Government
Institute for
Economic Research

Publications 65

Industrial organization studies on pharmaceutical markets

Tanja Saxell

VATT PUBLICATIONS

65

Industrial organization studies on pharmaceutical markets

Tanja Saxell

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Helsinki 2013

ISBN 978-952-274-099-1 (nid.)
ISBN 978-952-274-100-4 (PDF)

ISSN 0788-4990 (nid.)
ISSN 1795-3332 (PDF)

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Arkadiankatu 7, 00100 Helsinki, Finland

Edita Prima Oy
Helsinki, November 2013

Cover design: Niilas Nordenswan

Industrial Organization Studies on Pharmaceutical Markets

by

Tanja Saxell

Discipline of Economics
Department of Political and Economic Studies
Faculty of Social Sciences
University of Helsinki

Academic dissertation to be presented, by the permission of the Faculty of Social Sciences of the University of Helsinki, for public examination in Economicum Lecture Hall, Arkadiankatu 7, on January 10th 2014, at 12 p.m.

Abstract

My thesis consists of three essays on the industrial organization of pharmaceutical markets. In Chapter 1, I introduce the three essays and present the main results. In Chapter 2, I quantify how uncertainty affects medical decision-making by physicians. I estimate a dynamic model of demand where physicians may learn about the effectiveness of drug treatments from their prescription experiences. In the model, physicians may want experiment new drugs for their patients to get information that is valuable for their future drug choices. At the same time, risk aversion can make physicians reluctant to try less well-known, but potentially superior products. Using a rich Finnish data on cholesterol drug prescriptions, I study the roles of experimentation and learning in drug demand. I find that the effectiveness of cholesterol drugs varies across patients which creates uncertainty to medical decision-making. My results suggest that uncertainty and risk aversion create substantial switching costs in drug demand. I also find that if physicians became more willing to experiment with their treatment choices, the process of learning would improve and the efficiency of medical decision-making would increase.

In Chapter 3, I develop a framework for analyzing demand for experience goods where agents can learn product quality both from their own experiences and from the past behavior of their peers. I modify the standard theoretical models with social learning, by allowing agents to make repeated choices. I focus on the medical decision-making of physicians under uncertainty about the effectiveness of drug treatment. I ask whether information on the past choices of other physicians improves the efficiency of drug choices. My estimates from the Finnish market for cholesterol drugs suggest that treatment patterns relying heavily on the past choices of other physicians can lead to over-prescribing. I show that continuity of care - in the sense of a patient repeatedly consulting the same doctor - is an efficient policy to limit over-prescribing and to promote learning.

Finally, in Chapter 4, I explore how intellectual property rights change the competitive environment and technology flows between firms. Traditionally, stronger patents have been viewed to have an essential role in promoting innovation. Economic theory predicts that longer patents may hinder rather than stimulate innovation by increasing competition during the patent period. The theory also suggests that broad patents increase the costs of imitation and thus decrease competition. I test the relationship between patent strength and competition during patent protection. I consider the Finnish markets for pharmaceuticals that provide rich variation in both patent length and breadth across innovations. I find that patent breadth, rather than length, prevents imitation. Patent rights have no effect on the risk of parallel trade.

Acknowledgments

I would like to express deepest gratitude to my supervisor, Otto Toivanen. I have greatly benefited from our numerous discussions and your excellent advices on all aspects of my research. Maybe most importantly, you have always encouraged me to push up my limits and have taught me the keys of success in academia - ambition and patience. It is a great privilege to have you as my supervisor. I am sure that you will have a great impact on my future work, too.

I am also deeply indebted to my thesis examiners, Gregory Crawford and Tuomas Takalo. They have provided me numerous important insights and suggestions on how to improve the chapters of my thesis. I could have not wished for better advices. Special thanks to Tuomas for being my opponent.

It has been an honor to work with several other talented and ambitious researchers. Your example have greatly contributed to my work, whether it comes to participating seminars and reading groups or insights about research. Specifically, I am very grateful to Matti Liski, Pauli Murto and Juuso Välimäki whose impressive intuition and knowledge of economic theory I admire. Outside of Hecer, I would like to thank Ville Aalto-Setälä guided me to work with pharmaceutical markets at the early stage of my studies. I would also like to thank my fellow graduate students and colleagues for stimulating discussions and encouragement. I am especially thankful to Mats Godenhielm, Saara Hämäläinen, Anna Sahari, Nelli Valmari, Tuukka Saarimaa and Janne Tukiainen for being great friends.

I am very grateful to FDPE and Yrjö Jahnsson Foundation for funding. I would also like to thank the Social Insurance Institution of Finland for providing me data.

This dissertation would not have been possible without my husband Sami who has provided consistent support and love. Finally, I thank the most important person in my life - my son Anton.

Contents

Chapter 1

Introduction

"Life is short, and the Art long; the occasion fleeting; experience fallacious, and judgment difficult."

Hippocrates

1.1 Introduction

This thesis contains three empirical industrial organization (IO) studies on markets for pharmaceuticals. Economists have long been concerned with the question on whether medical care markets are different from other markets. The question goes back to a seminal paper by Arrow (1963) who explained that the specific feature of the medical care markets is uncertainty. In this thesis, I show that uncertainty has important implications on the efficiency of health care provision.

Uncertainty is particularly present in every dimension of the clinical judgment by a physician, from doing a diagnosis, to deciding a laboratory test, interpreting the patient's symptoms and to choosing a medical intervention. A disease involves often the complex, abnormal conditions of physiological mechanisms that depend on various factors, such as the patient's genotype, choices and environment. Quantifiable data and the physician's personal experience help to understand the disease and to choose a medical intervention. The data can be easily stored and transmitted between physicians through patient records, whereas the personal knowledge may not be explicitly measurable. This knowledge can be gathered during the course of the patient's therapy, by treating other patients with similar diseases, discussing with colleagues, reading academic journals and receiving information through advertisements.

To illustrate the significance of uncertainty in medical care markets further, consider cholesterol drugs called statins that are prescribed to millions of people globally to decrease the risk for cardiovascular events, such as heart attack and stroke. An extensive medical literature has documented that the benefits and side-effects of statins vary between patients, for example, by their age and gender. The physician may not thus know in advance the health effects of statins for a patient. Moreover, in the early 2000s, the use of a statin called Cerivastatin was linked to fatal kidney failures and to 385 nonfatal cases, most of whom required hospitalization. Given that Cerivastatin was estimated to have 700,000 users in the United States at that time, many physicians were not be able to fully anticipate the extent of serious adverse effects (Furberg and Pitt, 2001).

In this thesis, I consider the behavioral consequences of uncertainty. In economics, an

experience good is a product or service where product characteristics are not known in advance but may be learned through different channels, such as consumption or the behavior of other agents. Many situations fit into this category, such a choice between restaurants, a firm's decision on subcontractors and clinical trials where pharmaceutical firms aim to minimize patient deaths.

Let's first consider a single agent's decision between different alternatives under uncertainty about their quality. A theoretical workhorse is the multi-armed bandit problem (e.g. Gittins, 1979, Bergemann and Välimäki, 2006). In the problem, a gambler is in the front of slot machines, or "one-armed bandits". He decides which machines to play, how many times to play each machine and in which order to play them. When a certain machine is played, a random reward realizes from a distribution specific to that machine. The gambler's objective is to maximize the expected sum of rewards taking into account that her information will improve over time. In other applications, the slot machines can be replaced by, for example, products, services or medical interventions.

In the bandit problem, the agent has a trade-off between exploitation and exploration. In exploitation, she makes a decision that provides the highest current utility given her information, e.g. a physician prescribes a drug which clinical effectiveness for a patient is fairly easy to predict in advance. In exploration, the agent gathers more information about its quality by experimenting, e.g. the physician prescribes a new drug treatment with highly uncertain clinical effectiveness. If the agent is risk averse, she can be unwilling to try new, less well-known products, which decreases the price elasticity of demand. If she nevertheless switches between products, she undergoes direct (current) welfare losses but may benefit from new information in the future. The second chapter of this thesis quantifies the roles of exploration and experimentation in demand for cholesterol drugs.

In the third chapter, I consider how the private experience of a physician and the past choices of other doctors affect the process of learning and prescription choices. I show that the long-term doctor-patient relationship can improve the process of learning about the health effects of drug treatment and thereby increase the efficiency of medical decision-making. On the other hand, treatment patterns that rely heavily on the past choices of other doctors may lead to over-prescribing and eventually lower health.

To explain these predictions, consider the following setup. Physicians make one-shot decisions for the patient between two options, the drug treatment and the outside good (non-medical treatments). The drug has either high or low quality. For example, the drug can either improve patient health or produce serious side effects. The quality of the outside good is normalized zero, i.e. it is between high and low quality. Before making her

decision, the physician investigates the patient and privately observes the quality signal, or health effects, of the drug and the patient's prescription history. The signal of the first physician indicates that the drug is of high quality, so she chooses that. The second one also receives a high signal on the drug and, by seeing the action of the first one, she finds out her signal. These two positive signals make the second physician even more optimistic that the quality of the drug is high and so she chooses it. The third physician observes a low quality signal on the drug but the past choices of others make her to choose the drug treatment, and so forth. If quality is in reality high, observing the past choices of other doctors improves the process of learning. If quality is low, the past choices of other doctors increase the physician's optimism on quality and lead to over-prescribing. If the physician-patient relationship was long-term, the physician would sooner or later find out quality.[1]

Empirical observations from pharmaceutical markets are consistent with uncertainty and learning. The literature has demonstrated that the markets are characterized by significant first-mover advantages (Caves et al., 1991, Grabowski and Vernon, 1992, and Hollis, 2002) and persistence in demand that is driven by the physician's own experience and the past choices of other doctors (e.g. Hellerstein, 1998, Nair et al., 2010, Coscelli and Shum, 2004). When physicians are risk averse, they may continue to prescribe the brand name drug, as they and other doctors have got used to do, instead of considering new treatment alternatives. Correspondingly, patients who have used brand name drugs for many years may not be willing to switch to generic products. The brand name firm may have an incentive to exploit locked-in physicians and patients by setting a high price for its product. Consistent with this, empirical evidence suggests that brand name firms are often able to maintain high prices, or even raise them, in response to generic entry (e.g. Frank and Salkever, 1992, 1997, Regan, 2008). Policy makers have widely tried to control for high drug expenses with price ceilings and guided the behavior of physicians and patients with public insurance policies, such as reference pricing.

Uncertainty is also present in the development of new innovations. Firms may not know well the commercial success of their new drugs but only few innovations reach even the marketing authorization stage. It has been estimated that less than 1 % of compounds survive from pre-clinical period to human testing and 22 % from clinical trials to the US Food and Drug Administration (FDA) approval. As generic products are developed with substantially lower costs and risks than new drugs (Grabowski, 2002), patents have been often seen as the lifeblood of pharmaceutical innovation.

[1]See Chamley (2004) for the review of social learning models where agents learn by observing the past choices of other agents.

Pharmaceutical patents do not necessarily protect against competition for at least two reasons. First, imperfect intellectual property rights allow firms to invent around a patented innovation. In particular, rivals may imitate an innovation with an analogy process patent by inventing a (non-trivial) manufacturing processes that are not covered by the original patent claims. Second, competitors may import a patented innovation from another country without the permission of the intellectual property owner (e.g. Kyle, 2011).

Policy makers in the United States have been strengthening intellectual property rights during the past few decades so that patents have become easier to enforce in court and may be longer (Gallini, 2002). The rationale behind the reforms is that strong patents provide higher rents for the incumbent on innovation and stimulate R&D. Both theoretical and empirical evidence suggests that the effect of strong patent systems on innovation can be ambiguous or even negative (Gallini, 2002, Hassan et al., 2009). An explanation is that patent policy affects competition and technology transfer between firms. A longer patent may increase imitation (Takalo, 1998, Gallini, 1992, Kanniainen and Stenbacka, 2000) and parallel trade and thus may not much promote innovation. Broader patents may discourage follow-on invention, such as the development of non-infringing duplicates (Gallini, 1992). The fourth chapter takes the first step to test whether stronger patents affect competition, i.e. imitation and parallel trade, during the patent period. I consider the Finnish markets for pharmaceuticals that provide rich variation in both patent length and breadth across innovations.

The focus of this thesis is on the role uncertainty in demand. The literature has developed structural models in order to infer uncertainty from observed choices and to analyze the welfare consequences of learning. This thesis makes no exception in that respect. In the second section, I first discuss the benefits and drawbacks of structural modeling in general. I then present the existing literature on traditional discrete choice demand models where the product's quality is known by agents (e.g. physicians). After that, I discuss the literature on demand for experience goods. At the end of the chapter, I introduce the remainder of this thesis.

1.2 Structural demand models

The analysis of demand has a long tradition in empirical IO. Researchers have used and developed demand models for several purposes. First, the parameters of a utility function may be interesting. For example, a researcher may want to estimate uncertainty associated with the health effects of drugs and the risk aversion coefficient of physicians

(Crawford and Shum, 2005, Chapters 2 and 3). Second, she may be want to study the welfare effects of different policies on demand, such as the provision of information on the quality of the match between a patient and a drug treatment (Crawford and Shum, 2005, Chapter 2), the length of the doctor-patient relationship (Chapter 3), mergers (Nevo, 2000), the introduction of new goods (Petrin, 2002), insurer policies (Dickstein, 2011) or price regulation (Ericson and Starc, 2012). Third, demand analysis is often needed in to estimate the price-cost markups of firms (Berry, Levinsohn and Pakes, 1995, hereafter BLP).

In this subsection, I present a sample of the background literature on structural demand models. I begin by evaluating the benefits and limitations of structural econometric modeling in general.[2] I then discuss the traditional demand models where product qualities are observed. This discussion helps to understand the main setup of my demand analysis in Chapters 2 and 3. The traditional demand models are not, however, suitable for evaluating the welfare consequences of uncertainty and learning that are present in markets for pharmaceuticals. The second and third chapter of this thesis contribute to the literature on demand for experience goods that I present at the end of this section.

1.2.1 Structural models

Structural econometrics uses econometric theory to produce statements about relationships between endogenous variables $\mathbf{y}$ and exogenous variables $(\mathbf{x}, \omega)$ that may be observable $(\mathbf{x})$ or unobservable (ω) to the agents of the theoretical model or the econometrician. The relationships can be functions $\mathbf{y} = f(\mathbf{x}, \omega, \Theta)$ or inequalities, e.g. $\mathbf{y} \geq f(\mathbf{x}, \omega, \Theta)$, where Θ is a vector of parameters. Because econometric theory does not often provide a reasonable description of the data, statistical assumptions about the distribution of $(\mathbf{y}, \mathbf{x}, \omega)$ complete theoretical assumptions. These theoretical and statistical assumptions are used to form predictions that are fitted to data. After estimating the parameters, a structural model can be used to evaluate responses to counterfactual, or not-yet-observed, policies.

Non-structural approaches in economics include studies on the description of data, such as the measurement of the prevalence of diseases in patient population or using non-parametric techniques to estimate a medical expenditure density. At the other side are statistical models that are used to predict outcomes without using any economic theory about underlying relationships. For example, a researcher may want to predict demand

[2]This section is based on Reiss and Wolak (2007).

for drugs with an autoregressive model based on the previous demand. Both of these approaches are widely used in economics and statistics. In the middle between the structural and the "reduced-form" models are the "quasi structural" models that are only loosely connected to economic theory. One example is a Heckman's (1979) sample selection model. The model has been used to analyze, for example, a patient's decision to visit a doctor and the doctor's decision to choose treatment (see e.g. Jones, 2000).

Economic theory helps to formulate relationships between variables, to understand how they are affected by changes in institutional conditions and to identify causal relationships. Consider the following example where a researcher wants to evaluate the effects of new drugs on competition between pharmaceutical firms. Suppose that the researcher observes the products' demand, prices and characteristics and variation in the number of firms that are active in the market. A descriptive model that uses very little economic theory, besides specifying endogenous and exogenous variables, could predict how the market shares of firms change with the number of competitors. Without any other information on the nature of competition and demand, it could be hard to justify assumptions that would guarantee a causal relationship between the number of firms and the market shares.

With theoretical and statistical assumptions, the researcher could specify supply and demand for pharmaceuticals. She could then estimate the price elasticity of demand and the price-cost margins of pharmaceutical firms that are not directly observable in the data. Finally, the researcher could evaluate how the welfare of patients and the profits of firms change if some of the firms would exit the market. The counterfactual experiment can be performed without observing any changes in the number of firms.

The process of structural modeling involves many choices. These choices may or may not be credible from the viewpoint of statistical inference and knowledge on institutions and economic theory. For example, a researcher may want to impose assumptions on how marginal costs depend on product characteristics, whether firms decide on prices (Bertrand competition) or quantities (Cournot competition) and whether physicians are risk-averse, myopic or forward-looking while making their treatment choices for patients. Some of the assumptions may be tested with data. For example, the researcher may want to test if physicians are risk-averse or wants to evaluate whether Cournot or Bertrand competition provide a better fit to the data. Still, the functional form of a utility function is frequently taken as given even though it affects statistical inference and interpretation. A careful researcher should, when possible, experiment with different assumptions.

A convincing structural model should also respect both economic theory and institutional environment that generates data. To make this more concrete, suppose that price ceiling

regulation truncates the drug price distribution in the supply-demand -analysis. Then the consistent estimation of the effect of competitors on prices requires that this truncation is taken into account. When price ceilings are binding, demand exceeds supply. For this reason, an assumption on the equality of demand and supply does not hold (Reiss and Wolak, 2007).

1.2.2 Traditional demand models

This section considers the traditional demand models where product characteristics are observed by agents (e.g. physician or consumers) but not necessarily by the econometrician. To understand the benefits of discrete choice models that I apply in the analysis of demand for pharmaceuticals, I first discuss the neoclassical extensions of homogeneous goods demand models.

Early work in IO considered the estimation of demand in an industry where products are perfect substitutes (see e.g. Porter, 1983). The basic idea was to specify a system of demand equations that depend on prices, exogenous market variables and demand shocks. In the 1980s and the 1990s, researches became interested in demand for differentiated products (see e.g. Bresnahan, 1981 and 1987). Instead of estimating a one industry level demand equation, separate demand equations were often estimated for each product.

This approach created problems with the number of parameters that became often very large without any restrictions. For example, if there were 100 drugs, the number of estimated demand equations would be 100 and each of them would contain 100 price elasticity coefficients. The number of estimated parameters would be at least $10,000$ which requires large datasets and can cause computational challenges. One solution to avoid "the curse of dimensionality" is to make parameter restrictions, such as to assume that all cross-price elasticities are equal.[3] Parameter restrictions are, however, often ad hoc and affect price-cost markups in an undesirable manner.

A another solution is to relate the utility of an agent to a set of parameters and the attributes of the chosen product and the agent. For example, the utility of a patient from a drug treatment may depend on the patient's income, age and gender as well as on price, strength, drug form (e.g. a tablet) and route of administration. Market demand can then be aggregated from agent level choices. This approach avoids the curse of dimensionality because the computation of price elasticities is based on a much smaller set of utility parameters.

[3]See Reiss and Wolak (2007) for more extensive discussion about solutions to the curse of dimensionality.

The discrete choice literature goes back to 1970s and 1980s to the work of McFadden (1973, 1981, 1982, 1984) who developed conditional multinomial logit models. These models have received increased attention in empirical IO since the influential papers by Berry (1994) and Berry, Levinsohn and Pakes (1995) (e.g. Ackerberg and Rysman, 2005, Davis, 2000, Hendel, 1999, Nevo, 2000, Petrin, 2002).

To illustrate discrete choice demand models further, consider a physician who chooses a drug treatment for her patient i among J alternatives. For simplicity, assume that the physician is perfect agent for the patient, i.e. she maximizes the patient's utility.[4] Suppose that the number of potential patients that may need drug treatment at time, or prescription, t is N_t. The indirect utility of the patient for drug j, $j \in \{0, ..., J\}$, at time (or market) t is

$$U_{ijt} = U(\mathbf{Z}_{jt}, \omega_{ijt}, \Theta), \tag{1.1}$$

where $\mathbf{Z}_{jt}$ is a vector of covariates for the product at time t (including the price), ω_{ijt} is a variable that varies by patients, products and time periods and Θ is a vector of parameters. The alternative $j = 0$ denotes the outside good (e.g. non-medical treatments).

The physician chooses drug j at time t that maximizes the patient's utility,

$$j = \text{argmax}_{k \in \{0,...,J\}} U(\mathbf{Z}_{kt}, \omega_{ikt}, \Theta). \tag{1.2}$$

Following the early work, assume that heterogeneity across patients arises only in preference shocks and marginal valuations for characteristics are constant,

$$U_{ijt} = \mathbf{Z}_{jt}\beta + \omega_{ijt}, \tag{1.3}$$

where ω_{ijt} is the Type 1 extreme value distributed error term. This specification implies very restricted substitution patterns by assuming the independence of irrelevant alternatives (IIA). The assumption implies that the relative odds of choosing one alternative over another do not depend on the presence or absence of other "irrelevant" alternatives. For example, the relative probability of choosing between two cholesterol drugs does not change if a new cholesterol drug is added as an additional possibility. The assumption also implies that price elasticities are completely determined by prices and market shares, not by how different the products are (e.g. Berry, 1994).

[4] This assumptions allows me to drop the physician-index from the utility function. A demand model can easily be extended to allow for variation across both physicians and patients (see Chapter 3).

Random coefficients for the characteristics are often used to get more realistic substitution patterns. A vector of random coefficients can, for example, depend on the patient's demographics, $\mathbf{D}_i$, $\beta_i = \beta_1 + \mathbf{D}_i\beta_2 + \Sigma v_i$, where Σ is standard deviation and $v_i \sim P(v)$. For example, this specification allows for patients with diabetes to benefit more from cholesterol drugs and to respond less to changes in prices than healthy patients do. The indirect utility function is

$$\begin{aligned} U_{ijt} &= \mathbf{Z}_{jt}\beta_i + \theta_{jt} + \epsilon_{ijt} && (1.4) \\ &= \underbrace{\mathbf{Z}_{jt}\beta_1 + \theta_{jt}}_{\delta_{jt}} + \underbrace{\mathbf{Z}_{jt}(\mathbf{D}_i\beta_2 + \Sigma v_i) + \epsilon_{ijt}}_{\omega_{ijt}}, && (1.5) \end{aligned}$$

where θ_{jt} is the alternative and time specific random coefficient, δ_{jt} is the mean utility and ϵ_{ijt} is the preference shock. In the above expression, ω_{ijt} contains heterogeneity in marginal utilities and the preference shock across patients. An alternative is a discrete version of random coefficients where $\beta_i \in \{\beta_1, ..., \beta_K\}$ and $p_k = P(\beta_i = b_k|\mathbf{D}_i)$. For the estimation of these type of models with the simulated maximum likelihood or Gibbs sampling, see Train (2009).

BLP extended the random coefficient multinomial logit models by allowing for unobserved (to the econometrician) product characteristics θ_{jt} to be correlated with the observed characteristics, such as the price. They also assumed that neither the firms nor the econometrician observes ω_{ijt} in (5) but knows the distributions of the random coefficients.

The expected demand for drug j at time t is a sum of purchasing probabilities,

$$\begin{aligned} q^e_{jt}(\delta_t, \Theta) &= \sum_{n=1}^{N_t} E_{\omega_{ijt}}(U(\delta_{jt}, \omega_{ijt}, \Theta) \geq \max_{k\neq j}\{U(\delta_{kt}, \omega_{ikt}, \Theta)\}) && (1.6) \\ &= \sum_{i=1}^{N_t} P(\text{The physician chooses drug } j \text{ at time } t \text{ for patient } i), && (1.7) \end{aligned}$$

where N_t is the number of potential patients.

Let δ_t be a vector of mean utilities, $s_{jt}(\delta_t, \Theta) = q^e_{jt}(\delta_t, \Theta)/N_t$ be predicted market shares and S_{jt} be observed market shares. Because predicted market shares $s_{jt}(\delta_t, \Theta) = S_{jt}$ for drugs depend on the mean utility vector δ_t, the mean utility vector can be recovered by inverting the market shares, $\delta_t = s^{-1}(S_t, \Theta)$, where S_t is a vector of observed market shares. In the logit model (3), the inversion can be done analytically by $\delta_{jt} = \ln s_{jt} -$

$\ln s_{0t}$, where s_{0t} is the market share of the outside good (Berry, 1994). With the random coefficients, numerical methods can be used to compute the predicted market shares and the inversion.

Several papers have generalized discrete choice demand models further. For example, Ackerberg and Rysman (2005), Bajari and Benkard, (2001, 2005) have modified the assumptions of BLP on the functional form of the utility function or the distribution of agent heterogeneity. In Gentzkow (2007), agents may choose multiple products, that can be either substitutes or complements, at the same time. The traditional demand models also assume that agents know product quality θ_{jt}. The assumption is not often realistic in markets for pharmaceuticals. In these markets, the effectiveness and side effects of drugs can differ across patients which creates uncertainty to the physician's medical decision-making. Next, I consider demand models that take into account uncertainty in quality.

1.2.3 Demand models with unobserved quality

This section presents the literature on discrete choice demand models for experience goods. In the models, an agent does not know in advance the quality of available options that may be learning by consumption or by observing the past choices of other agents. In this thesis, I focus on demand models with Bayesian learners.[5]

To describe the setup of the learning models, consider again a physician's (she) prescription choice when she is a perfect agent for her patient (he). Assume that the physician does not know in advance how sensitive the patient is for the effectiveness and side effects of available drugs. Denote the quality of the match between product j and patient i by θ_{ij}.[6] The quality has the prior distribution $F(\theta_{ij}, \mathbf{Z}^q_{ij1})$ that may depend on a vector of observables at the beginning of the drug therapy, $\mathbf{Z}^q_{ij1}$.

While investigating the patient, the physician observes certain health effects associated with the use of the prescribed drug. Let $x_{ijt} = \theta_{ij} + \sigma v_{ijt}$, where σ is standard deviation and $v_{ijt} \sim F(v)$, measure health effects, or signals, that are realized after the patient has taken the drug in period t. Denote a set of the state variables by S_{it}. The set contains preference shocks ϵ_{ijt}, health effects x_{ijt} and characteristics $\mathbf{Z}_{ijt}$ that are observed by the beginning of period t.

[5]For example, in macroeconomics there is the extensive literature on non-bayesian learning, see e.g. Evans and Honkapohja (2011).

[6]In this thesis, I analyze the physician's learning about the patient-specific quality of drugs. For learning about the average health effects of drugs across patients, θ_j, see e.g. Coscelli and Shum (2004).

The physician maximizes the patient's discounted expected utility conditional on the state variables. The value function for drug j at time t is

$$U_j(S_{it}) = E(u(x_{ijt})|S_{it}) + \mathbf{Z}_{ijt}\beta + \epsilon_{ijt} + \rho E[\max_{k\in\{0,\ldots,J\}} U_k(S_{i(t+1)}|S_{it}, j)], \tag{1.8}$$

where ρ is the discount factor and $u(x_{ijt})$ is a sub-utility function for health effects. Conditional on the state variables, the first element of the value function is the expected per-period utility for drug j and the second one is the continuation value given that drug j was chosen.

The physician updates her beliefs about quality based on observed health effects, creating dynamics to pharmaceutical demand. When health effects are normally distributed, the updating process is simple because the posterior distribution of θ_{ij} at the beginning of period t, $f(\theta_{ij}|\{x_{ijt'}, t' < t\})$, is also normal (see DeGroot, 1970). Over time, more health realize and the physician's belief about quality become more precise (see e.g. Chapters 2 and 3 and Crawford and Shum, 2005).

To simplify the problem further, several papers (e.g., Coscelli and Shum, 2004, Ching, 2009, Chernew et al., 2009) have assumed agents are myopic, i.e. the discount factor equals zero. This assumptions abstracts away incentives to experiment with new, less well-known products. Despite of the computational challenges, many papers have estimated a dynamic demand model, starting from Erdem and Keane (1996) (e.g. Ackerberg, 2003, Chan et al., 2006, Crawford and Shum, 2005, Ching, 2009, Dickstein, 2011, and Kim, 2010). The second chapter fits into this literature and estimates the importance of learning through experimentation in the Finnish market for cholesterol drugs.

The structural learning literature has typically assumed that agents learn from their own experience or observe the signals of their predecessors perfectly. In many situations, agents may infer quality by using their own experiences and the past choices of peers. For example, physicians want to prescribe products that other doctors have previously prescribed (see the Section 3 of this thesis), restaurants full of customers are thought to sell high quality food, people want to be friends with those who are popular and an unemployment period is often believed to reveal information to the employer about the quality of the job applicant.

The literature on structural social learning models is still very limited, with a few exceptions (Cipriani and Guarino, 2012, Zhang, 2010, Knight and Schiff, 2010).[7] Cipriani and

[7] To be more specific, Cipriani and Guarino (2012) investigate the herding behavior of investors, Zhang (2010) analyzes kidney replacement decisions and Knight and Schiff (2010) consider elections.

Guarino and Knight and Schiff assume that agents are myopic, whereas Zhang allows for agents to be forward looking. The literature assumes that quality is constant across agents who make once-in-a-lifetime decisions.[8] Unlike the previous literature, I take into into account both private and social learning, by allowing for agents make repeated decisions, and allow for heterogeneity in quality (Section 3).

1.3 A summary of chapters

This section provides a summary of chapters. The second chapter presents a dynamic model of demand for pharmaceuticals where a physician does not know ex-ante the average health effects of drug treatments for a patient. With the model, I investigate the value of experimentation in the Finnish market for cholesterol drugs. In the third chapter, I provide a structural model of demand for pharmaceuticals with learning from the physician's personal experience and the past choices of other doctors. With the counterfactual experiments, I analyze how the length of the doctor-patient relationship affects learning and the efficiency of medical decision-making. In the final section, I ask whether stronger patents prevent competition during patent protection.

1.3.1 Chapter 2: Experimentation and Learning in Pharmaceutical Demand: Evidence from the Cholesterol Drug Market

Uncertainty and learning can have important behavioral implications on drug treatment choices by a physician. The physician can have incentives to experiment with new products to get more information about the effectiveness of drugs. At the same time, uncertainty and risk-aversion can make the physician reluctant to try new drug treatments for the patient. The second chapter quantifies the roles of learning and experimentation in pharmaceutical demand. I estimate a structural model of medical decision-making under uncertainty about the health effects of drug treatments for a particular patient. After taking the drug, the physician observes two health effects: the first effect affects the patient's symptoms and the second one affects the probability that the drug therapy continues. If the patient's therapy continues, this new information helps the physician to form more precise predictions about the quality of the match between the patient and the drug. I

[8] In this case, a set of state variables, S_{it}, includes a (one) experience signal x_{it} of agent i and the past choices of other agents.

focus on the Finnish market for cholesterol drugs. In this market, exit rates are high and physicians are very unwilling to change the drug treatment of their patients.

The parameter estimates imply that much of the uncertainty regarding to health effects is resolved after the first prescription. As physicians are estimated to be risk-averse, they are unwilling to prescribe new treatment alternatives for their patients. My findings suggest that if doctors became more willing to take risks in their treatment choices, the process of learning would improve. I also show that if a physician does not learn, the patient's welfare decreases and the quit rate from the drug therapy increases. These findings indicate that information provision about the average health effects of drug treatments should be promoted.

1.3.2 Chapter 3: Private Experience and Observational Learning in Pharmaceutical Demand

Uncertainty about the quality of a product is present in many markets. Agents may apply their own experiences and the past choices of their peers to re-evaluate how well the product matches with their preferences. Somewhat surprisingly, previous work about the roles of private and social learning in demand has remained very limited. In the third chapter of this thesis, I consider these issues in medical decision-making under uncertainty about the effects of drug treatment on patient health. I apply the standard theoretical models with social learning (Chamley, 2004), by allowing for physicians to take repeated choices. I test whether the long-term doctor-patient relationship is more efficient than providing information on the past choices of other doctors through patient records. Policies that guarantee continuity of care are commonly used in primary care to promote the process of learning and improve medical decision-making (e.g. Scott, 2000).

My data from the Finnish market for cholesterol drugs confirms that prescriptions are highly responsive to the length of the doctor-patient relationship. I explain this finding by showing that the number of interactions between the physician and the patient have important implications on pharmaceutical demand. Specifically, I find that the long-term treatment relationship promotes the process of learning and improves physician decision-making. I demonstrate that treatment patterns relying on the past choices of other doctors hinder learning and may lead to over-prescribing for a fraction of patients. This is so, since an inexperienced physician becomes more optimistic about quality, or the average health effects, if other doctors have prescribed the drug for the patient previously. Overall, my findings suggest that providing information on the past choices of other doctors does not compensate for the lack of the long-term relationship.

1.3.3 Chapter 4: Do Stronger Patents Protect Against Competition? Evidence from the Pharmaceutical Industry

The main goals of the patent system is to stimulate innovation and to encourage firms to disclose their innovations. The system is often claimed to be inefficient in achieving these goals, in part because patents do not provide exclusive rights for the innovator. Competitors are frequently inventing around patented innovations (Boldrin and Levine, 2005). In certain industries, such as software, music and pharmaceuticals, the resale of goods between countries (so called parallel trade) may also arise without the authorization of the owner of the intellectual property (Kyle, 2011). Over the past decades, policy makers in the United States have been strengthening patent protection such that patents have become longer for some innovations and easier to enforce in court (Gallini, 2002). The rationale behind these reforms is that stronger patents increase the profits of an innovator and promote innovation. Economic theory predicts that longer patents may hinder rather than stimulate innovation by increasing competition during the patent period. Broad patents increase the costs of imitation and thus decrease competition. In the fourth chapter, I test the theory on the relationship between patent strength and competition during patent protection.[9]

I consider the Finnish markets for pharmaceuticals that provide rich variation in patent length and breadth across innovations. With this variation, I analyze how the patent rights of an incumbent innovation affect the risks of imitation and parallel trade. My results suggest that patent breadth - measured by the number of claims - discourages imitation. I find no evidence that patent length would increase the risk of imitation during patent protection. The effects of both patent length and breadth on the rate of parallel trade are also insignificant. These findings suggest that policy makers should promote broader, rather than longer, patents if they aim to decrease imitation incentives and to guarantee higher rents for the incumbent on its R&D efforts. Still, further work is required to understand the role of patents in a cumulative innovation process.

