there is no time pattern in switching behavior. The important dynamics of competition are in the switches from non-BKK members towards the BKKs.

[28]Schut et al. (2003) found increasing price elasticities over time.

5.7 Conclusion

After the natural experiment in 1996 switching dynamics in the German statutory health insurance system have been on the increase. The company-based sickness funds (BKKs) were the main beneficiaries of these dynamics. This is likely to be due to the remarkably lower contribution rates, even after risk adjustment. The increasing contribution rate differential from 1998 onwards together with the different (additional) benefits packages of the BKKs gives rise to the conjecture that BKKs were actively, and successfully, engaged in risk selection.

Using 1995 to 2000 GSOEP data we tested this hypothesis by analyzing switches within and between non-BKKs and BKKs. We set up a recursive model and estimated the self-assessed health status by an ordered probit model at the first stage and fitted a multinomial logit model for health plan choice at the second stage. This procedure enabled us, first, to come up with a single health coefficient that can easily be compared across switching types, and, second, to disentangle switching costs and active risk selection.

We found that better health significantly increases the probability of switching. Sick people simply have higher switching costs. There is no significant difference in the (positive) health effects on switching behavior. As this is true for both samples, non-BKK members and BKK members, there is no indication of any (successful) risk selection activities in the German public health insurance market. The benefit packages thus seem to be no means of selection. Instead, differences may simply map the preferences of the enrollees. Although there is anecdotal evidence of funds using other selection devices we found no evidence that these are actually successful.

Due to the much larger flow towards the BKKs and lower switching costs of low risk consumers there is nevertheless risk separation. This originates in the historically better risk structures of BKKs together with incomplete risk adjustment. Public policy should thus be directed towards mitigating comparative advantages stemming from this asymmetry, i.e. an improvement of the risk adjustment mechanism is required. The planned consideration of morbidity measures as risk adjusters from 2007 onwards goes in that direction.

5.8 Appendix

In this Appendix we provide the explanation of variables and more detailed information on estimation results.

Variable	Explanation
HEALTH	self-assessed health status, 0 = bad, 1 = not so good,
	2 = satisfactory, 3 = good, 4 = very good
SWITCH[a]	0 = no switch, 1 = switch to a BKK, 2 = switch to a non-BKK
SWITCH[b]	0 = no switch, 1 = switch to a non-BKK, 2 = switch to a BKK
LNNET	natural logarithm of net income
LNGROSS	natural logarithm of gross income
EDU	years in the education system
AGE	age in years
VISITS	number of visits to doctors during the last three months
dummies:	
HEALTHX	1 = health status is X=0,...,4, HEALTH0*, HEALTH1*
SUBST	1 = membership in a substitute fund
OTHER	1 = membership in other funds
REGIONAL*	1 = membership in a regional fund
FULLTIME	1 = full time employed
LESSTIME*	1 = short working hours, part time contract, maternity leave,...
NOSINGLE	1 = no single household
YEARXX	1 = year 19XX, YEAR95*
DISAB	1 = disability or incapacity to work
WHITEC	1 = white collar employee
UNEMPL	1 = unemployed
FEMALE	1 = female
FOREIGN	1 = non-German nationality
IMPAIR	1 = health status prevents from completing everyday tasks
DOCTOR	1 = at least one visit to a doctor during the last three months
HOSPITAL	1 = hospital stay during the last year
SICK6	1 = work disability for longer than 6 weeks during the last year
SPORTS	1 = active sport at least once a month

Table 5.8: Explanations of variables. Note: * indicates that the variable is a reference category in our estimation, [a] non-BKK members, [b] BKK members.

	non-BKK members		BKK members	
Variable	**Mean**	**SD**	**Mean**	**SD**
HEALTH	2.5846	.8143	2.6013	.8351
SWITCH = 1	.0280	–	.0739	–
SWITCH = 2	.0556	–	.0369	–
LNNET	7.7008	.4843	7.8975	.4233
LNGROSS	8.1380	.5021	8.3385	.4244
EDU	11.8831	2.4412	11.7083	2.4138
AGE	38.4865	8.1378	38.5223	8.0558
VISITS	2.0167	3.4027	2.1832	4.0595
HEALTH0	.0114	.1064	.0118	.1082
HEALTH1	.0837	.2770	.0862	.2807
HEALTH2	.3057	.4607	.2978	.4574
HEALTH3	.5071	.5000	.4972	.5001
HEALTH4	.0921	.2891	.1070	.3092
SUBST	.4796	.4996	.0000	.0000
OTHER	.0988	.2984	.0000	.0000
FULLTIME	.8361	.3701	.9096	.2869
NOSINGLE	.6804	.4663	.6908	.4623
YEAR95	.1089	.3115	.1051	.3068
YEAR96	.2270	.4189	.1965	.3974
YEAR97	.2181	.4130	.2012	.4010
YEAR98	.2202	.4143	.2386	.4264
YEAR99	.2257	.4181	.2585	.4379
DISAB	.0451	.2075	.0559	.2297
WHITEC	.5238	.4995	.4446	.4970
UNEMPL	.0598	.2372	.0199	.1396
FEMALE	.4512	.4976	.3101	.4627
FOREIGN	.1307	.3371	.1544	.3614
IMPAIR	.2528	.4346	.2472	.4315
DOCTOR	.6285	.4832	.6383	.4806
HOSPITAL	.0828	.2556	.0715	.2577
SICK6	.0408	.1977	.0417	.1999
SPORTS	.3384	.4732	.3561	.4789
observations	15,728		2,112	

Table 5.9: Sample statistics. Source: GSOEP 1995-2000, own calculations.

Health status	Full sample			Stayers			Within switchers			Out switchers		
	N	(%)	Perc.	N	(%)	Perc.	N	(%)	Perc.	N	(%)	Perc.
Non-BKK members												
0 (bad)	180	(1.000)	.0114	168	(.933)	.0117	7	(.039)	.0080	5	(.028)	.0114
1 (not so good)	1,317	(1.000)	.0837	1,273	(.967)	.0883	29	(.022)	.0332	15	(.011)	.0341
2 (satisfactory)	4,808	(1.000)	.3057	4,435	(.922)	.3077	262	(.054)	.2998	111	(.023)	.2523
3 (good)	7,975	(1.000)	.5071	7,225	(.906)	.5012	490	(.061)	.5606	260	(.033)	.5909
4 (very good)	1,448	(1.000)	.0921	1,313	(.907)	.0911	86	(.059)	.0984	49	(.034)	.1114
All	15,728	(1.000)	1.0000	14,414	(.916)	1.0000	874	(.056)	1.0000	440	(.028)	1.0000
Average health			2.58			2.57			2.71			2.76
BKK members												
0 (bad)	25	(1.000)	.0118	19	(.760)	.0101	1	(.040)	.0128	5	(.200)	.0321
1 (not so good)	182	(1.000)	.0862	169	(.929)	.0900	5	(.027)	.0641	8	(.044)	.0513
2 (satisfactory)	629	(1.000)	.2978	582	(.925)	.3099	16	(.025)	.2051	31	(.049)	.1987
3 (good)	1,050	(1.000)	.4972	922	(.878)	.4909	48	(.046)	.6154	80	(.076)	.5128
4 (very good)	226	(1.000)	.1070	186	(.823)	.0990	8	(.035)	.1026	32	(.142)	.2051
All	2,112	(1.000)	1.0000	1,878	(.889)	1.0000	78	(.037)	1.0000	156	(.074)	1.0000
Average health			2.60			2.58			2.73			2.81

Table 5.10: Health status and switching behavior. Source: GSOEP 1995-2000, own calculations.

	Prob($y_2 = 1$)		Prob($y_2 = 2$)	
Variable	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**
Constant	−7.5514***	1.1153	−3.0056***	.7485
HEALTH2	.5429**	.2594	.8180***	.1898
HEALTH3	.7694***	.2696	.9034***	.1977
HEALTH4	.7173**	.3065	.8155***	.2266
IMPAIR	−.0108	.1513	.0264	.1027
DISAB	.1569	.2637	.2633	.1800
DOCTOR	.0153	.1167	.0899	.0836
VISITS	.0008	.0201	.0240**	.0120
HOSPITAL	−.1734	.2093	.1079	.1369
SICK6	.4560*	.2549	−.7297***	.2704
SPORTS	.1592	.1036	.1427*	.0750
WHITEC	.3157**	.1363	.0734	.0957
FULLTIME	.1053	.1829	.2997**	.1269
UNEMPL	−.3580	.2610	.1753	.1420
SUBST	.2653*	.1355	−.1121	.0918
OTHER	.7257***	.1568	−.1447	.1290
NOSINGLE	−.1331	.1082	.0377	.0790
FOREIGN	−.1234	.1710	−.9632***	.1616
FEMALE	.0624	.1228	.0109	.0869
AGE	−.0316***	.0070	−.0292***	.0050
EDU	−.0114	.0224	.0670***	.0159
LNGROSS	.4188***	.1412	−.0982	.0953
YEAR96	.4812*	.2522	.0469	.1334
YEAR97	.8032***	.2446	.0033	.1355
YEAR98	.8865***	.2429	.2784**	.1306
YEAR99	1.2069***	.2377	.2419*	.1315
Observations	15,728			
Log likelihood	−5,164.83			
Pseudo R^2	0.0359			
LR chisq(50)	384.55			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 5.11: Single equation estimates for switching behavior of non-BKK members.

	Prob($y_2 = 0$)		Prob($y_2 = 1$)		Prob($y_2 = 2$)	
Variable	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**
Constant	2.9879***	.4248	−1.6674***	.2888	−1.3205***	.3569
HEALTH2	−.4896***	.1025	.1131*	.0597	.3765***	.0928
HEALTH3	−.5771***	.1067	.1631***	.0629	.4140***	.0970
HEALTH4	−.5258***	.1229	.1524**	.0708	.3734***	.1093
IMPAIR	−.0097	.0581	−.0027	.0341	.0125	.0480
DISAB	−.1538	.1017	.0324	.0594	.1214	.0845
DOCTOR	−.0443	.0465	.0024	.0263	.0419	.0391
VISITS	−.0111	.0071	−.0001	.0045	.0112**	.0056
HOSPITAL	−.0121	.0783	−.0403	.0473	.0524	.0641
SICK6	.2355*	.1359	.1110*	.0583	−.3466***	.1288
SPORTS	−.0992**	.0417	.0342	.0235	.0649*	.0353
WHITEC	−.1011*	.0536	.0703**	.0313	.0308	.0448
FULLTIME	−.1594**	.0711	.0203	.0412	.1390**	.0600
UNEMPL	−.0034	.0871	−.0826	.0593	.0861	.0666
SUBST	−.0056	.0519	.0610**	.0310	−.0555	.0430
OTHER	−.0892	.0690	.1651***	.0382	−.0760	.0605
NOSINGLE	.0113	.0438	−.0304	.0245	.0192	.0370
FOREIGN	.4661***	.0807	−.0169	.0386	−.4493***	.0825
FEMALE	−.0183	.0486	.0139	.0277	.0044	.0406
AGE	.0201***	.0027	−.0068***	.0017	−.0133***	.0025
EDU	−.0281***	.0089	−.0033	.0051	.0315***	.0077
LNGROSS	−.0448	.0539	.0955***	.0330	−.0507	.0447
YEAR96	−.1244	.0823	.1079*	.0579	.0165	.0624
YEAR97	−.1734**	.0817	.1809***	.0581	−.0076	.0634
YEAR98	−.3168***	.0797	.1966***	.0582	.1202*	.0615
YEAR99	−.3687***	.0790	.2692***	.0595	.0995	.0618

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, all entries were multiplied by 10.

Table 5.12: Single equation marginal effects for non-BKK members.

	Prob($y_2 = 1$)		Prob($y_2 = 2$)	
Variable	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**
Constant	−7.3060***	1.1034	−2.2789***	.7404
Y1F	.1350*	.0758	.0599	.0520
WHITEC	.3286**	.1357	.0963	.0955
FULLTIME	.1038	.1822	.2813**	.1262
UNEMPL	−.3514	.2610	.1519	.1418
SUBST	.2715**	.1350	−.1020	.0915
OTHER	.7278***	.1566	−.1276	.1286
NOSINGLE	−.1381	.1075	.0417	.0784
FOREIGN	−.1354	.1708	−.9863***	.1614
FEMALE	.0512	.1212	.0220	.0859
AGE	−.0312***	.0074	−.0309***	.0053
EDU	−.0092	.0224	.0716***	.0158
LNGROSS	.4190***	.1402	−.0854	.0949
YEAR96	.4846*	.2517	−.0062	.1328
YEAR97	.8101***	.2441	−.0506	.1349
YEAR98	.9030***	.2425	.2260*	.1300
YEAR99	1.2212***	.2370	.1873	.1308
Observations	15,728			
Log likelihood	−5,194.69			
Pseudo R^2	0.0303			
LR chisq(32)	324.84			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 5.13: Second stage parameter estimates for switching behavior of non-BKK members.

