ORGANIZATIONAL FORM AND THE PROVISION OF QUALITY IN HEALTHCARE MARKETS

by

Dmitry Dubrovinsky

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2009

Abstract

The quality of healthcare is an important issue in any society. In the US the question "Which hospitals provide higher quality of care: Non-Profit (NP) or For-Profit (FP) ones?" is interesting both from the consumer and the policymaker perspective. The Arrow-Hansmann hypothesis states that NPs provide a higher level of quality, because of the lower incentive to exploit the opportunity to increase profits by reducing quality. Substantial empirical literature, however, rejects the main implication of this hypothesis: that quality of output is higher at NP organizations than at FP organizations. This thesis proposes a theory of information asymmetry in healthcare markets that is consistent with the empirical evidence and generates additional testable implications.

US hospital data on 2006-2007 mortality rates (a measure of quality) support the central implication of the model: FPs have a higher variance of quality than NPs; 2006 data on US emergency room waiting times support the second implication: NPs have longer waiting times than FPs. Three additional implications find support in the data as well.

Since the theory finds support in the data, I use it as a basis for policy recommendations. I review the various policy alternatives toward FP healthcare provision in the US. The theory developed herein supports government policies that encourage conversions of low-quality FPs into NPs.

The analysis is built on two innovations. In the model presented doctors-entrepreneurs self-select between the NP and FP sectors based on ability and not the degree of altruism, as in the previous literature. In the empirical analysis related to quality I use a newly constructed dataset

and test implications, not at the nation-wide level, as in the previous literature, but at the market level. The last feature makes my testing of NP theories the most direct to date. These innovations reveal a new picture of the healthcare markets that is in accordance with all of the existing empirical evidence that was previously regarded as contradictory.

Table of Contents

Appendices

List of Tables

List of Figures

Acknowledgements

I acknowledge the indispensable assistance and advice of my Thesis committee chair Ralph Winter, and committee members: Patrick Francois and Mariano Tappata; researchers who offered their valuable time and experience: Ambarish Chandra, Robert Evans, Leemore Dafny, Jill Horwitz, Charles Weinberg, Veikko Thiele and Linda Peritz; and my classmates who offered moral support and valuable advice: Alberto Romero, Pablo Moran, Victor Cui, Dror Hermel and Hamed Hasheminia throughout my work on this Thesis.

Dedication

To the memory of my grandfather whose insights about math and science seem truer and truer the longer I study; and to my loving family.

Chapter 1

Introduction

1.1 Motivation

Since the mid-1980s a substantial number of non-profit (NP) healthcare providers in the U.S. have converted to for-profit (FP) status. This process sparked the debate as to whether these conversions should be allowed. A number of well-publicized scandals contributed to the debate. One of the early large-scale scandals came to light in 1997 when a large FP chain of hospitals, Columbia/HCA, was found to practise fraudulent insurance billing (see Cutler and Horwitz, 2000 for details).

A decade later still more scandals are being uncovered. A New York Times article entitled "At Many Homes, More Profit and Less Nursing,"[1] describes in dramatic terms how a number of nursing homes in Florida decreased quality of care they offer to their patients after they were purchased by a group of investors, changing their status from NP to FP. Not only were regulatory violations observed, but a number of the elderly clientele died due to neglect. These events reinforced the side of the debate that opposes reliance on profit in healthcare and conversions to FP organizational form.

NPs are not immune to criticism. A Wall Street Journal report cited two nursing homes, one a NP and the other a FP, using anti-psychotic drugs to save on resources when caring for the elderly.[2]

These scandals show how acute the debate over FP healthcare provision is in the U.S. and how

[1] Duhigg, Charles (2007) "At Many Homes, More Profit and Less Nursing," New York Times, September 23, 2007.
[2] Lagnado, Lucette (2007) "Prescription Abuse Seen In U.S. Nursing Homes: Powerful Antipsychotics Used to Subdue Elderly; Huge Medicaid Expense," The Wall Street Journal, December 4, 2007.

important policy toward healthcare quality can be. The goal of any policy intervention is to induce provision of high quality of healthcare, while containing costs. The first step toward this goal is ensuring that the production of healthcare services is efficient, i.e. there is no resource waste, and every extra dollar put into the healthcare system is used to increase the quality and access to care for the patients (and not entirely for enrichment of the medical staff, or medical equipment producers).

The policy levers that are available to the U.S. government in the context of the composition of the healthcare market in terms of organizational forms are: discouraging FP provision, discouraging conversions from NP to FP, and eliminating the benefits of the legal status of NP organizations. The issue of the composition of the healthcare market in terms of organizational forms is important in the Canadian context as well, as there is an on-going debate about a shift to FP provision of healthcare in this country (see Deber, 2002).

The scope of the 1990s conversions varied across the different sub-sectors of the U.S. healthcare. Conversions among all three organizational forms (government owned, NP and FP) were observed. However, an overall shift toward FP healthcare is the most evident. In Rehabilitation Hospitals sub-sector, for example, the percent of FP hospitals jumped from 15% in the mid 1980s to above 55% in the late 1990s. In HMOs the percent of FP hospitals increased from 35% to just below 65%. Not withstanding the alarming headlines, the nursing homes is the only sub-sector that experienced a reduction in FP penetration from 75% to 65%. Within Acute Care Hospitals the increase was small: 16% to 17%. Although an increase in the percentage of FP facilities was evident in all sub-sectors, except for nursing homes, the total percentage of FPs in many sub-sectors remains small (late 1990s data): 19% (an increase from 2%) in Hospice Programmes, 5% (an increase from 1%) in Mental Health Centres, and 5% (an increase from 4%) in Residences for ED Children (Schlesinger and Gray, 2006). Currently, 20% of all hospitals in the U.S. are FP (Sultz and Young, 2009 p. 76).

The fears over a reduction in quality of care as a result of conversions from NP to FP status

are well grounded in the existing economic theory (Hirth, 1999; Glaeser and Shleifer, 2001). The existing economic theory, however, is not entirely supported by empirical evidence. Specifically, there are empirical studies showing FP hospitals to have higher quality of care, on average, than NP ones (Rosenau and Linder, 2003; Schlesinger and Gray, 2006). These findings contradict the current concerns over reduction in quality as a result of conversions. The goal of this thesis is to extend the existing theory of NP organizations in order to accommodate the existing empirical results (Chapter 2). I test the empirical implications of my theory on an original dataset (Chapter 3). Once my theory achieves empirical support, I use the insights of my model in order to inform the policy maker about the effect of each of the policy levers just discussed on the overall efficiency of healthcare delivery (Chapter 4).

In this introductory chapter, first I briefly review the history of the U.S. healthcare system and describe its current complex structure. Not all aspects of the current healthcare system are relevant to the economic analysis offered in this thesis, but this review is necessary to provide the context to my findings. Section 1.3 presents the existing economic theory of NPs. In section 1.4 I cite the existing empirical evidence on quality of NPs and FPs and argue that it is inconsistent with the existing theory. Section 1.5 describes my extension to the theory of NPs that is more consistent with the empirical evidence than the existing theory. Section 1.6 concludes the introduction.

1.2 Emergence and Structure of the Current Day U.S. Healthcare System

In 2006 16% of the U.S. population were uninsured, 5% had individual private insurance, 51% had employer-based private insurance and 28% of the U.S. population relied on government financing to pay for their healthcare needs (Bodenheimer and Grumbach, 2009 p. 6). Originally, the U.S. healthcare was administered entirely by physicians in private clinics on fee-for-service (FFS) basis

that patients paid out of their pockets. Whenever a more complex procedure was needed, and once such technologies became available, the physicians would refer their patients to hospitals. In the hospitals the very same physicians appeared as surgeons and treated their patients using the facilities of the hospital, again on FFS basis. On top of these fees the patients had to pay FFS to hospitals for each procedure granted by the hospital, e.g. laboratory tests, bed-day, all out-of-pocket.

Today, there is still a substantial number of physician offices and hospitals that work on FFS basis. However, currently it is by no means the universal structure, and even within the FFS system only a very small minority of patients pays directly out-of-pocket. In this section I review the evolution from purely FFS independent practices financed by patients out-of-pocket to the current day highly mixed structure of the U.S. healthcare system, which I describe in detail.[3]

During the Great Depression, the ability of the U.S. population to pay for hospital and physician services diminished. Many hospitals were not able to maintain themselves financially. As a solution to the crisis, hospitals and physicians joined forces in various localities to offer health insurance on a monopoly basis (one insurer for each state). These insurance schemes for hospital treatment became known as "Blue Cross", and for physician visits "Blue Shield" (the latter now operates mainly in California and is less popular overall). These insurance schemes were operating as NPs, as were the majority of the hospitals that established them (Bodenheimer and Grumbach, 2009).

Hospital management and physicians gradually realized that the risk of health incidents is negatively correlated with income. When they administered health rating on regional basis, all of the inhabitants of a region who wished to acquire insurance paid the same premium regardless of their income or health status (risk of disease). In this way the wealthier and healthier individuals subsidized the poorer and sicker ones, while hospitals and physicians benefited from the cross-subsidies as the demand for their services increased (they had no preference for the source of funds,

[3]The exposition of this section is based mainly on Bodenheimer and Grumbach (2009) and Barr (2007).

as long as their services were paid for).[4]

In such markets, where the wealthier and healthier clientele subsidized the poorer and sicker one, there was obvious advantage to entry by FP independent insurers. These insurers could cream-skim by offering lower premiums for healthier and wealthier individuals, thus, leaving the "Blue Cross" and "Blue Shield" with the sicker and poorer clientele. However, entry by the FP insurers happened only after the initial initiative by the hospitals and physicians (subsequently, as a response, the "Blue"s switched to individual risk-rating in setting the premiums).[5] With immediate out-of-pocket payments significantly below costs, the utilization of healthcare increased. This was the emergence of health insurance in the U.S. .[6]

The spread of the health insurance happened due to the following phenomenon. During WWII the U.S. government did not allow businesses to increase wages to their employees. However, fringe benefits were not considered as a part of "wage", thus unions could bargain for these benefits. Additionally, the U.S. government did not tax the fringe benefits. As healthcare is a major concern for any family, one of the main fringe benefits became health insurance (either directly administered by a large enough employer, or through a contract with a private insurer), which was implicitly subsidized by the government through reduced taxation. This subsidy induced the employees to buy insurance beyond their optimal amount (Barr, 2007).

In 2006 51% of the U.S. population were insured through an employer-adminisitered scheme.

[4]It may seem not incentive compatible to insure the poor. Indeed, a profit maximizing insurer could charge the highest fee that the wealthy are willing to pay and insure only them. In this way the insurer would earn maximum profits. However, in a NP system the only way to extract surpluses from the insurer is through claims payment. Recall that the medical staff worked on FFS basis and were the main beneficiaries from the claims payment. Hence, although insuring solely the wealthy generates highest surpluses, it does not allow their sufficient extraction. In order to extract the surpluses, one needs a frequent realization of the claim-state of the world. The poorer population provide higher frequency of pay-outs if they are included in the insurance scheme, as they experience more frequent incidents of disease. As the medical staff is paid on FFS basis, insuring the poor and making the wealthy subsidize them is the most efficient arrangement from the point of view of the medical profession who established the system.

[5]In June 1994, the requirement to organize the Blue Cross and the Blue Shield as a NP was eliminated. Subsequently, conversions of these insurance schemes to FP were observed as well (Claxton et al., 1997).

[6]This description is based, primarily, on Bodenheimer and Grumbach (2009).

The tax subsidy has never been cancelled, however, the ban on wage increases was lifted after the war. All of the employer administered insurance schemes, and the ones entitled "Blue" operated on FFS basis, which was extremely profitable for doctors, who had an incentive to expend utilization as much as their leasure-consumption trade-off dictated. The insurers were not able to use their bargaining power to control fees and utilization because of their less organized structure when dealing withe American Medical Association (AMA) that promoted physicians' interests (Blue Shields and Blue Crosses were directly controlled by hospitals and physicians). Aditionally, the contracts under the employer insurance schemes were such that any underwriting loss in a given year had to be covered by an increase of premiums in the following year, otherwise, few insurers were willing to insure healthcare (where there is a large moral hazard both on the side of the consumer and on the side of the provider who determines the occurance of the claim-payment state of the world itself).[7]

As an alternative to the FFS healthcare provision (both directly, or though insurance schemes), integrated systems of delivery developed in the U.S. as well. During the Great Depression a considerable number of large-scale construction projects were undertaken in the U.S. by large private contractors. The construction itself frequently took place in remote areas, as it was concerned mainly with infrastructure. In these areas there was a lack of hospitals and physicians. As a result, some of the private construction contractors sponsored the building and operation of a system of physician offices and hospitals that treated employees (and their families) for a fixed monthly payment by the employer per individual covered. This type of compensation for healthcare provision is called "capitation", and it combines both insurance and delivery and relies on the law of large numbers, i.e. a large number of "enrollees" for whom the capitations are paid. The service providers themselves (physicians, nurses, and medical staff in general) are compensated on a salary basis that

[7]Barr (2007) and Bodenheimer and Grumbach (2009).

may include bonuses (either positive or negative) for quality of service and budget handling.[8]

For example, the industrialist Henry J. Kaiser collaborated with Dr. Sidney Garfield in establishing a capitation-based healthcare delivery system that became known as "Kaiser Permanenete Health Plan". Under the system of capitation, the provider (who is an insurer as well) has an incentive to reduce utilization rather than expend it, as it is in the case of FFS compensation (where the provider and the insurer are separate). The bargaining power of the employer is needed to keep the quality of care high and the capitation rates low. Capitation-based provision was proven to have lower healthcare expenditures than FFS, which is of importance in the U.S., that strives to contain healthcare costs (Garber and Skinner, 2008). These systems of healthcare provision were called Prepaid Group Practice (PPGP) or "managed care" in general, and were made illegal in many localities, as the AMA pressured the legislators to outlaw these practices and physicians not to join them (Barr, 2007).

As the managed care form of healthcare provision (the PPGPs) where it was legal showed to be cheaper than the FFS system, in 1973 the U.S. passed the "HMO Act". According to this act, the U.S. government made PPGPs legal all over the U.S., required all employers who offer health insurance to offer a Health Maintenance Organizations (HMOs, as the PPGPs became to be called) as an option (where available). The U.S. government required all of the HMOs to be organized as NPs and offered subsidies for their establishment (Barr, 2007 pp. 91-93).

In order to qualify as an HMO a health services organization had to satisfy one of the following criteria in terms of its structure

- Staff-model HMO: The HMO owns hospitals and employs physicians. The physicians are paid fixed salary without any financial incentive to provide more care; hospitals are run on an annual budget. The insured directly, or through their employers, make premium (capitation) payments directly to the HMO. All patients are restricted to receive services from the HMO-

[8]The capitation-salary-bonuses system is popular within the National Health Service in the U.K. .

owned physician practices and hospitals (unavailable services are bought on FFS basis - it is the same for all HMO structures). The whole Staff-HMO as one unit bears all the risk by pooling the risks of the individual enrollees.

- Group-model HMO: The HMO does not own hospitals or hire physicians. The HMO acts as an insurer. It contracts with groups of physicians and chains of (or individual) hospitals. Providers bear the risk, as they receive capitations (reduced premiums). Within medical groups and (chains of) hospitals the risk can be shifted from individual staff to the institutional (or group) level through salaries to physicians and global budgets for hospitals within chains. Alternatively, the same risk shift (with adverse incentives) can be achieved through FFS physician compensation. A part of the capitation income covers these FFS, while a portion is directed to a separate pull, that is used at the end of the year for bonuses, or to cover extra costs (depending on whether the physicians went over budget when referring patients to covered external procedures).[9] Under the same risk-shifting-adverse-incentives policy hospitals are compensated on per-diem or per-admission basis. Finally, the capitations may be paid to individual providers directly, however, in this case the physicians in individual practices bear all the risk.

- IPA (Independent Provider Organization)-model HMO: Physicians form and run a corporation (HMO-IPA). The HMO-IPA contracts with its physicians directly either through (annual) capitations or through discounted FFS. The physicians bear all the risk, however, they are the main entrepreneurs here as well.

In the course of life of the managed care (after 1973), three additional structures of health

[9]Bodenheimer and Grumbach (2009) report that this arrangement, which encourages free-riding by one physician on the cost-saving attempts of others, is preferred by the medical profession to individual physician-level pools. Higher utilization (the free-riding incentive) not only increases physicians' compensation, it also protects them better against potential malpractice suits.

services organizations developed in the U.S. (Barr, 2007 pp. 94-97).

- PPO (Preferred Provider Organization): The insurer contracts with physicians and hospitals ("preferred providers"), who in turn offer discounts to the PPO-insured patients (the only incentive to use these providers). The physicians usually have a financial stake in the success of the PPO.

- POS (Point of Service): Care at HMO facilities costs to the insured (deductible) $\$10-20$ per visit. For care by "preferred providers" that are not a part of the HMO the patient pays $20-30\%$ of the fee (FFS type fees set by these physicians) out of her pocket. For care by any other provider the patient is responsible for $40-50\%$ of the charge. This is a triple option plan that has spanned out of managed care, however, it substantially weakens incentives to reduce utilization.

- PHO (Physician-Hospital Organization): An association of physician practises with hospitals, it does not provide "de jure" insurance, however, the system is capitation based (global annual budget if contract with large employers). If a PHO enrollee requires care that is not available at the PHO, the PHO must arrange for this care and cover its costs (the patient does not bear any extra costs). The PHO commits to service all of those enrolled (often insurers decide on treatment and this inefficiency is eliminated here). Caveat: it has been estimated that a PHO must include hundreds of physicians and more than one hospital for the "law of large numbers" to hold and the insurance to be viable.

Before the introduction of the "HMO Act" those U.S. citizens who were insured, or could afford it, were used to high-utilization healthcare. They viewed any attempt to reduce utilization by managed care organizations as a reduction in quality (this is an obvious conclusion if quality is measured through process of care measures[10]). This discontent is referred to as "managed care

[10]For a preliminary discussion of "process of care measures" see Section 1.4 of this chapter. For a detailed review

backlash". Due to political pressures, during the 1980s the requirements on attaining HMO status were loosened and the NP requirement was lifted.

FP HMOs and PPOs that compensate their physicians on FFS basis employ direct utilization control by having medical staff on the payroll of the insurer. These nurses and physicians approve treatments and decide when patients can be released from hospitals (this was one of the main contributors to the "managed care backlash").[11]

In 2003 2/3 of the HMO enrolment was at FP HMOs and the majority of PPOs were FP as well (Barr, 2007, p. 168). POS and PHO forms of managed care options seem to be much less prevalent than HMOs and PPOs. In 2005, for example, 3% of the U.S. employees were covered by conventional employer administered FFS health insurance plans, 21% were covered by HMOs, 61% by PPOs, and 15% by POSs (Sultz and Young, 2009 p. 244). The percentage of employees on a conventional health plan has decreased over time and it is at its historic low in 2005. Both HMOs and POSs peaked at first (whith HMOs always dominating in the percentage of enrolment), but their coverage fell over time in favour of PPOs, that are at their historic peak in 2005 (Sultz and Young, 2009 p. 244).

Table 1.1 breaks down the enrolment in managed care in the U.S. in 2008 into HMOs and PPOs. Table 1.2 shows the penetration of managed care overall. (Medicare is a U.S. government administered health insurance for those over the age of 65. The basic part of it is free. Additional parts: B, C, and D for drugs, are administered by private insurers, subsidized by the U.S. government, and partially paid for by the patients. Medicaid is a U.S. government funded and administered insurance for the poor.) The tables were taken from `http://www.mcareol.com/factshts/factnati.htm`

of those measures refer to Section 3.2 in Chapter 3.

[11] This phenomenon is also a source of many confusions and general inefficiencies of the system. As physicians typically fulfil two or more roles (treating physician for one network and panel physician, who oversees utilization, for another, which may repeat itself for more networks) procedures are not being performed in short enough time frames, which harms patients (Bodenheimer and Grumbach, 2009 pp. 65-66).

(accessed on 11 March 2009 at 13:44:30), and were, reportedly, based on reliable data sources.[12]

Table 1.1: U.S. Managed Care Enrolment 2008; Source:[12] `http://www.mcareol.com/factshts/factnati.htm`

Type of Care	Enrolment
HMO (1)	66.8 million
PPO	59.7 million
Total (1)	126.5 million

Table 1.2: U.S. Managed Care Penetration (numbers in millions); Source:[12] `http://www.mcareol.com/factshts/factnati.htm`

Segment	U.S. Population	U.S. Population (%)	Managed Care	Managed Care (%)
Medicare (2)(3)	44.2	14.6%	8.7	19.7%
Medicaid (4) (5)	58.9	19.4%	38.3	65.0%
Commercial	152.9	50.5%	101.9	66.6%
Uninsured (6)	47.0	15.5%	0	0%
TOTAL (7)	303.0	100.0%	148.9	49.1%

There is a discrepancy between the shares of HMO enrolment in the figures reported by Sultz and Young (2009) for 2005 and the numbers in Table 1.1. One way to reconcile this is to note that Sultz and Young (2009) report enrolment by employees only, while Table 1.1 reports the total number of people enrolled. The figures in Table 1.1 include the families of the employees as well

[12]Data for the table construction, according to `http://www.mcareol.com/factshts/factnati.htm` (accessed on 11 March 2009 at 13:44:30), came from:
(1) HealthLeaders-InterStudy, data as of July 2007. Total Managed Care Plans data reported in HealthLeaders-InterStudy press release dated December 19, 2007. HMO enrolment data published on 7/29/08 on Kaiser Family Foundation's State Health Facts web site (`www.statehealthfacts.org`).
(2) Centers for Medicare and Medicaid Services (CMS), data as of January 1, 2008. `www.cms.gov`
(3) Medicare Advantage Fact Sheet, The Kaiser Family Foundation, June 2007. `www.kff.org`.
(4) Urban Institute and Kaiser Commission on Medicaid and the Uninsured estimates based on data from Medicaid Statistical Information System (MSIS) reports from the Centers for Medicare and Medicaid Services (CMS), 2008. Data as of FY 2005.
(5) Medicaid Managed Care percentage data as of June 30, 2006. Medicaid Managed Care Penetration Rates by State as of June 30, 2006. Centers for Medicare and Medicaid Services, retrieved June 2007. Available at `http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/Downloads/mmcer06.pdf`
(6) Health Insurance Coverage: 2006, U.S. Census Bureau.
(7) Total U.S. Population data as of February 7, 2008, U.S. Census Bureau

and those purchasing the insurance independently and not through their employer as well as the employee numbers.

Presumably, the difference between the 148.9 million people in the U.S. who are enrolled in managed care (Table 1.2) and 126.5 million who are enrolled in HMOs and PPOs specifically (Table 1.1) accounts for the POS and PHO types of plans (22.4 million people). According to Table 1.2, 50.9% of the U.S. population are not a part of any of the managed care arrangements. Since 15.5% of the population are uninsured, it leaves 35.4% of the U.S. population who are treated outside of the managed care systems.

Although FFS compensation occurs in the managed care segment as well, outside of it this is the sole method of compensation. In general, very little oversight over utilization and expenses happens outside of the managed care system. Therefore, one can deduce that 35.4% of the U.S. population participates directly in a physician primary care market which is very similar to private investor-owned (i.e. FP) firms.

This calculation is somewhat biased. There are government administered delivery systems to the indigenous population, the armed forces and the veterans. This implies lower demand for traditional delivery than 35.4%. On the other hand, some of the uninsured are wealthy enough to demand FFS based services and pay for them out-of-pocket. Additionally, those covered by inadequate (to their taste) insurance may seek FFS services and pay out-of-pocket for them. All of these imply higher demand for traditional FFS delivery. For convenience purposes I refer to the demand in the FFS market as 35% of the population.

In the FFS market (the demand for which represents roughly 35% of the U.S. population), participants in which seek treatment outside of the managed care system, developed a phenomenon called "Physician Concierge".[13] These schemes work in the following manner. Patients pay $\$1,500-$

[13]This concept was first described in a New York Times article: Zipkin A. "The concierge doctor is available (at a price)," New York Times January 7, 2005.

2,500 to be included in physician's practice. This enrolment fee only provides access to care and not care itself. The physician may offer basic service through a health plan, for which an insurer is billed. Those who paid the "access fee" have the ability to receive a better service and are responsible for payment of the total fee. A part of the service fee may be reimbursed through an insurance firm. Additionally, the physician does not commit to accepting patient's insurance. Medical services delivered through this arangement tend to have a significantly better quality (at least higher utilization and patient satisfaction) than the services administered through the managed care system (Barr, 2007, pp. 183-184).

Beginning in the 1970s, with peak activity in the 1990s, physicians in many localities across the U.S. have been opening speciality hospitals that treat patients only with a specific diagnosis. These hospitals are usually organized on a FP basis and perform more profitable procedures (Barr, 2007, p. 181). The hospitals in Table 1.3 are such hospitals in the data I use (2006-2007) in the empirical analysis in Chapter 3. Hence, within the hospital market we find phenomena similar to "physician concierge". Both at the primary practices level, and at the hospital level we are encountering high quality FP providers. This regularity goes against existing economic theory, as I explain in section 1.3, and constitutes the central phenomenon that must be explained by the theory I propose.

In order to apply rigorous empirical analysis to questions concerning quality of care, one must measure quality of hospitals. Chapter 3 of this thesis discusses such measures in detail. The most satisfying, and commonly used in the literature, measure of hospital quality are risk-adjusted (standardized by patient characteristics) mortality rates (RSMR) for Heart Attacks (Acute Myocardial Infarction, or AMI) (see McClellan and Staiger, 2000; Shen, 2002). I use 30 day post-admission RSMR for Heart Attacks as a measure of hospital quality in the discussion below. This is also the main measure utilized in the tests reported in Chapter 3.

Table 1.3 presented 32 FP hospitals that have the highest quality of care in their market (i.e. have higher quality than *all* of their NP counterparts). Most of them are "speciality" hospitals

that focus on heart surgeries. These hospitals have lower mortality rates for Heart Attack (AMI) diagnosis (computed from 2006-2007 data) - the measure of quality, than the lowest mortality for the same procedure at a NP hospital in their Metropolitan Statistical Area (MSA), which I adopt as the working definition of a market.[14,15]

As the next section shows, existing economic theories are unable to predict the existence of FP hospitals that have higher quality of care than their NP counterparts. This empirical regularity is important for answering the question whether FP healthcare provision should be discouraged by means of government policy. Therefore, a new theory of NPs is needed that would be consistent with these examples. I propose such theory in Chapter 2 of this thesis.

To sum up, there are four main ways of healthcare delivery in the U.S. today, although the system is extremely diverse at every level.

- FFS paid directly out-of-pocket - the oldest system. It still exists today and is represented by "physician concierge" and speciality hospital services at the high end and, presumably, even more varieties at the lower end.

- FFS covered by an insurer, without insurer's oversight of the procedures - the oldest private insurance scheme. In 2005 only 3% of the U.S. employees were a part of this delivery system. Yet more individuals purchase such insurance plans privately (Gruber, 2008).

- FFS covered by an insurer with insurer's oversight of the process. The most common today, as it accounts for all the PPOs (61% of the U.S. employees in 2005) and most of the HMOs. It is wasteful in terms of resources and most confusing, as two physicians are needed for each procedure: one to perform and the other to oversee. The same physician may be overseeing one procedure and performing another at the same hospital/practice, as "networks" overlap.

[14] I discuss the market definition in detail in Chapter 3.

[15] Data from the U.S. Department of Health & Human Services, Hospital Compare that was used in the empirical investigation reported in Chapter 3 of this thesis.

Table 1.3: Highest Quality (Lowest Mortality for Heart Attacks) FPs in their MSA[15]

Hospital Name	MSA Name	Mort	Min NP Mort MSA
HEART HOSPITAL OF NEW MEXICO	Albuquerque, NM MSA	12.9	13.5
RAPIDES REGIONAL MEDICAL CENTER	Alexandria, LA MSA	16	16.4
AIKEN REGIONAL MEDICAL CENTER	Augusta–Aiken, GA–SC MSA	16	16.5
HEART HOSPITAL OF AUSTIN	Austin–San Marcos, TX MSA	13.1	13.4
BAKERSFIELD HEART HOSPITAL	Bakersfield, CA MSA	15.7	16
HARLINGEN MEDICAL CENTER	Brownsville–Harlingen–San Benito, TX MSA	15.3	15.4
VALLEY REGIONAL MEDICAL CENTER	Brownsville–Harlingen–San Benito, TX MSA	15.3	15.4
DOCTORS HOSPITAL	Columbus, GA–AL MSA	15.9	16.65
PLAZA MEDICAL CENTER OF FORT WORTH	Dallas–Fort Worth, TX CMSA	13.9	14.4
DAYTON HEART HOSPITAL	Dayton–Springfield, OH MSA	14.5	14.6
FLAGSTAFF MEDICAL CENTER	Flagstaff, AZ–UT MSA	15.7	15.8
SUMMIT MEDICAL CENTER	Fort Smith, AR–OK MSA	15.9	16.7
ST VINCENT HEART CENTER OF INDIANA, LLC	Indianapolis, IN MSA	13.9	14.1
HEART HOSPITAL OF LAFAYETTE	Lafayette, LA MSA	14.4	15.3
LAKE WALES MEDICAL CENTER	Lakeland–Winter Haven, FL MSA	15.5	16.1
HEART OF LANCASTER REGIONAL MEDICAL CENTER	Lancaster, PA MSA	16.3	16.5
SUMMERLIN HOSPITAL MEDICAL CENTER	Las Vegas, NV–AZ MSA	15.7	15.8
ARKANSAS HEART HOSPITAL	Little Rock–North Little Rock, AR MSA	12.5	14.7
LOS ROBLES HOSPITAL MEDICAL CENTER	Los Angeles–Riverside–Orange County, CA CMSA	13.1	13.2
LUBBOCK HEART HOSPITAL LP	Lubbock, TX MSA	14.9	14.95
MOUNT SINAI MEDICAL CENTER	Miami–Fort Lauderdale, FL CMSA	12.8	13.5
CAMDEN CLARK MEMORIAL HOSPITAL	Parkersburg–Marietta, WV–OH MSA	16.1	16.2
HOSPITAL METROPOLITANO DR TITO MATTEI	Ponce, PR MSA	15.6	16
ST MARK'S HOSPITAL	Salt Lake City–Ogden, UT MSA	15.1	15.2
TEXSAN HEART HOSPITAL	San Antonio, TX MSA	13.3	13.7
VENICE REGIONAL MEDICAL CENTER	Sarasota–Bradenton, FL MSA	15.3	15.4
LAKEWOOD RANCH MEDICAL CENTER	Sarasota–Bradenton, FL MSA	14.5	15.4
SAINT FRANCIS HOSPITAL SOUTH	Tulsa, OK MSA	15.1	15.26
ASPIRUS WAUSAU HOSPITAL	Wausau, WI MSA	17.3	17.4
JFK MEDICAL CENTER	West Palm Beach–Boca Raton, FL MSA	14.1	14.4
KANSAS HEART HOSPITAL	Wichita, KS MSA	15.2	15.3
YAKIMA REGIONAL MEDICAL AND CARDIAC CENTER	Yakima, WA MSA	13.7	14.4

- Integrated systems that hire physicians and other medical staff directly on a salary basis, own hospitals, and operate them within global budgets. The enrollees (or their employer) pay premium directly to the system. Such premium is referred to as "capitation" in the medical jargon. This seems to be the most efficient method of healthcare delivery from the economic perspective. Incidentally, it is also the most opposed to delivery system by physicians' interest groups in the U.S., such as AMA. Integrated delivery systems are most common in the industrialized world today. In other OECD countries these delivery systems are usually universal. Universality increases economic efficiency even further due to the "law of large numbers" associated with the insurance function of the system.

1.3 Existing Economic Theories of Non-Profits

The main reason for opposition to conversions from NP to FP organizational forms in healthcare provision is the belief that FPs have an incentive to reduce quality. This fear is supported by the existing economic theory. Arrow (1963), Hansmann (1980), Hirth (1999) and Glaeser and Shleifer (2001) all predict that in markets with uncertainty of information about quality, which is a typical characteristic of healthcare markets, FP firms will shirk on the quality of their output. In this section I summarize the existing economic theory of NPs. The most relevant for my goals summary of current economic theory and its performance in terms of data can be found in Malani, Philipson, and David (2003), on which my exposition is based. Additionally, Glaeser (2003) gives an overall overview of NPs from strictly economic perspective that is relevant for this study.

First, I introduce the main distinctions between NP and FP firms. Keeping in mind that the economic theory is rich with modelling the latter, one has to find compelling reasons why the existing FP theories cannot be directly applied to the study of NPs. NP firms by definition are not allowed to distribute profits. They do not have residual claimants or owners the way FPs do, and

their operations are controlled by a board of trustees who appoint a manager (CEO).

Although portions of NPs' budgets come from sales of goods and services, unlike FP firms, they also receive donations. In some sectors these donations constitute large portions of NPs' revenue: 68% in International NPs and 57% in religious NPs in the U.S. in 2000 (Boris and Steuerle, 2006), however, it is not clear how important donations are for the NP sector overall, as these sub-sectors are small. The two largest sub-sectors of the NP sector in terms of total revenue are education and healthcare. Education and healthcare received 17% and 4% donations as percentage of revenues in 2000, respectively (Boris and Steuerle, 2006). Healthcare is the largest among the two, and is the focus of my study. Although donations may be important for other volunatry organizations, they are not an important factor for NPs that operate in the healthcare sector.

NPs in the U.S. are tax exempt.[16] Hospitals specifically are exempt from corporate income taxes, state and local property taxes and have access to tax-exempt bond financing. Empirically, these tax exemptions do not seem to give NPs overall advantages when competing with FPs for investors' funds, or in operations (Gentry and Penrod, 2000). Therefore, the tax asymmetry between NP and FP organizations does not seem as an important factor one has to take into account when studying the performance of NP healthcare providers.

In summary, there exist a number of conceptual distinctions between NPs and FPs, that may be important from theoretical perspective. The inability to distribute profits rules out profit maximization as the objective function for NPs. Donations and tax exemptions have the potential of explaining direct advantages of the NP organizational form. Although these may be important in other sectors, in healthcare neither seems to be important for explaining NPs. Due to apparent little empirical importance of tax asymmetries and donations in healthcare, my own model which I review in section 1.5 of this chapter, takes as the main characteristic of NPs their inability to

[16] Although this is generally true, Claxton et al. (1997) report that many NP HMOs, Blue-Shield and Blue-Cross plans are not subject to the tax exemption under section 501(c)(3) of the U.S. Internal Revenue Code, which usually grant such exemption to voluntary organizations.

distribute profits.

There are three paradigms for analysing NP organizations. These paradigms are: "Altruistic", "Employee Cooperative" and the Arrow-Hansmann hypothesis. Research in Contract Theory on altruistically motivated agents (employees) provides a framework that may be classified as a fourth paradigm (Francois and Vlassopoulos, 2008). The main postulates of the "altruistic agents" approach are somewhat similar to the more traditional Altruistic paradigm, while its implications about the effect of organizational form (NP/FP) on quality are in line with the Arrow-Hansmann hypothesis. For this reason, although I review this literature as well, I am reluctant to call it a fourth paradigm explicitly.

The "altruistic agents" stream of literature is important to review because my own model stems from it. I employ the altruistic approach of modelling employee (physician) preferences. I abstract from the agency conflict within organizations, however, in order to focus on the agency conflict between the patient and the provider. My model is closer to the Contract Theory literature than the more traditional Altruistic paradigm of NPs. However, my model adds to the Contract Theory literature the market interaction between firms. In what follows I review the literature on the three paradigms of NP theory and the fourth application of "altruistic agents"-Contract Theory to the theory of NPs.[17]

The Altruistic paradigm of NP theories was pioneered by Newhouse (1970). This paradigm postulates that quantity or quality of output enter directly, as arguments, into the objective function of a NP organization. Those in control of the organization have preferences over quality or quantity of output. Since consumers are the ones who benefit from both quality and quantity of output, theories stemming from this paradigm are called "altruistic". Altruistic theories of NPs assume that the manager maximizes usage (or quantity) and quality of output (weighted in a certain

[17]The exposition of the first three paradigms of NP theories is based mainly on Malani, Philipson, and David (2003), while my overview of the "altruistic agents"-Contract Theory approach follows Francois and Vlassopoulos (2008).

manner) with respect to quantities of inputs (and the price of the output if a monopoly) subject to a break-even constraint that has to be binding.[18]

What are the market-level implications of the Altruistic paradigm? Newhouse (1970) postulates that NPs enter markets first and provide either output of higher quality than is feasible by the FP sector, or at prices below average costs, that are subsidized through donations. However, the capacity of the NP sector tends to be constrained either by the scarcity of entrepreneurs with altruistic motivations, or scarcity of donations. The demand grows and once it exceeds NP capacity FPs start entering these markets and meet the extra demand.

Empirical research seems to support the predictions of the Altruistic paradigm (Horwitz and Nichols, 2007; Liu and Weinberg, 2009). However this approach offers too little implications for relative quality of NPs and FPs. As the marginal entering firm is a FP, it could be the case that it provides the same quality as the NPs, however, it could also provide lower quality than the NPs (see Malani, Philipson, and David, 2003). On the other hand, it is unlikely that it provides output of higher quality than NPs, as in this case it would have entered the market beforehand.[19]

The Employee-Cooperative paradigm of NPs was introduced by Pauly and Redish (1973). These authors view a NP organization (they focus on a hospital) as a co-operative[20] of employees (physicians). Due to lack of residual claimant who is interested in resource allocation that increases her wealth, the physicians who operate the hospital will take charge. The superior knowledge of physicians with respect to treatment techniques enables them to maximize their own objectives at the expense of the board of trustees that initially set up the hospital. This approach seems to be confirmed by casual observations.

Glaeser (2003) argues that any NP at first is dominated by donors, who exercise control to

[18] See Weinberg (1980) and Liu and Weinberg (2009) for examples of such modelling approach.

[19] The duopoly model in Liu and Weinberg (2009) focuses on maximization of quantity as NP's objective, thus, it has no implications for quality. However, it can be extended to incorporate the quality dimension of output.

[20] Sometimes referred to as a "cartel" in less flattering summaries.

make sure that their donations are utilized in their best interests (this part follows the altruism paradigm). However, as the NP becomes wealthier, it de-facto shifts to control by employees (or top employees, perhaps even managers). Glaeser (2003) provides examples of NP universities and even the Catholic Church as NPs outside of the healthcare sector that became dominated by their employees. However, the empirical implications generated by the employee-control paradigm do not seem to be convincingly supported by the data. Additionally, there are no strait-forward implications for differences in quality provision between FPs and NPs (Malani, Philipson, and David, 2003).

Two extremely influential articles by Arrow (1963) and Hansmann (1980) gave birth to a third paradigm of economic theorizing on NPs. Arrow-Hansmann hypothesis says that in some markets products have unobservable and non-contractible quality aspects. If provision of these aspects of quality is costly, profit oriented firms will shirk in their provision. Every dollar saved as a result of quality reduction (less taxes) accrues to the owners as profits. On the other hand, NP organizations, due to their inability to distribute any surplus, have weaker incentives to lower costs by reducing quality. Thus, consumers will be more willing to make purchases from the latter. Neither of the fathers of the paradigm had provided mathematical models of the phenomena. However, this paradigm is, potentially, the most helpful for the current policy debate on FP organizations in healthcare. Arrow-Hansmann hypothesis can aid at determining the relationship between ownership form (NP or FP) and quality of firm's output because it addresses quality directly.

Glaeser and Shleifer (2001) analyze the Arrow-Hansmann hypothesis in a rigorous model of a monopolistic NP. The authors postulate an own-income maximizing entrepreneur and show that if she opens a NP firm, the organizational form can be used as a signal of quality to the consumer. As a result, the NP can charge a price that is higher than the one a FP monopolist would be able to charge in such circumstances (consumers have unit demands). The entrepreneur can extract the operating surplus of a FP in cash, while in the case of a NP only through perquisites. Although the

entrepreneur values the perquisites less than cash, the price premium of the NP (over the optimal price of a FP monopolist) generates operating surplus that is enough to compensate her for inability to extract the surplus in cash. In this way Glaeser and Shleifer (2001) show the plausibility of the Arrow-Hansmann hypothesis in a rigorous modelling setting. However, their model highlights an additional implication: if the entrepreneurs maximize their own income, the NP output must sell at a premium relative to the FP one (this implication will be instrumental as well in rejecting the model, as the next section shows).

Hirth (1999) follows the tradition of the Arrow-Hansmann hypothesis. He, however, recognizes that if one is to make any predictions on the effect of organizational form on quality in a mixed market (in which NP and FP firms operate side by side, as it happens in healthcare), one has to model interactions between the two ownership forms explicitly, and not each entity in isolation. Additionally, Hirth (1999) argues that in reality it is likely that there exists a *variation* in information asymmetry about quality between producers and consumers. Hirth (1999) shows that NPs provide weakly higher quality in a mixed market populated by both informed and uninformed consumers. Additionally, the paper predicts that entry by NPs increases the average quality provided in the FP sector.

As my model assumes a variation in the information asymmetry (between consumers and producers) and captures the interaction between organizational forms, it is important to emphasize the distinction between my model and the one developed by Hirth. Hirth (1999) allows the producers to choose only between two levels of quality. He studies how exogenously determined organizational form impacts firm's choice among the two levels. My model adds three extra dimensions: 1) Endogenous choice of organizational form, 2) Unobservable provider type that affects the cost of quality provision, 3) Choice of quality along a continuum. In section 1.4 I show that the empirical implications of Hirth (1999) are rejected by some empirical research. By construction, Hirth (1999) is both unable to predict higher observed quality of FP firms, or to explain why an entrepreneur

chooses a NP or a FP organizational form. Therefore, the three extra dimensions are needed for consistency with the empirical evidence and an "explanation" to "why" NPs arise.

Francois and Vlassopoulos (2008) summarize a stream of literature in Contract Theory which combines the altruism hypothesis at the employee level with a phenomenon that resembles the Arrow-Hansmann hypothesis. Bénabou and Tirole (2006) identify two types of altruism: "pure" - task oriented, and "impure" - effort oriented. In the case of impure altruism, an economic agent derives utility directly from effort exertion, while purely altruistic agent cares only about output. The latter will exert effort only as long as the marginal contribution to output exceeds her marginal cost of effort (when measured in units of output). The distinction matters when a principal is able to substitute agent's effort for other inputs.

The altruistic-agent branch of Contract Theory predicts that purely altruistic employees, who care about output, will shirk on effort if the worker knows that the residual claimant is able to substitute her effort for other costly inputs (Francois and Vlassopoulos, 2008). The worker will not provide effort at the efficient level, as increments in her effort will not translate into higher output at the marginal product of effort, but will accrue as increased profits (due to cost savings) to the residual claimant. A firm that cannot distribute profits (a NP) will have a lower incentive to substitute away from other inputs (as in Arrow-Hansmann case: due to lower pressure to make profits). Therefore, it is cheaper for a NP firm to induce any given level of effort by its employees than for a FP firm. It is likely that the (constrained) optimal contract from the point of view of the principal will induce lower level of effort at a FP firm than at a NP. If effort directly translates into quality of output, the NP firm will have higher quality of output than a FP one.

It turns out that in infinitely repeated settings the need for the NP status to signal quality disappears both for the Arrow-Hansmann hypothesis and for the altruistic employees theories (Francois and Vlassopoulos, 2008). If the consumers and the producers interact repeatedly, the consumers learn the reputation of the firm in the aspect of quality (Vlassopoulos, 2009). The consumers simply

purchase at a premium the output of more reputable firms, regardless of their organizational form. As operational surpluses provide stronger incentives for a FP organizational form, the threat of their loss that enables the commitment to provision of high quality in an infinitely repeated setting is more powerful for a FP firm than for a NP one. Under some conditions the quality of output of a FP in a repeated setting is higher than the static quality of a NP in a one-shot game. Thus, the FP firm both provides high quality and offers higher payoff to the entrepreneur, and the NP status is no longer needed to signal quality (Vlassopoulos, 2009). If the quality of output can be deduced after consumption, a reputation can be built and NP organizational form will not arise in these markets. One of the main postulates of my theory is that quality of healthcare cannot be deduced even after consumption, which hinders the ability of reputation building.

In a repeated setting applied to Altruistic-Contract Theory, the employees are able to learn in a relational contract setting the reputation of their employer not to substitute their effort for other inputs (Francois and Vlassopoulos, 2008). Although, in a repeated setting a FP firm can commit not to substitute the effort of its altruistic employees for other inputs, the per-period profits must be positive for the incentive compatibility constraint to be satisfied. NP does not require positive per-period profits, thus NP can commit not to substitute at lower costs. Hence, the altruistic employees theories of NPs are slightly more robust to repetition than the simple form of the Arrow-Hansmann hypothesis.

In order to avoid the criticism that the results hold only when no communication between players, or longer time horizon are not possible, Hirth (1999) uses a "credence" goods setting that seems most applicable to healthcare. Darby and Karni (1973) was the first paper to introduce "credence" goods, which are goods that the consumer does not learn their quality even after the consumption. This is also the approach I take in my model, as it seems the most plausible in healthcare. Both the Arrow-Hansmann hypothesis and the Altruistic-Contract Theory give similar predictions about the impact of the organizational form on the quality of output in a static setting,

when reputation cannot be built. However, in the next section I show that precisely this prediction is rejected by some of the data.

Biglaiser and Ma (2007) and Delfgaauw (2007) examine private versus public provision of healthcare. In a setting with no information asymmetry over the quality of care they show that altruistically oriented physicians, who care for the welfare of the patient would choose to remain in the public sector even when private (FP) provision is allowed. Physicians who care only about their own fee net of provision costs prefer to treat patients in the private sector, where they act as profit maximizers. Both studies find that it is possible that the quality of care is higher in the private sector, as the altruistic doctors are constrained by government regulation (that may come through compensation) in terms of the highest quality of care they may provide.

Although the nature of the analysis of Biglaiser and Ma (2007) and Delfgaauw (2007) is rather removed from the NP setting, one can interpret the public provision in these papers as NP provision of healthcare. Once this parallel is made, one can claim that the results of both Biglaiser and Ma (2007) and Delfgaauw (2007) can be interpreted as predicting the possibility that the quality of output (care) is higher in the FP sector relative to the NP sector. This result is not surprising, however, as it is well known that in a market with no asymmetries of information shirking in the FP sector is not likely to arise. Notice that the self-selection of players between the sectors (public and private, or NP and FP in the new interpretation) is based on the degree of altruism of the doctors and not their ability, as doctors do not vary in the cost of healthcare provision. The self-selection of players between the NP and FP sectors based on ability (and not the degree of altruism) is the main methodological innovation of the model developed in Chapter 2 of this thesis.

To summarize, there are four approaches to economic theorizing of NPs: Altruism at the organizational level, Employee-Cooperatives, Arrow-Hansmann hypothesis and the altruistic-employee Contract Theory. However, only the Arrow-Hansmann hypothesis and the Altruistic-Contract Theory have implications for the quality of output. All theories falling within these approaches have

the same quality implication: higher NP quality of output relative to FPs.[21] The rigorous models stemming from the Arrow-Hansmann hypothesis have two additional empirical implications: existence of positive NP price premium (Glaeser and Shleifer, 2001), and rising FP quality in the share of NPs in the market (Hirth, 1999).

As the next section shows, the rigorous empirical predictions of Hirth (1999) and Glaeser and Shleifer (2001) do not seem to find support in the data. Specifically, the implications for quality of the economic theory reviewed here are not in accordance with a considerable number of empirical studies. Yet, the effect of organizational form on quality of output is in the heart of the policy debate over FP healthcare delivery. This failure of theory to match evidence underscores the need for an extension to the current theories of NPs. This thesis offers such an extension. I review the main aspects of the new theory in Section 1.5 of this chapter and fully develop the model in Chapter 2.

1.4 Existing Empirical Evidence on Quality of Non-Profits vs. For-Profits

In order to answer the policy question of whether conversions from NP status to FP should be allowed for hospitals and other healthcare providers, one can draw on the existing economic theory. However, a proper empirical discipline accepts only those theories which have the most agreement with empirical evidence. Additionally, it could be the case that the effect of organizational form on quality is an empirical question. For these reasons it is important to review the empirical evidence on the relationship between organizational form and quality of output. In this section I review the relevant evidence through the lens of falsification of existing economic theories. However, I also draw direct implications to the policy debate, especially, the challenges that the evidence raise.

[21] According to Hirth (1999) this holds only *on average*.

Convenient for the purposes of the policy debate on FP hospitals, most of the empirical research on NPs comes from the U.S. healthcare market. The reasons for such high focus of the research are availability of data and the importance of the sector to the society. The healthcare sector is important economically: in 2006 it accounted for 15.3% of the U.S. GDP (Chandra and Skinner, 2008).

When one attempts to answer the question of which organizational form provides higher quality in the hospital industry, the first issue that arises is how to measure quality of a hospital. Schlesinger and Gray (2006) provide an overview of current literature that uses measures of quality. One can summarize the existing measures of quality of healthcare services as falling into two main groups: process measures and outcome measures. Process measures focus on the treatment technology used, ideally it is ranked by physicians specializing in this area before the study. Hospitals are subsequently ranked based on their adherence to the prescribed treatment technology.

Outcome measures are the richest in terms of categories and include: legal measures, consumer satisfaction measures and objective measures. Legal measures include regulatory violations and malpractice law suits. Measures of customer satisfaction, usually obtained by asking hospitalized or recently discharged patients to fill out questionnaires, were used in a small number of studies, and were more common for nursing homes rather than acute-care hospitals. Objective measures focus on observable dimensions of patient outcomes. The most common measure used are mortality rates within some number of days after hospitalization or discharge.[22] However, adverse outcomes, re-admissions and in-hospital mortality rates are used as well. For nursing homes specifically, measure of functional improvement was used in 3 reported studies (Schlesinger and Gray, 2006).

The number of empirical studies comparing quality of hospitals is quite overwhelming. Rosenau

[22]The raw mortality rates are taken for an urgent to treat diagnosis (to reduce the impact of hospital selection by patients), e.g. heart attacks, and adjusted for patients' characteristics (referred to as "risk") to obtain Risk Standardized (also called "Adjusted") Mortality Rates (RSMR). To show some evidence in this section, following most of the literature, I use 30 days post-admission for a heart attack RSMR (averaged across patients) as a measure of quality for a hospital (as a whole), see Section 3.2 in Chapter 3.

and Linder (2003) perform a meta-analysis of empirical evidence on NP versus FP quality. The authors use a sample of 69 studies analysing data from 1980 and on (presumably up to the date of the study, as the cutoff period is not reported). A 1979 literature review (1960s-1970s data) found no difference in quality of care between NP and FP providers (Rosenau and Linder, 2003). In their large sample of papers since 1980 Rosenau and Linder (2003) find 59% of studies supporting the Arrow-Hansmann hypothesis (NPs having higher quality than FPs). 12% of studies show higher FP quality, while the remaining (29% of) studies found no difference in quality between the two organizational forms. Schlesinger and Gray (2006) list 38 empirical studies comparing quality of healthcare services provided by NP and FP institutions (21 of which overlap with Rosenau and Linder, 2003). Only 14 of those studies (36.84%) confirm the prediction of the Arrow-Hansmann hypothesis that NPs provide higher quality than FPs. 20 studies (constituting 52.63% of total) find no difference in quality of output between NPs and FPs . While 4 studies (10.53%) report higher FP quality. Hence, overall empirical evidence on the prediction of the Arrow-Hansmann hypothesis is mixed.

As I explain in Section 1.5, due to their methodology, the vast majority of the studies just discussed, in fact, test the implication of "higher NP than FP quality on *average*" that can be deduced from Hirth (1999). Therefore, those studies that reject the Arrow-Hansmann hypothesis and Glaeser and Shleifer (2001), readily rejected the main implication of Hirth (1999) as well. An interesting regularity is that the *percentage* of studies that find superior FP quality seems to increase with the sample size of studies included in a review. Whilst findings of no difference between NP and FP quality can be attributed to low statistical power, the ones finding FPs superior are a clear rejection of the main implications of the existing economic theories of NPs in the dimension of quality of output. Additionally, such mixed findings raise serious challenges in the search for the optimal policy toward FP hospitals.

In addition to existence of evidence that rejects the main predictions of economic theories, a

number of empirical studies reject additional implications of the rigorous models as well. The additional implication of Glaeser and Shleifer (2001) states that NP output must sell at a premium over the FP one, as the NP status signals higher quality. Philipson (2000) finds no NP premium in the available data.[23] The additional implication of Hirth (1999) states that FP quality increases as the share of NPs in the market rises. This prediction is rejected somewhat indirectly by the findings of Cutler and Horwitz (2000) and Horwitz and Nichols (2007). These studies find that entrance by FP hospitals is associated with a reduction in NP quality. The model of Hirth (1999) implies that the quality of NPs is independent of existence of FPs and influences the latter rather than being influenced by it. Therefore, the two major models of NP behaviour that are relevant to the quality aspect are rejected by the data based on their additional implications as well.