[9]See Grönqvist, 2009, for the effect of patent length on the private value of patents in Finland.

Bibliography

Ackerberg, D. (2003) Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination. *International Economic Review,* Vol. 44, pp. 1007-1040.

Ackerberg, D. – Rysman, M. (2005) Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects. *RAND Journal of Economics*, Vol. 36, pp. 771-788.

Arrow, K. (1963) Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, Vol. 53, pp. 941-973.

Bajari, P. – Benkard, L. (2001) Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches. *Working manuscript*, Stanford Graduate School of Business.

Bajari, P. – Benkard, L. (2005) Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, Vol. 113, pp. 1239-1276.

Bergemann, D. – Välimäki, J. (2006) Bandit Problems. *The New Palgrave Dictionary of Economics*, Second Edition.

Berry, S.T. (1994) Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics*, Vol. 25, pp. 242-262.

Berry, S.T. – Levinsohn, J. – Pakes, A. (1995) Automobile Prices in Market Equilibrium. *Econometrica*, Vol. 63, pp. 841-890.

Boldrin, M. – Levine, D. (2005) Against Intellectual Monopoly.
http://levine.sscnet.ucla.edu/general/intellectual/against.htm

Bresnahan, T.F. (1981) Departures from Marginal-Cost Pricing in the American Automobile Industry: Estimates for 19771978. *Journal of Econometrics*, Vol. 11, pp. 201-227.

Bresnahan, T.F. (1987) Competition and Collusion in the American Automobile Market: The 1955 Price War. *Journal of Industrial Economics*, Vol. 35 pp. 457-482.

Caves R.E. – Whinston M.D. – Hurwitz M. – Pakes, A. – Temin, P. (1991) Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry. *Brookings Papers on Economic Activity. Microeconomics*, pp. 1-66.

Chamley, C. (2004) Rational Herds – Economic Models of Social Learning. *Cambridge University Press*, USA.

Chan, T. – Hamilton, B. (2006) Learning, Private Information, and the Economic Evaluation of Randomized Experiments. *Journal of Political Economy*, Vol. 114, pp. 997-1040.

Chernew, M. – Gowrisankaran, G. – Scanlon, D. (2009) Learning and the Value of Information: Evidence from Health Plan Report Cards. *Journal of Econometrics*, Vol. 144, pp. 156-174.

Ching, A. (2010a) A Dynamic Oligopoly Structural Model for the Prescription Drug Market After Patent Expiration. *International Economic Review*, Vol. 51, pp. 1175-1207.

Ching, A. (2011b) Consumer Learning and Heterogeneity: Dynamics of Demand for Prescription Drugs After Patent Expiration. *International Journal of Industrial Organization*, Vol. 28, pp. 619-638.

Cipriani, M. – Guarino, A. (2012) Estimating a Structural Model of Herd Behavior in Financial Markets. Forthcoming in *American Economic Review*.

Coscelli, A. – Shum, M. (2004) An Empirical Model of Learning and Patient Spillovers in New Drug Entry. *Journal of Econometrics*, Vol. 122, pp. 213-246.

Crawford, G. S. – Shum, M. (2005) Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, Vol. 73, pp. 1137-1173.

Davis, P. (2000) Demand Models for Market–Level Data. Working manuscript. MIT Sloan School.

DeGroot, M. (1970) Optimal Statistical Decisions. *McGraw–Hill*, New York.

Dickstein, M. (2011) Efficient Provision of Experience Goods: Evidence from Antidepressant Choice. *Job Market Paper*, Harvard University.

Erdem, T. – Keane, M. P. (1996) Decision–Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science,* Vol. 15, pp. 1-20.

Ericson, K. – Starc, A. (2012) Pricing Regulation and Imperfect Competition on the Massachusetts Health Insurance Exchange. Working paper.

Evans, G. – Honkapohja, S. (2011) Learning as a Rational Foundation for Macroeconomics and Finance. *Bank of Finland Research Discussion Papers*, Vol. 8, pp. 1–59.

Frank R. G. – Salkever, D. S. (1992) Pricing, Patent Loss and the Market for Pharmaceuticals. *South Econ J*, Vol. 59, pp. 165-79.

Frank R. G. – Salkever, D. S. (1997) Generic Entry and the Pricing of Pharmaceuticals. *J Econ Manage Strategy*, Vol. 6, pp. 75-90.

Furberg, C. – Pitt, B. (2001) Withdrawal of Cerivastatin from the World Market. *Curr Control Trials Cardiovasc Med.*, Vol. 2, pp. 205-207.

Gallini, N. (1992) Patent Policy and Costly Imitation. *RAND Journal of Economics*, Vol. 23, pp. 52-63.

Gallini, N. (2002) The Economics of Patents: Lessons from Recent U.S. Patent Reform. *Journal of Economic Perspectives*, Vol. 16, pp. 131-154.

Gentzkow, M. (2007) Valuing New Goods in a Model with Complementarity: Online Newspapers. *American Economic Review*, Vol. 97, pp. 713-744.

Gittins, J. C. (1979) Bandit Processes and Dynamic Allocation Indices. *Journal of the Royal Statistical Society*, Series B (Methodological), Vol. 41, pp. 148-177.

Goldberg, P.K. (1995) Product Differentiation and Oligopoly in International Markets: The case of the US Automobile Industry. *Econometrica*, Vol. 63, pp. 891-951.

Grabowski, H. G. – Vernon, J. M. (1992) Brand Loyalty, Entry, and Price Competition in Pharmaceuticals After the 1984 Drug Act. *Journal of Law and Economics*, Vol. 35, pp. 331-350.

Grabowski, V. (2002) Patents and New Product Development in the Pharmaceutical and Biotechnology Industries. *Working Paper*, pp. 1-29.

Grönqvist, C. (2009) Empirical Studies on the Private Value of Finnish Patents. *Bank of Finland, Scientific Monographs*, E:41, pp. 1-167.

Hassan, E. – Yaqub, O. – Diepeveen, S. (2009) Intellectual Property and Developing Countries, a Review of the Literature. *RAND Europe*, prepared for the UK Intellectual Property Office and the UK Department for International Development, pp. 1-68.

Heckman, J. (1979) Sample Selection Bias as Specification Error? *Econometrica*, Vol. 47, pp. 153-161.

Hellerstein, J. K. (1998) The Importance of the Physician in the Generic Versus Trade-Name Prescription Decision. *The RAND Journal of Economics*, Vol. 29, pp. 108-136.

Hendel, I. (1999) Estimating Multiple–Discrete Choice Models: An Application to Computerization Returns. *The Review of Economic Studies*, Vol. 66, pp. 423-446.

Hollis A (2002). The Importance of Being First: Evidence from Canadian Generic Pharmaceuticals. *Health Economics*, Vol. 11, pp. 723-734.

Jones, A. (2000) Health Econometrics. Chapter 6 in *Handbook of Health Economics*, Vol. 1, Part A, pp. 265-344.

Kanniainen, V. – Stenbacka, R. (2000) Endogenous Imitation and Implications for Technology Policy. *Journal of Institutional and Theoretical Economics*, Vol. 156, pp. 360-381.

Kim, M. (2010) Physician Learning and New Drug Diffusion. *Job Market Paper*, University of Chicago.

Knight, B., Schiff, N. (2010) Momentum and Social Learning in Presidential Primaries. *Journal of Political Economy*, Vol. 118, pp. 1110-1150.

Kyle, M. (2011) Strategic Responses to Parallel Trade. *The B.E. Journal of Economic Analysis & Policy*, Vol. 11, pp. 1-32.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics Academic Press*, New York, pp. 105-142.

McFadden, D. (1981) Econometric Models of Probabilistic Choice.*Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA.

McFadden, D. (1982) Qualitative Response Models. *Advances in Econometrics: Invited Papers for the Fourth World Congress of the Econometric Society*, 1-37, Cambridge University Press, Cambridge, UK.

McFadden, D. (1984) Econometric Analysis of Qualitative Response Models. *Handbook of Econometrics,* Vol. 2, 1395-1457, Amsterdam, North Holland

Milgrom, P. – Roberts, J. (1986) Price and Advertising Signals of Product Quality. *Journal of Political Economy,* Vol. 94, 796-821.

Nair, H. S. – Manchanda, P. – Bhatia, T. (2010) Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders. *Journal of Marketing Research*, Vol. 47, pp. 883-895.

Nevo, A. (2000) Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry. *RAND Journal of Economics*, Vol. 31, pp. 395-421.

Petrin, A. (2002) Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy*, Vol. 110, pp. 705-729.

Porter, R.H. (1983) A Study of Cartel Stability: The Joint Executive Committee, 18801886. *Bell Journal of Economics*, Vol. 14, pp. 301-314.

Regan, T. (2008) Generic Entry, Price Competition, and Market Segmentation in the Prescription Drug Market. *International Journal of Industrial Organization*, Vol. 26, pp. 930-948.

Reiss, P. – Wolak, F. (2007) Structural Econometric Modeling: Rationales and Examples from Industrial Organization. *Handbook of Econometrics*, Volume 6A, Chapter 64, pp. 4277-4415.

Scott, A. (2000) Economics of General Practice. Chapter 22 in *Handbook of Health Economics*, Volume 1, Edited by A.J. Culyer and J.P. Newhouse, 2000, Elsevier Science B.V.

Takalo, T. (1998) Innovation and Imitation Under Imperfect Patent Protection. *Journal of Economics*, Vol. 67, pp. 229-241.

Train, K. (2009) Discrete Choice Methods with Simulation. *Cambridge University Press*, Second edition, New York, USA.

Yiannaka, A. – Fulton, M. (2006) Strategic Patent Breadth and Entry Deterrence with Drastic Product Innovations. *International Journal of Industrial Organization*, Vol. 24, pp. 177-202.

Zhang, J. (2010) The Sound of Silence: Observational Learning in the U.S. Kidney Market. *Marketing Science*, Vol. 29, pp. 315-335.

Chapter 2

Experimentation and Learning in Pharmaceutical Demand

Buyers do not necessarily observe the quality of products in advance but may learn them through consumption. Such uncertainty creates incentives to experiment with new products to gain more information. At the same time, uncertainty makes risk-averse buyers reluctant to try new, less well-known products. Still, traditional demand models ignore uncertainty and learning. In this chapter, I estimate a dynamic model of demand for cholesterol drugs that allows for learning through experimentation. The results suggest the average health effects of cholesterol drugs are heterogeneous across patients which creates uncertainty to medical decision-making. My analysis also identifies drugs that induce higher exit rates from the cholesterol drug therapy. I find that uncertainty and risk aversion make physicians unwilling to try new drug treatments for their patients. These results suggest that if doctors became more willing to experiment with their treatment choices, the process of learning would improve and the efficiency of medical decision-making would increase.

Keywords: learning, structural modeling, unobserved quality, demand, physician behavior

2.1 Introduction

In markets for experience goods, agents may learn about the unknown quality of products through experimentation. In each period, an agent chooses a product that best matches with her preferences by knowing that her purchase will reveal new information about its quality. Through repeated choices, information accumulates and uncertainty diminishes. In many of these markets, the degree of risk version may have significant implications on behavior. One particularly interesting example is medical decision-making by physicians about risky treatment alternatives for their patients. When the health effects of drugs are uncertain, risk-aversion can significantly increase the costs of uncertainty and slow down the process of finding the best treatment alternative for a patient.

In this chapter, I study the roles of experimentation and risk aversion in the drug choices of physicians by using data from the Finnish market for cholesterol drugs called statins. The market is particularly interesting for several reasons. First, an extensive medical literature has shown that the health benefits and side effects of statins vary between patients.[1] This heterogeneity creates uncertainty about the effectiveness of the statin therapy for a patient. Second, my data show that physicians are very unwilling to change the drug treatments of their patients. I explain this finding with risk aversion and learning about the average health effects of cholesterol drugs. Third, potential improvements in the patients' health are substantial, as statins is one of the world's largest selling drug groups. Still, a third of patients in my data exit the drug therapy after the first prescription.

I develop my analysis as follows. A patient (she) comes to a physician (he) to seek a drug treatment for her medical condition. The physician diagnoses first the (fixed) medical condition of the patient. Conditional on the diagnosis, the physician chooses a cholesterol drug under uncertainty about the average health effects, or match values, of available drugs for this particular patient. After the patient has taken the drug, the physician observes two health effects. The first one captures the effect of the drug on the patient's symptoms (side effects). The second one measures how the drug treatment affects the exit rate. The therapy may end, for example, because the short-term therapy decreased the patient's total cholesterol under the desired level (5mmol/L) or lifestyle changes were more effective than the prescribed drug in reducing the patient's risk for cardiovascular events.[2] If the therapy continues, the physician makes a prescription. The decision takes into account the initial diagnosis and the physician's beliefs about match values that are based on the observed health effects. These steps are repeated until the patient's drug

[1] See for example National Institute for Health and Clinical Excellence (2006).

[2] Crawford and Shum (2005) interpreted the exit probability as the probability of recovery.

therapy ends.

Uncertainty and learning about the effectiveness of medical treatments have two important implications on prescription behavior. Risk aversion makes the physician reluctant to try a new treatment compared with a drug that has more certain, but possibly lower, effectiveness.[3] On the other hand, a forward-looking physician may have an incentive to experiment with a new treatment to get information about its health effects. While the former causes persistence, or "switching costs", in drug choices, the latter creates diffusion in demand.

This chapter relates to the growing literature on demand for experience goods. A significant portion of the existing literature has assumed that agents are myopic, i.e. their discount factor is zero (Coscelli and Shum 2004, Chernew et al., 2008, Ching 2008, Narayanan and Manchanda 2009). The assumption abstracts away incentives to experiment. A few papers (e.g. Crawford and Shum, 2005, hereafter CS, Ackerberg, 2003, Ching, 2009, Dickstein, 2011, Kim, 2010) study the demand of forward-looking agents under uncertainty. Finally, this paper is related to the literature that identifies risk preferences from the observed choices of myopic agents (e.g. Cohen and Einav, 2007, Chetty, 2006).

The first objective of this paper is to replicate the dynamic model of CS in the Finnish market for cholesterol drugs. A difference between cholesterol drugs and anti-ulcer drugs analyzed in CS is that cholesterol drugs are used as preventive treatments for cardiovascular diseases whereas anti-ulcer drugs treat ulcers in the stomach and the upper part of the small intestine. Second, after confirming that the results are qualitatively similar with CS, I perform counterfactuals that have been ignored in much of the existing literature on demand for experience goods. First, I evaluate the implications of risk aversion on demand and efficiency. Risk aversion may create habit persistence, slow down the learning process and decrease incentives to experiment.[4] Second, in order to understand the role of experimentation in demand, I evaluate how treatment outcomes and costs change when physicians become myopic.[5] In this case, physicians do not take into account the consequences of their treatment choices on the patients' future health.

The parameter estimates indicate that the average health effects of cholesterol drugs

[3] In practice, the risk aversion parameter captures factors that cause strong persistence in the prescription choices of a physician for a patient. Besides risk aversion, these factors may include, for example, time constraints and marketing efforts directed at the physician.

[4] See Cohen and Einav (2007) for the implications of risk aversion for individual behavior and pricing in insurance contracts.

[5] On the contrary, CS evaluated this issue with a policy experiment where the physician is restricted to prescribe the first drug to his patient in every period until she is healed.

are heterogeneous across patients, creating uncertainty to medical decision-making by a physician. The results suggest that market leaders Simvastatin and Atorvastatin perform reasonably well in the symptomatic dimension. My analysis, however, indicates that the use of Simvastatin induces higher exit rates than the use of other cholesterol drugs. Physicians learn fast as much of the uncertainty dissipates after the first prescription. At this stage of the therapy, the physician observes how effectively the first prescription decreased the patient's cholesterol levels and whether it caused any side effects.

The counterfactuals show that the provision of information about the average health effects of drugs can significantly increase the efficiency of medical decision-making. Consistent with high persistence in demand for cholesterol drugs, I find that risk aversion makes physicians unwilling to experiment new, less well-known treatment alternatives for their patients. The results also suggest that if physicians became more willing to take risks in their treatment choices, the process of learning through experimentation would improve and the welfare of a patient would increase.

The rest of this chapter is organized as follows. The second section presents the dataset and descriptive results. Section 2.3 goes through the theoretical model. Section 2.4 presents the likelihood function and discusses identification. Section 2.5 presents the estimation results. Section 2.6 evaluates the fit of the model and shows findings from the counterfactual experiments. The final section of this chapter concludes.

2.2 Market and data description

2.2.1 The Finnish cholesterol drug market

I study the prescriptions of physicians in the Finnish market for cholesterol drugs. Cholesterol drugs are used to lower cholesterol levels in the blood by reducing the production of cholesterol by the liver. Abnormal cholesterol levels, where the concentration of LDL-cholesterol ("bad" cholesterol) is high and the concentration of HDL-cholesterol ("good" cholesterol) is low, are one of the risk factors of cardiovascular diseases (CVD), such as coronary heart disease, heart attack and stroke. High morbidity to cardiovascular diseases has made cholesterol drugs as one of the world's largest selling drug groups.

I focus on the choices of physicians between different active ingredients that are referred to "drugs". Corresponding to the United States, 6 active ingredients are on the Finnish statin market: Simvastatin (brand-name Zocor), Lovastatin (Mevacor), Pravastatin (Pravachol

or Selektine), Fluvastatin (Lescol, Canef or Vastin), Atorvastatin (Lipitor) and Rosuvastatin (Crestor).[6] Active ingredients differ to some extent in their effectiveness, side effects and prices. Patients respond differently to statins which creates uncertainty about to the effectiveness of statins for a patient (the Finnish current care for dyslipidemia, 2011, Jousilahti, 2004).[7] Some individuals may also have more side effects with a one statin than another.[8]

A physician's statin treatment decision is based on the evaluation of the patient's risk for CVDs. This evaluation is based on several factors, including the patient's gender, age, blood pressure and cholesterol levels. In the model, the initial evaluation affects the prior beliefs of a physician about the quality of the match between the patient and cholesterol drugs.[9] The choice of a cholesterol drug is based on the (expected) benefits and adverse effects of cholesterol drugs. The main objective of cholesterol drug treatment is to decrease the total cholesterol level below 5 mmol/L or LDL-cholesterol below 3 mmol/L. If a cholesterol drug causes side effects for the patient, the physician decreases the dosage, experiments with an another statin or ends the cholesterol drug therapy (the Finnish current care for dyslipidemia, 2011). In the model, the physician observes health effects associated with the use of a cholesterol drug. These effects affect the patient's current utility from the drug (side effects) and the probability that the statin therapy ends.

The Pharmaceuticals Pricing Board, that is subordinated to the Ministry of Social Affairs and Health in Finland, regulates drug prices with price ceilings. The regulation decreases variation in drug prices across years. Price ceilings were likely to be binding for Fluvastatin, Atorvastatin and Rosuvastatin that remained under patent protection during the whole observation period $2003 - 2006$ and thus did not face fierce competition from generics. In the empirical application, I assume that drug prices are constant over time in order to reduce computational burden.[10] Table 2.1 shows that the averages prices of cholesterol drugs vary still across products. In my sample, that is described below, the

[6] Within the group of an active ingredient, statins differ also in drug forms, package sizes, strengths and prices. Besides these active ingredients, combination preparations of a statin and another active ingredient are also in the market.

[7] For example, the statin therapy is useful for men, post menopausal women and patients who have arterial disease or diabetes. The risk of side-effects can increase with genetic susceptibility and certain drug interactions. Approximately 5% of patients have muscular symptoms (the Finnish current care for dyslipidemia, 2011).

[8] See "Controlling Cholesterol with Statins" by U.S. Food and Drug Administration (2010).

[9] Lifestyle changes, including exercising and changes in diet, are often adequate for a low-risk patient. However, patients are often unwilling to change their lifestyles (Johnston, 1999).

[10] This assumption implies that prices are not included in a set of state variables, which simplifies the computation of a dynamic demand model.

average price of the oldest active ingredient, Simvastatin, was 85% lower than the average price of the most expensive statin, Atorvastatin.

Table 2.1: Market and sample description

Active ingredient	ATC[1]	Marketing authorization date[2]	Average cost[3]	Market share[4]
Simvastatin	C10AA01	1992/02	16.15	0.46
Lovastatin	C10AA02	1988/06	47.85	0.01
Pravastatin	C10AA03	1992/04	95.63	0.04
Fluvastatin	C10AA04	1995/11	79.11	0.04
Atorvastatin	C10AA05	1997/04	110.12	0.28
Rosuvastatin	C10AA07	2003/03	83.23	0.18
	Mean	Std	Min	Max
Nbr of prescriptions	2.81	1.72	1.00	10.00
Difference between prescriptions (months)	6.79	6.22	0	38
Nbr of different drugs	1.26	0.48	1.00	3.00
Censoring indicator (1: yes, 0: no)	0.48	0.50	0.00	1.00
Non-rational expectations indicator (1: yes, 0: no)	0.45	0.50	0.00	1.00
Individuals	1000			
Observations	2812			

1. The Anatomical Therapeutic Chemical (ATC) Classification of an active ingredient.
2. The date of the first marketing authorization. Source: National Agency for Medicines.
3. An average over sample period.
4. Share/total prescriptions.

2.2.2 The dataset

I use a rich dataset of all purchased cholesterol drug prescriptions received by Finnish patients between January 1 in 2003 and December 31 in 2006.[11] The data contains the date of the prescription and the characteristics of patients, their physicians and products.[12]

I follow CS to prepare the dataset for the empirical analysis. To simplify the theoretical model, I remove patients with multiple statin prescriptions within the same day from the

[11]The data is provided by the Social Insurance Institution of Finland which is responsible for the provision of public social security benefits to Finnish residents.

[12]The unit of observation in the data is the prescription of a patient.

data. To avoid left censoring, I study "new" patients, who had their first prescription after the first six month of the observation period, i.e. after June 2003. I define a patient as right-censored if his last prescription in the data was received during the last six months of the observation period, i.e. during $7/2006 - 12/2006$.[13] For computational reasons, I draw a random sample of 1000 patients and exclude other cholesterol drug prescriptions than statins from the sample.[14] In my sample, the total number of patients is 1000, the number of observations is 2812 and the share of censored patients is 48% (Table 2.1).

The lower panel of Table 2.1 describes the drug therapy of patients in the sample. The average difference between two prescriptions was 6.8 months.[15] On average, patients had 1.3 different active ingredients and 2.8 prescriptions. Figure 2.1 shows that there is significant heterogeneity in the number of prescriptions across patients. The distribution is very skewed to the left, as 29% of patients had only one prescription, 21% had two prescriptions, 22% had three prescriptions and 29% had more than 4 prescriptions. The results are fairly similar for non-censored patients.[16]

The empirical literature has found that physicians are often unwilling to prescribe new treatments for their patients (e.g., Hellerstein, 1998, Coscelli, 2000, CS). The literature has explained this persistence with uncertainty and learning about the health effects of drugs. Consistent with this explanation, the probability to switch an active ingredient from the previous prescription is only 0.15 in the sample (Table 2.2).

The switching of a drug may be caused by both experimentation and learning. The incentive to experiment is the strongest at the beginning of the medical therapy when the physician has the least information about the average health effects of different drugs for a single patient. Switching at a later stage of the therapy can be induced by learning, especially if the patient has been using one drug for a long time.

I next investigate whether the data is consistent with learning and experimentation. Table 2.2 presents the probability of switching at different phases of the drug therapy, conditional on the total number of the patient's prescriptions. The results show that the

[13] If the censoring interval is too short, the exit rate is overrated and hence the estimation results may be biased. This is particularly true for patients whose prescriptions are received at the end of the sample period and who have more than two prescriptions.

[14] These excluded cholesterol drugs belong to the group of lipid modifying agents and have a market share of 2.17% in my sample. In the future, my plan is to increase the sample size by reducing the number of drugs through aggregation and perform robustness checks for the estimation results.

[15] Given the time difference, a 6 month censoring window may not be realistic. In the future, my plan is increase the window.

[16] To be more specific, 36% of non-censored patients had only one prescription, 23% had two prescriptions, 21% had three prescriptions and 20% had more than 4 prescriptions.

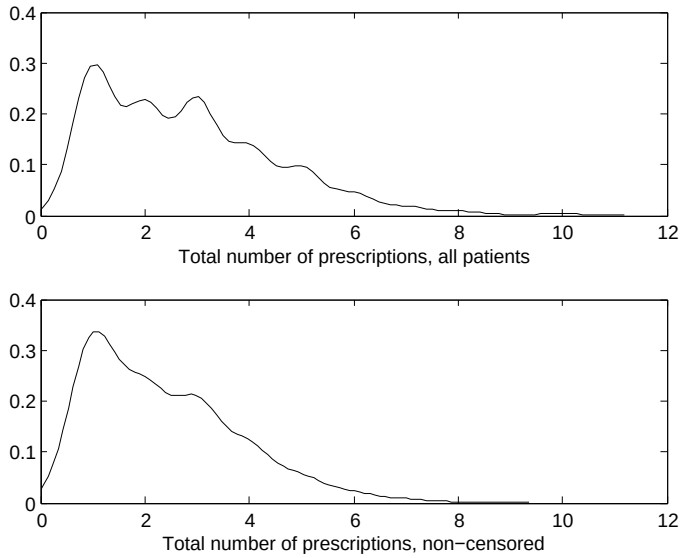

Figure 2.1: Kernel densities for the total number of prescriptions for all patients (the higher panel) and non-censored patients (the lower panel) in the sample

Table 2.2: The probability of switching from the previous prescription during the drug therapy in the sample of patients

Prescription nbr/Treatment length[1]	2	3	4	≥5	Total
2	0.122	0.119	0.188	0.122	0.133
3		0.179	0.188	0.098	0.155
4			0.234	0.140	0.182
≥5				0.176	0.176
Total	0.122	0.149	0.203	0.141	0.154

1. Total treatment length: the maximum number of the patient's prescriptions in the sample.
2. Patients with multiple prescriptions/physicians at some time point are excluded.

probability of switching is the highest both at the beginning and at end of the patient's drug therapy. The high switching probability at the beginning is consistent with experimentation, whereas the high switching probability at the end is consistent with learning. The results remain still very indicative without a structural model that isolates the roles of uncertainty and learning in demand.

2.3 The theoretical model

This section describes the dynamic model of demand under uncertainty that I then estimate. Consider a patient (she) who comes to the physician (he) to seek drug treatment for her medical condition. First, the physician makes an initial diagnosis about his fixed severity of illness, or type. The severity of illness affects both the symptomatic effects of drugs and the exit rate from the drug therapy. Conditional on the diagnosis, the physician evaluates the patient's risk for CVDs and side effects associated with drug treatments. This evaluation affects the physician's prior beliefs about the effects of drugs on the patient's symptoms and exit rate. Based on his prior beliefs, the physician selects a drug treatment that best matches the patient's medical condition.[17]

After the patient has taken the prescribed drug, she revisits the physician. The physician observes two health effects. The first effect measures the effects of the drug on the patient's symptoms. The second health effect captures the effect of the drug on the probability that the therapy ends, such as how effectively the drug and the patient's life style changes decreased her cholesterol levels. Then, if the patient's therapy does not end, the physician makes a new treatment decision. This decision is based on the physician's beliefs regarding the average health effects, or match values, of drugs, conditional on observed health effects. Again, the patient takes the drug and revisits the physician who observes the health effects of the drug. Conditional on the health effects that affect the exit rate, the patient's drug therapy either ends or continues. These steps are repeated until the patient exits from the therapy. During the course of the patient's therapy, the physician may learn the average health effects of drugs from the patient's treatment history.

Following CS, I assume the physician maximizes solely the expected discounted utility of the patient. In the model, all physicians have the same probability of choosing a drug treatment for the patient. This assumption abstracts away potentially important agency issues.[18] The model also ignores the possibility of physician specific effects. Such effects

[17] I follow CS and consider patients who have received at least one drug prescription.

[18] See for example Iizuka (2007) and (2011) for the empirical analysis of agency issues in the pharma-

can arise for example if the personal experience of a physician affects prescription behavior (see Chapter 3). The main focus of this paper is on learning by a physician about the quality of the match between a particular patient and different drugs. As most of the cholesterol drugs have been on the Finnish market for almost two decades (see Table 2.1), learning across patients is not likely to have a big role in my application.[19]

Next, I present the details of the dynamic demand model. I begin by defining the severity of illness. Then, I present the physician's decision-making problem for a patient. After that I define health effects, go through the learning process and present the exit probability and the set of state variables. Finally, I provide the physician's value function.

The severity of illness

Assume that the patient comes to seek a drug treatment for her medical condition for the first time. First, she is randomly matched to a physician who makes an initial diagnosis about her fixed severity of illness, or type. The probability that the patient is of type k, $k = 1, ..., K$, is given by p_k such that $\sum_{k=1}^{K} p_k = 1$. Illness types capture heterogeneity in the medical conditions of patients, such as lifestyle patterns or the amount of LDL cholesterol in blood, that affects the distributions of health effects. The illness type is observed by a physician but not by the econometrician. In my empirical application, I assume that $k = 2$.[20]

A drug treatment choice

Conditional on the illness severity type k of patient j, the per-period utility function of the physician (or the patient) is assumed have a constant absolute risk aversion sub-utility function for the symptomatic effect of the product n at time t, x_{jknt}.[21] I assume that the current utility is linear in the price, p_n, and the Type I extreme value distributed error term, e_{jknt}. To be more precise, I consider the following per-period utility function for the patient[22]

ceutical market.

[19] For learning across patients, see Coscelli and Shum (2004) and Kim (2010).

[20] In CS, the number of types is 4. My plan evaluate the robustness of results to the number of types.

[21] As all physicians of the patient are alike, the physician-index does not enter the utility function.

[22] I also make a conditional independence assumption $p(z', e'|z, e, n, \theta) = p(z'|z, n, \theta_2)p(e'|\theta_3)$, where z' are the other random state variables than e. The assumption is commonly made in the dynamic discrete choice literature (see e.g. Aguirregabiria and Mira, 2010). Note that the transition probability for the

$$u_{jknt} = -e^{-r \cdot x_{jknt}} - \alpha \cdot p_n + e_{jknt}, \tag{2.1}$$

where $r > 0$ is the degree of risk aversion and α is the price coefficient.

The physician makes the drug treatment choice among different active ingredients n, $n \in \{1, ..., N\}$, such that the chosen product maximizes the present discounted utility of the patient. This implies that the outside option is not in the physician's choice set.[23] Instead, I follow CS and assume that the probability that the patient's therapy ends evolves endogenously with the physician's prescriptions. The prescription is made conditional on a set of state variables at time t, S_{jkt}, that is specified below. The Markovian decision problem of the physician is

$$\begin{aligned} V(S_{jkt}) &= \max_n \{E\big[u(x_{jknt}, p_n, e_{jknt}) + \\ &\quad \beta(1 - w_{jkt}) E\big[V(S_{jk(t+1)}) | x_{jknt}, y_{jknt}, n\big] | S_{jkt}\big], \forall n\} \\ &= \max_n \{\overline{V}_{jknt} + e_{jnkt}, \forall n\}, \end{aligned} \tag{2.2}$$

where w_{jkt} is the indicator variable that takes value 1 if patient j exits from the drug therapy after period t, β is the discount factor and $\overline{V}_{jknt} = \overline{V}^k(S_{jnt})$ is the choice specific value function. The expectation in (2)-(3) is taken over two health effects, x_{jknt} and y_{jknt}. Whereas x_{jknt} captures the effects of the drug on the patient's symptoms, y_{jknt} affects the probability that the patient exits from drug the drug therapy. Because health effects are observed after the patient has taken drug n but before period $t + 1$, the expectation of the value function at $t + 1$ is conditional on x_{jknt} and y_{jknt}.

Health effects

After the patient has taken drug n at the end of period t, the physician observes health effects that affect her symptoms and exit rate from the drug therapy,

$$x_{jknt} = \mu_{jkn} + \sigma_{ex} e^x_{jknt}, \text{ where } \mu_{jkn} \sim N(\underline{\mu}_{kn}, \underline{\sigma}^2) \tag{2.3}$$

and

$$y_{jknt} = \nu_{jkn} + \sigma_{ey} e^y_{jknt}, \text{ where } \nu_{jkn} \sim N(\underline{\nu}_{kn}, \underline{\tau}^2) \tag{2.4}$$

states depends on the previous history via the most recent values of the state variables.

[23] In Section 2.3, I include the outside option to the physician's choice set.

where e^x_{jknt} and e^y_{jknt} are $N(0,1)$ distributed independent random variables.

Note that the match value means can vary between patient types and drugs. The variances of the match values are assumed to be constants.[24] For computational reasons, I follow CS and assume that health effects are uncorrelated, i.e. $\text{cov}(x_{jknt}, y_{jknt}) = 0$.[25] This assumption implies that experimenting with one cholesterol drug does not change the physician's view about another, possibly similar cholesterol drug. In my application, the assumption may not hold for at least two reasons: first, side effects can affect both the patient's symptoms and his exit rate from the drug therapy. Second, cholesterol drugs do not have significant clinical differences in reducing cardiovascular (National Institute for Health and Clinical Excellence, 2006). Thereby correlated learning can be relevant. Assuming zero correlation between health effects may bias results.

Conditional on realized health effects, the physician updates his beliefs about the unobserved match values, μ_{jkn} and ν_{jkn}. The posterior beliefs for the mean and variance of the symptomatic match value, $\underline{\mu}_{kn}$ and $\underline{\sigma}^2$, are (similarly for $\underline{\nu}_{kn}$ and $\underline{\tau}^2$)

$$\mu_{jknt} = \begin{cases} \frac{\sigma^2_{ex}\mu_{jkn(t-1)} + \sigma^2_{jn(t-1)}x_{jknt}}{\sigma^2_{ex} + \sigma^2_{jn(t-1)}} \text{ if drug n is taken at time } t, \\ \mu_{jkn(t-1)} \text{ otherwise,} \end{cases} \tag{2.5}$$

$$\sigma^2_{jnt} = \begin{cases} \frac{\sigma^2_{ex}\underline{\sigma}^2_n}{\sigma^2_{ex} + l_{jn(t-1)}\underline{\sigma}^2_n} \text{ if drug n is taken at time } t, \\ \sigma^2_{jn(t-1)} \text{ otherwise,} \end{cases} \tag{2.6}$$

where l_{jnt} the number of times patient j has tried the drug n up to (and including) time t.