	Prob($y_2 = 1$)		Prob($y_2 = 2$)	
Variable	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**
Constant	5.8667***	1.9054	−4.3067	2.9780
Y1F	.2693**	.1197	.2431	.1715
WHITEC	.2825	.2106	.5590*	.2883
FULLTIME	.4698	.3776	−.2643	.5044
UNEMPL	1.0754***	.4139	.5669	.7535
NOSINGLE	−.0526	.1936	−.0700	.2639
FOREIGN	.3039	.2343	−.8355*	.4883
FEMALE	−.4486**	.2243	−.0596	.2999
AGE	.0024	.0117	−.0491***	.0182
EDU	.0253	.0420	−.0275	.0564
LNGROSS	−1.2144***	.2559	.0576	.3763
YEAR96	−.1631	.3275	2.0567**	1.0407
YEAR97	−.2200	.3289	2.1344**	1.0364
YEAR98	.4965*	.2952	2.4777**	1.0245
YEAR99	−.4332	.3244	1.9892*	1.0308
Observations	2,112			
Log likelihood	−834.46			
Pseudo R^2	0.0563			
LR chisq(28)	99.66			
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$				

Table 5.14: Second stage parameter estimates for switching behavior of BKK members.

	Prob($y_2 = 0$)		Prob($y_2 = 1$)		Prob($y_2 = 2$)	
Variable	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**	**Coeff.**	**s.e.**
Constant	−2.3775*	1.3664	3.6243***	1.2575	−1.2468	.8609
Y1F	−.2184***	.0832	.1586**	.0761	.0598	.0477
WHITEC	−.3044**	.1474	.1610	.1289	.1434*	.0829
FULLTIME	−.2104	.2609	.2888	.2318	−.0785	.1352
UNEMPL	−.7719**	.3251	.6404**	.2617	.1315	.2016
NOSINGLE	.0482	.1347	−.0306	.1169	−.0177	.0699
FOREIGN	.0286	.1854	.1985	.1443	−.2272	.1426
FEMALE	.2782*	.1550	−.2703*	.1405	−.0079	.0795
AGE	.0108	.0084	.0023	.0071	−.0131**	.0058
EDU	−.0080	.0291	.0158	.0254	−.0078	.0150
LNGROSS	.6988***	.1812	−.7355***	.1847	.0367	.1002
YEAR96	−.4139	.2996	−.1349	.1993	.5489*	.3337
YEAR97	−.3998	.2980	−.1707	.2007	.5705*	.3358
YEAR98	−.9055***	.2804	.2565	.1819	.6490*	.3441
YEAR99	−.2386	.2958	−.2971	.2017	.5357	.3298

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, all entries were multiplied by 10.

Table 5.15: Second stage marginal effects for BKK members.

Chapter 6

Summary (in German)

In der Einleitung dieser Dissertationsschrift wurden zunächst einige Charakteristika des Gesundheitsmarktes des beschrieben. Das sich jeweils ergebende Marktversagen rechtfertigt staatliche Eingriffe in den Gesundheitsmarkt. Da ein sozialer Planer in der Regel mehrere Probleme gleichzeitig angeht bzw. angehen muss, greifen regulierende Maßnahmen mitunter ineinander, können aber durchaus auch sich widersprechende Wirkungen auslösen. Die Multidimensionalität macht es für einen Regulator schwer, wenn nicht sogar unmöglich, eine effiziente Politik zu implementieren. Die Multidimensionalität setzt aber auch der Wissenschaft Grenzen. So wird sich die ökonomische Analyse in der Regel auf Partialmodelle beschränken müssen.

Anschließend wurden, den Inhalten der einzelnen Kapitel folgend, vier Teilaspekte isoliert betrachtet und problematisiert. Die nachfolgenden vier Abschnitte fassen die in diesen Kapiteln erzielten Resultate zusammen.

6.1 Impfungen (Kapitel 2)

Das Standardbeispiel für externe Effekte im Gesundheitsmarkt sind Impfungen. Entscheidet sich ein Individuum sich impfen zu lassen, so hat dies zwei Effekte. Einerseits kann er sich selber nicht mehr mit der Krankheit infizieren, d. h. er ist immun.[1] Andererseits kann er in der Regel andere Personen nicht mehr mit der Krankheit anstecken. Der zweite Effekt ist

[1]Aus Gründen der besseren Lesbarkeit des Textes, und nur aus diesen, wird fast ausschließlich die männliche Form verwendet.

die Quelle einer Ineffizienz, da ein Individuum bei seiner persönlichen Impfentscheidung diesen positiven externen Effekt seiner Impfung auf andere nicht berücksichtigt. Statt dessen wägt er seinen persönlichen Vorteil aus der Impfung mit seinen individuellen Nachteilen, z. B. Zeitkosten, Schmerzen oder potentielle Nebenwirkungen, ab. Dies führt zu einem zu geringen Immunisierungsgrad und rechtfertigt staatliche Eingriffe.

In Kapitel 2 wurde ein Modell entwickelt, dass die Externalität des Impfens durch einen negativen Einfluss der Immunisierungsrate auf die Zahlungsbereitschaft für Impfungen, abbildet. Je höher der Anteil der Immunen in der Bevölkerung, um so geringer ist die Wahrscheinlichkeit, dass man auf einen Infizierten trifft und sich ansteckt. Zudem hängt die Zahlungsbereitschaft, empirischen Beobachtungen folgend, positiv vom Einkommen ab. In diesem Modellrahmen wurde das Preissetzungsverhalten eines Monopolisten für Impfstoffe untersucht.

Durch den bio-technologischen Fortschritt wurden während des letzten Jahrzehnts eine Vielzahl neuer Impfstoffe entwickelt. Die aktuelle Politik des Europäischen Patentamtes erlaubt zudem die Patentierung von lebenden Organismen oder genetischen Codes, so dass die Firmen ihre neuen Produkte relativ leicht durch Patente schützen können. Die Annahme von Marktmacht auf der Anbieterseite wird also durch empirische Fakten gestützt.

Ein Monopolist für einen Impfstoff hat im Wesentlichen zwei Anreize. Erstens hat er ein Interesse, die Krankheit am Leben zu erhalten. Denn sollte die Krankheit ausgerottet werden, so gehen neben der Monopolstellung auch die Monopolgewinne verloren. Zweitens wird der Monopolist die Abhängigkeit der Zahlungsbereitschaft vom Immunisierungsgrad strategisch nutzen. Er wird Individuen mit geringem Einkommen - und damit geringerer Zahlungsbereitschaft - ungeimpft lassen, um damit über die erhöhte Ansteckungsgefahr die Zahlungsbereitschaft der Reichen zu erhöhen. Das Angebot ist somit geringer als im Standardmonopolfall ohne Externalität. Der Anreiz das Angebot zu verknappen ist um so größer, je größer der Einfluss des externen Effektes ist. Beachte, dass die resultierende Ineffizienz zwei in dieselbe Richtung wirkende Ursachen hat, den externen Effekt und die Monopolmacht.

Es wurde zudem gezeigt, dass, im Gegensatz zum Standard-Monopolmodell ohne Externalität, ein perfekt preisdiskriminierender Monopolist nicht notwendiger Weise zu einer effizienten Allokation führt.[2] Bietet der Monopolist dem marginalen ungeimpften Individuum die Impfung zu seiner

[2] In unserem Modellrahmen ist die vollständige Immunisierung der Bevölkerung, d. h. die Ausrottung der Krankheit, effizient.

individuellen Zahlungsbereitschaft an, so sinkt nicht nur der Preis an dieser Grenze, sondern auch die Zahlungsbereitschaft aller anderen Individuen, da die Ansteckungsgefahr durch den höheren Immunisierungsgrad sinkt. Die Vereinten Nationen unterstützen mit ihrer UN Acceleration Access Initiative Preisdiskriminierung, da sie den Zugang der dritten Welt zu Impfstoffen verbessert. Das Modell zeigt, dass diese Einschätzung zwar richtig ist, aber auch, dass dies mitunter nicht ausreicht um Effizienz herzustellen. In diesen Fällen wären weitere Maßnahmen erforderlich, beispielsweise könnten Impfprogramme durchgeführt werden.

Ein sozialer Planer sieht sich einem „doppelten" Marktversagen gegenüber. Wie bereits oben erwähnt, wirken der externe Effekt und die Monopolmacht in dieselbe Richtung, d. h. sie reduzieren den Immunisierungsgrad. Eine Subventionierung des Impfstoffes könnte nun für beide Ursachen korrigieren und Effizienz herstellen. Im Allgemeinen sind Subventionen im Impfmarkt von geringerer Wirkung als in Märkten ohne Externalität. Der reduzierte Preis führt zu einem positiven Nachfrageeffekt. Dieser dämpft sich jedoch selbst, da durch die erhöhte Nachfrage auch das Ansteckungsrisiko und damit die Zahlungsbereitschaft für Impfungen sinkt. Neben diesem negativen „Prävalenzeffekt" ergibt sich noch ein zweiter negativer Effekt, wenn man die Finanzierung der Subventionen berücksichtigt. Eine entsprechende Besteuerung der Individuen löst einen negativen Einkommenseffekt aus. Ist dieser hinreichend groß, so kann eine positive Subvention das Marktergebnis sogar verschlechtern, d. h. mit Subvention impfen sich weniger Individuen als ohne Subvention. In diesen Fällen wäre eine Besteuerung von Impfstoffen optimal. Die Angemessenheit von Subventionen ist also mit einem Fragezeichen zu versehen.

Bei Impfprogrammen verhält es sich mitunter ähnlich. Liegt der Anteil der Bevölkerung, der durch das Impfprogramm abgedeckt ist, unter 100 Prozent, so senkt das Impfprogramm die Zahlungsbereitschaft derer, die nicht Teil des Programms sind. Deshalb ist es schwierig volle Immunisierung zu erreichen, wenn nicht die gesamte Bevölkerung durch das Programm abgedeckt ist. Diese Problematik entschärft sich, wenn der soziale Planer die individuellen Einkommen beobachten kann. In diesem Fall könnte er die Armen (z. B. die Dritte-Welt-Länder) in das Programm einbeziehen und die Reichen (z. B. die Industrieländer) vom Monopolisten bedienen lassen. Die Zahlungsbereitschaft der Reichen wird zwar durch das Impfprogramm ebenfalls geschmälert, jedoch bleibt sie wegen des Einkommenseffektes relativ hoch. Volle Immunisierung kann deshalb auch mit unvollständigen Impfprogrammen erreicht werden. Diese Tatsache bietet eine effizienzbasierte Begründung für Impfprogramme, die typischer Weise von den Vereinten

Nationen, der Weltbank oder der Weltgesundheitsorganisation unterstützt werden.

6.2 Preisregulierung (Kapitel 3)

Wegen der Besonderheiten von Gesundheitsgütern sind die Preise für medizinische Leistungen in der Regel reguliert. Zudem sind Patienten gegenüber Preisen nicht sehr sensitiv, da die Kosten medizinischer Behandlungen in der Regel (zum Großteil) von Krankenversicherungen getragen werden. Im Wettbewerb um Patienten können Ärzte somit den Preis nicht strategisch nutzen. Dementsprechend verlagern die Ärzte den Wettbewerb auf Nicht-Preis-Variablen, beispielsweise die Qualität medizinischer Behandlungen (vertikale Produktdifferenzierung) und die Spezialisierung auf bestimmte Krankheiten bzw. den Praxistandort (horizontale Produktdifferenzierung).

Die Anreizwirkungen von Preisregulierung wurden in einem 3-stufigen nicht kooperativen Spiel analysiert, in dem die Spieler die folgenden sequentiellen Entscheidungen zu treffen hatten: Markteintritt, Standortwahl und die Qualität medizinischer Leistungen. Wir verwendeten das Kreismodell von Salop (1979) und beschränkten uns auf symmetrische Gleichgewichte, d. h. insbesondere, dass sich die Standorte der Ärzte symmetrisch auf dem Kreis verteilen. Auch wenn wir die strategische Beziehung zwischen Standortwahl und medizinischer Qualität detailliert untersuchten, kam der Standortwahl an sich wegen der Symmetrieannahme nur geringere Bedeutung zu. Stattdessen konzentrierte sich die Analyse auf die Qualität medizinischer Versorgung und auf die Arztdichte.