However, the empirical evidence does not end there. There is a number of empirical examples that come in contradiction to some of the existing theory. McClellan and Staiger (2000) conduct 3 case studies of hospital quality in 3 U.S. cities over a period of 10 years. The authors find in one city out of the three (coded "Case 1" in the paper) that, out of the three hospitals active in the market, the two FP hospitals have the lowest mortality rates adjusted for patient characteristics (RSMR), while the only NP hospital having the highest. High mortality rates are an indicator of low quality of service. The existence of "physicians concierge," described in Section 1.2, points to specific high quality FP healthcare providers. The list of 32 highest quality FP hospitals in their respective markets in Table 1.3 adds to the empirical examples which contradict the implications of existing economic theory.

In Figure 1.1 I examine the raw data I use later in my empirical investigation of Chapter 3. These data seem to reject the notion that the quality of care at NP hospitals is systematically higher than at FP ones. I calculated the cross-hospitals average (risk-adjusted) mortality rates for heart attacks (as a measure of hospital quality) for NP hospitals and separately for FP hospitals

[23]There may be specification and control variables problems in the estimations, see Vogt (2000).

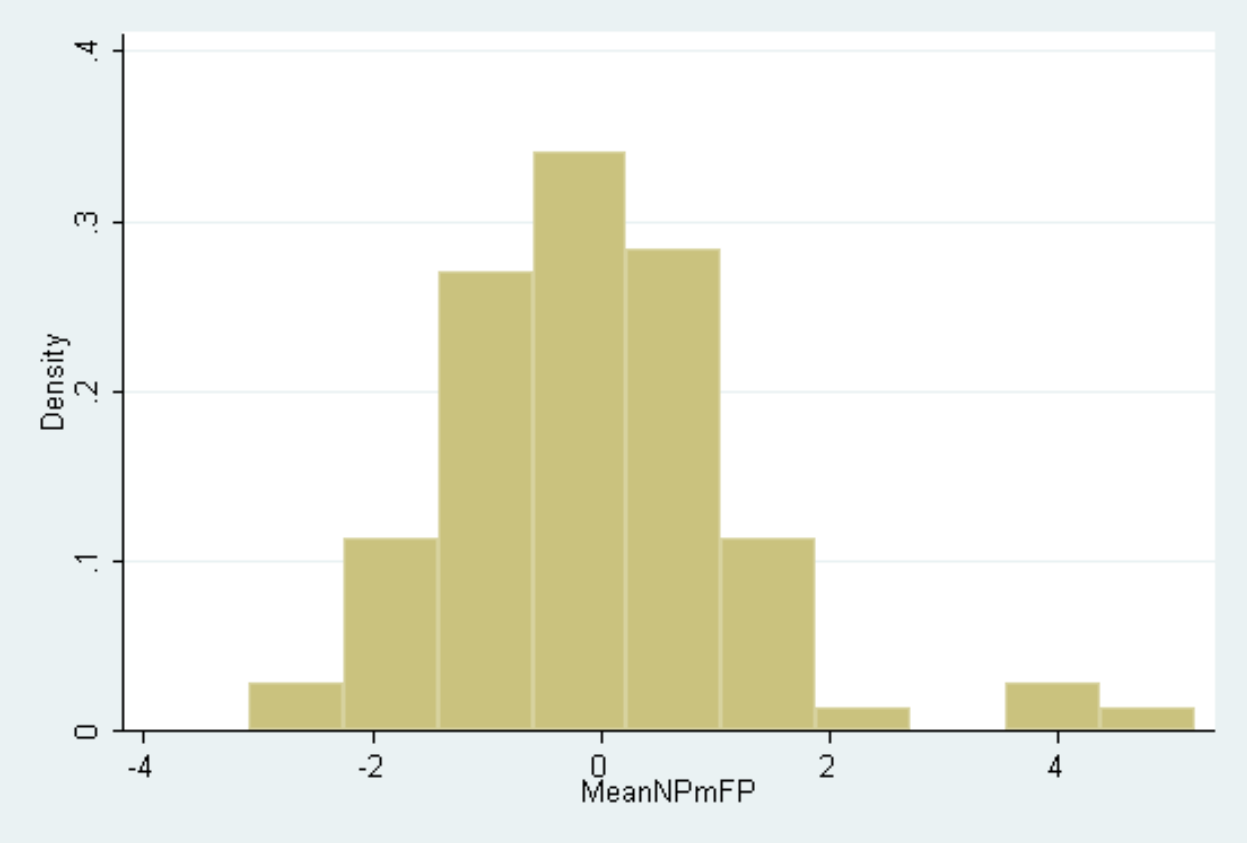


Figure 1.1: Mean Mortality Differences (for Heart Attacks) Between NP and FP Hospitals by MSA[15]

within each MSA-market.[15] Subsequently, I subtracted the mean mortality for FPs from the mean mortality for NPs and plotted the distribution of the differences across MSAs in Figure 1.1. The existing economic theories of NPs predict that one should observe a resulting distribution in which majority of the observations fall in the negative part of the horizontal axis: FP hospitals should have on average higher mortality rates than NP ones in every MSA (implying lower mean FP quality than mean NP quality). Figure 1.1, however, shows an almost symmetric distribution around zero.[24] Initial examination of the data seems to contradict the predictions about quality of

[24]The t-test for mean differences between organizational form market-level means ($H_1 : Mean_{FP}[RSMR] >$

existing economic theories of NPs.

The rich empirical literature finds mixed results regarding the main prediction of the economic theory and the main question of interest to the policy maker about the effect of organizational form of hospitals on quality of care. The rigorous empirical analysis coupled with more casual evidence implies that the existing evidence does not provide a viable tool for policy maker to decide on the desired composition of the healthcare market in terms of organizational forms. This raises the need for an extension to existing theory of NPs that would enable the policy maker to make sense of the conflicting evidence. Such new theoretical contribution will inform the policy maker on the potential outcomes of policies that affect the composition of the healthcare market in terms of organizational forms. This thesis performs precisely this task in Chapter 2. However, it also tests the new theory empirically in Chapter 3 and informs the policy maker in detail on the impacts of such an intervention while drawing on the insights of the newly developed theory, in Chapter 4. The main characteristics of the theoretical extension that I propose in Chapter 2 are summarized in the next section.

1.5 Extension to the Theory of Non-Profits

My extension to the economic theory of NPs analyzes the link between two fundamental properties of healthcare markets: the asymmetry in information about quality of the service provided in these markets, and the mix of organizational forms in them. In a perfect market, the quality of a product would be set by suppliers and known by buyers. If the market were competitive, or if buyers all had the same preferences, quality would be optimal. Quality is not provided optimally in markets with asymmetric information.

Asymmetry in information about quality is more profound in markets for healthcare than in

$Mean_{NP}[RSMR]$) has the p-value of 0.38 (84 observations): equality of the market-level means could not be rejected at conventional significance levels, see Section 3.5 in Chapter 3 for details.

virtually any other set of markets. The quality of a health service cannot be observed directly, or perfectly contracted for at the time of purchase; quality cannot even be determined reliably after the experience, since a patient cannot identify precisely the impact of the health service on any improvement in her health. Even examining the inputs cannot uniquely determine the quality of the outcome, due to high variation in technologies of healthcare provision even within one country (Skinner and Wennberg, 2000).[25]

Turning to the second fundamental property, in healthcare markets, more than in other markets, we observe a mix of organizational forms. Schlesinger and Gray (2006) report that in the late 1990s in the U.S., for example, in Dialysis Centres FP firms had a 70% market share, with NPs serving 30% of the market. On the other hand, Hospice Programmes, during the same period, had slightly less than a 20% FP market share, while NPs had a 75% market share (with government owned facilities serving the remaining of the market).

The literature argues for a tight link between the two fundamental properties: asymmetry in information about quality, and the mix of organizational forms. The Arrow-Hansmann hypothesis, introduced by Arrow (1963) and Hansmann (1980), states that when quality is unobservable by consumers, FP firms will have a higher incentive to shirk on quality than NPs.[26] Rigorous treatment of this postulate by Hirth (1999) and Glaeser and Shleifer (2001) generated the empirical implication that NPs must provide higher quality than FPs, at least on average.[27]

A typical empirical study that compares quality of NP and FP hospitals uses the following

[25]Due to the credence nature of healthcare, as defined by Darby and Karni (1973), the insight of Vlassopoulos (2009) about the possibility of reputation building in the quality dimension breaks down in my setting. One may argue that if the econometrician is able to measure quality to some extent, as Chapter 3 of this thesis asserts, some reputation building may occur in the healthcare sector. It is plausible, however, that similar type of analysis that is available to the econometrician is too costly (even to understand) at least for some individual consumers. For this reason the model of Chapter 2 assumes existence of a small share of perfectly informed consumers.

[26]Every dollar saved by a FP firm through reducing quality automatically accrues to the owners or managers of the firm. On the other hand, under NP status due to the non-distribution constraint, it is costlier (if possible at all) to transfer any operational surpluses to those in control of the firm.

[27]In a structural Econometric model Gowrisankaran and Town (1997) find the same for a theoretically defined quality variable.

statistical methodology: estimate a regression of mortality rates for a certain procedure at a hospital, as a measure of hospital quality, on a set of hospital characteristics and two dummies for the organizational form (usually NP and government ownership, with FP omitted) (see Schlesinger and Gray, 2006). The results are interpreted by the significance and the sign of the coefficient on the NP dummy. Although most of these studies never state it explicitly, they measure *average* quality of NP relative to FP hospitals. As stated above, the studies find all three possible results: no difference (the coefficient on the NP dummy is not statistically significantly different from zero at conventional levels), NP superiority (negative coefficient: recall that the dependent variable is mortality rate) and FP superiority (positive coefficient).

When one examines any economic data she always finds a *distribution* of values. Therefore, it is hard to argue with the fact that there is a non-degenerate distribution of hospitals by quality both within the NP and within the FP sector in any dataset. My innovation comes from the fact that I interpret the empirical results as pointing to the regularity that the first moment does not provide enough information to characterize these distributions of quality. There are two distributions of hospitals by quality: one within the NP sector and the other within the FP sector. If there is enough overlap in the support of these distributions, no meaningful inference can be made by examining the first moments.

At first glance it may seem that the overlap between the distributions of quality for NPs and the one for FPs implies that no meaningful answer can be obtained to the question: is there a difference between these distributions. I argue in this thesis that the opposite is true. My extension to the theory of NPs (Chapter 2) draws a new picture of mixed markets (such as the healthcare market): in a distribution of (all) firms by quality within a market populated by both NP and FP firms there is a higher mass of FPs on the tails of the distribution than in the middle of it.

The prediction regarding higher mass of one organizational form within a particular part of the distribution comes from the stylized nature of the model. In reality, unless the means of quality

are equal between NPs and FPs, this feature is unlikely to hold.[28] One can test, however, a direct implication of the above prediction: the *variance* of quality is higher in the FP sector relative to the NP sector. This implication can be easily tested by comparing second moments of the sector-level distributions of hospitals by quality within geographical markets, applying relevant controls. I do this in Chapter 3.

I depart from the existing theory in recognizing not simply imperfect information on the part of consumers, but a *variation* in the extent of information across consumers.[29] Without a doubt, some patients are less informed than others about the quality of service provided by a doctor. Those consumers who are better informed are able to receive high quality services from FP firms,[30] as they are willing to pay for any increase in quality. Those ones who are less informed receive low quality services from FPs due to shirking by providers, as the Arrow-Hansmann hypothesis suggests. If these effects are strong enough, NP firms' quality of output should fall in the middle of these two extremes. Although NPs do not suffer from the shirking problem,[31] they are less attractive to high ability doctors due to their lower flexibility in distributing earnings. In short, the question I address is not "Which organizational form provides higher quality?", but "How are the *distributions* of quality and organizational forms related in a mixed market?"

My theory predicts that we should see in a mixed market a quality distribution across hospitals having basic characteristics as depicted in Figure 1.2. At the upper tail of the distribution (right hand side of the graph) we should find high concentration of FP hospitals . The high concentration

[28]The number of hospitals within geographical markets tends to be too small to observe "tails" of the distribution. See Figure $B.2$ in Section $B.2$ of Appendix B for the distribution of the number of hospitals by Metropolitan Statistical Area (MSA). The empirical analysis in Chapter 3 treats each MSA as one market.

[29]Although Hirth (1999) allows for informed consumers, the empirical implications of his work are not in accordance with the data. Additionally, more recent papers by Glaeser and Shleifer (2001) and Vlassopoulos (2009) do not employ this assumption.

[30]Luft et al. (1990) provide evidence that choice of hospital is strongly influenced by quality, even when quality measures are not publicly available.

[31]A weak degree of physician-altruism is needed for this result to hold. The specific structure of the model is discussed in detail in Chapter 2 of this thesis.

Figure 1.2: The Distribution of Quality in a Mixed Market - New Model

of NP hospitals should occupy the middle of the quality distribution. Both of these groups of hospitals produce medical services efficiently, without shirking on quality. At the lower tail of the distribution of quality I expect to see a mass of FP hospitals that systematically shirk on quality that they provide to their patients (the shirking is the reason for a gap in the support between NPs and the left hand side FPs).

Figure 1.2 shows the main implication of the model: in the distribution of firms by quality the FP firms are more heavily represented on the tails than in the middle of the distribution, and the NP ones more in the middle than on the tails. I do not expect this to hold in reality, however, as this separation is likely to be driven by the stylized nature of the model.[32] The direct consequence of the differences in the relative prevalence of organizational forms along the distribution of quality are differences in *second moments* between the distribution of quality within each organizational form. I test empirically in Chapter 3 the following central prediction: FP firms have a higher variance of quality than NP ones.

How does my theory reconcile the seemingly contradictory empirical results presented in section 1.4? The key point is realizing that the existing empirical studies measure *average* differences between FP and NP quality due to their statistical methodology (see above), while the datasets they use contain under FP observations both hospitals serving informed patients (high quality) and the ones serving uninformed (low quality). Therefore, my theory is able to explain the evidence from these studies as: depending on the relative sizes of the pools of high quality FPs, NPs, and low quality FPs, *on average* either organizational form can have higher quality of output. Although

[32]Only in markets where the mean FP and NP quality are equal the over-representation of FPs in the tails is likely to show up.

the relationship of the first moments of FP and NP quality is non-determinate, my theory predicts a very specific structure: in the distribution of firms by quality, FPs are more heavily represented on the tails than in the middle of the distribution. The central new testable implication of my theory is: the *variation* in quality is higher among FPs than NPs.

In order to make sense of the conflicting empirical evidence, and to have a theory that is consistent with data, I develop a new model of NPs. This section summarizes the main characteristics of this model. These features are: variation in the degree of information asymmetry about quality across consumers, selection of doctors into organizational forms by ability and choice of quality of care by doctors. These components (with a weak degree of altruism on the side of the doctors, discussed in Chapter 2) predict a distribution of hospitals by quality such that FPs are represented more heavily on the tails. The model is capable of explaining the existing seemingly contradicting empirical results (*on average* either organizational form can have higher quality of output) on one hand, and is easily falsifiable due to particular structure of the distribution of quality, on the other (the *variance* of quality is higher for FP hospitals). Chapter 3 reports tests of my theory. It turns out that the data supports the central implication of the model on second moments as well as additional indirect implications. Therefore, this model appears as most suitable up to date to address the policy debate over FP hospitals. However, it still remains to show how my model can be applied directly to the U.S. healthcare sector. I do this in the following section.

1.5.1 The Connection of My Model to the U.S. Healthcare

In Chapter 2 of this thesis I develop a model of a healthcare market. One of the implications of this model is the existence of a segment of physicians treating patients in a FP setting. The model predicts that informed patients who are treated by FP providers are receiving the highest available level of quality of care in the economy. This feature seems to be at odds with the professional folklore of healthcare markets. In order for the basic properties of my model to hold, the segment

of high quality FPs need not be large. In reality, this segment maps best into "physician concierge" practices and speciality hospitals discussed in Section 1.2. Table 1.3 presents 32 FP hospitals that have higher quality of care than any other hospital, including NPs, in their respective markets. This establishes basic consistency of the main prediction of my theory with the moder-day U.S. healthcare system. In this section I examine various components of the model and show their relationship to the U.S. healthcare system and its legal environment.

In Section 1.2 I discuss the various provider compensation schemes that have developed in the U.S. healthcare. For tractability purposes my model addresses only one service the need for which is certain. For this reason, it seems most natural to adopt the FFS paid by patients out-of-pocket as the system of compensation in this modelling approach. The modelled FFS system may be thought of as a "degenerate" capitation, since there is no uncertainty over the need for the procedure. Additionally, any insurance considerations will reduce the tractability of the model. For this reason, in my model doctors choose a fee to charge for their service, and patients' purchasing decisions are influenced directly by this fee.

I model NPs as having the legal requirement to break-even, which is monitored by the regulator. In reality NPs are not required to break-even on an annual basis. Nevertheless, it is reasonable to suppose, and many have argued, that there is less pressure in NP organizations to generate profits because of the absence of a group of residual claimants on profits who, in the conventional FP organization, can reward (and fire) managers on the basis of profits. Section 501(c)(3) of the U.S. Internal Revenue Code, however, requires an oversight of the fact that no outside party is capable to benefit from operating surpluses of NPs (Claxton et al., 1997). This means that NPs are unable to distribute profits. Hence, it must be the case that NPs break-even on average across a number of years. In accounting terms: NPs invest positive operating surpluses and disinvest when they have negative operating surpluses. My model captures this aspect in a simplified for tractability static setting through the requirement to break-even.

In the model developed in Chapter 2 the regulator oversees the fact that NPs charge a break-even fee. In the managed care sector of the U.S. healthcare insurers directly monitor the fees charged by providers. However, HMOs (the main representatives of the managed care sector) beyond the monitoring of the fees also try to reduce utilization. Whether the reduction in utilization is harmful to patients is a controversy. Many believe that it may even be beneficial, as over-utilization may have adverse effect on patients' health (Bodenheimer and Grumbach, 2009 pp. 65-66). Nevertheless, it is clear that there are some similarities between NP providers in my model and the HMOs in the U.S. healthcare. However, many hospitals are large enough to have bargaining power vis-à-vis the HMOs and it is unlikely that all of them break-even. As I test my model on hospital data, the NP and FP status should be better indicators of behaviour than contracts with an HMO.

A NP private provider of healthcare is prohibited by law to distribute profits. This fact reduces the pressure to make profits. In these circumstances, if the provider cares to some extent about the welfare of its patients (which the medical profession, with its strict professional code of ethics, professes to do) two things are likely to happen: 1) uninformed patients are able to gauge the quality through the fee charged (a higher fee is likely to imply higher quality), 2) consumer valuation of the service of NP providers is significantly higher than the fees NPs charge (due to service's superior quality). As a result, it could be the case that more patients will seek care with a NP provider than the provider's capacity to serve. Of course this is true only in a world where the amount of resources available for healthcare is below what is demanded at the price NP providers charge. This creates the need to ration healthcare and limit the access to it. It seems plausible that, even if the access to healthcare within the NP sector is equitable for all of the uninformed patients, those with superior understanding of the system (i.e. informed) are able to gain an advantage in such access.

Low income patients are very likely to be uneducated and less savvy about medical procedures and the overall structure of the healthcare system. In order to infer quality it is important to

be savvy in terms of the general structure of the medical system and have a reference who could recommend a good physician. In my theoretical model I allow patients to vary only in the dimension of information. Therefore, I find it plausible to assume that fully informed patients have an advantage over uninformed ones in access to care even in the NP sector.

Many studies find that, although there is no significant difference in actual quality of care in terms of outcomes at HMOs (managed care) versus the traditional FFS care systems, the patient satisfaction is lower at HMOs (Barr, 2007 Chapter 4).[33] Additionally, it also has been found that low-income patients have significantly worse outcomes in HMOs that in FFS system (Barr, 2007 p. 98). Some of the studies are done in experimental design, i.e. the physicians do not know if their patient is low- or high-income, and the person's actual insurance coverage does not differ by income either. One of the explanations for worse HMO outcomes was that low-income patients have difficulties in navigating the complex bureaucratic systems of HMOs. As I argued above, HMO insurance arrangements are somewhat similar to the NP providers in my model due to fee oversight. This supports the assumption of my model that informed patients have advantages over uninformed in access to care in the NP sector.

In my model in equilibrium all of the informed patients are treated by high quality FP providers. If the government decides to ban FP medical practises, the informed patients will still have access to the same doctors they had before the ban (who are now in the NP sector), due to the better familiarity with the system of those patients. Hence, it could be the case that a ban on FP hospitals will not improve access to care of uninformed (disadvantaged) patients.

My model provides the following key insight to the policy debate. Without specific government policy that discourages FP healthcare provision there are FP providers who offer high quality of care to informed patients. If a policy that discourages FPs takes effect, the medical staff of these

[33]These perceptions led to more and more restrictions on HMOs in reducing costs, that resulted in a resumed increase in the inflation of healthcare costs in the U.S., see Barr (2007) Chapter 4.

FP providers will be pressed to open NPs and charge lower fees for services of similar quality as before (in accordance with the NP business practice). Such a policy is likely to cause nothing but a transfer of wealth from medical staff to (informed and) well-off patients. I discuss this in detail in Chapter 4.

Many claim that NP organizations are less efficient than FP ones.[34] Cutler and Horwitz (2000) report that FP hospitals are better at collecting higher insurance claims than NP ones, however, this comes at the expense of employing large administrative staff. On the other hand, there is evidence that NPs learn these practises when FPs enter (Cutler and Horwitz, 2000). Additionally, a report by the National Academy of Sciences cited in Barr (2007) (p. 178) stated that FP hospitals were found to be slightly less efficient. Various studies that employ efficiency-frontier estimation technique claim that they find operating efficiency advantages at FP hospitals (however, this result is by no means universal). A methodological flaw of these studies is that they may account for unmeasurable quality as a waste (Sloan, 2000). As there is no evidence either way, I abstract from the differences in productive efficiency in order to focus on the differences between NPs and FPs in the degree of the conflict of interests between the managers of the firm and its consumers. I assume that both organizational forms are equally efficient (i.e. do not waste resources) and that NPs break-even according to some standard, while FPs may charge any fee they choose (weakly) above costs.

As I note in this section, (Group-Model and IPA) HMOs may have characteristics similar to the regulator in the proposed theory in the sense that they are monitoring the fees that the FFS physicians and hospitals are allowed to claim from them. It is interesting to note that the U.S. government-administered insurance schemes Medicare and Medicaid, do not act in the same fashion as the HMOs in forcing the hospitals to operate close to their break-even points. Therefore, the

[34]Support for such claims is usually grounded in traditional economic arguments of higher incentives to generate profits. The newer Contract Theory approach emphasizes the value of incentives achieved through employee shares in profits. As no profits can be distributed by NPs, no such monetary incentives can be given in NP organizations.

data I use to test the new theory, which come from Medicare and Medicaid cases, is a better showcase of the differences in behaviour across hospitals of different organizational forms than the general population. Only if the hospital is NP, it will be monitored by the regulator to ensure no distribution of profits. The government is weak at monitoring the actual expenses hospital incurs on its Medicare and Medicaid patients (Barr, 2007), therefore, a FP hospital is able to shirk on quality if the patient is uninformed.

In this subsection I have given factual justification for various postulates and predictions of the new model of NPs I introduce in Chapter 2 of this thesis. This model was described in the beginning of Section 1.5, before this subsection. I attempted to show that both postulates and predictions are grounded in the highly complex and fragmented modern-day U.S. healthcare system and its legal environment. These will enable me in Chapter 4 to apply the insights of my theoretical contribution to the U.S. healthcare policy after it is tested using new empirical data in Chapter 3.

1.6 Summary of the Introduction

The wave of conversions of U.S. hospitals from NP to FP status throughout the 1990s raises concerns over a reduction in quality of care by converting hospitals. Although these concerns are apparently supported by existing economic theory, there is mixed empirical evidence on the validity of both the concerns and the theories that justify them. In order to make an informed decision on policy toward conversions, the policy maker needs a reliable tool to predict the impact of various policy alternatives on the quality and access to healthcare and on the overall welfare in the economy. This thesis develops such a model (tool), shows its empirical validity, and reports the main insights that it generates for healthcare policy in the U.S. .

The empirical research finds evidence of all three possibilities of the impact of organizational form on the quality of healthcare: better NP quality of care, no difference in quality between NPs

and FPs and FP superiority quality-wise. The last result is at odds with existing economic theory of NPs. This creates the need for a new theory which is able to reconcile the contradictory empirical evidence, and that could be used to inform policymakers on the effect of various policy alternatives.

In this thesis I develop a new model of markets with NP and FP firms. According to this model, FP firms are over-represented both at the upper and the lower tail of the distribution of quality in these markets. The NP firms are concentrated in the middle of the distribution of quality. These results hinge on the fact that consumers vary in their information asymmetry about the quality of healthcare they receive from hospitals. This model explains the existing empirical evidence by the fact that either organizational form can have higher first moment of quality. My theory generates the central falsifiable prediction: the variance of quality is higher in the FP sector relative to the NP one. The main innovation of the model is that the players self-select between the organizational forms based on their ability, and not the degree of altruism, as in the previous literature. The fact that the model is consistent with existing empirical evidence enables to use this model to predict the impact of policies that encourage only one organizational form (or discourage conversions) on the distribution of quality of healthcare in the economy, and on access to it.

Anecdotal evidence from the U.S. healthcare system documents the existence of a phenomenon that is called "physician concierge". These physicians operate private FP practices. They offer very high quality of service for extra fees paid by patients who, presumably, can monitor that high quality is indeed delivered. Additionally, I present evidence on 32 FP hospitals in the U.S. that have lower mortality rates for Heart Attacks (a common measure of quality) than the best NP hospital in their market. All of these examples shed doubt on previous NP theories and support my extension. More rigorous tests are undertaken in Chapter 3 of this thesis.

In Chapter 2 of this thesis I setup the model that constitutes an extension to the existing theory of NPs, solve it and discuss the results. Chapter 3 derives comparative static (indirect) implications of the model. It proceeds to tests these implications, the central prediction of higher variance of

quality in the FP sector relative to the NP on an original dataset, and implications regarding waiting time length on recent data. After obtaining empirical support for the model, I use the theory developed to inform policymakers on potential welfare consequences of a policy intervention that impacts the composition of healthcare markets in terms of NP and FP organizational forms. I present these insights for policy evaluation obtained from the new extension to the theory of NPs in Chapter 4. Chapter 5 summarizes this thesis.

Chapter 2

Theory

2.1 Introduction

This chapter analyzes the link between two fundamental properties of healthcare markets: the asymmetry in information about quality of the service provided in these markets, and the mix of organizational forms in them. In a perfect market, the quality of a product would be set by suppliers and known by buyers. If the market were competitive, or if buyers all had the same preferences, quality would be optimal. Quality is not provided optimally in markets with asymmetric information.

Asymmetry in information about quality is more profound in markets for healthcare than in virtually any other set of markets. The quality of a health service cannot be observed directly, or perfectly contracted for at the time of purchase; quality cannot even be determined reliably after the experience, since a patient cannot identify precisely the impact of the health service on any improvement in her health. Even examining the inputs cannot uniquely determine the quality of the outcome, due to high variation in technologies of healthcare provision even within one country (Skinner and Wennberg, 2000).[35]

Turning to the second fundamental property, in healthcare markets, more than in other markets, we observe a mix of organizational forms. Schlesinger and Gray (2006) report that in the late 1990s

[35]Due to the credence nature of healthcare, as defined by Darby and Karni (1973), the insight of Vlassopoulos (2009) about the possibility of reputation building in the quality dimension breaks down in my setting.

in the U.S., for example, in Dialysis Centres For-Profit (FP) firms had a 70% market share, with Non-Profit (NP) ones serving 30% of the market. On the other hand, Hospice Programmes, during the same period, had slightly less than a 20% FP market share, while NPs had a 75% market share (with government owned facilities serving the remaining of the market).

The literature argues for a tight link between the two fundamental properties: asymmetry in information about quality, and the mix of organizational forms. The Arrow-Hansmann hypothesis, introduced by Arrow (1963) and Hansmann (1980), states that when quality is unobservable by consumers, FP firms will have a higher incentive to shirk on quality than NPs.[36] Rigorous treatment of this postulate by Hirth (1999) and Glaeser and Shleifer (2001) generated the empirical implication that NPs must provide higher quality than FPs, at least on average.[37]

Schlesinger and Gray (2006) list 38 empirical studies comparing quality of healthcare services provided by NP and FP institutions. Only 14 of those studies (36.84%) confirm the prediction of the Arrow-Hansmann hypothesis that NPs provide higher quality than FPs. 20 studies (constituting 52.63% of total) find no difference in quality of output between NPs and FPs . While 4 studies (10.53%) report higher FP quality. Hence, overall empirical evidence does not seem to support the predictions of Arrow-Hansmann.[38]

This paper explores a conceptually simple departure from the Arrow-Hansmann framework that can reconcile the existing empirical evidence on quality and organizational form with the fundamental fact of information asymmetry. The theory not only reconciles existing evidence with informational conditions in the markets, it also offers additional testable implications.

[36]Every dollar saved by a FP firm through reducing quality automatically occurs to the owners or managers of the firm. On the other hand, under NP status due to the non-distribution constraint, it is costlier (if possible at all) to transfer any operational surpluses to those in control of the firm.

[37]In a structural Econometric model Gowrisankaran and Town (1997) find the same for a theoretically defined quality variable.

[38]An earlier meta-analysis of empirical evidence on NP versus FP quality was performed by Rosenau and Linder (2003). In the sample of papers analyzed, the authors find 59% of studies supporting the Arrow-Hansmann hypothesis, while 12% of studies show higher FP quality (the remaining studies found no difference in quality between the two organizational forms).

I depart from the existing theory in recognizing not simply imperfect information on the part of consumers, but *variation* in the extent of information across consumers.[39] Without a doubt, some consumers are less informed than others about the quality of service provided by a doctor.[40] Those consumers who are better informed are able to receive high quality services from FP firms, as they are willing to pay for any increase in quality. Those ones who are less informed receive low quality services from FPs due to shirking by providers, as Arrow-Hansmann suggest.[41] If these effects are strong enough, NP firms' quality of output is in the middle of these two extremes. Although NPs do not suffer from the shirking problem, they are less attractive to high ability doctors due to their lower flexibility in distributing earnings. In short, the question I address is not "Which organizational form provides higher quality?", but "How are the *distributions* of quality and organizational forms related in a mixed market?"

The starting point in my reconciliation of economic theory with existing evidence is to recognize that the existing empirical studies measure *average* differences between FP and NP quality due to their statistical methodology.[42] The datasets they use contain under FP observations both hospitals serving informed patients (high quality) and the ones serving uninformed (low quality). Therefore, my theory is able to explain the evidence from these studies as: depending on the relative sizes of the pools of high quality FPs, NPs, and low quality FPs, *on average* either organizational form can

[39] Hirth (1999) assumes the existence of informed and uninformed consumers in his model. Other features of Hirth (1999) make this earlier work inconsistent with some of the empirical evidence.

[40] Luft et al. (1990) provide evidence that choice of hospital is strongly influenced by quality, even when quality measures are not publicly available.

[41] Hansmann (1987), prudently, pointed out that his theory of NPs does not apply to market segments in which consumers are informed. However, to the best of my knowledge, this key limitation has yet found its way into rigorous theoretical literature. Moreover, all of the empirical work up to date, that I am familiar with, has ignored this limitation of Hansmann's theory.

[42] A typical study summarized by Schlesinger and Gray (2006) uses the following statistical tool: a regression of mortality rates for a certain procedure at a hospital, as a measure of hospital quality, on a set of hospital characteristics, one of them being a dummy for whether the hospital is a NP. The results were interpreted by the significance and the sign of the coefficient on the NP dummy. Although most of them never state it explicitly, these studies measure only *average* quality of NP relative to FP hospitals (or relative to all other organizational forms if government ownership dummy is not included).

have higher quality of output. The central new testable implication of my theory is: the *variation* in quality is higher among FPs than NPs. In the distribution of firms by quality, FPs are more heavily represented in the tails than in the middle of the distribution.[43,44]

The key ingredient of my model on the demand side is a variation across consumers in the extent of information asymmetry. To focus on the implications of a variation in quality, this paper takes this assumption to an extreme: some patients are uninformed and some patients are perfectly informed. The uninformed patents are disadvantaged in two ways: (a) they cannot identify the quality of service that any given doctor provides when choosing which doctor to hire; and (b) they cannot write a contract with any doctor that calls for a specified level of quality.

On the supply side of the model, in order to have the NP sector arise endogenously, I assume that doctors are altruistic. This perspective has recognition in the NP literature, for example, see Newhouse (1970) and Francois and Vlassopoulos (2008), who show that decision-makers (former) or employees (latter) who are relatively more concerned with quality performance and buyer welfare self select into the NP sector.

One possible link between organizational form and quality provision is to have doctors who care more about patient-welfare or quality delivery self-select into the NP sector. This feature has been studied extensively in the literature (e.g. Newhouse, 1970; Biglaiser and Ma, 2007; Delfgaauw, 2007; Francois and Vlassopoulos, 2008) and it generates the same prediction as the

[43]The literal implication of the model is: in the distribution of firms by quality the FP firms are more heavily represented on the tails than in the middle of the distribution, and the NP ones more in the middle than on the tails. I do not expect this to hold in reality, however, as this separation is likely to be driven by the stylized nature of the model. See Section 1.5 in Chapter 1 for a discussion of this point.

[44]In a setup where employees care about output even when they do not contribute to its production (public good), and the degree of altruism varies across employees, Francois (2007) finds that NPs should have two pulls of workers: high effort ones and low effort ones. The public good nature of the output, absent in my model, is the main reason for these findings. Although these results may seem at odds with mine, the author does not consider strategic interactions between firms, additionally, no conclusion about FPs in the same market can be made. Public good results seem less applicable to the healthcare sector. Doctors are likely to care about the welfare of the patient they treat more than the welfare this patient could receive if she were treated by another doctor. Patient treated by a doctor contributes to the reputation of this doctor, while a patient of another doctor contributes to the reputation of her competitor.

Arrow-Hansmann hypothesis. As indicated above, however, the empirical evidence is mixed on this prediction. My results do not rely on the self-selection of more altruistically oriented individuals into the NP sector. In fact, all doctors in my model share the same degree of altruism.[45]

I allow doctors to vary in their inherent ability, which I interpret as variation in cost of providing any given level of quality. The model yields a prediction about the selection of doctors of different abilities into different organizational forms (the hidden information aspect of the information asymmetry) as well as qualities of care they offer (resolution of the moral hazard, or hidden action problem). But it is not a simple bifurcation of doctors/healthcare providers into NPs providing higher quality than FPs. The model is consistent with either organizational form offering higher quality on average. My extension differs from the previous literature in the fact that doctors self-select based on ability and not on the degree of altruism. On top of the central testable prediction discussed, an extension to the baseline model generates yet additional hypothesis that waiting times are longer for NP healthcare providers relative to FP providers.

2.1.1 For-Profits

Some FPs sell to uninformed patients, and because of both the hidden action, or moral hazard problem (the choice of quality by the provider in its own interest instead of in the joint interest of both contractual parties), and the hidden information problem (the actual quality level, chosen by the doctors, cannot be identified by the patient), the quality of these FPs is low. These providers "shirk" on quality of care they offer to uninformed patients.

Other FPs provide services to informed patients. These providers can perfectly contract with patients over quality, which results in their ability to offer first-best[46] quality levels to the informed

[45] Francois (2003) also considers selection of agents into FP and NP organizational forms. In his setting the goal of both the NP and FP firms is to provide a public good. Employees care about public goods, therefore, one may think of the agents as altruistic and sharing the same degree of altruism. The model of Francois (2003), however, does not consider the provision of quality.

[46] As will be explained later in detail, "first-best" here refers to a world without altruism.

patients.

2.1.2 Non-Profits

In my model NPs perform the Arrow-Hansmann role: they provide high quality of care to uninformed patients. The entrepreneurs in NPs, however, are no more altruistic than FP entrepreneurs. Altruism simply gives a reason for NPs to exist. Entrepreneurs care about income net of effort cost as well as consumer welfare and there is no automatic meeting of consumer objectives.

The role of NPs in the model is not a mechanism by which entrepreneurs can commit to higher quality, as Glaeser and Shleifer (2001) suggest, for example. Instead, the role of NPs follows from the defining characteristic of this organizational form: NPs do not distribute profits. In my static model, I take this feature to an extreme, in assuming that NPs face a regulatory constraint that fees must just cover costs. The non-distribution requirement for NPs is the legal reality (at least in the U.S.), however, in the static setting it implies that no positive surpluses can be generated. Given the constraint against profits, NPs each produce the first-best quality;[46] since even a modest concern with consumer welfare leaves it as the only criterion governing the choice of quality. In this way the hidden action problem is resolved by the NP organizational form.

Because of the zero profit constraint for NPs, patients are able to infer quality exactly from the service fee charged. Thus, NPs eliminate the hidden information problem as well: the regulatory constraint solves the information problem for the consumers. The two classes of asymmetric information problems in economics: hidden action and hidden information, are both necessary for my model to generate the predictions that are consistent with empirical evidence, and both are eliminated in the transactions performed under the NP organizational form.

The assumption of zero profit in the NP sector may seem too restrictive especially since it implies a rather strong result of perfect inference of quality through fees. In the FP sector the relationship between costs and prices depends on the competitiveness of the market and the in-

formation distribution in it. In complex markets that often suffer from little competition, such as healthcare (Abraham, Gaynor, and Vogt, 2005), this relationship is not likely to be strong. In the NP sector on the other hand, as there is less ability to benefit from higher operating margins, fees are likely to be more strongly related to costs. When one observers comparable NPs which charge different fees, and comparable FPs that charge different fees, it is more likely that the higher fee NP has higher costs that went toward quality provision, while it is less likely that the same is true for FPs. The following three paragraphs provide three additional justifications to the zero profit assumption.

First, section 501(c)(3) of the U.S. Internal Revenue Code states that no outside party is allowed to benefit from operating surpluses of NPs (Claxton et al., 1997). This feature justifies the non-distribution constraint. The zero profit condition is merely an implication of the non-distribution constraint in a static setting. With the inability to distribute surpluses there are two other potential uses for them: investment in real (or financial) capital, or retainment to cover future operating deficits. The firm lives for only one period. As there are no future periods, the firm will not be able to enjoy the benefits of investment, or suffer from operating deficits. If the firm must cover its cost of operation, it must be the case it breaks-even, or resources will be wasted. The argument is even stronger for an altruistic firm that has the welfare of its consumers in its objective function. The best use of resources, from the point of view of such a firm, is transferring any surpluses, that the firm is unable to benefit from them directly, to its consumers.

Second, the goal of my model is to compare NPs and FPs. As explained above, informed patients monitor some FPs to ensure that they indeed receive the quality for which they pay. A priori there is little reason to suspect that informed patients are able to monitor hospitals significantly better than the regulator. Therefore, if the employees of hospitals collect operating surpluses simply through shirking on effort in their work, there is little reason to expect that it happens more in the NP

sector.[47] For this reason, in order to focus on the *differences* between NPs and FPs it is beneficial to abstract from the additional shirking by employees at hospitals that have superior quality in the first place (NPs and FPs which treat informed patients). This abstraction is especially useful since shirking is modelled explicitly for those FPs which treat uninformed patients.

Third, one of the main objections to perfect monitoring of the zero profit constraint could be the fact that the regulator cannot observe doctor's ability, which impacts the cost of provision of any given level of quality. The de-facto monitoring seems more realistic if ability proxies for education or seniority and there are accepted standards for expected quality of care from all doctors who share the same degree of seniority or education. This completes the three main justifications of the break-even assumption for NPs.

Section 2.6 discusses the consequences of relaxing the zero profit assumption. I assume in that extension that the monitoring by the regulator is probabilistic, i.e. with small probability NPs can circumvent the non-distribution constrain and collect operating surpluses. This leads to shirking on quality in the NP sector by those who are lucky to escape the monitoring. I show that the main characteristics of the equilibrium remain the same, if the organizational form choice is not allowed to change as a result of the possibility of shirking under the NP organizational form. Only firms which choose the NP status when shirking is not possible may shirk under the NP status. The inference of quality through fees by uninformed patients in the NP sector becomes imperfect. The predictions regarding the main characteristics of the distribution of quality in a mixed market are retained.

Is variation in ability across doctors essential for my results? If all doctors were of the same ability, then the first-best level of quality would be equal across doctors. In this case, both FPs who serve informed consumers, and NPs, would provide the same level of quality. As there is

[47] Although FP hospitals tend to have higher powered compensation for their management than NP ones, the difference between the two sectors in the power of the compensation falls as the rank of the employee falls (Erus and Weisbrod, 2003).

an additional segment of FPs that provide quality below the first-best (due to the information asymmetry), on average NPs would always have higher quality that FPs. This prediction is not supported by the empirical evidence. Therefore, variation in ability across doctors is crucial to my theory.

The rest of this chapter is organized as follows. Section 2.2 sets up the model. In Section 2.3 I solve for an equilibrium. I discuss the equilibrium strategies of doctors-entrepreneurs in each market segment (NP, low quality FPs, and high quality FPs) separately. Then, I discuss the allocation of patients between the three market segments. Finally, I derive the distribution of quality between the organizational forms in the mixed market. Section 2.4 provides the main welfare properties of the equilibrium. In Section 2.5 I discuss the sizes of the NP and FP sectors. Section 2.6 discusses the implications of relaxing the central assumption of the model that all clinics that choose NP status always break-even. Section 2.7 concludes. The proofs and the mathematical formulation of the equilibrium are in Appendix A.

2.2 The Game

2.2.1 Players

Consider a healthcare market with doctors and patients. Each patient buys zero or one unit of healthcare. Each doctor provides zero or one units of healthcare. Services doctors offer may differ in quality, denoted $q \in \mathbb{R}_+$. Quality is the only aspect of the service that is of value to the patients. Patients agree on the ranking of services by quality, and have identical preferences and willingness to pay. This enables to measure quality in the units of willingness to pay for a given quality level, i.e. dollars.

Provision of quality is costly for doctors, and the cost is increasing with quality. Doctors differ in their ability. Higher ability doctors can deliver the same level of quality at lower cost. Let θ

denote doctor's type, such that high θ implies high ability, or low costs. θ is uniformly distributed (call this CDF $G_\theta[\theta]$) over the interval $[\underline{\theta}, \overline{\theta}]$, of length D. The cost of providing quality at level q is $c(q;\theta)$. I assume that $c(q;\theta)$ is twice continuously differentiable in both variables, falling in θ, increasing and strictly convex in q and has negative cross partial derivative. Both $c(q;\theta)$ and $c_q(q;\theta)$ (the subscript denotes partial derivative) are zero for $q = 0$, for all θ.

There is a measure I of informed patients who can both observe and enforce any contract over quality of service. There is a measure U of uninformed patients who can neither observe quality, nor enforce it through a contract. These patients can observe only the treatment fee, which is specified in the potential contract they may accept, and the organizational from each doctor chose for her clinic: FP or NP. NP organizational form implies that the clinic has to make zero profits. FP organizational form implies that the clinic has to make non-negative profits, there is no upper limit on these profits. Communication between patients and monitoring of quality of service by informed, when the treatment is received by the uninformed, is not possible.

There is some consensus in the literature that producers (or employees) may care to some degree about the welfare of their consumers. This assumption seems even more compelling in the healthcare industry, where doctors are on the producer side and patients are on the consumer side.[48] I define $\alpha \in (0, 1)$, degree of altruism, to be the weight in doctor's utility function on the welfare of a patient the doctor treats. α is the same across all doctors.

It is well accepted in the literature that information asymmetry plays a major role in healthcare markets. However, a key ingredient in my model is *variation* in information levels across consumers. To capture this feature, I assume that information is scarce, or the measure of informed consumers is significantly smaller than the measure of doctors, formally: $D \gg I$.

For the purposes of the main implication of the model, it is enough to assume that there are

[48]The first paper with a similar assumption in healthcare is Newhouse (1970). General altruistic stream in the recent literature on theory of the firm is reviewed by Francois and Vlassopoulos (2008). Finally, altruistically oriented doctors assumption was adopted in a recent paper by Biglaiser and Ma (2007).

equal measures of patients and doctors in the economy (formally: $I + U = D$). However, in an extension to the model, I derive an empirical implication for relative length of waiting time for procedures between FPs and NPs. For this purpose I assume in the extension that the total measure of patients exceeds the total measure of doctors, while maintaining the assumption on the measure of informed patients relative to the one of doctors. The assumption in the extension is more realistic. As will become clear, however, this assumption does not change any of the qualitative results, although, it complicates the model.

2.2.2 Timing

Doctors move first and simultaneously choose whether to open a clinic or stay out of the market. If a doctor decides to open a clinic, at the same time she chooses one and only one contract of one of the following contract types:

1. "Non-Profit contract", denoted NP: a contract that specifies the fee, F, a patient would pay if treated by the doctor. The contract includes a requirement (monitored by the regulator) to set the quality such that $F = c(q; \theta)$;[49] in a compact manner I can denote any contract of this type generically as $[q, F; NP]$

2. "Discretionary For-Profit contract", denoted FP: a contract that specifies the treatment fee F, and leaves the quality choice q to the discretion of the doctor; compactly: $[q, F; FP]$

3. "Full For-Profit contract", denoted $\overline{FP}$: a contract that specifies both fee F and quality q; compactly: $[q, F; \overline{FP}]$

Note that if a doctor selects to stay out of the market, she can always choose contract of type $\overline{FP}$ and set the fee high enough and quality low enough, that no informed patient would be willing

[49] Although the regulator does not observe θ, it is assumed that she is able to determine perfectly ex-post whether the NP earned any positive profits. If indeed positive profits were made the punishment is assumed to be sever enough to discourage deviations from the break-even constraint ex-ante.

to select her as the healthcare provider (e.g. $F > q$), while no uninformed patient is even able to do so. Therefore, there is no need to specify an additional action for doctors uniquely for staying out of the market.

Patients move second. Patients observe the contract choice of each doctor, and simultaneously accept one contract among the ones doctors offered, or choose not to be treated. Uninformed patients cannot accept a contract of type $\overline{FP}$, as for them quality is uncontractable. As each doctor offers only one contract, acceptance of a contract by a patient is equivalent to provider choice. Moreover, offering a contract by a doctor is equivalent to opening a clinic with characteristics specified by the contract. Therefore, I use the three terminologies of contract acceptance and doctor or clinic selection by patients interchangeably throughout the paper.

2.2.3 Rationing Mechanism

Capacity constraints are a major feature of the healthcare sector. These constraints are responsible for waiting times, which are common in healthcare markets (Wilper et al., 2008). Although I assumed equal numbers of doctors and patients, I allow patients to queue in my model (not surprisingly, there will be no queueing *in equilibrium* in the baseline model). I model queueing through a probabilistic treatment, in order to retain the static nature of the model. To simplify the analysis of the model, I assume, without loss of generality, that one doctor can treat only one patient.

If more than one patient chooses the same doctor, the following rationing will occur: if there are informed patients who chose this doctor, then no uninformed patients will be treated. This feature proxies for the relative advantage in access to healthcare services of informed patients. Patients of the same type (who chose the same doctor) all have equal (uniform) probability of being treated (this is true for uninformed, only in the absence of informed patients who selected this doctor).

If more than one patient accepted the same contract of NP type (i.e. the probability of treatment by this doctor is less than 1 after all patients have made their choices), and there exists

a doctor who chose NP type contract that was not selected by any patient, one of the patients who selected a clinic where the probability of treatment is less than 1, will be automatically directed to the NP clinic without any patient.

In order to keep the baseline version of the model simple, I need to assume additionally the following: Once all of the NP clinics are selected by one patient, if an extra uninformed patient accepts a NP type contract, she is automatically directed to the clinic that offers the lowest consumer surplus if treatment were assured among the NP ones (it will be shown later that the consumer surplus when treatment is assured of a NP clinic can be perfectly inferred in equilibrium through the regulated fee). This assumption is not needed in the extension where the measure of patients exceeds the measure of doctors, however, it ensures that an equilibrium exists without imposting more structure on the cost function in the baseline model. While the baseline model, in turn, illustrates the main point of the paper in a much simpler way than its extension.

2.2.4 Payoffs

A patient receives zero payoff if she is not treated and

$$CS = q - F \tag{2.2.1}$$

if she receives a treatment of quality q and pays F for it. All patients have identical preferences. Additionally, patients are risk neutral both with respect to uncertainty over the occurrence of treatment (the rationing process) and quality of treatment (uninformed can neither observe nor contract over it).

If a doctor does not treat any patients, her payoff is zero. In case she treats a patient, she receives

$$v(q, F; \theta) = F - c(q; \theta) + \alpha CS(q, F) = F - c(q; \theta) + \alpha(q - F) \tag{2.2.2}$$

Recall, that α is the degree of altruism, equal across all doctors. All doctors have identical preferences.

2.2.5 Equilibrium

I use the concept of Perfect Bayesian Equilibrium (PBE) to solve this game. In general, PBE is a set of strategies and beliefs (one strategy and one belief function for each player), such that strategies are sequentially rational given beliefs, and beliefs are derived by Bayes rule on the equilibrium path. I proceed to defining these objects for the case of my model, thus describing the equilibrium.

Each doctor observes the realization of her type (the cost parameter θ) and chooses a contract offer to the patients that maximizes her payoff given patients' and other doctors' strategies. Formally, doctor's strategy can be represented as a mapping from her type space into to a space of contract choices, I restrict all doctors to use the same type-contingent strategy $s_D(\theta)$.

$$s_D : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}^2_+ \times \{NP, FP, \overline{FP}\}$$

under the restrictions outlined in Subsection 2.2.2. A typical member of the range of s_D is a vector $[q, F; \tau]$, where τ is the contract type: $\{NP, FP, \overline{FP}\}$ (refer to Section 2.2.2 for specification of each type of contract), q is quality and F is the fee choice, such that $(q, F) \in \mathbb{R}^2_+$.

There are two types of patients in the economy: informed and uninformed. Informed patients can observe all of the relevant parameters of the contract. Therefore, an informed patient's strategy is a mapping from all of doctors' choices made in the previous stage of the game into the entire set of contracts.

Uninformed patients observe only fee and contract type choices and can accept only contracts of types $\{FP, NP\}$. Uninformed patient's strategy is a mapping from doctors' quality and contract type choices into a subset of contracts, the ones of types $\tau \in \{FP, NP\}$. Additionally, uninformed

patients must form beliefs about the unobservable and payoff-relevant variable to them - quality.

A patient's belief function is a subjective probability that a doctor who has chosen a contract of type τ and set her fee at F has set quality less or equal to some q. Since all doctors of the same type θ choose the same contract (I focus on symmetric equilibria), the only relevant information for determining the subjective probability of quality provision by a doctor are the values of F and τ in her contract. Due to continuous nature of q, the belief function must depend on q as well, as it is a CDF. Denote this subjective CDF as $\mu(q; F, \tau)$. These beliefs are restricted to be the same across all uninformed patients.

This completes the definition of strategies and beliefs needed to specify an equilibrium for this game. Now, I move to a preliminary description of equilibrium values of beliefs μ, based on the degrees of information asymmetry that exist in equilibrium.

In many markets in the FP sector prices are weakly related to costs, due to market frictions. On the other hand, as NPs are not allowed to distribute profits, if consumers observe a variation in prices between various providers, these must indicate some differences in underlying quality, through differences in costs. This should make the relationship between prices and costs of inputs stronger in the NP sector. Stronger relationship between prices and costs mitigates the information asymmetry about quality in the NP sector. In my model, the NP organizational form fully eliminates the information asymmetry, while at FP clinics the information asymmetry is at its maximum. I show both of these features in the order just presented.