When the number of prescriptions l_{jnt} increases and information on the average health effects accumulates, the variances of the posterior distributions decrease towards zero and the physician learns the distributions of the match values.

I assume that the physician has rational expectations about other drugs besides Rosuvastatin that have been long on the market. The assumption implies that the physician knows the prior distributions of drugs, i.e. $\mu_{jkn1} = \underline{\mu}_{kn}$, $\nu_{jkn1} = \underline{\nu}_{kn}$, $\sigma^2_{jn1} = \underline{\sigma}^2_n$ and $\tau^2_{jn1} = \underline{\tau}^2$. The rational expectations assumption may not be true for Rosuvastatin that has been

[24] I have also experimented with the model where the variance of the symptomatic match value was allowed to differ across products, i.e. $\underline{\sigma}^2_n$. The results suggested that the variance estimates were almost the same for all products. Therefore, I assume in the empirical part of this chapter that $\underline{\sigma}^2$ is a constant across products. In the future, my plan is to decrease the number of products through aggregation and allow for $\underline{\sigma}^2_n$ to vary across products.

[25] For correlated learning, see Dickstein (2011).

marketed since the beginning of the sample period (17.6.2003). For this reason, I assume that the physician did not know the prior distribution of the symptomatic effects of Rosuvastatin[26] if his patient had her first prescription before 2005 (i.e. $7/2003 - 12/2004$).[27] Thus, physicians may have common, non-correct predictions for the distributions of the match values for Rosuvastatin during the first years. By allowing for non-rational expectations, I control for the possibility that physicians may learn about the average health effects of Rosuvastatin across patients by reading medical journals and attending conferences.

The exit probability

Assume that the patient has taken drug n and she revisit the physician at the end of period t. While investigating the patient, the physician observes health effects, y_{jknt}, that are associated with the use of the drug and affect the patient's exit rate from the drug therapy. Conditional on these health effects and the patient's exit rate at the previous prescription, I assume that the exit rate of patient i with type k at the end of period t is

$$h_{jkt} = \frac{\frac{h_{jk(t-1)}}{1-h_{jk(t-1)}} + d_{jnt}y_{jknt}}{1 + \frac{h_{jk(t-1)}}{1-h_{jk(t-1)}} + d_{jnt}y_{jknt}}, \tag{2.7}$$

where d_{jnt} is the indicator variable that takes value 1 if the patient takes drug n at time t and $h_{jk0} = \theta_k$ is the initial value of the exit rate. For $h_{jkt} \in [0, 1)$, the higher the health effects, y_{jknt}, are, the more likely the patient exits from the therapy at time t.[28] Correspondingly, the higher the previous exit probability, $h_{jk(t-1)}$, is, the higher the corresponding period-t probability is.

A set of state variables

A set of state variables for patient j at time t, S_{jkt}, consists of the exit rate, h_{jkt}, and the following drug specific variables: the number of prescriptions, l_{jnt}, the posterior means,

[26] In the empirical application, a non-rational expectation about the exit match value was very imprecisely estimated. Therefore, I allow non-rational expectations only for symptomatic effects.

[27] In $2005 - 2006$, I assume that the physician has rational expectations about Rosuvastatin. In my sample, physicians had non-rational expectations for 45% of patients.

[28] As noted by CS, very large and negative values of y_{jknt} lead to negative values of h_{jkt}. In this case, h_{jkt} is no longer a valid probability. However, the simulations of the model did not produce any negative values of h_{jkt}.

μ_{jknt} and ν_{jknt}, and the preference shocks, e_{jknt}, for drugs n, $n = 1, ..., N$.

The value function

By using the law of iterated expectations and the moment generating function of the normal distribution, the value function of the physician can be expressed as

$$V(S_{jkt}) = \max_n \{-e^{-r\mu_{jknt}+\frac{1}{2}r^2(\sigma^2_{jnt}+\sigma^2_{ex})} - \alpha p_n + e_{jknt} + \tag{2.8}$$

$$\beta E\big[(1 - h_{jkt}(h_{jk(t-1)}, y_{jknt}))E[V(S_{jk(t+1)})|x_{jknt}, y_{jknt}, n]|S_{jkt})\big]\}. \tag{2.9}$$

When the risk-averse physician is choosing a new treatment for the patient, he faces a trade-off between having a low present utility caused by the "risk-premium" $\frac{1}{2}r^2(\sigma^2_{jnt}+\sigma^2_{ex})$ and the option value that contains new information through health effects x_{jknt} and y_{jknt} that are realized after the patient has taken drug n. The more risk aversive the physician is, the more the risk-premium decreases the probability of choosing drug n. Moreover, learning about the average symptomatic effects of the drug for the patient, μ_{jkn}, decreases the variance of the posterior distribution σ^2_{jnt} and the risk premium. For this reason, learning increases the probability of choosing drug n again.

Because the optimization problem has a stationary Markovian structure, only the values of current state variables in S_{jk} affect the expectation of the physician about the future. The value function can thereby be expressed as

$$\begin{aligned} V(S_{jk}) &= \max_n \{E\big[u(x_{jkn}, p_n, e_{jkn}) + \\ &\quad \beta(1 - h_{jk})E\big[V(S'_{jk})|x_{jkn}, y_{jkn}, n\big]|S_{jk}\big], \forall n\}, \end{aligned} \tag{2.10}$$

where S'_{jk} is a set of state variables in the next period.

To decrease the dimensionality of the state space, I take the expectation of the value function in (10) over preference shocks, $e_{jk1}, ..., e_{jkN}$ (see e.g. Aguirregabiria and Mira, 2010, Rust, 1987)

$$E(V(S_{jk})) = \gamma + \log(\sum_{n=1}^{N} \exp(\overline{V}^k(S_{jk}))), \tag{2.11}$$

where $\gamma = 0.5772$ is Euler's constant. I follow CS and adapt a method by Keane and Wolpin (1994) to approximate the expected value function.[29]

2.4 The econometric model

In this section, I describe the simulated log-likelihood function and discuss identification. I use the following data to compute the simulated likelihood function: the vector of indicator variables, $d_{j1t}, ..., d_{jNt}$ where d_{j1t} equals 1 if the patient takes drug n in period t, the number of the patient's prescriptions for drug n by time t, l_{jnt}, the average price of the drug, p_n, the indicator for whether the drug therapy of the patient is censored at the end of the observation period, c_j, and the length of the cholesterol drug therapy, T_j.

The likelihood function contribution of (non-censored) patient i includes probabilities for chosen drugs in each period, $1, ..., T_i$, and the probabilities that the drug therapy is continued up to period T_i and ended at the end of period T_i,

$$L_j^{nc} = \sum_k^K p_k E[\underbrace{\prod_t^{T_j-1}((\prod_n \lambda_{jknt}^{d_{jnt}})(1-h_{jkt}))(\prod_n \lambda_{jnT_jk}^{d_{jnT_j}})}_{\text{Continue the therapy of patient } i \text{ with type } k \text{ up to } T_i} \quad h_{jkt}], \tag{2.12}$$

where $\lambda_{jknt} = \frac{e^{\overline{V}_{jknt}}}{\sum_{n'=1}^N e^{\overline{V}_{jn'tk}}}$ is the choice probability for drug n at time t. The expectation in (12) is taken over health effects x_{jnkt} and y_{jnkt} and preference shocks e_{jknt}[30] that are unobserved to the econometrician. As patient types k are also unobserved to the econometrician, their effects on the likelihood contribution must be averaged out.

The likelihood contribution of a censored patient is otherwise the same as (12), except that her exit from the drug therapy is not observed,

$$L_j^c = \sum_k^K p_k E[\prod_t^{T_j-1}((\prod_n \lambda_{jknt}^{d_{jnt}})(1-h_{jkt}))\prod_n \lambda_{jnT_jk}^{d_{jnT_j}}]. \tag{2.13}$$

Because the computation of the likelihood function would require integration over the distribution with a very high dimension and the choice probabilities do not have a closed

[29] The basic idea of the method is to solve the dynamic programming problem recursively at a subset of state space points and approximate in other points by using interpolation. See CS and Keane and Wolpin (1994) for details.

[30] These error terms are needed to compute the predicted choices of physicians for their patient.

form solution, the likelihood function is approximated with Monte Carlo integration. The simulated likelihood contribution of a non-censored patient is calculated by replacing the expectation with an average over S simulations,

$$L_j^{nc,s} = \frac{1}{S}\sum_{s=1}^{S}\sum_{k=1}^{K} p_k\Big[\prod_{t}^{T_j-1}\big((\prod_{n}(\lambda_{jknt}^s)^{d_{jnt}})(1-h_{jkt}^s)\big)\Big](\prod_{n}(\lambda_{jnT_jk}^s)^{d_{jnT_j}})h_{jkt}^s \tag{2.14}$$

and similarly for a censored patient.

The simulated log-likelihood function is given by

$$\ln L^{N,s}(\theta) = \sum_{j=1}^{N}[(1-c_j)\ln L_j^{nc,s}(\theta) + c_j \ln L_j^{c,s}(\theta)]. \tag{2.15}$$

To get the simulated version of the likelihood function, I first draw the signals and preference shocks of a patient for all products, conditional on her severity of illness. Next, I compute choice specific value functions for each simulation, product and illness type. Then, I calculate the simulated counterpart of the value function at time $t = 1$, update the beliefs by using the simulated signals and the updating formulas presented in equations (2.7)-(2.8) and compute the simulated exit rate. I repeat these steps for all periods when the patient received a prescription. I estimate the model using 10 simulations for each patient[31]. Because the number of simulation is small, the MSL estimator is inconsistent.[32] The results must thus be interpreted with caution.

Identification

Variation in drug choices with the number of prescriptions across patients help to identify the model parameters. Initial prescriptions identify the prior means of symptomatic match values, $\underline{\mu}_{kn}$, because the physician has not yet observed any symptomatic effects. Prescriptions in the early vs. the late stages of the treatment, or learning, identify the variance of symptomatic match values, $\underline{\sigma}_n^2$. Because risk aversive physicians are more reluctant to switch a drug treatment, persistence in drug choices identifies the risk aversion coefficient, r. Changes in prescription with the number of times the physician has

[31] My plan is to increase the number of simulations. CS estimated the model by using 30 simulations.

[32] This is because the simulated likelihood contribution of individual i $\ln\hat{f}_i$ is biased for $\ln f_i$ even if $\hat{f}_i$ is unbiased for f_i (see e.g. Cameron and Trivedi, 2005).

prescribed the drug n, l_{jnt}, identifies the variance of symptomatic health effects, σ^2_{ex}.[33] I normalize the price of the cheapest drug, Simvastatin, to zero to identify the price coefficient α.

I then consider variation that identifies the parameters of health effects that affect the exit rate from the drug therapy. The law of iterated expectations implies that the exit rate conditional on the state variables is

$$E(h_{jkt}|S_{jkt}) = E_{\nu_{jkn}} E_{y_{jknt}|\nu_{jkn}} \Big[\frac{\frac{h_{jk(t-1)}}{1-h_{jk(t-1)}} + d_{jnt} y_{jknt}}{1 + \frac{h_{jk(t-1)}}{1-h_{jk(t-1)}} + d_{jnt} y_{jknt}} | S_{jkt} \Big], \tag{2.16}$$

where the first expectation is taken over the mean of health effects that affect the exit rate, ν_{jkn}, and the second one is taken over health effects conditional on the mean, $y_{jknt}|\nu_{jkn}$. Because the effect of y_{jknt} on the exit rate varies with the number of prescriptions and is different across patients, the mean and variance of the exit match value, $\underline{\nu}_{kn}$ and $\underline{\tau}^2$, and the variance of y_{jknt} conditional on ν_{jkn}, σ^2_{ey}, are identified. The parameter of the initial exit rate, θ_k, is identified because it affects the exit rate in the previous period, $h_{jk(t-1)}$.

2.5 Estimation results

Table 2.3 describes the estimation results of the dynamic matching model. The first panel presents the estimates of initial exit and type probabilities, h_{jk1} and p_k for types $k = 1$ (line 1) and $k = 2$ (line 2). The second panel contains the means and the variance of the symptomatic match values for drugs n, $n = 1, ..., 6$ (rows), and patient types k, $k = 1, 2$ (columns), i.e. $\underline{\mu}_{kn}$ and $\underline{\sigma}^2$. Analogously, the third panel includes the means and the variance of the exit match values for drugs n, $n = 1, ..., 6$ (rows), and patient types k, $k = 1, 2$ (columns), i.e. $\underline{\nu}_{kn}$ and $\underline{\tau}^2$. Recall that due to the non-rational expectations of the physician, the mean symptomatic match value of Rosuvastatin is allowed to be different from the true match value mean if the first prescription was taken before 1.1.2005. The fourth panel presents the variances of the symptomatic and exit rate health effects conditional on the match values, i.e. σ^2_{ex} and σ^2_{ey}. The fifth panel contains price and risk aversion coefficients, α and r. I follow CS and fix the discount rate to 0.95.

[33] Prescription choices change because the posterior beliefs for the mean and variance of the symptomatic match value, $\underline{\mu}_{kn}$ and $\underline{\sigma}^2$, change with the number of prescription for drugs n (see (5) – (6)).

Table 2.3: Estimates from the dynamic demand model in the sample of patients

	Est.	Std.err.	Est.	Std.err	Mean, over types
	Exit prob. h_{j1k}		Type prob. p_k		
θ_1 (Type 1)	0.493	0.037	0.614	0.039	
θ_2 (Type 2)	0.494	0.037	0.386		
Symptomatic match values, μ_{jkn}	Type 1		Type 2		
Means:					
Simvastatin μ_{1k}	0.782	0.578	0.011	0.003	0.485
Lovastatin μ_{2k}	-0.013	0.016	-0.014	0.021	-0.014
Pravastatin μ_{3k}	-0.012	0.033	0.017	0.032	-0.001
Fluvastatin μ_{4k}	0.005	0.005	0.009	0.001	0.007
Atorvastatin μ_{5k}	0.067	0.073	-0.012	0.001	0.037
Rosuvastatin:					
rational expect. μ_{6k}	-0.111	0.187	0.007	0.025	-0.066
non-rational expect. μ_{j61k}	0.008	0.001	-0.023	0.000	-0.004
Variance σ^2	3.416	1.351			
Exit match values, ν_{jkn}	Type 1		Type 2		
Means:					
Simvastatin ν_{1k}	0.046	0.183	0.000	0.001	0.028
Lovastatin ν_{2k}	-0.001	0.005	0.001	0.007	0.000
Pravastatin ν_{3k}	0.001	0.011	0.001	0.010	0.001
Fluvastatin ν_{4k}	-0.001	0.002	-0.002	0.000	-0.001
Atorvastatin ν_{5k}	0.001	0.023	0.000	0.000	0.001
Rosuvastatin ν_{6k}	-0.008	0.059	0.005	0.008	-0.003
Variance τ^2	0.999	1.556			
Signal variance					
Symptomatic signal σ^2_{ex}	0.975	0.725			
Exit signal σ^2_{ey}	1.633	3.568			
Price coefficient, α	0.126	0.021			
Risk-aversion parameter, r	0.975	0.713			
Discount factor, β	0.95				
Number of individuals	1000				
Draws/individual	10				
Log likelihood function	5244				

The estimation results for the distribution of patient types provides evidence on heterogeneity in the severity of illness across patients with high cholesterol. Recall that the illness type captures factors observed by the physician but not by the econometrician that affect the distributions of health effects. The first set of parameters in Table 2.3 shows that type 1 and 2 patients have an equal changes of exit at the beginning of the drug therapy. The exit rate h_{jkt} can still differ across patients, depending on observed health effects and the physician's drug choices. I also find that patients are of type 1 with probability 0.61 and of type 2 with probability 0.39.

The second set of parameters in Table 2.3 provides evidence on heterogeneity in the distributions of health effects across patient types. The results suggest that the means of symptomatic match values, $\underline{\mu}_{1k}, ..., \underline{\mu}_{6k}$, differ between patient types but most of the means are somewhat imprecisely estimated. The rank of drugs based on the mean of symptomatic match values is (1,5,4,3,2,6)[34] for type 1 patients and (3,1,4,6,5,2) for type 2 patients. On average, the cheapest drug Simvastatin (drug 1) performs the best for type 1 patients and Pravastatin (drug 3) for type 2 patients. Patients of type 1 have the worst match on average with Rosuvastatin (drug 5) and type 2 patients with Lovastatin (drug 2). The estimated prior means of Rosuvastatin under non-rational expectations indicate that physicians were initially too optimistic about the average symptomatic effects of Rosuvastatin for type 1 patients and too pessimistic for type 2 patients. On average, patients have the best symptomatic match with Simvastatin and the worst with Lovastatin.

In Figure 2.2, I illustrate heterogeneity in the symptomatic match value distributions further. When a patient is of type 1, the efficiency of Rosuvastatin is much worse than that of other statins. For type 2 patients, the distributions of the match values overlap much more than for type 1 patients. The standard deviation estimate of symptomatic health effects is large (3.42) compared with the estimates of the mean symptomatic match values for both patient types ($-0.11 - 0.78$). Heterogeneity in symptomatic health effects implies that physicians face substantial uncertainty about the symptomatic effects of cholesterol drugs for their patients. Because the match values values are not known to physicians at the beginning of therapy, learning may significantly help physicians to find the best drug treatments for their patients.

[34] Active ingredients: 1: Simvastatin, 2: Lovastatin, 3: Pravastatin, 4: Fluvastatin, 5: Atorvastatin, 6: Rosuvastatin

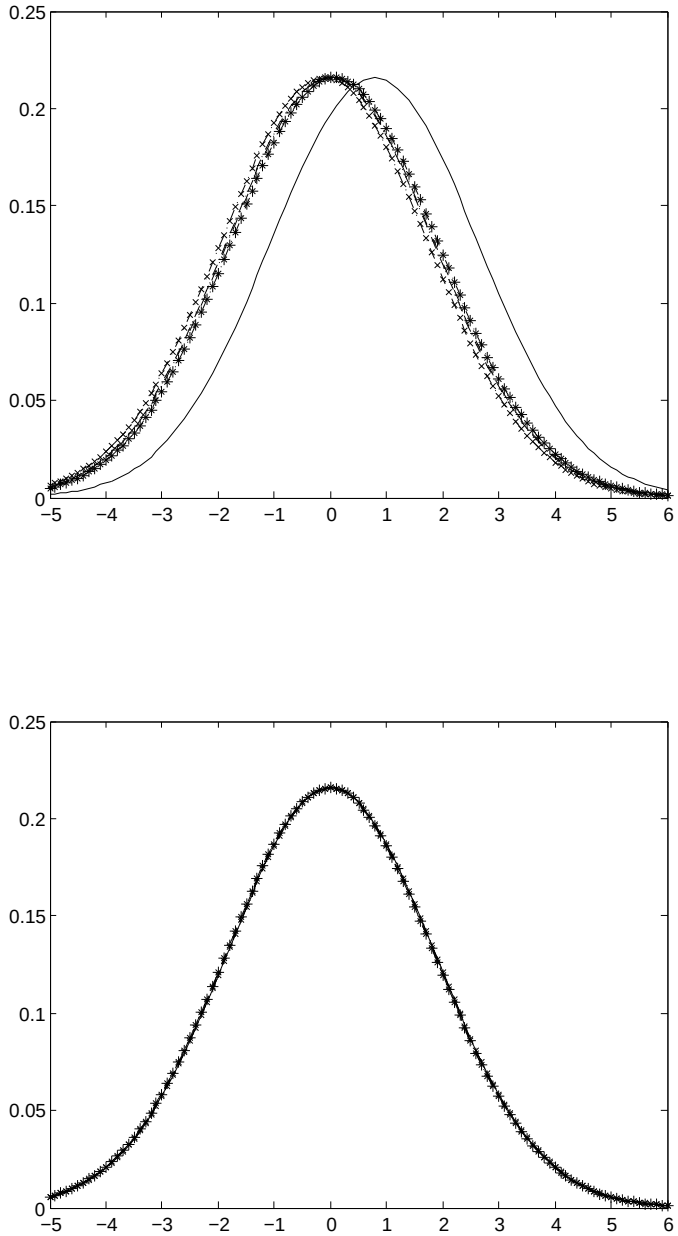

Figure 2.2: Symptomatic match value distributions for Types 1 (higher figure) and 2 (lower figure): Simvastatin (-), Lovastatin (–), Pravastatin (:), Fluvastatin (-.), Atorvastatin (-.*), Rosuvastatin (-.x). in the sample of patients

The third set in Table 2.3 presents the estimates for the type-specific distributions of

match values that affect exit rates. The results show that the average exit match values, $\underline{\nu}_{1k}, ..., \underline{\nu}_{6k}$, are again heterogeneous across patient types. The rank of drugs in this dimension is (1,5,3,4,2,6) for type 1 patients and (6,2,3,5,1,4) for type 2 patients where higher ranks indicate higher exit rates. On average, patients using Simvastatin have the highest and patients using Rosuvastatin have the lowest exit rates from the cholesterol drug therapy.

The estimates for variances σ^2_{ex} and σ^2_{ey} are large (0.98 and 1.63, respectively) compared with the estimates of average health effects. This may suggest that the physician faces significant uncertainty about health effects even after learning the patient and drug specific match values. Variances σ^2_{ex} and σ^2_{ey} are, however, imprecisely estimated (standard deviations 0.73 and 3.57) which prevents from making any stronger conclusions.

The final set of parameters contains price and risk aversion coefficients. Physicians are estimated to be insensitive to changes in the prices of cholesterol drugs since the point estimate of the price coefficient, α, is only 0.13. My findings also indicate that physicians are risk averse: the point estimate of the risk aversion coefficient, r, is 0.98 with a standard deviation of 0.71. In the next section, I will show that risk aversion and uncertainty about the health effects of cholesterol drugs decrease the incentives of physicians to experiment with new treatments.

Overall, the estimation results indicate that the market leader Simvastatin has relatively good symptomatic efficiency compared with the other statins. Atorvastatin has the second market highest share and it performs reasonably well in both the symptomatic and exit rate dimension. The health effects of Rosuvastatin, that has a market share of 15% in the sample, differ between patient types. My results also show that the health effects of the cholesterol drugs are heterogeneous across patients. These findings are consistent with the results of CS from the Italian anti-ulcer market.

2.6 Model fit and counterfactual experiments

In this section, I analyze the roles of learning and experimentation in demand for cholesterol drugs. To do this, I perform several counterfactual policy simulations and evaluate how they affect the patient's expected discounted utility, the length and costs of the drug therapy, experimentation incentives and the market shares of the drugs measured.[35] First, I replicate the policy experiments performed by CS. To be more precise, I investigate what

[35] I measure the market share of a drug by its prescription share of total prescriptions.

happens if a physician knew the match values of a patient with drug treatments.[36] Without uncertainty, the physician has higher incentives to prescribe new, less well-known drug treatments for the patient. Second, I force the physician to prescribe the first active ingredient for the patient every period during the drug therapy. In this experiment, the physician is not allowed switch to an another drug after learning the match quality of the first drug.

To understand in-depth the consequences of learning and experimentation in the market for statins, I perform the following new counterfactuals. First, I make the physician myopic (the discount factor equals zero) in his choices. This experiment removes experimentation incentives because the physician does not take into account the consequences of his actions on the patient's future health. Next, I analyze how risk aversion affects incentives to experiment. To do this, I decrease the risk aversion coefficient from 0.98 to 0.50. This experiment decreases switching costs caused by uncertainty and increases incentives to experiment. Finally, I consider the implications of the policy that prevents learning on pharmaceutical demand. This experiment corresponds to the situation where the physician does not investigate the patient and decides on a new prescription based on his prior knowledge. The results are compared to the raw data and the baseline case implied by the estimated parameters. To perform the counterfactuals, I simulate the sequences of treatments for 5000 patients and use the prior means of Rosuvastatin under rational expectations.

I begin by investigating the speed of learning by a physician about the average health effects of cholesterol drugs for his patient. Figures 2.3 and 2.4 present the evolution of the posterior means and variances when the number of prescriptions for each drug increases by one every period. The results suggest that uncertainty regarding to the average health effects decreases fast. The posterior variance of the symptomatic match value drops 44% and the posterior variance of the exit match value drops 28% after the first health effects have realized.[37] The diminishing of uncertainty slows down after the first prescription.

[36] Specifically, I set the variances of posteriors, σ^2_{jnt} and τ^2_{jnt}, to zero. Recall that there is still heterogeneity among patients in the exit rate and symptomatic effects x_{jknt} and y_{jknt} because the variances of health effects are non-zero.

[37] Even though this effect is substantial, CS found that in the anti-ulcer market, the posterior variance of the symptomatic match value decreased over 70% after a single prescription.

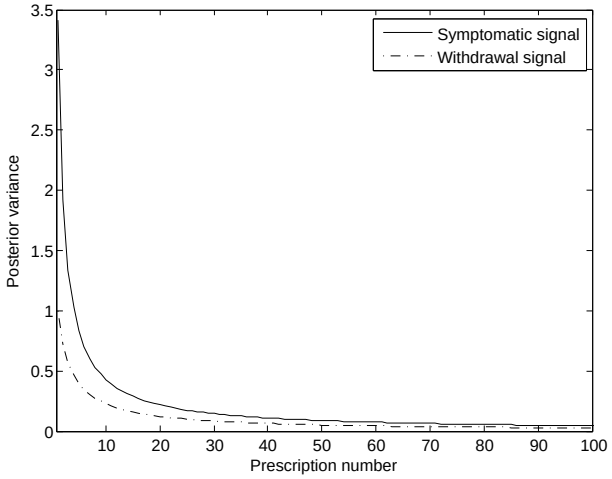

Figure 2.3: The evolution of posterior variances σ^2_{jnt} and τ^2_{jnt} when l_{jnt} increases by one every period, an average over patients, products and types in the sample

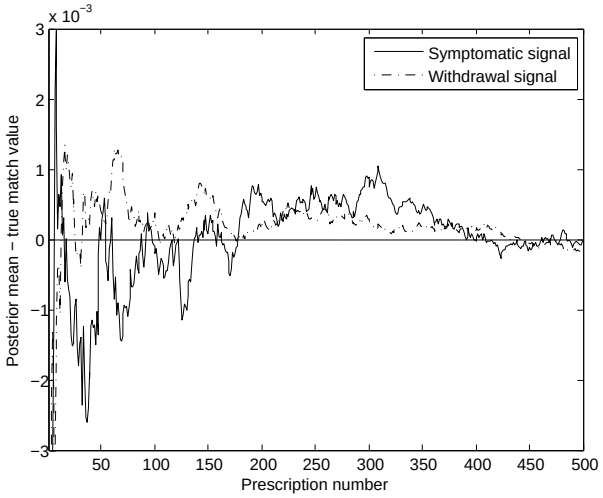

Figure 2.4: Differences between posterior means and true match values, $\mu_{jknt} - \mu_{jkn}$ and $\nu_{jknt} - \nu_{jkn}$, when l_{jnt} increases by one every period, an average over patients, products and types in the sample

Before analyzing the consequences of the policy experiments on demand for cholesterol drugs, I first assess how well the structural model fits data. The results in Table 2.4 suggest that the predicted outcomes are somewhat similar with those that are observed in data (Table 2.4).[38] The average length of the statin therapy is 2.8 prescriptions in the data whereas the one predicted by the model is 2. The predicted number of different drugs taken during the therapy is the same (1.3) as in the data. The predicted costs of the therapy relative to the average cost of Simvastatin is, however, much smaller than the observed total costs of the therapy in the sample. The difference arises because the market shares of expensive drugs, Atorvastatin and Rosuvastatin are underestimated and the market share of the cheap drug, Lovastatin is overestimated. The results must thus be interpreted with this caveat.

[38] In the future, my plan is to evaluate how the fit of the model could be improved.

Table 2.4: Counterfactual experiments and model fit in the sample of patients

Model	Data	Baseline	Complete info[2]	Use the first drug	Myopic	Risk aversion r=1/2	No learning[3]
Average discounted utility		-1.380	3.195	-4.687	-3.857	2.699	-8.869
Treatment length	2.812	1.972	1.975	1.972	1.972	1.973	1.972
Total costs (eur)[4]	152.305	49.299	70.482	43.235	49.222	75.582	42.954
Different drugs	1.257	1.256	1.597	1.000	1.254	1.600	1.347
Market share[5]							
Simvastatin	0.532	0.600	0.419	0.646	0.600	0.376	0.651
Lovastatin	0.003	0.108	0.174	0.099	0.108	0.188	0.092
Pravastatin	0.026	0.074	0.097	0.066	0.073	0.104	0.065
Fluvastatin	0.051	0.090	0.120	0.077	0.089	0.129	0.077
Atorvastatin	0.243	0.055	0.081	0.045	0.055	0.086	0.043
Rosuvastatin	0.145	0.074	0.109	0.068	0.074	0.118	0.072

[1] Number of simulated individuals=5000. The sample of 1000 patients is described in Section 2.2 of this chapter.

[2] Complete information: σ^2_{jnt} and τ^2_{jnt} are set to zero.

[3] No learning: physicians do not receive signals about the match values of the patient.

[4] Total costs (eur), the price normalized with the price of Simvastatin

[5] Market share: the share of prescriptions for a product from the total number of prescriptions.

The results from the counterfactual experiment with complete information suggest that uncertainty has substantial effects on treatment outcomes and costs (Table 2.4). When there is no uncertainty about the patient's average health effects, the expected discounted utility is higher than in the baseline case. The average treatment length (2 prescriptions) does not change from the baseline. The physician is more willing to prescribe different drugs when there is no uncertainty. To be more specific, the average number of different drugs is 1.6 under complete information and 1.3 in the baseline. The market share of the cheapest drugs, Simvastatin and Lovastatin, decreases 16% from the baseline.[39] As a result, the total costs of the statin therapy (relative to the average cost of Simvastatin) are 43% higher under the complete information scenario than in the estimated baseline.

In the second counterfactual experiment, the physician is forced to prescribe the first drug to the patient every period during the drug therapy. This policy rules out experimentation but lets the physician to learn from realized health effects. Unsurprisingly, the expected discounted utility decreases because the physician cannot switch to a bet-

[39] This happens because removing uncertainty decreases the risk-premium and increases the role of idiosyncratic preference shocks e_{jknt} in the utility function, making market shares more evenly distributed.

ter treatment alternative after getting more information about the quality of the first drug. Still, market shares, total costs and treatment length do not change much from the estimated benchmark.

In the next experiment, I investigate the significance of experimentation by making the physician myopic. The results are similar with the previous experiment. Even though the expected per-period utility of the patient is smaller than in the baseline, treatment outcomes, costs and market shares remain almost the same as in the estimated benchmark case. These results suggest that the disutility caused by uncertainty affects the medication choices of the physician more than the information gains of experimentation. This indicates that experimentation incentives do not matter much in the market for cholesterol drugs.

Then, I decrease the risk aversion parameter from 0.98 to 0.50. When the physician becomes less risk averse, the number of different drugs increases 27%, the process of learning improves and the patient's expected welfare increases from the estimated baseline scenario. Even though the drug therapy length does not change, the total costs of the statin therapy increase by 53%. To explain this, the lower risk aversion coefficient decreases the relative differences in sub-utilities $-e^{-r\underline{\mu}_{nk}+\frac{r^2}{2}(\underline{\sigma}_n^2+\sigma_{ex}^2)}$ between drugs and thus preference shocks e_{jknt} affect prescriptions more.

In the last simulation experiment, the physician does not observe health effects and thus he cannot learn. The expected discounted utility decreases because the variance of the symptomatic match value does not decrease over time. The market share of Simvastatin increases slightly because it has relatively good performance for both patient types. The total costs of the statin therapy and the average therapy length remain almost the same as in the benchmark case.

Then, I study in detail how incentives to experiment vary across policy experiments. Figure 2.5 presents the simulated probability of switching a drug treatment from the previous prescription, conditional on the number of prescriptions.[40] As suggested by Table 2.3, the probability of switching is higher in the experiment without learning than in the benchmark case. This is because uncertainty and learning make risk-averse buyers reluctant to try new, less well-known products. When the physician is myopic, the probability of switching is almost the same than in the baseline scenario.

[40]Note that the sample of patients varies with the prescription number.

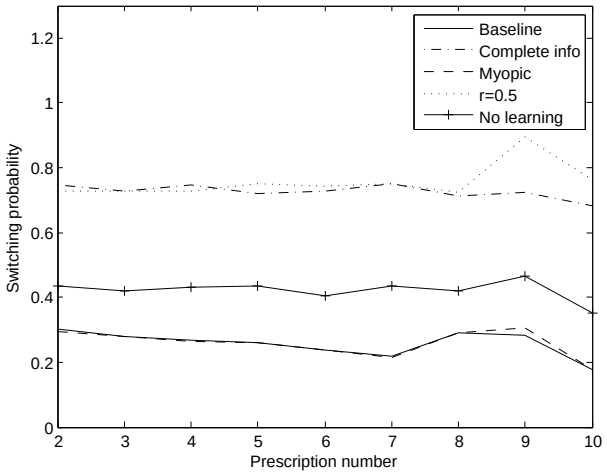

Figure 2.5: The probability of switching an active ingredient from the previous prescription, an average over patients, products and types in the sample

Finally, Figure 2.6 presents the market share of the market leader and the cheapest drug, Simvastatin, in different policy experiments. The results suggest that the average market share decreases over the course of the therapy. This may happen since those patients who are on the statin therapy long have the worst symptomatic match with Simvastatin. When the number of prescriptions is high, the estimated market share does not differ much from the market share of the policy with complete information since physicians have probably learned the average health effects of Simvastatin. The probability of prescribing Simvastatin is the highest, around 60%, in the experiment without learning.