Zunächst untersuchten wir das strategische Verhalten der Ärzte bei exogen gegebenem Preis. Die im Gleichgewicht bereitgestellte Qualität hängt positiv vom Preis ab. Je höher der Preis, desto höher ist der Anreiz eines Arztes über eine Erhöhung der Qualität die Patienten der Konkurrenten zu gewinnen. Dieser Qualitätswettbewerb ist jedoch kostspielig. Die Ärzte können diesem teilweise ausweichen bzw. diesen abschwächen, indem sie ihre Standorte weit voneinander entfernt wählen. Dies ist in einem Kreismodell mit endogenem Markteintritt nur dann möglich, wenn weniger Firmen in den Markt eintreten. Folglich führt ein höherer Preis für medizinische Leistungen zu einer niedrigeren Arztdichte. Genau dieses Phänomen lässt sich in Deutschland beobachten. Kapitel 3 liefert somit eine mögliche Erklärung für die im deutschen Markt für niedergelassene Ärzte gemachten Beobachtungen.

Anschließend wurden verschiedene Szenarien der Preissetzung betrachtet. Zunächst unterstellten wir das Gleichgewicht mit vollständiger Selbstbindung des Regulators. In diesem Fall konnte sich der Regulator vor der Markteintrittsentscheidung der Ärzte bindend auf einen Preis für medizinische Leistungen festlegen. Ist der Preis sehr niedrig, so treten zu viele Firmen in den Markt ein und gleichzeitig ist die gleichgewichtige Qualität zu gering. Eine Preiserhöhung steigert somit die soziale Wohlfahrt. In der second-best Lösung lohnt sich eine weitere Preiserhöhung nicht mehr. Das second-best Optimum ist durch eine zu hohe Arztdichte und übermäßige Qualitätsbereitstellung charakterisiert.

Im zweiten Szenario erfolgte die Preissetzung nach der Standortwahl jedoch vor der Qualitätsentscheidung der Ärzte. Die Fähigkeit des Regulators sich bindend auf einen Preis festzulegen war somit beschränkt. Nachdem die Ärzte ihren Standort gewählt haben, kann der Regulator nur noch die Qualität durch die Preissetzung beeinflussen. Folglich resultierte die effiziente Qualität. Dies ging jedoch mit einer (viel) zu hohen Arztdichte einher. Da der Regulator nach den Qualitätsentscheidungen keinen Anreiz hat den Preis zu ändern, bezeichneten wir dieses Szenario als zeitkonsistente Regulierung. Da die soziale Wohlfahrt in dieser zeitkonsistenten Lösung geringer ist als bei second-best Preisregulierung, würde sich der Regulator gerne bindend auf den second-best Preis festlegen. Dieser ist jedoch nicht zeitkonsistent, da der Regulator nach Markteintritt und Standortwahl einen Anreiz hat, den Preis zu verändern (zu senken), um die effiziente Qualität zu induzieren.

Im Allgemeinen kann die effiziente Allokation nicht implementiert werden, da der Regulator nur über eine Regulierungsvariable verfügt (den Preis), jedoch zwei Regulierungsziele hat (die optimale Arztdichte und die effiziente Qualität). Verfügt der Regulator über eine zweite Regulierungsvariable resultiert Effizienz. Es wurden verschiedene Variablen diskutiert und die jeweils optimale Politik ermittelt. Beispielsweise könnte der Regulator Lizenzen verkaufen oder Qualitätsuntergrenzen vorschreiben. Auch variable Vergütungen können Effizienz herstellen.

Angesichts des Regulierungs- bzw. des Selbstbindungsproblems könnte der Regulator erwägen, die soziale Wohlfahrt durch eine Stärkung des Wettbewerbs zu steigern. Er könnte beispielsweise die Preise freigeben. Wählen nun die Ärzte Preise und Qualitäten simultan, so ergibt sich dieselbe Allokation wie bei zeitkonsistenter Regulierung. Da sich auch Wähler einem Selbstbindungsproblem gegenüber sehen, ist die strategische Delegation der Preissetzung an die Wähler nicht zielführend. Im Gegenteil, das Marktergebnis verschlechtert sich gegenüber der zeitkonsistenten Politik sogar noch.

6.3 Hausarztprinzip (Kapitel 4)

Das Vereinigte Königreich und Skandinavien sind Beispiele für Länder, die den Hausärzten eine Lotsenfunktion im Gesundheitswesen zugewiesen haben, d. h. das „Hausarztprinzip" wird praktiziert. Patienten haben keinen direkten Zugang zu Fachärzten oder Krankenhäusern, sondern benötigen eine entsprechende Überweisung von ihrem Hausarzt. Das Hausarztprinzip ist Teil der aktuellen deutschen Gesundheitsreform, die zum 1. Januar 2004 in Kraft treten soll. Die Gesetzlichen Krankenkassen sollen verpflichtet werden, sogenannte Hausarztprogramme anzubieten. Patienten, die sich in einem solchen Programm einschreiben, sollen Beitragsnachlässe gewährt werden können. Im Gegensatz dazu gibt es in Schweden Überlegungen, den direkten Zugang zu Krankenhäusern, also ohne Überweisung des Hausarztes, zu ermöglichen.

Im Allgemeinen werden dem Hausarztprinzip drei positive Wirkungen bescheinigt. Erstens, trägt das Hausarztprinzip zur Kostendämpfung bei, da unnötige Behandlungen unterbleiben bzw. mit geringerer Rate auftreten. Diese Behauptung trifft überraschender Weise auf breite Akzeptanz, obwohl jegliche empirische Evidenz fehlt. So findet beispielsweise Barros (1998) keinen signifikanten Einfluss des Hausarztprinzips auf die aggregierten Gesundheitsausgaben.[3] Zweitens steigt die Effizienz im System, da eine Verlagerung der Behandlungen von Fachärzten bzw. Krankenhäusern hin zu Hausärzten stattfindet wann immer ein Patient besser vom Hausarzt behandelt werden kann (Hausärzte kennen ihre Patienten und deren Krankengeschichte typischer Weise besser als Fachärzte oder Krankenhäuser). Drittens steigt die Effizienz im Markt für Fachärzte und die Effizienz im stationären Bereich, da die Hausärzte in der Regel besser über die angebotene Qualität in diesen Märkten informiert sind als die Patienten. Es kommt so zu einer besseren Zuordnung von Patienten zu Fachärzten bzw. Krankenhäusern. In Kapitel 4 konzentrierten wir uns auf den dritten Effekt und analysierten die Auswirkungen auf Wettbewerb und soziale Wohlfahrt wenn durch das Hausarztprinzip mehr Informationen in den Markt gelangen. Diese Wettbewerbseffekte blieben bisher unberücksichtigt.

Wie in Kapitel 3 auch unterstellten wir Preisregulierung und Nicht-Preis-Wettbewerb. Im Unterschied zu Kapitel 3 betrachteten wir ein Hotelling (1929) Modell mit zwei Firmen (Fachärzte oder Krankenhäuser), die in Qualitätswettbewerb stehen und zudem ihre Spezialisierung (ihren Standort auf

[3] Abstrahiert man von Selektionseffekten, könnten Beitragsrabatte im Rahmen der Hausarztprogramme in der Gesetzlichen Krankenversicherung Deutschlands ab 2004 die bisher fehlende Evidenz liefern.

der Hotelling-Linie) strategisch wählen können. Jeder Standort entspricht also einer Krankheit. Alternativ lässt sich die Spezialisierung als physischer Standort der Praxis oder des Krankenhauses interpretieren.

Das Hauptanliegen von Kapitel 4 war das Herausarbeiten der Wettbewerbswirkungen der durch das Hausarztprinzip generierten Informationen. Wir unterstellten, dass einige Patienten uninformiert sind, d. h. sie kennen weder ihre eigene Krankheit noch die Spezifika des Fachärztemarktes. Gehen diese Patienten direkt zum Facharzt oder ins Krankenhaus, können Fehlentscheidungen entlang von zwei Dimensionen auftreten. Einerseits könnte ein Patient den Anbieter wählen, der die geringere Qualität bereit stellt. Andererseits besteht die Gefahr, dass er ein Krankenhaus wählt, dass nicht hinreichend für die eigene Krankheit spezialisiert ist bzw. schlechter in der Lage ist die Krankheit zu behandeln als das andere Krankenhaus. Es wurde unterstellt, dass der Hausarzt alle relevanten Informationen beobachten kann, d. h. die Krankheit des Patienten, die Qualität der Krankenhäuser und deren Spezialisierung, so dass die genannten Fehler nicht mehr auftreten. Zudem nahmen wir an, dass die Hausärzte keinerlei Kosten verursachen und ihre Informationen wahrheitsgemäß den Patienten offenbaren. Diese Annahmen entsprechen im Wesentlichen dem oben genannten dritten positiven Effekt des Hausarztprinzips.

Wir untersuchten zwei Varianten des Grundmodells. In der ersten Variante konnte der Regulator wählen, ob er auf Hausärzte gänzlich verzichtet oder ob er das Hausarztprinzip für alle Versicherten bzw. Patienten verbindlich vorschreibt.[4] Entscheidet er sich zur Einführung des Hausarztprinzips, sind alle Patienten vollständig informiert. Folglich reagiert die gesamte Population sensitiv auf Qualitätsinvestitionen seitens der Krankenhäuser. Im Vergleich zur Situation ohne Hausärzte wird der Qualitätswettbewerb erheblich gestärkt. Da dieser Wettbewerb mit substantiellen Kosten verbunden ist, versuchen die Krankenhäuser diesem auszuweichen indem sie ihre Produkte stärker (horizontal) differenzieren. Implizieren die Parameter des Modells starken bzw. kostenintensiven Wettbewerb, führt die Einführung des Hausarztprinzips zu einer Verschlechterung des Marktergebnisses - und das obwohl das Hausarztprinzip keine direkten Kosten verursacht. Die Zahlungsbereitschaft für Informationen ist also negativ.

Kann der Regulator zusätzlich den Preis für Krankenhausbehandlungen

[4] Die Spielstufen sind wie folgt: 1. Der Regulator entscheidet über die Einführung des Hausarztprinzips und über die Höhe der Vergütung. 2. Die Krankenhäuser wählen ihre Spezialisierung. 3. Die Krankenhäuser wählen ihre Qualität. 4. Die Patienten fragen Krankenhausleistungen nach. Unter Umständen suchen sie vorher noch ihren Hausarzt auf.

festlegen, so ist das Hausarztprinzip immer wünschenswert. Die potentiell negativen Auswirkungen der zusätzlichen Informationen können durch Preissenkungen ausgeglichen werden. Die Informationen entfalten dann nicht mehr die beschriebenen negativen Effekte, da die Anreize in Qualität zu investieren aufgrund des niedrigeren Preises verschwinden. Die Einführung des Hausarztprinzips sollte demnach von einer geeigneten Preisregulierung begleitet werden. Dennoch ist eine effiziente Allokation in Allgemeinen nicht erreichbar, da der Preis optimale Qualität und optimale Spezialisierung induzieren muss.

In der zweiten Version des Grundmodells entschieden die Patienten selbst, ob sie einen Hausarzt aufsuchen oder nicht.[5] Die Endogenisierung des Hausarztbesuchs hat zur Folge, dass der Regulator durch geeignete Wahl des Preises ein generelles Hausarztprinzip induzieren kann. Ist der Preis hoch, so sind nicht nur die bereitgestellten Qualitäten hoch sondern auch die Produktdifferenzierung. Die Wahl des falschen Krankenhauses ist somit mit hohen Kosten verbunden. Da diese Kosten durch einen Hausarztbesuch vermieden werden können, suchen viele Patienten einen Hausarzt auf. Ist der Preis hinreichend groß, konsultieren alle Patienten vor einem Krankenhausbesuch ihren Hausarzt. Jedoch ist ein solcher Preis nicht notwendiger Weise optimal, da er übermäßige Qualität und/oder übermäßige Spezialisierung implizieren kann. Dies ist, analog zur ersten Variante des Grundmodells, genau dann der Fall, wenn die Parameter des Modells starken Qualitätswettbewerb implizieren. Der optimale Preis balanciert dann die Information im Markt (Anteil der Hausarztpatienten) optimal gegen die Qualitäts- und Spezialisierungsanreize. Das Marktergebnis ist im Allgemeinen ineffizient, da der Regulator nur über eine Stellschraube verfügt (den Preis), jedoch drei Regulierungsziele erreichen möchte (generelles Hausarztprinzip, optimale Qualität und optimale Spezialisierung). Insbesondere wird das Ziel des generellen Hausarztprinzips verfehlt, wenn der Wettbewerb im Krankenhausmarkt hinreichend intensiv ist.

