

# ESSAYS ON HEALTH INSURER AND PROVIDER INTERACTIONS

(Spine title: Essays on Health Insurer and Provider Interactions)

(Thesis format: Integrated Article)

by

Daniel Montanera

Graduate Program in Economics

SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES  
THE UNIVERSITY OF WESTERN ONTARIO  
LONDON, CANADA  
OCTOBER 2012

© Daniel Montanera, 2012

THE UNIVERSITY OF WESTERN ONTARIO  
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor

\_\_\_\_\_  
Dr. Al Slivinski

Supervisory Committee

\_\_\_\_\_  
Dr. James B. Davies

\_\_\_\_\_  
Dr. Gregory S. Zaric

Examiners

\_\_\_\_\_  
Dr. John Wilson

\_\_\_\_\_  
Dr. Monali Malvankar

\_\_\_\_\_  
Dr. Peter Streufert

\_\_\_\_\_  
Dr. Elizabeth Caucutt

The thesis by

**Daniel Montanera**

entitled

**Essays on Health Insurer and Provider Interactions**

is accepted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Date \_\_\_\_\_

\_\_\_\_\_  
Chair of the Thesis Examining Board

# Abstract

This thesis consists of three chapters on insurer and provider interactions. The first addresses an important policy question: whether or not the fear of medical malpractice liability induces physicians to over-utilize medical services and/or avoid treating risky patients. Commonly known as defensive medicine, such behaviour, if it occurs, implies that liability costs borne by physicians can adversely affect the cost and quality of health care. Despite widespread reports of defensive medicine in surveys of physicians, empirical investigations have produced conflicting evidence that defensive medicine is practiced on a significant scale. In several countries, the United States in particular, this lack of empirical verification has confounded efforts to formulate, implement, and evaluate various tort reforms intended to lower costs and improve access to medical care. This paper develops an innovative model of the interaction between patients, physicians, and health insurers that provides a unified framework within which the existing empirical findings can be understood. The model generates two types of equilibrium, and predicts the opposite effects of changes in the malpractice environment on health care expenditure and quality to emerge in each type. In particular, when malpractice pressure is low, increasing pressure causes increases in both health care quality and expenditure. At high levels of malpractice liability, however, further increases in pressure induce quality and expenditure to decrease. These non-monotonic predictions provide an

explanation for the apparent conflicts and inconsistencies in the existing empirical literature. The model also provides policy guidance as the two equilibrium types are fully distinguished by the level of patient access to physicians. Thus, where measures of access (waiting times, incidences of late treatment, distance travelled for medical procedures) are at feasibly low levels, decreases in malpractice pressure cause reductions in health care expenditures and some loss of quality, but where these measures are high, the same reductions have the opposite effect. This implies that efforts at tort reform should be informed by data on patient access to physicians services in order to accurately anticipate effects on quality and expenditure.

The next chapter expands the model by making physicians mobile, and includes a location decision as a new margin of defensive behaviour. The empirical literature on the subject of defensive medicine includes studies utilizing data at the state or county level. Such jurisdictional studies typically find the least evidence that rising malpractice liability costs induce cost-increasing or quality-reducing practices by physicians. A key assumption in these studies is that changes in malpractice pressure have no cross-jurisdictional effects on health care spending and quality. This could be a strong assumption where physicians are mobile and malpractice pressure influences their location decisions. If malpractice pressure influences physicians' location decisions, then physician mobility represents a potential channel for such cross-jurisdictional effects. This paper constructs a theoretical model where insurers compete to provide consumers with health insurance while facing mobile physicians. Analytical and numerical results show that, through this mobility channel, changes in malpractice pressure unique to one jurisdiction do influence health care spending and quality in other jurisdictions. This introduces the possibility of omitted variable bias in estimates of the effects of changes in malpractice pressure, and drives a wedge between the direct and aggregate effects.