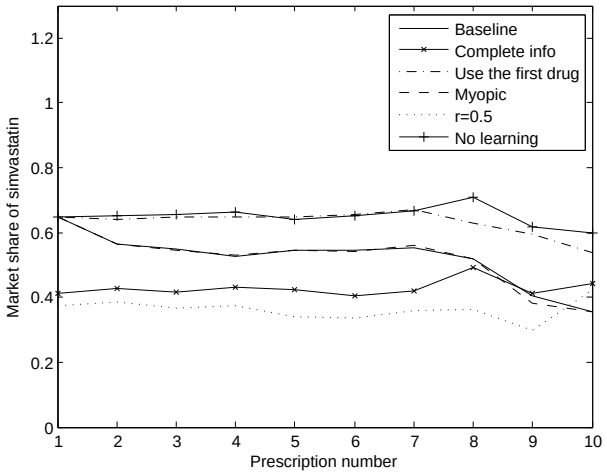

Figure 2.6: The market share of Simvastatin in each period, an average over patients, products and types in the sample

Overall, my results suggest that promoting in the process of learning can significantly improve medical decision-making. This finding implies that policy makers should provide information about the health effects of drugs. Consistent with high persistence in demand for cholesterol drugs, I find that physicians are not willing to experiment with new, less well-known treatment alternatives. If physicians became less risk averse, their learning would improve and the welfare of a patient would increase. To achieve these goals, one might imagine that treatment recommendations could encourage doctors to more risks in their treatment choices.

2.7 Conclusions

I analyzed the role of experimentation in demand for pharmaceuticals. I estimated a dynamic matching model of medical decision-making that incorporates uncertainty and learning about the average health effects of pharmaceuticals. After the patient has used a drug, the physician observes how the drug treatment affected both the patient's symptoms and the probability of ending the drug therapy. The structural model was estimated using rich data from the Finnish market for cholesterol drugs.

The parameter estimates implied that patients respond differently to cholesterol drug treatments. I also found that physicians are risk averse and face substantial uncertainty about the health effects of statins. These findings suggested that information and learning may have significant value in this market. The counterfactuals showed that the provision of information on the average health effects of drug treatments increases the patient's welfare from the estimated benchmark case. I also found that uncertainty affects medical decision-making more than the information gains of experimentation. If doctors became willing to take more risks in their treatment choices, the process of learning would improve.

A key assumptions of the model is that physicians are identical. In the next chapter, I analyze how physician's own experience of a patient and learning from past choices of other doctors affect her behavior.

Bibliography

[1] Ackerberg, D. (2003) Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination. *International Economic Review,* Vol. 44, pp. 1007-1040.

[2] Aguirregabiria, V. - Mira, P. (2010) Dynamic Discrete Choice Structural Models: A Survey. *Journal of Econometrics*, Vol. 156, pp. 38-67.

[3] Cameron, A. - Trivedi, P. (2005) Microeconometrics Methods and Applications. [41], New York, USA.

[4] Chan, T. - Hamilton, B. (2006) Learning, Private information, and the Economic Evaluation of Randomized Experiments. *Journal of Political Economy,* Vol. 114, pp. 997-1040.

[5] Chernew, M. – Gowrisankaran, G. – Scanlon, D. (2008) Learning and the Value of Information: Evidence from Health Plan Report Cards. *Journal of Econometrics*, Vol. 144, pp. 156-174.

[6] Chetty, R. (2006) A New Method of Estimating Risk Aversion. *American Economic Review*, Vol. 96, pp. 1821-1834.

[7] Ching, A. (2009) A Dynamic Oligopoly Structural Model for the Prescription Drug Market after Patent Expiration. Forthcoming in *International Economic Review.*

[8] Cohen, A. - Einav, L. (2007) Estimating Risk Preferences from Deductible Choice. *American Economic Review*, Vol. 97, pp. 745-788.

[9] Conventional Treatment Medication (2009) Cholesterol and Blood's Other Lipid's Perturbations (Dyslipidemia).

[41] Cambridge University Press

[10] Coscelli, A. (2000) The Importance of Doctors' and Patients' Preferences in the Prescription Decision. *The Journal of Industrial Economics*, Vol. 48, pp. 349-369.

[11] Coscelli, A. – Shum, M. (2004) An Empirical Model of Learning and Patient Spillovers in New Drug Entry. *Journal of Econometrics*, Vol. 122, pp. 213-246.

[12] Crawford, G. S. – Shum, M. (2005) Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, Vol. 73, pp. 1137-1173.

[13] Dickstein, M. (2011) Efficient Provision of Experience Goods: Evidence from Antidepressant Choice. Job Market Paper, Harvard University.

[14] Hellerstein, J. K. (1998) The Importance of the Physician in the Generic Versus Trade-Name Prescription Decision. *The RAND Journal of Economics*, Vol. 29, pp. 108-136.

[15] Iizuka, T. (2007) Experts' Agency Problems: Evidence from the Prescription Drug Market in Japan. *RAND Journal of Economics,* Vol. 38, pp. 844-862.

[16] Iizuka, T. (2011) Physician Agency and Adoption of Generic Pharmaceuticals, *American Economic Review*, 2011, accepted.

[17] Johnston, D. (1999) Lifestyle Changes After a Myocardial Infarction. Editorial in *Heart,* Vol. 82, pp. 543-544.

[18] Jousilahti, P. (2004) Current Care for Dyslipidemias is Renewed (in Finnish: Dyslipidemioiden Käypä hoito uudistuu). Kansanterveys-lehti, National Institute for Health and Welfare.

[19] Kansanterveyslaitos (2009) Sydän- ja verisuonisairaudet. http://www.ktl.fi/portal/suomi/tietoa_terveydesta/terveys_ja_sairaudet/sydan-_ja_verisuonisairaudet

[20] Keane, M. – Wolpin, K. (1994) The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation:Monte Carlo Evidence, *Review of Economics and Statistics,* Vol. 76, pp. 648672.

[21] Kela (2008) Lääkekorvaukset ja -kustannuksen lääkeryhmittäin. Lääkekorvaustilastot.

[22] Klaukka, T. – Helin-Salmivaara, A. – Huupponen, R. - Idänpään-Heikkilä, J. (2007) Statiinien käyttäjämäärä kasvaa mutta kustannukset ovat vähentyneet. *Suomen Lääkärilehti*, Vol. 62, pp. 3188-3191.

[23] Klaukka, T. – Helin-Salmivaara, A. – Huupponen, R. - Idänpään-Heikkilä, J. (2006a) Kolmannes yli 65-vuotiaista käyttää statiineja. *Suomen Lääkärilehti*, Vol. 35, pp. 3472-3474.

[24] Klaukka, T. – Helin-Salmivaara, A. – Huupponen, R. - Idänpään-Heikkilä, J. (2006b) Simvastatiinin asema vahvistui vuonna 2005. *Suomen Lääkärilehti*, Vol. 61, pp. 1998-2001.

[25] Klaukka, T. – Helin-Salmivaara, A. – Huupponen, R. - Idänpään-Heikkilä, J. (2006c) Statiinit valtimotaudin ehkäisyssä ja hoidossa. *Suomen Lääkärilehti*, Vol. 61, pp. 4228-4231.

[26] Kim, M. (2010) Physician Learning and New Drug Diffusion. Job Market Paper, University of Chicago.

[27] Lundin, D. (2000) Moral Hazard in Physician Prescription Behavior. *Journal of Health Economics*, Vol. 19, pp. 639ä662

[28] National Center for Health Statistics (2011) Health, United States, 2011: With Special Feature on Socioeconomic Status and Health. Hyattsville, MD. 2012.

[29] National Institute for Health and Clinical Excellence (2006) Statins for the Prevention of Cardiovascular Events. NICE Technology Appraisal 94, pp. 1-45.

[30] Narayanan, S. – Manchanda, P. (2009) Heterogeneous Learning and the Targeting of Marketing Communication for New Products. *Marketing Science,* Vol. 28, pp. 424-441.

[31] Peltonen, M. – Harald, K. – Männistä, S. – Saarikoski, L. – Peltomäki, P. – Lund, L. – Sundvall, J. – Juolevi, A. – Laatikainen, T. – Aldän-Nieminen, H. – Luoto, R. – Jousilahti, P. – Salomaa, V. – Taimi, M. – Vartiainen, E. (2008) Kansallinen FINRISKI 2007 -terveystutkimus. Tutkimuksen toteutus ja tulokset. *Kansanterveyslaitoksen julkaisuja,* B34/2008, pp. 1-72.

[32] Phelps, C. (2000) Information Diffusion and Best Practice Adoption. *Handbook of Health Economics Volume 1*, Part 1, pp. 223-264.

[33] Rust, J. (1996) Numerical Dynamic Programming in Economics. *Handbook of Computational Economics,* ed. by D. K. H. Amman and J. Rust. Amsterdam: North-Holland, Chapter 14.

[34] Rust, J. (1987) Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, Vol. 55, pp. 999-1033.

[35] Scott, A. (2000) Economics of General Practice. Chapter 22 in *Handbook of Health Economics,* Volume 1, Edited by A.J. Culyer and J.P. Newhouse, 2000, Elsevier Science B.V.

Chapter 3

Private Experience and Observational Learning in Pharmaceutical Demand

I quantify the roles of the physician's own experience and the past choices of other doctors in pharmaceutical demand. I develop a model of medical decision-making under uncertainty about the quality of the match between the patient and the drug treatment. Unlike previous demand models, I take into account both private and social learning, and allow heterogeneity in quality across patients. I test whether information on the past choices of other physicians improves drug choices. Using rich data from the market for cholesterol drugs, I show that treatment patterns relying heavily on the past choices of other doctors can lead to demand beyond the efficient level. My results suggest that continuity of care, where a patient repeatedly consulting the same doctor, is an efficient policy to limit such behavior.

Keywords: social and private learning, structural modeling, unobserved quality, asymmetric information, demand, information diffusion, physician behavior

3.1 Introduction

Agents may use their own experiences and the past choices of their peers to learn about the quality of a product. Yet there is very little previous work quantifying whether private learning and observing the behavior of other agents help to reduce uncertainty around choices. In this chapter, I explore these issues in pharmaceutical demand under uncertainty about the effects of the drug treatment on patient health. I will show that that treatment patterns relying heavily on the past choices of other doctors can lead to over-prescribing in terms of welfare. I analyze whether continuity of care - in the sense of a patient repeatedly consulting the same doctor - is an efficient policy to limit such behavior. The policy is commonly used in primary care to promote the process of learning and improve medical decision-making:

> However, there are other aspects to the doctor-patient relationship that have important implications on efficiency. The distinctive feature of general practice agency is that the doctor-patient relationship is usually long-term and more likely to be characterized by repeated transactions [...] In general practice repeated transactions are also potentially beneficial because the GP becomes more aware of the context of the patients' health problems, and has more information about the patients' medical history, social circumstances, values and preferences.
> Anthony Scott (2000), *Handbook of Health Economics*

I develop a model of medical decision-making under uncertainty about the quality of the match between the patient and the drug treatment. I modify the standard theoretical models with social learning (Chamley, 2004, Bikhchandani, Hirshleifer and Welch, 1992), by allowing agents to learn product quality from their own experiences. I ask whether continuity of care is preferable to providing information on the past behavior of other doctors through patient records. I focus on the Finnish market for cholesterol drugs that are used to decrease the risk for cardiovascular diseases. Benefits from improvements in the drug treatment of hyper-cholesterolemia (high cholesterol) can be substantial, as heart disease and stroke alone are among the most widespread and costly diseases. Still, many doctors claim that cholesterol drugs are prescribed to low-risk patients beyond the level of clinical effectiveness.[1]

Empirical evidence shows that private experience and peer effects have important roles in demand for pharmaceuticals. First, the extensive literature in medicine and economics

[1] See e.g. Franklin (2011), Adams (2011), Joelving (2011), BBC (2011).

(see e.g. Weiss and Blustein, 1996, Scott, 2000, King et al., 2008) has documented the positive relationship between continuity of care and treatment outcomes. My data confirms that prescriptions are highly responsive to changes in the length of the doctor-patient relationship.[2] Second, prescription behavior by inexperienced physicians is significantly affected by the choices of prominent physicians, or "opinion leaders" (Nair et al., 2010). In my data, the previous choices of peers affect prescribing behavior especially if a physician does not have much own experience of the patient.

To model these findings, I develop my analysis as follows. I consider a physician's (she) decision to continue the patient's (he) drug therapy in primary care where physicians may change.[3] The physician does not know ex-ante the efficacy and side effects of cholesterol drugs (referred as the "drug") for a patient. I analyze the physician's attempt to learn the quality match between the patient and the drug from her own experience and the past choices of other doctors. At the beginning, the physician evaluates the patient's risk for cardiovascular diseases based on his observed characteristics, such as gender and age. This evaluation forms the prior belief of physicians on the average health effects, or quality, of the cholesterol drug for the patient. In the follow-up, a physician performs diagnostic procedures and medical tests to evaluate whether the drug affected the patient's abnormal cholesterol levels and caused side effects. While interpreting her findings, the physician privately observes the health effects of cholesterol drugs that affect her personal opinion (private belief) on quality. Simultaneously, she looks at patient records to see the past choices of other doctors. With this information, the physician forms the public belief on quality. The physician takes into account her own opinion, the patient's prescription history and the prior belief when she decides on the continuation of the patient's therapy.

The model helps to understand whether continuity of care can improve the efficiency of drug choices. When only one physician is treating a patient, the physician becomes over time more familiar with the patient's disease and her perceptions on the distribution of health effects become more precise. She may thus learn whether the drug is on average good or bad for the patient which improves her medical decision-making. If physicians change frequently, the process of learning slows down and becomes heavily dependent on the past choices of other doctors. An inexperienced physician may belief that the drug treatment must perform well for the patient who has used the drug for many years. This

[2]Specifically, I consider the choices of physicians working in the Finnish public primary care. In this market, the physicians of a patient change frequently for exogenous reasons, such as due to the shortage of physician labor. See section 3.2.1 for details.

[3]The model can be extended to allow multiple inside goods. This is very straightforward if the health effects of only one drug group, say patented products, are uncertain.

optimism on quality leads to over-prescribing when the drug is of low quality.

A vast majority of the literature on demand for experience goods assumes that agents can only learn the quality of a product from their own experience (e.g. Crawford and Shum, 2005, Kim, 2010, Dickstein, 2011, Chan and Hamilton, 2006, Chernew et al. 2008) or that all information is public (e.g. Ackerberg, 2003, Ching, 2009). A few recent papers also look at the social learning of an agent who makes a once-in-a-lifetime decision (Cipriani and Guarino, 2012, Knight and Schiff, 2010, Zhang, 2010). My main contribution is that I take into account both private and social learning in demand, allowing agents to receive multiple experience signals. With my framework, I can analyze how the own consumption experiences of an agent interact with information received from the past choices of peers in her learning process.[4] Furthermore, because private and social learning may induce divergent beliefs about quality, a demand model should capture them both in order to produce reliable estimates on product quality and on the effects of policy experiments on choices.[5] Finally, unlike the previous work on social learning, I allow heterogeneity (among patients) in quality.

I find that the average health effects of the cholesterol drug treatment are heterogeneous across patients. Particularly, the quality of the match is on average high for 72% of patients and low for the remainder. The estimates also imply that most of the uncertainty associated with quality vanishes when the patient has used the cholesterol drug treatment once. Even if quality was known, uncertainty regarding to health effects remains significant. These results have implications on efficiency.

The counterfactual experiments suggest that information on the patient's prescription history does not compensate for the lack of the long-term treatment relationship. If the patient had only one physician, the physician learns fast and better health outcomes realize. If quality is high (low), the long-term doctor-patient relationship increases (decreases) demand for cholesterol drugs. Information on the past choices of other doctors for a patient promotes learning about high quality, but not as efficiently as continuity of care. If quality is low, observing the patient's prescription history increases demand over the level of efficient prescribing.

[4] Traditional private and social learning models are special cases of my framework.

[5] If there is private information unobserved by the econometrician, but all information is assumed to be public, quality estimates become biased. Specifically, when quality is in reality high, quality estimate is downwards biased because private information slows down learning and decreases the probability of choosing the product. Low quality estimate is, on the other hand, upwards biased because social learning makes agents too optimistic about quality which increases the probability of choosing the product.

The rest of the chapter is organized as follows. Section 3.2 describes the dataset and provides descriptive evidence on the effects of physician's own experience and the past choices of other doctors on medical decision-making. Section 3.3 goes through the structural model and Section 3.4 discusses estimation and identification. Section 3.5 presents estimation results, the fit of the model and the results from the counterfactual experiments. Section 3.6 concludes.

3.2 Market and data description

3.2.1 Cholesterol drug markets

Cardiovascular diseases (CVD), such as heart attacks, stroke and high blood pressure, affect millions of people globally. Heart disease and stroke alone are among the most common and costly health problems in Europe and the United States.[6] Patients who have experienced CVDs have to deal with high medical expenditures, lost wages and lower productivity.

I analyze the Finnish market for cholesterol drugs that are used to decrease the risk for cardiovascular events. I focus on statins (HMG-CoA reductase inhibitors) that is the most popular group of cholesterol drugs globally.[7] Statins decrease high serum LDL-cholesterol ("bad" cholesterol) and increase HDL-cholesterol ("good" cholesterol) by inhibiting an enzyme in the liver that has an important role in the production of cholesterol.[8] High morbidity to CVDs and a large volume of diagnoses of dyslipidemia, i.e. an abnormal amount of lipids, such as cholesterol and fat, in the blood, have made cholesterol drugs one of the world's largest selling drug groups.

Corresponding to the United States, the following active ingredients are on the Finnish statin market: Atorvastatin (Lipitor and Torvast), Fluvastatin (Lescol), Lovastatin (Mevacor, Altocor, Altoprev), Pravastatin (Pravachol, Selektine, Lipostat), Rosuvastatin (Crestor)

[6] Around 12% of adults suffered from heart disease in 2009 – 2010 in the United States (National Center for Health Statistics, 2011). Every year, there are around 152 000 strokes in the UK (British Heart Foundation, 2013).

[7] See e.g. Herper, M. (2010) *"Why You May Need Cholesterol Drugs"*, Forbes, and U.S. Food and Drug Administration (FDA), 2010.

[8] When cholesterol levels are too high, cholesterol can grow on the walls of blood vessels transporting blood from the heart to other body parts. Over time, these blood vessels can be blocked, preventing the heart from getting enough blood.See e.g. "What is cholesterol?" by the National Heart, Lung and Blood Institute that is a division of the National Institutes of Health in the USA.

and Simvastatin (Zocor, Lipex).[9] I focus on a physician's decision to continue the patient's statin therapy for several reasons. First, uncertainty is probably the highest in the health effects of statins in general. Second, clinical differences between statins in reducing cardiovascular events have been claimed to be small (National Institute for Health and Clinical Excellence, 2006) and thus it is quite natural to consider statins as a one group. I thereby ignore important questions regarding to a physician's or patient's choice between branded and generic products (see e.g. Scott-Morton, 1999, Ching, 2010a and 2010b) and between different active ingredients (see Crawford and Shum, 2005).[10]

A treatment decision by a physician is based on the benefits and adverse effects of statins. The statin therapy is initiated if the patient has a high risk for CVDs. The evaluation of the risk is based on several factors, including the patient's gender, age, blood pressure and cholesterol levels. In my model, the initial evaluation is captured by the physician's prior belief on the average health effect of cholesterol drugs for a particular patient. In the follow-up of the drug therapy, a physician evaluates the realization of the treatment goals and sustains the patient's treatment motivation. The main goal of cholesterol drug treatment is to decrease the total cholesterol level below 5 mmol/L (LDL-cholesterol below 3 mmol/L). If the patient experiences side effects, the physician decreases the dosage, experiments with an another statin or suspends the cholesterol drug therapy (the Finnish current care for dyslipidemia, 2011).[11] As patients respond differently to statins (the Finnish current care for dyslipidemia, 2011, Jousilahti, 2004), a physician may not know the efficacy and side effects for a single patient.[12] I take the uncertainty into account and

[9] Within the group of an active ingredient, statins differ also in the form of drugs, package sizes, strengths and prices. I do not consider a combination preparations of a statin and an another active ingredient.

[10] I also assume that the physician decides to end the patient's medical treatment. In practice, the final decision to end the therapy can be done either by the physician or the patient or both.

[11] Lifestyle changes, including exercising and changes in diet, are often adequate for a low-risk patient. However, patients are often unwilling to change their lifestyles, even after having a significant shock in their life. Perhaps 45% of smokers stop smoking after a myocardial infarction which is between 2 or 4 times of the success rate of antismoking clinics. Results are not as good for other cardiovascular risk factors related lifestyle, such as physical exercise or diet. Patients can become even less active after infarction. There is also some evidence that changes in self-reported fat intake in one year after infarction can be small. (Johnston, 1999)

[12] For example, statins are reported being useful for men, post menopausal women and patients who have arterial disease or diabetes. It has also been shown that statins decrease by 15% the mortality rate of patients who were 60 years and older and initially clinically asymptomatic. Genetic susceptibility and certain drug interactions can increase the risk of side effects. For example, approximately 5% of patients have been reported suffering muscular symptoms and an increase in the activity of serum muscular enzymes appears for $0.5 - 2.0\%$ of statin users, even though its clinical significance is often uncertain.

let the physician to learn the average health effects of statins by observing realized health effects and the patient's past statin prescriptions.

Cholesterol drugs are also particularly interesting as there is no consensus on an appropriate level of cholesterol drug prescribing. Some doctors have claimed that there is a little evidence that statins reduce the CVDs of low-risk individuals. Doctors supporting the use of statins have said that they have prevented heart attacks and other CVDs.[13] In my model, physicians disagree on the health effects of statins, depending on their personal experience of the patient.

Two features of the Finnish market simplify my empirical analysis. The first is that a choice of a physician by a patient was very restricted in public primary care. During the observation period, the patient was not allowed to choose the health center. Within the health center, the patient's family physician was (exogenously) determined based on the patient's residential area (Finnish Medical Association, FMA, 2007).[14] However, due to the shortage of physician labor, patients were not often treated by their own family physicians.[15] I assume that a physician is exogenously determined for the patient in primary care.[16]

The second feature is that two characteristics of the Finnish statin market decrease variation in drug prices over time. First, drugs are subject to price cap regulation by the Pharmaceuticals Pricing Board that is subordinated to the Ministry of Social Affairs and Health in Finland. Second, the patents of Fluvastatin, Atorvastatin and Rosuvastatin remained effective during the whole observation period $2003 - 2006$. As patent protection limits competition, it is likely that the prices ceilings of the patented products were binding. In the empirical analysis, I follow much of the previous learning literature (e.g. Crawford and Shum, 2005) and assume that the drug prices are exogenous. The assumption simplifies the construction of the structural model as prices do not adjust with the

(The Finnish current care for dyslipidemia, 2011)

[13]See e.g. Adams (2011), Joelving (2011), BBC (2011).

[14]Family physician practices are widely adopted in many countries. For example in the USA, The American Academy of Family Physicians (AAFP) is one of the largest national medical organizations. See AAFP, http://www.aafp.org.

[15]For example in 2006, 9% of the appointments in health centers had a shortfall of physicians and almost the same share of working-age physicians were absent from their permanent jobs. In 46% of these cases, this was caused by staying abroad (FMA, 2006c). It has been estimated that 90% of family physicians treat other than their own patients every week (see FMA, 2005, 2006a, 2006c, 2007).

[16]To be more specific, I assume that the probability of getting a certain physician does not depend on the statin treatment or the health of the patient. This probability is needed to recover the choice probability for the outside good.

observed behavior of physicians.[17]

3.2.2 Information transmission between physicians

In the model, I assume that a physician has personal experience about the patient-specific quality of the drug treatment. As MD Epstein (1999) illustrates in *the Journal of the American Medical Association*: "Clinical judgment is based on both explicit and tacit knowledge. Medical decision-making, however, is often presented only as a conscious application to the patient's problem of explicitly defined rules and objectively verifiable data. [...] Seasoned practitioners also apply to their practice a large body of knowledge, skills, values, and experiences that are not explicitly stated by or known to them. [...] While explicit elements of practice are taught formally, tacit elements are usually learned during observation and practice." In this section, I evaluate the validity of the assumption on private information further by discussing the information content of patient records and communication between physicians.

Patient records

A patient record documents and transfers information on a single patient's medication between physician. If all relevant information for medical decision-making is available in the record, a physician does not have any private information of the patient. To see whether this is the case, I next consider the information content of patient records.

The focus of patient records is on the patient's medical condition and medication.[18] To see what type of information is stored in patient records, consider an example of a patient record for a dispensary admission in Appendix B. The patient record provides a compact description of the patient's health status and the plan, the goal and the follow-up of the treatment. It also includes the name of the physician, the list of current medication and a brief justification for starting a medical treatment. In general, patient records may also

[17] In the financial market application of Cipriani and Guarino (2012), bid and ask prices (prices at which a trader can buy and sell) are endogenous because they reflect public information containing the history of trades and prices.

[18] Patient records regarding to medication include entries about the need of pharmacotherapy and medical foundations, a prescription and given medical treatment, including the name, quantity, form, dosage, dosage form, the date and time of issue of a drug and the name of the physician who has given or prescribed the drug (The Ministry of Social Affairs and Health, 2005).

contain information on whether medication is permanent and reasons for a physician's decision to end the patient's drug therapy.[19]

Patient records do not perfectly transfer all relevant information for medical decision-making between physicians. The case example demonstrates that the continuation of drug therapy is not justified (Appendix B). According to an interviewed specialist, this is a very common practice, at least in routine cases. Records do not include physician-specific factors, such as the physician's own preferences for medication and information on whether her medical decision-making is based on medical literature, advertising and treatment recommendations. The physician's accumulated knowledge of the patient's preferences, values and circumstances is rarely recorded (see Guthrie et al., 2008). The specialist also claimed that a narrative text format complicates the interpretation of records that may impede information transmission. The registering of information takes the physician's time that may decrease her incentives to record all relevant information.

Communication

I evaluate next whether all relevant information for medical decision-making is transfered through communication. A physician who cares about her patient may want to consult her colleagues before deciding on the continuation of the treatment. Because communication is time-consuming, consultation does not probably happen in routine cases. On the other hand, the patient, who wants to get as good medical treatment as possible, may want to communicate all relevant information to her physicians. It is, however, unlikely that medical decision-making by physicians is exclusively based on information received from the patient (see e.g. Epstein, 1999).

The theoretical cheap-talk[20] literature (see for example Crawford and Sobel, 1982, Olszewski, 2004) has shown that the truthful information revelation of a consultant (a sender, here: other physicians or a patient) to a decision maker (a receiver, here: a physician) is only one of many possible outcomes, even if there is no disagreement between participants. If the preferences of the consultant are even slightly misalligned with the preferences of the decision maker, there is some information loss in all equilibria (Craw-

[19]Essential information in electronic patient documents are reported in the following guidebook and its updated versions (in Finnish): "Opas Ydintietojen, otsikoiden ja näkymien toteuttaminen sähköisessä potilaskertomuksessa", version 1.1, 28.2.2006.

[20]In a typical cheap-talk game, the sender may, often costlessly, convey her private information through messages to the receiver. The receiver then takes an action that together with sender's signal affects the payoffs of both players.

ford and Sobel, 1982). If the consultation effort of the physician is unobserved to the patient, incentives for consultation may not be high.

Finally, if all physicians of a patient share the same information, they should have the same probability of choosing the medical treatment. As it turns out in the next section, this is not the case.

3.2.3 Data

Sample selection

I use a rich dataset of all purchased cholesterol drug prescriptions in Finland from January 1 in 2003 to December 31 in 2006. The data is provided by the Social Insurance Institution of Finland which is responsible for the provision of public social security benefits to Finnish residents. The data identifies patients, their physicians and cholesterol drugs.[21]

I prepare my data for the empirical analysis in the following steps. First, to follow patients from the beginning of cholesterol drug therapy and to avoid left-censoring, I focus on "new" patients who did not have any prescriptions during the first 6 months of the observation period i.e., before July 2003.[22] Second, I ignore patients with multiple prescriptions or physicians within a day to simplify the analysis further. Third, I consider patients whose physicians are primarily working in public health centers. Ideally, I would like to concentrate on patients who have only used the services of public health centers but unfortunately the data does not include this information. As a proportion of physicians work for both the public and the private sectors[23], some patients in the sample may have used private health care services. Fourth, I concentrate on patients who belong to the working-age (15-64 years) population because the data does not allow me to distinguish the death of a patient from the ending of the statin treatment. Finally, for computational reasons, I draw a random sample of 10000 patients from the sample of new working-age patients whose physicians are working in primary care.

[21] Other characteristics than the primary job of a physician (public health center/public hospital/other) received from the survey conducted by the Finnish Medical Association (FMA) are from the registers of the Social Insurance Institution of Finland. The response rate of the yearly survey has been very high. For example, in 2006, the response rate of physicians who received the survey was 80% (FMA, 2006c).

[22] This six months' time window has been also used by Crawford and Shum (2005).

[23] In 2006, 19.6% of physicians, who were primarily working in health centers, had a sideline job (FMA, 2006c).

Descriptive evidence

In this section, I provide the descriptive analysis of the sample. The results in Table 3.1 demonstrate that the sample consists of very heterogeneous patients. Most of the patients in my sample were relatively old at the time of the last prescription (an average 51 years) and almost half of the patients were men. The number of diagnosis varies[24] in substantially around its mean (0.7).[25] A significant portion of patients (55%) were censored in the sample i.e., they had their last prescription within the last six months of the observation period.

[24]The number of diagnosis is observed if the patient was on sick-leave.

[25]Information on the number of diagnosis is observed if a patient received sickness benefits from the Social Insurance Institution of Finland.

Table 3.1: Descriptive statistics for the sample of patients[1]

	Mean	Std.Dev.	Only non-censored patients	At the time of the last prescription
Patient characteristics				
Age	55.03	7.20	No	Yes
Gender (1: male, 0: female)	0.49	0.50	No	Yes
Nbr of diagnosis	0.73	1.31	No	Yes
Censoring indicator (1: yes, 0: no)	0.52	0.50	No	Yes
Patient's medical treatment				
Treatment ending (1: yes, 0: no)	0.34	0.47	Yes	No
Nbr of prescriptions	1.93	1.17	Yes	Yes
Nbr of physicians	1.28	0.58	Yes	Yes
Prescriptions of a current physician	1.676	1.072	No	No
Visit a physician specialized in internal diseases	0.01	0.09	No	No
Visit a non-specialized physician	0.69	0.46	No	No
Total number of physician's prescriptions	1.65	1.07	No	No
Physician change (1: yes, 0: no)[2]	0.33	0.47	No	No
Active ingredient change (1: yes, 0: no)[2]	0.17	0.38	No	No
Price, eur	46.32	49.16	No	No
Number of observations	22 021			

[1] The relevant population consists of new working-age patients who have used statins and the services of public health centers. The size of the random sample is 10 000 patients.

[2] Note that here the number of prescriptions is at least 2 because the change in the value of the variable from the previous prescription is computed by using the difference between its current and lagged value.

Following Crawford and Shum (2005), I assume that the drug therapy of a non-censored patient ends after the last prescription in the data. If the patient is censored, the end of the therapy is not observed. If the censoring interval is too short, the estimation results may be biased. This is particularly true if the patient's drug treatment is prescribed at the end of the observation period and he has more than two prescriptions.[26] Dickstein

[26] As a robustness check, I used a one-year censoring interval and defined a patient to be "new" if he did not have prescriptions during the first year. Then, the probability that the patient is censored

(2011) used an alternative approach where the treatment episode of a patient ends at the last prescription if there was a gap of 90 days within the treatment history. A patient appearing in the data again after the gap is then treated as a new patient.

The cholesterol drug therapies of non-censored patients in the sample were on average relatively short, approximately 2 prescriptions (Table 3.1). The probability that the patient's therapy ends at any stage of therapy is 0.34. The average number of physicians per patient was 1.3 and the total number of prescriptions received from a particular physician was 1.65. Most of the patients (70%) were treated by a non-specialized physician. The average price of a prescription was 41 eur.

Table 3.2 presents the distribution of the total number of prescriptions and physicians at the time of the (non-censored) patient's last prescription. Most of the non-censored patients (52%) had only one prescription and 80% of the patients were in a permanent physician-patient relationship. Even though the distributions of the total number of prescriptions and physicians are skewed to the right, 48% of non-censored patients had more than one prescription and 20% were treated by more than one physician.

Table 3.2: The percentage share of non-censored patients in the sample conditional on the total number of prescriptions and physicians at the last prescription

	Physicians			
Prescriptions	1	2	3-	Total
1	51.91	.	.	51.91
2	18.55	8.37	.	26.93
3	6.77	4.80	1.45	13.02
4-	3.13	2.95	2.07	8.15
Total	80.36	16.12	3.52	100.00

I consider next the incidence of a physician change in the sample of patients. Table 3.1 illustrated that the breakdown of the physician-patient relationship was very common. The probability that the patient's physician changes from the previous prescription was 33%. A high standard deviation also indicates significant diversity among patients in the incidence of a physician change.

Then, I analyze how the number of interactions between a physician and a patient affects prescriptions. I consider first how the probability of continuing the (non-censored) patient's statin therapy depends on the lagged number of physicians (Figure 3.1). I find that

was somewhat higher (0.73) than with the original censoring interval. The probability that the patient's treatment ends was 0.40 which is fairly close to the corresponding probability with other definition (0.34).

the continuation probability is 50% for patients who have only one physician, i.e. who do not have any physician switches. The choice probability decreases to 42% for patients having two physicians and further to 33% for patient with three physicians.

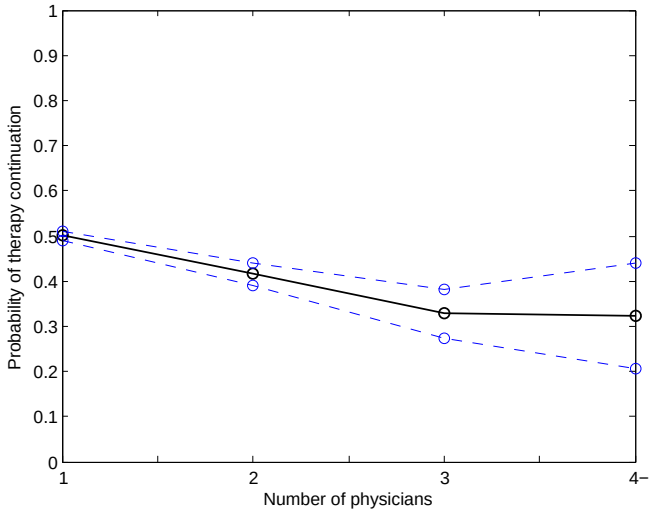

Figure 3.1: The probability of treatment continuation and its 95% confidence intervals by the number of physicians for non-censored patients, sample averages

I investigate next whether the decreasing pattern between the choice probability and the number of physicians is driven by the phase of the patient's therapy. To see if this is the case, I estimate the following linear probability model for the continuation of the (non-censored) patient's statin therapy,

$$\mathrm{a}_{it} = \alpha + X_{i(t-1)}\beta + e_{it},\ t > 1,$$

where a_{it} is an indicator variable that gets value 1 if the statin therapy of patient i is continued at time, or prescription, t and 0 otherwise[27], $X_{i(t-1)}$ is a vector of lagged explanatory variables and e_{it} is the error term.