Auch wenn in Kapitel 4 wichtige Aspekte des Hausarztprinzips vernachlässigt wurden, so abstrahierten wir von den beiden erstgenannten positiven Wirkungen, dokumentiert die Analyse, dass eine differenziertere Betrachtung dieses einschneidenden Reformschritts geboten erscheint. Die Ausblendung jeglicher Wirkungen von Informationen auf das Marktergebnis führt zu einer verzerrten Einschätzung der durch das Hausarztprinzip er-

[5] Die Spielstufen sind wie folgt: 1. Der Regulator legt den Preis fest. 2. Die Krankenhäuser wählen ihre Spezialisierung. 3. Die Krankenhäuser wählen ihre Qualität. 4. Einige Patienten suchen einen Hausarzt auf und erhalten alle relevanten Informationen. 5. Alle Patienten fragen Krankenhausleistungen nach.

warteten Gewinne. Es wurde gezeigt, dass, selbst wenn das Hausarztprinzip kostenlos ist, eine generelle Einführung über den verschärften Wettbewerb schlecht sein kann. Es ist theoretisch denkbar, dass der Wettbewerbseffekt alle - d. h. inklusive der im Modell vernachlässigten - positiven Wirkungen des Hausarztprinzips überkompensieren.

6.4 Risikoselektion (Kapitel 5)

Im Rahmen des Gesundheitsstrukturgesetzes aus dem Jahre 1992 wurde der Wettbewerb in der Gesetzlichen Krankenversicherung (GKV) erheblich gestärkt.[6] Der Wettbewerbsdruck entsteht durch das den Versicherten seit 1996 eingeräumte Recht, ihre Krankenkasse im Wesentlichen frei zu wählen. In Kapitel 5 überprüften wir, ob dieser Wettbewerb gut organisiert ist oder ob über die bestehenden Regulierungen hinausgehende Eingriffe erforderlich sind. Im Mittelpunkt stand dabei die Frage, ob Betriebskrankenkassen erfolgreich Versicherte mit geringem Krankheitsrisiko selektieren können und sich damit gesellschaftlich unerwünschte Wettbewerbsvorteile sichern.

Ökonomen sprechen dem Wettbewerb im Krankenversicherungsmarkt im Allgemeinen drei positive Effekte zu. Erstens verspricht man sich durch Wettbewerb eine erhöhte Qualität medizinischer Versorgung. Zweitens erhofft man sich eine verbesserte Effizienz im System insgesamt und, drittens, eine stärkere Kundenorientierung der Krankenkassen. Dürfen die Versicherer mit den Anbietern medizinischer Leistungen Einzelverträge schließen, sind diese Effizienzgewinne leicht zu verwirklichen. So könnten die Krankenkassen nur mit kostengünstigen Anbietern hoher Qualität Versorgungsverträge abschließen. Aber auch wenn die Versicherer nur Kollektivverträge schließen können, wie es weitgehend in Deutschland durch das Monopol der Kassenärztlichen Vereinigungen der Fall ist, kann Wettbewerb wünschenswert sein. So verlieren ineffiziente Krankenkassen, d. h. Krankenkassen mit hohen Verwaltungskosten, durch ihre hohen Beitragssätze Mitglieder. Im Extremfall werden sie aus dem Markt verdrängt.

In der GKV gilt das sogenannte Diskriminierungsverbot, nach dem eine Krankenkasse allen Mitgliedern denselben prozentualen Beitragssatz abverlangen muss. Anders als in der privaten Krankenversicherung sind dem individuellen Krankheitsrisiko angepasste Prämien verboten. Hinter dieser Regulierung verbirgt sich ein Gerechtigkeitsargument. Zum großen Teil wird das individuelle Krankheitsrisiko als jenseits der Verantwortung des Individuums angesehen. So müssten beispielsweise die Beiträge für chronisch

[6] Dieser Abschnitt ist eine leicht veränderte Version von Nuscheler (2003).

Kranke um ein Vielfaches höher sein als die von nicht chronisch Kranken. Aus der Ex-ante-Sicht, d. h. wenn den Individuen ihr Krankheitsrisiko bei Abschluss der Versicherung noch nicht bekannt ist, ist das Diskriminierungsverbot auch aus Effizienzgründen wünschenswert. Es versichert die Individuen dagegen, sehr hohe Beiträge zahlen zu müssen, falls sich ein hohes Krankheitsrisiko herausstellt.

Jedoch hat die Gerechtigkeit ihren Preis. Dürfen die Krankenkassen keine risikoabhängigen Prämien verlangen, so erzielen sie mit Versicherten mit niedrigem Krankheitsrisiko Gewinne, durch Beitragszahler mit hohem Krankheitsrisiko dagegen Verluste. Die Krankenkassen haben also einen Anreiz, Risikoselektion zu betreiben, d. h. bevorzugt Beitragszahler mit geringem Krankheitsrisiko versichern zu wollen. Eine Krankenkasse könnte beispielsweise bewusst Verträge ausschließlich mit schlechten Ärzten schließen. Die geringe Qualität würde gerade Versicherte mit hohem Krankheitsrisiko abschrecken, da diese mit höherer Wahrscheinlichkeit eine Behandlung benötigen. Die Möglichkeit Einzelverträge zu schließen kann also zur Risikoselektion missbraucht werden und somit die erhofften Effizienzgewinne des Wettbewerbs in ihr Gegenteil verkehren.

Um mögliche Gewinne der Krankenkassen durch Risikoselektion zu reduzieren, wurde 1994 ein Risikostrukturausgleich (RSA) eingeführt. Die Prämieneinnahmen einer Krankenkasse sollen so der tatsächlichen Risikostruktur der Kasse angepasst werden. Im Optimalfall gäbe es keine Selektionsanreize mehr. Der deutsche RSA verwendet fünf Variablen zur Anpassung des Risikos: Einkommen, Alter, Geschlecht, Behinderung und Krankengeldanspruch. Wie internationale Studien gezeigt haben, ist ein solch einfacher RSA hochgradig unvollständig, so dass substanzielle Anreize zur Risikoselektion bestehen bleiben.

Es ist somit wenig überraschend, dass weitere Regulierungsmaßnahmen getroffen wurden. So müssen gesetzliche Krankenkassen jeden Bewerber aufnehmen (Kontrahierungszwang). Direkte Risikoselektion ist also nur schwer möglich, auch wenn gelegentlich behauptet wird, dass Krankenkassen Versicherungsverträge während der Bearbeitung „verlieren". Indirekte Risikoselektion wird durch die weitgehende Regulierung der Leistungspakete erschwert, da 95 Prozent der Leistungen reguliert sind. So können Krankenkassen beispielsweise nicht jegliche Krebstherapie verweigern.

Trotz der den Wettbewerb begleitenden Regulierungen gibt es eine Vielzahl von Hinweisen, dass die Krankenkassen sich aktiv um die Versicherten mit geringem Krankheitsrisiko bemühen. So wird selektiv geworben, etwa in Universitätsmensen und im Internet. Gut ausgebildete und gut informierte Personen sind in der Regel gesünder und verursachen den Krankenkassen daher geringere (erwartete) Leistungsausgaben. Mitunter werden Leistungen bzw. der Service durch Ausdünnung des Geschäftsstellennetzes oder durch das Verzögern von Kostenerstattungen reduziert. Auch dies schreckt vorwiegend die Versicherten mit hohem Krankheitsrisiko ab, da diese mit höherer Wahrscheinlichkeit auf den Service angewiesen sind. Schließlich werden auch gerade diese Versicherten gezielt von der Möglichkeit eines Wechsels informiert. Diese Liste ließe sich beliebig fortsetzen.

Allerdings handelt es sich nur um Hinweise auf Risikoselektion und nicht um „harte Evidenz". Ein Blick auf die Entwicklung der Beitragssätze (siehe Abbildung 5.1) deutet jedoch darauf hin, dass es im Verlauf des Wettbewerbs zu einer Trennung von Versicherten mit hohen und niedrigen Krankheitsrisiken gekommen ist. Wettbewerb und RSA führten zu einer Annäherung der Beitragssätze der Nicht-Betriebskrankenkassen (z. B. Allgemeine Ortskrankenkassen und Ersatzkassen), während die Betriebskrankenkassen (BKKs) deutlich geringere durchschnittliche Beitragssätze aufweisen. Es ist daher wenig verwunderlich, dass die BKKs eine Vielzahl von Mitgliedern gewinnen konnten - seit Einführung des Wettbewerbs 1996 ca. vier Millionen.

Eine Erklärung für die niedrigen Beitragssätze der BKKs liefert die Unvollständigkeit des RSA zusammen mit der - historisch gewachsenen - besseren Risikostruktur der BKKs. Die Divergenz der Beitragssätze seit 1998 deutet zudem darauf hin, dass die BKKs insbesondere die Gesunden für sich gewinnen konnten. Und tatsächlich sind die über das regulierte Leistungspaket hinausgehenden Leistungen der BKKs eher für Gesunde attraktiv (z. B. Vorsorgeuntersuchungen) und die der Nicht-BKKs eher für Kranke (z. B. Krebstherapie, Chiro-Therapie und Ernährungsberatung). Es wäre jedoch verfrüht, diese Befunde als Risikoselektion zu deklarieren und damit den Wettbewerb in der GKV zu diskreditieren.

Die Beobachtungen stehen (auch) in Einklang mit einer anderen Erklärung: Gesunde haben geringere Wechselkosten und wechseln daher bei gleichen Vorteilen eines Wechsels häufiger als Kranke. Die höheren Wechselkosten für Kranke haben mehrere Ursachen. So haben Kranke in der Regel wichtigere Dinge zu tun als ihre Krankenkasse zu wechseln, beispielsweise an Rehabilitationsmaßnahmen teilzunehmen. Außerdem weisen Kranke mit höherer Wahrscheinlichkeit ein hohes Krankheitsrisiko auf als Gesunde und sind daher durch wiederholten Konsum medizinischer Leistungen besser über ihre Krankenkasse informiert als Gesunde (Wie bekomme ich meine Kosten erstattet? Welche Extraleistungen bieten Sie an?). Diese Informationen gingen im Falle eines Wechsels verloren.

In Kapitel 5 wurden Daten des Sozio-Oekonomischen Panels (SOEP) von 1995 bis 2000 verwendet, um zu testen, ob die BKKs tatsächlich erfolgreich Risikoselektion betreiben. Dabei war die Trennung von Risikoselektion vom Transaktionskostenargument entscheidend. Als Annäherung des Risikotyps wurde ein auf Grundlage des selbsteingeschätzten Gesundheitszustands konstruierter Gesundheitsindex verwendet. Der Datensatz wurde in Nicht-BKK-Mitglieder und BKK-Mitglieder geteilt und alle relevanten Wechselbewegungen wurden untersucht.

Allgemein zeigte sich, dass Nichtwechsler einen signifikant schlechteren Gesundheitszustand aufweisen als Wechsler. Wie bereits oben ausgeführt, haben schlechte Risiken höhere Wechselkosten und wechseln dementsprechend seltener. Die Trennung von Risikoselektion von eben diesen Transaktionskosten erforderte die Betrachtung einer Wechslerreferenzgruppe. Für die Nicht-BKK-Mitglieder ergab sich ein stärkerer Effekt des Gesundheitszustandes auf die Wechselwahrscheinlichkeit zu einer BKK als zu einer Nicht-BKK. Mit anderen Worten, die Wechsler zu den BKKs sind gesünder als die Wechsler innerhalb der Nicht-BKKs. Dieses Ergebnis deutet auf Risikoselektion seitens der BKKs hin, da es offensichtlich einen über die Transaktionskosten hinausgehenden Gesundheitseffekt gibt. Jedoch ist der beobachtete Unterschied nicht statistisch signifikant.

Dennoch könnte die Hypothese, dass BKKs Risikoselektion betreiben, gestützt werden, wenn unter den BKK Mitgliedern die Gesunden innerhalb der BKKs wechseln, während die Kränkeren (oder die weniger Gesunden) eher zu einer Nicht-BKK wechseln. Aber auch hier ließ sich kein signifikanter Unterschied in den Gesundheitseffekten ausmachen.