The final chapter investigates the responses of payers and providers to an

innovation in provider compensation: competitive bidding for patients. The rise of managed care and pressure to reduce health care spending have made prospective payment and capitation the dominant methods for health insurers to compensate health care providers in the United States. This can be problematic where providers (but not insurers) can observe patient heterogeneity within payment categories. This informational advantage can induce providers to practice risk selection: retaining only those patients believed to be low cost patients, and leaving expected high-cost patients without treatment or relegated to expensive and inefficient emergency room care. This problem can be modelled as a game between a principle (health insurer) and multiple agents (providers), an environment where auctions have proven useful at inducing agents to reveal private information. This paper constructs a multi-stage provider reimbursement system wherein unselected patients are allocated using competitive bidding. This modelling approach allows the analysis of the trade-off between selection and efficiency under competitive bidding. The study finds that a mixed system of competitive bidding and capitation dominates a pure bidding system, and sufficient conditions for dominance of a pure capitation system. Overall, competitive bidding eliminates risk selection in equilibrium without sacrificing efficiency.

# Acknowledgements

I wish to acknowledge the invaluable contribution to my doctoral studies of my supervisor, Dr. Al Slivinski. In addition to his guidance in assembling the following chapters of this thesis, Dr. Slivinski taught me how to recognize interesting research questions, build the models to go about answering them, and communicate the importance of my findings. He graciously accepted and corrected several years of my early attempts at writing, which is a testament to his patience. I intend to follow his example as I begin my career in academia, to the benefit of my future colleagues and students if I am able.

The other members of my supervisory committee, Jim Davies and Greg Zaric, have my gratitude for the many hours they spent on developing my research and preparing me for the job market. The faculty members in the Department of Economics; especially Ben Lester, Jim MacGee, and Charles Zheng; gave graciously of their time and have my sincere thanks. Much appreciation to Javier Cano Urbina for introducing me to LaTeX and MATLAB. Thoughtful and constructive comments on my research were provided by the attendees of the 2nd Annual Midwest Health Economics Conference, 10th Annual Meeting of the Canadian Health Economics Study Group, and the 4th Biennial Conference of the American Society of Health Economists. David Fieldhouse, Nick Bedard, Hugh Cassidy, Jon Rosborough, and Ke Xian Au Yong are overwhelming evidence that you learn as much from your classmates as you do from your professors.

*Personal Acknowledgements:* In January of this year, with much of this thesis yet to be written, my family and I suffered a tragic loss. That I have my three chapters to defend, and an assistant professorship, is due to a great deal of support. From my advisor, Al Slivinski, who went above and beyond the job description. From my parents, Walter and Mary Lou Montanera, who put the rest of us before themselves, and from my brothers Chris and Mike. From my fiancée, Jassie Grewal, whose love and faith make everything worth it. And finally, from my sister, Lisa, for every equation that balanced and term I could sign. We all miss you, love you, and will see you again someday.

*For my little sister,*

*Lisa Marie Montanera*

*August 7, 1986 – January 18, 2012*



# Table of Contents

<b>Certificate of Examination</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>Table of Contents</b>	<b>ix</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Appendices</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 The Hidden Evidence of Defensive Medicine</b>	<b>4</b>
2.1 Introduction . . . . .	4
2.2 Empirical Literature . . . . .	8
2.3 Theoretical Literature . . . . .	13
2.4 The Model . . . . .	17
2.5 Equilibrium: Physician . . . . .	25
2.6 Equilibrium: MCO . . . . .	27
2.7 Full-Access Equilibrium . . . . .	36
2.8 Limited-Access Equilibrium . . . . .	41
2.9 Discussion . . . . .	47
2.10 Summary and Implications . . . . .	53
2.11 Bibliography . . . . .	56