The results presented in Table 3.3 suggest that the continuation probability increases by 13% when the number of previous physicians increases by one. The lagged length of the

[27] To be more precise, $a_{it} = 0$ only once when the patient's statin therapy ends.

doctor-patient relationship has an opposite effect on the continuation probability. These findings may suggest that physicians do not share the same information about the health effects of the cholesterol drug treatment for a patient.

Table 3.3: Descriptive regressions for the probability of therapy continuation in the sample of non-censored patients

Variable[1]	Model (1)	Model (2)
Constant	0.672***	0.704***
	(0.167)	(0.169)
Own experience:		
prescriptions/current physician		-0.126***
		(0.013)
Nbr of physicians	0.129***	
	(0.017)	
Prescription nbr	-0.156***	-0.046***
	(0.008)	(0.010)
Gender	0.0288**	0.0277*
	(0.0109)	(0.0109)
Age	0.000	0.000
	(0.001)	(0.001)
Nbr of diagnosis	0.006	0.006
	(0.004)	(0.004)
Cost, eur	0.000***	0.000***
	(0.000)	(0.000)
Reimbursement	-0.000***	-0.000***
	(0.000)	(0.000)
Fixed effects:		
physician, ATC-code, hospital district	yes	yes
N	10031	10031
adj. R^2	0.093	0.100

[1] Explanatory variables are lagged by a one prescription.

[2] Variables are for cholesterol drug prescriptions.

[2] Standard errors in parentheses.

[3] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To get further evidence on peer effects and the role of private experience in demand, Table 3.4 illustrates how medical spending in the sample depends on the length of the physician-patient relationship, after controlling for observed characteristics. When the number of

physicians increases by one, the total costs of the therapy at any stage decreases by 7 euros which is 15% of the average costs of statins in the sample. Table 3.4 also shows that the more the physician has experience of the patient, the less the previous choices of peers - measured by the number of cholesterol drug prescriptions provided by other doctors to a single patient - affect an average medical spending at any phase of the therapy.[28] When the physicians of a patient change frequently relative to the stage of the drug therapy, the effect of physician's own experience on the total costs becomes small. These results are consistent with "asymmetric peer effects" where inexperienced physicians rely on experienced doctors to decrease uncertainty around their prescription decisions (see e.g. Nair et al., 2010). Still, the findings remain very indicative without putting any structure in the model that helps to isolate the effects of personal experience and social learning on medical decision-making.

[28] I measure the physician's own experience with the number of interactions with the patient.

Table 3.4: Descriptive regressions for treatment costs in the sample of patients

Explained variable	Total cost,	Total cost, eur[1]	Cost, eur	Total cost, eur[1]
Constant	-92.48***	-102.5***	12.52***	-168.2***
	(17.54)	(18.04)	(2.173)	(23.10)
Nbr of physicians	-7.106*			
	(3.312)			
Own experience:				
prescriptions/current physician		60.63***	0.549***	23.25***
		(2.952)	(0.119)	(5.537)
Other physicians' experience:				
prescriptions/previous physicians		60.20***	0.572***	
		(4.542)	(0.152)	
Own experience*others' experience		-3.718	-0.126**	
		(3.499)	(0.044)	
Physicians/prescriptions				86.71***
				(13.73)
Own experience*				
physicians/prescriptions				-31.65*
				(12.88)
Nbr of prescriptions	57.93***			55.60***
	(3.014)			(4.182)
Reimbursement	0.028***	0.028***	0.019***	0.029***
	(0.001)	(0.001)	(0.000)	(0.001)
Prescription date	0.002***	0.002***	-0.000***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Min prescription date	-0.002***	-0.002***	-0.000	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Age, years	0.020	0.018	-0.015	0.035
	(0.092)	(0.092)	(0.009)	(0.091)
Gender	1.722	1.775	0.165	1.499
	(1.531)	(1.513)	(0.149)	(1.486)
Nbr of diagnosis	-0.0260	-0.0343	-0.151**	-0.0927
	(0.479)	(0.489)	(0.059)	(0.463)
Fixed effects:				
physician, ATC-code, hospital district	yes	yes	yes	yes
N	22183	22183	22183	22183
adj. R^2	0.715	0.716	0.974	0.723

[1] Total (cumulative) costs at a given stage of the therapy.

[2] Variables are for cholesterol drug prescriptions.

[2] Standard errors in parentheses.

[3] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.3 A theoretical model of pharmaceutical demand

3.3.1 Overview

In this section, I present a structural model of medical decision-making with private experience and observational learning. In each period during the drug therapy, the patient (he) is randomly matched to a physician (she). After an initial treatment choice, the physician investigates the patient and gets private information about the quality of the match between the patient and the drug treatment. During the course of the patient's therapy, the physician may learn quality from from her own experience and the previous choices of other doctors for this particular patient.[29]

Consider patient i who comes for the first time to a public health center to seek drug treatment for her medical condition. After entrance, a physician is randomly assigned to the patient. As the sensitivity of patients to cholesterol drugs differ, the physician does not know ex-ante the average health effects, or quality, of the drug treatment for this particular patient. To form the prior belief on quality, the physician evaluates the patient's risk for CVDs based on the patient's observed characteristics. The physician takes the prior belief and her privately observed idiosyncratic preferences into account when she decides whether to initiate the cholesterol drug therapy.

In the follow-up of the drug therapy at time (or prescription number) t, patient i comes again to the health center where he is randomly matched physician l. First, the physician performs a diagnostic procedure, physical examination and tests for the patient to privately evaluate the efficacy and side effects of the drug treatment. This evaluation is modeled by an experience signal x_{ilt}. Simultaneously, she looks at patient records to see how long the patient has been using the drug. Conditional on the prior, the past choices of other doctors indexed by $l_1, ..., l_{t-1}$, $h_{it} = \{a_{il_1 1}, ..., a_{il_{t-1}(t-1)}\}$, and all private experience signals that the physician has received during the course of the patient's drug therapy up to and including time t, she updates her belief about its quality.

Recall that in previous social learning models (Cipriani and Guarino, 2012, Knight and Schiff, 2010, Zhang, 2010) agents can receive only one experience signal. Based on this posterior belief and her private preference shocks for the drug treatment and the outside good, v_{il1t} and v_{il0t} respectively, the physician makes a decision on the continuation of the

[29] A relatively easy extension of the model is to enrich the choice set of physicians that could include other medical treatment alternatives, such as non-patented products, with the known (to physicians) but possibly random quality. An extension that allows several inside goods with uncertain qualities comes at the cost of computation.

patient's therapy. Further decisions follow until any physician decides to end the drug therapy. The timing of events is summarized by Figure 3.2.

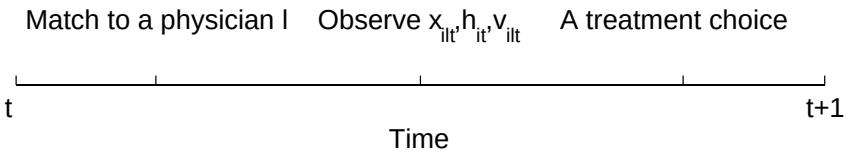

Figure 3.2: The timing of events in period t during the follow-up of the therapy: 1.) a patient is first matched to a physician, 2.) the physician observes a new signal x_{ilt} and the past choices of other doctors h_{it} and private idiosyncratic preference shocks, v_{il1t} and v_{il0t}, 3.) the physician makes a treatment choice on all her private signals received up to and including time t, public information h_{it} and private preference shocks.

In the long-term treatment relationship, the physician learns about the average health effects of the drug treatment from her own experience. If the relationship breaks down, a physician attempts to infer quality from the past choices of other doctors. The less the physician has own experience of the patient, the more the past choices of peers affect her prescription behavior. If the patient has used the drug treatment long, an inexperienced physician may perceive that the drug must be effective. When the drug is of high quality, observing the past choices of other doctors improves learning. On contrary, the optimism on quality leads to over-prescribing when the drug is of low quality.

To keep the model tractable and to avoid the salient computational burden, I assume that a physician maximizes her expected per-period utility. The assumption of myopic behavior is often made in the structural learning literature (e.g. Coscelli and Shum, 2004, Ching, 2009, Chernew et al., 2008) and it abstracts away incentives to experiment with the drug treatment to get new information about quality in the next period (see e.g. Crawford and Shum, 2005).[30]

Following e.g. Crawford and Shum (2005) and Dickstein (2011), the model does not take into account learning across patients.[31] This type of learning could be incorporated to the

[30] My future plan is to estimate a dynamic version of the model.

[31] For learning across patients, see Kim (2010) and Coscelli and Shum (2004). Note also that Crawford

model by using the entry of a new active ingredient, Rosuvastatin. This extension comes again with the cost of computation and tractability because physicians and the econometrician have to keep track on the posteriors of all doctors. Because many cholesterol drugs have been on the market since the end of the 1980s or the early 1990s, learning about the distribution of health effects across patients does not probably have a significant role in my application.

In the following sections, I present the model in detail. I first formulate a deterministic process governing the assignment of a physician for a patient.[32] Because the physician is not forward-looking in her treatment continuation choices, the assignment, or matching, probability does not affect her behavior. Then, I describe a therapy continuation choice under uncertainty and the information structure, including the distribution of signals (health effects) and the patient-specific quality. Finally, I derive the posterior belief of the physician about quality, conditional on her private experience and the patient's prescription history.

3.3.2 The theoretical model

Physician and patient matching

In each period until the therapy ends, patient i is assigned to a physician. The physician is either "new" i.e., she does not have the previous treatment relationship with the patient, or is any of the previously drawn "old" physicians $1, ..., N_{it}$. The number of old physicians at time $t+1$ increases by one, $N_{i(t+1)} = N_{it} + 1$, if the new physician treats the patient at time t, and otherwise it remains unchanged, $N_{i(t+1)} = N_{it}$.

I assume that the patient is assigned to the new physician with probability κ_i and to the old physician with probability $(1 - \kappa_{it}) \times \frac{1}{N_{it}}$. This specification implies that each old physician is randomly selected for the patient from the pool of the previously drawn physicians with the same probability $\frac{1}{N_{it}}$.[33]

I assume the following functional form for the matching probability of patient i:

and Shum (2005) allow the possibility of non-rational expectations, because in their model physicians' *prior* beliefs for one particular drug, Omeprazole, can evolve over time, which captures common changes in priors, for example, due to advertising. However, posteriors may also vary through a different type of mechanism, namely based on the previous medication decisions of a particular physician or other doctors.

[32] The assignment probability is used to recover the probability of the outside good (see Section 4.1).

[33] Note that only 3.5% of patients had more than 2 physicians in my data (see Table 3.2).

$$\kappa_i = P_i(d_{it} = 1) = \frac{e^{y_i}}{1 + e^{y_i}}. \tag{3.1}$$

In the above expression, y_i is $N(\theta^y, \sigma_y^2)$- distributed patient level random coefficient. The variance of the random coefficient, σ_y^2, measures the magnitude of heterogeneity in matching probabilities across patients. The heterogeneity is potentially important because the probability of a physician change can differ between patients, for example, by residential area.

A therapy continuation choice under uncertainty

Assume that physician l is drawn for patient i at time t. The physician decides whether to continue the drug therapy of patient i, $a_{ilt} = 1$, or end the therapy for good, $a_{ilt} = 0$, conditional on her information at that time, I_{ilt}. In the perfect Bayesian equilibrium, the physician chooses to continue the medical therapy if the expected utility from the medical treatment exceeds the utility from the outside option (the non-purchase option),

$$a_{ilt} = 1 \Leftrightarrow E(u_{il1t}|I_{ilt}) \geq u_{il0t}. \tag{3.2}$$

I assume that the per-period utility received from the medical treatment, u_{il1t}, depends on the quality signal, or health effects, x_{ilt}, and a vector of control variables, $\mathbf{Z}_{il1t}$. The controls include, for example, the (average) price of statins, observed patient level characteristics and the time trend capturing general market level changes over time due to advertising. These controls are observed by both physicians and the econometrician. Because patient records do not contain information on preference shocks, I assume that the physician's idiosyncratic, Type 1 extreme value distributed tastes for the drug treatment and the outside option, v_{il1t} and v_{il0t}, are her private information. Following the previous literature (e.g. Crawford and Shum, 2005), I assume a Constant Absolute Risk Aversion (CARA) sub-utility specification for the health effects. To be more specific, I consider the following utility function,

$$u(x_{ilt}, \mathbf{Z}_{il1t}, v_{il1t}) = -e^{-r \cdot x_{ilt}} + \mathbf{Z}_{il1t}\boldsymbol{\alpha} + v_{il1t}, \tag{3.3}$$

where $r > 0$ is the risk aversion coefficient.

I assume that the utility of the outside good for the physician l of patient i at time t, u_{il0t}, is a function of a vector of observed characteristics, $\mathbf{Z}_{il0t}$, and the physician's private preference shock, v_{il0t},

$$u(\mathbf{Z}_{il0t}, v_{il0t}) = \mathbf{Z}_{il0t}\boldsymbol{\beta} + v_{il0t}. \tag{3.4}$$

To ensure identification in the discrete choice model, I make a typical restriction that the constant of the outside option is zero. Recall that the utility of the outside good varies with the patient's observed characteristics (see Chan and Hamilton, 2006, for a similar approach). For example, cholesterol drugs prevent coronary events in the long-run after the patient's drug therapy has ended.[34] I control this with the number of prescriptions.

Health effects

The quality of the match between the patient and the drug treatment (referred as "quality"), θ_i, is without loss of generality either high θ_1 or low θ_0 with prior probabilities $p_i(\theta_1)$ and $1 - p_i(\theta_1)$, respectively. [35] The variance of random quality, $\text{Var}(\theta_i) = E(\theta_i^2) - (E(\theta_i))^2 = p_i(1 - p_i)(\theta_1^2 + \theta_0^2 - 2\theta_1\theta_0)$, measures prior uncertainty regarding to quality. The prior is uninformative when it equals 1/2.

The prior probability is common knowledge for physicians but it may vary across patients, depending on the patient's observed characteristics. I assume that each physician has the following prior belief that the treatment has high quality for patient i:

$$p_i(\theta_1) = \frac{e^{\gamma_0 + \mathbf{Z}_i^p \boldsymbol{\gamma}_1}}{1 + e^{\gamma_0 + \mathbf{Z}_i^p \boldsymbol{\gamma}_1}}, \tag{3.5}$$

where $\mathbf{Z}_i^p$ is a vector of patient level characteristics at the time of the first prescription.

In the follow-up of the patient's drug therapy at time $t > 1$, the physician observes an experience signal, or health effects associated with the use of cholesterol drugs. I assume that health effects are independent and normally distributed conditional on the true quality,

$$x_{ilt}|\theta_i \sim N(\theta_i, \sigma^2), \tag{3.6}$$

[34] The literature has explained this with the stabilization of existing plaque and the slowing of the progression of coronary artery disease (Ford et al., 2007).

[35] The model could be generalized to allow a continuous quality level but the computation of the posterior probability for quality θ conditional on information at time t I_t, $f(\theta|I_t)$, becomes more difficult than in the binary case as it would involve integration over quality levels θ.

where σ^2 measures uncertainty regarding to health effects. The distributions of signals and priors are common knowledge and θ_1, θ_0, σ^2, γ_0 and $\boldsymbol{\gamma}_1$ are parameters to be estimated.[36]

Because prior beliefs are heterogeneous across patients, the unconditional (mixture) density of health effects, $f(x_{ilt})$, depends on the observed characteristics of the patient. This means that the sensitivity of patients on the efficacy and side effects of statins may differ for example by their gender and age, as the medical literature suggests (see Section 3.1).

A physician's information set

Because signals are private information to physicians, a physician's information set for the patient at time t, I^{θ}_{ilt}, includes her own private experience of the patient and the previous therapy continuation choices of other physicians. Formally, $I^{\theta}_{ilt} = \mathbf{x}_{ilt} \cup h_{it} \setminus \{a_{ilt'}, t' < t\}$ where $\mathbf{x}_{ilt}$ is the set of signals that physician l has received up to (and including) time t and $h_{it} \setminus \{a_{ilt'}, t' < t\}$ is the patient's prescription history, $h_{it} = \{a_{il_1 1}, ..., a_{il_{t-1}(t-1)}\}$, without the physician l's actions, $\{a_{ilt'}, t' < t\}$. Because the preference shocks of physician l are her private information, the final information set of physician l at time t for patient i is given by $I_{ilt} = I^{\theta}_{ilt} \cup \mathbf{v}_{ilt}$ where $\mathbf{v}_{ilt}$ is the set of preference shocks that physician l has received up to (and including) time t.

The expected utility

The expected utility of physician l associated with the continuation of the drug therapy for patient i conditional on her information at time t, I_{ilt}, can be written as:

$$\begin{aligned} E(u_{il1t}|I_{ilt}) &= E_{\theta_i|I} E_{x|\theta_i,I}(-e^{-r x_{ilt}}) + \mathbf{Z}_{il1t}\boldsymbol{\alpha} + v_{il1t} \\ &= E_{\theta_i|I}(-e^{-r\theta_i + \frac{1}{2}r^2\sigma^2}) + \mathbf{Z}_{il1t}\boldsymbol{\alpha} + v_{il1t} \\ &= -\lambda_{ilt} e^{-r\theta_1 + \frac{1}{2}r^2\sigma^2} - (1-\lambda_{ilt}) e^{-r\theta_0 + \frac{1}{2}r^2\sigma^2} + \mathbf{Z}_{il1t}\boldsymbol{\alpha} + v_{il1t}. \end{aligned} \tag{3.7}$$

$\lambda_{ilt} = Pr(\theta_1|I_{ilt})$ is the posterior probability that quality is high. The first equality follows from the law of iterated expectations and the second one from the moment generating function of the normal distribution.

The expected utility of the risk averse physician decreases with uncertainty about the effect of the drug therapy on the patient's health, σ^2. The risk aversion parameter increases the

[36] The model could be extended to allow unobserved heterogeneity. In this case, the mean and variance of a signal can differ depending on the type of the patient that is observed by his physicians.

expected utility through quality parameters θ_1 and θ_0 and decreases it through the risk premium $\frac{1}{2}r^2\sigma^2$. Clearly, the latter effect starts to dominate when either σ^2 or the risk aversion parameter r is large enough, namely $r > \frac{2\theta_k}{\sigma^2}$, $k \in \{0, 1\}$.

Public and private beliefs

In this section, I describe how the physician updates her beliefs about the quality of the drug treatment. I find that the posterior belief about quality, λ_{ilt}, is a function of the prior and the physician's private and public beliefs. The private belief is the probability of quality, conditional on physician's accumulated private experience of the patient, $\mathbf{x}_{ilt}$. The public belief is the probability of quality, conditional on the past choices of other doctors. I show that the private experience affects the private belief through a sum of signals. It turns out that this property decreases the computational burden of the model substantially. Even though the physician does not observe the private information of other doctors, she tries to infer quality from their past therapy continuation choices.

The posterior belief

Let $P_i(\theta_1|\mathbf{x}_{ilt})$ denote the private belief of physician l that quality is high for patient i at time t conditional on her private experience $\mathbf{x}_{ilt}$. I denote by $q_{ilt} = P(\theta_1|l, h_{it})$ the corresponding public belief that is conditional on the previous therapy continuation decisions of other physicians $l' \neq l$.

Conditional on health effects $\mathbf{x}_{ilt}$ and the past choices of other doctors for patient i, physician l updates her beliefs about the quality of the treatment for patient i using Bayes' rule and the iid nature of the health effects,

$$\begin{aligned} \lambda_{ilt} &= P_i(\theta_1|l, h_{it}, \mathbf{x}_{ilt}) \\ &= \frac{P(h_{it}|l, \theta_1) f(\mathbf{x}_{ilt}|\theta_1) p_i(\theta_1)}{P(h_{it}|l, \theta_1) f(\mathbf{x}_{ilt}|\theta_1) p_i(\theta_1) + P(h_{it}|l, \theta_0) f(\mathbf{x}_{ilt}|\theta_0) p_i(\theta_0)}. \end{aligned} \tag{3.8}$$

In the above expression, $P(h_{it}|l, \theta)$ is the probability of other doctors' treatment continuation choices for the patient and $f(\mathbf{x}_{ilt}|\theta)$ is the probability of health effects, conditional on the true quality of the drug, $\theta \in \{\theta_0, \theta_1\}$.

The posterior can be linked to the prior, private and public beliefs as follows:

$$\begin{aligned}\lambda_{ilt} &= \frac{q_{ilt}f(\mathbf{x}_{ilt}|\theta_1)}{q_{ilt}f(\mathbf{x}_{ilt}|\theta_1)+(1-q_{ilt})f(\mathbf{x}_{ilt}|\theta_0)} \\ &= \frac{q_{ilt}P_i(\theta_1|\mathbf{x}_{ilt})/p_i(\theta_1)}{q_{ilt}P_i(\theta_1|\mathbf{x}_{ilt})/p_i(\theta_1)+(1-q_{ilt})P_i(\theta_0|\mathbf{x}_{ilt})/p_i(\theta_0)},\end{aligned} \tag{3.9}$$

where the first equality follows from (8). To see this, multiply and divide (8) by $1/P(l,h_{it})$ and note that $q_{ilt}=\frac{P_i(h_{it}|l,\theta_1)p_i(\theta_1)}{P(l,h_{it})}$ where $P(l,h_{it})$ is the probability of the public medication history of the patient without the physician l's actions. The second equality in (9) follows from the first one by dividing and multiplying the first equality by $1/f(\mathbf{x}_{ilt})$ and by observing that $\frac{f(\mathbf{x}_{ilt}|\theta)}{f(\mathbf{x}_{ilt})}=\frac{P(\theta|\mathbf{x}_{ilt})}{p(\theta)}$ for $\theta\in\{\theta_0,\theta_1\}$.

The posterior belief is determined by the prior, $p_i(\theta_1)$, and private and public beliefs, $P_i(\theta_1|\mathbf{x}_{ilt})$ and q_{ilt}. When the public (private) belief is uninformative (equals 1/2), the posterior belief depends only on the private (public) and prior beliefs. When the physician puts weight only on her prior and private experience, the model corresponds to a traditional structural learning model where agents learn only from their private experience (see e.g. Coscelli and Shum, 2004, Crawford and Shum, 2005, Ackerberg, 2003). Recall also that the posterior is an increasing function of private and public beliefs. Hence the higher these beliefs are, the more confident the physician becomes that the quality of the medical treatment is high.

The last step is to derive the evolution of private and public beliefs.

The private belief

First, I describe how the physician learns from her private experience. Assume that the physician has seen the patient S times in the follow-up of the therapy and has observed health $x_{il1},...,x_{ilS}$. Denote by $f(x_{il1},...,x_{ilS}|\theta)$ the joint probability of health effects $x_{il1},...,x_{ilS}$ conditional on θ for $\theta\in\{\theta_0,\theta_1\}$. By using the normality and independence of health effects, the physician updates her private belief about θ_1 for patient i according to Bayes' rule:

$$
\begin{aligned}
P_i(\theta_1|x_{il1},...,x_{ilS}) &= \frac{f(x_{il1},...,x_{ilS}|\theta_1)p_i(\theta_1)}{f(x_{il1},...,x_{ilS}|\theta_1)p_i(\theta_1)+f(x_{il1},...,x_{ilS}|\theta_0)p_i(\theta_0)} \\
&= \frac{\prod_{s=1}^{S}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_{ils}-\theta_1)^2}{2\sigma^2}}p_i(\theta_1)}{\prod_{s=1}^{S}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_{ils}-\theta_1)^2}{2\sigma^2}}p_i(\theta_1)+\prod_{s=1}^{S}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_{ils}-\theta_0)^2}{2\sigma^2}}p_i(\theta_0)} \\
&= \frac{1}{1+e^{\sum_{s=1}^{S}\frac{-(x_{ils}-\theta_0)^2+(x_{ils}-\theta_1)^2}{2\sigma^2}\frac{p_i(\theta_0)}{p_i(\theta_1)}}} \\
&= \frac{1}{1+e^{\frac{1}{2\sigma^2}(-2(\theta_1-\theta_0)X_{ilS}+S(\theta_1^2-\theta_0^2))\frac{p_i(\theta_0)}{p_i(\theta_1)}}}. \qquad (3.10)
\end{aligned}
$$

The posterior[37] depends on signals $x_{il1},...,x_{ilS}$ only through their sum $X_{ilS}=\sum_{s=1}^{S}x_{ils}$, which is also normally distributed given the true quality,

$$X_{ilS}|\theta_i \sim N(S\theta_i, S\sigma^2). \qquad (3.11)$$

The result generalizes to continuous, normally distributed quality, $\theta_i \sim N(\theta,\sigma^2)$.

A physician learns the true quality through her own experience when the number of signals is large enough. Assume that quality is high.[38] In this case, the joint probability for signals converges to zero more slowly than the corresponding probability for low quality. To see this, examine the denominator in (10) that can be rewritten as

$$1+e^{\frac{1}{2\sigma^2}(-S(\theta_1-\theta_0)^2-2(\theta_1-\theta_0)\sigma\sum_{s=1}^{S}e_{ils})}\frac{p_i(\theta_0)}{p_i(\theta_1)} \qquad (3.12)$$

when $x_{ils}=\theta_1+\sigma e_{ils}$ for $e_{ils}\sim N(0,1)$. Because the expected value of e_{ils} is zero, the denominator approaches one when the number of signals S increases.

At the patient population level, the weights of the exponential terms increase when the priors of patients, $p_i(\theta_1)$, $\forall i$, decrease. This delays private learning about high quality and increases variation in private posteriors across patients. Note also that for high enough signal realizations i.e., $X_{ilS}>\frac{S((\theta_1)^2-(\theta_0)^2}{2(\theta_1-\theta_0)}$, the private posterior decreases with the uncertainty parameter σ^2, making physicians less likely to continue the drug therapy.

[37]Note that this is a valid probability distribution as the posterior of signals given the true state is restricted between zero and one.

[38]Private learning on low quality is analogous.

The public belief

Next, I consider the social learning of the physician from the past choices of other doctors. After observing the action of physician $-l$, a_{i-lt}, the physician l (and all other physicians except physician $-l$) updates her posterior belief about high quality by using the following Bayes formula:

$$q_{il(t+1)} = \frac{P(a_{i-lt}|h_{it}, \theta_1) q_{ilt}}{P(a_{i-lt}|h_{it}, \theta_1) q_{ilt} + P(a_{i-lt}|h_{it}, \theta_0)(1 - q_{ilt})}. \tag{3.13}$$

The public posterior belief at time $t+1$ is determined by the (conditional) choice probabilities for high and low qualities and the public belief of physician l at time t. Given that the public beliefs correspond to priors at the beginning of the therapy, $q_{il1} = p_i(\theta_1)$, the final step is to compute the probability of a physician $-l$'s choice, conditional on the patient's prescription history and true quality, $Pr(a_{i-lt}|h_{it}, \theta)$ for $\theta \in \{\theta_0, \theta_1\}$. This is done in two steps.

First assume that physician l observes the physician $-l$'s signals, but not her preference shocks. Let's define a threshold for the difference of private valuations $v_{i-l0t} - v_{i-l1t}$ for which physician $-l$ is indifferent between the continuation and ending of the drug therapy,

$$W_{i-l1t} - W_{i-l0t} = \overline{v}_{i-l0t} - \overline{v}_{i-l1t},$$

where $W_{i-l1t} = E(u_{i-l1t}|I_{i-lt}) - v_{i-l1t}$ is the expected mean utility of the treatment and $W_{i-l0t} = u_{i-l0t} - v_{i-l0t}$ is the corresponding mean utility from the outside good.

Conditional on her signals, the public belief and control variables, a physician's optimal action is to continue the drug therapy if and only if the difference in private valuations is less or equal to the threshold, $v_{il0t} - v_{il1t} \leq \overline{v}_{il0t} - \overline{v}_{il1t}$. If physician l observes that physician $-l$ continued the therapy, she infers that the realization of the difference in private valuations must have been less or equal to this threshold. The larger the threshold, the larger the probability that the drug therapy is chosen.[39]

With the assumption on the distribution of $v_{i-l0t} - v_{i-l1t}$, the conditional choice probability $P(a_{i-lt}|X_{i-lt}, h_{it})$ can be recovered from the thresholds $\overline{v}_{il0t} - \overline{v}_{il1t}$ for all X_{i-lt}. Equivalently, when private valuations are Type 1 extreme value distributed, the conditional probability that physician $-l$ chooses the drug therapy is

[39] See Goeree et al., 2005 for theoretical work with one private signal.

$$
\begin{aligned}
P(a_{i-lt}=1|X_{i-lt},h_{it}) &= P(E(u_{i-l1t}|I_{ilt}) \geq u_{i-l0t}|X_{i-lt},h_{it}) \\
&= \frac{e^{W_{i-l1t}}}{e^{W_{i-l0t}}+e^{W_{i-l1t}}}. \qquad (3.14)
\end{aligned}
$$

As physician l does not observe the physician $-l$'s private experience, the second step is to compute the choice probability, conditional on the patient's prescription history and quality. The conditional choice probabilities for θ_0 and θ_1 are calculated by using the law of iterated expectations,

$$
P(a_{i-lt}=1|h_{it},\theta) = \int \frac{e^{W_{i-l1t}}}{e^{W_{i-l0t}}+e^{W_{i-l1t}}} dF(X_{i-lt}|\theta) \text{ for } \theta \in \{\theta_0,\theta_1\}. \qquad (3.15)
$$

where I average out the effect of the sum of signals on the physician's behavior. Without the property that the private belief depends on signals through their sum, the computation of the conditional choice probability would involve S integrals, instead of one. I compute the choice probability numerically by using Simpson's method with 100 uniform grid points.

When physician $-l$ decides to continue the drug therapy of patient i, the public belief of physician l at time $t+1$, $q_{il(t+1)}$, increases from q_{ilt} and hence she becomes more optimistic about quality. To see this, note first that the sum of signals X_{i-lt} is higher under θ_1 than θ_0. The expected utility associated with the continuation of the drug therapy for physician $-l$, $E(u_{i-l1t}|I_{i-lt})$, is increasing with the posterior belief λ_{i-lt}. The higher the sum of signals X_{i-lt} is, the more confident the physician becomes that quality is high i.e., $\frac{\partial \lambda_{i-lt}}{\partial X_{i-lt}} \geq 0$. Therefore, $P(a_{i-lt}=1|X_{i-lt},h_{it})$ in (19)-(20) is at least as high when quality is θ_1 than θ_0. Because $F(X_{i-lt}|\theta_1)$ has first-order stochastic dominance over $F(X_{i-lt}|\theta_0)$ for $\theta_1 > \theta_0$, $P(a_{i-lt}=1|h_{it},\theta_1) \geq P(a_{i-lt}=1|h_{it},\theta_0)$. As a result, the public posterior of physician l increases from the previous period i.e., $q_{il(t+1)} \geq q_{ilt}$.

3.4 The econometric model and identification

In this section, I present the simulated likelihood function of the structural learning model and discuss identification. I use the following data to compute the simulated likelihood function: 1.) the total number of physician visits for patient i, T_i, where the statin therapy of patient i was continued in periods $1,...,T_i-1$ and the outside option was chosen in

period T_i if the patient is non-censored, 2.) the number of patient i's "old" physicians at time t, N_{it}, 3.) an indicator variable if a previously chosen physician l is drawn for patient i again among N_{it} old physicians, d_{ilt}^{old}, 4.) a vector of control variables affecting utilities received from the statin therapy and the outside good, $\mathbf{Z}_{ilt}$, 5.) the censoring indicator, c_i, and 6.) the characteristics of patient i at the beginning of the therapy, $\mathbf{Z}_i^p$, that affect the prior probability.

3.4.1 The likelihood function

The likelihood contribution of censored patient i contains the following probabilities for each period $t \in \{1, ..., T_i - 1\}$ and physician $l \in \{1, ..., N_{it} + 1\}$ who is drawn for the patient at the beginning of period t: 1.) the probability that physician l is matched to patient i and 2.) the probability that physician l chooses the statin therapy for patient i conditional on the sum of signals and the patient's prescription history, $p_{ilt} = Pr(a_{ilt} = 1 | X_{ilt}, h_{it})$. Because health effects x_{ilt}, preference shocks v_{ilkt}, $k \in \{0, 1\}$, and random coefficients y_i are unobserved by the econometrician, their effects to the likelihood contribution of patient i must be integrated out.

The likelihood contribution of censored patient i is

$$L_i^c \equiv E(\tilde{L}_i^c) = E \prod_{t=1}^{T_i-1} \prod_{l=1}^{N_{it}} \underbrace{[\frac{1-\kappa_i}{N_{it}} p_{il1t}]^{d_{ilt}^{old}}}_{\text{a previously drawn doctor}} \underbrace{[\kappa_i p_{i(N_{it}+1)1t}]^{1-d_{ilt}^{old}}}_{\text{a new doctor}}, \tag{3.16}$$

which consists of the likelihood contributions of the patient's previously drawn and new doctors. For example, $\frac{1-\kappa_i}{N_{it}}$ is the probability that old physician l is drawn for the patient at the beginning of period t and p_{il1t} is the probability that the treatment of patient i is continued at time t by this physician l.

The data does not contain any information on the identity of the physician who decided to end the therapy. To tackle this problem, I first form the joint probability that a certain physician is drawn for the patient and the same physician chooses to end the drug therapy. Then I sum these joint probabilities over the physicians of the patient to recover the probability that any physician ends the therapy at time T_i.