Suppose that at least one patient accepts a doctor's contract in equilibrium, then the payoff function of a doctor who chooses a contract of type NP is

$$v(q, F; \theta) = v(q, c(q; \theta); \theta) = \alpha[q - c(q; \theta)]$$

if she treats a patient in equilibrium. Recall that the doctor must set the quality such that she

breaks even (i.e. the treatment fee must equal the cost of provision), if she chooses a contract of type NP. The doctor maximizes this payoff w.r.t. (with respect to) quality, q. The FOC (First Order Condition) of the maximization problem is

$$c_q(q;\theta) = 1 \tag{2.2.3}$$

where subscript denotes partial derivative.[50] Define q^θ as the value of q that satisfies this equation (it will depend on θ). Thus, doctors who choose contract of type NP set

$$q(\theta) = q^\theta$$

As $c_q(q;\theta)$ is monotonically increasing in q and falling in θ, there exists at unique q^θ for each θ. Therefore, there exists a one to one mapping between $c(q^\theta;\theta)$ and q^θ. More formally,

$$c^{-1}[F; c^{-1}[q^\theta(\theta); F] = \theta(F)] = q^\theta$$

is unique,[51] hence, through the regulated fee, uninformed patients are able to perfectly infer the

[50] Equation (2.2.3) defines the socially efficient level of quality in a world without altruism. The welfare maximizing level of quality can be found in an altruistic world as well. Moreover, there exists a unique welfare maximizing level of F too, as beyond the transfer nature of the fee, it reduces CS, that doctors care about. The following two equations describe the socially optimal level of quality and fee as a function of doctor's type θ in an altruistic world

$$c_q(q;\theta) = 1 + \alpha \tag{2.2.4}$$
$$F = 0 \tag{2.2.5}$$

However, throughout the paper I use the term "socially optimal" or "first-best" level of quality referring to the one in a world without altruism. This level of quality is characterized by (2.2.3). In a world without altruism F acts as a transfer, and has no welfare effect, therefore, only one condition describes the social optimum.

[51] To understand the last expression consider finding a general functional relationship between F and θ. First, one needs to express q^θ as a function of θ, denote this $q^\theta(\theta)$. Then plug $q^\theta(\theta)$ into $c(q;\theta) = F$ instead of q to get the new one-variable (derived cost) function $c(q^\theta(\theta);\theta) \equiv \tilde{c}(\theta) = F$. By inverting $\tilde{c}(\theta)$, one can find the value of θ, as F is known. This series of operations was compactly denoted above as $c^{-1}[q^\theta(\theta); F] = \theta(F)$. The next step would be plugging the value of θ into the cost function again. Now it is a one-variable function of q. As F is known, it can be inverted again to find q^θ. Consider the following simple linear example: $c(q;\theta) = \frac{q^2}{4\theta}$ $(\theta > 0)$, then $q^\theta(\theta) = 2\theta$,

quality provided by doctors who chose the NP organizational form.

This feature gives the first property of the equilibrium belief function of uninformed patients. Following the convention, I denote equilibrium values of strategies and beliefs with a superscript star. Thus,

$$\mu^*(q; F, NP) = \begin{cases} 0 & \text{if } q < c^{-1}[F; c^{-1}(q^\theta(\theta); F)], \\ 1 & \text{otherwise} \end{cases}$$

Which implies perfect inference.

What is the value of μ^* for FP type contracts, $\mu^*(q; F, FP)$? Suppose that only uninformed patients accept FP type contracts in equilibrium.[52] Any inference about quality, chosen by doctors who offer these contracts, can come through signalling. Consider all signalling possibilities. The only feasible signalling instrument for a doctor who chose contract of type FP in order to differentiate herself from all other doctors who chose a contract of type FP, is the fee (the only observable parameter of the contract beyond the contract type). Charging higher fees is profitable for all doctors, irrespective of quality they set, as $v_F(q, F; \theta) = 1 - \alpha > 0$, from expression (2.2.2) (the subscript denotes a partial derivative). This is due to the fact that doctors care about their net revenue more than patient's payoff. There is a hidden action problem, on top of the hidden information: for any given fee doctors can always lower quality to reduce their cost. Thus, the fee signal is not informative.

To strengthen the above conclusion, consider a case where doctors can commit to some lower bound of q that varies with θ, that is incentive compatible as long as they are able at least to break-even (and it will be shown later to be true), however, are not guaranteed a compensation for any increase in costs, as in the case of NP type contracts. This feature eliminates the hidden

$F = c(q^\theta; \theta) = \theta$, therefore, $q^\theta = 2F$ is the simple inference.

[52] This assumption might seem too restrictive at first glance. However, I can always relax it by noting that the following analysis applies only to those FP type contracts not accepted by informed patients in equilibrium.

action aspect of the problem. The next paragraph shows that even in this case, no signalling is possible in equilibrium.

Since even FP clinics must at least cover their cost of quality provision, it is enough to consider only pricing at the break-even level as a signal of quality. Even if the belief function is such that by pricing at the break-even level (when quality is set at the incentive compatible lower bound) the doctors can perfectly reveal their type (and, therefore, quality), there is no way for consumers to induce the doctors to do so, as there exist no other sources of compensation on top of the fee. Doctors always have the incentive to set the fee at the level of the highest possible cost of provision of the incentive compatible lower bound of quality across doctor types $\theta \in [\underline{\theta}, \overline{\theta}]$. Therefore, the fee cannot act as a signal of quality even in the case of no hidden action problem. There is no equilibrium in which uninformed patients perfectly infer the quality set under contracts of type FP.

All doctors who offer contract of type FP in equilibrium set the same fee. Therefore, the value of F conveys no extra information under the contract of type FP. Formally, since $\mu^*(\cdot; F, FP)$ does not depend on F, I drop it, and with abuse of notation use $\mu^*(q; FP)$ in later discussions. The full description of equilibrium beliefs and strategies appears in Appendix $A.2$. The economic insights obtained along the solution path are summarized in the following section.

2.3 Solution

2.3.1 General Structure

My goal is to find the distribution of service quality levels across NP and FP organizational forms. I focus on an equilibrium where the highest levels of quality of service in the market are provided by highest ability doctors to informed patients under $\overline{FP}$ type contracts, the medium range of quality levels is provided to uninformed patients by the second tier of doctors in terms of ability

under NP type contracts. Finally, the lowest levels of quality are provided by the lowest ability doctors under FP type contract to uninformed patients. The remaining of this chapter shows that this structure can be supported in an equilibrium.[53] Section $A.1$ of Appendix A proves that the equilibrium proposed indeed satisfies these properties.

In Subsection 2.2.2 I outlined three possible contract types for each doctor. It turns out that there is a unique doctor-payoff maximizing contract within each contract type $\tau \in \{NP, FP, \overline{FP}\}$ for each doctor type $\theta \in [\underline{\theta}, \overline{\theta}]$. Once these contracts are found for each contract type as a function of θ, each doctor's optimization problem is reduced to choosing the contract type that provides her with the highest payoff among the three.

As I will show, the single crossing property holds for derived utilities (i.e. when expressing doctor's payoff as a function of her type, by plugging in the optimal choices of F and q expressed as a function of θ, and repeating this for each contract type),[54] and every contract dominates staying out of the market. Each contract type dominates the other two for a continuous subinterval of doctor types $[\underline{\theta}, \overline{\theta}]$. An interval of doctor types θ all choose the same contract type in equilibrium. For example, an interval of doctor types $[\theta^{NH}, \overline{\theta}]$, where $\theta^{NH} \in (\underline{\theta}, \overline{\theta})$ is some marginal doctor to be found later, all choose a contract that can be compactly denoted $[q, F; \tau] = [q^{\theta}, q^{\theta} - s_I; \overline{FP}]$, where q^{θ} was defined above to be the socially optimal quality for doctor of type θ (i.e. a scalar once θ is known) and s_I is a constant do be found below.

In equilibrium some patients accept NP type contracts, while others FP and $\overline{FP}$ type ones. As shown toward the end of Subsection 2.2.5, the quality and fee choices of doctors who choose a contract of type NP do not depend on the strategies of other doctors or patients and can be pinned down first. Intuitively, one might say that once a doctor chooses a NP type contract she stops behaving strategically due to the binding break-even constraint imposed by the regulator.

[53] Section $A.2$ of Appendix A proves that the combination of strategies and beliefs derided here is indeed an equilibrium.

[54] Each pair of the derived utilities crosses only once on the $[\underline{\theta}, \overline{\theta}]$ interval.

This fact considerably simplifies the analysis of the game.

If a patient who accepted a FP or $\overline{FP}$ type contract were to deviate to accepting a NP type contract she is guaranteed some payoff. This payoff will differ between informed and uninformed patients and does not depend on strategies of other doctors (by the conclusion of the last paragraph). Call the expected consumer surplus from this deviation for uninformed s_U and for informed s_I.[55]

Doctors have the first mover advantage and care about their payoff more than the consumer surplus of the patient treated by them ($\alpha < 1$). Therefore, those doctors who choose FP and $\overline{FP}$ type contracts set their qualities and fees in such a way that a patient would be just indifferent between accepting their contract and deviating to accepting some NP type contract.

When deriving an equilibrium, I must find fees and quality levels for contracts of type FP and $\overline{FP}$ that give each patient that accepts them her payoff from deviation. This approach guarantees a selection of every doctor who offered FP and $\overline{FP}$ type contracts by one and only one patient in equilibrium, on one hand, and acceptance of every NP contract, on the other hand, as patients are not able to earn higher payoffs by accepting FP or $\overline{FP}$ type contracts. I start deriving the equilibrium contracts in such a way that I ensure that they are accepted by (at least one) patient in equilibrium.

2.3.2 Non-Profit Contracts

First, I repeat the unique doctor-utility maximizing contract for NP type contract. As shown in Subsection 2.2.5, this contract is of the form

$$[q, F; \tau] = [q^\theta, c(q^\theta; \theta); NP]$$

[55]The rationing mechanism, described in Subsection 2.2.3, together with the suppositions that only uninformed accept NP type contracts in equilibrium and that every doctor is selected by only one patient in equilibrium, give the following values of s_I and s_U. s_I equals the total surplus from treatment by the highest ability doctor who offers a NP type contract in equilibrium. While s_U equals to half of the total surplus from treatment by the lowest ability doctor who offers a NP type contract in equilibrium.

This contract gives the derived utility function for doctors who chose a contract of type NP, denoted u^N as follows

$$u^N(\theta) = \alpha[q^\theta - c(q^\theta; \theta)] \tag{2.3.1}$$

Thus, I have found the first of the three derived utility functions that arise from choosing each of the contract types NP, FP and $\overline{FP}$. Ultimately, I will find the range of θ over which u^N dominates the other derived utility functions from choosing contracts of types FP and $\overline{FP}$. I proceed to finding the remaining two derived utilities.

2.3.3 Full For-Profit Contracts

Only informed patients can select $\overline{FP}$ type contracts. These patients perfectly observe the quality level and value every increase in quality by one unit as an increase in their willingness to pay by one dollar. Let s_I be the highest consumer surplus an informed patient can get by deviating to a doctor who chose a contract of type NP. The fee and quality set by a doctor under $\overline{FP}$ type contract must satisfy the following relationship

$$q - F \geq s_I$$

if the doctor is to be selected by at least one patient in equilibrium. The partial derivative w.r.t. to F of doctor's utility function in (2.2.2) is

$$v_F(q, F; \theta) = 1 - \alpha$$

This expression is positive as α is less than 1. Therefore, unless a doctor were constrained by patient's willingness to pay, her optimal fee would be infinite. Since doctors care about their payoff more than patient's payoff ($\alpha < 1$), they will set $F = q - s_I$, at the highest possible level as long

as at least one patient is willing to accept their contract.

This allows me to write out the utility function of a doctor who chose a contract of type $\overline{FP}$ and is charging the highest feasible fee, under the constraint that she is selected by (an informed) patient as

$$v(q, q - s_I; \theta) = q - s_I - c(q; \theta) + \alpha s_I$$

Doctor of type θ chooses quality of care q to maximize this expression. The FOC of this maximization problem is

$$c_q(q; \theta) = 1 \tag{2.3.2}$$

note that doctor's choices of fee and quality have no influence on s_I (due to lack of strategic responses by doctors who chose NP type contracts).

However, the last condition is identical to equation (2.2.3), that defined q^θ, the socially efficient quality level as a function of doctor's type θ (in a world without altruism). By the arguments in Subsection 2.2.5, it is also unique for each θ. Therefore, doctors who choose a contract of type $\overline{FP}$, set the quality level at its socially optimal level $q(\theta) = q^\theta$. Plugging this quality level back into the constraint that this doctor is selected by at least one patient, I find the optimal fee: $F(\theta) = q^\theta - s_I$. Therefore,

$$[q, F; \tau] = [q^\theta, q^\theta - s_I; \overline{FP}]$$

is the unique optimal contract choice (as a function of θ) for contract type $\overline{FP}$.

This allows to write out the derived utility function for doctors who selected a contract of type $\overline{FP}$ as a function of their type θ

$$u^H(\theta) = (q^\theta - s_I) - c(q^\theta; \theta) + \alpha s_I \tag{2.3.3}$$

I turn to finding the utility maximizing contract for contract type FP.

2.3.4 Discretionary For-Profit Contracts

I am looking for an equilibrium in which all informed patients accept a contract of type $\overline{FP}$. Hence, I need to consider only uniformed patients for accepting contracts of type FP in this equilibrium. Uninformed patients cannot observe the quality chosen by doctors. This leads to a possibility that doctors may find it profitable to signal their quality choices to patients. However, in Subsection 2.2.5 I have already shown that no such credible signalling exists. Therefore, uninformed patients form an expectation about quality for each doctor who chose a contract of type FP. This expectation is not affected by doctor's actions. Therefore, doctor's actions cannot affect directly valuation of her service by a patient.

The quality level that maximizes doctor's utility under this type of contract is independent of the fee charged. Dependence of optimal quality on the fee can come from consumer valuation, or direct entrance of the fee into doctor's utility function such that cross-partial derivatives are not zero. The first possibility is ruled out by the conclusion of the previous paragraph: doctors cannot influence consumer valuation. The second possibility is ruled out by examining the doctors' utility function, expressed in (2.2.2). Therefore, when solving doctor's maximization problem to find the optimal quality level q I can treat the fee F as a constant (i.e. independent of q).

The FOC of doctor's maximization of the expression in (2.2.2) w.r.t. to q, while treating F as a constant, is

$$c_q(q;\theta) = \alpha \tag{2.3.4}$$

recall that α is a predetermined constant equal across all doctors. Call the level of q that satisfies equation (2.3.4), $q^{\alpha,\theta}$. In general, it depends on doctor's type θ. There is a unique $q^{\alpha,\theta}$ for each θ, by strict convexity of $c(q;\theta)$ in q, and the fact that $c_q(q;\theta)$ is monotonically falling in θ.[56] Therefore, the unique utility maximizing level of q under contract type FP is $q(\theta) = q^{\alpha,\theta}$.

[56]These arguments are identical to the ones for the uniqueness of q^{θ}.

I need to find the utility maximizing fee under contracts of type FP. As discussed above, doctors cannot signal their quality, therefore, uninformed patients value quality of all doctors who chose a contract of type FP at the same level for all of these doctors. Call q^e the expected quality across all doctors who chose a contract of type FP. Formally

$$q^e = \int_A q^{\alpha,\theta} d\mu(q; FP) \tag{2.3.5}$$

Where A is the set of all doctor types who select a contract of type FP in equilibrium.

All uninformed patients value the quality of doctors who selected a contract of type FP at q^e. Call s_U the (highest expected) payoff that an uninformed patient can obtain by deviation to a doctor who chose a contract of type NP. Then, at least one patient will accept a contract of type FP if the doctor sets the fee such that

$$q^e - F \geq s_U$$

As shown earlier, doctors care about their direct payoff more than about the consumer surplus of their patients. Doctors have an incentive to raise the fee as long as at least one patient accepts their contract in equilibrium. Therefore, doctors who chose a contract of type FP all set $F(\theta) = q^e - s_U$. Notice that this fee is independent of doctor's type θ.

Now, I can derive the final indirect utility function for doctors from choosing a contract of type FP. Note that the unique optimal contract under this contract type is

$$[q, F; \tau] = [q^{\alpha,\theta}, q^e - s_U; FP]$$

and the derived utility is

$$u^A(\theta) = (q^e - s_U) - c(q^{\alpha,\theta};\theta) + \alpha[q^{\alpha,\theta} - (q^e - s_U)] \tag{2.3.6}$$

Equations (2.3.1), (2.3.3) and (2.3.6) describe the payoffs doctors get from choosing contracts of types NP, $\overline{FP}$ and FP, respectively. Each doctor will choose the type of contract that gives her the highest payoff. As u^N, u^H and u^A all depend on doctor's type θ, it is the case that there is only one optimal contract type for each type of doctor.

It turns out that under assumptions[57]

$$q^{\alpha,\theta} = \alpha q^{\theta} \tag{2.3.7}$$

$$c(\alpha q^{\theta};\theta) = \alpha^2 c(q^{\theta};\theta) \tag{2.3.8}$$

the following two properties hold

1. Any subset of $[\underline{\theta},\overline{\theta}]$ over which one contract type dominates the other two is an continuous interval.

2. The intervals over which one contract type dominates the other two are located in the following order along $[\underline{\theta},\overline{\theta}]$: $[\underline{\theta},\theta^{AN}]$, where FP dominates the other two contract types; $(\theta^{AN},\theta^{NH})$, where NP dominates; finally, $[\theta^{NH},\overline{\theta}]$, where $\overline{FP}$ dominates (θ^{AN} and θ^{NH} are the threshold values of θ).

These properties lead to the following proposition, which summarizes the main result of the model

Proposition 1 *Highest levels of quality of care in the economy are provided by FPs to informed*

[57] See Section $A.1$ of Appendix A.

patients. A second tier of quality levels is provided by NPs to uninformed patients. A third tier of quality levels is provided by FPs to uninformed patients.

For the proof see Appendix $A.1$.

In order to see the content of Proposition 1 (and the structure of the equilibrium) more intuitively, it is useful to examine Figure 2.1, which plots the indirect utilities u^A, u^N and u^H as a function of θ. Under contracts of type $\overline{FP}$ any increase in the total surplus from the transaction as a result of a more able doctor is appropriated by the doctor by raising F accordingly. Hence, u^H has the steepest slope in Figure 2.1. Under a contract of type NP the doctor appropriates only fraction α of the rise in the total surplus. Thus, u^N has the intermediate slope. Finally, under contracts of type FP, an increase in doctor's ability contributes to her payoff even less, as she cannot even raise F in order to match the increase in cost of provision (as F is already set at its highest level that the uninformed patients would be willing to pay under the information asymmetry). This generates the lowest slope of the three indirect utilities.[58]

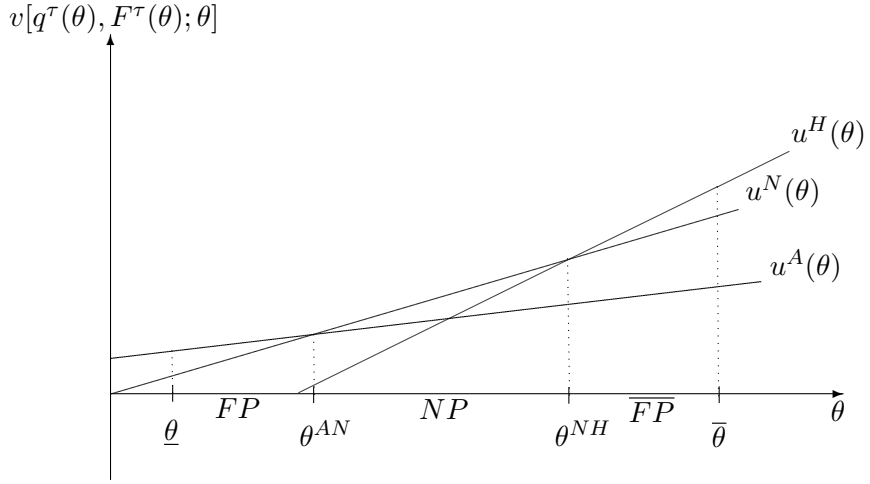


Figure 2.1: Partition of Doctor-Types Into Type-Optimal Contract-Classes

All doctors whose type falls into a segment of $[\underline{\theta}, \overline{\theta}]$ where u^A dominates both u^N and u^H offer

[58]The slope of u^A is positive due to the altruistic enjoyment of the increase in the consumer surplus.

contracts of type FP, doctors in the segment where u^N dominates the other two offer NP type contracts in equilibrium. In a similar manner, those of the highest ability falling into the segment where u^H dominates the other two functions offer $\overline{FP}$ type contracts. This is labelled on the vertical axis of Figure 2.1.

Figure 2.2 plots the choices of q that correspond to the doctor types in Figure 2.1. This generates the desired distribution of quality of care, where FP clinics (those doctors who chose contracts both of type FP and $\overline{FP}$) are on the tails of such distribution, while NP clinics are in the middle. The gap between the left FP segment and the NP segment comes from the fact that those doctors who offer contracts of type FP (who happened to be of the lowest ability in the economy) provide the shirking level of quality $q^{\alpha,\theta}$, while their colleagues who offer both NP and $\overline{FP}$ type contracts all provide their first-best level of quality q^{θ}. As both $q^{\alpha,\theta}(\theta)$ and $q^{\theta}(\theta)$ are continuous in θ, no other gaps are observed. I did not plot the distribution function, however, it should be some transformation of the uniform density, taking into account the functional forms of $q^{\alpha,\theta}(\theta)$ and $q^{\theta}(\theta)$.[59]

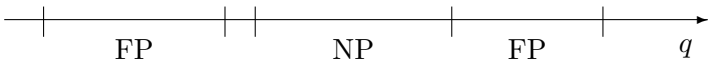


Figure 2.2: Distribution of Providers by Quality

2.3.5 Extension to Scarce Capacity

It is well known that waiting times for procedures are a common phenomenon in the healthcare industry. The existence of waiting times points to scarce capacity, otherwise all the patients could be treated the moment the seek care. One of the aspects of quality of care that most patients care about is the length of waiting times for procedures. As the usual measure of quality of care for a

[59]For the special case where both $q^{\alpha,\theta}(\theta)$ and $q^{\theta}(\theta)$ are linear in θ, the quality is distributed uniformly.

hospital are mortality rates (Shen, 2002), as long as the waiting times are below some threshold, they have little influence on this measure. For this reason and in order to simplify the main part of my model did not assume capacity constraints.

It is, however, of both theoretical and empirical interest to know which organizational form is more likely to have longer waiting times, or queues. In order to answer this question I extend my model to accommodate the case of scarce capacity of healthcare provision. In addition to answering the theoretical question, the predictions generated by this extension will provide an additional empirical implication, that will enable to test my model better.

Assume that $D \ll I + U$, while maintaining the assumption that $I \ll D$. Now, I do not need to assume that if there is more than one uninformed patient accepting NP type of contract, she is always sent to the lowest consumer surplus doctor. In the next paragraphs I show that qualitatively, the overall equilibrium structure does not change. The only difference is that now there is rationing in the NP sector, and no rationing in the FP sector.

I derive the new equilibrium by following these steps

1. Fix a FP or $\overline{FP}$ type contract.

2. Show that if more than one patient accept this contract, it must be the case that the doctor who offered it has a profitable deviation to raising her fee.

3. Argue that in equilibrium it cannot be the case that more than one patient accept either FP or $\overline{FP}$ type contract.

4. Argue that in equilibrium it must be the case that more than one patient accept each of the NP type contracts, or put differently: in equilibrium there is rationing in the NP sector.

Where point 2 leads to 3 and 3 leads to 4. It is crucial for the logical step between 3 and 4 to show that the payoff a patient can get by accepting a NP type contract is always positive. If this does

not hold, it could be the case that some patients stay out of the market. In this instance 4 does not follow from 3. For this reason, I start by considering the rationing mechanism in equilibrium in the NP sector. Then I show that only one patient accepts each of the FP and $\overline{FP}$ type contracts. Finally, I argue that the rest of the general structure of doctors' strategies (including the partition of their type space) does not change.[60]

If rationing is to occur in the NP sector in equilibrium, the rationing process in the NP sector must be such that the expected payoff across doctors is equalized. If this is not true, then patients would have a profitable deviation within the NP sector. Consider the following candidate for equilibrium probability of treatment by doctor of type θ who chooses a NP type contract

$$\pi(\theta) = \frac{D_{NP}}{U_{NP}} \times \frac{1}{\widehat{CS}(\theta)}$$

where D_{NP} is the measure of doctors who offer NP type contracts, U_{NP} is the measure of patients who accept NP type contracts, and $\widehat{CS}(\theta)$ is the consumer surplus doctor of type θ offers to the patient who is selected to be treated through the rationing process. Then the expected payoff from accepting a NP type contract is

$$\pi(\theta)\widehat{CS}(\theta) = \frac{D_{NP}}{U_{NP}} \tag{2.3.9}$$

which ensures equality of expected payoffs across NP clinics (as the expression is not a function of θ). This implies that the expected payoff of accepting any NP contract is simply the ratio of the measure of doctors who offer NP type contracts to the measure of patients who accept NP type contracts, and this is the value of s_U in a world with scarce capacity.

The expression in (2.3.9) is always be positive.[61] Hence, the payoff from a highest payoff

[60] The value of the marginal θ that is indifferent between FP and NP type contracts might change. Qualitatively, however, the equilibrium structure remains the same.

[61] See Lemma 3 in Appendix $A.1$ that implies that there will always be doctors who offer NP type contracts in equilibrium.

alternative for a patient selecting a contract of type FP or $\overline{FP}$ is always positive. This establishes the logical step between points 3 and 4 above.

In order to gain insights into the queueing process with respect to FP clinics, it is useful to consider the following lemma and its corollary.

Lemma 1 *In equilibrium, each patient who chooses a FP clinic is indifferent between her equilibrium clinic choice and deviation to a NP clinic.*

Corollary 1 *In equilibrium, every FP clinic that is selected by a patient, is selected by only one such patient.*

To see why Lemma 1 holds, consider the following arguments. In equilibrium, if a patient considers accepting a contract there are two possibilities: she strictly prefers accepting it, or she is just indifferent between the contract in question and a contract offered by another doctor. A doctor who opens a FP clinic (offers FP or $\overline{FP}$ type contract) is free to set her fee at any level above her cost of provision. As the doctor cares about her own net revenue more than the realized consumer surplus of the patient who is treated by her ($\alpha < 1$), she will always be willing to raise her fee as long as at least one patient chooses this doctor as the healthcare provider in equilibrium. If the expected consumer surplus this doctor offers is higher than the highest expected payoff a patient can get by accepting another contract, the doctor has a profitable deviation to raising her fee by some small $\epsilon > 0$, as the patient will still select her as the provider. Only if every patient who accepts a contract of type FP or $\overline{FP}$ is indifferent between accepting this contract and accepting another contract that gives her the highest payoff among all other contracts, no player has a profitable deviation. Therefore, in equilibrium, all patients who choose a FP clinic must be indifferent between their equilibrium play and a highest payoff alternative.

Corollary 1 can be restated in the context of the discussion of Lemma 1 above, as: if more than one patient of the same type (e.g. two informed, or two uninformed) select the same FP clinic,

given the equilibrium strategies of all other doctors and patients, there exists a higher fee that attracts one and only one patient. As $\alpha < 1$, it is obvious that this implies that there exists a profitable deviation for the doctor to raising her fee.

To see this more clearly, consider the case of one doctor selected by only two patients of the same type: either both informed or both uninformed. These patients have identical and risk-neutral preferences, hence, with the rationing mechanism, both value the contract at half of the consumer surplus this doctor offers to the patient selected by the rationing mechanism (the probability of treatment is $\frac{1}{2}$). Since two patients selected the same doctor, by Lemma 1, half of the consumer surplus if treatment were assured, must be at least as high as the payoff from a highest payoff alternative. This alternative must be identical for both patients, due to symmetric preferences. The doctor has a profitable deviation to raising her fee by some small $\epsilon > 0$. With the higher fee half of the CS this doctor offers to the patient selected by the rationing mechanism is less than the payoff from the alternative, however, by continuity of CS in F, CS is still higher then the payoff from the alternative. Therefore, in the next stage of the game (when patients move) one and only one patient will choose this doctor, while the doctor's net revenue (and, hence, utility, as $\alpha < 1$) increases.[62]

By Lemma 1, doctors who offer FP and $\overline{FP}$ contracts must at least match the rents their patients can get from deviation to a NP clinic. It will always be the case in equilibrium that only one patient accepts each FP or $\overline{FP}$ type contract. All of the informed patients still accept $\overline{FP}$ type contracts. Which implies that the measure of uninformed patients in access of the measure of doctors who select FP type contracts in equilibrium, must be accepting a NP type contract and participating in the rationing process in the NP sector.

The unique doctor-payoff maximizing contracts within each contract type, derived in Subsections 2.3.2 through 2.3.4, were not dependent upon the assumption on the relative sizes of the

[62]The doctor does not suffer any loss from the fact that one of the two patients no longer accepts her contract.

measures of doctors and patients. The only parameter that changes in the extension is the value of s_U. Therefore, the main properties of the equilibrium examined, stated in Proposition 1, still hold. However, now more than one uninformed patient accepts each NP type contract. By the arguments above, still only one patient accepts each contract of types FP and $\overline{FP}$. This gives rise to the empirical prediction summarized in the following proposition.

The following proposition summarizes the main result of the extension of the model to scarce capacity setting.

Proposition 2 *In equilibrium, NPs have longer queues than FPs.*

A number of empirical studies found some evidence suggesting that NPs indeed have longer queues than FPs. For example, Spector, Selden, and Cohen (1998) on p. 646 state that their results suggest that queues are longer at NP nursing homes. Weisbrod (1988), cited by Hirth (1999), finds that NPs are more likely to have waiting lists than FPs. Wilper et al. (2008) in their examination of waiting times in U.S. emergency rooms find NPs to have longer waiting times than FPs.

2.4 Welfare Analysis

The following propositions establish the welfare properties of the equilibrium just discussed (both in the baseline model and its extension).

Proposition 3 *In equilibrium, NPs provide treatment with socially efficient levels of quality in a world without altruism.*

Proof of Proposition 3 follows directly from expression (2.2.3).

My result that NP provide socially efficient level of quality is consistent with the Arrow-Hansmann hypothesis, which was treated rigorously by Hirth (1999) and Glaeser and Shleifer

(2001). However, recall that these simple models are not supported by empirical evidence, therefore, it is an important finding that the result holds in a more complex setting, crucial to understanding the existing empirical evidence.

FPs shirk on quality they provide to uninformed. This is an important feature of my model and it is in agreement with the Arrow-Hansmann hypothesis as well.

Proposition 4 *In equilibrium, those FPs which treat uninformed provide services with lower than socially efficient levels of quality.*

Proof of Proposition 4 follows directly from expression (2.3.4) and the fact that $\alpha < 1$.

Proposition 5 *In equilibrium, those FPs that serve informed patients provide services with socially efficient levels of quality.*

Proof of Proposition 5 follows directly from expression (2.3.2).

Proposition 5 is the central feature that makes my model more consistent than the previous ones with existing empirical evidence.

2.5 The Sizes of the Non-Profit and For-Profit Sectors

There has been a considerable discussion in the literature on the limits to the NP sector. Newhouse (1970) suggests that limits to the number of altruistically motivated individuals in the economy (e.g. donations), could determine the size of the NP sector. Steinberg (2006) lists inherent productive inefficiencies and appearance of firms that are able to distribute profits through perquisites (presumably, by raising costs; they are referred to as "FPs in disguise") as factors limiting NP provision. My findings allow me to conclude that there could be endogenous determinant of the potential size of the NP sector (eliminating the need to refer to various exogenous factors). The size of the NP sector is constrained by entrepreneurs' other contracting options.

My model enables to bound doctors who self-select into the NP sector. The upper bound on doctor ability in the NP sector comes from the fact that any doctor prefers to sign $\overline{FP}$ type contracts with informed patients over NP type contracts. Only the most able doctors get to contract with informed in equilibrium, as informed patients are scarce. Therefore, the highest ability doctor who is unable to find an informed patient to treat will open an NP clinic. This doctor's type constitutes the upper bound on ability in the NP sector.

On the other hand, the lower bound of the NP sector comes from the option to contract with uninformed patients through "discretionary FP" contracts. Under such a contract the doctor shirks on quality. If the doctor were able to appropriate the entire surplus from the transaction under FP type contract, every doctor would prefer FP type contract to NP (but not to $\overline{FP}$). However, in order to induce patients to accept FP type contracts, the doctors must forgo some of the surplus to the patient. The gap between total FP surplus and NP payoff is decreasing in doctor type. Starting from certain doctor type, this gap will no longer compensate for the amount of surplus that the doctor has to transfer to her patient to make her accept the FP contract, and NP type contract will be preferred by all higher types. In this way the FP type contract generates a lower bound (in terms of ability) on the size of the NP sector.

Doctor's payoff from choosing FP type contract (and optimal (q, F)) can be written as[63]

$$[q^e - c(q^{\alpha,\theta};\theta)] + \alpha(q^{\alpha,\theta} - q^e) - (1-\alpha)s_U \tag{2.5.1}$$

The first term (in square brackets) is the highest potential net income that a doctor can appropriate under information asymmetry, if the patient's highest payoff from deviation were zero

[63] See expression (2.3.6).

(initial reservation utility). If one is willing to assume

$$c_q(q^{\alpha,\theta};\theta)q^{\alpha,\theta\,\prime}(\theta) > |c_\theta(q^{\alpha,\theta};\theta)| \tag{2.5.2}$$

i.e. that the reduction in cost due to higher ability (RHS) is less than the increase in cost due to higher quality (motivated by altruism) as the ability rises (LHS), $c(q^{\alpha,\theta};\theta)$ is higher for more able doctors. As q^e (patients' expectation) does not rise with type, the highest potential net income is falling with type.

The second term in expression (2.5.1) is the altruistic impact of provision of quality below (negative) or above (positive) patients' expectation. Finally, the last term is the consumer surplus that the doctor must give up in order to have her contract accepted in equilibrium (due to altruism, a fraction α of it occurs back).

The last term in (2.5.1) does not depend on the doctor's type. Therefore, as θ rises only the first two terms change. Under FP type contract the doctor is compensated for any (voluntary) increase in quality only through altruism (the second term), while bearing all of the cost of such increase (the first term). On the other hand, under NP type contract the doctor can raise her fee with any increase in cost of provision while maintaining the same altruistic effect (which might even rise if $q^{\alpha,\theta\,\prime}(\theta) < q^{\theta\,\prime}(\theta)$). Therefore, the relative advantage of FP type contract relative to the NP one is falling with type. Starting from certain θ, NP type contract is preferred to FP, constituting the lower boundary of the NP sector.

Finally, I address the following comparative statics question: If we start with an economy where there are very few doctors, what happens to the equilibrium set of contracts as the measure of doctors increases? If $D < I$, then only $\overline{FP}$ contracts will be seen in equilibrium and $s_I = 0$. Once D becomes slightly greater than I, all of the informed patients accept $\overline{FP}$ type contracts with doctors in $[\theta^{NH}, \overline{\theta}]$, where $\theta^{NH} \equiv \overline{\theta} - I$. However, there are doctors in the economy such

that $\theta \in [\underline{\theta}, \theta^{NH})$. Think of a case where this set consists of only one doctor (measure zero). If this doctor chooses to offer NP type contract, she will get $\alpha[q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH})]$ (as her type is infinitely close to θ^{NH}). On the other hand, if she chooses to offer FP type contract her payoff will be $[q^{\alpha,\theta^{NH}} - c(q^{\alpha,\theta^{NH}}; \theta^{NH})]$ (see expression (2.5.1), $s_U = 0$ in this case). Obviously, she chooses whichever one is greater. Excluding pathological cases and non-monotonicities, the same contract will be chosen initially by all doctors in the set $[\underline{\theta}, \theta^{NH})$ as the measure of this set grows, but remains small. Hence, only when D is sufficiently greater than I we would expect to see the equilibrium market structure analyzed in this chapter, and the NP sector boundaries just discussed.

To make the claims of the last paragraph more concrete, think of a case where assumptions (2.3.7) and (2.3.8) about the relationship between q^{θ} and $q^{\alpha,\theta}$, and their respective costs of delivery, hold. As long as the measure of the interval $[\underline{\theta}, \theta^{NH})$ is small enough ($\underline{\theta}$ sufficiently high), such that

$$q^e = \int_{\underline{\theta}}^{\theta^{NH}} q^{\alpha,\theta}(\theta) d\mu(\theta; FP) > \alpha[q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH})]$$

there will be no NP type contracts offered in equilibrium.[64] For $\overline{FP}$ type contracts $s_I = q^{\alpha,\theta^{NH}} - q^e$, and for FP: $s_U = 0$. Only when the measure of the interval $[\underline{\theta}, \theta^{NH})$ becomes large enough such that $q^e \leq \alpha[q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH})]$, there will appear a NP sector in the economy (NP type contracts will be offered in equilibrium).[65]

[64] To see why no NP contract will be offered, think of a case with only one doctor not offering $\overline{FP}$. Then q^e is perfectly revealed to be arbitrarily close to $q^{\alpha,\theta^{NH}}$. In this case under FP type contract, the doctor can set $F = q^{\alpha,\theta^{NH}}$, and comparing the payoff from FP contract with the one for NP shows the dominance of the later by assumptions (2.3.7) and (2.3.8): $\alpha q^{\theta^{NH}} = q^{\alpha,\theta^{NH}}$, but $\alpha c(q^{\alpha,\theta^{NH}}; \theta^{NH}) = \alpha^2 c(q^{\theta^{NH}}; \theta^{NH}) < \alpha c(q^{\theta^{NH}}; \theta^{NH})$.

[65] If there is only one doctor offering a NP type contract, the expected payoff to an uninformed patient from accepting such contract is close to zero, as there will be rationing and the measure of uninformed accepting this contract is very large. Therefore, s_U, the payoff from deviation by U patients from accepting FP contract to NP contract, will be close to zero as well, which makes the system well-behaved even when the measure of doctors offering NP type contracts is arbitrarily small, but positive.

2.6 Extension to the Model: For-Profits in Disguise

It has been suggested in the literature that there may exist NP firms that exploit their NP status and maximize profits (see Hirth, 1999). These firms are referred to as "FPs in disguise" in the literature. A firm may register as a NP, however, if the regulator does not affectively prevent it from distributing operating surpluses, and it is not motivated not to do so due to altruism or reputational concerns, it is plausible that its managers, or the entrepreneurs who set it up, will be able to extract operating surpluses. These operating surpluses can be extracted in three ways: 1) by shirking on the effort the managers-entrepreneurs put into production itself or the organization of it, 2) through perquisites such as larger offices or a corporate jet, or 3) by increasing their own financial compensation. If those in control of the NP organization are able to extract the operating surpluses, they will be motivated to maximize these surpluses. This issue is important for the model presented in this chapter, since the assumption that NPs break-even may seem crucial for the results summarized in Proposition 1 to hold. This section examines the case when this assumption is not satisfied. I start with a short literature review to set up the stage.

If all of the NPs in the world offer maximized operating surpluses to those in control, the reason for the NP existence would disappear. It must be the case that some NPs indeed do not distribute operating surpluses, and, hence, do not act as profit maximizers. In such a case the FPs in disguise free-ride on the signal of the NP status of the "honest" NPs. Both Hirth (1999) and Glaeser and Shleifer (2001) assume that the extraction of profits from a NP firm is costlier than from a FP firm. In this sense all of the NPs of Glaeser and Shleifer (2001) are FPs in disguise; the NP status, nevertheless, provides a viable signal of quality. Hirth (1999), on the other hand, assumes that there exist honest NPs as well, in this setting he finds that the informativeness of the NP price as a signal of quality depends on it being too low for of FPs in disguise to choose the NP status. One of the main characteristics of the model presented in this chapter is the perfect inference of quality by

the consumers from the prices NPs charge. If the non-distribution (and, hence, zero profit in the static setting) is not enforced, this feature of perfect inference of quality through prices no longer holds.

Biglaiser and Ma (2007), as a part of their model of public healthcare provision, consider auditing by the regulator of quality that doctors provide in this sector. The doctors of Biglaiser and Ma (2007) choose a quality of treatment and report the cost of it to a national health authority. The national health authority reimburses the doctor for the cost she reports, unless the doctor is audited. If the doctor is audited, she is reimbursed exactly the cost she incurs. Doctors do not vary in their ability, hence, the provision cost of each level of quality is the same for all doctors. Doctors, however, vary in their degree of altruism: either high or low. In equilibrium, those doctors who are low on altruism choose lowest allowed quality, while those who are high on altruism choose the highest. The authors show that the low-altruism doctors free-ride on the high cost reports of the altruistic doctors, by randomizing between this report and a report of cost that is weakly below the cost of the audit. The main conclusion of this part of the model of Biglaiser and Ma (2007) is that it is optimal for the regulator to audit the cost reports corresponding to the highest allowed quality with some positive probability strictly less than 1. I rely on this result in what follows.

As argued before, the model of Biglaiser and Ma (2007) can be interpreted in terms of the NP theory in the following manner: doctors who work for the public service can be thought of as operating private NP clinics. In this sense, auditing by the regulator is equivalent to the enforcement of the break-even constrain in my model. The model presented in this chapter so far can be thought of as having an audit probability of 1. Based on the results of Biglaiser and Ma (2007), in order to relax this assumption it is enough to assume that the regulator in my model enforces the break-even constraint with probability strictly less than 1. Therefore, this framework is equivalent to relaxing the seemingly over-restrictive assumption that all NPs break-even.

It is plausible that when a new NP is set-up the non-distribution, and by implication the break-

even constraint is perfectly monitored. In the course of the life of the NP, however, more and more loopholes appear that allow to circumvent the non-distribution constraint. Glaeser (2003) suggests that any NP is at first dominated by donors, who closely monitor its performance, and the fact that no employee extracts any benefits. Gradually, however, the NP turns into an employee cooperative, as top-level employees gain more and more control over the organization. It is most likely that during this stage the non-distribution constraint becomes less effective.

To capture this idea consider an economy that lives for two periods. In the first period the game described in Section 2.2 is played. In the second period a new spell of disease occurs and it is distributed in the population such that the measures of informed (I) and uninformed (U) patients remain the same. Doctors move in the beginning of the second period. All doctors must continue operating clinics under the same organizational form as in the previous period, or leave the market. Doctors who chose FP or $\overline{FP}$ type contracts in the first period are allowed to offer any of those contract types in the second period. NP clinics may change their fee offers (as they cannot contract on quality, see Subsection 2.2.2). Any NP clinic that changes the fee offer is audited with probability 1 to ensure that it indeed breaks-even. If the fee of a NP does not change it is audited with probability $\gamma \in (0, 1)$, following the results of Biglaiser and Ma (2007). Patients move second and choose one clinic where to be treated, under the same rationing mechanism as described in Subsection 2.2.3, or opt not to be treated at all. At the end of the second period the payoffs are realized.

Assume that the discount factor between the two periods is equal across all doctors. This discount factor is high enough such that doctor's decisions in the second period cannot influence her choice of organizational form in the first period. As I focus on the second period in this section, one can think of this extension to the model as a one period game in which the organizational form is inherited by a new doctor who takes over the clinic. The heir is selected in such a way, however, that she shares the same ability θ with the retiring doctor.

As the second period cannot influence the actions of doctors in the first period, by assumption, I analyze only the game of the second period. I show that the an equilibrium of the second period possesses the properties summarized in Proposition 1. As argued in Subsection 2.3.1, the strategies of doctors in the NP sector are not influenced by the strategies of their peers in the FP sector. I start the analysis with the NP sector. No doctor in the NP sector changes the fee she charges between the two periods: if she is audited, this is her optimal fee; if she is not audited the change of fee will trigger an audit, hence, it is not profitable to change the fee. Doctors in the NP sector for whom the non-distribution constraint is no longer enforced may opt to lower the quality they provide to their patients in order to save on operating cost. Since all doctors maintain their degree of altruism, they would choose to set the quality at the lowest incentive compatible level $q^{\alpha,\theta}$, as shown in Subsection 2.3.4. All doctors for whom the no-audit state of thew world is realized will lower their quality, as they place a higher weight on their own net revenue than on the consumer surplus of the patient they treat. These clinics are FPs in disguise in my model.

In what follows I discuss first the inference of quality through fees in the NP sector in the second period. Then I turn to finding the second period equilibrium belief function of uninformed patients about the quality of care in the NP sector. I derive the equilibrium payoffs of uninformed patients in the second period. Finally, I derive the new equilibrium strategies of the doctors in the FP sector. I argue that the main results of the model summarized Proposition 1 hold even when some NPs do not break-even.

When the uninformed patients know that with probability γ a NP is not audited, and provides quality $q^{\alpha,\theta}$ rather than q^{θ}, inference of the quality of NPs from the fee becomes imperfect. When an uniformed patient observes a fee of a NP clinic (in the second stage), she can invert the cost function, knowing that the fee is set at the cost of provision of q^{θ}.[66] In this manner the consumer infers the value of θ. The patient does not know, however, whether the doctor has been audited or

[66] See Section 2.2.5 for the details of this inference process.

not. Therefore, with probability γ: $q = q^{\alpha,\theta}$, and with the complimentary probability: $q = q^{\theta}$. The patient forms an expectation about the quality of care for this doctor, which equals $\gamma q^{\alpha,\theta} + (1-\gamma)q^{\theta}$. Now I can turn to deriving the new equilibrium belief function of uninformed patients about quality conditional on the NP status of the clinic.

The equilibrium belief function of the uninformed patients about quality of clinics in the NP sector in the second period is

$$\mu_2^*(q; F, NP) = \begin{cases} 0 & \text{if } q < q^{\alpha,\theta}(c^{-1}(q^{\theta}(\theta); F)), \\ \gamma & \text{if } q < c^{-1}[F; c^{-1}(q^{\theta}(\theta); F)], \\ 1 & \text{otherwise} \end{cases}$$

With probability γ the NP is shirking on quality, and there is zero probability that the quality is less than that. By inverting the cost function for quality at the level q^{θ}, the patient learns the type θ of the doctor, which she plugs into the mapping $q^{\alpha,\theta}(\theta)$ to determine the shirking level of quality. With probability $(1-\gamma)$ the doctor indeed provides the level of quality that enables it to break-even, hence, the consumer can simply invert the cost function (knowing the doctor's type) to find the actual quality. Recall that for all θ, $q^{\alpha,\theta} < q^{\theta}$, this explains the second value of the belief function. On the equilibrium path no doctor sets her quality at any value not equal to $q^{\alpha,\theta}$ or q^{θ}, hence, the beliefs for this case can be specified arbitrarily, I set them to be the lowest possible quality.

What is the expected consumer surplus from a treatment by a doctor in the NP sector for an uninformed patient? I start with the baseline model, where the measure of patients equals the measure of doctors. The expected consumer surplus of a patient who chooses to be treated by a doctor who charges fee F in the NP sector is $\gamma q^{\alpha,\theta}(c^{-1}(q^{\theta}(\theta); F)) + (1-\gamma)c^{-1}[F; c^{-1}(q^{\theta}(\theta); F)] - F$. What is the payoff of an uninformed patient if she deviates from the FP sector to a NP clinic? If an uninformed

patient deviates to the NP sector, she receives half of the expected consumer surplus of the lowest ability doctor in the NP sector, according to the rationing mechanism from Subsection 2.2.3. The expected consumer surplus of the lowest ability doctor in the NP sector θ^{AN}, expressed in terms of doctor's type is $\gamma q^{\alpha,\theta^{AN}} + (1-\gamma) q^{\theta^{AN}} - c(q^{\theta^{AN}}; \theta^{AN})$, the expected quality net of the fee, which is set at the cost of provision of the first-best quality for θ^{AN}. Therefore, s_U, the payoff to an uninformed patient from a deviation from the FP to the NP sector in the second period when the measure of doctors exactly equals the measure of patients is: $s_U = \dfrac{\gamma q^{\alpha,\theta^{AN}} + (1-\gamma) q^{\theta} - c(q^{\theta^{AN}}; \theta^{AN})}{2}$.

In the extension of the model to scarce capacity the measure of patients exceeds the measure of doctors, while the measure of informed patients is less than the measure of doctors. I turn to finding the expected payoff of an uninformed patient in the NP sector in the second period, which also equals the payoff from deviation from the FP sector to a NP clinic for uninformed patients. Recall that in the FP sector the fees are set in such a way that only one patient selects each doctor, otherwise doctors have a profitable deviation to raising their fees. Hence, the rationing in the second period must occur only in the NP sector as it was in the first. I focus on the equilibrium probability of treatment discussed in Subsection 2.3.5:

$$\pi_2(\theta) = \frac{D_{NP}}{U_{NP}} \times \frac{1}{\widehat{CS}(\theta)}$$

where D_{NP} is the measure of doctors who offer NP type contracts, U_{NP} is the measure of patients who accept NP type contracts, and $\widehat{CS}(\theta)$ is the consumer surplus doctor of type θ offers to the patient who is selected to be treated through the rationing process. This expression for the probability of treatment does not change between the two periods. The value of $\widehat{CS}(\theta)$, however, becomes $\widehat{CS}(\theta) = \gamma q^{\alpha,\theta} + (1-\gamma) q^{\theta} - c(q^{\theta}; \theta)$. This implies that the expected payoff from a deviation of an uninformed patient from FP to NP sector and the equilibrium payoff from treatment in the NP sector, s_U, retains its first period value: $s_U = \dfrac{D_{NP}}{U_{NP}}$, the ratio of the doctors in the NP sector

to the patients there.

How do the equilibrium strategies of doctors in the FP sector change in the second period? First notice that no doctor would find it profitable to change her contract type between the two periods, as higher ability doctors can always outbid the lower ability ones when attracting informed patients. The higher ability doctors, however, offered $\overline{FP}$ type contracts (that only informed can accept) already in the first period. I begin with the FP type contracts. Notice that the level of quality selected by each doctor who chooses a FP type contracts in equilibrium does not change between the two periods. This is true since it was already set at the lowest incentive compatible level in the first period. For this reason the expected quality of treatment in this sector, q^e, does not change either. The expected payoff doctors offer under FP type contracts must equal s_U. If this payoff is below s_U, no patient would accept a FP type contract. If this payoff is higher than s_U, the doctor has a profitable deviation to raising her fee. Therefore, the expression for the fee choice under FP type contracts does not change between the two periods and equals: $q^e - s_U$, the value of s_U changes in the baseline, however, as derived above.

Now I turn to the equilibrium fee and quality choices in the second period under $\overline{FP}$ type contracts. As argued in Subsection 2.3.3, under $\overline{FP}$ type contracts doctors can benefit from setting the quality at the socially efficient level q^θ and raising the fee until the (informed) patient is indifferent between accepting the contract and deviating to the NP sector. This is achieved through setting the fee at $F = q^\theta - s_I$, where s_I is the payoff from a deviation to the NP sector for an informed patient. s_I equals the highest payoff an informed patient can obtain in the NP sector. In the second period its value can be expressed as $s_I = \max\left\{q^{\max(N)} - c(q^{\max(N)}; \max(N)), q^{\alpha,\max(Q)} - c(q^{\alpha,\max(Q)}; \max(Q))\right\}$, where N is the set of all doctor types θ who are operating NP clinics and breaking even in the second period, and Q the is the set of all doctor types θ who are operating NP clinics and shirking on quality in the second period. This completes the derivations of the doctor equilibrium strategies in the second period.

Do the equilibrium strategies of patients change between the two periods? Informed patients follow the same strategy in the second period as in the first one. If the measure of patients and doctors are equal, uninformed patients as well follow the same strategy as in the first period. If the measure of patients exceeds the measure of doctors, uniformed select NP clinics such that the equilibrium probability of treatment for each doctor, $\pi_2(\theta)$ as defined in this section, is satisfied, and accept FP type contracts such that one patients accepts each contract. This completes the derivation of the equilibrium for the second period (notice that as the quality choices of doctors under FP type contracts do not change between periods, the equilibrium belief function of uninformed about quality under these contracts does not change between the periods).

The main purpose of this section is to discuss the implications for the main properties of the solution of relaxing the break-even constraint for NPs. First notice that no doctor in the FP sector changes the quality she offers between the two periods. The quality offered in the NP sector changes. Particularly, the average quality in the latter sector falls in the second period. Recall, however, that all doctors in the NP sector have higher ability than the ones who offer FP type contracts. Therefore, even those doctors who shirk on quality in the second period in the NP sector still provide higher quality than the doctors in the low quality FP sector (as the shirking level of quality rises in ability, under assumptions (2.3.7) and (2.3.8)). Additionally, the doctors in the high quality FP sector are of higher ability than any doctor in the NP sector and all provide first-best quality, which rises in ability. For this reason the quality of any given clinic in the NP sector remains bounded by the quality of the lowest quality clinic in the high quality FP segment and the highest quality clinic in the low quality FP segment. This ensures that the distribution of quality in the market retains its main properties in the second period, which are summarized in Proposition 1.

In this section I examined the consequences of relaxing the assumption of perfect oversight by the regulator over the fact that no operating surplus can be extracted by the entrepreneurs-managers

in NPs. Economic literature refers to NPs that maximize and pass their operating surpluses to their employees or managers as "FPs in disguise". I reviewed this literature. Relying on previous results, I model FPs in disguise as a consequence of probabilistic enforcement of the break-even constraint. To capture the fact that NPs are less likely to be perfectly overseen in the later stages of their existence relative to the earlier ones, I allow the probabilistic enforcement only in the second period of the economy, after the organizational forms were already selected by the doctors one period earlier. Those NPs which are lucky enough to escape an audit by the regulator decrease their quality. The uninformed patients expect this and revise their equilibrium strategy in a world where the measure of patients exceeds the one of doctors such that the probabilities of treatment adjust in the NP sector to equalize the expected payoffs across doctors. FP firms anticipate the decreases in quality in the NP sector and the reduction in the welfare there as a result (in a world with equal measures of patients and doctors) and raise their fees accordingly. The main properties of the predicted distribution of quality in a mixed market (of NP and FP firms) do not change.