Es ließen sich somit keine Hinweise auf (erfolgreiche) Risikoselektion finden, weder seitens der BKKs noch seitens der Nicht-BKKs. Die Wechselbewegungen begründen sich somit aus Transaktionskosten, d. h. aus dem positiven Zusammenhang von Wechselkosten und den erwarteten Kosten medizinischer Leistungen. Dennoch führt der unvollständige Risikostrukturausgleich zu einem systematischen Wettbewerbsvorteil der BKKs. Da dieser den positiven Wirkungen des Wettbewerbs entgegenwirkt, ist eine Verbesserung des RSA zwingend. Die für 2007 geplante Einführung eines morbiditätsorientierten RSA, der die erwarteten Leistungsausgaben durch Einbeziehung zusätzlicher Variablen besser abbildet als bisher, zielt in diese Richtung.

Die Verzerrungen des Wettbewerbs sind somit weniger auf den Wettbewerb selbst zurückzuführen (Risikoselektion) als auf dessen schlechte Organisation (unvollständiger RSA). Dem Wettbewerb in der GKV sollte mithin mehr „Vertrauen“ entgegen gebracht werden. Die Tatsache, dass beispielsweise Geschäftsstellennetze ausgedünnt werden oder dass manche Extraleistungen angeboten und manche nicht angeboten werden, kann durchaus den Vorstellungen der Versicherten entsprechen und ist damit ein positiver und deshalb gewünschter Effekt des Wettbewerbs. Spätestens nach der Verbesserung des RSA sollte daher eine deutliche Stärkung des Wettbewerbs angestrebt werden. Mögliche Dimensionen sind beispielsweise die Zulassung von Einzelverträgen, d. h. die Abschaffung des Monopols der Kassenärztlichen Vereinigungen, und die Ausdehnung von Selbstbehalttarifen auf die Pflichtversicherten in der GKV. Die Einführung von Bonusprogrammen ab Januar 2004 kann also nur als erster (kleiner) Schritt in Richtung mehr Wettbewerb verstanden werden.

Bibliography

[1] AFW Dienstleistungsgesellschaft, 2002. http://www.gesetzlichekrankenkassen.de/leistungsvergleich/leistungsvergleich.html

[2] Andersen, Hanfried H. and Johannes Schwarze, 1999, Kassenwahlentscheidungen in der GKV, Eine empirische Analyse, Arbeit und Sozialpolitik 5-6, 10-23.

[3] Arrow, Kenneth, 1963, Uncertainty and the Welfare Economics of Medical Care, American Economic Review 53, 941-973.

[4] D'Aspremont, Claude; Gabszewicz, Jean J. and Jacques-François Thisse, 1979, On Hotelling's "stability in competition", Econometrica 47, 1145-1150.

[5] Barros, Pedro Pita, 1998, The Black Box of Health Care Expenditure Growth Determinants, Health Economics 7, 533-544.

[6] Barros, Pedro Pita, 2003, Cream-Skimming, Incentives for Efficiency and Payment System, Journal of Health Economics 22, 419-443.

[7] Barros, Pedro Pita and X. Martinez-Giralt, 2002, Public and Private Provision of Health Care, Journal of Economics and Management Strategy 11, 109-133.

[8] Baye, Michael R. and John Morgan, 2001, Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets, American Economic Review 91, 454-474.

[9] Bensaid, Bernard and Jean-Philippe Lesne, 1996, Dynamic Monopoly Pricing with Network Externatilies, International Journal of Industrial Organization 14, 837-855.

[10] Benstetter, Franz and Achim Wambach, 2001, Strategic Interaction in the Market for Physician Services: the Treadmill Effect in a Fixed Budget System, CESifo Working Paper No. 427, Munich.

[11] Bester, Helmut, 1998, Quality Uncertainty Mitigates Product Differentiation, RAND Journal of Economics 29, 828-844.

[12] Bound, John, 1991, Self-Reported versus Objective Measures of Health in Retirement Models, Journal of Human Resources 26, 106-138.

[13] Brekke, Kurt R.; Nuscheler, Robert and Odd Rune Straume, 2002, Quality and Location Choices under Price Regulation, Discussion Paper FS IV 02-28, Wissenschaftszentrum Berlin.

[14] Brekke, Kurt R.; Nuscheler, Robert and Odd Rune Straume, 2003, Gatekeeping in Health Care, Working Papers in Economics, No. 10-03, University of Bergen.

[15] Bresnahan, Timothy F., 1981, Duopoly Models with Consistent Conjectures, American Economic Review 71, 934-945.

[16] Breyer, Friedrich and Mathias Kifmann, 2001, Optionen der Weiterentwicklung des Risikostrukturausgleichs in der GKV, DIW Discussion Papers, No. 236, Berlin.

[17] Breyer, Friedrich; Zweifel, Peter and Mathias Kifmann, 2003, Gesundheitsökonomie, 4th ed., Springer-Verlag, Berlin.

[18] Brito, Dagobert L.; Sheshinski, Eytan and Michael D. Intriligator, 1991, Externalities and Compulsory Vaccinations, Journal of Public Economics 45, 69-90.

[19] Buchner, Florian and Jürgen Wasem, 2003, Needs for Further Improvement: Risk Adjustment in the German Health Insurance System, Health Policy 65, 21-35.

[20] Bundesministerium für Arbeit und Sozialordnung, 2001, Statistisches Taschenbuch 2001.
http://www.bma.bund.de/download/statistiken/stat2001/stat2001.zip

[21] Bundesministerium für Gesundheit (BMG), 1999, Daten des Gesundheitswesens, Nomos Verlag, Baden-Baden.

[22] Bundesministerium für Gesundheit (BMG), 2001, Daten des Gesundheitswesens 2001, Schriftenreihe des BMG, Band 137, Nomos Verlagsgesellschaft, Baden Baden.

[23] Burdett, Kenneth and Kenneth L. Judd, 1983, Equilibrium Price Dispersion, Econometrica 51, 955-969.

[24] BusinessWeek Online, 2002, December 9, Vaccines are Getting a Booster Shot
http://www.businessweek.com/magazine/content/02_49/b3811060.htm

[25] Cabral, Luis M. B.; Salant, David J. and Glenn A. Woroch, 1999, Monopoly Pricing with Network Externalities, International Journal of Industrial Organization 17, 199-214.

[26] Calem, Paul S. and John A. Rizzo, 1995, Competition and Specialization in the Hospital Industry: An Application of Hotelling's Location Model, Southern Economic Journal 61, 1182-1198.

[27] Crossley, Thomas F. and Steven Kennedy, 2002, The Reliability of Self-Assessed Health Status, Journal of Health Economics 21, 643-658.

[28] Cutler, David M. and Sarah J. Reber, 1998, Paying for Health Insurance: The Trade-off Between Competition and Adverse Selection, Quarterly Journal of Economics 113, 433-466.

[29] Cutler, David M. and Richard J. Zeckhauser, 1998, Adverse Selection in Health Insurance, in: A. Garber, ed., Frontiers in Health Policy Research, Vol. 1, MIT Press, Cambridge, Massachusetts, 1-31.

[30] Cutler, David M. and Richard J. Zeckhauser, 2000, The Anatomy of Health Insurance, in: A. J. Culyer and J. P. Newhouse, eds., Handbook of Health Economics, Vol. 1, North-Holland, Amsterdam, 563-643.

[31] Darby, Michael R. and Edi Karny, 1973, Free Competition and the Optimal Amount of Fraud, Journal of Law and Economics 16, 67-88.

[32] Desmond, Christopher and Robert Greener, 2003, The Strategic Use and Potential Demand for an HIV Vaccine in Southern Afrika, World Bank Policy Research Working Paper 2977.

[33] Donabedian, Avedis, 1980, Explorations in Quality Assessment and Monitoring, Volume 1, The Definition of Quality and Approaches to its Assessment, Health Administration Press, Ann Arbor, Michigan.

[34] Dranove, David; Kessler, Daniel; McClelland, Mark and Mark Satterthwaite, 2003, Is More Information Better? The Effects of "Report Cards" on Health Care Providers, Journal of Political Economy 111, 555-588.

[35] Economides, Nicholas, 1984, The Principle of Minimum Differentiation Revisited, European Economic Review 24, 345-368.

[36] Economides, Nicholas, 1986, Minimal and Maximal Product Differentiation in Hotelling's Duopoly, Economics Letters 21, 67-71.

[37] Economides, Nicholas, 1989, Quality Variations and Maximal Variety Differentiation, Regional Science and Urban Economics 19, 21-29.

[38] Economides, Nicholas, 1993, Quality Variations in the Circular Model of Variety-Differentiated Products, Regional Science and Urban Economics 23, 235-257.

[39] Ellis, Randall P., 1989, Employee Choice of Health Insurance, Review of Economics and Statistics 71, 215-223.

[40] Ellis, Randall P., 1998, Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins, Journal of Health Economics 17, 537-555.

[41] Emons, Winand, 1997, Credence Goods and Fraudulent Experts, RAND Journal of Economics 28, 107-119.

[42] England, Sarah; Kaddar, Miloud; Nigam, Ashok and Matilde Pinto, 2001, Practice and Policies on User Fees for Immunization in Developing Countries, World Health Organization, Geneva.

[43] European Observatory on Health Care Systems, 1999, Health Care Systems in Transition: United Kingdom, Kopenhagen.

[44] European Observatory on Health Care Systems, 2000, Health Care Systems in Transition: Germany, Kopenhagen.

[45] EUROSTAT, 2001, EUROSTAT Yearbook 2001.
http://europa.eu.int/comm/eurostat/

[46] Ferris, Timothy G.; Chang, Yuchiao; Blumenthal, David and Steven D. Pearson, 2001, Leaving Gatekeeping Behind - Effects of Opening Access to Specialists for Adults in a Health Maintenance Organization, The New England Journal of Medicine 345, 1312-1317.

[47] Foster, E. Michael, 1997, Instrumental Variables for Logistic Regression: an Illustration, Social Science Research 26, 487-504.

[48] Francis, Peter J., 1997, Dynamic Epidemiology and the Market for Vaccinations, Journal of Public Economics 63, 383-406.

[49] Frey, Bruno S. and Alois Stutzer, 2002, What Can Economists Learn from Happiness Research?, Journal of Economic Literature 40, 402-435.

[50] Gabszewicz, Jean J. and Jacques-François Thisse, 1980, Entry (and Exit) in a Differentiated Industry, Journal of Economic Theory 22, 327-338.

[51] Gabszewicz, Jean J. and Jacques-François Thisse, 1992, Location, in: Handbook of game theory, eds. Aumann, R.J. and S. Hart, North-Holland, 281-304.

[52] Geoffard, Pierre-Yves and Tomas Philipson, 1997, Disease Eradication: Private versus Public Vaccination, American Economic Review 87, 222-230.

[53] Glied, Sherry, 2000, Managed Care, in: A. J. Culyer and J. P. Newhouse, eds., Handbook of Health Economics, Vol. 1, North-Holland, Amsterdam, 707-753.

[54] Gravelle, Hugh, 1999, Capitation Contracts: Access and Quality, Journal of Health Economics 18, 315-340.

[55] Gravelle, Hugh and Giuliano Masiero, 2000, Quality Incentives in a Regulated Market With Imperfect Information and Switching Costs: Capitation in General Practice, Journal of Health Economics 19, 1067-1088.

[56] Gravelle, Hugh and Ray Rees, 1992, Microeconomics, 2nd edition, Longman, London.

[57] Greene, William H., 1997, Econometric Analysis, 3rd edition, Prentice-Hall, New Jersey.

[58] Grossman, Michael, 1972, On the Concept of Health Capital and the Demand for Health, Journal of Political Economy 80, 223-255.

[59] Hillman, Alan L.; Pauly, Mark V. and Joseph J. Kerstein, 1989, How Do Financial Incentives Affect Physician's Decisions, Resource Use and Financial Performance in Health Maintenance Organizations?, New England Journal of Medicine 321, 86-92.

[60] Hinloopen, Jeroen and Charles van Marrewijk, 1999, On the Limits and Possibilities of the Principle of Minimum Differentiation, International Journal of Industrial Organization 17, 735-750.

[61] Holly, Alberto; Gardiol, Lucien; Domenighetti, Gianfranco and Brigitte Bisig, 1998, An Econometric Model of Health Care Utilization and Health Insurance in Switzerland, European Economic Review 42, 513-522.