Formally, the likelihood contribution for the observed data of non-censored patient i is

$$L_i^{nc} = E(\tilde{L}_i^c \cdot [\frac{1-\kappa_i}{N_{iT_i}} \sum_{l=1}^{N_{iT_i}} p_{il0T_i} + \kappa_i p_{i(N_{iT_i}+1)0T_i}]), \tag{3.17}$$

where $\frac{1-\kappa_i}{N_{iT_i}} p_{il0T_i}$ is the joint probability that an old physician l is drawn and she decides to end the treatment and $\kappa_i p_{i(N_{iT_i}+1)0T_i}$ is the corresponding joint probability for new physician $N_{iT_i}+1$.

Because expectations over signals in the likelihood function contributions are difficult to compute numerically, I use their simulated counterparts $L_i^{c,s}$ and $L_i^{nc,s}$. For example, for non-censored patients,

$$L_i^{nc,s} = \frac{1}{S} \sum_{s=1}^{S} (\tilde{L}_i^{c,s} \cdot [\frac{1-\kappa_i^s}{N_{iT_i}} \sum_{l=1}^{N_{iT_i}} p_{il0T_i}^s + \kappa_i^s p_{i(N_{iT_i}+1)0T_i}^s]), \tag{3.18}$$

where S is the number of simulations. To compute the simulated likelihood function contribution for each patient, I draw S realization of random coefficients y_i^s governing physicians switching probabilities and $T_i \times S$ realizations of signals and preference shocks to get choice probabilities for each period and patient.[40]

Finally, the simulated log-likelihood function is

$$\log L^s(\theta) = \sum_{i=1}^{N} [c_i \log L_i^{c,s}(\theta) + (1-c_i) \log L_i^{nc,s}(\theta)]. \tag{3.19}$$

In general, simulation error increases the variance of the he maximum simulated likelihood (MSL) $\hat{\theta}_{MSL}$ estimator compared to the maximum likelihood (ML) estimator. This simulation error disappears asymptotically when the number of simulations increases at a rate higher than $\sqrt{N}$. As the estimation of the model is computationally intensive, I set the number of simulations per patients to ten.[41] Obviously, simulation error may be an issue when the number of simulations is small and therefore estimation results must be interpreted with this caveat. To get appropriate standard errors, I use the standard formula for the simulated estimate of the asymptotic variance which relies on the BHHH estimate for the information matrix. I estimate the model by using the derivative free simplex method (see e.g. Cameron and Trivedi, 2005).

[40] Note that only one physician makes a treatment choice each period and therefore in total $T_i \times S$ simulations of signals and preference shocks are needed for each patient.

[41] For example, Crawford and Shum, 2006, had 30 simulations per patient. I plan to experiment with the number of simulations to see how the results would change.

3.4.2 Identification

In this section, I briefly consider the structural assumptions of the demand model and the variation in the data that help identify the parameter vector $\Theta = (\theta_0, \theta_1, \sigma^2, \gamma_0, \boldsymbol{\gamma}_1, \boldsymbol{\alpha}, \theta_y, \sigma_y^2, \eta)$. To a large extent, identification relies on similar arguments that have been presented in the previous literature on demand for experience goods (see e.g. Crawford and Shum, 2005).

Market shares at the beginning of the therapy identify the parameters of the prior distribution, γ_0 and $\boldsymbol{\gamma}_1$, because the treatment choice of the physician is then governed by her prior belief. Because the private learning of the physician decreases uncertainty associated with the quality of the medical treatment, choice probabilities at the end of the long-term drug therapy identify parameters for unobserved quality, θ_0 and θ_1. This is particularly true if the patient is in a long-term treatment relationship with his physician. The identification of quality parameters can be also seen from the expected utility of the drug treatment (equation (3.7)). After fixing the parameters of the prior distribution, γ_0 and $\boldsymbol{\gamma}_1$, and the variance of signals, σ^2, changes in the posterior belief λ_{ilt} with the number of prescriptions identify the quality parameters. Heterogeneity in the choices of physicians both across patients and over time identify the standard deviation of signals. Because quality has two possible values θ_0 and θ_1, it is not possible to separately identify the quality parameters and the risk aversion coefficient, r. I normalize the risk aversion parameter to one which is close to the parameter estimate of Crawford and Shum (2005).[42]

3.5 Results

In this section, I present results from the estimation of the structural learning model and describe the fit of the model. Because the risk of cardiovascular diseases increases with age and is higher for men than for pre-menopausal women, I allow the prior probability to depend the log of age at $t = 1$ and gender. The prior depends also on an indicator variable for whether the patient was treated by an internal disease specialist at the time of the first physician visit. It is likely that the patient, who used the services of the specialist, is more severely ill and gains more from cholesterol drugs.

I allow the utilities associated with the statin treatment and the outside good to depend on several observed variables. First, I let the utility from therapy continuation to depend

[42] An alternative is to interpret parameters θ and σ^2 relative to risk aversion coefficient r, e.g. $\hat{\theta}_1 = r\theta_1$, where $\hat{\theta}_1$ is the estimated parameter.

on for the average price of statins at time t. I also control for a time trend in months since January 2003 because market level changes, such as advertising, might as well affect the utility from statins. Because the patient's health might deteriorate when he becomes older, I let the utility without cholesterol drugs to depend on age at time t. As cholesterol drugs prevent coronary events in the long-run after the patient's drug therapy has ended, I allow the outside good utility to vary with the number of prescriptions.[43]

Discussion of the results and the fit of the model

Table 3.5 presents the parameter estimates and their standard errors. The first set contains the key parameters of the model: quality levels θ_0 and θ_1 and the standard deviation of health effects, σ (see 3.6). Figure 3.3 presents the conditional and unconditional distributions of signals, $f(x_{ilt}|\theta_0)$, $f(x_{ilt}|\theta_1)$ and $f(x_{ilt})$, for the estimated parameters and the average of priors $p_i(\theta_1)$.

[43] Alternatively, the controls of the outside good could be included in a vector of inside good controls.

Table 3.5: Parameter estimates for the learning model in the sample of patients

Parameter	Estimate	Std.Err.
Signal (x_{ilt}) parameters		
Low quality (θ_0)	-0.220	0.001
High quality (θ_1)	1.338	0.002
Std. Dev. (σ)	1.049	0.003
Prior parameters		
Constant (γ^0)	-0.003	0.001
log(Age in years at t=1)	0.120	0.000
Gender	0.093	0.001
Visit an internal disease		
specialist at t=1 (1: yes, 0: no)	0.067	0.443
Prior mean and std	0.717	0.012
Physician matching probability		
Random coefficient		
Constant (θ^y)	-0.049	0.001
Std. Dev. (σ_y)	1.057	0.004
Physician switching probability,		
mean and average std	0.491	0.217
Control variables		
Patient's deductible, eur	-0.021	0.000
Time trend in months/10	-0.028	0.000
Outside good controls		
Patient's age/10 years	-0.089	0.000
Number of prescriptions/10	0.107	0.000
Number of observations	22 021	
Number of patients	10 000	
Number of simulations[1]	10	
Simulated log-likelihood function	30 555	

[1] The number of simulations per patient and physician visit.

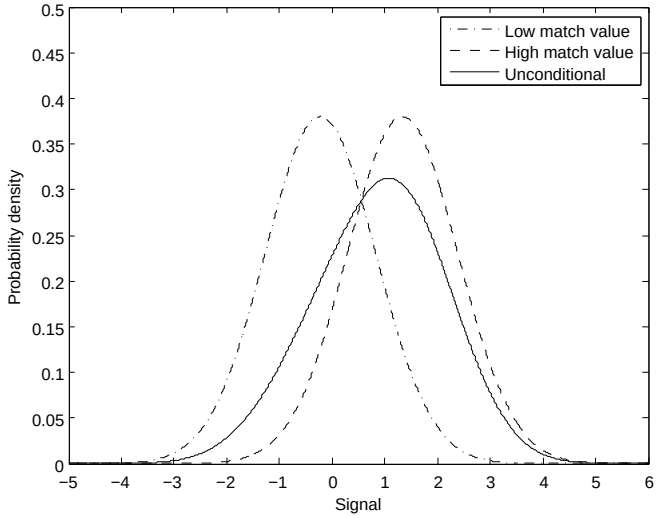

Figure 3.3: The conditional and mixture probability densities of signals, $f(x_{ilt}|\theta_0)$, $f(x_{ilt}|\theta_1)$ and $f(x_{ilt})$, for estimated parameters and the average prior in the sample of patients

The results demonstrate substantial uncertainty and heterogeneity among patients in the quality and health effects of the statin treatment. The parameter estimate for high quality θ_1 (1.34) is in absolute terms over 6 times higher than the estimate of low quality, θ_0 (-0.22). The variance estimate of signals, σ^2, implies that physicians face significant uncertainty about the health effects of statins even if quality was known. To be more precise, the variance of signals is 5 times higher than the low quality estimate $\hat{\theta}_0$ and 82% of the value of the high quality estimate $\hat{\theta}_1$.

Heterogeneity in health effects implies that information and learning may significantly improve medical decision-making by a physician. Without uncertainty about quality, the incentives of the physician to continue the patient's therapy may be much higher when quality is high rather than low. A high uncertainty in health effects decreases the expected utility of a risk-averse physician, slows down her learning[44] and diminishes her incentives to continue the patient's statin therapy.

The second set of parameters in Table 3.5 includes estimates for the physician's prior

[44]This can be seen from the denominator of equation (3.15) in which iid physician l's shocks e_{ils}, $s \in \{1, ..., S\}$, for patient i get more weight when standard deviation σ increases.

belief that the quality of the statin treatment is high, $p_i(\theta_1)$. As expected, the physician has a higher prior probability if her patient is older and male and thus has a higher risk of CVDs compared with other patients. Quite intuitively, the prior belief is higher if the patient has visited an internal disease specialist at the time of the first prescription.

Depending on the characteristics of patients, the prior probability varies across patients from 65% to 75% and has a mean of 72% with a small standard deviation. At the beginning of the therapy, the physician beliefs that quality is more likely to be high than low. Because the average prior belief is fairly uninformative, the posterior belief of the physician λ_{ilt} is mostly determined by her private and public beliefs. This, coupled with a relatively large variance of signals, σ^2, implies that the learning of the physician from her private experience may take some time.

Third, I report the parameters of the random coefficient y_i that affects the probability that the patient is assigned for a new physician, κ_i. The set of parameters for the random coefficient includes the constant, θ_y, and the standard deviation, σ_y. The results suggest that the estimated standard deviation $\hat{\sigma}_y$ (1.06) is much higher than the estimated mean $\hat{\theta}^y$ (-0.05). These findings imply that the probability of getting a new physician varies substantially (0-99%) around its mean (49%). The (average) standard deviation of κ_i is 0.19 that is 32% of the estimated mean of κ_i. Heterogeneity in assignment probabilities across patients can arise for several reasons, including differences between municipalities in their ability to recruit permanent physician labour.

The final set of variables includes control variables affecting utilities associated with the statin therapy and the outside option. The price of statins has a very small, negative effect on the expected utility from the statin treatment. A physician can be insensitive to changes in average prices because a significant part of expenses is covered by the national health insurance. Over time, the expected utility of the physician from the statin treatment decreases. This may reflect changes in advertising by pharmaceutical firms over a product's life cycle and other market level changes. Physicians whose patients are older, and hence have a higher risk of having more severe diseases, are less likely to end the statin therapy as their patients gain less from the outside alternative. The utility associated the outside good increases with the number of prescriptions. This may happen because the statin therapy is likely to have long-term effects on the patient's health even after the statin therapy has ended.

Finally, I consider the model fit by comparing average predicted and observed choice probabilities. For each physician-patient pair, I compute the predicted probability of choosing the statin treatment, conditional on the sum of signals and the patient's prescription his-

tory, $P(a_{ilt} = 1|X_{ilt}, h_{it})$. I then compare the corresponding observed choice probabilities to these predicted probabilities, as presented in Figure 3.4. The model fits the data relatively well even though it slightly over-predicts the observed average choice probability at the beginning of the treatment and under-predicts after that. At the aggregate level, the model fits the data reasonably well: the average observed probability of choosing the statin therapy is 79% which is close to the predicted probability, 81%. The average predicted probability of getting a new physician is lower (49%) than the corresponding observed probability (60%).

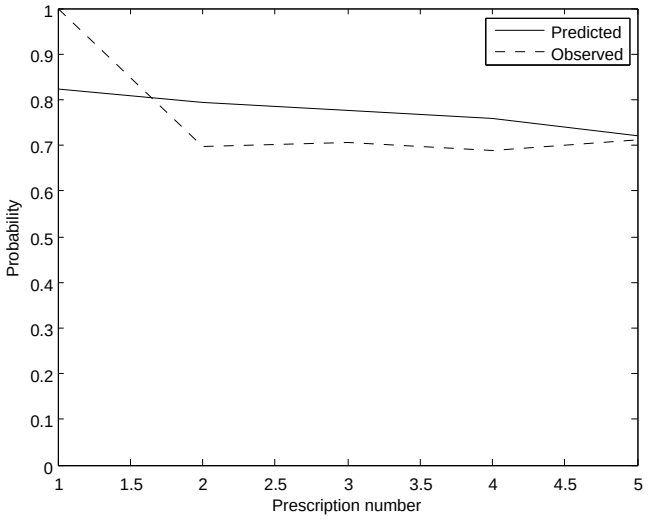

Figure 3.4: Difference between observed and predicted choice probabilities by the number of prescriptions in the sample, an average over patients, physician visits and simulations

3.6 Counterfactual experiments

After estimating the parameters, I quantify the roles of private and observational learning in medical decision-making. The main objective is to evaluate to the length of the doctor-patient relationship affects the process of learning and the efficiency of medical decision-making. To be more specific, I evaluate whether the policy promoting continuity of care is preferable to providing information on the past choices of other doctors.

I first investigate what happens if the patient had only one physician. In this case, the

physician learns only from her private experience. Next, I investigate whether information on the past choices of other doctors compensates for the lack of continuity of care. To do this, I compare treatment outcomes and costs in the long-term treatment relationship with the policy where the patient has a different physician every period. A physician has then a one-shot opportunity to investigate the patient to get information on the health effects of cholesterol drugs but she observes the patient's treatment history. To understand the role of peer effects in demand, I study how the behavior of the physician changes if information on the past choices of other doctors was not available. In this experiment, the physician has to rely only on her private experience and the prior belief. Finally, I evaluate the consequences of the policy where the physician does not learn. In this case, the physician decides about the continuation of the patient's therapy without investigating him. I compare the results with the baseline scenario predicted by the estimated model. To perform the policy experiments, I simulate 10 prescription paths for each patient in the observed sample of 10 000 patients used in the estimation of the model.[45]

I begin by describing the development of posterior beliefs over time and dispersion among patients under different policy experiments. I then investigate how treatment adherence, expected utilities and costs change when the length of the treatment relationship and the amount of available information were changed. I measure adherence by the predicted length of the drug therapy and the probability of choosing the statin therapy conditional on the information of the physician, $P(a_{ilt} = 1|I^{\theta}_{ilt})$ (see Dickstein, 2011, for the similar approach).

The speed of learning

Figure 3.5 describes the development of the average posterior belief over patients, physicians and simulations, conditional on high quality. At the beginning of the therapy, a physician is fairly pessimistic about the effect of the drug treatment on patient health since the average prior for low quality is 28%. Most of the uncertainty regarding to quality vanishes after the first physician visit. At this stage of the therapy, the physician has observed how well the first prescription decreased the patient's cholesterol levels and whether any side effects realized. In the long-term treatment relationship, the physician

[45]When the number of predicted prescriptions is less than the observed one, I use the observed characteristics of patients. Otherwise, I assume that patients come to seek treatment for high cholesterol once a year. The time trend increases by 12 months, the patient's age by a one year and the number of prescriptions by one in period $t + 1$ from the previous period t. An exception is the average price of statins at time t which I replace with the average over time, products and patients.

learns quality fast, by the eighth physician visit. In short-term relationships, physicians become more optimistic on quality during the course of the patient's therapy, but learning is slower than in the long-term relationship. The bottom half of Figure 3.5 presents the standard deviation of posterior beliefs. At the first prescription, variation in posteriors arises because prior beliefs are heterogeneous across patients. Reflecting high variation in health effects, the standard deviation increases to 0.2 at the second prescription. As expected, learning diminishes the variances of the posteriors gradually.

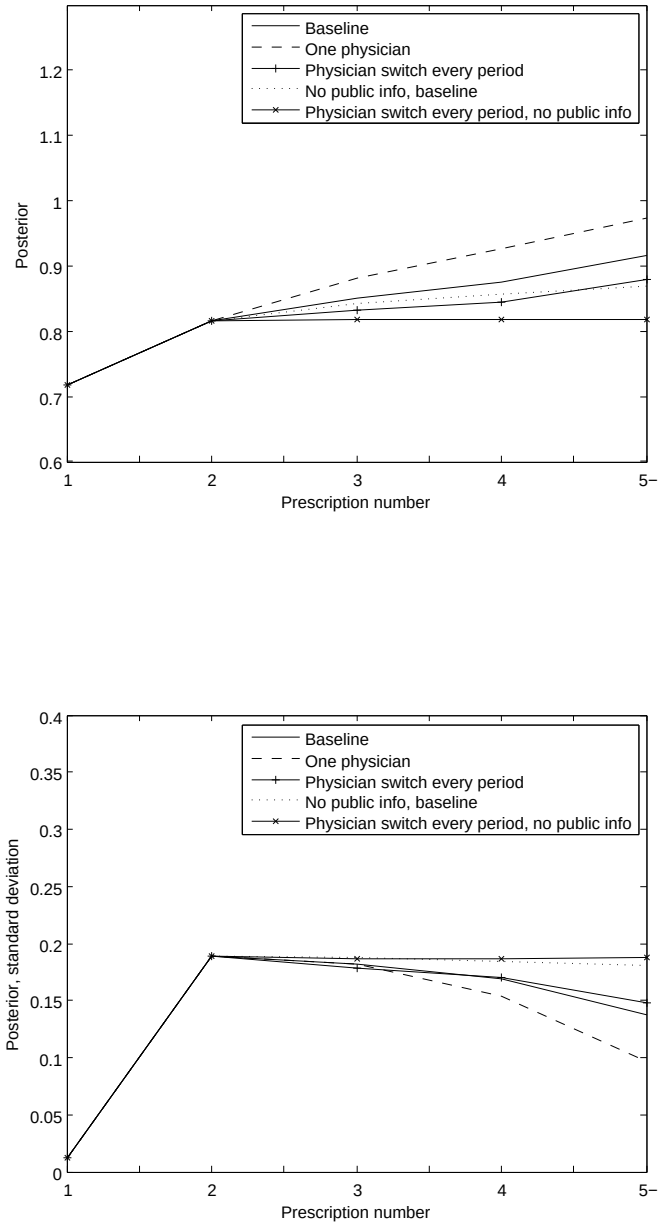

Figure 3.5: The mean (higher figure) and variance (lower figure) of the posterior belief $\lambda_{ilt} = Pr(\theta_1|I_{ilt})$ given that true quality is high ($\theta_i = \theta_1$) in the sample of patients

The top of Figure 3.6 illustrates the development of the average posterior when the patient-specific quality is low. In this case, large differences in average posteriors between different

scenarios arise. In the long-term treatment relationship, the physician learns again fast. If physicians change frequently, the average posterior starts to increase after a few prescriptions. Again, the physician becomes more optimistic about quality when other doctors have chosen the drug treatment for the patient previously. The bottom part of Figure 3.6 shows that heterogeneity in posteriors at the aggregate level is higher among patients when quality is low rather than high. The standard deviation of posteriors are fairly similar in the counterfactual experiments. In particular, a high variation in the posteriors remains also in the permanent treatment relationship, even though the posterior belief is decreasing over time.[46]

[46]Note that the exponential term in equation (3.10) is $e^{S(\theta_0-\theta_1)^2-2(\theta_1-\theta_0)\sigma\sum_{s=1}^{S}e_{ils}}$ if $\theta_i = \theta_0$. When $2(\theta_1-\theta_0)\sigma\sum_{s=1}^{S}e_{ils}$ is high relative to constant term $S(\theta_0-\theta_1)^2$, there can be much variation in the posterior beliefs of physicians among patients.

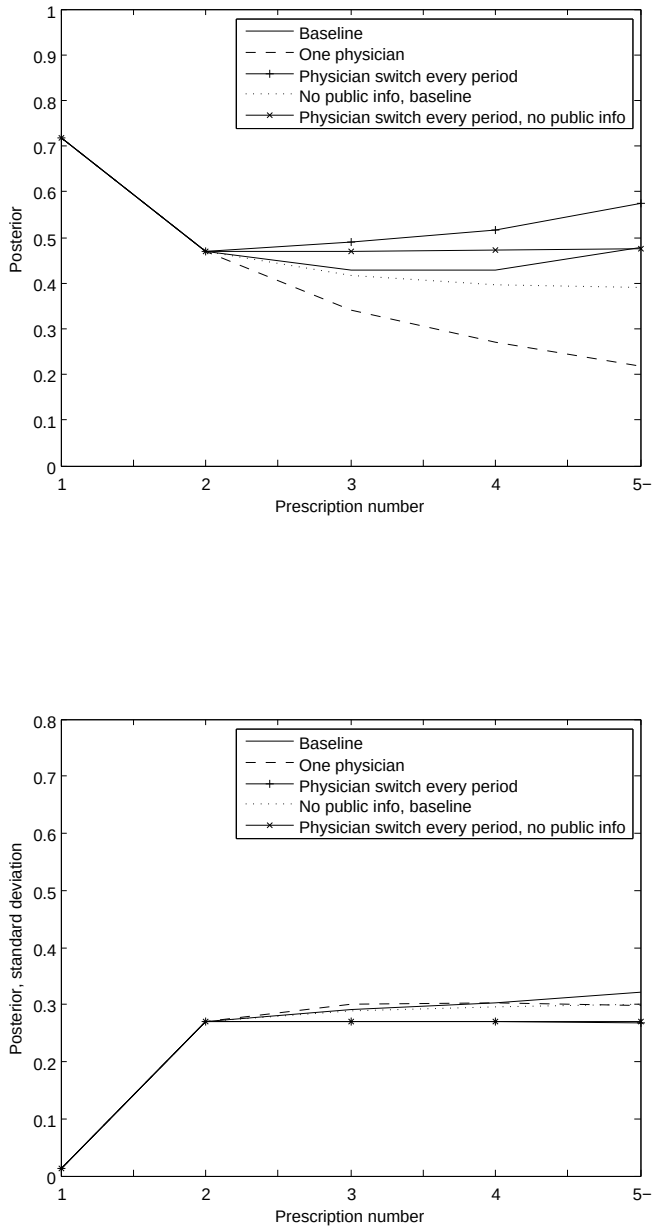

Figure 3.6: The mean (higher figure) and variance (lower figure) of the posterior belief $\lambda_{ilt} = Pr(\theta_1|I_{ilt})$ given that true quality is low ($\theta_i = \theta_0$) in the sample of patients

Overall, the results suggest that the long-term doctor-patient relationship promotes the process of learning about quality. The physician becomes optimistic about quality when

she observes the past choices of other doctors. When quality is high, information on the prescription history improves learning, but not as efficiently as the long-term relationship. When quality is low, such information slows down learning. These results have implications on prescriptions, costs and efficiency.

The length of the doctor-patient relationship

I first examine how the long-term doctor-patient relationship affects outcomes and costs. Table 3.6 presents averages for the expected utility, the adherence of the treatment and the total costs, conditional on quality. The results suggest that continuity of care promotes learning and improves medical decision-making by a physician. Consider first the patient with high quality of the match with cholesterol drugs. In this case, the long-term physician-patient relationship leads to the highest expected utility and the treatment adherence among evaluated experiments. Still, the treatment adherence and the total costs increase only slightly from the estimated benchmark. When quality is low, I find that continuity of care decreases treatment length by 5% and the total costs of the drug therapy by 5% compared with the estimated benchmark. This is so, since the physician learns fast that the treatment does not suit well for the patient.

Table 3.6: Counterfactual simulations in the sample of patients

Outcome variable	Baseline[2]	One physician[3]	Physician change every period	No public info, info[4]	No public physician change	No learning[5]
True quality: θ_1:						
Expected utility	-2.612	-2.594	-2.624	-2.618	-2.632	-2.704
Expected utility, ≥ 5 prescriptions	-2.791	-2.765	-2.808	-2.802	-2.824	-2.873
Treatment length	4.327	4.420	4.420	4.272	4.199	4.013
Probability of statin therapy	0.770	0.774	0.766	0.766	0.763	0.751
Total cost/100 eur	2.883	2.945	2.844	2.846	2.797	2.674
True quality: θ_0:						
Expected utility	-2.815	-2.844	-2.798	-2.823	-2.808	-2.692
Expected utility, ≥ 5 prescriptions	-3.034	-3.104	-3.006	-3.056	-3.031	-2.865
Treatment length	3.406	3.243	3.503	3.361	3.439	3.967
Probability of of statin therapy	0.706	0.692	0.715	0.703	0.710	0.748
Total cost/100 eur	2.270	2.160	2.335	2.240	2.292	2.644

[1] These values are calculated by using the observed sample of 10 000 patients and 10 simulated prescription sequences per patient.

[2] The baseline scenario is predicted by the model estimates.

[3] In this experiment, the physician-patient relationship is permanent.

[4] Public information on the previous treatment continuation choices of other physicians is not available.

[5] Learning is prevented and hence the posterior of physician l for patient i at time t, λ_{ilt}, equals to the prior belief $p_i(\theta_1)$.

I next investigate the consequences of the policy where a new physician treats the patient in every period (Table 3.6). When the physician does not have much own experience of the patient, she relies more on the past choices of other doctors. Consider first the patient with high quality of the match with statins. In this case, continuity of care does not much improve drug choices or change treatment outcomes compared to the policy where treatment relationships are short-term but the prescription history is observed.

Consider then the patient with the low quality of the match in Table 3.6. In this case, the length of the treatment relationship has more pronounced effect on treatment outcomes and costs. This happens because social learning increases the optimism of the physician about the quality and can lead to over-prescribing. The results show that the policy with the short-term relationship increases the adherence by 3% and the total costs by 8% from

the experiment with continuity of care. Table 3.6 demonstrates that the physician would be slightly better-off, in terms of efficiency, without information on the prescription history when physicians change frequently. Specifically, when treatment relationships are short-term, providing information on the past choices of other doctors increases the adherence by 1% and the total costs by 2% from the policy without such information. Again in terms of efficiency, even worse outcomes arise if learning is not possible.

The results have several policy implications. Continuity of care helps the physician to find out sooner the health effects of the drug treatment. This reduces the costs of uncertainty and improves her medical decision-making, as suggested by the existing reduced-form literature (Weiss and Blustein, 1996, Scott, 2000, King et al., 2008). The second conclusion is that information on the patient's prescription history does not compensate for the lack of the long-term treatment relationship. When the treatment suits well for the patient, prescription records promote learning, but not as efficiently as continuity of care. If a physician does not have much own experience, treatment patterns based on the observed medication history of the patient may hinder learning and lead to over-prescribing for a fraction of patients.

3.7 Conclusions

I quantified the roles of private experience and the past choices of other doctors in pharmaceutical demand. I constructed a structural model of demand for pharmaceuticals under uncertainty about the quality of the match between the patient and the drug treatment. I analyzed whether continuity of care is more efficient than the policy where information on the past choices of other doctors is observed but treatment relationships are short-term.

Using rich data from the market for cholesterol drugs, I found that prescriptions are highly responsive to the length of the doctor-patient relationship. I illustrated that the number of interactions between the physician and the patient have important implications on pharmaceutical demand. My analysis suggested that treatment patterns relying heavily on the past choices of other doctors may lead to over-prescribing for a fraction of patients, in terms of efficiency. I also showed that the long-term treatment relationship can limit over-prescribing and improve medical decision-making.

The structural model can be extended to allow the other important features of the pharmaceutical market. The first extension is to make physicians forward-looking in their decision-making, creating incentives for experimentation to get more information. Second, the model can be broadened to incorporate several inside goods. The framework

can be also applied in other experience good markets, such as financial markets, where traders are investing in assets with uncertain returns.

A typical example of a patient story for one dispensary visit:

The reason of entry
A patient comes with the referral of physician X due to atrial fibrillation

At issue a 65 years old retired gymnastics teacher. In an anamnesis 2003 acute coronary thrombosis, angioplasty RCA. Discovered then also a decreasing diverticulum of an aorta ad 50mm, controls in fall. In the Doppler-ultrasound-research of neck veins in 2005 was discovered in left arteria carotis interna stenosis less 50%. Discovered year 2007 COPD. The patient smoked over 30 years, quit 6 years ago. In a tolerance test 8/07, no coronary ischaemia.

The patient has visited in the health center of X due to dizziness. Discovered elevated blood pressure, irregular beat. Hear enzymes and other laboratory values normal, pro-BNP over 500. Patient's medication at this moment Pravachol 20mg x 1, Linatil 20mg x 1, Carvedilol 12.5mg x 2. Started Marevan due to atrial fibrillation, aiming to do cardioversion.

Today taken INR, only 1.3. Hence cardioversion cannot be done now. Pulse also fairly fast 80-90/min, RR-level 180-170/110-100. Carvedilol ad raised 25mg + 12.5mg. INR-controls will continue in the side of outpatient treatment. Phone contact after a month.

Physician X

Bibliography

[1] Ackerberg, D. (2003) Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination. *International Economic Review,* Vol. 44, pp. 1007-1040.

[2] Adams, S. (2011) Millions Taking Statins 'Needlessly'. *The Telegraph.*

[3] Anderson, L. – Holt, C. (1997) Information Cascades in the Laboratory. *American Economic Review*, Vol. 87, pp. 847-862.

[4] Athey, S. – Stern, S. (2002) The Impact of Information Technology on Emergency health care Outcomes. *RAND Journal of Economics*, Vol. 33, pp. 399-432.

[5] Avery, C. – Zemsky, P. (1998) Multidimensional Uncertainty and Herd Behavior in Financial Markets. *American Economic Review*, Vol. 88, pp. 724-748.

[6] BBC (2011) Questions over statin prescribing.

[7] Bikhchandani, S. – Hirshleifer, D. – Welch, I. (1992) Theory of Fads, Fashion, Custom, and Structural Change as Informational Cascades. *Journal of Political Economy*, Vol. 100, pp. 992-1026.

[8] Cameron, A. – Trivedi, P. (2005) Microeconometrics - Methods and Applications. *Cambridge University Press*, New York, USA.

[9] Celen, B. – Kariv, S. (2004) Distinguishing Informational Cascades from Herd behavior in the Laboratory. *American Economic Review*, Vol. 94, pp. 484-498.

[10] Chan, T. – Hamilton, B. (2006) Learning, Private Information, and the Economic Evaluation of Randomized Experiments. *Journal of Political Economy*, Vol. 114, pp. 997-1040.

[11] Chamley, C. (2004) Rational Herds - Economic Models of Social Learning. *Cambridge University Press*, the United States of America.

[12] Chaudhry, B. – Wang, J. – Wu, S. – Maglione, M. – Mojica, W. – Roth, E. – Morton, S. – Shekelle, P. (2006) Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Annals of Internal Medicine,* Vol. 144, pp. 742-752.

[13] Chernew, M. – Gowrisankaran, G. – Scanlon, D. (2008) Learning and the Value of Information: Evidence from Health Plan Report Cards. Journal of Econometrics, Vol. 144, pp. 156-174.

[14] Chevalier, J. – Ellison, G. (1999) Career Concerns of Mutual Fund Managers. *Quarterly Journal of Economics*, Vol. 114, pp. 389-432.

[15] Ching, A. (2009) A Dynamic Oligopoly Structural Model for the Prescription Drug Market after Patent Expiration. Forthcoming in *International Economic Review.*

[16] Ching, A. (2010a) A Dynamic Oligopoly Structural Model for the Prescription Drug Market After Patent Expiration. *International Economic Review*, Vol. 51, pp. 1175-1207.

[17] Ching, A. (2010b) Consumer Learning and Heterogeneity: Dynamics of Demand for Prescription Drugs After Patent Expiration. *International Journal of Industrial Organization*, Vol. 28, pp. 619-638.

[18] Cipriani, M. – Guarino, A. (2012) Estimating a Structural Model of Herd behavior in Financial Markets. Forthcoming in *American Economic Review.*

[19] Coscelli, A. – Shum, M. (2004) An Empirical Model of Learning and Patient Spillovers in New Drug Entry. *Journal of Econometrics*, Vol. 122, pp. 213-246.

[20] Crawford, G. S. – Shum, M. (2005) Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, Vol. 73, pp. 1137-1173.

[21] Crawford, V. – Sobel, J. (1982) Strategic Information-Transmission. *Econometrica*, Vol. 50, pp. 1431-1451.

[22] Dickstein, M. (2011) Efficient Provision of Experience Goods: Evidence from Antidepressant Choice. Job Market Paper, Harvard University.

[23] Drehmann, M. – Oechssler, J. – Roider, A. (2005) Herding and Contrarian Behavior in Financial Markets: An Internet Experiment. *American Economic Review*, 95, pp. 1403-1426.

[24] Epstein, R. M. (1999) Mindful Practice. *Journal of American Medical Association*, Vol. 282, pp. 833-839.

[25] Finnish Medical Association (FMA) (2006a) Physician Situation in Health Centers 2006 (In Finnish: Terveyskeskusten lääkäritilanne, 2006).

[26] Finnish Medical Association (FMA) (2006b) Ennalta Ehkäisevä Terveydenhuolto Vaarantuu: Omalääkärijärjestelmä Uhkaa Romuttua. *FMA Newsletter*, 2.5.2005.

[27] Finnish Medical Association (FMA) (2006c) Physician Survey. www.laakariliitto.fi/files/laakarikysely2006.pdf (In Finnish)

[28] Ford, I. – Murray, H. – Packard, C. – Shepherd, J. – Macfarlane, P. – Cobbe, S. (2007) Long-Term Follow-up of the West of Scotland Coronary Prevention Study. *New England Journal of Medicine*, Vol. 357, pp. 1477-1486.

[29] Franklin, C. (2011) America's Epidemic of Over-Prescribing. *Guardian*.

[30] Goeree, J. – Palfrey, T. – Rogers, B. (2005) Social Learning with Private and Common Values. *Economic Theory, Research Articles*, Vol. 28, pp. 245-264.

[31] Golman, G. (1990) The Tacit Dimension of Clinical Judgment. *The Yale Journal of Biology and Medicine*, Vol. 63, pp. 47-61.