2.7 Conclusion

In this chapter I develop a model of a healthcare market in which altruistic doctors offer NP and FP contracts to patients. When selecting the type of contract to offer, the doctors have to choose also the level of (costly) quality of care to provide and the fee to charge. Patients vary in the extent to which they can observe (and contract on) the quality of care offered by doctors. The ability to offer NP contracts enables doctors to reveal the quality they chose to uninformed patients through the break-even constraint.

I showed that, in equilibrium, the informed patients accept FP contracts that offer the highest quality of care in the economy. The NP contracts are accepted by uninformed patients and the doctors who chose them offer second tier of quality levels. Finally, FP contracts intended for

uninformed patients were chosen by doctors who offer the lowest quality of care. The equilibrium strategy of the latter two groups of doctors (NP and low quality FP) is in accordance with the Arrow-Hansmann hypothesis: FP providers shirk on the quality they offer to uninformed consumers, while NP ones do not. The innovation of my model is due to the fact that there exists the first group of providers who offer high quality of care under the FP organizational form. This feature of the model makes it more consistent with the existing empirical evidence on relative quality of FPs and NPs than the traditional Arrow-Hansmann hypothesis.

The main innovation of my model is the fact that, contrary to the previous literature, doctors self-select between the NP and FP sectors based on ability and not the degree of altruism. The model demonstrates that altruism is needed for NPs to arise, however, the differences in quality between the NP and FP sectors can be explained through differences in doctors' ability between the sectors. The self-selection in terms of types allows me to address the sizes of the two sectors and their boundaries in a rigorous framework without relying on exogenous factors, the way the previous literature does (Newhouse, 1970; Steinberg, 2006).

The new model has two main empirical implications: FPs are over-represented on the tails of the distribution of firms by quality in a market where NPs and FPs operate side by side; and NPs have longer queues, or waiting times for procedures, than FPs.[67] Chapter 3 derives additional comparative statics implications employing the theoretical framework developed here: 1) An increase in the proportion of informed patients increases the mean FP quality for small proportions of informed, and decreases the mean FP quality for large proportions; 2) An increase in the proportion of informed patients reduces the variance (standard deviation) of NP quality; 3) Higher quality NP hospitals have longer queues than lower quality NPs. Chapter 3 proceeds to test both the two main hypotheses and the three additional comparative statics implications on a

[67]The discussion in Subsection 2.3.5 also implies that higher quality NPs have longer waiting times (lower probability of treatment) than lower quality ones.

newly contracted dataset and additional recent data on waiting times.

Chapter 3

Empirical Analysis

3.1 Introduction

In the U.S. healthcare markets, more than in other markets, we observe a mix of organizational forms. In the late 1990s in the U.S., in the Dialysis Centres, for example, For-Profit (FP) firms had a 70% market share, with Non-Profits (NP) serving 30% of the market. On the other hand, in Acute Care Hospitals during the same period FPs had an 18% market share, while NPs had a 60% market share with government owned facilities serving the remaining of the market (Schlesinger and Gray, 2006). Beginning in mid-1980s a large number of NP healthcare providers in the U.S. has converted to FP status.[68] Recent newspaper articles raised concerns over reductions in quality of care, mostly in conversions from NP to FP status.[69]

A significant number of empirical studies has aimed to determine the relationship between organizational form an quality (Rosenau and Linder, 2003; Schlesinger and Gray, 2006). The empirical research finds evidence of all three possibilities of the impact of organizational form on the (average) quality of healthcare:[70] better NP quality of care, no difference in quality between NPs and FPs and FP superiority quality-wise (Rosenau and Linder, 2003; Schlesinger and Gray, 2006). Not only that the three empirical results seem inconsistent with each other, the last result

[68] Shen (2002) reports: "between 1985 and 1999 more than 700 hospitals in the United States changed their ownership status. These hospital conversions, many of them from not-for-profit to for-profit status, have captured much attention."

[69] See the newspaper citations in the beginning of Section 1.1 of Chapter 1.

[70] See the discussion on the methodology of existing studies in Section 1.5 of Chapter 1.

is at odds with existing economic theory of NPs as well (Hirth, 1999; Glaeser and Shleifer, 2001).

In Chapter 2 of this thesis I develop a new theory of healthcare markets that allows to reconcile the seemingly contradictory results of the empirical studies. In this chapter I test the central implication of the new theory as well as additional comparative statics implications. The following hypotheses are tested:

1. FP hospitals have higher variance (standard deviation) of quality than NP hospitals.
2. An increase in the proportion of informed patients increases the mean FP quality for small proportions of informed, and decreases the mean FP quality for large proportions.
3. An increase in the proportion of informed patients reduces the variance (standard deviation) of NP quality.
4. NPs have longer queues that FPs.
5. Higher quality NP hospitals have longer queues than lower quality NPs.

The rest of the Introduction is organized as follows. Subsection 3.1.1 reviews the model developed in Chapter 2. Subsection 3.1.2 derives the implications of the model that lead to the hypotheses just outlined. The model provides additional empirical implications that will not be tested in this thesis, however, they are discussed here as well. In subsection 3.1.3 I briefly review the methodology that is used to test the hypotheses above.

3.1.1 Theoretical Model of NPs

Chapter 2 develops a model of a healthcare market in which doctors supply zero or one unit of healthcare service and patients demand zero or one unit. The healthcare service may vary in quality, which patients value. There are two type of patients in the model: "informed" who are able to

observe and contract on quality, and "uninformed" who can neither observe nor contract over it. The provision of quality is costly for the doctors, and the cost depends on doctor's unobservable ability. The doctors are located along an interval of abilities. For higher ability doctors it is cheaper to provide the same level of quality.

Doctors move first and offer contracts for service provision to patients. Doctors choose among three types of contracts (each doctor offers no more than one contract): NP without contracting on quality, FP without contracting on quality, or FP with a contractual obligation to provide a certain level of quality. Additionally, doctors choose the quality and fee at which to offer the service. All contracts must specify a service fee. If a doctor decided to open a NP clinic[71] she must set the fee in such a way that it exactly equals the cost of provision of quality.[72] The patients move second and accept a contract (each patient accepts no more than one contract). Uninformed patients are not able to accept contracts that specify quality of care (some of the FP contracts). In the last stage of the game treatments and payoffs are realized.

Doctors are not able to earn positive profits if they offer a NP contract. Therefore, one needs to make an additional assumption to ensure that NPs do appear in equilibrium. The model assumes that each doctor has a positive weight (strictly lower than one) in her utility function on the welfare of the patient she treats. This weight, or degree of altruism, is equal across all doctors.

The stylized model just outlined yields the following main result. In equilibrium, the highest ability doctors open FP clinics and treat informed patients. The medium interval of abilities chooses NP clinics, finally, the remaining low-ability doctors choose FP clinics. The equilibrium quality choice of a doctor rises in her ability. The model predicts the distribution of qualities across the organizational forms as depicted in Figure 3.1.

[71] For the purposes of the exposition it is more intuitive to think that offering a contract is equivalent to opening a private practice, or a clinic. I draw implications for hospital markets, as clinics can be seen as a stylized way of modelling hospitals under certain conditions, such as constant returns to scale.

[72] The plausibility of this assumption is discussed in detail in Section 2.1 of Chapter 2.

Figure 3.1: Distribution of Hospitals by Quality

The model is setup in such a way[73] that any doctor prefers treating informed patients under FP organizational form to the other two contracting options (treating uninformed under NP or FP). Informed patients are willing to pay for any increase in quality. When treating informed under FP organizational form doctors maximize the total surplus from the transaction. The first mover advantage allows the doctors to appropriate portions of this surplus. This results in provision of the socially optimal level of quality on one hand, and most desirable for doctors surplus extraction (subject to informed patients' payoff from deviation) on the other.

The uninformed patients cannot deduce quality that they actually receive (signalling is impossible), therefore, they must form expectations of it. If doctor's (optimal) quality choice happens to equal the one expected by the uninformed, the doctor prefers to treat uninformed patients under FP organizational form to treating those patients under NP one. In this case, however, the doctor will shirk on quality. Doctors vary in quality they set when they treat uninformed in FPs due to symmetry in altruism (non-zero quality), but asymmetry in the cost of provision of quality: the doctor-optimal level increases with doctor's ability. This feature causes doctors who provide higher than the expected (by uninformed) level of quality to be "under-compensated" for it. If the doctor's (optimally) chosen level of quality is significantly higher than the expectation of the uninformed patients, she prefers to open a NP clinic.

Under NP organizational form the uninformed patients are certain that the fee charged reflects actual cost of the quality provision (due to the break-even requirement), thus, they are willing to pay

[73] In order to grasp the intuition of the model in a simpler way, one should ignore any strategical considerations on the side of patients beyond what is stated in the current paragraph. Simply assume that they accept the contracts. The proof that the proposed structure is indeed an equilibrium is sketched in Section $A.2$ of Appendix A.

any fee the doctor may charge. Commitment to the NP organizational form implies that doctor's compensation is merely the altruistic enjoyment of the total surplus from the transaction that she offers to her patient. Therefore, doctors maximize such surplus and provide socially optimal level of quality. This explains the gap between the lower bound of quality in the NP segment, versus the upper bound of the FP one in Figure 3.1. The upper bound of the left FP segment is the highest shirking level of quality of a doctor whose ability is almost the same as the one located on the lower bound of the NP segment. The doctor (at the lower bound) in the NP segment, however, provides her first-best level of quality.

In Figure 3.1 the FP segment on the right corresponds to the doctors who treat informed patients. The lower bound of this right segment is determined by the fact that no doctor, for whom it is optimal to set quality anywhere below this bound, will be able to attract an informed patient, as all of them are already being treated by her more able colleagues.

It is easy to see how this model is consistent with all of the existing empirical evidence. Datasets that come from hospital markets populated by both organizational forms contain under FP observations both hospitals serving informed patients (high quality) and the ones serving uninformed (low quality). Therefore, my theory is able to explain the evidence from these studies as: depending on the relative sizes of the pools of high quality FPs, NPs, and low quality FPs, *on average* either organizational form can have higher quality of output. I proceed to discussing the new empirical implication that this model offers.

3.1.2 Empirical Implications of the Model

After explaining the structure of the theoretical model of Chapter 2 I turn to the derivations of the Hypotheses 1 to 5 as well as additional empirical implications. Figure 3.1 highlights the central implication of the model: FP firms are over-represented on the tails of quality distribution, while their NP counterparts tend to be in the middle of such distribution.

Similar to most of the empirical literature on NPs, I test the predictions of the model on the U.S. hospital data. Due to small number of hospitals in geographical markets in the U.S.[74] even if I expected the model to hold precisely, it would not be possible to test the implication regarding the relative prevalence of the organizational forms along the distribution of quality. I believe, however, that this stark implication is generated due to the stylized nature of the model. A direct consequence of the relative prevalence implication, nevertheless, should hold: the variance of quality should be higher in the FP sector relative to the NP one, as summarized in Hypothesis 1 above.

Any empirical implications about the effect of changes in proportion of informed patients can be derived by noting that an increase in this proportion will shift the lower bound of the right FP segment in Figure 3.1 to the left. In other words, more doctors would be able to attract informed patients (as now there are more informed patients), however, these doctors will be of lower ability (hence, provide lower quality) than the ones who were treating informed patients before such change. I focus on the baseline model, where the lower bound of the NP sector does not depend on the proportion of informed.[75]

I start by deriving Hypothesis 2. At first, the mean quality of the FP sector (which is computed over both the left and the right segments) falls somewhere below the lower bound of the right FP segment. As some of the uninformed patients become informed, more and more doctors are able to contract with informed. Due to the competition between doctors, only the highest ability (and quality) doctors (who are not treating informed initially) are able to attract the new informed patients. The quality offered by the marginal doctor who treats informed is falling as more patients

[74] See Figure $B.2$ in Section $B.2$ of Appendix B for the distribution of the number of hospitals by Metropolitan Statistical Area (MSA). In the empirical analysis that follows I treat each MSA as one market.

[75] In the baseline model the measures of doctors and patients in the economy are equal. There are two marginal doctors types: one between left FP interval and the NP interval - θ^{AN}, and the other between the NP interval and right FP interval - θ^{NH}. The value of s_U - the equilibrium payoff of uninformed patients at FPs (the left FP segment), is set at half of the (first-best) consumer surplus of θ^{AN} by assumption. Hence, s_U does not depend on θ^{NH}. Thus, the payoffs from deviations do not change for either the uninformed patients or the doctors types around θ^{AN}. For this reason I treat θ^{AN} as fixed in the analysis below.

become informed. By the supposition that initially mean FP quality falls in the interior of the NP interval (below the marginal doctors treating informed), the new FP clinics offer higher than the average FP quality. This implies that the mean FP quality rises for small proportions of informed patients. However, when the right FP segment is large enough, the mean FP quality will fall within it. The marginal FP clinic that treats an informed patient provides quality below the FP average in this case. By implication, the mean FP quality decreases as the proportion of informed patients rises for high proportions of informed. This leads to Hypothesis 2.

The second comparative statics prediction is summarized in Hypothesis 3. As more patients are becoming informed, more doctors are able to open FP clinics and treat them. It is fairly easy to see in Figure 3.1 that when the lower bound of the right FP segment moves to the left, as long as the lower bound of the NP segment is unchanged,[75] the length of the NP segment will decrease. In distributional terms, it implies that the support of the distribution of quality of service for NPs is decreasing, which further implies that the variance of NP quality falls with the proportion of informed patients.[76] This feature generates Hypothesis 3.

Finally, the baseline model allows to generate the following slightly more complex prediction: a disproportionately higher cost shock for lower quality hospital will 1) cause conversions from FP to NP organizational form, 2) increase the variance of NP quality, 3) decrease the mean of NP quality. Assume that clinics in the economy experience a cost shock. Suppose this cost shock is not uniform across doctors. Specifically, increase in cost for lower ability (higher cost) doctors is higher relative to the increase in cost for higher ability doctors. This shock will reduce the attractivity of the FP organizational form for doctors in the left FP segment in Figure 3.1 and the ones close to the upper bound will convert to NPs. As a result, the lower bound of the NP segment will shift to the left.

[76] It is easy to show that for any random variable with bounded support for which mean and variance exist the following is true: the standard deviation of the random variable is bounded from above by the length of the support. Therefore, the statement must be true at the limit. The statement is trivially true if the quality is distributed uniformly (e.g. because the quality choice is a linear transformation of doctor's ability).

This causes the mean NP quality to decrease (as the "new" NP clinics are of lower quality). At the same time, variance of the NP quality increases, as the NP interval becomes longer.[77] I do not test this implication.

I turn to Hypothesis 4. This hypothesis is derived from the extension to the baseline model, which assumes that the demand for healthcare could not be satisfied due to capacity constraints if treatment fees were zero. First, I introduce the rationing mechanism to which I refer as queueing.[78] Queues impose utility costs to consumers. Due to the static nature of the model, I model queueing through probabilistic treatment. I impose equal probability of treatment for all patients queueing at the same provider. Thus, the valuation of treatment for a queueing patients is reduced from its total value. If the queue is at least 2 patients long the probability of being treated is less than 1 (in fact, it never exceeds $\frac{1}{2}$).

A FP provider is able to raise the fee she charges in a way that allows her to appropriate the utility losses to the consumers as a result of queueing. In equilibrium only one patient chooses such a provider.[79] NPs are required to break-even. For this reason they are not able to raise their fees like their FP counterparts. Additionally, the consumer surplus NP providers offer is high (they provide first best quality and charge a break-even fee). If the number of patients in the economy exceeds the number of doctors, however, the number of informed patients does not, queues of uninformed patients will form. These queues will exist solely at NP providers. The empirical prediction summarized in Hypothesis 4 takes into account the abstractions of the model made for tractability, therefore, it is less stark than the result itself.

My model generates yet additional hypothesis regarding queue length: higher quality NP hospitals have longer queues than lower quality NPs. In equilibrium the payoff to a patient from

[77]The upper bound of the NP interval is determined solely by the proportion of the informed patients in the economy. Hence, it will not change as a response to such cost shock.

[78]This discussion by no means implies that queues indeed arise in equilibrium, it simply describes the rationing mechanism if they were to arise.

[79]See Lemma 1 and Corollary 1 in Chapter 2.

selecting a NP provider has to be equal across all NP providers (otherwise some patients selecting NP providers will have a profitable deviation to another NP provider).[80] As queueing is modelled through probabilistic treatment and all consumers are risk neutral, it has to be the case that the expected consumer surplus is equal across all NP clinics. Under NP contracts the "lucky" patient, who is chosen to be treated, receives the entire surplus from the transaction. This ex-post consumer surplus varies across doctors (since they differ in ability and, hence, in the quality they they choose in equilibrium). To keep the ex-ante payoff equal, the probability of treatment must differ across NP providers. As the consumer surplus (total surplus from the treatment in this case) rises with doctor's ability, one can conclude that higher ability doctors have lower probabilities of treatment, or longer queues. Ability is unobservable, while chosen quality is measurable empirically (see the discussion in Subsection 3.1.3) and it monotonically rises with ability in equilibrium. The following empirical prediction is generated: higher quality NP hospitals have longer waiting times.

In the next subsection I review the main methodological issues that arise in the testing of Hypotheses 1 through 3. These issues are discussed in detail in Sections 3.2 through 3.4. Hypotheses 4 and 5 are tested on a separate dataset. These tests constitute a contribution to an entirely different stream of literature that is fairly sparse at this stage. I deffer the discussion of the empirical strategy for the tests of the last two hypotheses until Section 3.6 of this chapter. The data seems to support Hypotheses 1, 2 and 4. Hypotheses 3 and 5 are supported rather weakly.

3.1.3 Empirical Methodology

There are three main methodological issues that I address in order to test Hypotheses 1 through 3: 1) how to measure hospital quality, 2) how to bound geographical markets, 3) at what level to perform the analysis (patient, hospital, or market). For Hypotheses 4 and 5 I need to find a

[80] Recall that in equilibrium only uninformed patients are treated in the NP sector. These patients, however, can infer the quality of care perfectly, as NPs set their fees at the break-even level (see footnote 51 in Subsection 2.2.5 of Chapter 2).

measure of waiting time length. I review those issues briefly in this subsection.

Hypotheses 1 through 3 involve hospital quality. Previous empirical literature on hospital quality has devised measures of quality. I review those measures in detail in Section 3.2. There seems to be a convergence in the recent literature toward measuring hospital quality in terms of mortality rates: higher mortality implies lower quality (McClellan and Staiger, 2000; Shen, 2002). Studies that examine the effect of organizational form on hospital quality estimate a regression of mortality rates for a certain procedure at a hospital on a set of hospital characteristics and two dummies for the organizational form (usually NP and government ownership, with FP omitted) (Schlesinger and Gray, 2006). I use 30-day post-admission risk-adjusted (standardized by patient characteristics) mortality rates (RSMR) for Heart Attacks, obtained from the U.S. Department of Health & Human Services, Hospital Compare.[81]

It has been recognized in the literature that one needs to control for market-level variables when estimating the impact of organizational form on quality (McClellan and Staiger, 2000; Shen, 2002). Market-level controls usually include demographic and market structure variables. In order for the controls to be meaningful, the researcher must consistently bound geographical markets for hospitals. For the purposes of my study I use Metropolitan Statistical Area (MSA) boundaries as the boundaries of geographical markets for hospitals. All hospitals located within a given MSA fall into one geographical market. Advantages of this market definition include easy access to demographic data and lower likelihood of defining markets too narrowly. It is less likely that a significant amount of patients seeks emergency care for Heart Attacks outside of the MSA of their residence. At the same time all hospitals within the MSA (that treat Heart Attacks) have the potential for admitting patients that suffered a heart attack within the boundaries of the MSA (Horwitz and Nichols, 2007).

Initially, empirical studies on hospital quality used patient-level data directly in the estimating

[81] `http://www.hospitalcompare.hhs.gov/` accessed on 3 December 2008 at 13:07.

equations (Shen, 2002). Shen (2002) recognizes the advantages of estimating mortality rates for each hospital in the first stage and using the estimated values in the organizational form regressions in the second stage. This procedure treats each hospital as one observation. My innovations comes from further aggregation of the unit of analysis to the market level. I treat each MSA (market) as one observation. The justification for this approach is closer correspondence between the hypotheses being tested and the empirical analysis.

Hypotheses 4 and 5 involve length of waiting times for procedures. I managed to obtain data on waiting times for patients from the arrival to an emergency department until they see a physician. The issues related to the tests of the last two hypotheses are fairly removed from the issues involved in the test of Hypotheses 1 through 3. Hypotheses 1 through 3 constitute the main implications of the model and are concerned with hospital quality of care. The quality of care is the central issue addressed in this thesis. For this reason I discuss in detail the issues of hospital quality and the methodology for the tests of Hypotheses 1 through 3 first. I defer the discussion of the methodology and the data for the test of Hypotheses 4 and 5 until Section 3.6.

There exist contradictory empirical evidence regarding the impact of organizational from of hospitals on the quality of care. This thesis proposes an extension to the current theory of NPs that has the potential to reconcile the existing empirical results. This theory is stated in the rigorous model I develop in Chapter 2. The implications of this theory lead to Hypotheses 1 through 5. A test of these hypotheses will validate the theory and the explanation of the previous empirical results that were seen as contradictory. If the model finds support in the data, it can be viewed as a valuable extension to the economic theory of NP organizations. In order to perform the empirical investigation I must address the issues of measurement of hospital quality, market boundaries and level of analysis. My innovation on the empirical side comes from the use of a whole market as a unit of analysis, which has not been done before.

The remaining of this chapter is organized as follows. Section 3.2 addresses the issue of how

to measure hospital quality. Section 3.3 offers a literature review that focuses on how the previous studies solve the methodological issues I encounter in the current analysis. Section 3.4 describes the data and the estimation strategy. Section 3.5 reports the results of the estimations. In Section 3.6 I present the methodology and the results of the tests of Hypotheses 4 and 5 above. Section 3.7 concludes.

3.2 Measuring Hospital Quality

Chapter 2 of this thesis started with the notion that "the quality of a health service cannot be observed directly ... (or) ... even be determined reliably after the experience, since a patient cannot identify perfectly the impact of the health service on any improvement in her health." However, in order to test Hypotheses 1 through 3 outlined is Section 3.1, I need to measure precisely this quality. Many econometricians have already addressed the problem of measuring hospital quality in the past, in this section I review these approaches briefly and explain the rational for my particular choice of mortality rates for Heart Attacks as a measure of hospital quality.

The existing measures of quality can be summarized as falling into two main groups: process measures and outcome measures (see Feldman and Scharfstein, 2000 and Schlesinger and Gray, 2006 for a survey). Process measures focus on the treatment technology utilized. Ideally, this technology is ranked by physicians specializing in the area before the study. Outcome measures are the richest in terms of categories and include legal measures, consumer satisfaction measures and objective measures. Legal measures include regulatory violations and malpractice law suits. Measures of customer satisfaction are usually obtained by asking hospitalized or recently discharged patients to fill out questionnaires. These measures are more common for nursing homes rather than acute-care hospitals. Objective measures focus on observable dimensions of patient outcomes. The most common objective measure is mortality rates within a fixed number of days after hospitalization

or discharge. However, adverse outcomes, re-admissions and in-hospital mortality rates are used as well. For nursing homes specifically, the measure of functional improvement was used in a number of studies (Schlesinger and Gray, 2006).

I continue by discussing the disadvantages of each family of quality measures just outlined. There is an obvious preference toward objective outcome measures of quality both within the medical and the economics professions (Feldman and Scharfstein, 2000; Schlesinger and Gray, 2006). Outcome measures are viewed as proxies for the latent variable: "goal of treatment". Therefore, the main question I address while assessing the various measures of quality is how well process and subjective measures proxy for the (latent) outcome measure.[82]

First I address process measures of quality of care. Skinner and Wennberg (2000) find that treatment techniques for the same medical condition vary substantially between U.S. regions. Specifically, regions where the spending is higher, survival rates of patients are not necessarily higher as well, and frequently even lower, after controlling for the underlying population health. Variation in technology of treatment seems to affect costs in the predictable direction, i.e. more intensive input utilization implies higher costs. However, more intensive inputs rarely lead to better outcomes. This finding suggests that process measures of quality, which actually measure the inputs into healthcare production, have little to say about the quality of outcomes.[83] Thus, deductions about quality of outputs based on inputs have low validity for hospitals.

Legal measures of hospital quality suffer from a number of shortcomings. Both regulatory violations and malpractice law suits are quite rare events for an institution, therefore, any such

[82]Skinner and Wennberg (2000) argue that patients' preferences usually involve a trade-off between how intrusive the treatment is (i.e. the quality of their life during the treatment period) and the quality of life after the treatment (i.e. the improvement, or the outcome). It is hard to for researchers to measure patient-preferences. On the other hand, in order to form this preference ordering, patients need to be able to predict the impacts of procedures on their well-being during the treatment period and the impacts on their health (and well-being) after the treatment. Both of these are hard to for patients to predict, and, hence, to form a meaningful preference ordering.

[83]Kaestnera and Guardado (2008) show that exogenous changes in hospital funding have little systematic effect on both staffing and outcomes.

measure will most likely have low signal to noise ratio, if measured for each institution across time. It is unlikely that every hospital (nursing home) in the market engages in regulatory violations or treatments that can attract malpractice suits. This fact implies that a hospital (nursing home) that engages in such is an "outlier" and is affected by factors that may lie outside of conventional economic theories that do not study such behaviour explicitly. Ironically, if indeed every firm in the market were to engage in some form of legal violations, this would suggest that either society's standards or enforcement strategies may need amendment. When the vast majority of players engages in regulatory or legal violation, the fact that any particular player engages in these violations, or attracts malpractice suits, stops being a relevant signal of quality altogether. Therefore, malpractice suits and regulatory violation are poor measures of quality of care.

Consumer satisfaction measures are subject to various problems common to the situations when economists try to measure economic variables through complex surveys. First, obtaining large datasets is costly, which brings to the previous critique of low signal to noise ratio. Second, it is not clear how the surveyed interpreted the questions and whether there were misunderstandings common to large subgroups of subjects. However, the main criticism is from the theoretical point of view: if the credence nature of healthcare is true,[84] a consumer does not have significantly more information about the quality of a healthcare service even after the consumption. Therefore, survey results must be interpreted with much caution. There may be systematic biases due to perceptions, framing of expectations, and not the actual underlying treatment. Thus, consumer satisfaction surveys are a poor measure of quality as well.

The objective outcome measures of healthcare quality seem to be the most satisfying both from the economic and the professional medical perspective. Outcomes can be measured more accurately and on a wider scale than virtually any of the other measures proposed. Good outcome measures are based on existing medical records. These measures aggregate data that was carefully recorded

[84]See Darby and Karni (1973) for a definition and discussion of "credence" goods.

by professionals. The careful records imply precision and the aggregation creates information that is unavailable to a single patient, hence, the superiority over patient satisfaction measures. It is not surprising that the majority of studies interested in quality of healthcare employ outcome measures. However, even these measures are not without problems.

It is known that higher patient volume is associated with higher quality of care. Thus, Feldman and Scharfstein (2000), for example, use volume[85] as a measure of quality, arguably the simplest to obtain objective outcome measure. Some argue for two potential reasons for the positive correlation between quality and volume: learning by doing (i.e. volume affects quality), or patient selection of hospitals based on quality of care (quality affects volume) (Feldman and Scharfstein, 2000; Gowrisankaran, Ho, and Town, 2006). I examine each of the two explanations both jointly and separately and drive implications about the consistency of volume as a measure of quality.

Gowrisankaran, Ho, and Town (2006) study the learning by doing explanation for the correlation between volume and hospital quality of care. Empirical findings suggest that learning by doing causality between volume and quality is strongly dependent on the specific medical procedure examined and it is hard to generalize across procedures.

If both of the statements are true: volume affects quality and quality affects volume, we should observe a concentration of patients within a small group of hospitals after a certain period of time. On the other hand, we would observe almost empty hospitals in the same markets, especially, if physicians are "forgetting" by "not doing", as the results of Gowrisankaran, Ho, and Town (2006) suggest.[86] This extreme bi-polarization of hospitals by patient numbers does not happen in reality

[85]The authors simply count the number of surgeries each physician performed in a given year (1995 in the sample used). Then they conclude that the healthcare plan that refers patients to physicians who performed higher number of surgeries in a given year is of better quality. This conclusion has similar bias to process measures, as it is likely that plans that refer patients to higher volume physicians simply practice higher utilization. Unfortunately, this bias may be particularly important for Feldman and Scharfstein (2000) as the authors compare Fee-For-Service systems versus Managed Care (see Section 1.2 of Chapter 1 for a detailed discussion of these systems of delivery), where the former are known to have an incentive for increased utilization.

[86]In economic terms this implies an extreme case of "economies of scale".

(see Figure 3.2 in Appendix B). This feature strongly suggests that causality behind the positive association between volume and quality is likely to be procedure dependent. If for some procedures volume of treatment is correlated with quality, yet no learning by doing takes place, there must be hospital selection.

If patients select healthcare providers based on quality unobservable to the Econometrician, measuring hospital quality through volume may lead to significant biases in estimations. These biases are likely to be severe in demand-side estimations, for example. If the association between volume and quality is solely determined through hospital selection, this relationship will break-down in settings where pricing is more flexible, as lower prices can substitute for higher quality. For this reason, if other objective outcome measures of quality are available, in setting where there is a possibility of hospital selection, the use of volume as a measure of quality should be discouraged.[87]

The objective outcome measure most commonly used is mortality rates. McClellan and Staiger (2000) examine this measure in detail. First they focus on its most important disadvantage - it is very rare, and, hence, noisy. In order to overcome this problem, they introduce a filter that puts more weight on hospitals with more patients. Another objection one may raise to the mortality measure is endogeneity: sicker patients maybe referred to higher quality hospitals. In this case the quality measure for high quality hospitals will be biased downwards and lower quality ones upwards. In order to overcome this bias McClellan and Staiger (2000) suggest focusing on severe enough condition where timely medical intervention is crucial for survival and patients are more likely to be transported to the nearest hospital. One such condition, is Heart Attacks, often referred

[87] I can suggest yet a third explanation for the correlation between volume and quality. If hospitals are financially constrained, it is plausible that those hospitals with higher volume of patients are able to obtain higher revenue and invest it in better equipment, which improves quality of care. NP hospitals are more likely to be financially constrained (Cutler and Horwitz, 2000; Gentry and Penrod, 2000). Therefore, in order to test this explanation one needs to show that exogenous variation in patient volume affects quality more in the NP sector than in the FP sector. In order to perform this estimation, however, one needs to be able to measure both the technological inputs (equipment) used and the resulting quality of care (the impact of better equipment on quality of care needs to be shown empirically as well!). The measure of quality has to be an objective outcome measure other than volume.

to as Acute Myocardial Infarction (AMI)in the medical literature. I use mortality rates for Heart Attacks as the measure of quality in this study.

I employ 30-day post-admission Risk-Standardized (adjusted for patient characteristics and health status)[88] Mortality Rates (RSMR) for Heart Attacks,[89] as the main measure of quality in this study. However, for some estimations I supplement this measure by mortality rates for Heart Failures. This is done in order to increase the statistical power, due to limited number of observations. When a patient experiences a Heart Attack it is, usually, crucial to transport her to the hospital within minimum amount of time. Therefore, it is unlikely that there is enough time to select a better hospital in case of a Heart Attack. However, when an individual suffers from Heart Failure, the time lag to the beginning of treatment may have lower impact on health,[90] hence, hospital selection is more likely. There is little cost to using mortality rates for Heart Failures as a measure of quality, as it can only bias *against* finding significant differences between hospitals.

In this section I discuss various measures of quality for hospitals. I use risk-adjusted post admission mortality rates from Heart Attacks as the main measure of quality in this study. The goal of this section was to show that, although this measure has its caveats, it is superior to other measures of quality used in the literature. My study aims to exploit variation across markets to estimate the relevant parameters, as the empirical implications are all stated at the market level. Therefore, the other central methodological issue that I need to address is market definition. In the next section I address this issue while reviewing the existing empirical literature.

[88] Differences in case-mix between hospitals need to be controlled for for proper measurement of quality. Risk adjustment aims to control for for such differences (McClellan and Staiger, 2000).

[89] The exact procedure used to compute the RSMR in the data set I use in this investigation is discussed in detail in Section $B.1$ of Appendix B.

[90] The "Heart Failure" diagnosis if fairly broad, therefore, both high and low emergency cases may fall in it.

3.3 Analysis Level and Market Definition in Hospital Studies

A considerable amount of empirical research has been done on hospital markets. Rosenau and Linder (2003) review 69 studies that deal specifically with quality and organizational form. Schlesinger and Gray (2006) discuss 38 such studies of which 21 overlap with the previous review.

Any study that employs objective patient-outcome measures of hospital quality (e.g. mortality or readmissions) has to address analysis-level choice. There are three possibilities for analysis levels: patient-level - each admission (release) is an observation, hospital-level - each hospital is one observation, or market-level - each (geographical) market is one observation. My work is the first to introduce the market-level analysis, as it seems to provide more direct tests for theories that address market-level moments of the distribution of quality by organizational form (like the one developed in Chapter 2). The market-level analysis constitutes a methodological innovation, to the best of my knowledge.

In any study on hospital markets (even if it is not related to quality) it is important to define geographical markets for hospitals examined. It is particularly important to have satisfying market boundaries if one is to use market-level analysis, as I do in this chapter. I use Metropolitan Statistical Areas (MSA) as geographical markets in this study.In this section I address various solutions to the analysis level and market definition in the existing literature.

In order to give an example of how others solve the key methodological issues of analysis level and market definition, I focus on two recent papers within the field of Economics that seem to be representative of empirical research on hospital markets. Shen (2002) belongs to the McClellan-Staiger school (McClellan and Staiger, 1999; McClellan and Staiger, 2000) and studies the effect of organizational form on quality. Shen uses the McClellan-Staiger database, where mortality rates (a measure of quality) were adjusted for patient characteristics and filtered as discussed in McClellan and Staiger (1999). The other example is an extensive yet unpublished work by Horwitz and Nichols

(2007) that focuses on competition between NP and FP hospitals.

Shen (2002) summarizes the existing empirical literature on hospital quality as falling under one of the two levels of analysis: patient-level and hospital-level. Patient-level analysis constitutes of estimating a regression, where each patient constitutes one observation. A binary mortality variable (equals to 1 if patient died t days after admission (release), and zero otherwise) is regressed on a set of controls. Most common controls are market characteristics, hospital characteristics and patient demographics, if available. The main criticism of patient-level analysis is that the variation needed for identification of the coefficients of interest, especially the organizational form (NP, FP or government ownership), comes from hospital-level variation, hence, this procedure generates artificially low standard errors. For this reason the author advocates for a hospital-level estimation, where each hospital is one observation.

Shen (2002) uses data where patient-level admission information was aggregated to hospital-level mortality rates. This enables the author to perform the analysis without biasing against finding statistically insignificant estimates of hospital characteristics. In order to measure the effect of organizational form on quality the researcher regresses the computed mortality rates on NP and government ownership dummies (FP status omitted), hospital characteristics: size (by admissions), teaching status, market characteristics: Herfindahl-Hirschman Index (HHI) (by inpatient discharges), existence of FP in the market (not including the given hospital), urban dummy, county-level: per capita income and population size. A market was defined as all hospitals falling within the 15 mile radius around the zip code of a given hospital.[91]

[91]The study of Shen (2002) is representative of other studies on hospital quality (Schlesinger and Gray, 2006) (with the exception that earlier studies used patient-level data together with hospital or market level variables). Due to the mixed results on hospital performance by organizational forms across these studies (Schlesinger and Gray, 2006), there arose the need to examine the impact of organizational form on quality of care more carefully from the theoretical perspective. The model developed in this thesis addresses offers the solution to the puzzle of the mixed results by suggesting that there are two types of hospitals that fall under FP status: low quality FPs that treat less informed patients, and high quality FPs that treat better informed consumers (See the discussion in Subsection 3.1.1). The model of Chapter 2 suggests that the results of an estimation where a measure of hospital quality is regressed on a dummy variable for organizational form and controls (e.g. Shen, 2002) will depend on the proportion

As some previous theories of NPs (e.g. Hirth, 1999), the theory developed in this thesis has empirical implications at the market level. Hypotheses 1 through 3 (Section 3.1) involve market-level moments of the distribution of quality (for organizational forms) and the share of informed consumers in the market. Therefore, exploiting the variation across markets seems a natural starting point for the empirical tests. The argument of Shen (2002) can be applied again: the main source of variation in the independent variable ("share of informed consumers") comes from cross-market and not cross-hospital or cross-patient variation. For this reason I choose market-level analysis, where each market (MSA) is one observation.

Horwitz and Nichols (2007) do not address quality of care. However, this study is relevant to the analysis of this chapter, because the authors address the issue of market boundaries. In order to study most competition issues require binding markets in the dataset. Horwitz and Nichols (2007) study the impact of competition with FP hospitals on the mix of services and operating margins that (NP, FP and government owned) hospitals offer. One of the innovations employed by Horwitz and Nichols (2007) is more flexible definition of the market. As the researchers were interested only in computing a share of each "market" that is served by a hospital they used a weighting approach.

Horwitz and Nichols (2007) include hospitals in the market of a given hospital by assigning a weight to every hospital around it. This weight is proportional to the admissions of a candidate hospital and inversely proportional to the distance from the hospital for which the market was being computed.[92] Although this method may appear appealing intuitively, there are non-trivial costs associated with such calculations. The benefits of the weighted approach for my purposes may be limited. My goal is constructing a dataset where each market is an observation, while the weighted-distance definition eliminates the discrete nature of observations. Moreover, some demographic data is meaningful only at the Metropolitan Statistical Area (MSA), or county level,

of high-quality FP hospitals in the sample analysed. If the proportion of high quality FPs is small, one is likely to find NP dominance over FPs in terms of quality. If this proportion is large enough, the result will reverse.

[92] The weights decline with distance fairly rapidly, see Horwitz and Nichols (2007) for details.

and demographic regressors will play a crucial role in the tests I perform.[93] I use MSA as the working definition of a market.

MSA specification of the market was used as a robustness check by Horwitz and Nichols (2007) and was shown to yield similar results to the more sophisticated weighted-distance specification. Horwitz and Nichols (2007) list several advantages of MSA specification: MSAs were designed by the U.S. Census Bureau in such a way as to reflect barriers to travel such as mountains and rivers. Neither the fixed-radius approach (Shen, 2002) nor the weighted-distance one account for these barriers, as they are constructed arbitrarily from the point of view of the topography of the earth. The barriers to travel are a major determinant of hospital admissions, especially for Heart Attacks where timely intervention is crucial. This makes it less likely for a patient to be transported to a hospital that, although, falls within a certain radius, is located across a river or a mountain. Therefore, the MSA specification of markets is particularly advantageous for my purposes.

There are three candidates, suggested by the literature, for a computationally simple and discrete definition of hospital markets: 15 mile radius around a given hospital, counties and MSAs. Both counties and MSAs perform better than 15 mile radii in terms of earth topography that takes into account transportation issues. 15 mile radii and counties, however, seem too narrow as market definitions for hospitals when one considers market structure (i.e. competition).[94] Large U.S. metropolitan areas (e.g. Los Angeles) encompass a number of counties and 15 mile radii. Patients are frequently transported across county and arbitrary 15 mile radii borders within the same city for treatment,[95] which implies that counties and 15 mile radii may not capture the full extent of

[93]For example, although proportion of population with graduate or professional degree (see Section 3.4) may be available at the zip code level from the U.S. Census Bureau, it is likely that the noise in this variable is too high for any meaningful estimation.

[94]The model I develop in Chapter 2 is about competition between healthcare providers. Horwitz and Nichols (2007) examine competition issues empirically and use MSAs and not counties, as a robustness check to their market definition.

[95]Athey and Stern (2000) report an average of 0.0072 hospitals per square mile (U.S. based study). In 1995 the average time of an emergency ambulance ride to a hospital was 13.235 minutes (Athey and Stern, 2000).

hospital markets in urban areas. In rural areas it is often the case that only one hospital is available for treatment of severe conditions as Heart Attacks in a large geographical area. For this reason meaningful inferences in terms of quality distribution are harder to obtain in rural markets. Therefore, restricting the analysis to urban areas and treating each MSA as one market is the approach I take in this study.[96]

The main methodological issues that one needs to address in order to test the predictions of the model developed in Chapter 2 (especially Hypothesis 1 through 3 in Section 3.1) are: measurement of hospital quality, market definition and selection of analysis level. I discuss measures of hospital quality in Section 3.2 and choose 30 day post-admission RSMR for Heart Attacks as the main measure of quality in this study. I define markets as MSAs due to the geographical-transportational advantages suggested by Horwitz and Nichols (2007). The main variation in the variables of interest for my analysis (market-level moments of quality distribution and the share of informed patients) is at the market level, hence, I treat each market as one observation (adapted version of the argument by Shen, 2002). The next section develops the empirical strategy.

3.4 Empirical Strategy and Data

In order to test Hypotheses 1 through 3 in Section 3.1 I construct a dataset of market-level moments of the distribution of hospital quality for each organizational form, and control variables. This section begins with an exposition of the specifications I estimate. After presenting the regression specifications I move to a discussion of the importance of the controls used, and potential biases due to omitted variables. I describe the data available and the steps for the construction of the final dataset that is used to estimate the specifications in Subsection 3.4.1; the summary statistics appear in this subsection. A weighting procedure is developed in order to correct potential biases

[96] I have replicated the analysis discussed in this chapter at the county level. The results were mostly qualitatively similar to the ones reported at the MSA level, however, the estimates tend to be less precise (larger standard errors).

in the original dataset, it is introduced in Subsection 3.4.2.

The main implication of my model developed in Chapter 2 (summarized in Hypothesis 1) is *FP hospitals have higher variance (standard deviation) of quality than NP hospitals.* In order to test this hypothesis, I generate a sample where each market (MSA) appears up to three times: once with the corresponding moment for NPs, once for FPs and once for government owned hospitals.[97] I treat each MSA as one market, the rational for this was discussed in the previous section.

As explained in Section 3.2, quality is measured through mortality rates pre-adjusted for patient characteristics, RSMR for Acute Myocardial Infractions (AMI) - Heart Attacks. These data were obtained from the U.S. Department of Health & Human Services, Hospital Compare.[98,99] I estimate the following specification to test Hypothesis 1

$$SDMort_{cmo} = \beta_0 + \beta_1 FP_{cmo} + \beta_2 GOV_{cmo} + \beta_3 X_m + \beta_4 HF_{cmo} + \epsilon_{cmo} \tag{3.4.1}$$

Where $SDMort_{cmo}$ is the standard deviation of RSMR for condition c in MSA m for hospitals of organizational form o. There are three types of organizational (or ownership) forms in the dataset: NP, FP and Government. Dummy variables capture the effect of the organizational form. FP equals one if it comes from a FP hospital, and GOV equals one if the mortality rate comes from a government owned hospital (NP is the omitted category). X_m is a vector of market characteristics, intended to control for forces that are not incorporated into my theory, that would render markets comparable. HF is a dummy variable equal to one if the observation belongs to HF condition. ϵ is an error term coming from various idiosyncratic market-level shocks.[100] A positive and significant

[97] In some markets there may be no hospitals of a particular organizational form. Those markets will appear in the sample less than three times.

[98] `http://www.hospitalcompare.hhs.gov/` accessed on 3 December 2008 13:07.

[99] I use mortality rates for Heart Failures (HF) (in addition to AMI) to increase statistical power of the results in estimating specification (3.4.1). RSMR for HF have not been used before in economic literature, to the best of my knowledge. As I explain in Section 3.2, the inclusion of HF has the potential to bias only against finding support for Hypothesis 1.

[100] The error terms may be potentially correlated at the MSA-level, as each MSA appears as an observation in the

coefficient on the FP dummy ($\beta_1 > 0$) would be in line with the central prediction of the new theory of NPs.

One may object to specification (3.4.1) on the grounds that the small number of hospitals within MSAs[101] does not allow to estimate the standard deviation of quality precisely at the MSA level. The imprecise estimation of the dependent variable has the potential to introduce a considerable amount of noise in the estimation of specification (3.4.1). The right way to view the market-level moments of quality distribution, however, is as *population* moments rather than sample moments. Sample moments indeed estimate population moments imprecisely if the sample size is small. The dataset I use, however, contains almost all the relevant hospitals (acute care hospitals of significant capacity that treat patients who suffer from Heart Attack and Heart Failure) that fall within MSAs. Therefore, within markets I capture (almost) the entire population. On the other hand, since my study focuses on MSAs, and excludes rural areas, I use only a sample of all hospital markets in the U.S. . In the estimations of specification (3.4.1) the observations are constructed from population moments and constitute a sample of all U.S. markets, therefore, the precision of these estimations is not likely suffer from excessively noisy variables.

In order to test Hypothesis 2: *An increase in the proportion of informed patients increases the mean FP quality for small proportions of informed, and decreases the mean FP quality for large proportions*, I specify the following regression equation

$$MeanMortFP_m = \gamma_0 + \gamma_1 Grad_m + \gamma_2 Grad_m^2 + \gamma_3 X_m + \eta_m \tag{3.4.2}$$

Where $MeanMortFP_m$ is the mean RSMR (for Heart Attacks) for FP hospitals in MSA m. $Grad$ is the percentage of population in the MSA who have Graduate or Professional degree. This

dataset up to three times (once for each organizational form). To check for robustness I cluster the standard errors for the estimation that uses only Heart Attack RSMR and shows results supportive of Hypothesis 1. Clustering does not seem to matter (see Table 3.4).

[101] See Chapter $B.2$ in Appendix B for a discussion of the distribution of the number of hospitals by MSA.

variable proxies for the proportion of informed consumers in the population. I believe that higher level of education makes an individual more likely to belong to a social network that includes medical professionals. Those medical professionals are able to recommend the best hospitals and doctors. Additionally, higher level of education may improve individual's ability to make inferences about quality of hospitals from the information available. Following the non-monotonicity in the hypothesis I include $Grad^2$ in the regression as well. X_m is a vector of market-level controls (same as in (3.4.1)) and η_m is an error term coming from various idiosyncratic shocks that X_m does not control for. A negative coefficient on $Grad$ and positive one on its square is consistent with Hypothesis 2, as high mortality rates imply low quality.

Finally, to test Hypothesis 3: *An increase in the proportion of informed patients reduces the variance (standard deviation) of NP quality.*[102] I estimate the following specification.

$$SDMortNP_m = \delta_0 + \delta_1 Grad_m + \delta_2 X_m + \pi_m \tag{3.4.3}$$

Where $SDMortNP_m$ is the standard deviation of RSMR (for Heart Attack) for NP hospitals in MSA m. $Grad$ is the percentage of population in the MSA who have Graduate or Professional degree, again a proxy for the proportion of informed consumers. X_m are MSA-level controls and π_m is an error term coming from various idiosyncratic shocks. A negative coefficient on $Grad$ is consistent with Hypothesis 3. I turn to a discussion of the variables that enter in X_m, market-level controls.

It is well known that both income and education affect health outcomes, Jones and Wildman (2008) reports such evidence on income and DeWalt et al. (2004) offers a review of evidence on education. Therefore, it is important to control for income and education differences between MSAs. Higher income enables individuals to maintain healthier life style. On the other hand,

[102] I discuss the empirical strategy for the test of Hypotheses 4 and 5 in Section 3.6.

richer communities can afford to invest more in healthcare facilities. This is important in the U.S. particularly as much of the community investment comes from local funds.

A general measure of the level of education in the population is associated with better health outcomes, possibly due to better information about a healthy life style. I control for the general level of education with the variable "percentage of the MSA population with 9th to 12th grade (no high school diploma)". In most of the specifications this measure of education enters the regression more significantly than other measures. However, in some of the estimations that test the main implication of the model, "percentage of the population with high school degree or equivalent" is the most statistically significant measure of education. As the two measures seem fairly equivalent (the correlation between them is $\rho = 0.227$) I report the results with the one that enters the regression most significantly.

The variable *Grad* is defined as the "percentage of MSA population with graduate or professional degree", this variable proxies for the percentage of informed patients in the MSA. Including a measure of overall level education enables to estimate the coefficient on the variable *Grad* more consistently, as it is correlated with the measures of general level of education ($\rho = -0.627$ for "percentage of the MSA population with 9th to 12th grade (no high school diploma)", and $\rho - 0.472$ for "percentage of the population with high school degree or equivalent"). Therefore, it is important to control for the general level of education.

If there are economies of scale in healthcare (especially if they are exogenous to the firm), higher population in a market would imply higher levels of investments that are correlated with lower mortality (e.g. enhanced 911 service). This force should be even stronger if the population has higher income as well. The number of doctors is presumably highly correlated with population (provided that doctor-patient ratios do not vary significantly in urban U.S.). Hence, we should systematically observe that markets with higher population have better quality of care, regardless of the education level. This potentially biases upward the coefficient on *Grad* specification (3.4.2),

and in unpredictable direction the coefficient on *Grad* in specification (3.4.3). To prevent this bias I control for MSA population and medium household income.

Transportation takes longer in more congested cities relative to less congested ones. The speed with which patients can be delivered to the hospital has a significant effect on their survival after a Heart Attack. It is unlikely, however, that transportation systematically reduces the quality of one organizational form relative to the other. (3.4.1) attempts to discriminate between the different organizational forms within the same market. Hence, overall higher mortality rates in more congested markets should not bias specification (3.4.1). If the level of traffic congestion is correlated with the proportion of the population with graduate or professional degrees, estimation of specifications (3.4.2) and (3.4.3) maybe be biased. Including MSA population in the regressions is likely to reduce this bias. Due to lack of data I am unable to control for road congestion directly (see the discussion below on travel time to work).

It is common to use the following controls in investigations of effects of organizational forms on hospital characteristics (not limited to quality): hospital size and teaching status (whether the hospital is a teaching institution or not), membership in a system of hospitals; market-level controls: HHI (by patient discharges), existence of FP, urban dummy, Health Maintenance Organization (HMO) penetration; demographics (of potential patients): per capita income, population size, age, education, race, sex, marital status, employment by industry and travel time to work (Shen, 2002; Horwitz and Nichols, 2007).

My aim is to test the relationship of market-level *moments* of quality, which cancel out the effects of individual hospitals. For this reason I do not control for hospital size and teaching status directly.[103] Nevertheless, due to the small number of hospitals within MSAs the individual characteristics of a specific large hospital may have a disproportionate effect on the population moments. Teaching hospitals tend to be large and have higher quality of care (McClellan and

[103]Subsection 3.4.2 introduces indirect controls for hospital size.

Staiger, 2000).

It is likely that the presence of a teaching hospital implies increased variance of NP quality. Most teaching hospitals are NP (McClellan and Staiger, 2000), however, the majority NP hospitals in one MSA will not have teaching status. Since teaching hospitals have higher quality of care than other NPs, their presence potentially increases the variance of NP quality. This feature is likely to bias specification (3.4.1) against finding support for Hypothesis 1. MSAs that have teaching hospitals are likely to have higher percentage of population with graduate or professional degree (due to the personnel and the trainees of those hospitals), therefore, estimation (3.4.3) is potentially biased against finding support for Hypothesis 3. Specification (3.4.2) is not likely to be biased, since it is rare for FPs to engage in teaching (McClellan and Staiger, 2000). There is little harm in failing to control for teaching status, as it biases against finding the desired results.

HHI and HMO penetration control for competitiveness of the market. Both competitiveness and membership in a system of hospitals are endogenous choices for the players. In the model developed in Chapter 2 I address the market structure directly. The best test for consistency of the data with the predictions of the theory, in my view, is not controlling for competitiveness and membership in a system of hospitals. In fact, the lack of such controls biases against finding support for the theory, as it introduces additional variance not addressed in the model. Out of the three (HHI, HMO penetration and system membership), only HMO penetration is potentially correlated with the degree of informativeness of the consumers. However, the results should be interpreted as: better educated patients should be able to make more informed choices both when selecting an insurer (HMO) and when selecting a provider.

My specifications control for income, population size and education. I do not control for age, race, sex, and marital status demographics. The mortality rates in my data were calculated after controlling for the demographic characteristics of actual patients admitted to the hospitals (see Section $B.1$ in Appendix $B.1$). Therefore, a potential causality for a bias is the following: in

a market where Heart Attacks are more (less) prevalent due to the demographic composition of the population, there was a decrease (increase) in actual Heart Attacks due to some unobservable factor. In this case all hospitals may improve (deteriorate) their quality of treatment, say due to lower capacity utilization. There is little reason to expect differential impact by organizational form, however. Not withstanding this scenario, I do not control for the factor that causes the shift in disease prevalence. This feature already potentially biases the estimations of specifications (3.4.2) and (3.4.3). Sex race and marital status of non-patients are not commonly used in studies on hospital quality.