<b>3</b>	<b>Physician Mobility and the Differential Effects of Defensive Medicine</b>	<b>61</b>
3.1	Introduction . . . . .	61
3.2	The Model . . . . .	66
3.3	Equilibrium: Analytical . . . . .	73
3.4	Equilibrium: Numerical . . . . .	78
3.5	Results . . . . .	81
3.5.1	Part 1: Jurisdiction 2 at Full-Access . . . . .	82
3.5.2	Part 2: Jurisdiction 2 at Limited-Access . . . . .	87
3.5.3	Part 3: Heterogeneous Jurisdictions, Common Malpractice Pressure . . . . .	90
3.6	Summary and Implications . . . . .	94
3.7	Bibliography . . . . .	96
<b>4</b>	<b>Mitigating Risk Selection with Patient-Level Competitive Bidding</b>	<b>98</b>
4.1	Introduction . . . . .	98
4.2	Literature Review . . . . .	107
4.3	The Model . . . . .	111
4.4	Equilibrium: Capitation . . . . .	113
4.5	Equilibrium: Capitation-Plus . . . . .	118
4.6	Summary and Implications . . . . .	127
4.7	Bibliography . . . . .	129
<b>5</b>	<b>Conclusion</b>	<b>133</b>
<b>A</b>	<b>Chapter 2 Appendix</b>	<b>136</b>
A.1	Proof of Proposition 2.1 . . . . .	136
A.2	Proof of Proposition 2.2 . . . . .	142
A.3	Proof of Proposition 2.3 . . . . .	143
A.4	Proof of Proposition 2.4 . . . . .	145
A.5	Proof of Proposition 2.5 . . . . .	146
	<b>Curriculum Vitae</b>	<b>148</b>

# List of Tables

2.1 Summary of analytical results . . . . . 46

# List of Figures

2.1	Illustration of Assumption 1 . . . . .	31
2.2	Illustration of Assumption 2 . . . . .	31
2.3	Expansion Path . . . . .	35
2.4	Effect on Spending: Full-Access . . . . .	38
2.5	Effect on Spending: Limited-Access . . . . .	43
2.6	Numerical Example . . . . .	48
3.1	Limited-access solution . . . . .	76
3.2	Full-access-interior solution . . . . .	77
3.3	Full-access-corner solution . . . . .	77
3.4	Part 1: Rising $P_1$ , low $P_2$ . . . . .	83
3.5	Part 1: Physician movement . . . . .	84
3.6	Part 1: Effect on spending . . . . .	85
3.7	Part 1: Effect on quality . . . . .	85
3.8	Part 2: Rising $P_1$ , high $P_2$ . . . . .	87
3.9	Part 2: Physician movement . . . . .	88
3.10	Part 2: Effect on spending . . . . .	89
3.11	Part 2: Effect on quality . . . . .	89
3.12	Part 3: Common Malpractice Pressure . . . . .	91
3.13	Part 3: Physician movement . . . . .	92
3.14	Part 3: Effect on spending . . . . .	93
3.15	Part 3: Effect on quality . . . . .	93

4.1	Capitation system spending . . . . .	116
4.2	Capitation-plus system spending . . . . .	123
4.3	Spending differences between systems . . . . .	126
A.1	Existence . . . . .	138
A.2	Uniqueness . . . . .	140

# List of Appendices

- A.1 Proof of Proposition 2.1 . . . . . 141
- A.2 Proof of Proposition 2.2 . . . . . 143
- A.3 Proof of Proposition 2.3 . . . . . 145
- A.4 Proof of Proposition 2.4 . . . . . 146
- A.5 Proof of Proposition 2.5 . . . . . 147

# Chapter 1

## Introduction

A consistent, ongoing debate in countries all over the developed world concerns both the level of and growth in health care expenditure. Dissatisfaction is routinely expressed with the burden of health care expenditure on government budgets, the allocation of scarce medical resources and services, and the level of access to care enjoyed by vulnerable populations. Despite decades of debate and research into various ways to structure health care systems, and a variety of alternatives attempted in the past or in use today, no dominant structure has emerged. Further complicating the debate is the lack of a clear positive relationship between the amount of health care expenditure and various measures of system quality. This in turn leads to speculation that health care systems, as structured today, contain inefficiencies that produce suboptimal outcomes, and that “throwing money” at the problem is not a viable solution.

Economists have been making contributions to this debate for at least fifty

years. Kenneth Arrow, in his seminal paper *Uncertainty and the Welfare Economics of Medical Care* (1963), outlines several characteristics that make health care “special”, and distinct from other goods and services. As the title indicates, much of the distinction is made in the degree of uncertainty faced by health insurers, health care providers, and patients as they interact in health care systems. In the decades since this article’s publication, those working in the field of health economics have produced further insights into the nature of this uncertainty. In many cases, the uncertainty faced by one party in a health care system concerns private information held by another party. For example, health insurers must form expectations over the cost of a policyholder’s treatment, while policyholders may have some idea of their idiosyncratic health risk, or preference for utilizing health services. Depending on the structure of a health care system, policyholders can find it in their best interest to strategically withhold this private information from other parties. As Arrow’s article points out, these information problems call into question whether competitive market create efficient outcomes, and whether alternative structures can produce improvements.