[32] Graham, J. (1999) Herding among Investment Newsletters: Theory and Evidence. *The Journal of Finance*, Vol. 54, pp. 237-268.

[33] Guthrie, B. – Saultz, J. W. – Freeman, G. K. – Haggerty, J. L. (2008) Continuity of Care Matters. *BMJ,* Vol. 337, pp. -.

[34] Haggerty, J. – Reid, R. – Freeman, G. – Starfield, B. – Adair, C. – McKendry, R. (2003) Continuity of Care: a Multidisciplinary Review. *BMJ*, Vol. 327, pp. 1219-1221

[35] Hillestad, R. – Bigelow, J. – Bower, A. – Girosi, F. – Meili, R. – Scoville, R. – Taylor, R. (2005) Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, And Costs. *Health Affairs*, Vol. 24, pp. 1103-1117.

[36] Hong, H. – Kubik, J. – Solomon, A. (2000) Security Analysts' Career Concerns and Herding of Earnings Forecasts. *RAND Journal of Economics*, Vol. 31, pp. 121-144.

[37] Joelving, F. (2011) Statin Benefits Questioned in Low-Risk Patients. *Reuters Health.*

[38] Johnston, D. (1999) Lifestyle Changes After a Myocardial Infarction. Editorial in *Heart*, Vol. 82, pp. 543-544.

[39] Jousilahti, P. (2004) Current Care for Dyslipidemias is Renewed (in Finnish: Dyslipidemioiden Käypä hoito uudistuu). *Kansanterveys-lehti*, National Institute for Health and Welfare.

[40] Kim, M. (2010) Physician Learning and New Drug Diffusion. Job Market Paper, University of Chicago.

[41] King, M. – Jones, L. – Richardson, A. – Murad, S. – Irving, A. – Aslett, H. – Ramsay, A. – Coelho, H. – Andreou, P. – Tookman, A. – Mason, C. – Nazareth, I. (2008) The Relationship Between Patients' Experiences of Continuity of Cancer Care and Health Outcomes: a Mixed Methods Study. *British Journal of Cancer*, pp. 529-536.

[42] Klaukka, T., Helin-Salmivaara, A., Huupponen, R., Idanpää-Heikkilä, J. (2011) Simvastatiinin asema vahvistui vuonna 2005. *Finnish Medical Journal*, 18/2006, Vol. 61, pp. 1998-2001.

[43] Knight, B. – Schiff, N. (2010) Momentum and Social Learning in Presidential Primaries. *Journal of Political Economy*, Vol. 118, pp. 1110-1150.

[44] Ministry of Social Affairs and Health in Finland (2005) National Guide for Pharmacotherapy in Social and health care. *Handbooks of the Ministry of Social Affairs and Health, Finland*, 32, pp. 1-117.

[45] Nair, H. S. – Manchanda, P. – Bhatia, T. (2010) Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders. *Journal of Marketing Research*, Vol. 47, pp. 883-895.

[46] National Audit Office of Finland (2011) Sosiaali- ja Terveydenhuollon Valtakunnallisten IT- hankkeiden Toteuttaminen. *Valtiontalouden tarkastusviraston tuloksellisuustarkastuskertomukset*, 217/2011, pp. 1-274.

[47] National Center for Health Statistics (2011) Health, United States, 2011: With Special Feature on Socioeconomic Status and Health. Hyattsville, MD. 2012.

[48] National Health Service (2013) Cholesterol-Lowering Medicines, Statins - Side Effects.

[49] National Institute for Health and Clinical Excellence (2006) Statins for the Prevention of Cardiovascular Events. *NICE Technology Appraisal*, Vol. 94, pp. 1-45.

[50] National Institute of Health and Welfare (2011) Cardiovascular Diseases.

[51] Novack DH, Suchman AL, Clark W, Epstein RM, Najberg E, Kaplan C.(1997) Calibrating the Physician. Personal Awareness and Effective Patient Care. Working Group on Promoting Physician Personal Awareness, American Academy on Physician and Patient. *The Journal of American Medical Association*, Vol. 278, pp. 502-509.

[52] Olszewski, W. (2004) Informal Communication. *Journal of Economic Theory*, Vol. 117, pp. 180-200

[53] Park, A., Sabourian, H. (2010) Herding and Contrarian Behavior in Financial Markets. Forthcoming in *Econometrica.*

[54] Pisacano, N. J. (2007) History of the Specialty. *American Board of Family Medicine.*

[55] Scott, A. (2000) Economics of General Practice. Chapter 22 in *Handbook of Health Economics*, Volume 1, Edited by A.J. Culyer and J.P. Newhouse, Elsevier Science B.V.

[56] Scott-Morton, F. (1999) Entry Decisions in the Generic Drug Industry. *RAND Journal of Economics*, Vol. 30, 421-440.

[57] Smith, L. – Sorensen, P. (2000) Pathological Outcomes of Observational Learning. *Econometrica*, Vol. 68, No. 2, pp. 371-398.

[58] Sturmberg, JP. (2001) Continuity of Care Revisited. *Family Practice*, Vol. 18, pp. 461-462.

[59] U.S. Food and Drug Administration (FDA) (2010) Controlling Cholesterol with Statins. *FDA Consumer Health Information*, pp. 1-2.

[60] Vives, X. (2008) Information and Learning in Markets: The Impact of Market Microstructure. *Princeton University Press,* the United States of America.

[61] Weiss, L.J. – Blustein, J. (1996) Faithful Patients: The Effect of Long-Term Physician-Patient Relationships on the Costs and Use of Health Care by Older Americans. *American Journal of Public Health*, Vol. 86, pp. 1742-1747.

[62] Winblad, I. – Reponen, J., Hämäläinen, P. – Kangas, M. (2006) Informaatio- ja kommunikaatioteknologian käyttö. Tilanne ja kehityksen suunta. Stakes, Reports, 7/2006, pp. 1-83.

[63] Wold heart federation (2011) Cardiovascular disease risk factors. http://www.world-heart-federation.org/cardiovascular-health/cardiovascular-disease-risk-factors/

[64] Zhang, J. (2010) The Sound of Silence: Observational Learning in the U.S. Kidney Market. *Marketing Science*, Vol. 29, pp. 315-335.

Chapter 4

Do Stronger Patents Protect Against Competition? Evidence from the Pharmaceutical Industry

Policy makers have been strengthening intellectual property rights in several countries during the past decades. The rationale behind these reforms is that stronger patents increase the profits of an innovator and promote R&D. Economic theory predicts that longer patents may hinder rather than stimulate innovation by increasing competition during the patent period. Broad patents, on the other hand, increase the costs of imitation and decrease competition. I test the theory on the relationship between patent strength and competition during patent protection. I consider the Finnish markets for pharmaceuticals that provide rich variation in both patent length and breadth across innovations. The results suggest that patent breadth, rather than length, prevents imitation. Patent rights have no effect on the risk of parallel trade.

Keywords: intellectual property rights, imitation, parallel trade, pharmaceuticals, innovation

4.1 Introduction

Several countries, including the United States, have been strengthening their intellectual property rights during the past few decades (see e.g. Gallini, 2002). The economic rationale for stronger patents is to provide inventors larger rewards from R&D and to increase innovation incentives. The theoretical literature has challenged this view by showing that longer patents can increase competition during the patent period and thus decrease the profits of an innovator (Gallini, 1992, Takalo, 1998). On the other hand, an increase in patent breadth raises the costs of developing non-infringing duplicates and thus decreases the entry incentives of competitors.[1] In this chapter, I test empirically the theory on the relationship between patent strength and competition during the patent period.

I investigate markets for pharmaceuticals where competition during patent protection can arise for at least two reasons. First, competitors (so called parallel traders) may resale patented products between countries without the authorization of the owner of the intellectual property (Kyle, 2007). Second, rivals may invent around patented innovations. I consider the imitation of an analogy process patent where the competitor has invented new manufacturing processes to produce the patented innovation.[2]

The pharmaceutical industry provides a good setup to analyze the economics effects of stronger intellectual property rights, as patents have been viewed to have an essential role in promoting pharmaceutical innovation (Grabowski, 2002, Gallini 2002). The process of bringing a single new drug to the market - from its discovery to marketing approval - involves substantial risks, is time-consuming and costs from around 500 million to 2 billion US dollars (DiMasi, 2003, Adams and Brantner, 2006 and 2010).[3] Because generic compounds are developed with substantially lower costs and risks than new drugs (see Grabowski, 2002), it is not surprising that much of the pharmaceutical innovation does not provide significant increments on existing innovations.[4]

[1] It can be asked why a patent holder does not grant licenses to its competitors. Licensing may, however, fail for several reasons, such as due to informational asymmetries between the patent holder and the rival about the rival's current and potential future profits (see Bessen and Manskin, 2009).

[2] Analogy process patents have been used often in countries where product patents for drugs are not available (Domeij, 2000). Process patents are also used in other industries. For example, the share of process patents on all patents in the manufacturing industry was around $24-30$ in 1970s% (see e.g. Cohen and Klepper, 1996).

[3] It has been estimated that less than 1% of compounds survive from pre-clinical period to human testing and only 20% of the compounds entering clinical trials gain the US Food and Drug Administration (FDA) approval (DiMasi, 1995).

[4] For example, the National Institute of Health Care Management reported that only 35% of new

I specifically focus on the Finnish markets for pharmaceuticals. In these markets, the Supplementary Protection Certificate (SPC) system provides rich variation in patent length that is usually fixed to 20 years from the filing date of the patent application. An SPC is an intellectual property right that extends the period of exclusivity from zero to five years, depending on the time needed to obtain marketing authorization.[5] Using heterogeneity in both patent length and breadth across innovations and data on the entry decisions of firms, I analyze how intellectual property rights affect the risks of parallel trade and generic entry, or imitation.

A large theoretical literature has analyzed how intellectual property rights change the nature of competition (Gallini, 1992 and 2002, Choi, 1998, Bessen and Maskin, 2009, Takalo, 1998). Still, this question has received surprisingly little attention in the empirical patent literature. The previous work has estimated the effects of patent strength on innovation (e.g. Hall and Ziedonis, 2001, Kortum and Lerner, 1999, Noel and Schankerman, 2006, Sakakibara and Branstetter, 2001, Bessen and Hunt, 2007, Baldwin et al., 2000, Moser, 2005) and patenting (Lerner, 2002, Hall and Ziedonis, 2001). The literature has also studied the importance of patent characteristics in the risk of it to be involved in infringements and invalidity suits (Lanjouw and Schankerman, 2001, Lerner, 2010, Cremers, 2004). Even though there is a large empirical literature on the determinants of entry in pharmaceutical markets (e.g. Morton, 2000, Kyle, 2006, Danzon et al., 2005), the effect of patent rights on competition during the patent period has not been previously studied, despite of its importance. If broader or longer patents do not prevent competition, welfare gains from policies that improve patent strength may be very limited.

My results suggest that patent breadth - measured by the number of claims - prevents imitation. To be more precise, the hazard rate of imitation decreases by $11-13\%$ when the patent breadth of an incumbent innovation increases by one claim. I find no evidence that patent length would increase the risk of imitation during patent protection. Patent length and breadth have no effect on the rate of parallel trade.

The remainder of the chapter proceeds as follows. Section 4.2 describes the institutional environment and the dataset. Section 4.3 presents the econometric model and discusses the identification assumptions. Section 4.5 goes through the estimation results. Section

products had new active ingredients and only 23% of those had sufficient clinical improvements over existing products to get a priority rating from the agency. Only part of this non-drastic innovation is imitation during the patent protection of an original innovation.

[5]In the US, the Drug Price Competition and Patent Term Restoration Act of 1984, commonly known as the "Hatch-Waxman" Act, permits extensions similar with SPCs to compensate for market lost during the drug approval process by the Food and Drug Administration (FDA).

4.6 concludes.

4.2 The institutional environment and the dataset

This section presents the institutional environment and the data. In the empirical analysis, I measure (maximum) patent length in two ways: number of years either from the patent application date or from the date of grant of the patent. To understand how the patent length from the applicate date is determined, I first discuss the patenting process in Finland. The granting process at the Finnish patent office (National Board of Patents and Registration in Finland) is broadly speaking similar with the processes at the European Patent Office (EPO) and the United States Patent and Trademark Office.[6] Then, I present the characteristics of patents and discuss briefly entry regulation for pharmaceuticals. Finally, I describe a sample of markets. A market is defined as an active ingredient (or their combination) of a medical product that is protected by a patent.

4.2.1 The Finnish patent system

A patent owner has the right to exclude others from the commercial utilization of the invention. To get the patent, firms and individuals must prepare and submit an application to the Finnish patent office.

The patent application is published 18 months after the date of filing or in the priority date that is the date of filing of the first application. Contrary to the practices of EPO, certain identification details, such as the name of the applicant, the application number and the filing date, are published in Finland immediately. Then, the applicant has 6 months to decide whether or not to pursue the application by requesting a substantive examination. Alternatively, the applicant who has requested the examination previously has to confirm whether the application should proceed.

After requesting the examination, the patent office examines whether the innovation meets certain requirements and whether the patent can be granted. The most important requirements are novelty and non-obviousness. Exclusive rights are not accorded to an old,

[6]For further information, see "How to apply for a European patent" by European Patent Office (2013) and "Patents" by National Board of Patents and Registration in Finland (2013). Van Zeebroeck et al. (2009), however, show that there are major differences across countries in patent drafting styles. Allison and Lemley (2002) also argue that heterogeneity in patent system across countries has increased over time.

previously known technique. Besides being novel, the innovation has to differ from known innovations and this difference has to be so big that the idea is not obvious.

Patent length

In Europe, the maximum length of a general patent is 20 years from the actual date of filing an application.[7] In order to keep the patent in force, renewal payments must be paid (see National Board of Patents and Registration for details).

In European Union member countries, the supplementary protection certificate (SPC) system provides an extension for a general patent. The SPC system was introduced to compensate for the long time needed to obtain a marketing authorization. SPCs are available for various pharmaceutical and plant protection products. A certificate application can be made for any medical product which is protected by a basic patent in a European Union member state and has received a marketing authorization in the same member state. Minor changes to a medicinal product, such as use of a different salt, an excipient and a presentation, do not justify a new SPC.[8]

The maximum length of the SPC for the general patent depends on the time needed to get the marketing authorization,

$$\text{Max length SPC}_i = \max\{0, \min\{(\text{MA}_i^{EEA} - \text{AP}_i) - 5 \text{ years}, 5 \text{ years}\}\} \tag{4.1}$$

where MA_i^{EEA} is the date of the 1st marketing authorization in the European Economic Area for the active ingredient associated with patent i and AP_i is the application date of the patent. SPC duration varies across innovations from zero to five years and hence the maximum length of a patent varies from 20 to 25 years from the application date.[9]

In the empirical part, I measure patent length in two ways: with number of years either starting from the filing date of the patent application or from the date of grant of the

[7] In the United States, for utility patents filed since June 8, 1995, the maximum length of a general patent is 20 years from the earliest filing date of the application on which the patent was granted. For patents filed prior to June 8, 1995, he maximum length is either 20 years from the earliest filing date or 17 years from the issue date, depending on which is longer (the United States Patent and Trademark Office, 2012).

[8] See Case C-431/04, Proceedings initiated by Massachusetts Institute of Technology.

[9] The duration of the SPC can be extended to 5.5 years when the SPC relates to a human medicinal product for which data from clinical trials conducted in accordance with an agreed Paediatric Investigation Plan have been submitted.

patent. The reason is that patent length since grant, rather than application, is likely to have more significant impact on competition during patent protection since uncertainty about the scope and timing of patent rights is narrowed after the patent has been granted.

Incumbents may also create patent clusters containing several patents (e.g. for a process or a reformulation) in order to extend patent protection. Then the patent associated with the SPC, that is observed in the data, provides a lower bound for the patent length of the medical product, including all its extensions.[10]

Claims

Patent claims are a part of the patent that define the breadth of patent protection and legal basis in technical terms. Innovators determine claims together with patent examiners. The breadth of the claims may be broad or narrow. Narrow claims are typically more specific about a particular element or a product than broad claims.[11] Broad claims are often more valuable than the narrow ones but they may be more difficult to obtain and to enforce because there can be other patents invalidating them. I follow Lerner (1994) and use the number of claims as a proxy for patent breadth.[12]

Other patent characteristics

A patent has several other characteristics, besides its length and breadth. First, an innovation can be patented in several countries. A set of patents in various countries for a single invention is called a patent family. Because a patent is costly to obtain and to keep in force, it is likely that the size of the patent family is higher for more valuable

[10]The anecdotal evidence from the quote of an originator company suggests that "Before end 80s: Products mainly NCEs which where [were?] protected by the one patent- [...] Late 80s early 90s[...] Expansion of the portfolio to cover lifecycle initiatives, to extend protection time for product and the brea[d?]th of the protection trying to keep competition further away". (The preliminary report of Pharmaceutical Sector Inquiry by European Commission.) Because the most of the innovations in the data are relatively old (the average application year is 1991), this may not be a big concern.

[11]See e.g. Soonwoo Hong: "Claiming what counts in business: drafting patent claims with a clear business purpose", the World Intellectual Property Organization.

[12]For example, Lanjouw and Schankerman (2001) suggested that the number of claims is an indicator of the complexity and value of an innovation. The number of claims can, however, reflect other factors besides breadth, such as resource constraints, patent application drafting style and effort and uncertainty about the value of an innovation (Allison and Lemley, 2002).

innovations (Harhoff et al., 2003, and Putnam, 1996). In the empirical part, I consider the number of patent family members for an innovation.

Second, the International Patent Classification (IPC) is a hierarchical patent classification system that is widely used to classify the technology of a patent. The highest hierarchical level of the IPC correspond to very broad technical fields (e.g. C denotes "Chemistry and Metallurgy"). Sections are divided into classes (e.g. class C21 denotes the "Metallurgy of iron"). Classes are again divided into more than 640 subclasses (e.g. class A21B denotes "Bakers' ovens; Machines or equipment for baking") and further into main groups and subgroups (the World Intellectual Property Organization). I follow Lerner (1994), Harhoff et al. (2003) and Lanjouw and Schankerman (2001) and use the number of the lowest level IPC classifications as a proxy for patent scope.

Third, a priority country is the country of the first application. The priority right allows the claimant to file a subsequent application in another country for the same invention with the date of filing the first application. Because the priority country is most often the country of an innovator, it can reflect regional variation in the costs of innovation.[13] I classify priority countries to Europe and non-Europe.

Fourth, a patent can cite previous patent documents. I consider the number of cited documents that help to evaluate how much the patent relies on past innovations. The number of backward citations may also correlate with the value of the patent (Hall et al. 2005, Harhoff et al., 2003).[14]

Pharmaceutical analogy process patents

I consider analogy process patents that have been commonly used for pharmaceuticals in many countries, where product patents have not been available.[15] Claims for an analogy process patent define manufacturing processes for a chemical substance. For a patent to be effective, the claims should include all feasible manufacturing processes. This may be very costly and time consuming, and often competitors have been found new ways to produce the drug. For example, competitors may have developed the new ways of synthesis. More often, they have made only small modifications, such as another pH, to existing innovations (Domeij, 2000).

[13]PutFME (1996) showed that priority country was the country of an innovator for 98% of US, 88% of German and 84% of French inventors.

[14]In the future, my plan is to use forward citations that indicates how much an innovation has contributed to the development of subsequent inventions (Lanjouw and Schankerman, 2001).

[15]In Finland, analogy process patents were granted for pharmaceuticals until the mid of 1990s.

For example, in the Danish court case on an analogy process patent for the anti-ulcer drug rantidine, the defendant's innovation had the same synthesis steps as in the original claims. The defendant had used somewhat different starting materials to create a process that was not defined in the claims. Then it added an additional reaction step to create the same final product, raniditine. Because the reaction step was new and unpredictable, the Danish Supreme Court decided the case for the favor of the defendant (Domeij, 2000).

Data exclusivity and entry regulation

In European pharmaceutical markets, data exclusivity has been supplementing intellectual property rights since 1994. During the data exclusivity period, an incumbent has exclusive rights to utilize research results associated with its marketing authorization. A competitor can still receive a marketing authorization on the basis of its own research results. Since the end of 2005, the length of the data exclusivity in the EU has been 8 years, and during the 10 years of marketing exclusivity, a generic product cannot be placed on the market.[16] For marketing authorizations submitted before the end of 2005, the data exclusivity granted to the original marketing authorization holder was either 6 or 10 years (6 in Finland) (European Commission, 2008). I control for changes in the data exclusivity regulation with the time trend.

The entry of pharmaceuticals is strictly regulated all over the world and procedures preceding it are fairly similar in Finland as in the USA. In order to enter the market, a firm must receive a marketing authorization for its product from the Finnish National Agency for Medicines (NAM) which corresponds to the FDA in the USA. To get a license for parallel importation, an original product must have authorizations in both the source and destination country. Parallel imported products must have the same chemical compositions, dosage forms and strengths as the original product in both countries.

[16]Time for the protection starts when the first authorization is granted in the EU area.

4.2.2 The dataset and patent characteristics

The dataset

The dataset is collected from several sources. I use the register data for pharmaceuticals that is provided by the Finnish Medicine Agency (FMA). The data contains information on the characteristics of a medical product such as its entry, exit and marketing authorization approval dates, the Anatomical Therapeutic Chemical (ATC) classification code of the World Health Organization,[17] and the indicator for whether the product is parallel imported. The data include all active ingredients that had valid marketing authorizations at some time point during $2003-2006$ and 2008. The data does not include information on products that both entered and exited markets during 2007.[18] I also use the dataset of FME on drugs that are substitutable in the generic substitution system in 28.1.2009.[19] The data contains information on the documentation type of a product, i.e. whether it the product is an incumbent or a generic.

The data for the SPCs of pharmaceutical process patents applied before January 2009 was obtained from the National Board of Patents and Registration in Finland.[20] The dataset includes the patent's identification number, application, granting and expiration dates and the name of the patent holder. With the identification number, I gathered information on the patent family, the number of claims and cited patent documents from Espacenet which is an international network of patent databases. I merged the patent data to the FME data by the name of the active ingredient. A firm who entered a market during patent protection with a generic product was interpreted as an imitator.

If a firm entered and exited a market before the FME data on substitutable drugs was compiled or a product was not substitutable, information on the documentation type was not available. I supplemented data on the documentation type by comparing the name

[17]The ATC Classification System is used for the classification of drugs. Different levels of the code indicate the following groups: 1.) 1st level: the anatomical main group, 2.) 2nd level: the therapeutic main group, 3.) 3rd level: the therapeutic/pharmacological subgroup, 4.) 4th level: the chemical/therapeutic/pharmacological subgroup, 5.) 5th level: the chemical substance.

[18]Data for year 2007 was not available in the web page of FMA at the time of data collection in 2008. The robustness analysis suggests that the results do not change if year 2007 is excluded from the sample.

[19]The generic substitution system was introduced in Finland on 1 April 2003. In the system, the prescribed medicinal product is substituted in a pharmacy by the cheapest, or close to the cheapest, generic alternative. Substitutable medicinal products contain the same quantity of the same active agent and are biologically equivalent.

[20]Pharmaceutical patents without SPCs are not included in the dataset. Excluded patents include the least valuable innovations which patents were not renewed.

of the SPC holder with the name of the marketing authorization holder. If it was the same, the marketing authorization holder was interpreted as the incumbent. If it was not, I tried to find information on a possible license agreement between SPC and marketing authorization holders from the U.S. Securities and Exchange Commission (SEC)[21] files, the web pages of pharmaceutical firms and Thomson Reuters Current Patent Gazettes.[22] If a firm entered a market before the expiration of a patent and information on an alternative patent was found, then the documentation type was denoted as an imitator. In 6% of markets, information on whether an entrant was an original innovator, a licensee or an imitator was not available.[23] In the empirical part, I analyze how long the incumbent's period of exclusivity is before either imitation or parallel importation or both occur. Therefore, I had to exclude from the data those markets where a firm with an unknown identity entered a market before an imitator.

Before presenting evidence on the role of patent rights in early competition, I provide descriptive statistics for the characteristics of innovations in Table 4.1. The average patent length from its application was 23 years and from its grant only 16 years, indicating long approval times. Importantly, patent length, measured either from the patent application or grant date, varied between innovations. For example, the variance of the average length from the application date was 13% of its mean. The number of claims was on average 19, with a high standard deviation across markets. The average number of patent family countries for a single innovation was 38, and the priority country of an innovation was most often in Europe. On average, a patent cited on average 16 patent documents and the mean number of inventors was 5. A second set of variables describe the other characteristics of active ingredients. The results show that 33% of products had restrictions in prescribing[24] and the share of "drastic" innovations (the first innovations in the chemical/therapeutic/pharmacological market)[25] was 25% in the sample of markets. Overall, the descriptive evidence illustrates that pharmaceutical innovations are very heterogeneous.

[21]SEC is responsible for enforcing the federal securities laws and regulating the securities industry, the nation's stock and options exchanges, and other electronic securities markets in the United States.

[22]Thomson Reuters is the world's leading source of intelligent information for businesses and professionals.

[23]Parallel importation is always observed in the data.

[24]For example, a specialized physician can prescribe certain drugs. See FMA for further details.

[25]This corresponds to the 4th level of the ATC code. Example: C10AA for HMG CoA reductase inhibitors (statins).

Table 4.1: Descriptive statistics for the sample of markets[1]

Variable	Mean	Std
Patent variables[2]		
Years from		
- patent application to SPC expiration	22.710	2.891
- patent grant to SPC expiration	15.589	3.561
Claims	19.045	14.611
Patent family size	37.639	39.056
Priority area:		
Europe (1: yes, 0: no)	0.529	0.501
International patent classifications	3.574	1.173
Cited documents	16.116	18.631
Inventors	5.465	4.023
Other characteristics		
Restrictions in prescribing right	0.329	0.471
The first active ingredient in ATC4 (1: yes, 0: no)	0.252	0.435
The share of markets with		
imitation:		
- All markets	0.056	0.231
- Years from the grant of a patent:		
shorter than the average	0.022	0.147
longer than the average	0.082	0.277
- Claims:		
less than the average	0.077	0.268
more than the average	0.000	0.000
The share of markets with		
parallel importation:		
- All markets:	0.224	0.419
- Years from the grant of a patent:		
shorter than the average	0.196	0.401
longer than the average	0.246	0.434
- Claims:		
less than the average	0.256	0.439
more than the average	0.138	0.351
The share of markets with competition[3]	0.271	0.447
Nbr of markets	107	

[1] Summary statistics are measured in 2008 for those markets where information on the number of claims is available.

[2] The maximum within a patent family for the following variables: international patent classifications, cited documents and the number of inventors.

[3] Competition: parallel importation or imitation.

Table 4.1 demonstrates that early competition in the Finnish markets for pharmaceuticals was very common. On average, the probability of imitation at any time point before patent expiration was 0.06 and the probability of parallel trade was 0.22.[26] The probability of early competition, i.e. either imitation or parallel trade, was 0.27. This means that either parallel trade or imitation but not both occurs in some markets.

Figure 4.1 illustrates how competition during the patent period shortens the effective patent life, i.e. the incumbent's monopoly period during the patent period. In markets with early competition, the effective patent life was on average very short (6.7 years), with substantial variation around the mean. Without competition, the average number of years since the entry of an incumbent to the expiration of a patent was 12 years. When only non-censored markets were examined, the average of the effective patent life in markets with competition was 8 years and without competition 11 years. The results suggest that early competition shortens the period of exclusivity substantially.

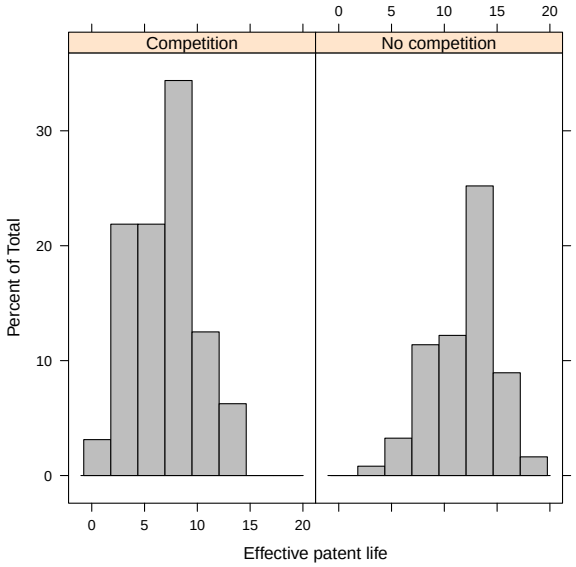

Figure 4.1: Effective patent life in years for markets with and without competition (entry of an imitator or a parallel importer) in the sample of markets

I then investigate how the prevalence of early competition depends on the length of a patent. The results of Table 4.1 suggest that the probability of early competition increases with patent length, measured from the application date. When a patent was longer than the average (referred as a "long patent"), 8% of markets had imitation and 25% parallel

[26] When also those markets that do not have information on the number of claims are included to the sample, 9% of markets had imitation and 20% parallel importation.

importation during the patent period. When the patent was shorter than the average, the probability of imitation was only 0.02 and the probability of parallel importation 0.20. When interpreting the results, it should be noted that the majority of markets (75%) were censored, i.e. the patent protection of an original innovation was still effective at the end of the observation period in 2008. When all 25 non-censored markets are analyzed, the risk of parallel trade was 0.33 and does not depend on the length of a patent since its grant. The probability of imitation remains to be much higher (0.13) in markets with a long patent than in other markets (0.08). Overall, these descriptive results indicate that an increase in patent length induces imitation during patent protection.

Finally, I examine how the number of claims affect the risks of imitation and parallel trade. When the number of claims was more than the average ("broad" patent), the results of Table 4.1 show that the rates of imitation and parallel trade during patent protection were both much lower than for innovations with narrow patents. These results may suggest that broad patents prevent competition during patent protection. The risk of parallel trade was higher than than risk of imitation in markets with narrow patents. This might indicate that parallel importers want to enter markets where the costs of imitation are high. When non-censored markets were investigated, the risk of parallel trade is lower in markets a high number of claims (0.29) than in markets where the number of claims is less than the average (0.35). The risk of imitation in non-censored markets is relatively similar between markets where the number of claims is less (0.10) or more (0.14) than the average.

To conclude, the descriptive results demonstrate that imitation and parallel trade are present in many pharmaceutical markets that are still under patent protection. The effective patent life of an incumbent remains often very short. The results suggest that the strengthening of patent rights decreases the risk of early competition and thus may help the incumbent to appropriate rents from its innovation efforts. Next, I use regression analysis that controls for the observed heterogeneity between innovations.

4.3 The econometric model and its identification

4.3.1 The econometric model

The setup of the econometric analysis is as follows. Each market i is at risk of experiencing imitation ($j = 1$) and parallel trade ($j = 2$). Let t_{ij} be the number of years from the entry of an incumbent to the jth event type (parallel importation or imitation) in

market i. Denote by T_{ij} the corresponding random variable that has the cumulative distribution function $P(t_{ij}) = P(T_{ij} \leq t_{ij})$. If patent protection is effective at the end of the observation period and competition has not yet occurred, the time of censoring t_{ic} is observed instead.

The hazard function for the occurrence of an event of type j (imitation or parallel importation) in market i is

$$\lambda_j(t_{ji}|\mathbf{x}_{jit}) = \lim_{\Delta t_{ij} \to 0} \frac{P[(t_{ji} \leq T_{ji} < t_{ji} + \Delta t_{ij})|T_{ij} \geq t_{ji}, \mathbf{x}_{jit}]}{\Delta t_{ij}}, \tag{4.2}$$

where $\mathbf{x}_{jit}$ is a vector of covariates that can vary over markets and time and can be different for imitators and parallel importers. The numerator of the hazard function is the conditional probability that the event occurs in the time interval $[t, t + dt)$, given that it has not occurred before, i.e. $T_{ij} \geq t_{ji}$. Dividing the denominator by the width of the interval gives a rate of event occurrence per unit of time. When the width of the interval goes to zero, the conditional probability becomes an hazard (instantaneous) rate of occurrence for the event.

The hazard rate $\lambda_j(t_{ji}|\mathbf{x}_{jit})$ of the event type j (imitation or parallel importation) is specified as follows,

$$\lambda_j(t_{ji}|\mathbf{x}_{jit}) = \lambda_{j0}(t) e^{\alpha_j \text{Length}_i + \gamma_j \text{Claims}_i + \tilde{\mathbf{x}}_{jit}\beta_j}, \tag{4.3}$$

where $\lambda_{j0}(t)$ is the baseline hazard function, Length_i is the patent length (either from the application or grant date), Claims_i is the number of claims and $\tilde{\mathbf{x}}_{jit}$ is a vector of control variables for market i at time t_{ij}.

I first investigate imitation and parallel importation as independent events. This means that I estimate separate Cox Proportional Hazard (CPH) models (3) for both event types. The parameters of the CPH model for a given event are estimated by maximizing the event-type specific partial likelihood function. The model is semi-parametric because the estimation of the coefficients of the explanatory variables does not require the simultaneous estimation of the baseline hazard function (see e.g. Cameron and Trivedi, 2005).

I then treat imitation and parallel trade as competing events and study how patent rights affect the combined risk of early competition. In this case, time T_{ij} is the effective patent life. The occurrence of imitation eliminates the market from the risk of facing parallel importation and the other way round. At most one complete duration for each market is

thus observed and the other event type censored.[27] As a market can face either imitation or parallel importation, there are two records of the data for the market at any time during the observation period until the entry of a competitor occurs. I estimate the stratified, or the grouped, CPH model (2) where I allow for the effects of patent length and claims to differ between event types. I estimate the parameters of the model by maximizing the partial likelihood that is a product of event type-specific partial likelihoods.

The model explains how the intellectual property rights of an incumbent innovation affect the risks of imitation and parallel trade during patent protection.[28] I am only interested in whether there is competition during the patent period, not in the identities or the number of competitors. Thus, the descriptive model is used instead of the structural approach (see e.g. Berry, 1992). The non-structural approach has often been used in the literature on pharmaceutical entry (e.g. Kyle, 2006, 2007, and Danzon et al., 2005) and in the studies of patent litigation (e.g. Lanjouw and Schankerman, 2001).