Employment by industry proxies for the insurance coverage of the population (Horwitz and Nichols, 2007). It is possible that mortality rates are lower in MSAs with higher percentage of the population with graduate or professional degree, if it is associated with higher percentage of persons working in industries that offer better insurance coverage. Strictly speaking, in this case specification (3.4.2) may be biased toward supporting Hypothesis 2. However, one may interpret the results more broadly: more educated people make more informed choices of both treatment and insurance coverage, in this sense this is not a bias. If higher percentage of MSA population with graduate or professional degree is associated with larger variety of industries in the MSA, which in turn offer varying levels of insurance, specification (3.4.3) will be biased against finding the desired result (Hypothesis 3). Employment by industry is not commonly used in studies on hospital quality.

Travel time to work proxies for willingness to drive in the population, that may influence hospital selection, according to Horwitz and Nichols (2007). In my estimations, due to the importance of immediate treatment of Heart Attacks, there is lower probability of hospital selection. Travel time to work, however, may be used as a proxy for the average length of an ambulance ride in the MSA, as it must be correlated with road congestion. Travel time to hospital after a Heart Attack has a strong impact on patient's survival. Failure to control for travel time may bias specification (3.4.2),

if the percentage of highly educated population is correlated with more congested roads. However, in this case the bias is toward rejecting Hypothesis 2. Travel time to work is not commonly used in investigations addressing hospital quality.

3.4.1 Data Description and Dataset Construction

After introducing and discussing the specifications, I turn to a description of the dataset I use to estimate them. I obtained the information on RSMR, ownership status, and addresses by Medicaire/Medicaid Provider ID from the U.S. Department of Health & Human Services via the Hospital Compare website.[104] The RSMR were calculated from admissions of Medicare and Medicaid insured patients that occurred between July 2006 and June 2007. For the details of RSMR estimations see Section $B.1$ in Appendix B.

From the U.S. Census Bureau[105] I obtained a dataset on Median Household Income in 1999 by MSA, Total Population by MSA, and the number of people with the following highest levels of education by MSA, sex and race: graduate or professional degree, bachelors degree, associate degree, some college, high school degree, grades 9th to 12th (no high school diploma), less than 9th grade. These data come from the 2000 census (the median household income is from 1999) and may be slightly dated, however, should be highly correlated with more current data. The lagged nature of these variables renders them more likely to be exogenous. I aggregate the education variables across sexes and races to the MSA level and divide them by the total MSA population to obtain percentages.

I construct the datasets used in the estimations in the following manner. First I map each Medicare/Medicaid provider ID into its characteristics: ownership status (organizational form) and zip code of its location. Then I map the zip codes into the FIPS codes of the MSAs. Finally, I

[104] `http://www.hospitalcompare.hhs.gov/` accessed on 3 December 2008 13:07.
[105] `www.census.gov` American Fact Finder [download centre] accessed on 4 December 2008 14:13.

merge the file with mortality rates and characteristics of each provider with the demographics of the MSA in which it is located. After this stage I perform various operations to construct two separate datasets with moments of mortality rates at the MSA-level, where each MSA is one observation (with the exception of the first dataset described below, where each MSA appears up to three times).

I use one dataset to estimate specification (3.4.1) and a separate dataset to estimate specifications (3.4.2) (3.4.3) due to the differences in the levels of observations between the specifications. Specification (3.4.1) must be estimated on a dataset where each MSA appears up to three times as an observation (for each condition): once with standard deviation (mean) of quality for NPs, once for FPs and once for government owned hospitals. The two dummy variables FP and GOV code the observations by organizational form. This procedure must be repeated for each of the conditions: Heart Attacks and Heart Failures. A dummy variable HF codes the condition to which the observation belongs.

Specifications (3.4.2) and (3.4.3) must be estimated on a dataset where each MSA appears as an observation only once. For each MSA one must construct one variable with the mean of mortality for NPs, another with the mean of mortality for FPs, and additional two variables with the standard deviation of mortality (one for NPs and one for FPs). I address the classification of hospitals to ownership (organizational form) status next.

The classification of hospitals to organizational forms was done in the following manner. Hospitals whose ownership was: "Government - Federal", "Government - Hospital District or Authority", "Government - Local" and "Government - State" are all classified as government owned. Hospitals with ownership type "Proprietary" I classify as FPs. Finally, all of the following ownership forms are classified as NP: "Voluntary non-profit - Church", "Voluntary non-profit - Other" and "Voluntary non-profit - Private".[106]

[106]Studies that do not investigate the differences between religious NPs and non-religious ones (for an example of

Table 3.1 shows the descriptive statistics for mortality rates. One mortality rate for a procedure correspond to a hospital. The number of NP hospitals is more than twice the number of FP and government owned ones. Both the means and the standard deviations of mortality rates are fairly

Table 3.1: Risk Adj. Mortality Rates by Condition and Org. Form (estimated from 2006-2007 Medicare/Medicaid admissions)

Condition	Org. Form	Obs.	Mean	Std. Dev.	Min	Max
AMI	FP	322	16.072	1.097	12.5	19.5
	NP	1112	15.968	1.208	12.4	20
	GOV	256	16.11	0.996	12.7	20.8
HF	FP	332	10.954	1.216	7.8	16.6
	NP	1129	10.875	1.39	6.7	16.8
	GOV	270	11.21	1.171	8.4	16.3

close between the NP and FP organizational forms. The NPs seem to have slightly lower average mortality, while FPs have slightly higher standard deviation of mortality. One may interpret these results as a weak support of previous theories, that in the spirit of Arrow-Hansmann hypothesis (Arrow, 1963; Hansmann, 1980; Hirth, 1999; Glaeser and Shleifer, 2001) predict the NPs to have higher quality of output than FPs. At the same time there seem to be a weak inconsistency with Hypothesis 1. One has to note, however, that the moments of the mortality rates in Table 3.1 were calculated from the entire U.S. sample, while the predictions of all the theories (including mine) are at the market level.

The rest of the summary statistics for the variables I use in the analyses below appear in Table 3.2. There is fairly high variation in the population of MSAs (standard deviation of over 2 million inhabitants). This shows the importance of the inclusion of the population control. The income dispersion between the highest and the lowest median income MSA is more than $ 50,000 a year.

a comparison between religious and non-religious NPs see Hansmann, Kessler, and McClellan, 2003) tend to classify all of these institutions under one NP category.

Table 3.2: Summary Statistics: Hospital Quality Estimations

Variable	Mean	Std. Dev.	Min	Max
MSA Population (mill.)	0.899	2.06	0.058	21.2
Med. HH Inc. MSA 1999 (000s $)	39.175	6.755	11.385	62.024
MSA Pop. with Grad (%)	0.081	0.029	0.029	0.21
MSA Pop. with 9th to 12th gr. (%)	0.114	0.028	0.034	0.186
MSA Pop. High School Degr. (%)	0.301	0.062	0.178	0.5
No. of Obs.	N=746			

This shows the importance of the income control. The percentage of population who have a high school degree or an equivalent as their highest level of education is the most variable among the three education measures used. In Subsection 3.4.2 I turn to a procedure that is intended to increase the precision of the calculation of the market-level moments of quality.

3.4.2 Weighting

In some of the estimations below I construct the dependent variables using a weighted quality (RSMR) for each hospital. The weight applied to a hospital's mortality is its share of patients within MSA, condition (Heart Attack/Heart Failure) and organizational form (NP/FP). I use the moments computed from the weighted mortalities as the dependent variables in specifications (3.4.1) through (3.4.3).

It is well documented that larger hospitals tend to have higher quality of care (see Feldman and Scharfstein, 2000 and Gowrisankaran, Ho, and Town, 2006 for a survey). This is related to economies of scale and potential learning by doing. The model developed in Chapter 2 does not address these factors. The U.S. Department of Health & Human Services, Hospital Compare employs a methodology (hierarchical regression model) that moves the mortality estimates for smaller hospitals (those with lower number of cases) closer to the U.S. mean mortality for the

condition (see Section $B.1$ in Appendix B for the details of RSMR calculations). Placing lower weight on the biased estimations reduces their impact on the market-level moments of quality.

The weighting of mortality rates was done within organizational form. One may object to assigning potentially different weights to two hospitals operating in the same MSA which admitted the same number of patients with a certain condition, but happened to have different organizational forms. Figure 3.2 shows that FP hospitals tend have systematically less patients than NP ones. Therefore, there is a danger of assigning systematically lower weights to FP hospitals if the weighting is done simply by hospital's share of patients within a market (and condition). If the RSMR of FP hospitals receive systematically lower weights, the distribution of weighted mortality rates for FPs will be biased downwards. Since this would imply biasing FPs toward higher quality, and potentially reducing the variance of quality, I decided to assign weights by the patients treated within organizational form.

The distribution of hospitals by number of patients within each organizational form is reported in Figure 3.2. FP hospitals tend to be smaller than NP ones, if size is measured by the number of patients. The t-test for mean differences in the number of patients per hospital admitted for a Heart Attacks between NPs and FPs ($H_1 : N_{NP} > N_{FP}$, where N is the number of patients) has the t-statistic value of 5.559 (322 FP observations and 1112 NP observations).

Although FPs have systematically less patients than NPs, it could be the case than within the same market fairly equal size hospitals locate. It turns out that this is not the case. Figure 3.3 plots the distribution of the mean number of patients per hospital by MSA (within organizational form). Even within MSAs NP hospitals have systematically more patients per hospital, on average than FPs. The t-test for mean differences in the average number of patients per hospital admitted for a Heart Attack across MSAs between NPs and FPs ($H_1 : N_{NP} > N_{FP}$, where N is the number of patients) has the t-statistic value of 4.097 (104 FP observations and 215 NP observations).[107]

[107]Not only that NP hospitals dominate in their numbers within MSAs, there is a larger number of MSAs where

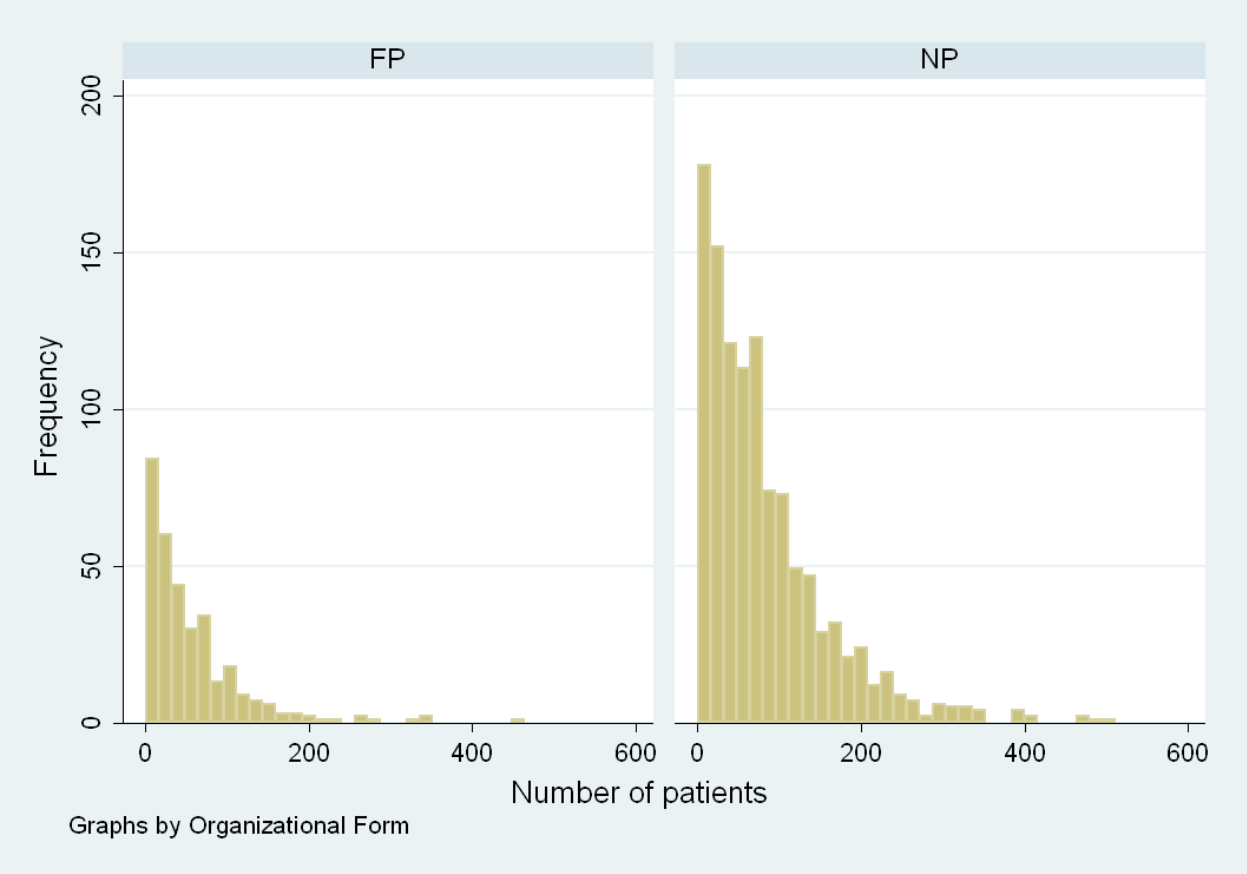


Figure 3.2: Distribution of the Number of (Heart Attack) Patients by Hospital for NPs and FPs (2006-2007 U.S. Hospitals)

The estimation methodology for mortality rates is likely to move a larger number of the estimates for FPs toward the U.S. mean than for NP hospitals. This feature implies that the variance of quality in the data is likely to be biased downward more within the FP sector than within the NP sector. The weighting aims to correct this bias. Note, that if there is not such bias, there is not reason to believe that the weighting will bias the relative variances of the organizational forms in the other direction. I show an example of the impact of the weighting on the entire distribution of

there is at least one NP hospital than the number of MSAs where there is at least one FP hospital (more than twice the amount: 104 MSAs for FPs versus 215 MSAs for NPs), see Section $B.2$ in Appendix B.

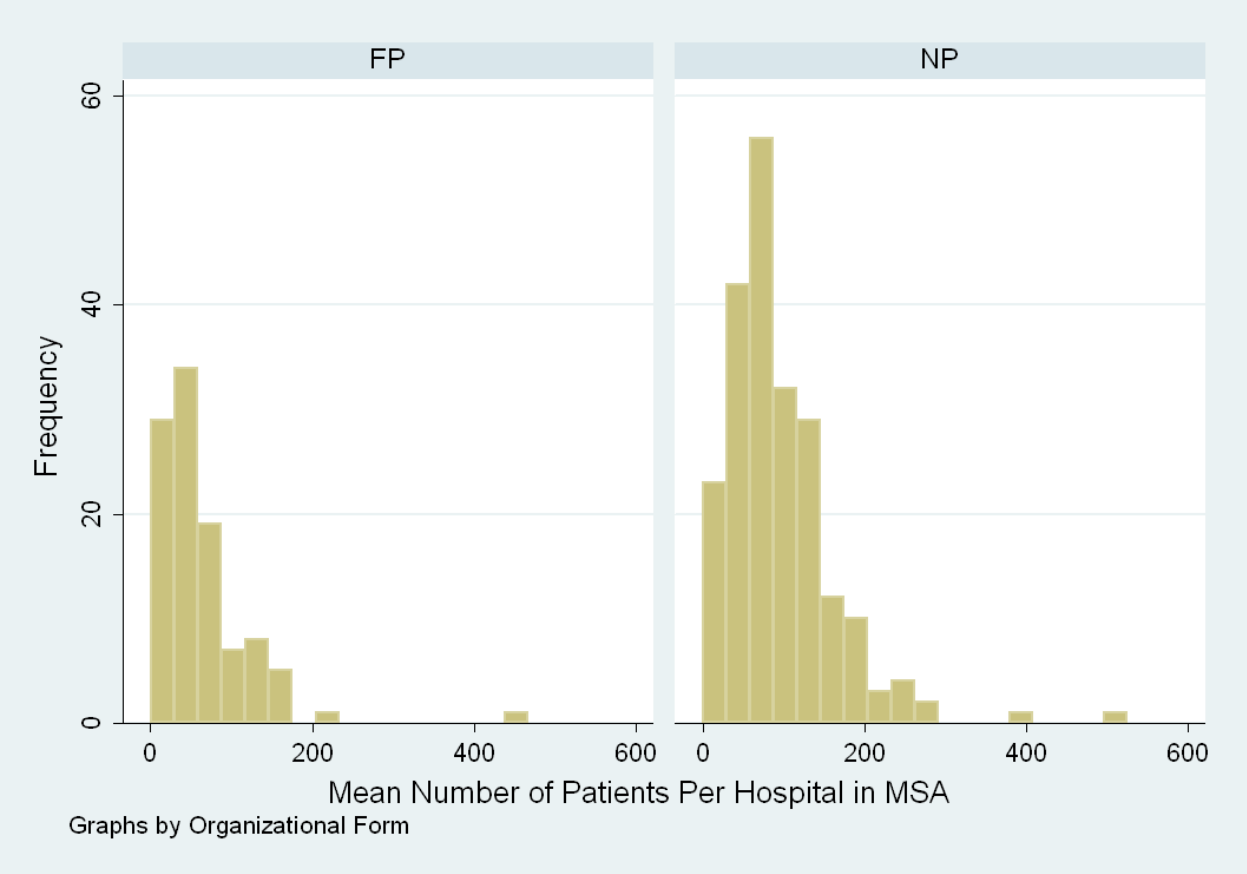


Figure 3.3: Distribution of the Average Number of (Heart Attack) Patients per Hospital by MSA for NPs and FPs (2006-2007 U.S. Hospitals)

mortality for one market in Appendix $B.3$.

I estimate each of the specification (3.4.1) through (3.4.3) twice: once with the market-level moments of mortality calculated from the unweighted data, and the second time with the moments calculated from the weighted mortality rates. I report the results for both types of estimations.

In this section I presented the specifications and the data for the tests of Hypotheses 1 through 3 (see Section 3.1). A weighting procedure was introduced to reduce biases due to systematic differences in hospital admissions between FPs and NPs. The next section reports the results of

these tests. The discussion of the empirical strategy and the test of Hypotheses 4 and 5 are delayed until Section 3.6.

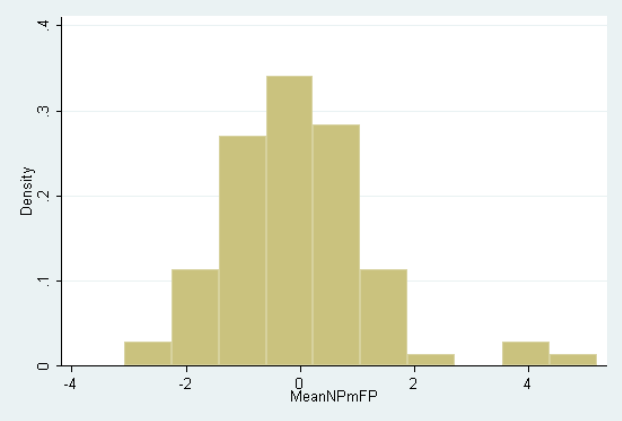
3.5 Estimation Results

In this section I discuss the empirical results. First, it should be instructive to examine the differences in market-level moments of the quality distribution between the NP and FP organizational forms. This gives a preliminary indication of whether the data is consistent with the economic theories of NPs: both existing and the one developed in this thesis. Later, I report the results of regression estimations and discuss their interpretation.

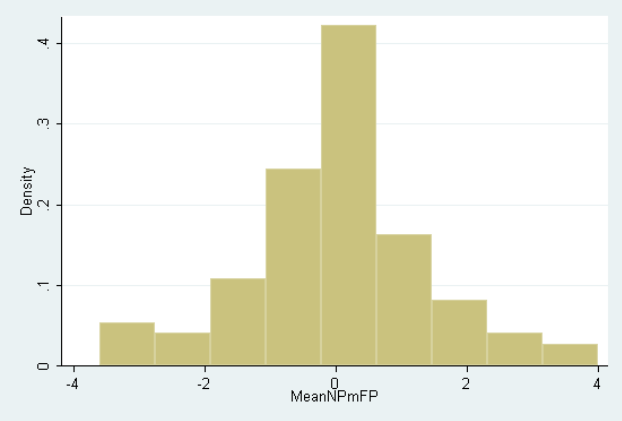
Previous theories of NPs stemming from the Arrow-Hansmann hypothesis predict that NPs provide higher quality of service than FPs (Hirth, 1999; Glaeser and Shleifer, 2001). To test preliminary consistency of the data with this prediction I compute the mean mortality for NP and FP hospitals separately in each market (MSA) and subtract the value for FPs from the one for NPs. I repeat this procedure for Heart Attacks and Heart Failures. Figure 3.4 plots the distribution of these differences by MSA. A distribution in which majority of the observations fall in the negative part of the horizontal axis would be consistent with the Arrow-Hansmann hypothesis that NPs have higher mortality rates than FPs.

In Figure 3.4 all distributions appear almost symmetric around zero. The t-test for mean differences between organizational form market-level means ($H_1 : Mean_{FP}[RSMR] > Mean_{NP}[RSMR]$) has the p-value of 0.38 for Heart Attacks (84 observations) and 0.531 for Heart Failures (86 obs.): equality of the market-level means could not be rejected at conventional significance levels.

In Figure 3.5 I repeat the above analysis after weighting the observations by patient shares within MSA, condition and organizational form, as discussed in Subsection 3.4.2. The market-level means were calculated from the weighted mortality rates.

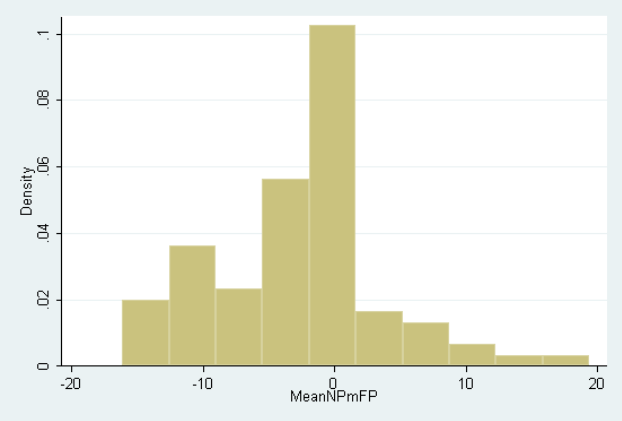


(a) Heart Attacks

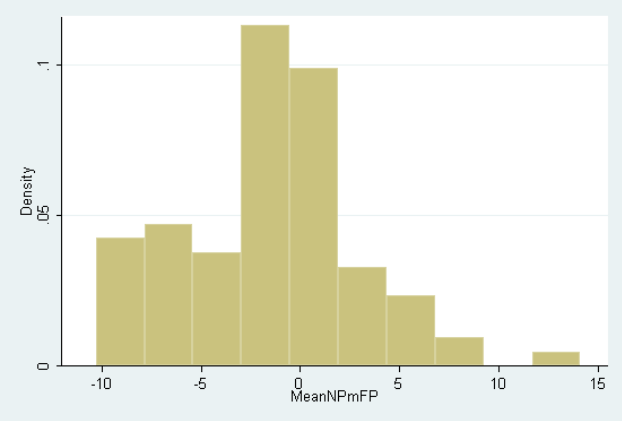


(b) Heart Failures

Figure 3.4: Distribution of $Mean_{NP}(RSMR) - Mean_{FP}(RSMR)$ by MSA for AMI and HF



(a) Heart Attacks

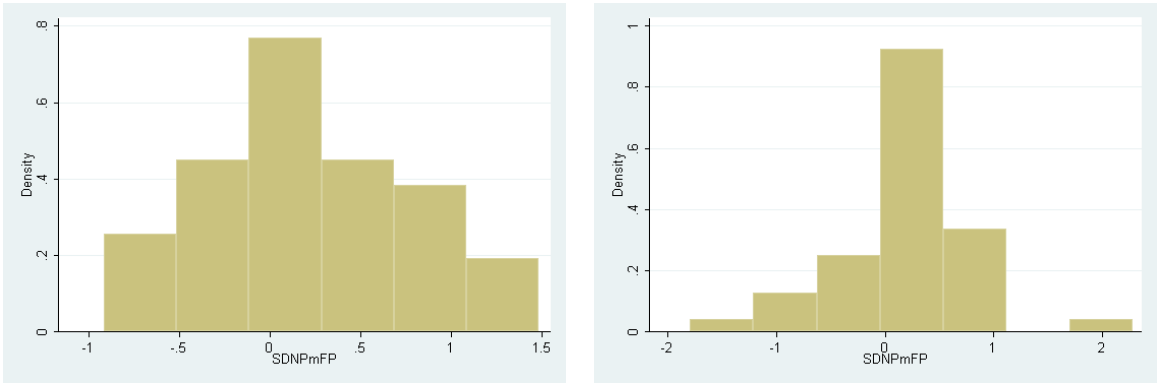


(b) Heart Failures

Figure 3.5: Distribution of $Mean_{NP}(RSMR) - Mean_{FP}(RSMR)$ by MSA for AMI and HF (RSMR Weighted by Patient Shares)

Figure 3.5 is more supportive of the Arrow-Hansmann hypothesis. The t-test for mean differences (of the market-level organizational form means: $H_1 : Mean_{FP}[RSMR] > Mean_{NP}[SMR]$) had the p-value of 0.0001 for Heart Attacks (84 observations) and 0.0005 for Heart Failures (86 obs.): equality of the means is rejected in favour of higher mean FP mortality at conventional levels. A large share of the existing empirical studies find lower quality of care at FP hospitals (Rosenau and Linder, 2003; Schlesinger and Gray, 2006). The consistency the findings here with a large share of the empirical literature supports the weighting procedure I employ, as similar results could not be replicated with unweighted mortality rates. It is likely that the weighting reduces measurement biases which are due to smaller hospital sizes.

I perform the same analysis with standard deviations. I calculate standard deviations of mortality for NP and FP hospitals separately in each market (MSA) and subtract the value for FPs from the one for NPs. I repeat the procedure for Heart Attacks and Heart Failures and plot the distribution of these differences by MSA in Figure 3.6.



(a) Heart Attacks

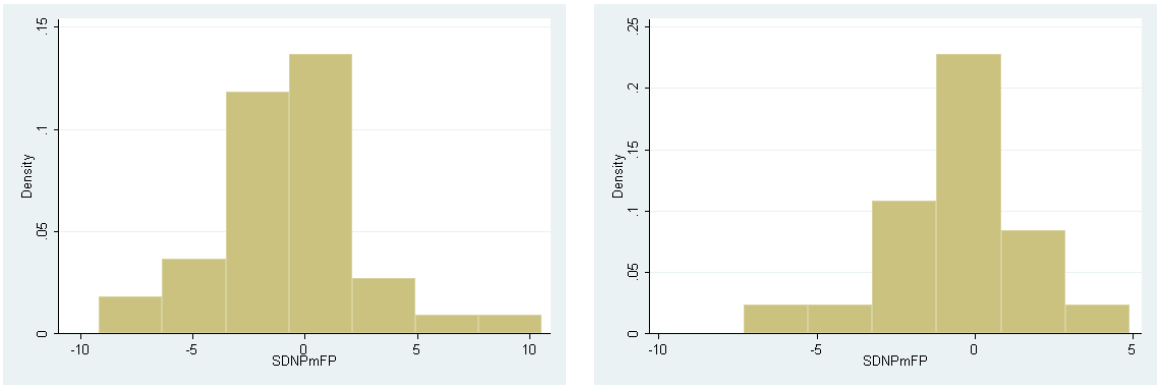
(b) Heart Failures

Figure 3.6: Distribution of $SD_{NP}(RSMR) - SD_{FP}(RSMR)$ for AMI and HF

The t-test for mean differences between market-level standard deviations for FP and NP quality ($H_1 : SD_{FP}[RSMR] > SD_{NP}[SMR]$) has the p-value of 0.977 for Heart Attacks (39 observations)

and 0.912 for Heart Failures (41 obs.): the hypothesis of mean equality can be rejected in favour of higher *NP* standard deviation. This feature may seem inconsistent with the theory proposed in this thesis (Hypothesis 1). It is likely, however, that the market-level standard deviations of FP quality are all biased downward, by the RSMR estimation procedure, due to systematically smaller sizes of FP hospitals relative to NP ones (see Subsection 3.4.2).

I repeat the procedure after weighting the RSMR by patient shares within MSA, condition and organizational form. Figure 3.7 presents the results of the analysis in which the standard deviations were computed using weighted mortality rates.



(a) Heart Attacks

(b) Heart Failures

Figure 3.7: Distribution of $SD_{NP}(RSMR) - SD_{FP}(RSMR)$ for AMI and HF (RSMR Weighted by Patient Shares)

Figure 3.7 seems more consistent with Hypothesis 1. The t-test for mean differences between the market-level standard deviations of FP and NP quality ($H_1 : SD_{FP}[RSMR] > SD_{NP}[RSMR]$) has the p-value of 0.065 for Heart Attacks (39 observations) and 0.053 for Heart Failures (41 obs.): although the sample sizes are small, one can reject the hypothesis of equality of standard deviations in favour of higher FP standard deviations at 10% confidence level.

The uncontrolled tests for differences in market-level moments of hospital quality by organiza-

tional form are in line with both the previous economic theories of NPs and the theory proposed in this thesis after reducing the effect of potential biases in the measurement of quality through weighting. The finding on differences in the first moment is consistent with a large number of previous studies. This increases the confidence in the weighting procedure introduced in Subsection 3.4.2. The differences in the second moments provide preliminary support to Hypothesis 1. I turn to more rigorous tests of Hypotheses 1 through 3.

3.5.1 Impact of the Organizational form on Mean Quality

Following previous empirical studies I examine the differences in the first moment of the quality distribution between NPs and FPs. In Section 3.3 I introduce the methodology of market-level analysis that has not been used previously. Subsection 3.4.2 develops a weighting procedures that is intended to put lower weight on hospitals for which the estimates of quality (RSMR) are less precise due to smaller number of patients. I test the newly developed methodology by repeating the analysis of the previous literature while using new tools. I estimate a specification similar to (3.4.1), however, here the dependent variable is mean quality instead of the standard deviation.

I calculate the mean RSMR (separately for Heart Attacks and Heart Failures) for FP, NP and government owned hospitals within each MSA and call this variable "Mean Mort." (measure of quality). This way each MSA appears up to three times in the dataset for each condition. I regress "Mean Mort." on a dummy variable for FP and government ownership (NP category omitted). The results are reported in Table 3.3.

Section 3.4 discusses the importance of controls. The literature argues for the importance of income in the determination of health outcomes (Jones and Wildman, 2008). Higher population of an MSA may allow better use of economies of scale exogenous to hospitals to increase overall quality of healthcare through community investments, for example. The education level is a significant determinant of health outcomes (DeWalt et al., 2004). I control for MSA population, medium

household income and the overall level of education. The variable "percentage of the MSA population with 9th to 12th grades (no high school diploma)" proxies for the general level of education in the population.

Column (1) of Table 3.3 reports the results of the estimation in which the dependent variable was computed from the mortality rates for Heart Attacks as supplied by Hospital Compare. None of the coefficients is statistically significantly different from zero at conventional levels and the model has low explanatory power. Specifically, the coefficient of interest on the FP dummy is positive, however, it is zero in statistical terms. In column (2) of Table 3.3 I include mortality for Heart Failures as well as for Heart Attacks in order to enrich the dataset. The model has a high explanatory power (R^2=0.84). The coefficient of interest, however, remains statistically insignificantly different from zero.

The negative and significant coefficient of MSA population in column (2) implies that the average quality of care increases with population size (this is true within FP organizational form as well - see Table 3.6). I propose two mechanisms through which population size impacts outcomes of care. 1) Economies of scale make investment in healthcare facilities that reduce mortality rates (e.g. enhanced 911 service) more profitable, implying higher investment in more populated ares. 2) More populated MSAs tend to have higher density of population, which increases the transportation time to the hospital on congested roads and increases mortality rates. From the results in column (2) I can conclude that the first mechanism is more likely to dominate.

The estimation of column (3) employs RSMR data for Heart Attacks that was first weighted by hospital's share of patients within its MSA and organizational form. It replicates the estimation of column (1) with the dependent variable calculated from the weighted mortality. The weighting seems to improve the fit, higher R^2 than in column (1). I use different control for the general level of eduction in column (3). The variable "percentage of the MSA population with high school degree or equivalent" enters the regression more significantly than the variable "percentage of the MSA

Table 3.3: The Effect of Organizational Form on Mean Quality (RSMR)

	(1)	(2)	(3)	(4)	MSA-Clustered S.E.
Dependent Variable:	Mean Mort.	Mean Mort.	Mean Weigh Mort.	Mean Weigh Mort.	Mean Weigh Mort.
FP	0.039	-0.031	2.232***	1.805***	2.232***
	(0.114)	(0.085)	(0.630)	(0.388)	(0.588)
GOV	0.036	0.018	3.971***	3.328***	3.971***
	(0.113)	(0.087)	(0.549)	(0.342)	(0.511)
MSA Population (mill.)	-0.010	-0.033***	-0.776***	-0.676***	-0.776***
	(0.011)	(0.010)	(0.129)	(0.082)	(0.207)
Med. HH Inc. MSA 1999 (000s $)	0.004	-0.004	-0.139***	-0.118***	-0.139***
	(0.008)	(0.006)	(0.035)	(0.022)	(0.041)
MSA Pop. with 9th to 12th gr. (%)	2.899	0.072			
	(2.459)	(1.801)			
HF	Excluded	-4.823***	Excluded	-3.056***	Excluded
		(0.071)		(0.297)	
MSA Pop. High School Degr. (%)			10.468**	8.480***	10.468**
			(4.236)	(2.620)	(4.549)
Constant	15.603***	16.265***	12.057***	11.956***	12.057***
	(0.550)	(0.411)	(2.021)	(1.249)	(2.173)
Observations	436	878	436	878	436
R-squared	0.01	0.84	0.28	0.33	0.28

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

population with 9th to 12 grade (no degree)", therefore, I use the more significant variable as the control.

The results replicate the conclusions of most of the empirical studies on the impact of organizational form on quality: NPs have on average higher quality of care than FPs (Rosenau and Linder, 2003; Schlesinger and Gray, 2006). The coefficient on the FP dummy in column (3) is positive and significant at 1% level, implying that FPs have higher average mortality rates than NPs. The results are also consistent with the findings of Shen (2002) that government owned hospitals have lower quality of care than NPs. More populated MSAs, those that have higher income and better educated population all benefit from lower mortality rates from Heart Attacks. This is consistent with previous studies that find positive association of quality of healthcare both with education and income (DeWalt et al., 2004; Jones and Wildman, 2008).[108]

The dependent variable in the estimation of column (4) includes market-level first moments of both Heart Failure and Heart Attack mortality weighted by patient-shares. Column (4) replicates the estimation of column (3) using a larger dataset. The results are qualitatively similar to column (3).

It could be the case, however, that the mean mortality rates of the three organizational forms are correlated within MSAs. In this case the estimates of the standard errors in the pooled OLS regressions are biased. Biased estimates of standard errors influence the power of the inference. Therefore, it possible that the significance of the coefficients in columns (3) and (4) of Table 3.3 is obtained solely due to biased estimates of the standard errors. For this reason I estimate specification (3) again while taking the clustering of the data into account. The results are reported in the last column of Table 3.3. Neither of the coefficients looses its significance relative to column (3). Clustering of the data does not seem to bias the inference.

[108]Recall that the model developed in Chapter 2 of this thesis is consistent with either of the organizational forms having higher quality of care on average.

In this subsection I replicate the analysis that has appeared in the empirical literature in the past using the new empirical methodology developed here. The results show that the dataset I use is consistent with the a large amount of previous findings that (on average) NP hospitals have higher quality of care than FPs. The fact that I am able to replicate the results of a large share of previous studies provides support for the new methodology. These findings are obtained only when I apply the weighting. This offers support to the weighting procedure specifically. Now, that I have a preliminary support for the methodology I continue to testing Hypotheses 1 through 3.

3.5.2 Test of the Main Implication of the Theory: Impact of the Organizational Form on Variance of Quality

The main implication of the theory developed in Chapter 2 of the thesis states that *FP hospitals have higher variance (standard deviation) of quality than NP hospitals* (Hypothesis 1). This implication is what allows my theory to reconcile previous contradictory empirical results on the ranking of NP hospitals versus FP hospitals by quality.

I estimate specification (3.4.1) to test Hypothesis 1. The dataset construction was described in detail in Section 3.4. I control for population, income and general level of education. The latter is proxied by the variable "percentage of the MSA population with 9th to 12th grades (no high school diploma)". Column (1) in Table 3.4 reports the results of the estimation with the market-level standard deviations calculated from mortality rates for Heart Attacks as the dependent variable. The coefficient of interest, on FP dummy, has a positive sign, as predicted, however, it is not significantly different from zero at conventional levels.

It could be the case that the number of observations is too small to identify the effect of interest. Column (2) reports the results for the regression in which the dependent variable includes market-level standard deviations of mortality for both Heart Attacks and Heart Failures. The coefficient of interest has the wrong sign and it remains zero in statistical terms.

Table 3.4: The Effect of Organizational Form on the Variance of Quality (RSMR)

	(1)	(2)	(3)	(4)	MSA-Clustered S.E.
Dependant Variable:	SD Mort.	SD Mort.	SD Weigh Mort.	SD Weigh Mort.	SD Weigh Mort.
FP	0.002	-0.108	0.736*	0.554**	0.736*
	(0.098)	(0.070)	(0.427)	(0.257)	(0.426)
GOV	-0.323***	-0.265***	1.521***	1.330***	1.521***
	(0.081)	(0.067)	(0.526)	(0.312)	(0.515)
MSA Population (mill.)	0.014	0.005	-0.249***	-0.203***	-0.249***
	(0.009)	(0.007)	(0.053)	(0.032)	(0.071)
Med. HH Inc. MSA 1999 (000s $)	0.009	0.004	-0.034	-0.030**	-0.034
	(0.007)	(0.005)	(0.021)	(0.013)	(0.022)
MSA Pop. with 9th to 12th gr. (%)	0.917	0.806			
	(2.197)	(1.655)			
HF	Excluded	0.223***	Excluded	-1.472***	Excluded
		(0.055)		(0.204)	
MSA Pop. High School Degr. (%)			4.292	2.993	4.292
			(3.515)	(2.109)	(3.324)
Constant	0.482	0.716*	3.935***	4.089***	3.935***
	(0.491)	(0.367)	(1.389)	(0.862)	(1.372)
Observations	232	470	232	470	232
R-squared	0.06	0.06	0.16	0.22	0.16

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

In Subsection 3.4.2 I introduce a weighting procedure that is intended to control for systematic differences in hospital size. The mortality rates are calculated by the U.S. Department of Health & Human Services, Hospital Compare in such a way that the estimates for smaller hospitals are moved toward the U.S. average mortality rate, thus reducing their variance. FP hospitals tend to be systematically smaller that NP ones (see Subsection 3.4.2), therefore, it is likely that the estimations in columns (1) and (2) are biased.

Columns (3) and (4) of Table 3.4 report the estimates when the dependent variable was calculated after weighting the mortality rates by patient shares within MSA, condition (Heart Attacks or Heart Failures) and organizational form. In column (3) only mortality rates from Heart Attacks are included, and in column (4) the rates from Heart Failures are included as well. In the weighted specifications I control for MSA characteristics through total MSA population, medium household income and "percentage of MSA population with a high school degree or equivalent". The latter variable controls for the overall level of education in the population. It enters the regression more significantly than "percentage of MSA population with 9th to 12 grade (no high school diploma)".

The coefficient on the FP dummy is positive and significant at 10% confidence level in columns (3) and (4). This implies that FPs have higher standard deviation of quality than NPs. This result is consistent with Hypothesis 1. One may be concenred about the reversal of the sign on the coefficient of the governement ownership dummy between columns (1)-(2) and (3)-(4). Government owned hospitals tend to be systematically smaller than both NP and FP ones. Therefore, the bias of the mortality estimates toward the U.S. mean is likely to be happen for an even lareger number of governement owned hospitals. There is no theoretical or empircal study that addresses the relative variance fo quality for government owned hosptials, to the best of my knowledge. Therefore, I am unable to determeine which sign on the government dummy is more resonable.

Total MSA population and medium household income seem to impact the standard deviation of quality negatively in Columns (3) and (4). The analysis of Subsection 3.5.1 shows that all hospitals

(irrespective of organizational form) improve their quality of care as the population and the income rise. Higher concentration of all hospitals at the upper tail of the distribution of quality reduces the observed variance. There is consistency between the findings of the two subsections.

It could be the case that the standard deviations of mortality rates of the three organizational forms are correlated within MSAs. In this case the estimates of the standard errors in the pooled OLS regressions are biased. It possible that the significance of the coefficients in columns (3) and (4) of Table 3.4 is obtained solely due to the biased estimates of the standard errors. For this reason I estimate specification (3) again while taking the clustering of the data into account. The results are reported in the last column of Table 3.4. None of the previously significant coefficients looses its significance. Specifically, the coefficient on the FP dummy remains significant at 10% level. This indicates that clustering of the data is unlikely to bias the infrerence.

The estimations reported in Table 3.4 are in line with Hypothesis 1 after applying the weighting procedure. This offers some support to the main implication of the new theory of NPs developed in this thesis. The new picture of healthcare markets where FP hospitals tend to be concentrated on the tails of the distribution of quality, while NPs are in the middle of it has some merit. I turn to the tests of Hypotheses 2 and 3 that involve comparative statics predictions.

3.5.3 Tests of Comparative Statics Hypotheses 2 and 3

I test Hypothesis 2 which states that *An increase in the proportion of informed patients increases the mean FP quality for small proportions of informed, and decreases the mean FP quality for large proportions* by estimating specification (3.4.2). The market-level mean mortality (for Heart Attacks) across FP hospitals is regressed on the "percentage of the MSA population with a graduate or professional degree". The latter variable proxies for the share of informed consumers in the populations, as higher education should be associated with both better inference of healthcare quality and higher chances of inclusion of the medical personnel in one's social circle.

Column (1) of Table 3.5 reports the results of a naive specification, without controlling for the characteristics of the MSA. The coefficient of the variable of interest and its square have the predicted signs, but are not significantly different from zero at conventional levels.

Table 3.5: Regression tests of Hypothesis 2

	(1)	(2)	(3)
Dependant Variable:	Mean Mort. FP	Mean Mort. FP	Mean Mort. FP
MSA Pop. with Grad (%)	-14.416	-23.740	-35.779**
	(11.668)	(17.310)	(17.890)
$Grad^2$	59.598	93.952	122.805*
	(50.092)	(69.354)	(69.511)
MSA Population (mill.)		-0.007	0.018
		(0.023)	(0.025)
Med. HH Inc. MSA 1999 (000s $)		0.016	0.005
		(0.017)	(0.017)
MSA Pop. with 9th to 12th gr. (%)			-10.735**
			(5.049)
Constant	16.848***	16.727***	19.180***
	(0.583)	(0.599)	(1.303)
Observations	104	104	104
R-squared	0.01	0.02	0.06

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

In column (2) I control for the population of MSA and medium household income in the MSA in 1999 (the data available). The coefficients of interest (on percentage of population with graduate and professional degree and its square) remain statistically insignificantly different from zero. However, they maintain he predicted signs. The variable "percentage of MSA population with graduate or professional degree" is correlated with the general level of education. Therefore, the coefficient estimates of the variables of interest are likely to be biased.

Column (3) of Table 3.5 presents the results of a regression where I control for the general level of education in the population. The coefficients on "percentage of MSA population with graduate

and professional degree" and its square become statistically significant and have the predicted sign. The data supports Hypothesis 2.[109]

It is interesting to note that the coefficient on the variable "percentage of the population with 9th to 12th grade of education (but no high school diploma)" is negative and significantly different from zero. This implies that the higher is the overall level of education, the lower mortality rates are expected at FP hospitals. Previous research finds positive impact of education on health (DeWalt et al., 2004). The results of column (3) innovate by showing that the generally accepted relationships also holds in markets where FPs choose to locate.

I replicate the analysis of Table 3.5 after weighting the mortality rates. I calculate the dependent variable from the weighted rates. The results of this analysis are reported in Table 3.6.

Once one controls for the differences in hospital sizes through weighting, the results are statistically significant and supportive of Hypothesis 2 in all specifications including the naive one. The specifications using weighted data have better fit than the unweighted ones (Table 3.5). Additionally, some of the coefficients are estimated more precisely in Table 3.6. This may be interpreted as an extra support for the weighting procedure.

I test Hypothesis 3 that *An increase in the proportion of informed patients reduces the variance (standard deviation) of NP quality.* by estimating specification (3.4.3). I regress the standard deviation of mortality for NP hospitals (at the market-level) on the "percentage of MSA population with graduate or professional degree" (proxy for the share of informed patients). Table 3.7 reports the results.

Column (1) presents the results when the specification was estimated without any controls. The coefficient is not significantly different from zero. In column (2) I control for the general level of education in the MSA. In column (3) the population of the MSA and the median household

[109]Each market enters the dataset only once, therefore, there is no potential clustering of the residuals at the MSA-level as in Subsections 3.5.1 and 3.5.2.

Table 3.6: Hypothesis 2 tested using RSMR weighted by hospitals' patient-shares

	(1)	(2)	(3)
Dependant Variable:	Mean Weigh Mort. FP	Mean Weigh Mort. FP	Mean Weigh Mort. FP
MSA Pop. with Grad (%)	-255.446***	-191.700**	-226.546**
	(68.486)	(86.464)	(91.035)
$Grad^2$	1,128.665***	947.739**	1,031.249***
	(318.484)	(370.655)	(374.987)
MSA Population (mill.)		-0.711***	-0.640***
		(0.192)	(0.197)
Med. HH Inc. MSA 1999 (000s $)		-0.008	-0.039
		(0.089)	(0.094)
MSA Pop. with 9th to 12th gr. (%)			-31.069
			(30.232)
Constant	23.016***	20.528***	27.630***
	(3.332)	(3.441)	(7.613)
Observations	104	104	104
R-squared	0.09	0.21	0.22

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.7: Regression tests of Hypothesis 3

	(1)	(2)	(3)
Dependant Variable:	SD Mort. NP	SD Mort. NP	SD Mort. NP
MSA Pop. with Grad (%)	0.532	0.557	-2.112
	(1.770)	(2.379)	(2.608)
MSA Pop. with 9th to 12th gr. (%)		0.041	0.848
		(2.725)	(3.307)
MSA Population (mill.)			0.020
			(0.014)
Med. HH Inc. MSA 1999 (000s $)			0.016*
			(0.009)
Constant	0.926***	0.919*	0.392
	(0.166)	(0.471)	(0.696)
Observations	136	136	136
R-squared	6e-04	6e-04	0.04

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

income are used as controls as well. Again, the coefficient on the proportion of the population with graduate or professional degree is zero in statistical terms in all of these specifications. In column (3), however, where all of the controls are employed, the coefficient of interest (on "percentage of MSA population with graduate or professional degree") has the predicted sign.

It is interesting to note that the coefficient on income in column (3) is significantly different from zero at 10% confidence level. The sign of this coefficient is positive. In Subsection 3.5.2 I test the impact of income on overall standard deviation of quality. It turns out that the results in Table 3.7 are at odds with the ones in Subsection 3.5.2. It is hard to reconcile the results beyond the marginal significance of the result here and the fact that the coefficient is significant only in column (4) of Table 3.4 and not in other specifications.

I replicate the analysis in Table 3.7 on mortality rates first weighted by the hospital's share of patients. The results of this analysis are reported in Table 3.8.

Table 3.8: Hypothesis 3 tested using RSMR weighted by hospitals' patient-shares

	(1)	(2)	(3)
Dependant Variable:	SD Weigh Mort. NP	SD Weigh Mort. NP	SD Weigh Mort. NP
MSA Pop. with Grad (%)	-16.778**	-21.078**	1.480
	(7.614)	(9.054)	(9.932)
MSA Pop. with 9th to 12th gr. (%)		-6.991	0.877
		(11.457)	(12.790)
MSA Population (mill.)			-0.338***
			(0.113)
Med. HH Inc. MSA 1999 (000s $)			-0.044
			(0.040)
Constant	4.856***	5.995***	5.480**
	(0.655)	(1.811)	(2.709)
Observations	136	136	136
R-squared	0.03	0.03	0.16

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Column (1) of Table 3.8 presents the naive specification. Without controlling for MSA characteristics, the coefficient on "percentage of MSA population with graduate or professional degree" is negative and statistically significant at 5% level. The naive specification is in line with Hypothesis 3.

Column (2) shows the results when one controls for the overall level of education. The significance of the coefficient of interest marginally increases. It is particularly important to control for the overall level of education, as it is correlated with the percentage of population who have obtained graduate or professional degrees and its omission biases the estimate of the coefficient on the latter.

Column (3) shows that controlling for medium household income and population makes the coefficient on the percentage of population with graduate degree loose its significance and reverse

the sign. On the other hand, the population of MSA seems to be negatively associated with the standard deviation of quality for NP hospitals. This is consistent with the results in Subsection 3.5.2 when the dependent variable is calculated from weighted mortality rates. As the results involve relationships of second moments, they are quite hard to interpret intuitively, without a model. However, it is plausible that as the population grows NP hospitals become more homogeneous (the standard deviation decreases, possibly because they all improve quality of care due to learning by doing) on one hand. On the other hand, the percentage of highly educated people tends to be higher in larger cities. For this reason it may be hard to separate the effect of the share of informed consumers from the impact of the population size.[110]

In this section I tested the predictions of the theory proposed in Chapter 2 of this thesis on a newly constructed dataset. I use risk adjusted 30-day post-admission mortality rates for Heart Attacks as the main measure of quality. The results are reported in Tables 3.4 through 3.8. The central implication of my theory (Hypothesis 1) is supported by the sample where the dependent variable was calculated from weighted mortality rates. Hypothesis 2 is supported rather strongly by the data, including the second derivative. Hypothesis 3 is supported only when one does not control for the market characteristics other than the general level of education. Overall, the implications of the theory involving the provision of quality find support in the data. There is some merit to the new picture of healthcare markets proposed in this thesis. The next section reports the test of Hypotheses 4 and 5 involving waiting times.

[110] An alternative explanation that in highly populated areas each hospital has more admissions, which enables to estimate mortality rates more precisely (with lower variance) is possible as well. However, this assumes that all of the NP hospitals have the same underlying quality (market-level moments are their sample estimates), which need not hold especially across markets.

3.6 Waiting Times

The theory developed in Chapter 2 of this thesis generates the following predictions regarding queues or waiting times: *NPs have longer queues that FPs* (Hypothesis 4) and *Higher quality NP hospitals have longer queues than lower quality NPs* (Hypothesis 5). In this section I review the literature and show the importance of the issue of waiting times. I discuss the empirical strategy employed to test these hypotheses. Finally, I describe the data used and report the results.

A recent report by the Fraser Institute[111] shows that the issue of waiting times for medical procedures is extremely acute in Canada. Waiting periods for most procedures have increased significantly between 1993 and 2008. The actual waiting times diverge from reasonable ones by a large amount for many procedures. The waiting times in this report are measured in weeks between appointment and actual treatment. Siciliani and Hurst (2005) offer a review of waiting times for elective surgeries for a number of OECD countries.

Economic literature has examined waiting times both from theoretical and empirical perspectives. Brekke, Siciliani, and Straume (2008), for example, study the impact of the degree of competition between hospitals on waiting times in a theoretical model of product differentiation. Siciliani, Stanciolec, and Jacobs (2009) examine the effect of waiting times on hospital costs in a sample of U.K. hospitals.

Surprisingly, economists have paid little attention to waiting times in the U.S. hospital industry. I found only one recent study on the issue. Wilper et al. (2008) examine the change in the time needed to see a physician after arrival in the U.S. Emergency Departments between 1997 and 2004. The data for the study comes from the National Hospital Ambulatory Medical Care Survey (NHAMCS) administered by the U.S. National Centre for Health Statistics. I review this work in detail, as my investigation uses the same data source (newer dataset), and follows this study

[111] Fraser Institute 2008, "Waiting Your Turn: Hospital Waiting Lists in Canada 2008 Report", by Nadeem Esmail and Maureen Hazel with Michael A. Walker, Studies in Health Care Policy, 18th Edition.

closely.