[62] Hotelling, Harold, 1929, Stability in Competition, Economic Journal 39, 41-57.

[63] Jacobs, Klaus; Reschke, Peter; Cassel, Dieter and Jürgen Wasem, 2002, Zur Wirkung des Risikostrukturausgleichs in der gesetzlichen Krankenversicherung, Schriftenreihe des BMG, Band 140, Nomos Verlagsgesellschaft, Baden Baden.

[64] Kassenärztliche Bundesvereinigung (KBV), 2000, Grunddaten zur Vertragsärztlichen Versorgung in der BRD 1999.
http://www.kbv.de/publikationen/grunddaten.htm

[65] Kassenärztliche Bundesvereinigung (KBV), 2001, Grunddaten zur Vertragsärztlichen Versorgung in der BRD 2000.
http://www.kbv.de/publikationen/grunddaten.htm

[66] Kauf, Teresa L., 1999, Price Discrimination and Bargaining Power in the U.S. Vaccine Market, Implications for Childhood Immunization Policy, The Quarterly Review of Economics and Finance 39, 249-265.

[67] Keeler, Emmett B.; Carter, Grace and Joseph P. Newhouse, 1998, A Model of the Impact of Reimbursement Schemes on Health Plan Choice, Journal of Health Economics 17, 297-320.

[68] Kessing, Sebastian G. and Robert Nuscheler, 2003, Monopoly Pricing with Negative Network Effects: the Case of Vaccines, Discussion Paper SP II 2003-06, Wissenschaftszentrum Berlin.

[69] Kifmann, Mathias, 2002, Community Rating in Health Insurance and Different Benefit Packages, Journal of Health Economics 21, 719-737.

[70] Kifmann, Mathias, 2003, Risk Selection and Supplementary Health Insurance: The Swiss Approach, University of Konstanz, mimeo.

[71] Krähmer, Daniel, 2003, Entry and Experimentation in Oligopolistic Markets for Experience Goods, International Journal of Industrial Organization 21, 1201-1213.

[72] Kremer, Michael, 2002, Pharmaceuticals and the Developing World, Journal of Economic Perspectives 16, 67-90.

[73] Laffont, Jean-Jacques and Jean Tirole, 1993, A Theory of Incentives in Procurement and Regulation, MIT Press, Cambridge, Massachusetts.

[74] Lambertini, Luca, 2001, Vertical Differentiation in a Generalized Model of Spatial Competition, The Annals of Regional Science 35, 227-238.

[75] Lauterbach, Karl W. and Eberhard Wille, 2001, Modell eines fairen Wettbewerbs durch den Risikostrukturausgleich. http://www.medizin.uni-koeln.de/kai/igmg/endgutachten_rsa.pdf

[76] Lommerud, Kjell Erik and Lars Sørgard, 2003, Entry in Telecommunication: Customer Loyalty, Price Sensitivity and Access Prices, Information Economics and Policy 15, 55-72.

[77] Lyon, Thomas P., 1999, Quality Competition, Insurance, and Consumer Choice in Health Care Markets, Journal of Economics and Management Strategy 8, 545-580.

[78] Ma, Ching-to Albert, 1994, Health Care Payment Systems, Journal of Economics and Management Strategy 3, 93-112.

[79] Ma, Ching-to Albert and James F. Burgess, Jr., 1993, Quality Competition, Welfare, and Regulation, Journal of Economics 58, 153-173.

[80] Mason, Robin, 2000, Network Externalities and the Coase Conjecture, European Economic Review 44, 1981-1992.

[81] McGuire, Thomas G., 2000, Physician Agency, in: A. J. Culyer and J. P. Newhouse, eds., Handbook of Health Economics, Vol. 1, North-Holland, Amsterdam, 461-536.

[82] Médecins Sans Frontières, 2003, Press Release Feb 13, 2003. http://www.msf.org/content/page.cfm?articleid=FE4CC291-AEC3-4E78-A15E41777D91BFDA

[83] Nelson, Philipp, 1970, Information and Consumer Behaviour, Journal of Political Economy 78, 311-329.

[84] Neven, Damien and Jacques-François Thisse, 1990, On Quality and Variety Competition, in: Economic Decision-Making: Games, Econometrics and Optimisation, eds. Gabszewicz, J.J., Richard, J.-F. and L.A. Wosley, North-Holland, 175-199.

[85] New York Times, 2001, AIDS Drug Battle Deepens in Africa, March 8.

[86] Newhouse, Joseph P., 1996, Reimbursing Health Plans and Health Providers: Efficiency in Production versus Selection, Journal of Economic Literature 34, 1236-1263.

[87] Nicholson, Sean; Polsky, Daniel; Bundorf, Kate and Rebecca Stein, 2003, Observed and Unobserved Risk Selection in Employer-Sponsored Health Plans, University of Pennsylvania, mimeo.

[88] Novshek, William, 1980, Equilibrium in Simple Spatial (or Differentiated Product) Models, Journal of Economic Theory 22, 313-326.

[89] Nuscheler, Robert, 2002, Physician Reimbursement, Time-Consistency and the Quality of Care, Discussion Paper FS IV 02-16, Wissenschaftszentrum Berlin.

[90] Nuscheler, Robert, 2003, Mehr Wettbewerb wagen, WZB Mitteilungen 102, Wissenschaftszentrum Berlin, 36-40.

[91] Nuscheler, Robert and Thomas Knaus, 2003, Risk Selection in the German Public Health Insurance System, Wissenschaftszentrum Berlin, mimeo.

[92] OECD, 1992, The Reform of Health Care, A Comparative Analysis of Seven OECD Countries.

[93] Pauly, Mark V., 1984, Is Cream Skimming a Problem for the Competitive Medical Market?, Journal of Health Economics 3, 87-95.

[94] Pauly, Mark V., 1988, Is Medical Care Different? Old Questions, New Answers, Journal of Health Politics, Policy and Law 13, 227-237.

[95] Pauly, Mark V., 2000, Insurance Reimbursement, in: A. J. Culyer and J. P. Newhouse, eds., Handbook of Health Economics, Vol. 1, North-Holland, Amsterdam, 537-560.

[96] Philipson, Tomas, 1996, Private Vaccination and Public Health: An Emprirical Examination for U.S. Measles, Journal of Human Resources 31, 611-630.

[97] Philipson, Tomas, 2000, Economic Epidemiology and Infectious Diseases, in: A. J. Culyer and J. P. Newhouse, eds., Handbook of Health Economics, Vol. 1, Elsevier Science, Amsterdam.

[98] Philipson, Tomas and Stéphane Mechoulan, 2003, Intellectual Property and External Consumption Effects: Generalizations from Pharmaceutical Markets, NBER Working Paper No. 9598, Cambridge, Massachusetts.

[99] Pohlmeier, Winfried and Volker Ulrich, 1992, Determinanten des Gesundheitszustandes: Ein empirischer Ansatz zur Outputmessung im Gesundheitswesen bei partieller Information, Zeitschrift für Wirtschafts- und Sozialwissenschaften (ZWS) 112, 219-238.

[100] Pope, Gregory C.; Adamache, Killard W.; Khandker, Rezaul K. and E. G. Walsh, 1998, Evaluating Alternative Risk Adjusters for Medicare, Health Care Financing Review 20, 109-129.

[101] Pratt, John W., 1964, Risk Aversion in the Small and in the Large, Econometrica 32, 122-136.

[102] Projektgruppe Sozio-oekonomisches Panel, 1995, Das Sozio-oekonomische Panel 1994, Vierteljahreshefte zur Wirtschaftsforschung, 64, 5-13.

[103] Reekie, Duncan, 2000, South Africa's Battle with AIDS and Drug Prices, National Center for Policy Analysis, Brief Analysis No. 334, Dallas, Texas.
http://www.ncpa.org/ba/ba334/ba334.html

[104] Reiss T. and E. Strauss, 1998, Vaccines - Patenting Dynamics of a Powerful Healthcare Tool, Expert Opinion on Therapeutic Patents 8, 951-958.

[105] Riordan, Michael R., 1986, Monopolistic Competition with Experience Goods, Quarterly Journal of Economics 101, 265-279.

[106] Ronning, Gerd, 1991, Mikroökonometrie, Springer-Verlag, Berlin.

[107] Rothschild, Michael and Joseph E. Stiglitz, 1976, Equilibrium in Competitive Insurance Markets: an Essay in the Economics of Incomplete Information, Quarterly Journal of Economics 90, 624-649.

[108] Russel, Phillip K., 2002, Developement of Vaccines to Meet Public Health Needs: Incentives and Obstacles, Risk 7. http://www.piercelaw.edu/risk/vol7/summer/russell.htm

[109] Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (SVRKAiG), 1998, Gesundheitswesen in Deutschland, Kostenfaktor und Zukunftsbranche, Band 2, Nomos Verlag, Baden-Baden.

[110] Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (SVRKAiG), 2002, Bedarfsgerechtigkeit und Wirtschaftlichkeit, Band 3, Über-, Unter- und Fehlversorgung, Nomos Verlag, Baden-Baden.

[111] Salop, Steven C., 1979, Monopolistic Competition with Outside Goods, Bell Journal of Economics 10, 141-156.

[112] Santerre, Rexford E. and Stephen P. Neun, 1996, Health Economics: Theories, Insights, and Industry Studies, Irwin, Chicago.

[113] Schellhorn, Martin, 2001, The Effect of Variable Health Insurance Deductibles on the Demand for Physician Visits, Health Economics 10, 441-456.

[114] Schulenburg, Matthias Graf von der and Wolfgang Greiner, 2000, Gesundheitskonomik, Mohr Siebeck, Tübingen.

[115] Schultz, Christian, 2002, Market Transparency and Product Differentiation, CIE Discussion Paper 2002-02, Centre for Industrial Economics, University of Copenhagen.

[116] Schultz, Christian, 2003, Transparency on the Consumer Side and Tacit Collusion, European Economic Review, forthcoming.

[117] Schut, Frederik T.; Greß, Stefan and Jürgen Wasem, 2003, Consumer Price Sensitivity and Social Health Insurer Choice in Germany and the Netherlands, International Journal of Health Care Finance and Economics 3, 117-138.

[118] Schwarze, Johannes and Hanfried H. Andersen, 2001, Kassenwechsel in der Gesetzlichen Krankenversicherung: Welche Rolle spielt der Beitragssatz?, Schmollers Jahrbuch: Zeitschrift fuer Wirtschafts- und

Sozialwissenschaften (Journal of Applied Social Science Studies) 121, 581-602.

[119] Scott, Anthony, 2000, Economics of General Practice, in: Culyer, A.J. and J.P. Newhouse (Eds.), Handbook of Health Economics, Vol. 1. North-Holland, Amsterdam, 1175-1200.

[120] Statistisches Bundesamt, 2001, Fachserie 12, Gesundheitswesen, Reihe S2, Ausgaben für die Gesundheit, Metzler-Poeschel, Stuttgart.

[121] Strombom, Bruce A.; Buchmueller, Thomas C. and Paul J. Feldstein, 2002, Switching Costs, Price Sensitivity and Health Plan Choice, Journal of Health Economics 21, 89-116.

[122] The Associated Press, 2001, October 24, Deal Reached on Cutting Cipro Price to Government.
http://sns.tribune.com/sns-anthrax-cipro.story

[123] Tirole, Jean, 1988, The Theory of Industrial Organization, MIT Press, Cambridge, Massachusetts.

[124] USA Today, 2001, October 29, Cipro Saga Exposes How Drugmakers Protect Profits. http://lists.essential.org/pipermail/ip-health/2001-October/002299.html

[125] USA Today, 2002, November 26, Smallpox Vaccine Carries a Dose of Risk. http://www.usatoday.com/news/health/2002-11-26-smallpox-side-usat_x.htm

[126] Van de Ven, Wynand P. M. M.; Beck, Konstantin; Buchner, Florian; Chernichovsky, Dov; Gardiol, Lucien; Holly, Alberto; Lamers, Leida M.; Schokkaert, Erik; Shmueli, Amir; Spycher, Stephan; Van de Voorde, Carine; van Vliet, René C. J. A.; Wasem, Jürgen and Irith Zmora, 2003, Risk Adjustment and Risk Selection on the Sickness Fund Insurance Market in Five European Countries, Health Policy 65, 75-98.

[127] Van de Ven, Wynand P. M. M. and Randall P. Ellis, 2000, Risk Adjustment in Competitive Health Plan Markets, in: A. J. Culyer and J. P. Newhouse, eds., Handbook of Health Economics, Volume 1, North-Holland, Amsterdam, 755-845.