The three chapters comprising this thesis are intended as a contribution to the study of the information and commitment problems that set health care apart from other goods and services, particularly those that arise during the interactions between health insurers or payers and health care providers. They use game theory to model these interactions, and examine the consequences



to the cost and quality of health care when these interactions are shocked by changes in external factors. The external factor examined in the first two chapters is the medical malpractice liability costs faced by physicians. These models investigate physician behaviour in response to changes in malpractice costs, and how health insurers would compensate for this behaviour by adjusting contracts and affecting physician incentives. The behaviour of both parties reveals predictions of the effects of changes in malpractice liability costs on health care spending and quality. With these predictions come the first explanation for seeming inconsistencies and discrepancies in the results of dozens of empirical investigations into these relationships, and offer clues regarding alternative model specifications for future studies. The third chapter investigates the selection problems that can arise when differences in payers' reimbursement rates do not capture all underlying heterogeneity across patients in expected treatment costs. It develops a model for comparing reimbursement by capitation with one that utilizes competitive bidding in the allocation of patients across providers. It shows the usefulness of competitive bidding in overcoming this selection problem, and reveals the conditions under which a system with competitive bidding provides higher quality and lower cost health insurance than a pure capitation system.

## **Chapter 2**

# **The Hidden Evidence of Defensive Medicine**

### **2.1 Introduction**

An important question among researchers and policymakers is whether or not medical malpractice liability costs affect either the cost or quality of consumer health insurance. Liability costs could affect health care spending directly, through the passing of physicians' malpractice insurance premiums on to patients and health insurers, or indirectly, by changing the way physicians practice medicine. The sum of all malpractice premiums each year makes up less than two percent of total annual health care spending. Therefore, if there exist any policy-relevant relationships between malpractice liability costs and

health care spending and quality, they must arise due to the effects on physician behaviour. The treatment decisions made by physicians primarily to avoid malpractice liability rather than benefit patients are commonly known as defensive medicine. Positive defensive medicine is the over-utilization of medical services in an effort to forestall claims of negligence, while negative defensive medicine is the avoidance of risky patients or procedures believed likely to result in malpractice claims. Such behaviour illustrates the potential for adverse relationships to exist between the liability costs born by physicians, often called “malpractice pressure”, and both the cost and quality of the health care enjoyed by consumers.

Widespread reports of defensive medicine in surveys of physicians have led to calls for liability-reducing reforms. The two most recent US Presidents, George W. Bush and Barack Obama, both promoted malpractice reform as a legitimate policy option for reducing health care spending in the 2007 and 2011 State of the Union Addresses, respectively. The goal of such policies is the reduction of wasteful medical spending, making health insurance more affordable and thus improving consumers’ access to care. Researchers have investigated the benefits of malpractice reform for two decades using utilization, spending, and quality data. Unlike the physician surveys, however, these studies have produced inconsistent and often conflicting findings. Depending on

the populations studied and measures used for key variables, different studies have uncovered positive, negative, and even no statistically significant relationships between malpractice pressure and spending or quality. Without consistent findings in the data, empirical evidence to corroborate the physician surveys has remained elusive. In several countries, the United States in particular, this lack of empirical verification has confounded efforts to formulate, implement, and evaluate tort reforms. As the US is in the midst of healthcare reform, and spending on health care approaches 20% of GDP, insight into the practice and effects of defensive medicine would be timely.