Wilper et al. (2008) investigate the impact of the patient and hospital level factors on the waiting time from patient's arrival (the first time recorded in the medical record) until her examination by a physician (time in minutes from patient data form), and how it changed between 1997 and 2004. The following patient level characteristics were used: age, ethnicity, sex, expected sources of payment, urgency of care (triage), patient's reasons for the visit (up to 3), primary diagnosis Acute Myocardial Infarction (AMI), whether the patient was evaluated by resident MD and whether the patient was admitted; these data come from a survey competed by the medical staff. The following hospital level controls were used: ownership status (NP, Government - non-federal, and FP) and whether the hospital belongs to an MSA. The ethnicity is determined by the hospital staff, unless the patient is asked directly.

Wilper et al. (2008) find that the waiting time increased between 1997 and 2004 from a median of 22 to a median of 30 minutes. Females waited longer than males, and non-White ethnicities waited longer than the Whites. In urban areas waits were longer. Consistent with Hypothesis 4 of this chapter, waiting times at NPs were longer as well.

Although the medical literature has addressed the issue of waiting times in the U.S. (emergency rooms) more than the economic one, it is fairly sparse. James, Bourgeois, and Shannon (2005) examine the effect of child's ethnicity on the waiting time to see a physician. The authors conclude that ethnicity has a significant impact on waiting times. Children of Hispanic background waited longer than either White or Black non-Hispanic children. Richardson and Hwang (2001) offer a literature review of emergency medicine. Only two of the reviewed articles are concerned with waiting times. The main finding is that waiting times have an adverse impact on patient's health status. Patients choose to leave emergency rooms without being treated when the waits are long. This often causes further deterioration in their health.

I obtained the data for my study from the National Hospital Ambulatory Medical Care Survey

(NHAMCS) administered by the U.S. National Centre for Health Statistics[112]. The data is similar to the one used by Wilper et al. (2008), however, it was collected for 2006 emergency department visits. In the following paragraphs I review the coding of the variables I use in the estimations. I change the values to missing if the respondent left the question blank or the value is unknown.[113]

The dependent variable in this empirical investigation is the waiting time measured as the difference in minutes between the time recorded on the medical record (patient's arrival) and her examination by a physician. Both of these values appear on patient data form. Wilper et al. (2008) argue that there is a potential for a number of outliers in the data with atypically long waiting times. The authors recommend applying a logarithmic transformation to the wait time in minutes. Other studies seem to follow this procedure as well (James, Bourgeois, and Shannon, 2005). I apply the same transformation.

One of the main relationships of interest in my study is the impact of organizational form on the waiting time. The dataset contains information on hospitals' organizational form (ownership). There are three types of ownership in the data. I code NP and government ownership through dummy variables and omit the FP dummy. The data comes from voluntary NP hospitals, government non-federal (federal hospitals tend to belong to specialized systems, such as veteran affairs) and proprietary (FP).

Hospital quality of care is important for Hypothesis 5. I measure quality through the average number of repeated visits of the emergency department within 72 hours.[114] More repeated visits imply lower quality. It is possible, however, that the reason for the repeated visit of the emergency department was a previous disease spell (potentially improperly treated at a different facility). I adjust the average repeated share of visits in a manner that only those patients who were not

[112] `http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm` accessed on 08 July 2009 22:16.

[113] The documentation on survey design and variable coding can be obtained from `ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHAMCS/doc06.pdf` accessed on 08 July 2009 at 22:10.

[114] I was not able to match hospitals in the NHAMCS sample with their mortality rates from the Hospital Compare sample, as Medicare/Medicaid provider IDs were not available for hospitals in the NHAMCS dataset.

hospitalized within the past 7 days are included in calculating the average. I call this variable "adjusted repeated visit".

Hypothesis 5 is concerned directly with the impact of hospital quality on the waiting time (in NP hospitals). I use "adjusted repeated visit"as a proxy for (bad) quality of care. Controlling for patient characteristics reduces the impact of patient-mix. One may claim that hospital selection is likely to occur. There are two possible directions of such selection: hospitals which treat patients faster receive more patients. In this case there is a bias against finding significant differences between hospitals, and support for the hypotheses. The other direction is that patients select hospitals on quality. The latter, however, is modelled explicitly in Chapter 2.[115] Therefore, it is important to allow such selection in the data to test Hypothesis 5.

The dataset contains information on patient's health status controls. There are 5 levels of urgency (triage) of treatment in the data. I code the urgency through dummy variables according to whether the urgency was immediate, could be delayed by 1 to 14 minutes, by 15 to 60 minutes, or by 1 to 2 hours, with non-urgent (up to 24 hour delay) category omitted. The more urgent the procedure is the less likely patients are to select a hospital. The degree to which patients are able to select the hospital is not addressed in the model of Chapter 2. Therefore, it is important to control for such flexibility.

Following the literature (Wilper et al., 2008), I control for patient's expected source of payment. The following categories are coded through dummies: private insurance, Medicare, Medicaid, worker compensation, self (out-of-pocket) and the cases for which there was no charge ("other" category omitted).

NHAMCS collects information on patients' demographics characteristics. It is important to control for these characteristics, as case-mix may differ systematically across hospitals. I code

[115]Recall that the uninformed patients infer the quality of NP hospitals perfectly (due to the non-distribution constraint). See Subsection 3.1.1.

5 categories of age through dummy variables: 15 to 24 years, 25 to 44, 45 to 64, 65 to 74 and over 75 (under 15 years of age is the omitted category). A female dummy controls for the sex of the individual. Ethnicity, unfortunately, is a significant determinant of waiting time in the U.S. emergency departments (James, Bourgeois, and Shannon, 2005; Wilper et al., 2008). I code the ethnicity variables in the following manner: whether the individual is White (non-Hispanics), Black (non-Hispanics), Hispanic, Asian only, Native Hawaiian or Other Pacific Islander only, Native American or Alaska Native, the multiple races category is omitted.

The dataset contains information on the demographics of the population in the zip code of the hospital. I code this information in the following manner. I introduce dummy variables for whether there is less than 5% poverty in the zip code, 5% to 9.99%, 10% to 19.99% and omit the 20% and over catergory. Higher poverty is likely to increase waiting times in the region due to higher indegent demand for uncompensated care.[116] I code the zip code level median household income through dummies as follows: income between $32,794 and $40,626, $40,627 to $52,387, $52,388 or higher; I ommit the category of incomes below $32,793. Income is an important determinant of population's health status and demand for healthcare (DeWalt et al., 2004). Following the literature (Wilper et al., 2008), the region of the U.S. in which the hospital is lcoated is controled for through Northeast dummy, Midwest and West dummy (South is omitted). Finally, whether the hospital belongs to an MSA is controlled through an urban dummy. As the population dancity is higher in urban areas, hospitals tend to be more conggested.

The NHAMCS data was collected through a survey prepared by the U.S. Census Bureau. The dataset was assembled using a four stage probability design.[117] The dataset provides appropriate weights needed to obtain U.S. national estimates. The goal of my investigation, however, is to test hypotheses generated by theory. For this reason I am more interested in partial correlations

[116]Every hospital in the U.S. is required by law to provide charity care upon patient's arrival.

[117]The documentation on survey design and variable coding can be obtained from `ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHAMCS/doc06.pdf` accessed on 08 July 2009 at 22:10.

between variables rather than in national estimates. The observations in the dataset are at the patient level, however. It is likely that the error terms are correlated at the hospital level. More crowded hospitals potentially impact the waiting time for all of their patients. In order to check for robustness of the standard error estimates, I report the results of an estimation in which the standard errors are clustered at the hospital level, as well.

Table 3.9 presents the summary statistics for the variables of interest to the testing of the hypotheses: wait time, organizational form dummies and adjusted percentage of repeat visits, which proxies for the quality of care; and zip code level controls. In order to save space I do not report summary statistics and the coefficient estimates of the patient level controls. There are 364 hospitals in the dataset (as the value of the zip code level demographic variables is assigned to each patient observation, the number of patient observations and zip code characteristics is equal in the table). Only 347 of those hospitals enter the regressions, due to missing variables.

Table 3.9: Summary Statistics: Wait Time Estimations

Variable	Obs	Mean	Std. Dev.	Min	Max
waittime	28391	58.387	88.89	0	1430
NP	28391	0.734	0.442	0	1
GOV	28391	0.179	0.383	0	1
NP × Adj. Repeat Visit (%)	28055	0.024	0.049	0	0.75
Adj. Repeat Visit (%)	28055	0.031	0.049	0	0.75
Med. HH Inc. $32,794 to $40,626	27128	0.25	0.433	0	1
Med. HH Inc. $40,627 to $52,387	27128	0.214	0.41	0	1
Med. HH Inc. $52,388 or more	27128	0.214	0.41	0	1
Poverty: Below 5%	27124	0.144	0.351	0	1
Poverty: 5% to 9%	27124	0.262	0.44	0	1
Poverty: 10% to 19%	27124	0.363	0.481	0	1
Region: Northeast	28391	0.239	0.426	0	1
Region: Midwest	28391	0.215	0.411	0	1
Region: West	28391	0.185	0.388	0	1
Urban	28391	0.883	0.322	0	1

One can deduce from Table 3.9 that the variable *waittime* is highly volatile, which justifies the logarithmic transformation suggested by the literature. Majority of the observations seem to come from the South and from MSA areas. Southern region is the omitted category, as the shares of the other regions sum up to 64%, this implies that 36% of the observations come from the Southern region. The lowest share of the observations comes from the Western region. Over 70% of the patients were treated at NP, and only 9% at FPs.

I estimate the following specification to test Hypothesis 4 that *NPs have longer queues that FPs* and Hypothesis 5 that *Higher quality NP hospitals have longer queues than lower quality NPs.*

$$LNwaittime_p = \gamma_0 + \gamma_1 NP_h + \gamma_2 GOV_h + \gamma_3 NP \times RepVis_h + \gamma_4 RepVis_h + \gamma_5 Z_h + \gamma_6 C_p + \vartheta_p \quad (3.6.1)$$

Where $LNwaittime_p$ is the natural logarithm of the waiting time in minutes between patient p's arrival and her examination by a physician. NP_h and GOV_h are dummy variables that equal 1 if hospital h has NP status, or is government owned, respectively (FP category omitted). $RepVis_h$ is the percentage of repeated visits of the hospital's emergency department adjusted to include only patients who were not discharged from a hospital in the past 7 days. Z_h is a vector of demographic characteristics of the population in the zip code where the hospital is located, and C_p is the vector of patient characteristics. ϑ_p is the error term that captures the impact of unobservable factors that impact waiting times. ϑ_ps are potentially correlated within hospitals, clustering is employed to check for robustness.

Estimation of specification (3.6.1) enables to test Hypotheses 4 and 5 in the following manner. Positive and significant coefficient on the NP dummy supports Hypothesis 4. Negative and significant coefficient on the interaction between the adjusted percentage of repeated visits and the NP dummy is in line with Hypothesis 5.

Previous literature controls for chest pain complaint, AMI diagnosis, and an examination by a

resident MD (Wilper et al., 2008). Due to problems in the dataset I was not able to code for AMI diagnosis and examination by a resident MD. I did not see a reason to differentiate between chest pain and other urgent conditions. If these conditions have an impact, they increase the noise in the data, which biases against finding significant results.

Table 3.10 reports the estimation results. Column (1) estimates the regression only with the variables of interest and no controls. The coefficient on the NP dummy is positive and significantly different from zero. This is in line with Hypothesis 4. The coefficient on the interaction between the adjusted percentage of repeated visits (a proxy for quality) and the NP dummy is negative and significantly different from zero. This fact supports Hypothesis 5. It is interesting to note that in general (i.e. not for the NP group), higher percentage of repeated visits is associated with longer waiting times.

In column (2) I control for zip code level characteristics. The coefficients of interest remain supportive of the hypotheses. It seems that waiting times are longer in areas with higher income. Poverty does increase wait times, however, the main difference is between high poverty areas (over 20% - the omitted category) and the rest. Waiting times are the longest in the South (the omitted region). In urban areas hospitals are significantly more conggested. Surprisingly, Wilper et al. (2008) find longer waiting times in the Northeast than in the South. My investigation ranks the Northeast right behind the South in terms of waiting time length. Other resutls are consistent across the two studies.

Column (3) reports the results of the estimation in which both the zip code level and the patients level characteristics appear as controls. The coefficient of interest remain significantly different from zero and supportive of the hypotheses. Qualitatively, the results do not differ from column (2). The results on patient level characteristics, not reported here, are consistent with Wilper et al. (2008).

In the last column I check for robustness of the standard error estimates through clustering at the hospital level. Clustering seems to matter. The coefficient on the NP dummy remains

Table 3.10: Test of Hypotheses 4 and 5

	(1)	(2)	(3)	Hospital-Clustered
Dependent Variable:	LNwaittime	LNwaittime	LNwaittime	LNwaittime
NP	0.166***	0.324***	0.331***	0.331***
	(0.029)	(0.030)	(0.030)	(0.096)
GOV	0.164***	0.227***	0.302***	0.302**
	(0.030)	(0.032)	(0.031)	(0.117)
NP × Adj. Repeat Visit (%)	-2.193***	-4.277***	-3.275***	-3.275*
	(0.521)	(0.519)	(0.504)	(1.986)
Adj. Repeat Visit (%)	2.191***	4.298***	3.549***	3.549*
	(0.506)	(0.505)	(0.487)	(1.901)
Med. HH Inc. $32,794 to $40,626		0.105***	0.088***	0.088
		(0.023)	(0.023)	(0.060)
Med. HH Inc. $40,627 to $52,387		0.175***	0.170***	0.170**
		(0.028)	(0.029)	(0.074)
Med. HH Inc. $52,388 or more		0.141***	0.131***	0.131
		(0.034)	(0.035)	(0.096)
Povrery: Below 5%		-0.278***	-0.146***	-0.146*
		(0.038)	(0.040)	(0.084)
Poverty: 5% to 9%		-0.237***	-0.134***	-0.134*
		(0.030)	(0.032)	(0.072)
Poverty: 10% to 19%		-0.154***	-0.098***	-0.098*
		(0.023)	(0.024)	(0.055)
Region: Northeast		-0.184***	-0.169***	-0.169*
		(0.019)	(0.020)	(0.086)
Region: Midwest		-0.283***	-0.234***	-0.234**
		(0.020)	(0.021)	(0.091)
Region: West		-0.199***	-0.113***	-0.113
		(0.022)	(0.024)	(0.087)
Urban		0.561***	0.460***	0.460***
		(0.022)	(0.024)	(0.099)
Patient level controls	No	No	Yes	Yes
Constant	3.312***	2.882***	3.223***	3.223***
	(0.028)	(0.038)	(0.093)	(0.215)
Observations	27290	26067	22010	22010
R-squared	0.002	0.04	0.13	0.13

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

significant at 1% confidence level indicating strong support for Hypothesis 4. The coefficient on the itneraction term, however, has the p-value of 10% exactly. This implies only weak suppor for Hypothesis 5.

It is interesting to note that after the correction for clustering of the standard errors the urban dummy is the only zip code level control that maintains its previous significance level. The clustering ranks the West closest to the South in terms of waiting time length (in fact there is no significant difference between the two). Poverty continues to impact waiting times. Income levels, however, seem to loose their differential impact. Only one of the income coefficients remains significant.

In this section I performed an empirical investigation of the impact of organizational form on the waiting time length for hospital procedures. There is only a handful of papers that address waiting time length in U.S. hospitals. Although it seems to be less of a concern in the U.S. relative to the other OECD countries, waiting times in the U.S. have a negative impact on the health status of the population (Richardson and Hwang, 2001). I use data on waiting time length in U.S. emergency departments in 2006 from NHAMCS administered by the U.S. National Centre for Health Statistics in this study. Hypothesis 4 finds strong support in the data. It is consistent with the findings of previous studies as well. Hypothesis 5 finds only weak support. The clustering of the observations within hospitals is important.

3.7 Conclusion

In this chapter I test the implications of the theory of NP hospitals that was developed in Chapter 2 of this thesis. The aim of this thesis is to investigate the impact of organizational form on the quality of healthcare. Hypotheses 1 through 3 generated by the theory address the issue of quality. Hypothesis 1, specifically, deals with the central implication of the new theory of NPs that *FP hospitals have higher variance (standard deviation) of quality than NP hospitals.* Support for this

hypothesis implies that the new picture of healthcare markets proposed in this thesis has some merit.

The implications of the theory extend beyond the quality of care, however. Hypotheses 4 and 5 address the effect of the organizational form on waiting times for hospital procedures. As this is not the central topic in this thesis, I restrict the discussion of waiting times to one section of this chapter (Section 3.6).

The analysis in this chapter is based on a central methodological innovation. When investigating the impact of organizational form on quality of care, I examine its effect on the market-level moments of the quality distribution. This feature offers the most direct test of NP theories up to date. I exploit the variation at the market level to estimate the parameters of interest. Although previous NP theories can be interpreted as generating only market-level predictions, no similar analysis has been performed so far, to my knowledge.

There exists an abundance of measures of hospital quality, which I review in detail. Following the most recent economic literature, I use risk adjusted (or standardized) mortality rates for Heart Attacks (and in some cases for Heart Failures) obtained from the Hospital Compare database as a measure of quality. The measures of mortality rates for smaller hospitals are biased toward the U.S. national mean due to the computation procedure employed by the researchers at the U.S. Department of Health & Human Services. A waiting procedure was needed to reduce the impact of this bias in the estimations.

I construct an original dataset and calculate the relevant moments in the distribution of quality within markets for each organizational form separately. Treating each market as one observation I estimate a number of specifications to test Hypotheses 1 through 3. Hypotheses 1: *FP hospitals have higher variance (standard deviation) of quality than NP hospitals* and Hypothesis 2: *An increase in the proportion of informed patients increases the mean FP quality for small proportions of informed, and decreases the mean FP quality for large proportions* seem to find support in the

data. The data is weakly in line with Hypothesis 3: *An increase in the proportion of informed patients reduces the variance (standard deviation) of NP quality.*

Section 3.6 discusses waiting time length. Although the issue is important in the U.S. as well, the literature on waiting times in this country is quite sparse. I follow the literature by using the time elapsed between patient's arrival to an emergency department and her examination by a physician as a measure of waiting times. I follow closely a recent study on the subject (Wilper et al., 2008), introducing additional variables required for my investigation. I employ the dataset from National Hospital Ambulatory Medical Care Survey, administered by the U.S. National Centre for Health Statistics on 2006 emergency department admissions. It is the most recent dataset to be analysed, to the best of my knowledge. The data seem to support Hypothesis 4: *NPs have longer queues that FPs.* The data is weakly supportive of Hypothesis 5: *Higher quality NP hospitals have longer queues than lower quality NPs* as well.

The data is supportive of the implications of the theory proposed in this thesis. This implies that the new picture of healthcare markets where NP hospitals tend to occupy the middle of the distribution of quality and FPs are more concentrated on the tails has some merit. The new theory of NPs has a wider range of empirical implications than previous theories. Some of those implications are more consistent with the evidence than the prediction of previous theories. In particular, the quality implications of my theory are consistent with *all* of the existing evidence I am aware of, while previous theories with only a part of them. For this reason, I believe that the newly developed theory should be used as a tool to compare various policy alternatives. I examine policy alternatives and develop some recommendations in the next chapter.

Chapter 4

Policy Implications

4.1 Introduction

Since the mid-1980s a substantial number of non-profit (NP) healthcare providers in the U.S. have converted to for-profit (FP) status. This process sparked a debate as to whether these conversions should be allowed. A number of well-publicized scandals inflamed the debate. One of the early scandals came into light in 1997 when a large FP chain of hospitals, Columbia/HCA, was found to practise fraudulent insurance billing (see Cutler and Horwitz, 2000 for details). More recently, these headlines appeared in newspapers: "At Many Homes, More Profit and Less Nursing,"[118] which described in dramatic terms how a number of nursing homes in Florida decreased quality of care they offer to their patients after they were purchased by a group of investors, changing their status from NP to FP. Not only were regulatory violations observed, but a number of the elderly clientele died due to neglect. These events reinforced the side of the debate that opposed reliance on profit and conversion into FP organizational forms in healthcare. NPs are not immune to criticism, however. A Wall Street Journal report cited two nursing homes, one a NP and the other a FP, using anti-psychotic drugs to save on resources when caring for the elderly.[119]

In this chapter I review the relevant policy alternatives available to the various levels of government in the U.S. regarding organizational forms of hospitals, and offer insights about possible

[118]Duhigg, Charles (2007) "At Many Homes, More Profit and Less Nursing," New York Times, September 23, 2007.
[119]Lagnado, Lucette (2007) "Prescription Abuse Seen In U.S. Nursing Homes: Powerful Antipsychotics Used to Subdue Elderly; Huge Medicaid Expense," The Wall Street Journal, December 4, 2007.

outcomes of three sets of policies, based on the theory presented in Chapter 2 of this Thesis. The policies I review are intended to influence the provision of healthcare by altering the composition of the healthcare market in terms of the organizational form of providers. I focus on those policies that influence the determination of the scope of NPs versus FPs. One set of policies aims to discourage FP hospitals altogether; another set of policies strives to discourage conversions from NP to FP status; and a third set of policies discourages NP status.

The use of economic theories to inform policy making has, of course, a long tradition. Any useful economic theory that has the potential to offer insights into policy issues necessarily relies on simplifying assumptions. Due to these assumptions the predictions of these theories often appear too unrealistic for application to the complex issues of the real world.[120] However, the valuable insights the theories offer, specifically because of their simplicity, have proven to be of value in informing policy making. The theory developed in Chapter 2 offers predictions that are quite stark, due to its simplifying assumptions (such as only two extremes of information asymmetry: informed and uninformed patients). I use the broad insights of the model in order to analyse policy alternatives. I recognize that this analysis offers only one view on the complex set of issues that policymakers in healthcare must deal with.

The proposed theory of NPs offers a unique perspective: conversions of a particular subset of the FP hospitals to NP status may be viewed as beneficial. The theory supports a specific criterion by which one decides which hospitals should convert: that the quality of care offered by the hospital falls below the average quality that comparable NPs offer. The theory predicts a concentration of FP hospitals on the tails of the distribution of quality. Those FP hospitals located on the lower tail of the distribution are damaging to welfare. Encouraging their conversion to the NP organizational form should, therefore, induce an improvement in their quality of care.

To implement this policy recommendation one needs to be able to measure quality of care in

[120]See the introductory chapter in Friedman (1953) for a discussion of this issue.

hospitals. The empirical literature on hospitals suggests a number of ways to measure quality of care. I reviewed those measures in Chapter 3. The most common one in research and the most satisfying from both medical and economic perspective are post-admission mortality rates (usually for Heart Attacks - Acute Myocardial Infarction, see McClellan and Staiger, 2000), adjusted for patient characteristics. Mortality rates are an easily measurable occurance, moreover, they are already available for the majority of U.S. hospitals through the Hospital Compare website (`http://www.hospitalcompare.hhs.gov/`) of the U.S. Department of Health & Human Services. After a Heart Attack the patient is less likely to be able to select a better hospital. Hospital selection is important to consider because when worse cases are treated in better quality hospitals, the measure of mortality will be biased. This measure assumes that the quality of all treatments within a hospital is correlated to the hospital's ability to perform heart attacks; nevertheless, such assumption is acceptable among the medical profession, as mortality rates for Heart Attacks are widely used as a measure of quality.

The policy recommendation discussed may appear too particular and dependent on how well my theory fits the specific market under consideration. I therefore offer an alternative, although potentially less effective recommendation. Discouraging conversions from NP to FP status dominates a policy of no intervention, according to my theory. Additionally, economic analysis offers a warning: if a type of discouragement of FPs policy is to be instituted, one must ensure that other jurisdictions (e.g. other states in the U.S.) engage in similar measures, otherwise, policy interventions may lead to exodus of qualified medical staff. Physicians are already likely to be leaving Canada, where FP hospitals are mostly disallowed, for the U.S. (Barer and Webber, 1999; Dauphinee, 2005).

I start by reviewing the background on conversions. Then I review potential effects of conversions that have been suggested in the literature and gauge at a general level their possible impact on overall social welfare. After presenting this general background I turn to the evaluation of policy

alternatives through the lens of my new theory of NPs.

4.2 Background

The scope of the 1990s conversions varied substantially across sub-sectors of the U.S. healthcare. While conversions among all three organizational forms (government owned, NP and FP) were observed, an overall shift toward FP healthcare can be inferred. In Rehabilitation Hospitals sub-sector, for example, the percent of FP hospitals jumped from 15% in the mid 1980s to above 55% in the late 1990s. In HMOs the percent of FP hospitals increased from 35% to just below 65%. Not withstanding the alarming headlines referred to above, the nursing homes is the only sub-sector that experienced a reduction in FP penetration from: 75% to 65%. Within Acute Care Hospitals the increase was small: 16% to 17%. Although an increase in the percentage of FP facilities was evident in all sub-sectors, except for nursing homes, the total percentage of FPs in many sub-sectors remains small (late 1990s data): 19% (an increase from 2%) in Hospice Programmes, 5% (an increase from 1%) in Mental Health Centres, and 5% (increase from 4%) in Residences for ED Children (Schlesinger and Gray, 2006). Currently, 20% of all hospitals in the U.S. are FP (Sultz and Young, 2009 p. 76).

Conversions from government owned facilities to FP or NP status occurred as well. These conversions are from a public institution to a private one. The set of issues involved in these conversions is quite different from the conversions that occur within the private sector from NP to FP (Deber, 2002). Privatization of hospitals is an issue of major policy debates in countries where the healthcare is publicly funded and mostly publicly administered, however, the private versus public provision issue does not surface as often in policy debates in the U.S. . Within the U.S. the acute care government hospitals that are open to the population as a whole[121] are mostly hospitals

[121]In the U.S. there are systems of government healthcare provision. One system treats only armed forces personnel and their families, while the other caters to veterans. Theses systems own hospitals and clinics, and seem to be

of last resort that provide low quality of care to the poor. For this reason, following most of the U.S. policy debates, in the analysis presented in this chapter I focus only on conversions of hospitals within the private sector (between FP and NP organizational forms).

From the news reports it may appear that main issue one needs to consider when deciding on policy toward conversions is whether the new FP organizations are more likely to shirk on quality. This is, in fact, the focus of my analysis in this chapter. A number of authors, however, have suggested other economically relevant issues to consider. Even if we find that the conversions actually have no effect on quality, there may be other reasons to discourage such practices. For example, conversion of a hospital from NP to FP organizational form may reduce the amount of community benefits the hospitals provides. On the other hand, if there are substantial economic gains from conversions, such as an increase in tax revenue that can be used for more productive purposes, it may be efficient from the society point of view to tolerate a slight decrease in quality.

In the next two sections I review overall possible impacts of conversions on welfare. First I focus on the impact of conversions on productive efficiency of hospital care. Then I isolate those aspects of conversions that do not affect production efficiency and review them in detail. This is done in order to set the background to the main analysis of the chapter. The rest of the chapter then examines the impact of policies that influence the composition of healthcare markets in terms of the organizational form (NP and FP) of the providers on social welfare.

4.3 Conversions and the Productive Efficiency of Hospital Care

Before turning to the review of the various aspects of conversions that affect welfare, I examine conversions from pure productive efficiency prospective. It is well understood in economics that the

successful and even outperform the rest of the U.S. healthcare system (Solberg, 2009). These systems combine insurance with provision, and fall in entirely different markets than the private NP and FP hospitals addressed in the policy debates. Psychiatric hospitals in the U.S. tend to be publicly owned as well.

generation of optimal social welfare is based on maximizing the difference between resource costs in production and the valuation of the output by the society in consumption. Feasible production sets are based on the current technological knowledge. Some combinations of input-resources within these production sets have lower costs of production than others for the same quality and quantity of output. The cost of production are important to consumers, as they are the ones who ultimately bare these costs (either through taxation or prices). To the extent that production activities are conducted within institutions, the lowest-cost combinations of resources may not be achievable due to particular agency structures in these institutions that do not align consumer and employee interests.

One way in which changes in organizational form are able to influence production efficiency is through the agency relationship between the employees of an institution and its customers. If one organizational form has an advantage in structuring agency relationships in such a way that lower cost input-resource combinations are implemented in production, conversions will have an impact (either positive or negative) on the production efficiency and on the social welfare.

An important question for the analysis is whether FPs are inherently more or less efficient than NPs in production. The evidence is mixed. Often NPs dominate (Barr, 2007 p. 178), contrary to beliefs of some economists. Efficiency advantages to FPs could come, however, from their ability to reduce the power of the unions and lower the wages of the hospital staff (Deber, 2002). To the extent that wages are a transfer, one cannot point to them as waste. On the other hand, the healthcare cost containment effort will point to lower wages as a beneficial outcome. In order for a FP to operate in a market it must earn adequate return on the investments of its shareholders. If the entire gain from wage reduction is absorbed into the return on investment that goes to the shareholders, for the society as a whole there is no gain from the process of conversions (in fact there is a loss, as conversions often involve litigation, or at least supervision, which use resources). Only if the incentives that are built into the compensation contracts better align the interests of

the employees and the *consumers* (and not the investors) under the FP organizational form (e.g. share of employees in profits) is one able point to a production efficiency gain.

Existing economic theory (Glaeser and Shleifer, 2001; Hirth, 1999) predicts that FP organizations have a higher incentive to reduce the aspects of quality that are harder for the consumers to observe. This is because the investors, who are (represented by) the stakeholders of a FP organization, are able to gain from any such savings directly through dividends. On the other hand the stakeholders of NPs are overseen by the regulator to make sure that they do not appropriate *any* operational surpluses, under section 501(c)(3) of the U.S. Internal Revenue Code (Claxton et al., 1997). For this reason the incentives to save costs are stronger for the stakeholders of FPs relative to those of stakeholders of NPs.

Some have suggested that conversions may generate *potential* efficiency gains in production, but gains that are not realized in practice due to the incentive to shirk on the quality of output.[122] This incentive exists because of the agency conflict between the consumers and the investors. Some analysts argue that in cases where the potential for efficiency gains exists, but is not realized, the government may be able to mitigate the conflict of interests through regulation. In order to monitor the quality of output of healthcare organizations, however, one needs to be able to measure it. In practice, due to heterogeneity and complexity of outputs such measurement may not be possible for *each* output, therefore, the quality reduction is always a concern when one examines conversions from NP to FP organizational form.[123]

Conventional contract theory points to the value of incentive schemes for employees. One form of incentives is conditioning compensation on profit. Due to legal constraints, however, only a FP institution is able to offer a share of its profits to employees. This implies that FPs have higher

[122]This discussion follows Deber (2002).

[123]As I argue in this chapter, the overall quality of care at a hospital can be measured through objective outcome measures, such as mortality rates. The argument here, however, is that an oversight of *every* activity of an organizational by the government is not feasible. Even if all outputs could be measured the costs of such regulation are likely to be substantial.

ability to offer incentives to employees,[124] which may lead to higher efficiency in production. On the other hand, Francois and Vlassopoulos (2008) review literature which suggests that if employees are altruistic (care about output), FPs have a stronger incentive than NPs to substitute away from other costly inputs for every extra unit of employee effort. As a result, NPs have higher ability to commit not to substitute, which makes it cheaper to achieve the same level of incentives under the NP organizational form.

Even if conversions do not generate efficiency gains in production either through improvement in production technology (i.e. innovation), or through improvements in incentives, there may be other impacts of healthcare provider conversions that effect social welfare. If conversions involve a transfer from one economic agent to another, and there is a reason to believe that the receiving agent is able to use the re-allocated resources in a manner that benefits the society more (or less), there will be a potential impact of conversions on welfare. In the case of hospitals, such transfers are: the direct proceeds from conversions, future taxation, "side" community benefits the institution delivers (e.g. uncompensated care, eduction, research) and change in prices that consumers face (to the extent that the demand is inelastic, as when the quantity/quality of healthcare services demanded changes, there are additional welfare effects). I review the specific impacts on welfare of these transfers caused by conversions in the next section.

4.4 The Costs and Benefits of Conversions

4.4.1 Institution Level Effects

When one considers the impact of conversions from NP to FP organizational form, a number of factors must be taken into consideration.[125] There are institution-level costs and benefits of a

[124] Indeed FP hospitals tend to have higher-powered compensations compared to their NP counterparts (Erus and Weisbrod, 2003).

[125] The following discussion draws upon Deber (2002) and Claxton et al. (1997).

conversion and there are market-wide ones. The latter influence the social welfare through market interactions, as a result of the change in the characteristics of the players. The institution-level effects are the following. After the conversion the new institution may offer less community benefits (Claxton et al., 1997), however, the proceeds from the conversion will go to the community, and the new institution will pay taxes. Therefore, when one examines conversions one must add the proceeds and the future tax revenues as benefits and subtract the potential loss in community services as costs. In what follows I examine these three effects first, and then turn to the market-level effects.

Although it is clear that there are a small number of NP hospitals that provide substantial community benefits through uncompensated care, education of medical staff and research, the majority of NPs do not seem to provide benefits significantly in access of their FP counterparts (Claxton et al., 1997). What matters for this discussion, however, is the amount of benefits that the converting NPs provide, and it is unclear how likely the high-community-benefits NPs are to convert.

When evaluating the impact of conversions on welfare, it seems that some policymakers and analysts in the U.S. emphasize particularly the effects on the immediate locality of the converting institution rather than the overall impact at the state or the national level. For example, the money transfered from the new investors to the board of trustees representing the converting NP institution is frequently more important in policy debates than the future tax revenue of the new FP. NPs do not pay taxes under section 501(c)(3) of the U.S. Internal Revenue Code (Claxton et al., 1997), while FPs do. All of the U.S. citizens benefit from federal tax proceeds, and all of the state residents benefit from state taxes. Gentry and Penrod (2000) report that FP hospitals paid $976 million in income taxes in 1995, for example. Hence, tax revenues seem to be an important issue to consider, one that is often overlooked in policy debates.

Various policy examinations have emphasized the need to set up an accountable trust that will be used to benefit the immediate community (Claxton et al., 1997). Somewhat surprising was the

outcry against appropriation of the proceeds by the government of the state of New York in the conversion of the Empire Blue Cross (Robinson, 2003), presumably, as an alternative to putting the money in a private trust. It is unclear, however, whether these trusts, which normally administer the proceeds, generate substantial community benefits or not.

The stream of tax revenues from FP hospitals seems quite a significant figure to take into account in welfare calculations. It is common, however, to consider only the portion of the taxes that is kept within the community. This portion tends to be small. It seems that overall, the institution-level effects of conversion neither through the benefits (taxes and proceeds) nor through the costs (loss of community benefits) are able to tip the scales either in favour or against conversions from NP to FP organizational form. For this reason one must examine market-wide and long-run effects of conversions as well.

4.4.2 Market Wide Effects

The market-wide welfare impacts of conversions occur through market interactions. Conversions may cause changes in costs and demand faced by the remaining NP and government-owned competitors of the converting institution. The converting institutions, by altering their post-conversion behaviour may influence market-aggregate variables such as overall distribution of output by quality, price level for medical services, and overall access to healthcare services. There are both short- and long-run effects, I discuss the short-run effects first.

Beyond idiosyncratic reasons for conversions, such as bankruptcy, the need for more capital investment, or inadequate management (Cutler and Horwitz, 2000), the extension to the theory of NPs introduced in this thesis predicts two additional more systematic reasons for conversions. One is the opportunity to cater to consumers who are willing to pay for high quality of care, and are able to monitor it, and the other is a (negative) cost shock that is more prominent for institutions that deliver lower quality of care. If the conversion is due to the opportunity to serve higher-end

consumers who can monitor quality, the converting institution will not change its quality of care. However, the price the provider charges is likely to rise. The new organization is likely to choose to focus exclusively on providing high-quality services to its clientele, who are more able to monitor the quality of care they receive, and will not treat patients who are not willing (due to inability to assess), or unable, to pay for the high quality of care. There is no concern over quality reduction in this case.

If a conversion is driven by a decrease in costs, the converting organization is likely to shirk on quality once it assumes FP status. This is due to the fact that its clientele is likely to be composed mostly of those who are not able to monitor the quality they receive. This would imply low sensitivity of the fees the consumers are willing to pay (either directly or through insurance) to quality reductions. However, such quality reductions lower the costs of provision. The owners of the new FP will aim to design the employee compensations schemes after the conversion in such a way that the cost savings will be maximized. There is room for a concern over quality reduction if the reason for the conversion is purely cost-saving.

When a new FP enters the market, or an existing institution converts to a FP status, it may engage in client selection more than its predecessor or other NPs in the market (Deber, 2002). As a result, there may arise a change in the characteristics of the demand that the non-converting NPs face. This may cause strains on the system, especially if some of the activities are paid for by the government, as the case is with the care for Medicare and Medicaid patients. If reimbursement is not adjusted by severity, which determines the costs of a treatment, it could be the case that the FPs are treating simpler and cheaper cases, and earning high profits, while the NPs are faced with the severe and expensive cases. Moreover, if the reimbursement is set in such a way that the simpler cases are intended to cross-subsidize the more severe ones (as often is the case), the NPs will tend to go bankrupt, while the new FPs will earn considerable profits. Although behavioural responses on the side of the NPs are able to mitigate these forces, the strain on the system may

still remain. On the other hand, the mitigating behaviour of the remaining NPs (e.g aggressive marketing, client selection, etc.) reduces consumer welfare.[126]

Conversions usually imply changes to organizational structures. Some NPs are allowed to operate FP subsidiaries, in these cases the conversions are not complete and the organizational structure may not change. However, often, as the case was with the Empire Blue Cross, the whole institution assumes a FP status (Robinson, 2003). In these cases the converting institution may divest itself, or simply break previous contractual agreements that somewhat integrated it vertically or horizontally with other NPs. After the conversion the new FP may merge into an existing network of other FPs, as it was the case with purchases of hospitals by Columbia/HCA (Cutler and Horwitz, 2000). The changes in organizational structures cause shifts in cost structures across sectors. Following a conversion there may be a cost increase in the NP sector, as a result of a loss of synergy benefits (e.g. economies of scope/scale, easier access to inputs). On the other hand, there may be a cost reduction in the FP sector as a result of the new synergy benefits. The interaction with the demand and the type of patient selection across the sectors will determine the welfare implications. However, a very likely outcome is where the NPs are facing more expensive cases, while suffering increases in costs, and the FP sector enjoys both simpler cases and a reduction in costs due to synergies. This will cause price increases in the NP sector, and possibly a fall in the quality that the NP sector is able to deliver (Deber, 2002). Both of these are damaging to the consumer welfare.

In addition to the higher likelihood of the new FPs to use various marketing techniques to attract favourable customers, as discussed above, they also tend to put more resources into the maximiza-

[126] Aggressive marketing may not seem to be welfare damaging from first glance. However, aggressive marketing is costly, and while it delivers benefits to those who engage in it, it does not benefit the society as a whole, beyond the communication of existence of a service or good. In an institution that does not distribute any benefits to outside parties (a NP) these costs are ultimately borne by the consumers. Additionally, there are casual costs to the society from advertising, such as environmental damage through waste of paper, use of lethal paint on ads, or by detracting from the enjoyment of a TV programme.

tion of proceeds from insurance claims (sometimes even beyond the legal boundaries as in the Columbia/HCA case, see Cutler and Horwitz, 2000). The remaining NPs that are already strained due to loss of revenue from more profitable consumers and losses of costs savings of synergies, are induced to exercise more "efficient" insurance billing, often through increases in bureaucracy (Cutler and Horwitz, 2000; Deber, 2002). As claim payments are only transfers, they have no additional welfare effect once the service is delivered, therefore, larger bureaucracies are a pure waste to the society as a whole.[127] Additionally, higher transfers to providers imply rises in costs of healthcare, while policymakers are aiming to contain the already inflated costs of the U.S. healthcare (at least as it compares to the rest of the developed countries, see Garber and Skinner, 2008). This shows that short-run market-wide effects of conversions are likely to be welfare damaging.

4.4.3 Long Run Effects of Conversions

Beyond the short-run effects of conversions there may arise additional long-run effects. One of the main long-run effects is on the labour supply. It is believed that FPs are better at reducing the power of the unionized labour (Deber, 2002), and, thus, lowering the cost of the payroll. The reduction in the payroll comes through a number of avenues. One possible avenue is institution-level reductions in wages, or a freeze on wage raises. However, in addition to that, the new FP may substitute non-physician staff for less educated and lower ability personnel.

In the short-run, a wage reduction may have little effect on production, as long as the right incentives on the margin are kept in place. However, when lower skilled personnel is hired, certain aspects of quality that are of value to the customer are likely to be undermined. Moreover, as the benefits of certain medical professionals decrease, there will be less entry into these professions in the future, as entry costs (mainly education) remain high. This will lead in the long run to labour

[127] Increased insurance billing creates perverse ex-ante incentives, however. Insurers are more likely to deny coverage or increase fees if the claims increase substantially.

shortages, which cause both lower access to healthcare, and, in complete opposite of the initial intentions, to an increase in employee benefits, creating a cost-shock to the industry (Deber, 2002). It is interesting to note that the tragedy of the commons is relevant to this situation. The benefits from a payroll reduction and substitution toward less educated employees and, hence, cheaper labour, occur immediately to the new FP, while the costs of lower supply are paid collectively by all of the players in the future.

If one views an economic and political system as a system of bargaining units, conversions change the distribution of the bargaining power. Specifically, new players are born and endowed with bargaining power. These new FP players have an incentive to form coalitions and lobby the government for higher funding for government administered healthcare programmes[128] (Medicare and Medicaid in the U.S., and considerably more extensive programmes in the rest of the developed world). On the other hand, the same coalitions of FP organizations have stronger incentive to bargain with the insurers to increase pay outs. To the extent that parts if the insurance industry are NP (e.g. Blue Cross), these fee increases are very likely to be passed on to the consumers (as an increase in costs of an organization that is breaking-even has to lead higher fees or bankruptcy, thus the pass-through of costs should be higher for NPs relative to FPs). This both increases the costs of healthcare and reduces consumer welfare.

From the studies reviewed above it seems unlikely that conversions of NP healthcare institutions to FP status are beneficial to the society as whole. In the remaining of the chapter I use the insights of the theory proposed in this thesis to examine in detail the possible outcomes of policies that aim at managing the NP and FP organizational forms: discouragement of FP provision; the discouragement of conversions from NP to FP; and the reduction in benefits to NP status. The same general impact of conversion I reviewed above repeats itself, and any intervention that limits FP organizations seems likely to increase welfare. I offer additional important insights for policymakers

[128]This point was suggested by Deber (2002).

applicable to any policy decision that limits FPs, especially, if they are surrounded by jurisdictions that are not subject to the same legislative process as theirs.

4.5 Impact of Policies that Discourage FP Healthcare Provision

It is critical from a policy perspective to understand the welfare implications of policies that discourages hospital organization as FPs. The extension to the theory of NPs I introduced in this thesis offers some insights into this issue. A key policy contribution of the new theory is the result that there may exist FPs that do not shirk on quality they provide to consumers, even when there are information asymmetries between producers and consumers. Those FPs serve consumers who are better able to assess the quality of service they receive and are willing pay accordingly. This may suggest that there is no benefit from discouraging FP hospitals; however, it turns out that it may not be the case. Policies that encourage conversions to NP status only for low quality FP hospitals find a theoretical foundation in my model. I start by reviewing the general forces in the hospital markets that are relevant from theoretical perspective to potential outcomes of government interventions. Then I turn to the implications of the discouragement of the FP provision.

4.5.1 Relevant Forces in the Healthcare Markets

U.S. hospital markets are fundamentally mixed markets, where NPs operate alongside FPs. To predict the outcomes of policies in mixed markets one must examine the behaviour of both of the organizational forms and the interactions between them. Mainstream economic theory predicts provision of the optimal level of quality by FP firms to well-informed consumers, if the markets are competitive. Consumers who cannot assess well the quality of medical services they demand, however, are likely to encounter FP providers that shirk on the unobservable dimensions of quality. This happens simply because providers can save resources and lower costs, while the fees they are

able to charge are hardly affected by these practices (which patients are unable to detect). A FP provider is able to transfer these savings to employees as salary bonuses and to its investors as a return on their funds.

Economic theory of NPs predicts that hospitals should provide adequate quality to every patient that chooses to seek healthcare from them, even she if she is less informed. Although reductions in quality may not be felt immediately (or even in the future, in many cases) by the patients, gains from neglect will not benefit the personnel much. The U.S. government, in accordance with section 501(c)(3) of the U.S. Internal Revenue Code, monitors NPs to make sure that no party benefits from the income these facilities generate. Hence, if medical staff has even the slightest amount of altruism toward their patients (which they should, according to the professional ethics), they would provide services of adequate quality to all patients: both informed an uninformed.[129] For this reason, there should not be significant differences in quality between FP hospitals that serve better informed consumers and the NP ones.[130] If a patient is less informed, however, and hence, is more likely to receive higher quality treatment in the NP sector, why should she bother risking a bad outcome of treatment in the FP sector in the first place?

Patients who are less informed are more likely to have better outcomes of treatment in the NP sector relative to the FP sector. On the other hand, NP hospitals are forced by the U.S. Internal Revenue Code to charge fees that enable them only to break-even after new investments are accounted for. These two conditions imply that less informed patients are offered considerably better value in the NP sector than the option of treatment at a FP facility by staff that may

[129]Some policy analysts have raised the argument that NP employees may be prone to reduce effort due to low-powered compensation. Economic theory, including the one in this thesis, has focused on active distortions relative to the low-incentive ex-ante NP outcome. If one accepts that NPs inherently shirk on quality, the next question should be whether FPs shirk more. As all of existing theory points out, the answer is affirmative for the case of less informed consumers.

[130]Some hospitals may still be better than others. The quality differences, however, cannot be attributed to the organizational form, but rather to the skill of the hospital staff. In fact, the model of Chapter 2 predicts that FPs serving well-informed patients also deliver higher quality care than NPs.

neglect certain (unobservable) aspects of their duty, and charge fees that enable to earn adequate return on investment. Due to this reason a large number of patients would be willing to seek treatment in the NP sector. This implies that NP hospitals should be more congested than FP ones, which increases the time patients wait for procedures.[131] Due to the inconveniences associated with waiting and crowded hospitals, even those patients who are more vulnerable to the adverse practices of the FP institutions would rather try their luck in the FP sector than wait for certain procedures (government owned hospitals that are open to all patients are particularly small in number and under-funded in the U.S.).

The FP hospitals are more flexible in setting treatment fees than NPs. Due to this reason, it is not necessary to "ration" healthcare in the FP sector. If a facility is over-crowded it can raise the treatment fees, which will reduce the number of patients seeking treatment there. The ability to set fees high enough to attract only the number of patients that the facility is able to treat in reasonable amount of time creates a more attractive treatment experience in terms of observable characteristics in the FP sector for the informationally disadvantaged patients. Therefore, one is able to find patients who are less able to assess quality of care in hospitals of both organizational forms. After reviewing the theoretically relevant aspects of the hospital markets, I turn to outcomes of the potential government intervention.

4.5.2 Implications of Discouragement of the FP Provision

If FP provision of healthcare is discouraged, some hospitals are likely to convert from FP to NP status. The new NPs would adhere to the non-distribution constraint. This feature will reduce the incentives of the medical staff at the converting facilities to lower the less observable aspects of quality of treatments, regardless of whether this was the common practice before the conversion

[131] Section 3.6 of Chapter 3 and existing literature (Wilper et al., 2008) show evidence of longer waiting times at NP hospitals relative to FPs.

or not (recall, that there are likely to be FP facilities that do not shirk on quality). From this perspective a policy that discourages hospital organization as FPs is likely to be beneficial to the society. However, the benefit is not as great as the previous economic theories predicted, as some FP facilities provided the best available treatment even before the policy was enacted, due to the informed nature of their patients. Therefore, the benefits will be limited only to a subset of hospitals that change their organizational form as a result of the policy.

It is important to determine the impact on access to care after of a policy that discourages the FP organizational form. If one accepts that in the NP sector access to care is equitable, at least for those who are not savvy enough to manipulate the system in their favour, equity of access to healthcare is likely to rise when the government discourages FP provision. If access to care is proxied through medical staff to patients ratio, after the policy takes effect, this ratio will rise in the NP sector. NP facilities are more crowded than FP ones (this is the reason uniformed patients turn to the FP sector). When FP facilities join the NP sector, although they bring new patients to this sector, they bring even more medical personnel. The length of waiting times should decline in the NP sector after conversions of FP hospitals. The increase in treatment capacity will come mostly from the lower quality converting FPs, however. The higher end converting FP hospitals are likely to retain their clientele, if well informed patients are able to manipulate the NP system to enjoy better access to care.[132] In emergencies, however, the highly skilled medical personnel previously employed in exclusive high quality FP practices is likely to be available for the treatment of less privileged patients as well. Both general improvement in access to care in the sector where it is rationed through waiting lists (NP), and possible access to higher quality medical staff in emergencies for a larger group of patients, are beneficial to the society.

[132]Barr (2007) (p. 98) cites evidence that more privileged patients have better access to care (at least in terms of outcomes) even in systems that should have equal access to all.

4.5.3 Main Policy Recommendation

As I showed above, from the perspective of quality of treatment and access to care, policies that discourage FP organization of hospitals should be beneficial to the society as a whole. However, one has to ask who would be most likely to oppose to this policy. It is particularly important to determine who would do so due to the fact that she is merely on the losing side, versus those who are able to bring up fairness arguments. The new theory clearly shows that all those medical professionals who were originally employed in FP hospitals would tend to oppose such a policy, both the high quality end and the low quality end. Those who were in the high quality end would have easier time arguing their cases, however, as explained in the next paragraph.

There is evidence of easier access of those more savvy about the inner working of the medical system to better doctors even in systems that should have equal access to all (Barr, 2007 p. 98). Hence, it is very likely that those privileged patients who were treated at high quality FP facilities before the policy was enacted will remain with their providers. However, under the new arrangement these patients pay lower fees, while the medical stuff of converted hospitals is receiving lower compensation. Therefore, the medical staff is able to argue that privileged patients, who are willing to pay more for high quality care they receive, gain from the policy at the expense of medical personnel. Policymakers and other advocates will be more inclined to see this blow to the income of the medical personnel as unjust.

The observations of the last paragraph point to the fact that a less extreme policy that is not targeted at all FP hospitals may still generate considerable welfare gains, while having a broader support of the public. Such policy is encouraging only those FP hospitals performing below certain standard on the scale of outcome measures of quality (e.g. mortality rates) to convert to NP status. The threshold standard for policy targeting should be close to the performance of a comparable NP, and above the malpractice level. Under such a policy the targeted hospitals are more likely

to be lower-end FPs catering to less informed patients. Those hospitals are particularly prone to shirk on quality. Therefore, the gains to the less medically savvy patients should be preserved at large, without denying some of the medical personnel income at the expense of more privileged patients.[133]

To summarize, the efficiency of healthcare provision in terms of quality is likely to increase if FP provision of healthcare is discouraged. The new theory suggests that this increase will not be as great as previously thought, as there exist higher end quality FPs that do not shirk on quality. Patients who are less savvy about the inner workings of the medical system, and cannot assess the quality of care they receive will be undoubtfully the main beneficiaries: due to higher quality (if previously treated at a FP facility), and shorter waiting times (if previously treated at a NP). Those patients who were treated at high quality FPs (and paid higher fees, or insurance premiums) before the policy was enacted will benefit mainly through fee reductions. As the medical personnel of FPs is the main loser from the change, it would be easier for them to object such policy focusing on benefits to more privileged consumers at their expense. The arguments of high end FP providers have the most weight since the quality of care at their facilities was at its highest before the policy was introduced. A viable solution is encouraging only FP hospitals that perform below comparable NPs on measures of outcomes care to convert to NP status, without penalizing the high quality facilities for their FP status.

[133]Some may argue that a policy that encourages low quality FP hospitals to convert to NP status offers a reward for poor hospital management. If after the enactment of such policy one observes higher tendency of hospitals to fall in the quality dimensions, it may be beneficial to consider a more authoritative policy that *forces* such conversions, while penalizing the management of the converting institution in the process.

4.6 Impact of Policies that Discourage Conversions from NP to FP Status

There are two additional sets of policy alternatives in addition to the one discussed in the previous section. One is a less extreme case of the alternative already discussed: discouragement of only future conversions to FPs. The other set of alternatives is: reduction in the benefits of the NP status for hospitals. Overall, the impact of a policy that discourages only conversions is of same direction, but lower degree relative to the impact of policies that discourage FP provision altogether. However, the comparison of no intervention versus a policy that discourages conversions is somewhat more interesting, since one has to address the question "why do conversions occur?" In this section I review the possible outcomes of a policy that discourages conversions from NP to FP status, and leave the next section for a discussion of policies that reduce the benefits of the NP organizational form.

The new theory presented in this thesis points to two systematic reasons (as oppose to idiosyncratic reasons, where a NP hospital has a bad draw of income/management, see Cutler and Horwitz, 2000) for conversions from NP to FP status. The first reason is an increase in the amount of patients in the economy who are able to evaluate the quality of healthcare services they receive. The other is a reduction in costs that and is experienced disproportionately more by lower quality hospitals. First I discuss the reasons for conversions in detail. Then, I conclude that the second reason for conversions can potentially justify discouragement of such practices.