4.3.2 Identification

The empirical model has some strong assumptions. First, survival times are independent across markets, event types and time periods. The independence assumption is analogous with the assumption on the Independence of Irrelevant Alternatives (IIA) of the multinomial logit model. As in that case, the violation of the independence assumption in the data may bias the estimation of parameters.[29] To correct for the market level correlation and multiple events, I use the robust (sandwich) variance estimate (Cameron and Trivedi, 2005). The results must still be interpreted with caution.

I also make a common, but criticized, assumption that the transition time t_{ij} and the censoring variable are independent (Putter et al., 2007). This means that the hazard of censored markets can be represented by the hazard of markets that remain without competition, after controlling for observed characteristics. The independence assumption may be realistic when the end of the study period causes the censoring of observations. In the competing risks framework, censoring can also be caused by the appearance of an event (say, imitation), which prevents occurrence of the another event (parallel importation).

[27] In the data, there is one market where the entries of a parallel importer and an imitator occurs in the same year. The Efron approximation is used to deal with the tie (Efron, 1977).

[28] This also means that imitation does not occur if an incumbent introduces a generic alternative during patent protection through its subsidiary.

[29] As discussed by McFadden, Train and Tye (1981), the IIA property implies that error terms are independent random variables.

Even though the main event of interest is the first event, the other event is still competing. To put it differently, for the time to imitation (parallel importation), all markets where parallel importation (imitation) appeared first are censored. Under the independence assumption, the entry of an imitator does not change the risk of parallel trade. Importantly, the assumption implies that the effects of patent characteristics on the risk of imitation are the same for markets that are still without competition as for markets where parallel importation has occurred first. In practice, the independence assumption does not hold if the entry of a parallel importer intensifies competition and thus decreases the entry incentives of an imitator.

I assume that the patent rights of the incumbent innovation do not correlate with the decisions of early rivals to enter the Finnish market. The exogeneity assumption is not realistic if patent rights are strongly correlated with the (unobserved) value of the innovation across countries, firms and years. The positive correlation can arise if the incumbent negotiates a broader patent for the innovation that will produce higher expected profits in Finland. This profitability may increase the rival's incentives to enter the market during patent protection which biases the effect of the patent breadth on early competition upwards. If the litigation risk increases with patent breadth, the value of the innovation can be negatively correlated with the patent breadth. The effect of the patent breadth may thus be downwards biased. Respectively, the effect of the patent length on competition during the patent period can be either upwards or downwards biased. The negative correlation between profitability and the patent length arises if the marketing authorization process takes longer for the drug that produces serious adverse effects for patients. The bias upwards arises if regulators and a drug firm speed up the marketing authorization procedure when the value of the innovation is high.

There are several reasons to believe that the endogeneity of the patent rights may not be a serious issue. First, the incumbent may not be able to anticipate the profitability of its innovation at the time when decisions about the patent's length and breadth are made. This may happen if the value of the innovation is very uncertain.[30] Second, the patent breadth is likely to be determined by the technological and human capital (innovators) advances of the innovative firm. In the empirical model, the general level of technology and human capital are controlled by the first publication year of the patent and the time trend. Due to technological development and the long time difference between the development of the original innovation and early competition, these advances may not

[30]There is the large literature on the role of uncertainty in the demand for pharmaceuticals. See e.g. Crawford and Shum, 2005.

affect the entry incentives of imitators and parallel importers.[31] Third, the patent length is also determined by several other factors that are not probably highly correlated with the entry decisions of imitators and parallel importers to Finland, including 1.) legal deadlines which account for roughly one month, 2.) the monitoring and control of regulators, 3.) competition between regulatory agencies and 4.) the overall number of applications.

4.4 Estimation

This section presents the estimation results of the survival analysis in Table 4.2. The table contains CPH models for imitation and parallel trade (models 1-4) and Cox competing hazard (CCH) models for the hazard rate of early competition (models 5-6). The patent length is measured in two ways: number of years either from the patent application date (models 1,3,5) or from the date of grant (models 2,4,6) to the date of SPC expiration. In the CPH models, I explain the number of years from the entry of an incumbent either to patent expiration or to the specific event (imitation or parallel trade). In the CCH models, I consider the incumbent's exclusivity period during patent protection, i.e. number of years from the entry of an incumbent either to patent expiration or to early competition. In the second subsection, I evaluate the robustness of the results.

[31] For this particular reason, Morton (2000) used years on patent as an instrument for brand advertising before patent expiration in the analysis of its effect on generic entry.

Table 4.2: Estimation results for Cox hazard models in the sample of markets

	Imitation	Imitation	Parallel importation	Parallel importation	Competing risks	Competing risks
Model	(1)	(2)	(3)	(4)	(5)	(6)
Patent length (years),						
from grant		0.156		0.059	0.211	
		(0.179)		(0.063)	(0.126)	
from application	0.028		0.007			0.155
	(0.190)		(0.073)			(0.183)
Claims	-0.104*	-0.106	-0.010	-0.005	-0.072*	-0.086*
	(0.048)	(0.065)	(0.034)	(0.036)	(0.032)	(0.034)
Interaction with a parallel trade indicator						
Patent length (years),						
from application						-0.109
						(0.195)
from grant					-0.133	
					(0.134)	
Claims					0.047	0.055
					(0.035)	(0.035)
Year	-0.206**	-0.195**	0.069	0.059	-0.021	-0.015
	(0.074)	(0.075)	(0.064)	(0.066)	(0.063)	(0.063)
Publication year[6]	0.200*	0.206	0.027	0.026	0.011	0.012
	(0.100)	(0.135)	(0.046)	(0.044)	(0.036)	(0.036)
Patent family size	0.050	0.054	-0.004	-0.003	-0.002	-0.003
	(0.028)	(0.044)	(0.012)	(0.013)	(0.011)	(0.010)
Priority country in Europe (1: yes, 0: no)	-2.459	-2.330	0.223	0.315	-0.242	-0.355
	(2.011)	(2.294)	(0.478)	(0.490)	(0.397)	(0.393)
Cited documents	-0.009	0.002	-0.003	-0.002	0.004	0.002
	(0.028)	(0.025)	(0.022)	(0.022)	(0.020)	(0.020)
Nbr of inventors	0.166	0.170	0.009	0.009	0.060	0.056
	(0.092)	(0.113)	(0.082)	(0.077)	(0.045)	(0.045)
Nbr of international patent classifications	0.282	0.359	0.184	0.163	0.229	0.230
	(0.444)	(0.635)	(0.183)	(0.185)	(0.168)	(0.166)
First in ATC4-group (1: yes, 0: no)	-0.542	-0.574	0.518	0.457	0.171	0.181
	(1.730)	(1.517)	(0.467)	(0.458)	(0.424)	(0.437)
Prescribing restriction (1: yes, 0: no)	-1.029	-0.850	-1.446	-1.443	-1.314*	-1.345*
	(1.360)	(1.273)	(0.809)	(0.791)	(0.640)	(0.655)
AIC	48.041	47.619	208.937	208.209	245.846	248.184
R^2	0.015	0.015	0.009	0.010	0.007	0.006
Max. R^2	0.044	0.044	0.182	0.182	0.115	0.115
Num. events	5	5	24	24	29	29
Num. obs.	860	860	979	979	1912	1912

[1] The effective patent life starts from the entry of an incumbent.

[2] The value of the variable is its maximum number within the family of a patent.

[3] In the competing risks model, separate baseline hazard functions are allowed for imitators and parallel importers.

[4] Significance starts for the P-value: $P \leq 0.10$ (·), $P \leq 0.05$ (*), $P \leq 0.01$ (**), $P \leq 0.001$ (***).

[5] Standard errors are clustered at an active ingredient level (ATC5).

[6] Minimum within a patent family.

[7] The number of international patent classifications.

4.4.1 Discussion

The patent length

I first discuss the effects of the patent length on the risk of imitation. The results indicate that both variables for patent length have economically significant but statistically imprecise effects on the hazard rate of imitation (models 1-2 in Table 4.2). I find that one year increase in the patent length starting from the application increases the imitation rate by $100 \times (e^{0.243} - 1) = 28\%$. When the patent length from the grant increases by one year, the rate of imitation increases by 35%.

There are several reasons why the patent length from its grant has a higher effect on imitation than the length from the application date. First, the patent application is not published immediately after the application. Thus it can take time until imitators or parallel traders receive information about the innovation. Second, the length and breadth of a patent are likely to be uncertain at the beginning of the patent grant process (see e.g. Gans et al., 2008). Third, less valuable or novel innovations may have longer application times. The effect of the patent length from the application date may thereby be downwards biased.

The results of models 3-4 suggest that the effects of patent length variables on the rate of parallel trade are both economically and statistically insignificant. For example, one year increase in the patent length from the application increases the hazard rate of parallel importation by 0.7%. The standard error of the patent length coefficient is 10 times larger than the estimated coefficient. Corresponding to the previous results for imitation, the point estimate for the effect of the patent length from the grant is somewhat higher, 6%, but it is also statistically insignificant.

Coefficient estimates of the CCR models (columns 5-6) show that one year increase in the length of a patent from its application (grant) increases the hazard rate that the first competitor is an imitator by 17% (23%), but the effect is not statistically significant. The explanation for why these effects are smaller than those from the CPH models for imitation (columns 2-3) is that in many cases the first entrant is a parallel importer. The effects of both patent length variables on the hazard rate of parallel trade are again economically and statistically insignificant. Overall, these results suggest that a longer patent may induce imitation but it does not affect the risk of parallel trade.

The patent breadth

The parameter estimates for the number of claims provide important evidence on the entry deterrence effect of the patent breadth on imitation. To be more precise, the yearly hazard rate of imitation decreases by $11 - 13\%$ when the number of claims increases by one (models 1 and 2). This effect is also statistically significant at the 5% significance level.

As expected, the number of claims does not seem to affect the rate of parallel importation as its coefficient estimates for models 3-4 are close to zero and imprecise. The results from the competing risks models (models 5-6) suggest that the effects of the number of claims on the risk that the first competitor in an imitator is smaller than the corresponding effects in the models 1-2. Again, these results are driven by the observation that the first entrant is more often a parallel importer than an imitator.

Other characteristics

The results for the year -variable suggest that the rate of imitation has decreased and the rate of parallel trade has not change over years.[32] The other characteristics of the innovation do not change the risk of early competition. To be more specific, the coefficient estimates the patent family size are close to zero and do not differ statistically significantly from zero.[33] Correspondingly, the priority area does not affect the rates of imitation and parallel importation. The point estimates of the number of cited documents suggest that the number of cited documents and the number of inventors do not affect the rate of parallel trade and are almost zero. Finally, the number of IPCs has a statistically insignificant but positive effect on the rates of parallel trade and competition. For example, when the number of IPCs increases by one, the hazard rate of imitation increases by $19 - 25\%$.

The final set of variables measures the market size and cost factors that can affect profits. The results show that "drastic" innovations (the first product in an ATC4-group) are less likely to be imitated but face more parallel trade. Even though the coefficients of the drastic innovation indicator have economically significant magnitudes, they are

[32]To be more specific, the longer the time in years from the entry of an incumbent to the end of the observation period is, the higher the rate of imitation is.

[33]The results should not be interpreted as causal effects because the patent family size over time is likely to be endogenous. For example, competition during patent protection can decrease the expected profits of the incumbent and thereby decrease incentives to patent the original innovation in other countries.

imprecisely estimated. The rate of early competition is lower for products that can not be prescribed by all physicians than other products without such restrictions.

4.4.2 Robustness

In this subsection, I evaluate the robustness of the results concerning the effects of the length and claims on the rates of imitation and parallel trade. First, in 34% of markets, information on the number of claims is missing. If this sample selection to non-missing and missing observations is not random, the estimation results can be biased. This is particularly true if the information is missing for, say, less profitable innovations. To evaluate the importance of this, I study how the point estimates of both patent length variables change when the number of claims is not controlled. The results in Table 4.3 suggest that one year increase in the length of a patent from grant (application) increases the rate of imitation by 12% (10%) which is less than when claims are controlled. The corresponding coefficient estimate for the rate of parallel importation is 0.058 (0.041). The point estimates of both patent length variables are statistically insignificant. Overall, the results regarding to the effects of the patent length variables do not change much when the number of claims is not controlled. This suggests that the missing observations of claims do not bias the results.

Table 4.3: Estimation results for Cox hazard models without claims in the sample of markets

	Imitation	Imitation	Parallel importation	Parallel importation	Competing risks	Competing risks
Patent length (years),						
from grant		0.114		0.058	0.053	
		(0.069)		(0.050)	(0.080)	
from application	0.094		0.041			-0.013
	(0.077)		(0.068)			(0.069)
Interaction with a parallel trade indicator						
from grant					0.029	
					(0.086)	
from application						0.095
						(0.094)
Year	-0.234*	-0.239*	0.081	0.078	-0.035	-0.026
	(0.104)	(0.096)	(0.058)	(0.058)	(0.058)	(0.056)
Publication year[6]	0.044	0.036	0.000	-0.003	0.006	0.013
	(0.061)	(0.059)	(0.025)	(0.025)	(0.022)	(0.022)
Patent family size	0.007	0.006	-0.003	-0.003	0.002	0.002
	(0.020)	(0.020)	(0.010)	(0.010)	(0.008)	(0.008)
Priority country in Europe (1: yes, 0: no)	-0.564	-0.594	0.035	0.015	-0.285	-0.243
	(0.498)	(0.492)	(0.383)	(0.383)	(0.304)	(0.304)
Cited documents	-0.005	0.002	-0.008	-0.005	-0.005	-0.008
	(0.021)	(0.020)	(0.018)	(0.018)	(0.016)	(0.016)
Nbr of inventors	0.065	0.072	-0.035	-0.026	0.034	0.023
	(0.069)	(0.063)	(0.084)	(0.080)	(0.048)	(0.053)
Nbr of international patent classifications	-0.172	-0.173	0.253	0.273	0.077	0.061
	(0.237)	(0.253)	(0.154)	(0.153)	(0.148)	(0.138)
First in ATC4-group (1: yes, 0: no)	-0.234	-0.250	0.300	0.298	0.079	0.082
	(0.506)	(0.510)	(0.383)	(0.372)	(0.314)	(0.317)
Prescribing restriction (1: yes, 0: no)	0.174	0.099	-1.727*	-1.734*	-0.891	-0.860
	(0.708)	(0.709)	(0.768)	(0.757)	(0.505)	(0.513)
AIC	160.502	159.255	339.694	338.851	464.519	466.066
R^2	0.006	0.007	0.008	0.008	0.003	0.002
Max. R^2	0.084	0.084	0.180	0.180	0.130	0.130
Num. events	17	17	36	36	50	50
Num. obs.	1721	1721	1679	1679	3240	3240

[1] The effective patent life starts from the entry of an incumbent.

[2] The value of the variable is its maximum number within the family of a patent.

[3] In the competing risks model, separate baseline hazard functions are allowed for imitators and parallel importers.

[4] Significance starts for the P-value: P $\leq$ 0.10 (·), P$\leq$ 0.05 (*), P $\leq$ 0.01 (**), P $\leq$ 0.001 (***).

[5] Standard errors are clustered at an active ingredient level (ATC5).

[6] Minimum within a patent family.

[7] The number of international patent classifications.

In 5% of analyzed market, a patent owner did not renew its SPC before the expiration date. In these markets, the values of innovations are likely to be lower than in markets where the maximum duration of SPCs were reached. Because the value of an innovation

may be correlated with the entry incentives of imitators and parallel importers, both patent length variables may be endogenous. When markets, where patent owners did not renew their SPCs, were removed from the sample, the results did not change much (Table 4.4). These findings indicate that the bias caused by non-renewed SPCs may not be a serious issue.

Table 4.4: Estimation results for Cox hazard models, a SPC was not renewed in the sample of markets

	Imitation	Imitation	Parallel importation	Parallel importation	Competing risks	Competing risks
Patent length (years),						
from grant	0.297		0.039			0.201
	(0.249)		(0.070)			(0.133)
from application		0.222		-0.065	0.122	
		(0.342)		(0.102)	(0.218)	
Claims	-0.120*	-0.134*	-0.007	-0.013	-0.086**	-0.072*
	(0.047)	(0.054)	(0.036)	(0.034)	(0.033)	(0.031)
Interaction with a parallel trade indicator						
Patent length (years),						
from application					-0.147	
					(0.242)	
from grant						-0.144
						(0.144)
Claims					0.053	0.045
					(0.034)	(0.034)
Year	-0.199***	-0.216*	0.055	0.067	-0.012	-0.022
	(0.059)	(0.088)	(0.066)	(0.062)	(0.061)	(0.063)
Publication year[6]	0.047	0.038	0.023	0.021	0.007	0.009
	(0.063)	(0.066)	(0.044)	(0.047)	(0.037)	(0.035)
Patent family size	0.008	0.001	-0.003	-0.003	-0.003	-0.002
	(0.017)	(0.021)	(0.013)	(0.013)	(0.010)	(0.011)
Priority country in Europe (1: yes, 0: no)	-1.253	-1.394	0.292	0.214	-0.366	-0.263
	(0.931)	(0.855)	(0.492)	(0.486)	(0.399)	(0.398)
Cited documents	0.034	0.026	-0.003	-0.004	0.001	0.003
	(0.026)	(0.027)	(0.022)	(0.021)	(0.019)	(0.020)
Nbr of inventors	0.084	0.067	0.008	0.011	0.055	0.058
	(0.068)	(0.062)	(0.077)	(0.084)	(0.045)	(0.045)
Nbr of international patent classifications	0.223	0.172	0.137	0.144	0.195	0.203
	(0.418)	(0.350)	(0.188)	(0.187)	(0.167)	(0.170)
First in ATC4-group (1: yes, 0: no)	-0.994	-0.970	0.427	0.485	0.147	0.139
	(1.254)	(1.361)	(0.457)	(0.460)	(0.433)	(0.424)
Prescribing restriction (1: yes, 0: no)	-0.630	-0.730	-1.441	-1.402	-1.342*	-1.327*
	(1.329)	(1.396)	(0.791)	(0.816)	(0.658)	(0.642)
AIC	59.689	60.971	207.272	207.224	246.542	244.766
R^2	0.013	0.011	0.009	0.009	0.006	0.007
Max. R^2	0.050	0.050	0.187	0.187	0.119	0.119
Num. events	6	6	24	24	29	29
Num. obs.	977	977	936	936	1826	1826

[1] The effective patent life starts from the entry of an incumbent.

[2] The value of the variable is its maximum number within the family of a patent.

[3] In the competing risks model, separate baseline hazard functions are allowed for imitators and parallel importers.

[4] Significance starts for the P-value: P $\leq$ 0.10 (·), P$\leq$ 0.05 (*), P $\leq$ 0.01 (**), P $\leq$ 0.001 (***).

[5] Standard errors are clustered at an active ingredient level (ATC5).

[6] Minimum within a patent family.

[7] The number of international patent classifications.

As a firm enters and exits the market in 2007 or before 2003, information regarding to it is missing as the data combines entry information from the registers of years $2003 - 2006$ and 2008. For this reason, I first restrict the sample to contain years before 2007. Table 4.5 suggests that the results are again fairly similar with those presented in Table 4.2.[34] I next focus on markets where the incumbent's entry was from year 1990 onwards and take into account observation before year 2007. The effects of the both patent length variables on the rate of imitation become negative but remain statistically insignificant (Table 4.6). To explain this, an entry of a competitor must have happened at the early stage of patent protection because innovations in the subsample are fairly new. These findings suggest that the results are fairly robust to the missing data due to sample selection.

[34]Small variation in the data does not allow for a further restriction that takes into account only those markets where an incumbent entered a market from year 2003 onwards.

Table 4.5: Estimation results for Cox hazard models, year≤ 2006 in the sample of markets

	Imitation	Imitation	Parallel importation	Parallel importation	Competing	Competing
Patent length (years),					risks	risks
from grant		0.156		0.125	0.169	
		(0.179)		(0.085)	(0.133)	
from application	0.028		0.128			0.122
	(0.190)		(0.120)			(0.203)
Claims	-0.104*	-0.106	-0.028	-0.018	-0.071*	-0.078**
	(0.048)	(0.065)	(0.031)	(0.033)	(0.028)	(0.029)
Interaction with a parallel trade indicator Patent length (years),						
from grant					0.016	
					(0.146)	
from application						0.107
						(0.242)
Claims					0.027	0.018
					(0.031)	(0.030)
Year	-0.206**	-0.195**	0.044	0.051	-0.067	-0.078
	(0.074)	(0.075)	(0.080)	(0.076)	(0.066)	(0.069)
Publication year[6]	0.200*	0.206	0.008	0.006	-0.007	0.000
	(0.100)	(0.135)	(0.054)	(0.052)	(0.044)	(0.045)
Patent family size	0.050	0.054	-0.026	-0.025	-0.014	-0.018
	(0.028)	(0.044)	(0.018)	(0.019)	(0.014)	(0.015)
Priority country in Europe (1: yes, 0: no)	-2.459	-2.330	0.068	0.227	-0.592	-0.764
	(2.011)	(2.294)	(0.506)	(0.540)	(0.449)	(0.443)
Cited documents	-0.009	0.002	0.016	0.018	0.021	0.017
	(0.028)	(0.025)	(0.018)	(0.017)	(0.019)	(0.022)
Nbr of inventors	0.166	0.170	0.017	0.022	0.093*	0.079*
	(0.092)	(0.113)	(0.097)	(0.092)	(0.037)	(0.035)
Nbr of international patent classifications	0.282	0.359	0.368	0.364	0.473*	0.448*
	(0.444)	(0.635)	(0.210)	(0.222)	(0.201)	(0.192)
First in ATC4-group (1: yes, 0: no)	-0.542	-0.574	0.974	0.975*	0.572	0.537
	(1.730)	(1.517)	(0.518)	(0.494)	(0.498)	(0.506)
Prescribing restriction (1: yes, 0: no)	-1.029	-0.850	-2.061*	-2.021*	-1.511*	-1.572*
	(1.360)	(1.273)	(1.045)	(1.020)	(0.728)	(0.738)
AIC	48.041	47.619	146.875	145.864	171.707	173.503
R^2	0.015	0.015	0.018	0.019	0.013	0.012
Max. R^2	0.044	0.044	0.154	0.154	0.097	0.097
Num. events	5	5	18	18	22	22
Num. obs.	860	860	838	838	1636	1636

[1] The effective patent life starts from the entry of an incumbent.

[2] The value of the variable is its maximum number within the family of a patent.

[3] In the competing risks model, separate baseline hazard functions are allowed for imitators and parallel importers.

[4] Significance starts for the P-value: P $\leq$ 0.10 (·), P$\leq$ 0.05 (*), P $\leq$ 0.01 (**), P $\leq$ 0.001 (***).

[5] Standard errors are clustered at an active ingredient level (ATC5).

[6] Minimum within a patent family.

[7] The number of international patent classifications.

Table 4.6: Estimation results for Cox hazard models, year≤ 2006, an incumbent's entry in year≥ 1990 in the sample of markets

	Imitation	Imitation	Parallel importation	Parallel importation	Competing	Competing
Patent length (years),					risks	risks
from grant		-0.012		0.149	0.123	
		(0.125)		(0.096)	(0.137)	
from application	-0.286		0.198			0.055
	(0.203)		(0.152)			(0.243)
Claims	-0.237*	-0.240	-0.049	-0.039	-0.080*	-0.087*
	(0.112)	(0.137)	(0.033)	(0.034)	(0.033)	(0.035)
Interaction with a parallel trade indicator Patent length (years),						
from grant					0.049	
					(0.156)	
from application						0.209
						(0.297)
Claims					0.030	0.023
					(0.033)	(0.033)
Year	-0.830*	-0.872	0.019	0.012	-0.023	-0.020
	(0.404)	(0.501)	(0.098)	(0.109)	(0.091)	(0.080)
Publication year[6]	0.204*	0.186*	0.014	0.008	0.008	0.014
	(0.101)	(0.089)	(0.054)	(0.055)	(0.044)	(0.045)
Patent family size	0.015	0.000	-0.026	-0.027	-0.007	-0.010
	(0.015)	(0.019)	(0.019)	(0.020)	(0.012)	(0.012)
Priority country in Europe (1: yes, 0: no)	-3.505	-3.191	-0.314	-0.147	-0.767	-0.933
	(2.492)	(2.535)	(0.503)	(0.551)	(0.517)	(0.484)
Cited documents	0.066	0.078	0.017	0.019	0.020	0.018
	(0.048)	(0.058)	(0.019)	(0.016)	(0.021)	(0.025)
Nbr of inventors	0.163	0.142	0.060	0.071	0.066	0.052
	(0.095)	(0.090)	(0.065)	(0.067)	(0.051)	(0.048)
Nbr of international patent classifications	-0.866	-0.887	0.490*	0.479*	0.360	0.354
	(0.475)	(0.507)	(0.224)	(0.226)	(0.203)	(0.201)
First in ATC4-group (1: yes, 0: no)	-2.058**	-2.073**	0.623	0.682	0.333	0.246
	(0.768)	(0.778)	(0.557)	(0.549)	(0.532)	(0.534)
Prescribing restriction (1: yes, 0: no)	2.232	2.258	-2.149	-2.061	-1.231	-1.311
	(1.177)	(1.319)	(1.112)	(1.085)	(0.722)	(0.760)
AIC	38.739	39.229	133.308	132.599	155.141	155.607
R^2	0.015	0.015	0.018	0.019	0.009	0.009
Max. R^2	0.036	0.036	0.151	0.151	0.091	0.091
Num. events	4	4	16	16	19	19
Num. obs.	791	791	764	764	1510	1510

[1] The effective patent life starts from the entry of an incumbent.

[2] The value of the variable is its maximum number within the family of a patent.

[3] In the competing risks model, separate baseline hazard functions are allowed for imitators and parallel importers.

[4] Significance starts for the P-value: P ≤ 0.10 (·), P≤ 0.05 (*), P ≤ 0.01 (**), P ≤ 0.001 (***).

[5] Standard errors are clustered at an active ingredient level (ATC5).

[6] Minimum within a patent family.

[7] The number of international patent classifications.

4.5 Conclusions

I studied empirically whether the strengthening of the incumbent innovation's patent rights prevents competition during patent protection. I considered two competing events - imitation and parallel trade - that may decrease the profits of the innovator on R&D investments. I used data from the Finnish markets for pharmaceuticals that provides rich variation in both patent breadth and length across innovations.

The results indicated that patent breadth prevents imitation but does not affect the rate of parallel trade. The first conclusion is that patent policy makers should acknowledge that rivals respond to changes in patent rights. The second implication of the results is that patent breadth, rather than length, could be used to strengthen intellectual property rights.

Further empirical work is required to understand the role of patent rights in competition and ultimately in incentives to innovate. The next step could be estimate the welfare consequences of increasing patent strength. It would be also interesting to study empirically whether stronger patents delay follow-on innovation in general.

Bibliography

[1] Adams C. – Brantner V. (2006) Estimating the Cost of New Drug Development: Is It Really 802 Million Dollars? *Health Affairs*, Vol. 25, pp. 420-428.

[2] Adams, C. – Brantner, V. (2010) Spending on New Drug Development. *Health Economics*, Vol. 19, pp. 130-141.

[3] Allison, J. R. – Lemley, M. A. (2002) The Growing Complexity of the United States Patent System. *Boston University Law Review*, Vol. 82, pp. 77-144.

[4] Baldwin J. R. – Hanel, P. – Sabourin, D. (2000) Determinants of Innovative Activity in Canadian Manufacturing Firms: the Role of Intellectual Property Rights. *Working Paper*, No. 122. Ottawa: Statistics Canada.

[5] Bessen, J. – Maskin, E. (2009) Sequential Innovation, Patents, and Imitation. *Rand Journal of Economics*, Vol. 40, pp. 611-635.

[6] Berry, S. (1992) Estimation of a Model of Entry in the Airline Industry. *Econometrica,* Vol. 60, 889-917.

[7] Bessen, J. – Hunt, R. (2007) An Empirical Look at Software Patents. *Journal of Economics and Management Strategy*, Vol. 16, pp. 157-189.

[8] Cameron, A. – Trivedi, P. (2005) Microeconometrics - Methods and Applications. *Cambridge University Press*, New York, USA.

[9] Choi, J. P. (1998) Patent Litigation as an Information-Transmission Mechanism. *American Economic Review*, Vol. 88, pp. 1249-1263.

[10] Cohen, W.M. – Klepper, S. (1996) Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D. *Review of Economics and Statistics*, Vol. 78, pp. 232-43.

[11] Cohen, W.M. – Nelson, R.R. – Walsh, J.P. (2000) Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not). Working Paper no. W7552, *National Bureau of Economic Research.*

[12] Crawford, G. – Shum, M. (2005) Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, Vol. 73, pp. 1137-1173.

[13] Cremers, K. (2004) Determinants of patent litigation in Germany. *Centre for European Economic Research (ZEW) Discussion Paper*, No. 04-72, pp. 1-29.

[14] Danzon, P. – Wang, Y. R. – Wang, L. (2005) The Impact of Price Regulation on the Launch Delay of New Drugs. *Health Economics*, Vol. 14, pp. 269-292.

[15] DiMasi, J. – Hansen, R. – Grabowski, H. (2002) The Price of Innovation: New Estimates of Drug Development Costs. *Journal of Health Economics*, Vol. 22, pp. 151-185.

[16] DiMasi, J. (1995) Success Rates for New Drugs Entering Clinical Testing in the United States. *Clinical Pharmacology and Therapeutics*, Vol. 58, pp. 1-14.

[17] Domeij, B. (2000) Pharmaceutical Patents in Europe. *Kluwer Law International*, Cambridge, U.S.A.

[18] European Commission (2008) Pharmaceutical Sector Inquiry, Preliminary Report. *DG Competition Staff Working Paper.*

[19] Efron, B. (1977) The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, Vol. 72, pp. 557-565.

[20] Gallini, N. (1992) Patent Policy and Costly Imitation. *RAND Journal of Economics*, Vol. 23, pp. 52-63.

[21] Gallini, N. (2002) The Economics of Patents: Lessons from Recent U.S. Patent Reform. *Journal of Economic Perspectives*, Vol. 16, pp. 131-154.

[22] Grabowski, V. (2002) Patents and New Product Development in the Pharmaceutical and Biotechnology Industries. *Working Paper*, pp. 1-29.

[23] Gans, J. S. – Hsu, D. H. – Stern, S. (2008) The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays. *Management Science*, Vol. 54, pp. 982-997.

[24] Hall, B. – Ziedonis, R. (2001) The Patent Paradox Revisited: an Empirical Study of Patenting in the U.S. Semiconductor Industry: 1979-1995. *RAND Journal of Economics*, Vol. 32, pp. 101-128.

[25] Harhoff, D. – Scherer, F. M. – Vopel, K. (2003) Citations, Family Size, Opposition and the Value of Patent Rights. *Research Policy*, Vol. 32, pp. 1343-1363.

[26] Kortum, S. – Lerner, J. (1999) What is Behind the Recent Surge of Patenting?, *Research Policy*, Vol. 28, pp. 1-22.

[27] Kyle, M. (2006) The Role of Firm Characteristics in Pharmaceutical Product Launches. *RAND Journal of Economics* Vol. 37, pp. 602-618.

[28] Kyle, M. (2007) Pharmaceutical Price Controls and Entry Strategies. *Review of Economics and Statistics,* Vol. 89, pp. 88-99.

[29] Lanjouw, J. – Schankerman, M. (2001) Characteristics of Patent Litigation: A. Window on Competition. *RAND Journal of Economics,* Vol. 32, pp. 129-151.

[30] Lerner, J. (1994) The Importance of Patent Scope: An Empirical Analysis. *RAND Journal of Economics,* Vol. 25, pp. 319-333.

[31] Lerner, J. (2001) 150 Years of Patent Protection. *American Economic Review Papers and Proceedings,* Vol. 92, pp. 221-225.

[32] Lerner, J. (2010) The Litigation of Financial Innovations. *Journal of Law and Economics, University of Chicago Press*, Vol. 53, pp. 807-831.

[33] McFadden, D. - Train, K. - Tye, W. B. (1981) An Application of Diagnostic Tests for the Independence From Irrelevant Alternatives Property of the Multinomial Logit Model. *Transportation Research Board Record*, Vol. 637, pp. 39-46.

[34] Morton, F. (2000) Barriers to Entry, Brand Advertising, and Generic Entry in the US Pharmaceutical Industry. *International Journal of Industrial Organization*, Vol. 18, pp. 1085-1104.

[35] Moser, P. (2005) How do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World Fairs. *American Economic Review*, Vol. 95, pp. 1214-1236

[36] Noel, M. – Schankerman, M. (2006) Patent Thickets and Software Innovation: Theory and Evidence from a Panel of U.S. Firms. *CEPR Working Paper*, Vol. 5701.

[37] Nordhaus, W.D. (1969) Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change. *Cambridge, Mass.*, MIT Press.

[38] Putnam, J. (1996) The value of international patent rights. Ph.D. Thesis. Yale University, Yale.

[39] Putter, H. – Fiocco, M. – Geskus, R. B. (2007) Tutorial in Biostatistics: Competing Risks and Multi-State Models. *Statistics in Medicine*, Vol. 26, pp. 2389-2430.

[40] Sakakibara, M. – Branstetter, L. (2001) Do Stronger Patents Induce more Innovation? Evidence from the 1988 Japanese Patent Law Reform. *RAND Journal of Economics*, Vol. 32, pp. 77-100.

[41] Takalo, T. (1998) Innovation and Imitation under Imperfect Patent Protection. *Journal of Economics*, Vol. 67, pp. 229-241.

[42] United States Patent and Trademark Office (2012) Patent Term. *Manual of Patent Examining Procedure.*

[43] van Zeebroeck, N. – van Pottelsberghe de la Potterie, B. – Guellec, D. (2009) Claiming More: the Increased Voluminosity of Patent Applications and its Determinants. *Research Policy*, Vol. 38, pp. 1006-1020.

VALTION TALOUDELLINEN TUTKIMUSKESKUS
STATENS EKONOMISKA FORSKNINGSCENTRAL
GOVERNMENT INSTITUTE FOR ECONOMIC RESEARCH

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
P.O.Box 1279
FI-00101 Helsinki
Finland

www.vatt.fi

ISBN 978-952-274-100-4 (PDF)
ISSN 1795-3332