[128] Van de Ven, Wynand P. M. M. and René C. J. A. van Vliet, 1992, How Can We Prevent Cream Skimming in a Competitive Health Insurance Market?, in: Peter Zweifel and H. E. Frech III, eds., Health Economics Worldwide, Kluwer Academic Publishers, Dodrecht, 23-46.

[129] Van Doorslaer, Eddy and Andrew M. Jones, 2003, Inequalities in Self-Reported Health: Validation of a New Approach to Measurement, Journal of Health Economics 22, 61-87.

[130] Van Vliet, René C. J. A., 1992, Predictability of Individual Health Care Expenditures, Journal of Risk and Insurance 59, 443-460.

[131] Varian, Hal R., 1980, A Model of Sales, American Economic Review 70, 651-659.

[132] Wagner, Gerd G.; Burkhauser, Richard V. and Friederike Behringer, 1993, The English Language Public Use File of the German Socio-Economic Panel, Journal of Human Resources 28, 429-433.

[133] Wagstaff, Adam and Eddy van Doorslaer, 1994, Measuring Inequalities in Health in the Presence of Multiple-Category Morbidity Indicators, Health Economics 3, 281-291.

[134] Weisbrod, Burton A., 1991, The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment, Journal of Economic Literature 29, 523-552.

[135] Wilson, Charles, 1977, A Model of Insurance Markets with Incomplete Information, Journal of Economic Theory 16, 167-207.

[136] Wolinsky, Asher, 1984, Product Differentiation with Imperfect Information, Review of Economic Studies 51, 53-61.

[137] Wooldridge, Jeffrey M., 2002, Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, Massachusetts.

[138] World Health Organization (WHO), 2000, World Health Report 2000, Geneva.

[139] World Health Organization (WHO), 2001, Measles, Mortality Reduction and Regional Elimination, Strategic Plan 2001-2005, Geneva.

[140] World Health Organization (WHO), 2003, Global Eradication Initiative, Progress 2002, Geneva.

[141] Zweifel, Peter and Friedrich Breyer, 1997, Health Economics, Oxford University Press, Oxford.

FINANZWISSENSCHAFTLICHE SCHRIFTEN

Band 1 Werner Steden: Finanzpolitik und Einkommensverteilung. Ein Wachstums- und Konjunkturmodell der Bundesrepublik Deutschland. 1979.

Band 2 Rainer Hagemann: Kommunale Finanzplanung im föderativen Staat. 1976.

Band 3 Klaus Scherer: Maßstäbe zur Beurteilung von konjunkturellen Wirkungen des öffentlichen Haushalts. 1977.

Band 4 Brita Steinbach: "Formula Flexibility" - Kritische Analyse und Vergleich mit diskretionärer Konjunkturpolitik. 1977.

Band 5 Hans-Georg Petersen: Personelle Einkommensbesteuerung und Inflation. Eine theoretisch-empirische Analyse der Lohn- und veranlagten Einkommensteuer in der Bundesrepublik Deutschland. 1977.

Band 6 Friedemann Tetsch: Raumwirkungen des Finanzsystems der Bundesrepublik Deutschland. Eine Untersuchung der Auswirkungen der Finanzreform von 1969 auf die Einnahmenposition der untergeordneten Gebietskörperschaften und ihrer regionalpolitischen Zieladäquanz. 1978.

Band 7 Wilhelm Pfähler: Normative Theorie der fiskalischen Besteuerung. Ein methodologischer und theoretischer Beitrag zur Integration der normativen Besteuerungstheorie in der Wohlfahrtstheorie. 1978.

Band 8 Wolfgang Wiegard: Optimale Schattenpreise und Produktionsprogramme für öffentliche Unternehmen. Second-Best Modelle im finanzwirtschaftlichen Staatsbereich. 1978.

Band 9 Hans P. Fischer: Die Finanzierung des Umweltschutzes im Rahmen einer rationalen Umweltpolitik. 1978.

Band 10 Rainer Paulenz: Der Einsatz finanzpolitischer Instrumente in der Forschungs- und Entwicklungspolitik. 1978.

Band 11 Hans-Joachim Hauser: Verteilungswirkungen der Staatsverschuldung. Eine kreislauftheoretische Inzidenzbetrachtung. 1979.

Band 12 Gunnar Schwarting: Kommunale Investitionen. Theoretische und empirische Untersuchungen der Bestimmungsgründe kommunaler Investitionstätigkeit in Nordrhein-Westfalen 1965-1972. 1979.

Band 13 Hans-Joachim Conrad: Stadt-Umland-Wanderung und Finanzwirtschaft der Kernstädte. Amerikanische Erfahrungen, grundsätzliche Zusammenhänge und eine Fallstudie für das Ballungsgebiet Frankfurt am Main. 1980.

Band 14 Cay Folkers: Vermögensverteilung und staatliche Aktivität. Zur Theorie distributiver Prozesse im Interventionsstaat. 1981.

Band 15 Helmut Fischer: US-amerikanische Exportförderung durch die DISC-Gesetzgebung. 1981.

Band 16 Günter Ott: Einkommensumverteilungen in der gesetzlichen Krankenversicherung. Eine quantitative Analyse. 1981.

Band 17 Johann Hermann von Oehsen: Optimale Besteuerung. *(Optimal Taxation)*. 1982.

Band 18 Richard Kössler: Sozialversicherungsprinzip und Staatszuschüsse in der gesetzlichen Rentenversicherung. 1982.

Band 19 Hinrich Steffen: Zum Handlungs- und Entscheidungsspielraum der kommunalen Investitionspolitik in der Bundesrepublik Deutschland. 1983.

Band 20 Manfred Scheuer: Wirkungen einer Auslandsverschuldung des Staates bei flexiblen Wechselkursen. 1983.

Band 21 Christian Schiller: Staatsausgaben und crowding-out-Effekte. Zur Effizienz einer Finanzpolitik keynesianischer Provenienz. 1983.

Band 22 Hannelore Weck: Schattenwirtschaft: Eine Möglichkeit zur Einschränkung der öffentlichen Verwaltung? Eine ökonomische Analyse. 1983.

Band 23 Wolfgang Schmitt: Steuern als Mittel der Einkommenspolitik. Eine Ergänzung der Stabilitätspolitik? 1984.

Band 24 Wolfgang Laux: Erhöhung staatswirtschaftlicher Effizienz durch budgetäre Selbstbeschränkung? Zur Idee einer verfassungsmäßig verankerten Ausgabengrenze. 1984.

Band 25 Brita Steinbach-van der Veen: Steuerinzidenz. Methodologische Grundlagen und empirisch-statistische Probleme von Länderstudien. 1985.

Band 26 Albert Peters: Ökonomische Kriterien für eine Aufgabenverteilung in der Marktwirtschaft. Eine deskriptive und normative Betrachtung für den Allokationsbereich. 1985.

Band 27 Achim Zeidler: Möglichkeiten zur Fortsetzung der Gemeindefinanzreform. Eine theoretische und empirische Analyse. 1985.

Band 28 Peter Bartsch: Zur Theorie der längerfristigen Wirkungen 'expansiver' Fiskalpolitik. Eine dynamische Analyse unter besonderer Berücksichtigung der staatlichen Budgetbeschränkung und ausgewählter Möglichkeiten der öffentlichen Defizitfinanzierung. 1986.

Band 29 Konrad Beiwinkel: Wehrgerechtigkeit als finanzpolitisches Verteilungsproblem. Möglichkeiten einer Kompensation von Wehrungerechtigkeit durch monetäre Transfers. 1986.

Band 30 Wolfgang Kitterer: Effizienz- und Verteilungswirkungen des Steuersystems. 1986.

Band 31 Heinz Dieter Hessler: Theorie und Politik der Personalsteuern. Eine Kritik ihrer Einkommens- und Vermögensbegriffe mit Blick auf die Leistungsfähigkeitstheorie. 1994.

Band 32 Wolfgang Scherf: Die beschäftigungspolitische und fiskalische Problematik der Arbeitgeberbeiträge zur Rentenversicherung. Eine Auseinandersetzung mit der Kritik an der lohnbezogenen Beitragsbemessung. 1987.

Band 33 Andreas Mästle: Die Steuerunion. Probleme der Harmonisierung spezifischer Gütersteuern. 1987.

Band 34 Günter Ott: Internationale Verteilungswirkungen im Finanzausgleich der Europäischen Gemeinschaften. 1987.

Band 35 Heinz Haller: Zur Frage der zweckmäßigen Gestalt gemeindlicher Steuern. Ein Diskussionsbeitrag zur Gemeindesteuerreform. 1987.

Band 36 Thomas Kuhn: Schlüsselzuweisungen und fiskalische Ungleichheit. Eine theoretische Analyse der Verteilung von Schlüsselzuweisungen an Kommunen. 1988.

Band 37 Walter Hahn: Steuerpolitische Willensbildungsprozesse in der Europäischen Gemeinschaft. Das Beispiel der Umsatzssteuer-Harmonisierung. 1988.

Band 38 Ulrike Hardt: Kommunale Finanzkraft. Die Problematik einer objektiven Bestimmung kommunaler Einnahmemöglichkeiten in der gemeindlichen Haushaltsplanung und im kommunalen Finanzausgleich. 1988.

Band 39 Jochen Michaelis: Optimale Finanzpolitik im Modell überlappender Generationen. 1989.

Band 40 Bernd Raffelhüschen: Anreizwirkungen der sozialen Alterssicherung. Eine dynamische Simulationsanalyse. 1989.

Band 41 Berend Diekmann: Die Anleihe- und Darlehenstransaktionen der Europäischen Gemeinschaften. 1990.

Band 42 Helmut Kaiser: Konsumnachfrage, Arbeitsangebot und optimale Haushaltsbesteuerung. Theoretische Ergebnisse und mikroökonometrische Simulation für die Bundesrepublik Deutschland. 1990.

Band 43 Rüdiger von Kleist: Das Gramm-Rudman-Hollings-Gesetz. Ein gescheiterter Versuch der Haushaltskonsolidierung. 1991.

Band 44 Rolf Hagedorn: Steuerhinterziehung und Finanzpolitik. Ein theoretischer Beitrag unter besonderer Berücksichtigung der Hinterziehung von Zinserträgen. 1991.

Band 45 Cornelia S. Behrens: Intertemporale Verteilungswirkungen in der gesetzlichen Krankenversicherung der Bundesrepublik Deutschland. 1991.

Band 46 Peter Saile: Ein ökonomischer Ansatz der Theorie der intermediären Finanzgewalten – Die Kirchen als Parafisci. 1992.

Band 47 Peter Gottfried: Die verdeckten Effizienzwirkungen der Umsatzsteuer. Eine empirische allgemeine Gleichgewichtsanalyse. 1992.

Band 48 Andreas Burger: Umweltorientierte Beschäftigungsprogramme. Eine Effizienzanalyse am Beispiel des "Sondervermögens Arbeit und Umwelt". 1992.

Band 49 Jeanette Malchow: Die Zuordnung verteilungspolitischer Kompetenzen in der Europäischen Gemeinschaft. Eine Untersuchung aufgrund einer Fortentwicklung der ökonomischen Theorie des Föderalismus. 1992.

Band 50 Barbara Seidel: Die Einbindung der Bundesrepublik Deutschland in die Europäischen Gemeinschaften als Problem des Finanzausgleichs. 1992.

Band 51 Ralph Wiechers: Markt und Macht im Rundfunk. Zur Stellung der öffentlich-rechtlichen Rundfunkanstalten im dualen Rundfunksystem der Bundesrepublik Deutschland. 1992.

Band 52 Klaus Eckhardt: Probleme einer Umweltpolitik mit Abgaben. 1993.

Band 53 Oliver Schwarzkopf: Die Problematik unterschiedlicher Körperschaftsteuersysteme innerhalb der EG. 1993.

Band 54 Thorsten Giersch: Bergson-Wohlfahrtsfunktion und normative Ökonomie. 1993.

Band 55 Li-Fang Chou: Selbstbeteiligung bei Arzneimitteln aus ordnungspolitischer Sicht. Das Beispiel der Bundesrepublik Deutschland. 1993.

Band 56 Harald Schlee: Einkommensteuerliche Behandlung von Transferzahlungen. Zur Neuordnung der Familienbesteuerung sowie der Besteuerung von Versicherungsleistungen und Sozialtransfers. 1994.