Both the discrepancy between physician surveys and empirical studies, as well as the mixed empirical findings, have been noted in the literature (Helland & Showalter 2009, Lakdawalla & Seabury 2009, Sloan & Shadle 2009, Avraham & Shanzenbach 2010, Reyes 2010, Cotet 2012). An explanation for these discrepancies, however, has been lacking. Aggregation could play a role, since studies utilizing broad data sets tend to find weak relationships while studies focused on particular medical specialties, patient demographics, or geographic regions tend to find strong relationships that conflict with one another. If the relationships vary qualitatively across narrowly defined populations, then the aggregation of populations in a broad study would cancel and produce weak or inconclusive relationships over the entire sample. This does not explain why changes in malpractice pressure would affect one population differently than another. Without knowing this, it is difficult to understand the differences in

the existing empirical findings or predict how different populations would be affected by proposed malpractice reforms.

This is the first paper to provide an explanation for these discrepancies. It presents an innovative model of the interaction between patients, physicians, and health insurers that provides a unified framework within which the existing empirical findings can be understood. Ultimately, the model is a decision problem on the part of an insurer involving whether and how to structure its contracts with physicians as malpractice pressure increases. There are three main analytical predictions of the model. First, two types of equilibrium can occur, one of which is a corner solution and the other an interior solution. Second, these two equilibrium types exhibit the opposite comparative statics from one another concerning malpractice pressure and both health care cost and quality. Finally, rising malpractice pressure can trigger a switch from one type of equilibrium to the other. This creates a non-monotonicity in the relationships between malpractice pressure and both health care quality and expenditure. Specifically, both health care spending and quality rise initially with malpractice pressure. These effects are positive up to a threshold, after which any further increases in pressure produce negative effects. These relationships are non-monotonic despite both the incentive for physicians to behave defensively and the actual practice of defensive medicine in equilibrium at all levels of malpractice pressure. It is the non-monotonicity that can leave evidence of defensive medicine hidden from empirical methods designed to investigate for

monotonic relationships.

These predictions are broadly consistent with the literature. Data focusing on specialties facing relatively low malpractice pressure tend to report positive relationships, while those with high malpractice pressure reveal negative relationships. In addition to the explanation for the empirical discrepancies, the model suggests a novel use for measures of access to distinguish observations in one type of equilibrium from those in the other, as well as the prevailing equilibrium type of a given population. This is useful in specifying models to evaluate past tort reforms and form expectations of the effects of any future reforms.

## **2.2 Empirical Literature**

Surveys of physicians regularly report the widespread practice of defensive medicine. Over 90% of physician respondents over-utilize medical resources due to liability concerns (Studdert et al 2005, Dove et al 2010, Paik et al 2011, Sethi et al 2012). In other surveys, between 16% and 64% of physicians limit the performance of high-risk procedures or avoid patients considered more likely to file claims (Bovbjerg et al 1996, Lumalcuri & Hale 2010, Reyes 2010). Fear of liability has been the most cited reason for stopping the practice of obstetrics (Lumalcuri & Hale 2010), and for medical residents to relocate (Mello & Kelly 2005). While these studies note that the high profile of malpractice

premiums in medical circles could inflate the importance of defensive medicine in survey responses, they indicate a consistent and genuine belief that liability costs produce adverse effects on the cost and quality of health care. It may therefore follow that reforms designed to reduce malpractice pressures would spur reductions in wasteful medical procedures, while also increasing access to care for high-risk procedures and patients. Based on this, the American Medical Association advocates for malpractice reform under the belief that doing so would significantly slow the rising cost of health care in the US while also increasing health care quality through improved access. Taken together, the survey responses and efforts to reform the malpractice system reflect a widely held belief in the medical establishment that defensive medicine is a serious problem and that malpractice reform would pay considerable dividends.

Contrary to physician surveys, the results of empirical studies investigating the practice of defensive medicine demonstrate considerable inconsistency. Over a span of two decades, empirical works have uncovered mixed evidence of any policy-relevant relationships between malpractice pressure and either the cost or quality of health care. Baicker & Chandra (2005) found little evidence that malpractice pressure affects physicians' location decisions or treatment choices. Later analysis turned up significant but relatively modest effects. At the state level, elasticities of total Medicare spending per beneficiary to malpractice premiums or payments are estimated at 0.06 to 0.1 (Baicker

et al 2007). Hellinger & Encinosa (2006) found that US states adopting legislation capping awards for noneconomic damages due to medical malpractice experienced between 3.25% and 3.4% lower per-capita health care expenditure than states without caps. Also utilizing state-level tort reform as exogenous reductions in malpractice pressure, Sloan & Shadle (2009) found that reforms generally had no significant effect on expenditure or quality of care enjoyed by Medicare patients suffering from heart attack, breast cancer, diabetes, or stroke. These studies suggest limited scope for tort reform to produce reductions in health care expenditure on a national scale.