The first systematic reason for conversions from NP to FP organizational form is an increase in the amount of patients in the economy who are able to evaluate the quality of healthcare services they receive. When more patients are becoming better able to evaluate quality of healthcare, more and more highly qualified medical staff would try to take advantage of being able to charge higher fees for higher quality of care under FP status, where they enjoy higher flexibility in setting the

fees. These conversions may be seen as damaging to welfare from the point of view of quality provision only if these medical personnel are leaving larger NP facilities, where they were available for emergency treatments to all patients, as the FPs would cater only to the more privileged clientele.[134] Additionally, unless there is some limit on the fees those high quality FPs are able to charge, there is no reason to expect active case-selection by FP providers, as more expensive cases can be priced accordingly, keeping them within the FP sector. Only in the case that patients are unwilling or unable to pay the higher fees in order to reduce waiting times for procedures (that are longer in the NP sector), the NP providers will face more expensive case mix than FPs. Therefore, the benefit from discouraging conversions of hospitals that convert due to their effort to cater to better informed patients may be small.

One may wonder whether a reduction in equity of access to care can be used as an argument to discourage conversions that occur due to increased number of patients who are able to evaluate the quality of care. It is likely that the actual clientele of highly skilled physicians and nurses, who push for conversions of the facilities where they are employed, may not change after the conversion. Conversions do not occur fast. They are likely to be slow responses to changes in the composition of the patient pool in terms of ability to assess quality of care. First the clientele is largely established and only then the conversion occurs. Therefore, within a number of years preceding the conversions, the better informed patients already had access to this highly skilled personnel more exclusively than their less informed peers, even within the NP system. Even if there is little incentive for their providers to convert (or switch to the FP sector in other ways) due to policy that discourages such conversions, the access of less informed patients to these NP providers is likely to be limited any way. Regardless of whether the conversion occurs or not, the damage in access to care still occurs. Therefore, potential gains in terms of access to care, from preventing conversion that are driven by

[134]It is possible that the growth of high quality speciality hospitals in the 1990s (see Section 1.2 in Chapter 1) was due to better access to information on the quality of care by consumers.

the ability to cater to better informed patients, are limited.

The second systematic reason for conversions from NP to FP organizational form, identified by the proposed theory, is a reduction in costs that is experienced disproportionately more by lower quality hospitals. This could happen as a result of a fall in the wages of less qualified medical staff.[135] Lower quality NP hospitals may find it beneficial to convert to FP status and reduce the quality they offer, as these hospitals cater to less informed patients, both before and after the conversion. The reduction in quality due to conversions stimulated by a fall in costs reduces the overall welfare.

I now examine the impact of conversions due to the second reason on access to care. The new FP facility will charge fees that reduce the demand for its services to the point that it is able to satisfy all of it. The number of patients that the converting facility transfers with it to the FP sector constitutes only a fraction of the number of those who were originally treated there. The ratio of medical personnel to patients in the NP sector decreases. Conversions that are driven by reduced cost of operating lower quality hospitals increase waiting times in the NP sector. The reduction in access to care in the NP sector implies reduced equity of access to healthcare for the underprivileged information-wise patients. Therefore, discouragement of such activities is likely to have important benefits to the society.

There are two systematic reasons for conversions, according to my model. The first reason for conversions, to cater to better informed patients, has little negative welfare impact in terms of access to care or quality provision.[136] The second reason for conversions, due to a reduction in

[135]Reduced power of labour unions may be responsible for such a scenario. It does not need to be the case that the wages fall only in the low quality FP segment. If the wages fall for NP sector as well, the conversions would still be profitable. This is due to the fact that, ceteris paribus, low quality FPs are more profitable. The fees they charge depend only on patients' *expectation* of (or average) quality (the idea is similar to Akerlof, 1970, market for "lemons"), which is unlikely to change much, if the number of converting facilities is small relative to the whole low quality FP segment. The costs of provision are now lower, however, which implies higher (potential) profitability of the new FPs.

[136]Conversions themselves, if occur, transfer wealth from privileged patients to their treating medical staff, as discussed in Section 4.5.

costs, is purely welfare damaging both in terms of access and quality. In practise, however, it may be hard to distinguish between the two motives for conversions; especially, since there is a high prevalence of conversions due to idiosyncratic reasons, such as access to investment funds (Cutler and Horwitz, 2000) as well. Therefore, discouraging all conversions from NP to FP organizational form may be beneficial to the society.

4.7 Impact of Policies that Reduce the Benefits of NP Status

It is only fair to review a final set of policy alternatives that some advocates are pushing for.[137] I address the impact of policies that cause a reduction in the benefits of the NP status for hospitals in this section. It is not surprising, in light of the findings above, that there is are potential welfare damaging effects from discouragement of the NP status. In fact, these effects are potentially far reaching from the perspective of access to care, quality, and distribution of wealth. Therefore it may be worthwhile to review them in detail, in order to warn policymakers about possible adverse effects of policies that discourage NP healthcare provision.

A majority of patients that are treated by NP hospitals tend to be less informed (Hirth, 1999 cites evidence for this from nursing homes, however, it should be applicable to hospitals as well). FPs tend to shirk on the quality they provide to less informed patients. Therefore, to the extent that hospitals retain their clientele after conversions from NP to FP status, due to discouragement of the NP form, the converting hospitals will reduce their quality of care. NP hospitals constitute 80% of the total number of hospitals in the U.S. (Sultz and Young, 2009 p. 76), therefore, depending on the effectiveness of the redcution in the benefits of NPs, a substantial number of U.S. hospitals is likely to reduce their quality of care.

However, on top of the quality reduction, explained in the last paragraph, a substantial shift of

[137]The fact that this issue has been discussed is reported by Gentry and Penrod (2000).

wealth from patients to the medical personnel and reduction in equity of access to care are likely to happen as well. Both the low-end and the high-end pre-existing FPs will raise their fees. Before policies that discourage NPs were implemented, the patients treated by both low and high quality FPs had the option to turn to the NP sector. Once the NP sector shrinks (or even disappears), as a result of the policy intervention, the only alternative to patients' current providers will become the no treatment option. Therefore, after the abolition of the NP status all patients (both better informed and less informed) will become "locked-in" by the providers, in such a way that most providers are likely to be able to raise their fees without losing patients that can be treated within a reasonable period of time.

Ironically, the waiting times should decrease dramatically after an abolition of the NP status. However, this will be due to higher fees and not more treatments. Many less-informed patients will voluntarily forgo treatments, as the fees will be simply too high. It turns out that the well informed patients are lucrative enough for the medical staff to make sure they still choose to be treated, however, a considerable appropriation of their wealth through fees is likely to happen. The fact that waiting times are likely to fall can be used as an argument in favour of a reduction in the benefits of the NP status. It is important to understand, that this measurable outcome may imply a severe welfare damage.[138]

I have reviewed three sets of policy alternatives that impact the composition of the healthcare markets in terms of the organizational forms. Policies that discourage FP healthcare provision, policies that discourage conversions from NP to FP organizational form, and policies that reduce the benefits of NPs (or encourage conversions to FPs) were reviewed. There are potentially considerable welfare damages to the patients from policies that discourage NP healthcare provision by reducing the benefits to the NP status. Although the last set of policies is likely to be highly lucrative to the

[138]The insight of the theory can be extended beyond the current context. The model developed in this thesis sheds light on the fact that a reduction in waiting times at hospitals may imply a reduction in demand due to higher prices, and a significant welfare deterioration, rather than an improvement in healthcare provision.

medical personnel, it is welfare damaging overall, according to the theory developed in this thesis.

Discouragement of FP healthcare provision is predicted to generate the highest benefits to the society as a whole. Discouraging conversions from NP to FP is likely to prevents only future damages. A more lenient set of policies that encourage conversions from FP to NP status for hospitals that are below a certain standard of quality is likely to generate less opposition than the discouragement of FPs altogether. My theory favours the latter set of policies that encourages conversion of low quality hospitals, as the inferiority to the best set of policies of discouraging all FPs, is relatively low welfare-wise, while the chances of gaining public support seem to be the highest.

4.8 The Case of Non-Universal Policies and Implications for Canadian Healthcare

4.8.1 Impact of Non-Universal Policies

The preceding analysis of this chapter reviews the possible effects on welfare of three sets of policies that have the potential to impact the composition of the healthcare market in terms of organizational forms in a world where only one jurisdiction exists. In other words, the policy maker is able to affect all the universe of hospitals that are relevant employers and treatment facilities for the population in her jurisdiction. However, more often policymakers live in a world where there exist other jurisdictions in close proximity over which they have no control. And it happens that a policy change in one jurisdiction affects the welfare of its neighbours and vice versa. One example is state level policymaking in the U.S., as many of the neighbours of the state in question are potentially affected due to movement of resources, medical personnel and patients between the neighbouring states. Additionally, Canada is directly affected by healthcare policies in the U.S., as the moving

costs for physicians (and other medical personnel) from Canada to the U.S. are quite manageable, while high rewards in terms of wage gains are awaiting on the other side of the border. In this subsection I examine the outcome of the above policy alternatives already discussed in a world where the medical personnel can escape regulation by moving to another jurisdiction. In the next subsection I address the implication for Canadian healthcare.

Consider a case in which bordering jurisdictions are comparable in terms of costs of operating hospitals and the salaries of medical staff. One of the jurisdictions passes a legislation that discourages FP organizational form for hospitals believing that it will increase the welfare of its patients and economic efficiency of healthcare delivery, as was shown in the previous sections. No other jurisdiction follows suit. What is likely to happen next? Medical staff that is qualified enough to attract better informed consumers in a neighbouring jurisdiction would prefer to move to that jurisdiction. In the new jurisdiction, the medical personnel would be able to join a FP hospital that will charge patients higher fees for higher quality services. The alternative for this highly qualified physician or nurse is switching to a NP hospital in her home jurisdiction, where the personnel is able to enjoy only break-even fees. Remaining at a FP and carrying the burden of the discouraging policy, is the third, probably, least attractive option. Therefore, the first consequence of a discouragement of FP healthcare provision is exodus of highly skilled medical personnel from the jurisdiction with the discriminatory policy toward FPs.

Are there any other potential adverse effects on welfare in the jurisdiction which initiated a legislation that discourages FP hospitals? The answer is affirmative again, specifically, due to the exodus of less qualified medical staff. Although loss of low-skilled physicians may not seem damaging at first glance, it creates shortages of medical staff and lowers access to care in the jurisdiction that discourages FPs. The reason for the exodus of the less qualified medical staff is the fact that fees in the lower-end FP sector, which caters to patients that are less able to assess the quality of care, reflect the average quality of similar hospitals rather than the specific quality

of care of any one institution (the idea is similar to Akerlof, 1970 market for "lemons"). Medical staff who are able to receive salaries that reflect break-even fees of NPs in the jurisdiction that discourages FPs may be able to benefit from providing services with quality very much below the price that patients pay (that is affected by the level of care that the patients *expect* to be receiving) in the neighbouring jurisdiction. Therefore, the exodus of not only the highly skilled, but of low skilled physicians from a jurisdiction that discourages FP healthcare delivery is quite likely.

There are two avenues for welfare reduction when a jurisdiction enacts policies that discourage FP provision, if the jurisdiction is located close enough to another jurisdiction, where FPs are not penalized by legislation. Both the highly skilled physicians and nurses are likely to leave for the jurisdiction without the discriminating policy, and the very low skilled ones will be tempted to migrate as well. For this reason, the proposal to encourage conversions from FP to NP only for low quality FPs has the potential to reduce the impact of exodus of physicians, and is likely to restrict it to those of lower qualifications.

4.8.2 Implications for Canadian Healthcare and Local Inter-Jurisdiction Relations

In Canada, although some FP hospitals exist, most hospitals are NPs. These hospitals negotiate global budgets with either local health authorities or provincial governments (depending on the region). Physicians who perform procedures in hospitals are compensated on fee-for-service (FFS) basis, and the rest of the medical staff is salary based. The salaries and the level of FFS are negotiated by representatives of the profession with hospital management, local health authority, or provincial government, depending on the circumstances. Overall, the system operates as if all of the facilities are NP, as the government reimburses always at the pre-negotiated (break-even) rate. Only a small number of FPs charge patients extra fees. Therefore, movement of physicians across

the border for higher compensations, or less work loads at FP facilities in the U.S. is quite likely.[139]

Discouraging conversions of hospitals from NP to FP organizational forms is the case of a less extreme intervention than encouraging the conversion of all existing FPs to NP status. Therefore, the adverse effects of a discouragement of conversions in the world where there exist neighbouring jurisdictions is less severe as well relative to a full or partial discouragement of FP facilities. Any movement of medical personnel between the jurisdiction as a result of a policy that discourages conversions occurs only if both jurisdiction experience same shocks to costs of healthcare provision, or an increase in the number of better informed patients (the two systematic drivers of conversions, identified in this thesis theoretically).[140] Interestingly enough, due to policy formulations, even if both jurisdictions discourage conversions, however, neither discourages the establishment of new FPs, or expansion of the capacity of existing ones, there may be exchanges of medical personnel and resources, such as equipment, between the jurisdictions. Those who wish to convert some services within a hospital, or the whole hospital to FP status, can move to the neighbouring jurisdiction and open a new FP, or join an existing one. These exchanges are likely to be asymmetric and more damaging to the more restrictive jurisdiction.

In summary, an overall discouragement of FP healthcare provision will be damaging to a jurisdiction which has among its neighbours jurisdictions where FPs are not disadvantaged through legislation. Both the highly skilled medical professionals and the lower skilled ones will move to a jurisdiction where they can work at FPs. A partial discouragement of only the low-quality FPs will induce migration only by the lower quality medical personnel, which is less damaging, although

[139]Barer and Webber (1999) report: "There is a general understanding that most of Canada's emigration of physicians is to the United States, although no reliable sources of data on destinations of departing physicians exist."; Dauphinee (2005) reports that in 1996 "726 [physicians] moved abroad [mainly to the U.S.] while 218 returned" to Canada "for a net loss of 508".

[140]If the shocks that are responsible for systematic reasons for conversions happen independently across jurisdictions, the anti-conversion policy will not induce any extra exodus of medical professionals from the jurisdiction that implements such a policy: medical staff will migrate to the jurisdiction(s) where FP become more lucrative irrespective of the policy.

does reduce assess to healthcare in the discriminating jurisdiction. If cost shocks to hospitals and increases in numbers of better informed patients are correlated across jurisdiction even a discouragement of conversions will induce migration of those willing to convert their institution/service. Paradoxically, if both jurisdictions discourage conversions but not expansion of FPs by other means, exchange of medical staff between the two are likely happen.

The lessons from this to the U.S. is that individual states must cooperate if they are willing to introduce policies that discourage FP healthcare provision. Additionally, there is an advantage from cheating and introducing discriminatory policies toward FPs a few years after the partner state. This has to be taken into account in negotiations between states. For Canada, which may be losing both high and low quality physicians and nurses to the U.S., unfortunately, the only cure to the loss of medical personnel may be either raising wages of healthcare professionals or allowing FP of provision of high-end healthcare in Canada,[141] if the U.S. does not change its policy toward FP hospitals. Allowing FP provision of healthcare in Canada will enable the Canadian government retain physicians within the country more cheaply, however, it will work up to the extent that Canadians are willing to pay no less than the U.S. citizens for healthcare services. If the U.S. citizens are able, or willing to pay more, some extent of exodus of physicians will remain without further increases in government funding.

4.9 Conclusion

Since the late 1980s a wave of conversions of government and NP hospitals to FP status has taken place in the U.S. . There is an on-going concern about the quality and raging costs of healthcare in this country. The wave of conversions has sparked a heated debate on whether conversions

[141] The analysis of this chapter points to substantial welfare losses from allowing low quality FP healthcare provision. On the other hand, it is likely that much fewer low skilled medical personnel migrates in order to work at low quality U.S. facilities.

should be allowed. This requires a careful analysis of one organizational form versus another. The question of conversions is a dual question in nature. One must first decide between public versus private provision of healthcare, and then within the private provision between FP and NP organizational forms. Public ownership in the U.S. general (not specialized networks) hospitals seems to be failing.[142] For this reason I focus on the policies that impact the composition of the hospital markets in terms of FP and NP organizational forms.

There are many aspects from which one organizational form may dominate another. There are also various costs and benefits to conversions, such as tax revenues from FP facilities, proceeds from conversions themselves and losses of community benefits that FPs tend to provide (on average) less than NPs. There is a debate over the relative efficiencies in operations of the two organizational forms, however, there is little empirical evidence to support any difference. Additional aspect that is needed to be considered is the provision of quality. My work in this thesis falls in the latter category. In this chapter I examine three sets of policy options: discouragement of FP healthcare provision, policies that discourage conversions from NP to FP, and policies that reduce the benefits of the NP organizational form for hospitals. I offer possible outcomes of each of these sets policies, as viewed through the lens of the theory introduced in this thesis.

In a world of a nation-wide decision making, discouraging FPs is the best option, as measured by overall social welfare. This option ensures high quality services for all patients. However, the medical profession may be opposed to it. Therefore, I propose a compromise of encouraging conversions from FP to NP status of hospitals whose performance falls below comparable NPs, as measured through outcomes such as mortality rates for a severe enough diagnosis (such as Heart Attacks) adjusted for patients' characteristics (common measures of quality are readily available through the U.S. Department of Health & Human Services, Hospital Compare).

[142]Needleman, Chollet, and Lamphere (1997) report that "Public hospitals were more likely than either nonprofit or for-profit hospitals to convert."

The other two sets of policy alternatives are likely have similar direction of welfare consequences to an overall discouragement of FP organizational form. Discouragement of conversions from NP to FP is welfare enhancing, but to lesser extent than policies that discourage FPs altogether. It is important to warn the policymakers against discouragement of the NP status (e.g. through reduction in its benefits). Such policy is damaging to patients; however, it may appear as desirable due to a fall in average waiting times for hospital procedures it is likely to generate. One has to remember that this reduction in waiting times is most likely to come through fee rises and patients opting not to be treated rather than real increases in treatment capacities, or more efficient technologies. Additionally, discouragement of the NP status may be pushed forward by the medical profession, as it generates considerable gains to the latter.

In a world where one jurisdiction borders another, a policy that discriminates against FP healthcare providers in one is likely to cause an exodus of medical professionals from the discriminating jurisdiction to its neighbours. Worse yet, the most qualified of the healthcare professionals in the discriminating jurisdictions are very likely to be the first ones to leave. The exodus of medical staff has detrimental effects on the welfare of the patients through access to care in the jurisdiction that initiated the legislation. Therefore, if a U.S. state wishes to pass a policy that discourages FPs, it must coordinate similar policies with its neighbours. Additionally, policymakers have to be aware of the incentives to delay the policy implementation by their neighbours, as in the meantime the neighbouring states are likely gain new highly skilled medical personnel.

The analysis of a world with more than one jurisdiction has implication for the Canadian healthcare as well. There is very likely to be quite an alarming volume of physicians, most of whom are highly skilled, who are leaving Canada for the U.S. (Barer and Webber, 1999; Dauphinee, 2005). If Canada wishes to retain those physicians, the Canadian policymakers should consider increases in compensations to the medical staff, or allowing high-end, and only high-end, FP healthcare provision in this country. This may be the sad reality until the U.S. starts introducing policies

that have the potential to discourage FP healthcare provision within its borders. These policies are likely to mitigate somewhat the "brain drain" within the medical profession. Discouraging FP provision, however, is likely to be beneficial for the U.S. healthcare as well, as the analysis in this thesis shows.

Chapter 5

Summary

5.1 Thesis Overview

In this thesis I examine the effect of organizational form on the provision of quality in the healthcare market. Private firms in the U.S. healthcare markets organize in one of two forms: For-Profit (FP) or Non-Profit (NP). Although there exist some contrary empirical evidence (Rosenau and Linder, 2003; Schlesinger and Gray, 2006), the notion that NP firms provide higher quality of output has been a popular folklore both in healthcare and in general economic theory of NPs (Hirth, 1999; Glaeser and Shleifer, 2001). I develop a model of a mixed-market, where FP firms (are predicted to) coexist with NP ones. This model is consistent with all the previous empirical evidence, that was previously thought of as contradictory. The model, however, also generates 5 testable hypotheses that, unlike previous theories, extend beyond the quality dimension. These are:

1. FP hospitals have higher variance of quality than NP hospitals.
2. An increase in the proportion of informed patients increases the mean FP quality for small proportions of informed, and decreases the mean FP quality for large proportions.
3. An increase in the proportion of informed patients reduces the variance of NP quality.
4. NPs have longer queues that FPs.
5. Higher quality NP hospitals have longer queues than lower quality NPs.

The data I analyse in this thesis are consistent with all 5 hypotheses.

The empirical research finds all three types of relationships between the relative quality of organizational forms, depending on the dataset analysed: NPs dominate on average; there is no difference in quality between FPs and NPs; and FPs dominate on average. My model suggests that there likely to exist both high and low quality FPs, if some of the consumers in the market can assess the quality of care they receive, while others cannot. The high quality FPs cater to the better informed consumers, while the low quality ones shirk on the quality they provide to the less informed ones. In the data, one always finds a distribution of firms by quality. To replicate this fact in the model, I assume the modelled doctors to vary in their costs of quality provision, which I interpret as ability. The model predicts FP firms to be concentrated on the tails of the distribution of quality, while NP ones in the middle, due endogenous selection of doctors into organizational forms by ability. Depending on the relative sizes of the pools of high quality FPs, NPs, and low quality FPs, when one compares the mean quality between organizational forms (pooling together *all* FPs), either organizational form may dominate. This feature explains the mixed empirical results.

My model is the first to have the players self-select between the NP and FP sectors based on ability. Doctors-entrepreneurs choose to open NP practices because of a slight degree of care about the welfare of the patient they treat. If a doctor is required to decide how to divide a dollar of total surplus from treatment between herself and her patient, she would always prefer to keep the entire dollar. This degree of altruisms is equal across all doctors in the economy. Contrary to previous literature, NPs do not simply attract the most altruistic individuals in the economy. NPs attract only the middle range of abilities in the population of doctors, due to the existence of FP contracting options. This feature implies that the main factor that explains differences in quality between the organizational forms is differences in ability and not differences in altruism.

The reason for the existence of NPs is that altruistic doctors of high enough ability choose to provide a higher than the average level of quality in the market, even when they shirk on quality.. The treatment fees in the low quality FP segment of the market, where less informed patients are treated, depend on consumers' expectation of quality, rather than actual quality each consumer receives. Although higher ability doctors provide higher quality, they receive the same level of compensation as those who provide significantly lower than the expected (average) quality. As the chosen (shirking) level of quality increases, the costs of its provision increase as well, and the non-altruistic part of doctor's utility falls. At a certain point, doctors who would choose to provide even higher (shirking) quality, find it more beneficial to open a NP.

NPs are characterized by a requirement to break-even. This requirement is responsible for two features that make the NP organizational form preferable to the FP one for higher skilled doctors who treat less informed patients. First, patients are able to infer that higher fee implies higher quality and (without wealth constraints) are willing to pay for quality improvements. Second, the break-even structure implies that doctors have to focus on maximization of their altruistic payoff, which is perfectly in line with consumer preferences. This payoff compensates for the loss of the advantages to shirking in FPs. I also show that even without perfect oversight of the regulator that all NPS break-even, the main results of the analysis still hold.

An extra feature that my model adds to the existing literature is the existence of patients who can assess well the quality of care they receive. All doctors would find it more lucrative to treat these patients under FP rather than NP organizational form, where there is higher flexibility in setting the fees. Patients are willing to pay for any improvement in quality, which prevents shirking even under FP organizational form. Without wealth constraints and when all patients have the same preferences, the informed ones are able to attract only the best doctors in the economy. This last feature of the model allows to reconcile the mixed empirical evidence. FPs that treat informed patients provide the highest quality of care in the market. As the size of the high quality FP

segment grows, the average FP quality rises as well (as long as the high quality FP segment is small enough). The existence of high quality FPs generates the central empirical implication of the model that provides a test for the explanation I propose. FP firms are predicted to have higher variance of quality than NP ones.

I perform empirical analysis to test the implications of the newly developed model. The central innovation of the empirical analysis in this thesis is that it analyses directly differences in moments of the quality distribution between the NP and FP sectors. The unit of analysis in the investigations intended to test the predictions related to the quality dimension are market-level moments of quality, contrary to the previous literature where individual hospitals are the unit of analysis. The central empirical implication is supported by the data I analyse. After a manipulation of the data, intended to reduce the impact of possible biases that the measure of quality for small hospitals introduces to the variance of quality at the market level, I find that FP hospitals tend to have higher variance of quality than NP ones.

I use mortality rates for Heart Attacks during 30 days after admission, already adjusted for patient characteristics, as the measure of quality in the empirical investigations. Higher mortality implies lower quality of hospital. I test the second implication of the model, that an increase in the proportion of informed patients increases the mean FP quality for small proportions of informed, and decreases the mean FP quality for large proportions. Although the prediction is non-monotonic, this hypothesis is supported rather strongly by the data, even without the adjustment that improves the precision of the inference both in this test and others. A third implication related to the variance of NP quality is supported by the data as well, however, rather weakly.

Although the main focus of this thesis is on the impact of organizational form on the provision of quality, the model developed for the purposes of the examination of this main effect is rich enough to generate additional predictions. The model predicts that in markets where there is limited capacity for production of output, or treatment of patients, NP firms will ration the access to their services

more than FPs. Effectively, longer waiting times are expected at NP hospitals than at FP ones. I use data on waiting time length in the U.S. emergency departments as a measure of waiting times for procedures (the degree of rationing). NP hospitals are found to have longer waiting times than FP ones, consistent with the theory. Additional hypothesis that higher quality NPs have longer waiting times than low quality ones is supported by the data somewhat weakly.

In summary, the theory proposed in this thesis offers an explanation of previous empirical results, some of which were not in line with both the healthcare folklore and existing economic theory. The model itself contains a number of theoretical innovations. The predictions of the model, however, extend beyond the quality dimension, for which it was intended. I test the quality related predictions on an original dataset employing yet another methodological innovation related to the level of analysis. All 5 hypotheses generated by the model seem to find a degree of support in the data. The fact that the model is supported by the data makes it a candidate for a tool to be used for policy evaluations. I perform such policy evaluation in Chapter 4 of this thesis. I shall return to the policy issue in Section 5.3 of this summary, after reviewing directions for future research.

5.2 Future Directions

In light of the findings of the current analysis, there are a number of possible future directions of research that have the potential to extend our understanding of markets with NP and FP players. In the 1990s in the U.S. a large number of physicians were opening speciality hospitals. These hospitals were usually organized as FPs and provided high quality treatments for a narrow range of conditions (Barr, 2007, p. 181). The theory developed in this thesis predicts that this phenomenon could be driven by increases in the number of patients who are able to assess the quality of care they receive. An empirical investigation that examines the association between increases in the

number of informed patients, or in information available to consumers in general, and the speed of growth of the speciality hospitals would provide an important test for the theory.

It is well known that the volume of patients that a hospital treats is positively associated with its quality of care (Feldman and Scharfstein, 2000; Gowrisankaran, Ho, and Town, 2006). Two potential reasons for the positive correlation between quality and volume have been proposed: learning by doing (i.e. volume affects quality), or patients select hospitals based on quality of care (quality affects volume). I offer a third explanation in a footnote in Chapter 3. If hospitals are financially constrained, it is plausible that those hospitals with higher volume of patients are able to obtain higher revenue and invest it in better equipment, which improves quality of care. NP hospitals are more likely to be financially constrained than FP ones (Cutler and Horwitz, 2000; Gentry and Penrod, 2000). Therefore, in order to test this explanation one needs to examine the effect of an exogenous variation in patient volume on quality for differences between the NP and FP sectors. It is important to perform such study in the future to improve our understanding of the association between volume and quality of care.

A slight modification of the model developed in this thesis predicts that when a policy that impacts the composition of healthcare markets in terms of organizational form is implemented, we are likely to observe movement of medical staff. Specifically, discriminatory policies toward FPs are likely to cause an exodus of highly skilled physicians from the jurisdiction that enacts such policies.[143] If one is able to obtain data on migration of doctors between the U.S. and Canada, or between states within the U.S., these implications can be tested. An association between the timing of various legislations that limit FPs in one state with movement of doctors to another, would provide crucial information for policymaking when deciding on policies toward FP healthcare provision. The problem of "brain drain" in Canada is believed to be severe. It is important to assess the exact scale of it and to determine whether policy intervention is needed. My theory shows that

[143]Not taking the movement of the medical staff into account, however, these policies are supported by my analysis.

changes to the regulations concerning organizational forms may reduce the "brain drain", while being potentially simpler to implement than increases in healthcare funding.

5.3 This Thesis and the U.S. Healthcare Policy

The issue of the impact of organizational form on quality is important from the policy perspective in the U.S. due to the on-going debate about conversions from NP to FP organizational forms within the healthcare sector. This wave of conversions had started in mid-1980s and received a wide attention both in the academic literature (e.g. Claxton et al., 1997; Needleman, Chollet, and Lamphere, 1997; Robinson, 2003) and in the general press (see the newspaper citations in the beginning of Section 1.1 in Chapter 1). The policy aspects of conversions are complex and extend beyond the issue of quality provision. Although the predictions of the model proposed in this thesis extend beyond the quality dimension as well, as any useful economic model, this model does not capture all aspects relevant to the policy debate. Nevertheless, it offers a unique perspective that enables the prediction of some of the market level effects of possible policy interventions. My analysis favours encouragement of conversions of FP hospitals that perform below comparable NPs on the quality dimension to NP status.

The implications of the theory developed in this thesis regarding waiting times appear particularly important for policy evaluation. The theory offers a warning that policies that discourage NP healthcare provision have the potential to lead to decreases in average waiting times for hospital procedures when the average is calculated over the entire pool of hospitals. This feature has the danger to appear desirable to some stakeholders. The reason for these decreases in waiting times, however, is likely not to be higher efficiency of healthcare delivery, but rather higher fees FPs charge that depress the demand (patients opting to forgo treatments). The last feature may have severe adverse welfare effects on the society.

As this thesis is being submitted, the U.S. president is pressing the legislative bodies to pass a bill that will constitute a reform to the funding of healthcare in the U.S. . According to a recent Reuters report,[144] the following changes are proposed: a universal requirement to obtain health insurance, disallowing insurance rating based on health history and gender in determination of coverage and fees, possible limits to insurance premiums, setting up a government run insurer to compete with the private players, creation of government run insurance exchanges for individuals and small businesses to buy insurance, abolishing the subsidies for insurance purchase for individuals with earnings above 300% of the poverty line (currently the limit to subsidies is set at 500%), increased taxation to fund the reforms for those with earnings in excess of $350,000.

If these reform proposals are implemented, the demand for healthcare in the U.S. is likely to rise. The increases in demand are likely to come mainly from individuals of lower socio-economic status. Even if some of the higher earning individuals, who currently refuse to purchase insurance, increase their consumption of healthcare services once they are forced to buy insurance, it is unlikely that their share of demand will be large, as the prevalence of diseases generally falls with income. It is likely that a large fraction of low income individuals are less able to assess the quality of medical procedures, due to lower levels of education, and fewer previous experiences with the medical system. Therefore, the increase in the demand for healthcare services will come mainly from less informed consumers.

As the reform has the potential to increase the demand for healthcare by patients who are less able to assess the quality of care, the welfare losses associated with the low quality FP providers are likely to increase. The fees in the low quality FP sector are likely to rise, which will disproportionately increase the demand and waiting times for procedures at NP facilities, according to my analysis. Both the fee and the waiting times increases are likely to raise the overall costs of

[144]Dunham, Will (editor) (2009) "Q+A: U.S. healthcare overhaul proposals and goals," Reuters, Jul 13, 2009 7:23pm EDT, `http://www.reuters.com/article/GCA-HealthcareReform/idUKTRE56C61720090713?sp=true` accessed on 15 July 2009 at 16:28.

healthcare provision that the U.S. has been trying to curb.[145] Especially in light of the reform, a policy intervention that would discourage low quality FP healthcare provision, supported by this thesis, may be even more crucial.

[145]This by no means implies that the reform does not have a welfare improving effect overall. If less patients forgo treatments as a result of the reform, a significant welfare improvement would be achieved.

Bibliography

Abraham, Jean Marie, Martin S. Gaynor, and William B. Vogt. 2005. "Entry and Competition in Local Hospital Markets." NBER Working Paper Series, Vol. w11649.

Akerlof, George A. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics* 84 (3):488–500.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53 (5):941–973.

Athey, Susan and Scott Stern. 2000. "The Adoption and Impact of Advanced Emergency Response Services." In *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, edited by David M. Cutler. NBER, The University of Chicago Press, 113–155.

Barer, Morris L. and William A. Webber. 1999. "Immigration and Emigration of Physicians to/from Canada." Centre for Health Services and Policy Research, The University of British Columbia.

Barr, Donald A. 2007. *Introduction to U.S. Health Policy: The Organization, Financing, and Delivery of Health Care in America*. Baltimore: John's Hopkins University Press.

Bénabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behaviour." *The American Economic Review Review* 96 (5):1652–1678.

Biglaiser, Gary and Ching-to Albert Ma. 2007. "Moonlighting: Public Service and Private Practice." *RAND Journal of Economics* 38 (4):1113–1133.

Bodenheimer, Thomas S. and Kevin Grumbach. 2009. *Understanding Health Policy: A Clinical Approach*. New York: McGraw-Hill Medical Publishing Division, 5th ed.

Boris, Elizabeth E. and C. Eugene Steuerle. 2006. "Scope and Dimensions of the Nonprofit Sector." In *The Nonprofit Sector: A Research Handbook*, edited by Walter W. Powell and Richard Steinberg, chap. 3. Yale University Press, second ed., 67–88.

Brekke, Kurt R., Luigi Siciliani, and Odd Rune Straume. 2008. "Competition and waiting times in hospital markets." *Journal of Public Economics* 92 (7):1607–1628.

Chandra, Amitabh and Jonathan Skinner. 2008. "Expenditure and Productivity Growth in Health Care." Dartmouth College, February. Forthcoming as an NBER Working Paper.

Claxton, Gary, Judith Feder, David Shactman, and Stuart Altman. 1997. "Public Policy Issues In Nonprofit Conversions: An Overview." *Health Affairs* 16 (2):9–28.

Cutler, David M. and Jill R. Horwitz. 2000. "Converting Hospitals from Not-for-Profit to For-Profit Status: Why and What Effects?" In *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, edited by David M. Cutler. NBER, The University of Chicago Press, 45–79.

Darby, Michael R. and Edi Karni. 1973. "Free Competition and the Optimal Amount of Fraud." *Journal of Law and Economics* 16 (1):67–88.

Dauphinee, W. Dale. 2005. "Physician migration to and from Canada: The challenge of finding the ethical and political balance between the individual's right to mobility and recruitment to

underserved communities." *Journal of Continuing Education in the Health Professions* 25 (1):22–29.

Deber, Raisa B. 2002. "Delivering Health Care Services: Public, Not-For-Profit, or Private?" Commission on the Future of Health Care in Canada Discussion Paper No. 17.

Delfgaauw, Josse. 2007. "Dedicated Doctors: Public and Private Provision of Health Care with Altruistic Physicians." Tinbergen Institute Discussion Paper TI 2007-010/1.

DeWalt, Darren A., Nancy D. Berkman, Stacey Sheridan, Kathleen N. Lohr, and Michael P. Pignone. 2004. "Literacy and health outcomes: A systematic review of the literature." *Journal of General Internal Medicine* 19 (12):1228–1239.

Erus, Burscay and Burton A. Weisbrod. 2003. "Objective Functions and Compensation Structures in Nonprofit and For-Profit Organizations: Evidence from the "Mixed" Hospital Industry." In *The Governance of Not-for-Profit Organizations*, edited by Edward L. Glaeser. NBER, The University of Chicago Press, 117–142.

Feldman, Sarah and David Scharfstein. 2000. "Managed Care and Provider Volume." In *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, edited by David M. Cutler. NBER, The University of Chicago Press, 229–248.

Francois, Patrick. 2003. "Not-for-Profit Provision of Public Services." *The Economic Journal* 113 (486):C53–C61.

———. 2007. "Making a difference." *RAND Journal of Economics* 38 (3):714–732.

Francois, Patrick and Michael Vlassopoulos. 2008. "Pro-social Motivation and the Delivery of Social Services." *CESifo Economic Studies* 54:22–54. URL `doi:10.1093/cesifo/ifn002`.

Friedman, Milton. 1953. *Essays in positive economics*. University of Chicago Press, 8 ed.

Garber, Alan M. and Jonathan Skinner. 2008. "Is American Health Care Uniquely Inefficient?" *Journal of Economic Perspectives* 22 (4):27–50.

Gentry, William M. and John R. Penrod. 2000. "The Tax Benefits of Not-for-Profit Hospitals." In *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, edited by David M. Cutler. NBER, The University of Chicago Press, 285–324.

Glaeser, Edward L. 2003. "Introduction." In *The Governance of Not-for-Profit Organizations*, edited by Edward L. Glaeser. NBER, The University of Chicago Press, 1–43.

Glaeser, Edward L. and Andrei Shleifer. 2001. "Not-for-Profit Entrepreneurs." *Journal of Public Economics* 81:99–115.

Gowrisankaran, Gautam, Vivian Ho, and Robert J. Town. 2006. "Causality, Learning and Forgetting in Surgery." Department of Economics, University of Arizona.

Gowrisankaran, Gautam and Robert J. Town. 1997. "Dynamic Equilibrium in the Hospital Industry." *Journal of Economics & Management Strategy* 6 (1):45–74.

Gruber, Jonathan. 2008. "Incremental Universalism for the United States: The States Move First?" *Journal of Economic Perspectives* 22 (4):51–68.

Hansmann, Henry. 1980. "The Role of Nonprofit Enterprise." *Yale Law Journal* 89:835–901.

Hansmann, Henry, Daniel Kessler, and Mark McClellan. 2003. "Ownership Form and Trapped Capital in the Hospital Industry." In *The Governance of Not-for-Profit Organizations*, edited by Edward L. Glaeser. The University of Chicago Press, 45–69.

Hansmann, Henry B. 1987. "Economic Theories of the Nonprofit Sector." In *The Nonprofit Sector*, edited by W. Powell. Yale University Press, 27–42.

Hirth, Richard A. 1999. "Consumer Information and Competition between Nonprofit and For-profit Nursing Homes." *Journal of Health Economics* 18:219–240.

Horwitz, Jill R. and Austin Nichols. 2007. "What Do Nonprofits Maximize? Nonprofit Hospital Service Provision and Market Ownership Mix." URL `http://www.nber.org/papers/w13246`. NBER Working Paper 13246.

James, Catherine A., Florence T. Bourgeois, and Michael W. Shannon. 2005. "Association of Race/Ethnicity with Emergency Department Wait Times." *Pediatrics* 115 (3):e310–e315.

Jones, Andrew M. and John Wildman. 2008. "Health, income and relative deprivation: Evidence from the BHPS." *Journal of Health Economics* 27 (2):308–324.

Kaestnera, Robert and Jose Guardado. 2008. "Medicare Reimbursement, Nurse Staffing, and Patient Outcomes." *Journal of Health Economics* 27 (2):339–361.

Krumholz, Harlan M., Yun Wang, Jennifer A. Mattera, Yongfei Wang, Lein Fang Han, Melvin J. Ingber, Sheila Roman, and Sharon-Lise T. Normand. 2006a. "An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With an Acute Myocardial Infarction." *Circulation* 113:1683–1692.

———. 2006b. "An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With Heart Failure." *Circulation* 113:1693–1701.

Liu, Yong and Charles B. Weinberg. 2009. "Pricing for Nonprofit Organizations." In *Handbook Of Pricing Research In Marketing*, edited by Vithala R. Rao, chap. 24. Edward Elgar.

Luft, H. S., D. W. Garnick, D. H. Mark, D. J. Peltzman, C. S. Phibbs, E. Lichtenberg, and S. J. McPhee. 1990. "Does quality influence choice of hospital?" *The Journal of American Medical Asscociation* 263 (21).

Malani, Anup, Tomas Philipson, and Guy David. 2003. "Theories of Firm Behavior in the Nonprofit Sector: Synthesis and Empirical Evaluation." In *The Governance of Not-for-Profit Organizations*, edited by Edward L. Glaeser. NBER, The University of Chicago Press, 181–215.

McClellan, Mark and Douglas Staiger. 2000. "Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals." In *The Changing Hospital Industry: Comparing Not-forProfit and For-Profit Institutions*, edited by David M. Cutler. NBER, The University of Chicago Press, 93–112.

McClellan, Mark B. and Douglas Staiger. 1999. "Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals." NBER Working Paper No. W7327.

Needleman, Jack, Deborah J. Chollet, and JoAnn Lamphere. 1997. "Hospital Conversion Trends." *Health Affairs* 16 (2):187–195.

Newhouse, Joseph P. 1970. "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital." *American Economic Review* 60 (1):64–74.

Pauly, Mark and Michael Redish. 1973. "The Not-for-Profit Hospital as a Physicians' Cooperative." *American Economic Review* 63 (1):87–99.

Philipson, Tomas. 2000. "Asymmetric Information and the Not-for-Profit Sector: Does Its Output Sell at a Premium?" In *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, edited by David M. Cutler. NBER, The University of Chicago Press, 325–345.

Richardson, Lynne D. and Ula Hwang. 2001. "Access to Care: A Review of the Emergency Medicine Literature." *Academic Emergency Medicine* 8 (11):1030–1036.

Robinson, James C. 2003. "The Curious Conversion Of Empire Blue Cross." *Health Affairs* 22 (4):100–118.

Rosenau, Pauline Vaillancourt and Stephen H. Linder. 2003. "Two Decades of Research Comparing For-Profit and Nonprofit Health Provider Performance in the United States." *Social Science Quarterly* 84 (2):219–241.

Schlesinger, Mark and Bradford H. Gray. 2006. "Nonprofit Organizations and Health Care: Some Paradoxes of Persistent Scrutiny." In *The Nonprofit Sector: A Research Handbook*, edited by Walter W. Powell and Richard Steinberg, chap. 16. Yale University Press, second ed., 378–431.

Shen, Yu-Chu. 2002. "The effect of hospital ownership choice on patient outcomes after treatment for acute myocardial infarction." *Journal of Health Economics* 21:901–922.

Siciliani, Luigi and Jeremy Hurst. 2005. "Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries." *Health Policy* 72 (2):201–215.

Siciliani, Luigi, Anderson Stanciolec, and Rowena Jacobs. 2009. "Do waiting times reduce hospital costs?" *Journal of Health Economics* In Press, Accepted Manuscript:–.

Skinner, Jonathan and John E. Wennberg. 2000. "How Much is Enough? Efficiency and Medicare Spending in the Last Six Months of Life." In *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, edited by David M. Cutler. NBER, The University of Chicago Press, 169–193.

Sloan, Frank A. 2000. "Not-for-profit Ownership and Hospital Behaviour." In *Handbook of Health Economics*, vol. 1, edited by Anthony J. Culyer and Joseph P. Newhouse, chap. 21. Elsevier, 1141–1174.

Solberg, Leif. 2009. "Lessons for non-VA care delivery systems from the U.S. Department of Veterans Affairs Quality Enhancement Research Initiative: QUERI Series." *Implementation Science* 4 (1):9.

Spector, William D., Thomas M. Selden, and Joel W. Cohen. 1998. "The Impact Of Ownership Type on Nursing Home Outcomes." *Health Economics* 7:639–653.

Steinberg, Richard. 2006. "Economic Theories of Nonprofit Organizations." In *The Nonprofit Sector: A Research Handbook*, edited by Walter W. Powell and Richard Steinberg, chap. 5. Yale University Press, second ed., 117–139.

Sultz, Harry A. and Kristina M. Young. 2009. *Health Care USA: Understanding Its Organization and Delivery.* Sudbury, Massachusetts: Jones and Bartlett Publishers, 6th ed.

Vlassopoulos, Michael. 2009. "Quality, Reputation and the Choice of Organizational Form." *Journal of Economic Behavior & Organization* In Press, Accepted Manuscript:–.

Vogt, William B. 2000. "Comment on Chapter 10." In *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, edited by David M. Cutler. NBER, The University of Chicago Press, 351–356.

Weinberg, Chargles B. 1980. "Marketing Mix Decision Rules for Nonprofit Organizations." *Research in Marketing* 3:191–234.

Weisbrod, Burton A. 1988. *The Nonprofit Economy.* Cambridge, MA: Harvard Univ. Press.

Wilper, Andrew P., Steffie Woolhandler, Karen E. Lasser, Danny McCormick, Sarah L. Cutrona, David H. Bor, and David U. Himmelstein. 2008. "Waits to see an emergency department physician: U.S. trends and predictors, 1997-2004." *Health Affairs (Project Hope)* 27 (2):w84 – 95.

Appendix A

Appendix to Chapter 2

A.1 Proof of the General Structure of Equilibrium: Proposition 1

In order to prove Proposition 1 in Chapter 2, I will show that there is one optimal contract type τ choice for each doctor type θ. This involves using the single crossing property twice. The optimal contract type choice will be the same for a continuous interval of doctor types (Lemma 2). Next, I need to show that the intervals of doctors for which the optimal contract type is the same are located along the interval $[\underline{\theta}, \overline{\theta}]$ in the order proposed. This reduces to proving that $\underline{\theta} < \theta^{AN} < \theta^{NH} < \overline{\theta}$ (Lemmas 4 through 6), where θ^{AN} is the marginal doctor type between contract types FP and NP, and θ^{NH} is the marginal doctor type between contract types NP and $\overline{FP}$. Thus all doctors of type $\theta \in [\underline{\theta}, \theta^{AN}]$ choose contracts of type FP,[146] doctors for whom $\theta \in (\theta^{AN}, \theta^{NH})$ choose contracts of type NP, and doctors of type $\theta \in [\theta^{NH}, \overline{\theta}]$ choose contracts of type $\overline{FP}$.[147] Together with Propositions 3 through 5, this ensures that Proposition 1 holds by continuity of $q^{\theta}(\theta)$ and $q^{\alpha,\theta}(\theta)$ in θ.

I start by proving the following theorem that will be useful in proving the lemmas that will

[146] Lemma 2 establishes what contract type dominates to the left of each marginal doctor type, and what contract type to the right of it.

[147] One may think that it could be the case that the marginal doctor type between contracts of type FP and $\overline{FP}$, call it θ^{AH}, is important as well. However, since all of the derived utility function are upward sloping, the single crossing property guarantees that θ^{AH} falls in the interval over which NP type contract dominates both of FP and $\overline{FP}$ types, therefore proving that $\underline{\theta} < \theta^{AN} < \theta^{NH} < \overline{\theta}$ is sufficient for Proposition 1 to hold.

follow.

Theorem 1 *If the following assumptions are satisfied*

1. $c(q;\theta)$ *is twice continuously differentiable in* q *and* θ
2. $c_q(0;\cdot) = 0$
3. $c_{qq}(q;\cdot) > 0 \ \forall q \in \mathbb{R}_+$
4. $c_\theta(\cdot;\theta) < 0 \ \forall \theta \in [\underline{\theta}, \overline{\theta}]$

then the socially efficient surplus from medical treatment in a world without altruism monotonically increases in doctor type θ*, for all* θ *in* $[\underline{\theta}, \overline{\theta}]$*.*

Proof Recall that the socially efficient surplus from medical treatment in a world without altruism is defined by the following FOC (first order condition)

$$c_q(q;\theta) = 1$$

as the Left Hand Side (LHS) is the marginal social cost of quality provision by a doctor, and the Right Hand Side (RHS) is the marginal social benefit of such provision. I defined q^θ to be the value of q for each θ that satisfies this condition. By Assumptions 2 and 3 of the theorem, q^θ exists. Additionally, Assumption 3 guarantees that the second order conditions for a local maximum are satisfied, and that q^θ is unique for each θ.

To prove the theorem it is enough to show that in general $\dfrac{d[q^\theta(\theta) - c(q^\theta(\theta);\theta)]}{d\theta} > 0$ for all θ. Differentiate this expression to get

$$\frac{d[q^\theta(\theta) - c(q^\theta(\theta);\theta)]}{d\theta} = q^{\theta\,\prime}(\theta) - [c_q(q^\theta;\theta)q^{\theta\,\prime}(\theta) + c_\theta(q^\theta;\theta)] = -c_\theta(q^\theta;\theta) > 0 \qquad \text{(A.1.1)}$$

By Assumptions 1 and 3 of the theorem, $q^{\theta\,\prime}(\theta)$ exists (in fact, it can be found by applying the implicit function theorem; Assumption 3 is needed to make sure there is no division by zero). The second equality holds by the fact that the derivative is taken at the optimal q, therefore, $c_q(q^\theta;\theta) = 1$. Finally, the inequality holds by Assumption 4 of the theorem. As the inequality is true for all θ, the socially optimal surplus is monotonically increasing in θ. This proves the theorem. ■

In order to show that the general structure of the equilibrium is as described in Proposition 1 holds, first I show that single crossing property holds for ratios of derived utility functions. This proves that: if θ^{AN} and θ^{NH} exist, there are continuous intervals of doctor types over which one contract type dominates the other two. Then I prove that $\theta^{AN} < \theta^{NH}$, and that both of the marginal types fall strictly in the interior of the interval $[\underline{\theta}, \overline{\theta}]$. This ensures that θ^{AN} and θ^{NH} indeed exist and that doctors self-select to contract types as described in Proposition 1. The final feature is the building stone of the distribution of quality in a market with both FP and NP healthcare provision.

As I have shown in Section 2.3 of Chapter 2, there is a unique doctor type-optimal contract choice within each contract type τ. Therefore, I can write doctors' derived utilities as a function of their type as[148]

$$u^A(\theta) = q^e - s_U - c(q^{\alpha,\theta};\theta) + \alpha[q^{\alpha,\theta} - (q^e - s_U)]$$

if a doctor chooses a contract of type FP

$$u^N(\theta) = \alpha[q^\theta - c(q^\theta;\theta)]$$

[148]These functions are derived utilities in the sense that, *given* the equilibrium strategies of all other players, these functions represent the highest level of utility a doctor of type θ can obtain in the game.

if she chooses a contract of type NP and

$$u^H(\theta) = q^\theta - c(q^\theta;\theta) - (1-\alpha)s_I$$

if she chooses a contract of type $\overline{FP}$.

Where q^θ is the quality choice of a doctor of type θ that is socially efficient in a non-altruistic world. And $q^{\alpha,\theta}$ is the quality choice that is below the socially efficient level, that a doctor chooses under contract of type FP. s_U and s_I are the equilibrium payoffs of uninformed and informed patients, respectively.

Assume the following relationships to hold

$$q^{\alpha,\theta} = \alpha q^\theta \tag{A.1.2}$$

$$c(\alpha q^\theta;\theta) = \alpha^2 c(q^\theta;\theta) \tag{A.1.3}$$

Therefore, u^A can be re-written as

$$u^A(\theta) = (1-\alpha)[q^e - s_U] + \alpha^2[q^\theta - c(q^\theta;\theta)]$$

Lemma 2 *If in addition to the assumptions in Theorem* 1 $c(0;\theta) = 0$ *hods as well, then for* $(i,j) \in \{(A,N),(N,H)\}$ *the following holds: if* u^i *crosses* u^j*, then*

1. *They cross only once.*

2. *To the left of the crossing point* u^i *dominates* u^j*, and to the right* u^j *dominates* u^i*.*

3. *The subsets of* $\mathbb{R}$ *over which one function dominates the other are continuous intervals.*[149]

[149]Note that $[\underline{\theta},\overline{\theta}] \subset \mathbb{R}$. Only if the crossing points fall in the interval $[\underline{\theta},\overline{\theta}]$, the functions cross. I show this in the next lemmas.

Proof I start with $(i,j) = (A,N)$. Note that $\frac{u^A(\theta)}{u^N(\theta)} = \frac{(1-\alpha)[q^e - s_U]}{\alpha[q^\theta - c(q^\theta;\theta)]} + \alpha$, therefore

$$\frac{d[u^A(\theta)/u^N(\theta)]}{d\theta} = -\left[\frac{-\alpha c_\theta(q^\theta;\theta)\{(1-\alpha)[q^e - s_U]\}}{\{\alpha[q^\theta - c(q^\theta;\theta)]\}^2}\right] < 0 \ \forall\theta$$

The derivative is taken at the optimal θ, therefore, as argued in Theorem 1, the derivative of the total surplus reduces to $-c_\theta(q^\theta;\theta)$, which is positive. $c(0;\theta) = c_q(0;\theta) = 0$ together with $c_{qq}(\cdot;\theta) > 0$ imply that $q^{\alpha,\theta} > 0$, therefore, $q^e - s_U = F \geq c(q^{\alpha,\theta};\theta) > 0$. Hence, the derivative of the ratio of the two derived utility functions is negative. This implies that if u^A and u^N cross, they cross only once.