Band 57 Alexander Spermann: Kommunales Krisenmanagement. Reaktionen baden-württembergischer Stadtkreise auf steigende Sozialhilfekosten und Einnahmenausfälle (1980-92). 1993.

Band 58 Otto Roloff / Sibylle Brander / Ingo Barens / Claudia Wesselbaum: Direktinvestitionen und internationale Steuerkonkurrenz. 1994.

Band 59 Claudia Wesselbaum-Neugebauer: Internationale Steuerbelastungsvergleiche. 1994.

Band 60 Stephanie Miera: Kommunales Finanzsystem und Bevölkerungsentwicklung. Eine Analyse des kommunalen Finanzsystems vor dem Hintergrund der sich abzeichnenden Bevölkerungsentwicklung am Beispiel Niedersachsens unter besonderer Berücksichtigung des Landkreises Wolfenbüttel und seiner Gemeinden. 1994.

Band 61 Wolfgang Scherf: Die Bedeutung des kaldorianischen Verteilungsmechanismus für die gesamtwirtschaftlichen Wirkungen der staatlichen Neuverschuldung. 1994.

Band 62 Rainer Volk: Vergleich der Vergünstigungseffekte der verschiedenen investitionsfördernden Maßnahmen. 1994.

Band 63 Hans-Georg Napp: Kommunale Finanzautonomie und ihre Bedeutung für eine effiziente lokale Finanzwirtschaft. 1994. 2., unveränderte Auflage 1994.

Band 64 Bernd Rahmann / Uwe Steinborn / Günter Vornholz: Empirische Analyse der Autonomie lokaler Finanzwirtschaften in der Europäischen Gemeinschaft. 1994.

Band 65 Carsten Kühl: Strategien zur Finanzierung der Altlastensanierung. 1994.

Band 66 Stephan Boll: Intergenerationale Umverteilungswirkungen der Fiskalpolitik in der Bundesrepublik Deutschland. Ein Ansatz mit Hilfe des Generational Accounting. 1994.

Band 67 Karl Justus Bernhard Neumärker: Finanzverfassung und Staatsgewalt in der Demokratie. Ein Beitrag zur konstitutionellen Finanztheorie. 1995.

Band 68 Christian Haslbeck: Zentrale versus dezentrale Internalisierung externer Effekte bei unvollständiger Information. 1995.

Band 69 Regina Müller: Horizontale oder vertikale Transfers zur Durchsetzung eines horizontalen Finanzausgleichs. 1995.

Band 70 Christian Hockenjos: Öffentliche Sportförderung in der Bundesrepublik Deutschland. Darstellung und finanztheoretische Analyse. 1995.

Band 71 Manfred Rosenstock: Die Kontrolle und Harmonisierung nationaler Beihilfen durch die Kommission der Europäischen Gemeinschaften. 1995.

Band 72 Christian Rüsch: Wohnungsbau- und Wohneigentumspolitik im Rahmen der Einkommensteuer. Eine Analyse unter steuersystematischen, verteilungspolitischen und fiskalischen Aspekten. 1996.

Band 73 Stephan Winters: Die kollektive Vorsorge für den Pflegefall im Alter. Eine Untersuchung am Beispiel der gesetzlichen Pflegeversicherung in den Niederlanden. 1996.

Band 74 Knut Blind: Allokationsineffizienzen auf Sicherheitsmärkten: Ursachen und Lösungsmöglichkeiten. Fallstudie: Informationssicherheit in Kommunikationssystemen. 1996.

Band 75 Barbara Petrick-Rump: Ökonomische Wirkungen von Steueramnestien. Untersuchung konkreter Erfahrungen ausgewählter Länder mit dem Einsatz von Steueramnestien anhand eines effizienten Steueramnestieprogramms. 1996.

Band 76 Georg Hirte: Effizienzwirkungen von Finanzausgleichsregelungen. Eine Empirische Allgemeine Gleichgewichtsanalyse für die Bundesrepublik Deutschland. 1996.

Band 77 Ulrike Kirchhoff: Die rheinland-pfälzischen Gemeinden im System des Finanzausgleichs. 1996.

Band 78 Kerstin Keil: Der soziale Mietwohnungsbau: Mängel und Alternativen. 1996.

Band 79 Bernhard Manzke: Kinderlastenausgleich versus verstärkte Einwanderung. Alternative Ansätze zur langfristigen Sicherung der Gesetzlichen Rentenversicherung. 1997.

Band 80 Hariolf M. Wenzler: Institutionenökonomik und öffentliche Finanzkontrolle. Eine Analyse am Beispiel der Europäischen Union. 1997.

Band 81 Joachim Nagel: Supply-Side Policy in den USA. Eine theoretische und empirische Analyse der angebotsorientierten Wirtschaftspolitik Reagans unter besonderer Berücksichtigung steuerlicher Aspekte. 1997.

Band 82 Heinz Lampert: Krise und Reform des Sozialstaates. 1997.

Band 83 Monika Hanswillemenke / Bernd Rahmann: Zwischen Reformen und Verantwortung für Vollbeschäftigung. Die Finanz- und Haushaltspolitik der sozial-liberalen Koalition von 1969 bis 1982. 1997.

Band 84 Berthold Fürst: Die Maastrichter Budgetkriterien im Konflikt mit der Verschuldungsautonomie der deutschen Gebietskörperschaften. 1997.

Band 85 Burkhard Pahnke: Einkommensorientierte Förderung des sozialen Mietwohnungsbaues. Bestandsaufnahme und Kritik. 1998.

Band 86 Judith Safford: Staatsverschuldung im Vereinigten Königreich. Die öffentliche Verschuldung unter der Konservativen Regierung von 1979-1994. Ursachen und Auswirkungen. 1998.

Band 87 Ralf Oberheide: Die Bekämpfung der Steuerumgehung. 1998.

Band 88 Achim Truger: Die neue Finanzwissenschaft zwischen Realitätsferne und Irrelevanz der Annahmen. Eine methodologische Analyse potentieller Verteidigungsstrategien der neuen Finanzwissenschaft gegen den Vorwurf der Realitätsferne ihres entscheidungstheoretischen Fundamentes. 1998.

Band 89 Karin Bickel: Familienbezogene Elemente im System der gesetzlichen Rentenversicherung. Unter besonderer Berücksichtigung von Ein-Eltern-Familien. 1999.

Band 90 Wolfgang Scherf: Schlüsselzuweisungen und Kreisumlage. Die Problematik der Finanzierung der Landkreise am Beispiel des kommunalen Finanzausgleichs von Rheinland-Pfalz. 1998.

Band 91 Sandra Ehmann: Familienpolitik in Frankreich und Deutschland – ein Vergleich. 1999.

Band 92 Hendrik Suermann: Einkommensteuerliche Behandlung von Währungsgewinnen und -verlusten. Eine finanzwissenschaftliche Analyse des Steuerrechts in den USA und in Deutschland. 1999.

Band 93 Rolf Bösinger: Die Neuordnung des bundesstaatlichen Finanzausgleichs 1995. Eine theoretische und empirische Analyse unter Berücksichtigung von allokationstheoretischen und polit-ökonomischen Gesichtspunkten. 1999.

Band 94 Ulrich Ermschel: Finanzwirtschaftliche Konsequenzen beim Übergang auf das Ursprungslandprinzip im Europäischen Binnenmarkt. Eine Untersuchung am Beispiel des unvollkommenen oligopolistischen Neufahrzeugmarktes. 1999.

Band 95 Ute Hansen: Überwälzte Leistungen der Administration. Eine empirische und theoretische Analyse. 2000.

Band 96 Hans-Werner Seiler: Zur Durchsetzung der Einmalbesteuerung deutscher Körperschaftsgewinne. Strategien zur Vermeidung der im deutschen Körperschaftsteuersystem angelegten Benachteiligung ausländischer Anteilseigner. Eine finanzwissenschaftliche Analyse. 2000.

Band 97 Steffen Meyer: Zwischenstaatliche Finanzzuweisungen im zusammenwachsenden Europa. Zur Gestaltung eines Finanzausgleichs für die Europäische Union. 2000.

Band 98 Marion Hübner: Ökodumping? Umweltpolitik in internationalen Oligopolmärkten. 2000.

Band 99 Christhart Bork: Steuern, Transfers und private Haushalte. Eine mikroanalytische Simulationsstudie der Aufkommens- und Verteilungswirkungen. 2000.

Band 100 Norbert Eichler: Die Probleme des Gemeindefinanzausgleichs im Kooperativen Föderalismus. Eine ökonomische Analyse am Beispiel des Bundeslandes Nordrhein-Westfalen. 2000.

Band 101 Wolfgang Scherf: Der Länderfinanzausgleich in Deutschland. Ungelöste Probleme und Ansatzpunkte einer Reform. 2000.

Band 102 Stefan Dietrich Josten: Staatsverschuldung, intertemporale Allokation und Wirtschaftswachstum. Eine theoretische Analyse staatlicher Verschuldungspolitik in Modellen exogenen und endogenen Wachstums. 2000.

Band 103 Axel Breitbach: Steuerhinterziehung und Schattenwirtschaft aus gesamtwirtschaftlicher Sicht. 2000.

Band 104 Alexander Spermann: Negative Einkommensteuer, Lohnsubventionen und Langzeitarbeitslosigkeit. 2001.

Band 105 Michael Broer: Der kommunale Finanzausgleich in Hessen. Historische Darstellung und ökonomische Analyse unter besonderer Berücksichtigung der Schlüsselzuweisungen. 2001.

Band 106 Jan-Paul Ritscher: Der Einsatz von Finanzderivaten unter einer modernisierten Schuldenstrukturpolitik des Bundes. 2002.

Band 107 Martin Gasche: Dynamische Fiskalpolitik. Makroökonomische Wirkungen der Fiskalpolitik in einem Real Business Cycle-Modell. 2003.

Band 108 Felix Brosius: Internationaler Steuerwettbewerb und Koordination der Steuersysteme. 2003.

Band 109 Claudia Hensberg: Eigennützige Regierungen im fiskalischen Wettbewerb um Kapital. 2003.

Band 110 Hans-Martin Grambeck: Konsumsteuerreformen und Konsumbesteuerung. Eine vergleichende Analyse und Bewertung verschiedener Konsumsteuermodelle unter besonderer Berücksichtigung der Probleme in offenen Volkswirtschaften. 2003.

Band 111 Antje Draheim: Probleme der finanzpolitischen Willensbildung in Europa. Eine kritische Analyse der europäischen Haushalts- und Finanzverfassung. 2004.

Band 112 Robert Nuscheler: On Competition and Regulation in Health Care Systems. 2005.

www.peterlang.de

Peter Lang · Europäischer Verlag der Wissenschaften

Katharina Janus

Managing Health Care in Private Organizations

Transaction Costs, Cooperation and Modes of Organization in the Value Chain

Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien, 2003.
380 pp., num. fig., num. tab.
ISBN 3-631-51567-7 / US-ISBN 0-8204-6505-4 · pb. € 56.50*

Health care delivery systems in all major industrial countries approach a crisis and there is no panacea so far to rising health care costs, in particular as an aging population and advanced medical technology further drive these costs. This dissertation does not attempt to provide a solution to the *macroeconomic* debate about health care costs. It rather focuses on *microeconomics* and how health systems (private entities) organize themselves – cooperate, integrate, and disintegrate – in a specific market, the San Francisco Bay Area, U.S.A. Based on transaction cost economics alternative modes of governance in the health care market are examined. For this purpose a long-term case study has been conducted in the Bay Area in order to evaluate *how health care transactions should be organized to economize on transaction costs, which experiences have been made over the last years and which strategies are recommendable for future reorganization* in the U.S. health care market and other countries. In addition, implications for the German health care market are discussed as well. A conclusion finally summarizes the findings and outlines further areas of research.

Contents: The Transaction Costs Theory as an Economic Basis · Assessing the Health Care Market with Respect to the Transaction Costs Theory and Modes of Governance · Concepts for Economizing on Transaction Costs in Hybrid Forms of Organization · Cooperation and Modes of Organization in the Bay Area Health Care Market – An Empirical Case Study · Pharmaceutical Companies – Taking a More Comprehensive approach to Health Care Delivery · Implications for the German Health Care Market

Frankfurt am Main · Berlin · Bern · Bruxelles · New York · Oxford · Wien
Distribution: Verlag Peter Lang AG
Moosstr. 1, CH-2542 Pieterlen
Telefax 00 41 (0) 32 / 376 17 27

*The €-price includes German tax rate
Prices are subject to change without notice

Homepage http://www.peterlang.de