The literature also contains studies finding relatively large effects of malpractice pressure on health care cost and quality. Kessler & McClellan (1996) found that certain tort reforms were able to lower hospital expenditures on elderly Medicare patients suffering heart attack and heart disease between 5% and 9%, without significantly affecting the quality of care. A later study (Kessler & McClellan 2002) uncovered similar effects, although stronger for diagnostic rather than therapeutic services. The authors interpret these findings as evidence that defensive medicine is practiced on a significant scale, and their estimates are heavily cited in arguments promoting the benefits of malpractice reform. Other studies produce similar large effects on health care spending. A recent working paper (Lakdawalla & Seabury 2009) utilizing county-level data mostly from California and New York found that rising malpractice pressure



was responsible for approximately 15% of the growth in health care spending over the 1990s. The authors also estimate, however, that the cost-savings would be outweighed by reductions in health care quality. State-level reforms designed to reduce liability exposure have been found to reduce various measures of health care utilization, such as hospital admissions, surgeries, outpatient visits (Cotet 2012), mechanical ventilation and length of stay in hospital (Moriya 2011). Overall, these studies indicate a positive effect of malpractice pressure on the cost of health care, and at least a weakly positive effect on health care quality through increased utilization.

In contrast to the studies positive effects, several others examining obstetrics have found the effects to be significantly negative. A widely cited paper using national US data (Dubay et al 2001) found that higher malpractice premiums were associated with a lower incidence of timely prenatal care. Where the timing of care is a vital component of health care quality, these findings indicate tort reforms would deliver improvements in quality rather than reductions. Delivery by cesarean section is often associated with defensive behaviour since it is an expensive alternative to vaginal birth and is widely believed among physicians to reduce the complications most likely to result in malpractice claims (Yang et al 2009). State-specific investigations into an effect of malpractice pressure on the use of cesarean section vary considerably, from positive (Localio et al 1993, Yang et al 2009, Shurtz 2010) to negative (Tussing & Wojtowycz 1992) to no apparent relationship (Baldwin et al 1995).

A well-cited national study by Currie & MacLeod (2008) found that limits on noneconomic damage awards increase the use of cesarean section, and tort reforms thought to make physicians more legally responsible for their own treatment decisions had the opposite effect. The negative relationships found indicate that, in at least some cases, malpractice pressure reduces expenditure in obstetrics by decreasing the use of expensive alternative treatments. Given the findings from studies using data from obstetrics, tort reforms could produce the opposite of their expected effect in certain medical specialties. This has created questions as to whether any benefits of malpractice reform are generalizable across patient populations (Congressional Budget Office 2004).

A relatively consistent aspect of the literature is the effect of changes in malpractice pressure on patients' access to care. The relationship appears to be weakly negative where access is measured in timeliness of prenatal care (Dubay et al 2001), physicians' hours worked (Helland & Showalter 2009), and rates of consumer health insurance coverage (Avraham & Shanzenbach 2010). The relationships become more negative as patients' socioeconomic status declines (Dubay et al 2001) and for patients living in rural areas (General Accounting Office 2003). While the empirical findings on access to care fit qualitatively with physicians' survey responses, the inconsistent and often conflicting results on spending and quality have led policymakers to question whether defensive medicine is an important factor in rising health care expenditure. Empirical work on the importance of defensive medicine and the viability of

malpractice reform continues, notably with recent advances in the anticipation effects of tort reform (Malani & Reif 2010), as well as physicians' sensitivity to framing in surveys and long-run specialization effects (Reyes 2010). Based on the existing body of empirical work, however, the evidence on the relationships between malpractice pressure and both the cost and quality of consumer health care is unclear, and certainly at odds with the beliefs of the medical establishment.