Since the ratio is decreasing, if the functions cross, it equals 1 at a certain point. To the left of this point the numerator u^A is greater than the denominator u^N, and to the right the denominator is greater than the numerator. Hence, to the left of the crossing point u^A dominates u^N, and to the right the opposite is true. By continuity of u^A and u^N in θ, the dominance regions are continuous intervals.

Now, let $(i,j) = (N,H)$. In a similar manner as in the previous paragraph: $\frac{u^H(\theta)}{u^N(\theta)} = \frac{1}{\alpha} - \frac{(1-\alpha)s_I}{\alpha[q^\theta - c(q^\theta;\theta)]}$, and

$$\frac{d[u^H(\theta)/u^N(\theta)]}{d\theta} = \frac{-\alpha(1-\alpha)s_I c_\theta(q^\theta;\theta)}{\{\alpha[q^\theta - c(q^\theta;\theta)]\}^2} > 0 \ \forall\theta$$

I have to rely here on the fact that $s_I > 0$ and the assumption that $c_\theta(\cdot;\theta) < 0$.[150] This implies that if u^N and u^H cross, they cross only once.

Applying the same arguments as in the case of u^A and u^N (only reversed due to the reverse signs of the derivatives): if the functions cross, the ratio equals 1 at this point. To the left of the

[150] In equilibrium $s_I = q^{\theta^{NH}} - c(q^{\theta^{NH}};\theta^{NH})$. The RHS of this expression is the socially optimal surplus that type θ^{NH} generates. By Lemma 3, one can always find $\overline{\theta}$ such that $\theta^{NH} \in [\underline{\theta},\overline{\theta}]$, which implies that her socially efficient surplus is positive: $c(0;\theta) = 0$ (for all $\theta \in [\underline{\theta},\overline{\theta}]$), $[q - c(q;\theta)]$ is maximized under q^θ, and its maximum (strictly) rises with type by Theorem 1. Since, $0 - c(0;\underline{\theta}) = 0$, $[q^{\theta^{NH}} - c(q^{\theta^{NH}};\theta^{NH})]$, and hence, s_I must be positive.

crossing point u^N dominates, and to the right of it u^H dominates. The dominance regions are continuous intervals. As I showed this for both pairs of functions, this proves the lemma. ■

The following four lemmas together imply that both of the pairs of functions in Lemma 2 indeed cross on the interval $[\underline{\theta}, \overline{\theta}]$.

Lemma 3 *If (A.1.2) and A.1.3 hold, then for any $s_U \in (0, \infty)$ there exists α such that $\theta^{AN} < \theta^{NH}$.*

Proof In order to prove Lemma 3, I first need a to derive expressions that characterize the marginal doctors types θ^{AN} and θ^{NH}, and to show that $\theta^{AN} < \theta^{NH}$. I start with is θ^{AN}. First define $\tilde{G_\theta}(\theta) \equiv \frac{G_\theta[\theta]}{G_\theta[\theta^{AN}]}$ The equation that characterizes this value is $u^A(\theta^{AN}) = u^N(\theta^{AN})$, or explicitly after noting that

$$q^e = \int_{\underline{\theta}}^{\theta^{AN}} q^{\alpha,\theta}(\theta) d\tilde{G_\theta}(\theta) = \int_{\underline{\theta}}^{\theta^{AN}} \alpha q^\theta(\theta) d\tilde{G_\theta}(\theta)$$

and re-arranging

$$\int_{\underline{\theta}}^{\theta^{AN}} q^\theta(\theta) d\tilde{G_\theta}(\theta) - \frac{s_U}{\alpha} = q^{\theta^{AN}} - c(q^{\theta^{AN}}; \theta^{AN}) \tag{A.1.4}$$

and $\int_{\underline{\theta}}^{\theta^{AN}} q^\theta(\theta) d\tilde{G_\theta}(\theta)$ is simply $E(q^\theta | \theta \in [\underline{\theta}, \theta^{AN}])$.

Similarly, I can derive the equation that defines θ^{NH}. It is characterized by $u^N(\theta^{NH}) = u^H(\theta^{NH})$, or explicitly[151]

$$s_I = q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH}) \tag{A.1.5}$$

In order to prove Lemma 3, I need to show that

$$q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH}) > q^{\theta^{AN}} - c(q^{\theta^{AN}}; \theta^{AN})$$

[151] This is simply the equilibrium value of s_I, as in equilibrium $\theta^{NH} = \overline{\theta} - I$.

which by Theorem 1 will imply that $\theta^{AN} < \theta^{NH}$. However, it is enough to show that

$$E(q^{\theta}|\theta \in [\underline{\theta}, \theta^{AN}]) - \frac{s_U}{\alpha} < q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH})$$

or

$$E(q^{\theta}|\theta \in [\underline{\theta}, \theta^{AN}]) < q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH}) + \frac{s_U}{\alpha}$$

Consider two cases:

1. $E(q^{\theta}|\theta \in [\underline{\theta}, \theta^{AN}]) < q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH})$
2. $E(q^{\theta}|\theta \in [\underline{\theta}, \theta^{AN}]) \geq q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH})$

In case 1, the above inequality trivially holds, as $s_U > 0$ and $\alpha > 0$.[152] In case 2, the following holds[153]

$$\lim_{\alpha \to 0} \frac{s_U}{\alpha} = \infty$$

Therefore, there exist $\alpha \in (0, 1)$ such that the inequality holds. From the two cases it follows that: for any $s_U \in (0, \infty)$ there exist (an interval of) α such that $\theta^{AN} < \theta^{NH}$. This completes the proof. ■

The next step is to show that $\theta^{NH} < \overline{\theta}$ and that $\theta^{AN} > \underline{\theta}$. These two properties will ensure that the derived utility functions cross on the interval $[\underline{\theta}, \overline{\theta}]$, and, therefore, all of the three contract types are chosen in equilibrium by intervals of doctor types such that Proposition 1 holds. The

[152] Recall that s_U is the equilibrium payoff of uninformed patients. It is set at the level of highest payoff from a deviation from their equilibrium behavioural strategy, due to the first-mover advantage of the doctors. Highest payoff from deviation is the payoff that uninformed patients can get by deviating to the NP sector (see Subsection 2.3.4 in Chapter 2), where the fees are set at the break-even level. There is a positive probability (strictly between 0 and 1) to get the entire socially efficient surplus (in a non-altruistic world) from treatment by a doctor who chose a contract of type NP. This surplus is always positive. s_U is a product of a strictly positive probability and the socially efficient surplus, therefore, it is positive as well.

[153] For completeness I need to ensure that $q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH}) > -\infty$. Footnote 150, however, shows that $q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH}) > 0$.

argument that Lemmas 4, 5 (for the extension to the baseline model) and Lemma 6 (for the baseline model) establish is the following: Pick any generic interval $[\underline{\theta}, \overline{\theta}]$ such that θ^{NH} lies within it, θ^{AN} is guaranteed to lie within it as well.

Lemma 4 *If $D > I > 0$, one can always find $\overline{\theta} \in \mathbb{R}$ such that $\theta^{NH} < \overline{\theta}$ holds.*

Proof Note, that the only condition that $\underline{\theta}$ and $\overline{\theta}$ must satisfy is $\overline{\theta} - D = \underline{\theta}$. Therefore, the choice of the actual $\underline{\theta}$ and $\overline{\theta}$ is rather arbitrary. One cal always set $\overline{\theta} = \theta^{NH} + I$. This would guarantee that $\theta^{NH} < \overline{\theta}$, as long as $I > 0$, or there are informed patients in the economy.[154]

■

Finally, it remains to prove that $\theta^{AN} > \underline{\theta}$. The proof for the baseline model differs from the proof of this condition for the extension to the model. First I prove this condition for the extension to the model, where the proof is more intuitive.[155] In order to prove this property in the baseline model I need to introduce an extra assumption on the relationship between the cost function and α. I do this in Lemma 6.

Lemma 5 *If $(A.1.2)$ and $(A.1.3)$ hold, then there exists U such that for any $D > 0$, $\theta^{AN} > \underline{\theta}$ in the extension to the model.*

Proof I will show that $u^A(\underline{\theta}) > u^N(\underline{\theta})$ even when $q^e = q^{\alpha,\underline{\theta}}$, or that the lowest doctor type $\underline{\theta}$ always prefers FP type contract to NP type contract, even when her shirking level of quality is perfectly identified. Then it has to follow that the type who is indifferent between FP and NP type contracts, θ^{AN} is to the right of $\underline{\theta}$. Preference for FP type contract is even stronger when

[154] In this case $\underline{\theta} = \theta^{NH} - (D - I)$ (recall that $D > I$). One may ask: what happens if we pick $\overline{\theta} \neq \theta^{NH} + I$? In this case θ^{NH} will adjust to satisfy this equality.

[155] In fact, what the proofs of Lemmas 5 and 6 establish is the following: for any arbitrary chosen $\underline{\theta}$, θ^{AN} must always lie above it.

$q^e > q^{\alpha,\underline{\theta}}$, as in such a case F is higher. $q^e > q^{\alpha,\underline{\theta}}$ will hold in equilibrium if there exists a doctor of type that is arbitrary close, but higher than $\underline{\theta}$ who also offers an FP type contract.

By assumptions $(A.1.2)$ and $(A.1.3)$

$$q^{\alpha,\underline{\theta}} - c(q^{\alpha,\underline{\theta}};\underline{\theta}) = \alpha q^{\underline{\theta}} - \alpha^2 c(q^{\underline{\theta}};\underline{\theta}) > \alpha[q^{\underline{\theta}} - c(q^{\underline{\theta}};\underline{\theta})]$$

in words: the total surplus from transaction under FP contract is higher than the fraction α of first-best total surplus for any doctor type, and for type $\underline{\theta}$ in particular (recall that $\alpha \in (0,1)$). Incidentally, fraction α of the first best surplus is also the payoff to a doctor under NP type contract. The last inequality, however, implies that there must exist U such that the following holds

$$q^{\alpha,\underline{\theta}} - c(q^{\alpha,\underline{\theta}};\underline{\theta}) > \alpha[q^{\underline{\theta}} - c(q^{\underline{\theta}};\underline{\theta})] + (1-\alpha)\frac{D}{U}$$

notice that $\lim_{U\to\infty}(1-\alpha)\frac{D}{U} = 0$. Therefore, at the limit, the last inequality always will be satisfied by assumptions $(A.1.2)$ and $(A.1.3)$,as D was assumed to be positive.

Next, note that $\frac{D}{U}$ is the upper bound on the value of s_U in the extension to the baseline model.[156] The ratio of D (the upper bound on the measure of doctors in the NP sector) to U (measure of patients in the NP sector) is the upper bound of the expected payoff to an uninformed patient from accepting a NP type contract (see Subsection 2.3.5 in Chapter 2).

The last inequality implies

$$u^A(\underline{\theta}) > u^N(\underline{\theta})$$

when $q^e = q^{\alpha,\underline{\theta}}$. Therefore, $\underline{\theta}$ always prefers FP type contract, even when her quality is perfectly identified by uninformed consumers. Preference for FP type contract is even stronger when $q^e > q^{\alpha,\underline{\theta}}$, as in such a case F is higher. By continuity of both u^A and u^N the inequality must hold for

[156]When only $\underline{\theta}$ chooses FP type contract and the rest of the doctors in the economy, who do not offer $\overline{FP}$ type contracts, all offer NP type ones $s_U = \frac{D-I}{U}$.

some type slightly to the right of $\underline{\theta}$. This reasoning implies that $\theta^{AN} > \underline{\theta}$, which proves the lemma.

■

At first glance Lemma 5 may seem incompatible with Lemma 3, however, the logic is following: for any arbitrary small s_U that satisfies Lemma 5, one can always find a (much smaller) α that satisfies Lemma 3.

Lemma 6 *If* $q^{\alpha,\underline{\theta}} - c(q^{\alpha,\underline{\theta}};\underline{\theta}) > [\alpha + \frac{1}{2}(1-\alpha)]\{q^{\underline{\theta}} - c(q^{\underline{\theta}};\underline{\theta})\}$, $(A.1.2)$ *and* $(A.1.3)$ *hold for all* $\alpha \in (0,1)$*, then* $\theta^{AN} > \underline{\theta}$ *in the baseline model.*

Proof The expression below follows directly by rearranging the first inequality in the assumptions of the lemma when $q^e = q^{\alpha,\underline{\theta}} = \alpha q^{\underline{\theta}}$ and $s_U = \frac{1}{2}[q^{\underline{\theta}} - c(q^{\underline{\theta}};\underline{\theta})]$

$$u^A(\underline{\theta}) > u^N(\underline{\theta})$$

Therefore, it must be the case that $\underline{\theta} < \theta^{AN}$, following the same logic as in the proof of Lemma 5.

■

In this section I proved three points: a) Each contract type dominates the other two for a continuous interval of doctor types (Lemma 2); b) the marginal doctor types θ^{AN} (to the left of it FP dominates NP type contract, and to the right vice-versa - Lemma 2) and θ^{NH} (to the left of it NP dominates $\overline{FP}$, and to the right vice-versa - Lemma 2) have the following relationship: $\theta^{AN} < \theta^{NH}$ (Lemma 3); c) the marginal types are in the interior of any genetic interval $[\underline{\theta}, \overline{\theta}]$ (Lemmas 4 through 6). a), b) and c) together imply that the equilibrium distribution of the quality of care is as described in Proposition 1: Highest levels of quality of care in the economy are provided by FPs to informed patients (under $\overline{FP}$ type contracts); a second tier of quality levels is provided by NPs to uninformed patients (under NP type contracts); a third tier of quality

levels is provided by FPs to uninformed patients (under FP type contracts). This establishes the proposition.

A.2 Characterization of the Perfect Bayesian Equilibrium

A.2.1 Statement of the Equilibrium

For a complete description of the Perfect Bayesian Equilibrium (PBE), I need to find the strategies for doctors and patients and uninformed patients' beliefs. Formally, doctors' equilibrium strategies (the choice of $[q, F; \tau]$ as a function of their type θ) are

$$s_D^*(\theta) = \begin{cases} [q^{\alpha,\theta}, q^e - s_U; FP] & \text{if },\theta \in [\underline{\theta}, \theta^{AN}] \\ [q^\theta, c(q^\theta;\theta); NP] & \text{if },\theta \in (\theta^{AN}, \theta^{NH}) \\ [q^\theta, q^\theta - s_I; \overline{FP}] & \text{if },\theta \in [\theta^{NH}, \overline{\theta}] \end{cases} \tag{A.2.1}$$

In the baseline model

$$s_U = \frac{q^{\theta^{AN}} - c(q^{\theta^{AN}};\theta^{AN})}{2}$$

In the extension to scare capacity its value is

$$s_U = \frac{\theta^{NH} - \theta^{AN}}{U - (\theta^{AN} - \underline{\theta})}$$

While s_I in both cases is

$$s_I = q^{\theta^{NH}} - c(q^{\theta^{NH}};\theta^{NH})$$

Uninformed patients' beliefs about quality of service when they sign FP type contract are

$$\mu(q; FP) = \frac{G_\theta[q^{\alpha,\theta\ -1}(q)]}{G_\theta[\theta^{AN}]} \tag{A.2.2}$$

Where $q^{\alpha,\theta}(\theta)$ exists and is invertible by twice continuous differentiability of $c(q;\theta)$ in both variables and if one is willing to assume that $c_{q\theta}(\cdot;\cdot) < 0$.

Uninformed patients' equilibrium strategies can be described as accepting NP or FP type contract, such that in equilibrium without capacity constraints every patient selects one of these contracts. With capacity constraints (the extension to the baseline model) uninformed patients select each doctor such that probability of treatment by each doctor of type $\theta \in (\theta^{AN}, \theta^{NH})$ is

$$\frac{s_U}{\overline{CS}[q^\theta, c(q^\theta;\theta)]}$$

and of type $\theta \in [\underline{\theta}, \theta^{AN}]$ is 1.

The equilibrium strategy of an informed patient in both the baseline model and the extension is accepting a $\overline{FP}$ type contract, such that every informed patient selects one of these contracts in equilibrium.

A.2.2 Sketch of a Proof that the Solution is an Equilibrium

Now, I can sketch a general proof that the solution above is an equilibrium. All I need to do is to show that there are no profitable deviations for either of the players and that uninformed patients' beliefs are consistent with play on the equilibrium path. I do this class by class: first for informed patients, then for uninformed (both the ones who accept FP type contracts in equilibrium and those who accepts NP type ones). As I have derived the doctors' equilibrium strategies as solutions to their optimization problems, they do no have profitable deviations by construction.

In what follows I establish that patients have no profitable deviations. I start with informed, then proceed to uninformed. Finally, I show that uninformed patients' beliefs about quality are consistent with play on the equilibrium path.

In equilibrium, all informed patients accept $\overline{FP}$ type contracts. These patients can deviate to accepting either a contract of type NP, FP, or another contract of type $\overline{FP}$. I consider first deviations to accepting NP type contract, then other $\overline{FP}$, and finally FP. The payoff an informed patient can obtain from a deviation to accepting a NP type contract (given the equilibrium play of all other doctors and patients) is the first-best surplus of the target doctor. The deviating informed patient will be treated with probability 1 due to her priority over uninformed. If the target doctor is of type θ^{NH}, this payoff is $q^{\theta^{NH}} - c(q^{\theta^{NH}}; \theta^{NH}) = s_I$, which the equilibrium payoff an informed patient. Deviation to accepting any other NP type contract will result in lower consumer surplus, by Theorem 1 in Section A.1. Therefore, accepting any NP type contract is not a profitable deviation for informed.

Consider a deviation to accepting another $\overline{FP}$ type contract. Such deviation will result in treatment with probability $\frac{1}{2}$ and in the event of treatment consumer surplus of s_I. Therefore, the expected payoff from deviation to accepting another $\overline{FP}$ type is $\frac{s_I}{2}$, which is less than the pre-deviation payoff of s_I. Hence, deviating to accepting any $\overline{FP}$ type contract is not profitable either. Finally, I need to show that informed patients do not have a profitable deviation to accepting any FP type contract.

Consider a deviation to accepting a FP type contact by an informed patient. Recall that the fee doctors charge under FP type contract is $F = q^e - s_U$, equal across all types who choose FP. Informed patients can observe the actual quality chosen by each doctor. Additionally, they have a priority in treatment over uninformed patients. The consumer surplus an informed patient can obtain by deviating to accepting a FP type contract with doctor of type θ^{AN} is $q^{\alpha,\theta^{AN}} - (q^e - s_U)$. This is also the highest obtainable surplus from a deviation to accepting FP type contract, as

$q^{\alpha,\theta}$ increases in θ by $c_{q,\theta}(\cdot,\cdot) < 0$. By Theorem 1 and Lemma 3 in Section $A.1$, the first-best surplus of θ^{NH}, which is also the equilibrium payoff of an informed patient, is higher than the first-best surplus of θ^{AN}, $q^{\theta^{AN}} - c(q^{\theta^{AN}}; \theta^{AN})$. $(q^e - s_U)$ must be higher than the cost of provision of quality for doctor θ^{AN}, $c(q^{\alpha,\theta^{AN}}; \theta^{AN})$, otherwise, θ^{AN} would prefer to offer a NP type contract. Therefore, the payoff from deviation must be lower than the total shirking surplus from treatment by θ^{AN}, $q^{\alpha,\theta^{AN}} - c(q^{\alpha,\theta^{AN}}; \theta^{AN})$. However, this surplus is even lower than the first-best surplus of θ^{AN} (Proposition 4 in Chapter 2, where "lower than socially efficient quality" implies that the total surplus is not maximized, therefore, it is lower than first-best). As it was established that the first-best surplus of θ^{AN} is lower than the equilibrium payoff of informed patients, it is not profitable to deviate to accepting FP type contract. As it is not profitable to deviate to accepting NP or another $\overline{FP}$ type contracts either, the informed patients have no profitable deviations. It remains to establish the same and consistency of beliefs for uninformed patients.

Uninformed patients can deviate to accepting either FP or NP type contract. These are also the contract types they are accepting in equilibrium. I consider deviations across sectors first, and then within. An uninformed patient who chose a contract of type FP can deviate to accepting some NP type contract, this deviation gives her a payoff of s_U in both the baseline model and the extension. This payoff is equal to her equilibrium payoff, therefore, this deviation is not profitable. Consider a deviation from accepting NP type contact to a FP type. The expected payoff from accepting FP type contract is $\frac{s_U}{2}$, there is probability half of being treated. This payoff is less than the equilibrium payoff of s_U. Thus, there are no profitable deviations for uninformed patients across sectors. Consider a deviation within NP sector (to another NP type contract). A payoff from such a deviation is s_U (in both the baseline model and the extension), this deviation is not profitable. Finally, consider a deviation within the FP sector. The payoff from such a deviation is $\frac{s_U}{2}$, due to rationing, again this is not profitable. Therefore, no profitable deviations exist for uninformed patients. It remains to examine their equilibrium beliefs.

The concept of PBE requires than in addition to the fact that no player has profitable deviations, the beliefs of (uninformed) players are derived using the Bayes rule on the equilibrium path. In the game analyzed in Chapter 2 there are no events from which uninformed patients can deduce quality chosen by doctors under FP type contracts. Therefore, the application of Bayes rule to this case amounts to showing that the beliefs of uninformed patients are consistent with the equilibrium play of the doctors. In Subsection 2.2.5 of Chapter 2, I showed the consistency of beliefs about quality for NP type contracts, which resulted in perfect inference of quality through the fee by uninformed patients. In Subsection 2.3.4 of Chapter 2 I showed that under FP type contract doctors set the quality of care using decision rule (2.3.4). The equilibrium belief function (A.2.2) incorporates this decision rule, therefore, it is consistent with doctors' equilibrium play.

In this subsection I have shown that: a) no doctor has a profitable deviation from her equilibrium strategy (by construction), b) no informed patient has a profitable deviation from her equilibrium strategy, c) no uninformed patient has a profitable deviation from her equilibrium strategy, d) equilibrium beliefs of uninformed patients are consistent with doctors' equilibrium play. These four points together imply that the solution I proposed to the model developed in Chapter 2 is indeed an equilibrium.

Appendix B

Appendix to Chapter 3

B.1 Calculation of Risk Adjusted (Standardized) Mortality Rates

In order to perform the estimations in Section 3.5 of Chapter 3 I use data on 30-day post admission risk-adjusted (or standardized) mortality rates (RSMR) as a measure of quality for a hospital. The following information on RSMR calculations comes from the U.S. Department of Health & Human Services via the Hospital Compare website.[157] "All people with Medicare aged 65 or over who were enrolled in Original Medicare (traditional fee-for-service Medicare) for the entire 12 months [from July 2006 through June 2007] prior to their hospital admission for heart attack or heart failure or pneumonia, and for whom complete administrative data for that 12-month period are available, are included in the [sample]." "All short-stay acute-care hospital discharges for heart attack or heart failure or pneumonia in the reference year based on a principal discharge diagnosis on the Medicare beneficiary's inpatient claim" were merged with a dataset of "all deaths (for all causes) within 30 days of admission. Hospital stays that lasted one day or less are excluded, provided the patient was discharged alive and not against medical advice." "Hospital mortality rates for heart attack are calculated based on all admissions for heart attack, even if a person with Medicare was hospitalized more than once for this condition during the 12-month period. However, for purposes

[157] `http://www.hospitalcompare.hhs.gov/Hospital/Static/InformationforProfessionals_tabset.asp` accessed on 24 June 2009 at 11:46.

of calculating heart failure and pneumonia mortality rates, if a beneficiary had multiple admissions during the 12-month period, one admission is chosen randomly for inclusion in the [dataset]."

After the data collection, "Medicare.gov - Hospital Compare: Information for Professionals",[158] reports the following estimations. Hierarchical regression model estimates the effect of patient-specific risk characteristics and "hospital's unique quality of care for all patients treated for that condition in that hospital... on mortality." "Predicted mortality" was defined as "the rate of deaths from heart attack or heart failure or pneumonia that would be anticipated in the particular hospital during the 12-month period, given the patient case mix and the hospitals unique quality of care effect on mortality."

To obtain estimates of "predicted mortality" for Heart Attacks the following logit specification (that clusters the observations at the hospitals level, with hospital specific random effects - this reduces the volatility of the estimate of RSMR for smaller hospitals by moving them toward the "U.S. National mortality rate" - see below) was estimated. Dependent variable equal to 1 if the patient died within 30 days of admission for Heart Attack and zero otherwise was regressed on: Age (years over 65), and dummy variables for: Male, History of Percutaneous Transluminal Coronary Angioplasty (PTCA), History of Coronary Artery Bypass Graft (CABG), History of heart failure (HCC 80), History of Acute Myocardial Infarction (AMI) (HCC 81), Anterior MI (ICD-9 410.00410.19), Inferior/lateral/posterior MI (ICD-9 410.20410.69), Unstable angina (HCC 82), Chronic atherosclerosis (HCC 83 and 84), Cardiopulmonary-respiratory failure and shock (HCC 79), Valvular heart disease (HCC 86), Hypertension (HCC 89 and 91), Stroke (HCC 95 and 96), Cerebrovascular disease (HCC 97, 98, 99, 103), Renal failure (HCC 131), Chronic Obstructive Pulmonary Disease (COPD) (HCC 108), Pneumonia (HCC 111, 112, 113), Diabetes (HCC 1520, 120), Protein-calorie malnutrition (HCC 21), Dementia (HCC 4950), either one of: Hemiplegia,

[158] `http://www.hospitalcompare.hhs.gov/Hospital/Static/InformationforProfessionals_tabset.asp` accessed on 24 June 2009 at 11:46.

paraplegia, paralysis, functional disability (HCC 100, 101, 102, 68, 69, 177, 178), Peripheral vascular disease (HCC 104, 105), Metastatic cancer (HCC 7, 8), Trauma in last year (HCC 154156, 158162), Major psychiatric disorders (HCC 54, 55, 56) and Chronic liver disease (HCC 25, 26, 27) (Krumholz et al., 2006a).

To obtain estimates of "predicted mortality" for Heart Failures the following logit specification (that clusters the observations at the hospitals level, with hospital specific random effects - this reduces the volatility of the estimate of RSMR for smaller hospitals by moving them toward the "U.S. National mortality rate" - see below) was estimated. Dependent variable equal to 1 if the patient died within 30 days of admission for Heart Failure and zero otherwise was regressed on: Age (years over 65), and dummy variables for: Male, History of Percutaneous Transluminal Coronary Angioplasty (PTCA), History of Coronary Artery Bypass Graft (CABG), History of heart failure (HCC 80), History of Myocardial Infarction (MI) (HCC 81), Unstable angina (HCC 82), Chronic atherosclerosis (HCC 83 and 84), Cardiopulmonary-respiratory failure and shock (HCC 79), Valvular heart disease (HCC 86), Hypertension (HCC 89 and 91), Stroke (HCC 95 and 96), Renal failure (HCC 131), Chronic Obstructive Pulmonary Disease (COPD) (HCC 108), Pneumonia (HCC 111, 112, 113), Diabetes (HCC 1520, 120), Protein-calorie malnutrition (HCC 21), Dementia (HCC 4950), either one of: Hemiplegia, paraplegia, paralysis, functional disability (HCC 100, 101, 102, 68, 69, 177, 178), Peripheral vascular disease (HCC 104, 105), Metastatic cancer (HCC 7,8), Trauma in last year (HCC 154156,158162), Major psychiatric disorders (HCC 54, 55, 56) and Chronic liver disease (HCC 25, 26, 27) (Krumholz et al., 2006b).

"Expected mortality" was defined as "the rate of deaths from heart attack or heart failure or pneumonia that would be expected if the same patients with the same characteristics had instead been treated at an 'average' hospital, given the 'average' hospital's quality of care effect on mortality for patients with that condition." Where "an 'average' hospital" refers to average overall U.S. hospitals in the time interval. To calculate the RSMR the researchers divided the "predicted 30-

day mortality" for each hospital by its "expected mortality" and multiplied the ratio by the "U.S. National mortality rate" for the condition.

Finally, the document emphasizes that "The hierarchical regression model also adjusts mortality rates results for small hospitals or hospitals with few heart attack or heart failure or pneumonia cases in a given year. ...the predicted mortality rate for a hospital with a small number of cases is moved toward the overall U.S. National mortality rate for all hospitals."[159] I obtained the RSMR dataset by Medicaire/Medicaid Provider ID from the U.S. Department of Health & Human Services via the Hospital Compare website.[160]

B.2 Distribution of the Number of Hospitals by MSA

The theory developed in Chapter 2 of this thesis has the central implication that involves relative prevalence of organizational forms (NP and FP) along the distributions of quality in one market. One of the impediments to testing this hypothesis is a small number of hospitals within geographical markets. I bound each market empirically as one MSA. Figure $B.1$ plots the distribution of the number of Medicare/Medicaid provider IDs that treated Heart Attack patients by MSA. Heart Attack is the most reliable measure of quality, as this condition requires immediate treatment, and hospital selection is unlikely to take place (McClellan and Staiger, 2000).

Figure $B.1$ shows that the majority of MSAs have six hospitals or less in total. For this reason no meaningful inference can be made with regard to the location of a single hospital along the distribution of quality. In Section 3.5 of Chapter 3 I test the direct consequence of the differences in the relative prevalence of organizational forms along the distribution of quality: differences in second moments between the distribution of quality within each organizational form.

[159] For more information refer to the original document `http://www.hospitalcompare.hhs.gov/Hospital/Static/InformationforProfessionals_tabset.asp` accessed on 24 June 2009 at 11:46.

[160] `http://www.hospitalcompare.hhs.gov/` accessed on 3 December 2008 at 13:07.

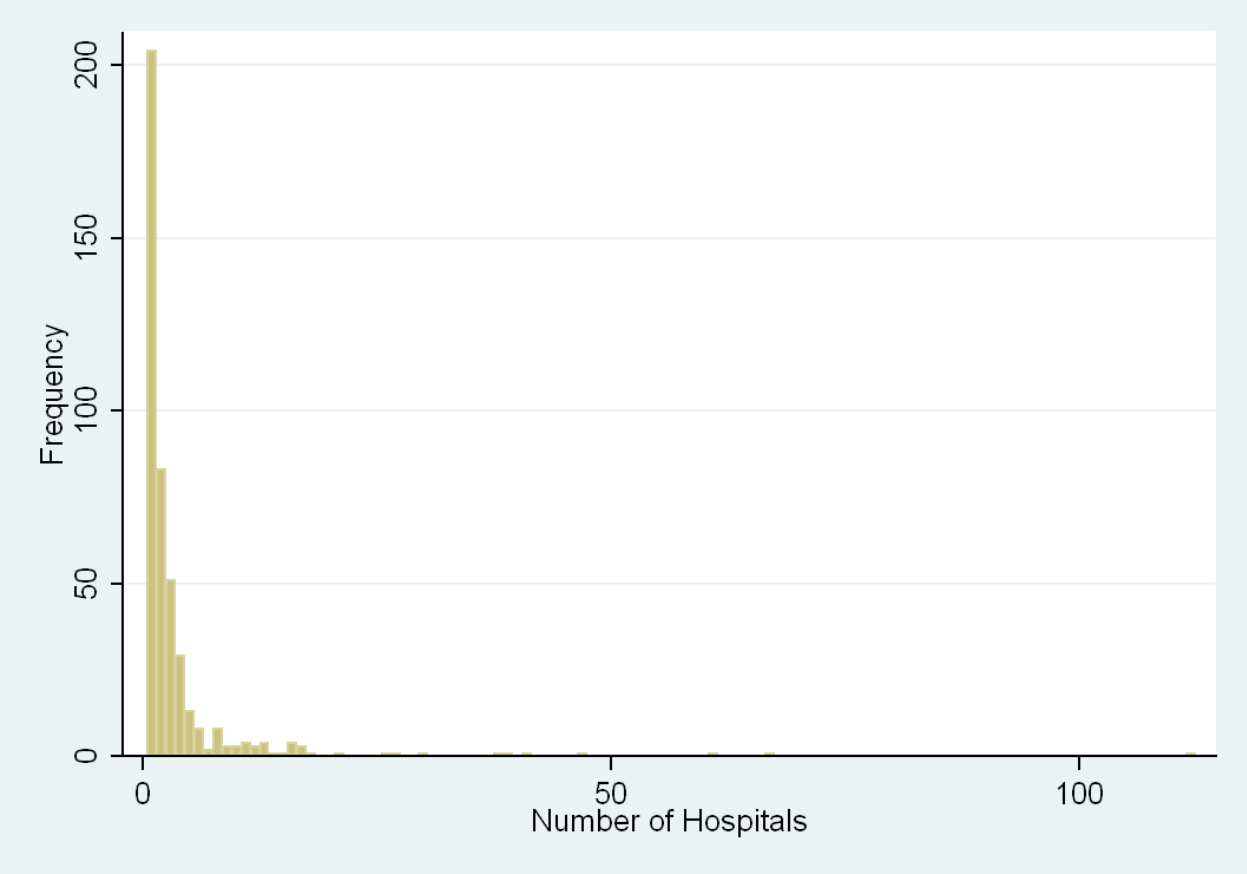


Figure B.1: Distribution of the Number of Hospitals (U.S. Medicare/Medicaid Provider IDs that treated Heart Attacks in 2006-2007) by MSA

Figure $B.2$ plots the distribution of the number of hospitals by MSA for each organizational form. This figure offers a more detailed picture of the prevalence of hospitals. One can infer that there are fewer FP and government owned hospitals within MSAs, relative to NP ones. However, there are also less MSAs that have government owned and FP hospitals. The NP organizational form seems significantly more prevalent relative to the rest.

In Section 3.5 of Chapter 3 I investigate the differences in the distributions of hospitals by quality between the organizational forms by focusing on their first and second moments. The

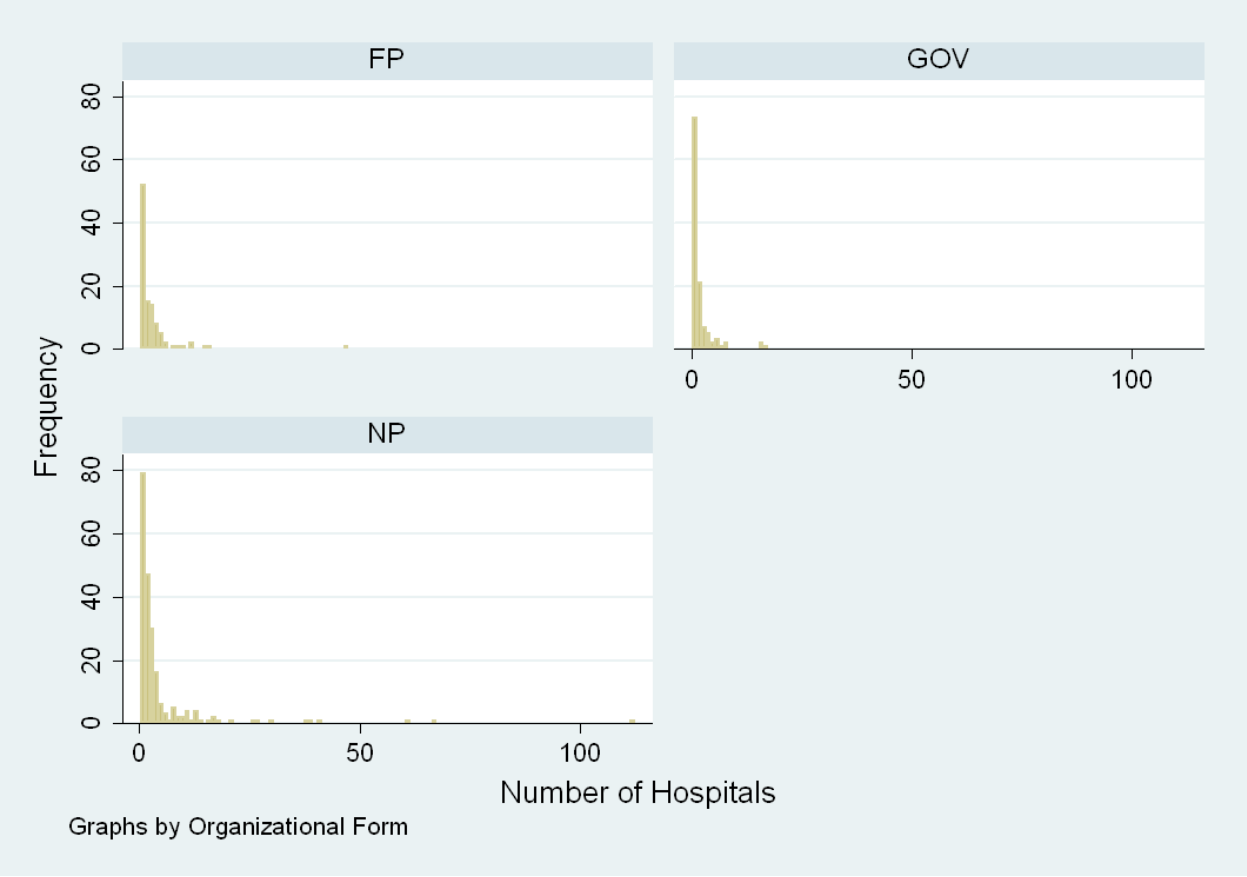


Figure B.2: Distribution of the Numbers of Provider IDs (Hospitals that treated Heart Attacks) by MSA within Org. Forms

large difference in the prevalences of the organizational forms in the sample potentially reduces the reliability of the results. Currently, only 20% of all hospitals in the U.S. are FP (Sultz and Young, 2009 p. 76). Hence, even if one obtains a larger sample by including rural hospitals, the percision of the inference is unlikely to improve.

B.3 Weighting Exploration

In the estimations I perform in Section 3.5 of Chapter 3 I apply weights to the RSMR based on Medicare/Medicaid provider ID share of patients within its MSA and organizational form (for each condition separately). One may argue that weighting the observations of RSMR by patient shares biases the results in a predictable direction. I address this concern in the current section.

Hospital selection is less likely to occur for Heart Attacks than for Heart Failures. Hospital selection by patients biases the actual mortality rates to the mean in the market. Hence, one is likely to find larger differences in mortality rates for Heart Attacks between hospitals than for Heart Failures. If the weighting biases the estimations, the impact of this bias is more likely to be higher for Heart Attacks. I explore the effect of the weighting procedure on the distribution of the mortality rates for Heart Attacks using an example from one MSA.

Figure $B.2$ in Section $B.2$ presents the distribution of the number of hospitals by MSA and organizational form. One of the most notable features of this distribution is that most MSAs have only several hospitals. Moreover, the number of FP hospitals is particularly low. For this reason it is impossible to draw a meaningful distribution of mortality rates for the majority of MSAs.

I choose an MSA with the highest number of FP hospitals in order to examine the impact of the weighting procedure on the distributions of mortality rates within the NP and FP sectors. The MSA with the highest number of FP provider IDs that treated patients for Heart Attacks was Los Angeles-Riverside-Orange County, CA, it had 47 such provider IDs. There were 61 NP provider IDs that treated Heart Attacks in this MSA as well. The FP hospitals treated the total of 1,286 patients for Heart Attacks. The NP hospitals treated the total of 2,693 for the same condition. The population of this MSA was 16.374 million in 2000.

The weights applied to the mortality rates are a function of the number of patients hospitals treated. Figure $B.3$ shows the distribution of the number of patients by hospital for each organi-

zational form in Los Angeles-Riverside-Orange County, CA. One recognizes the familiar pattern where NP hospitals tend to treat more patients.

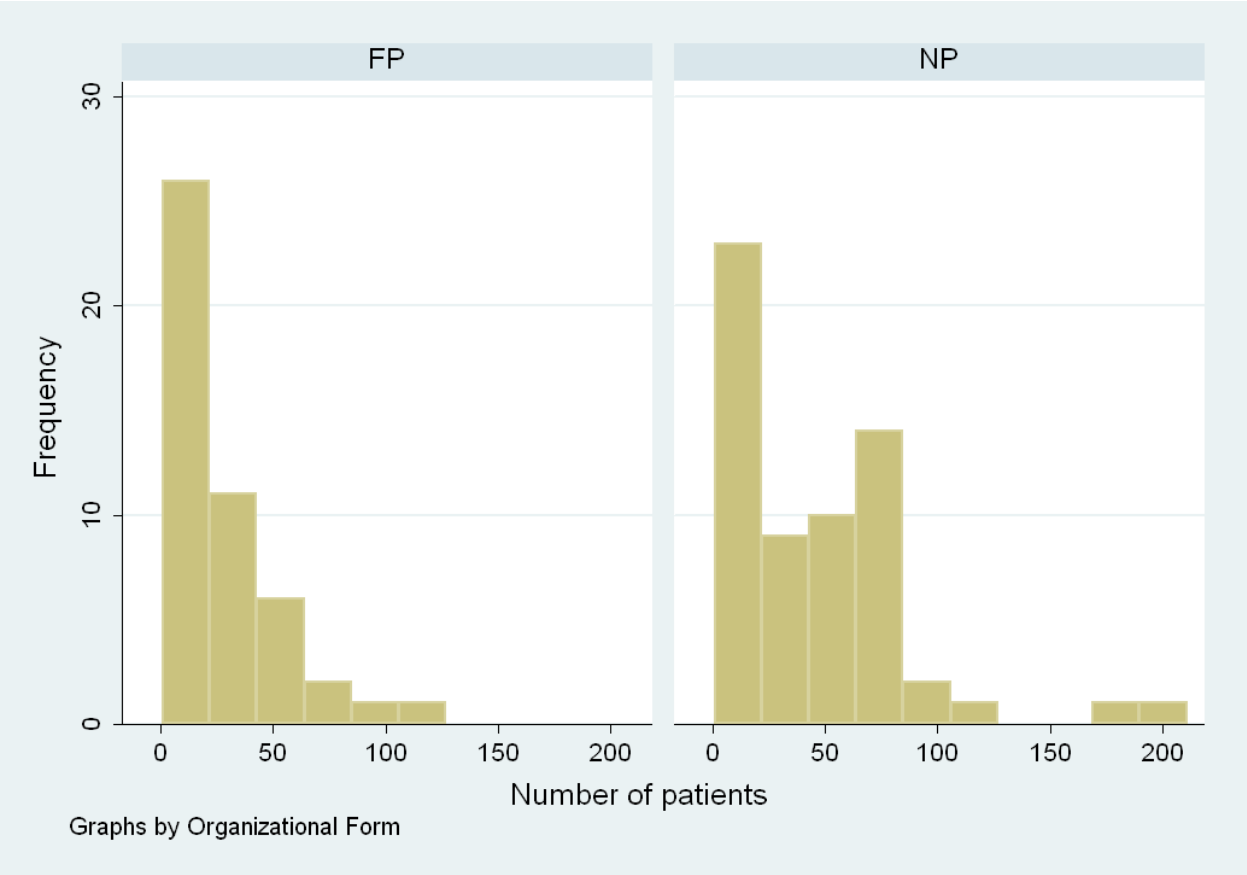


Figure B.3: Distribution of the Number of Patients by Hospital for Heart Attacks in Los Angeles-Riverside-Orange County, CA MSA by Org. Form

Figure $B.4$ plots the distribution of hospitals by mortality rates for Heart Attacks, as they were obtained from the Hospital Compare database. I separate between the two organizational forms of interest: NP and FP. In Figure $B.5$ I plot the same distributions after weighting each mortality rate by hospital's patient share within MSA and organizational form, as discussed above.

It is hard to draw strong conclusions from the graphs. The main finding is that the weighting

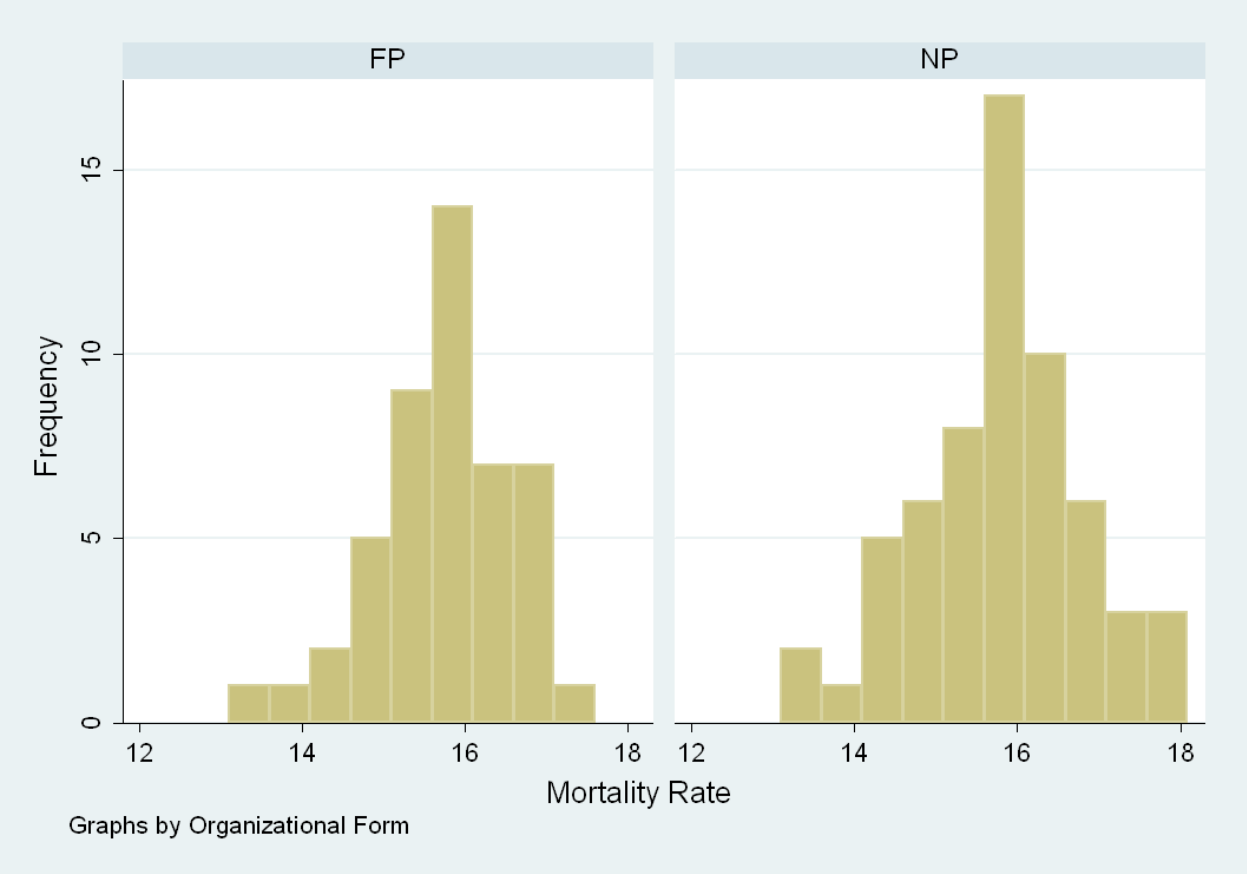


Figure B.4: Distribution of Unweighted RSMR for Heart Attacks in Los Angeles-Riverside-Orange County, CA MSA by Org. Form

procedure does not cause an obvious bias toward finding higher variance of quality withing the FP sector. Comparing the figures here with Figures 3.4 and 3.5 in Section 3.5 of Chapter 3, it may seem puzzling that there is a higher dispersion of differences in the market-level means in Figure 3.5, where RSMR were weighted, relative to Figure 3.4, where they were not weighted. In the case of Los Angeles-Riverside-Orange County, the weighting procedure seems to bound the mortality rates between zero and 1.5 for both organizational forms. Hence, the highest possible difference in mean mortalities should not exceed 1.5.

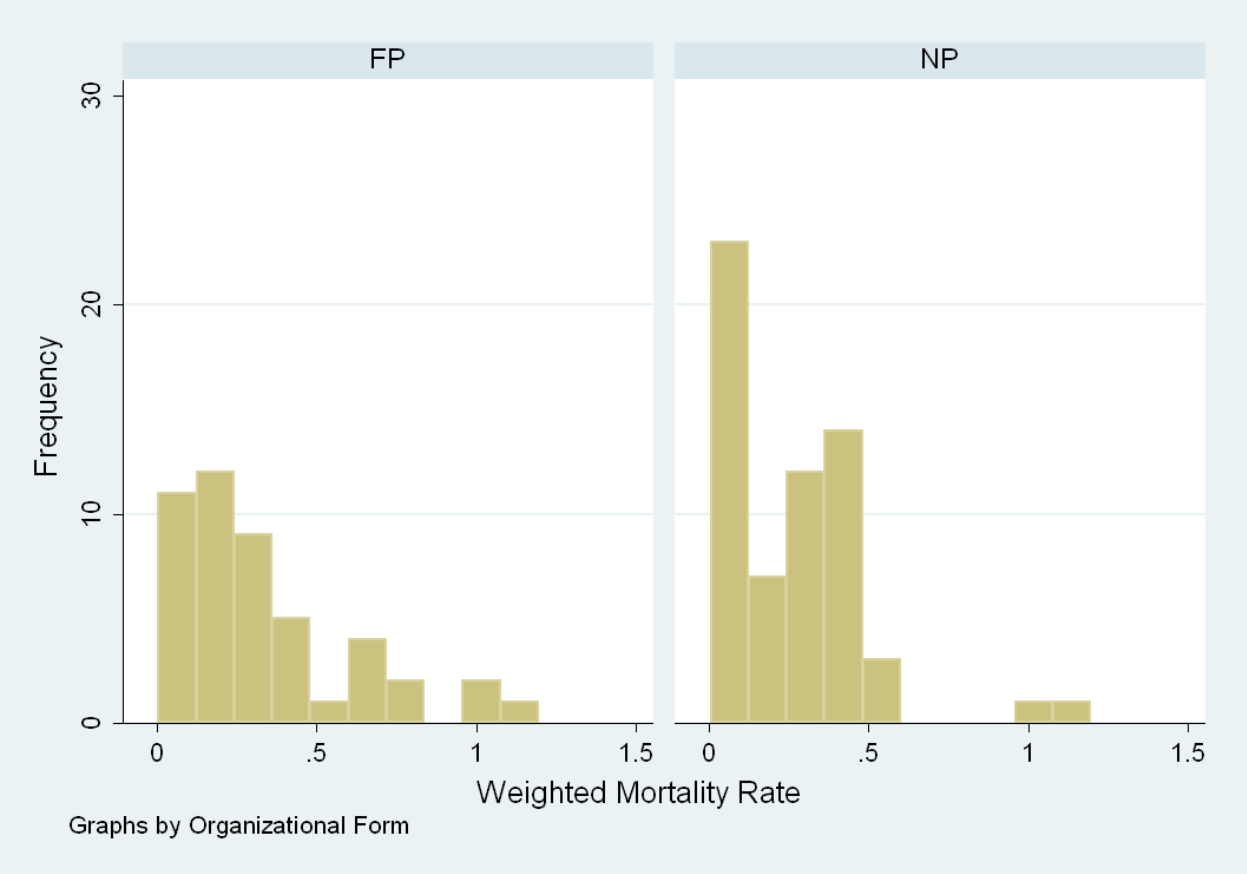


Figure B.5: Distribution of Weighted RSMR for Heart Attacks in Los Angeles-Riverside-Orange County, CA MSA by Org. Form

After re-examining the results, I am forced to conclude that it is likely that the example given here is fairly special. The large number of hospitals implies that each hospital treats a small share of patients. When hospitals are weighted by patient shares, virtually every hospital receives a low weight. The total number of patients exceeds the number of patients treated in the largest hospital more than 10 fold for both organizational forms: 1,286 total FP patients versus 108 at the largest FP; 2,693 total NP patients versus 211 at the largest NP (these numbers are extracted directly from the dataset). Hence, the differences in weights between the hospitals are small enough, such

that most hospitals are assigned a small (almost) equal weight.

In MSAs where the number of hospitals is much smaller it is often the case that one large hospital in each organizational form is significantly larger than the rest. This hospital receives most of the weight for the organizational form. Once the differences between the means are calculated (Figure 3.5) they reflect mostly the difference in mortality between the two largest (within each organizational form) hospitals. This is most likely to explain the larger differences in means between the organizational forms in Figure 3.5 relative to Figure 3.4. In the latter unweighted case the smaller hospitals provide "noise" that does not enable to find extreme differences.

The main implication of the theory developed in Chapter 2 of this thesis is *FP hospitals have higher variance (standard deviation) of quality than NP hospitals* (Hypothesis 1 in Section 3.1 of Chapter 3). The weighting has the potential to introduce a bias in the test of this implication through an artificial reduction in the variance of RSMR for NPs relative to FPs. The above example shows that this is less likely to happen. Similar analysis cannot be replicated for the majority of MSAs due to small numbers of (FP) hospitals.