## **2.3 Theoretical Literature**

A theoretical basis for the practice of defensive medicine has been established in the literature. Medical malpractice is classified under tort law, and under English common law, is subject to a negligence rule of liability. Health care providers as defendants can be held liable if plaintiffs can show that an injury occurred, that it was caused by the provider's medical care, and that such care deviated from due care (Danzon 2001). The law therefore recognizes that injuries can result from medical care, and that the cost of those injuries should be born by health care providers if the care was negligent, and alternatively by patients if the care was non-negligent. Economic theory shows that where physicians are held perfectly accountable for their treatment decisions under

a negligence rule, sufficiently high malpractice penalties would induce physicians to meet the standard of care and have no demand for malpractice insurance (Danzon 1985). Physicians, however, tend to demand full insurance, to the point of forming physician-owned mutuals to secure coverage whenever it became scarce in traditional markets (Danzon et al 2004). Real-world physician behaviour, therefore, differs from that predicted under a well-functioning negligence rule.

As Olbrich (2008) shows, evidentiary uncertainty and positive lawsuit costs can explain this behaviour. The standard of care itself is subjective, as it consists of the care that a competent physician would reasonably provide in the same circumstances. The complexity of medical care makes this standard unverifiable, and the reliance on beliefs and opinion thus introduces a random element onto patients' decisions to sue and the court's judgement of negligence. Determinations of liability depend more on economic loss (Feess 2012) and treatment success (Kessler 2011) rather than fault, and legal precedents allow juries to find even customary medical practice as unreasonable, and therefore negligent (Studdert et al 2004, Moffett & Moore 2011). In addition to the damages that a physician could be forced to pay, lawsuit costs include the legal costs of their defense as well as the reputational and emotional costs associated with an accusation of professional negligence (Keren-Paz 2010). Therefore, due to uncertainty and lawsuit costs, each patient the physician treats represents an

expected liability cost. The physician's treatment choice could reduce this liability cost, but it may be impossible or extremely costly to eliminate it entirely. This is the way malpractice liability costs are modeled in this paper, as well as in Gal-Or (1999), Currie & MacLeod (2008), Olbrich (2008), and Feess (2012). Rising malpractice costs could thus induce physicians to engage in defensive practices to lower their exposure to liability.

This paper follows several theoretical works modeling patients, health insurers, and physicians under the threat of medical malpractice. Examples include Danzon (1985), Gal-Or (1999), Leger (2000), Zeiler (2004), Arlen & MacLeod (2005), Olbrich (2008), Kpelitse (2010), and Feess (2012). The model presented here departs from these models in three ways. First, this paper is positive in focus whereas the objectives in the existing literature tend to be normative, such as determining optimal malpractice costs, liability rules, or contracts between physicians and insurers. While certainly addressing important issues with broad policy implications, these studies nonetheless investigate different questions than are posed in the empirical literature, and are thus of limited use in explaining the discrepancies that form the basis of this paper. Second, the workhorse principal-agent framework used in the literature contains an implicit assumption on the extensive margin of physician behaviour. Typically, a single insurer (principal) induces a single physician (agent) to provide a certain quality of treatment to a single patient. This

framework is useful for examining health care quality on the intensive margin, but implicitly assumes the number of patients treated by each physician to be determined exogenously. This is a strong assumption given the frequency with which physicians report choosing to limit the performance of certain procedures or refraining from treating certain types of patients due to the fear of malpractice liability. In order to investigate these choices, this model endogenizes physicians' caseloads as well as the intensity of treatment provided to the patient<sup>1</sup>. Finally, existing studies tend to model information asymmetry between physicians and patients or health insurers as a basis for the contractibility problems that necessitate the use of the principal-agent framework. Physicians tend to have private information over the health status of a given patient, thus requiring incentive compatibility constraints on any contract guaranteeing the appropriate treatment for a patient of a particular health status. In trading-off realism for tractability, this paper assumes that physicians and patients are homogenous and thus does not contain information asymmetry. The model instead exogenously assumes the contractibility problems that arise due to real-world information asymmetry without explicitly modelling it. Therefore, like the other principal-agent frameworks in the literature, health insurers are able to influence physician behaviour only through incentives rather than commands.

---

<sup>1</sup>Feess (2012) is another exception in that physicians choose both a degree of care for each patient in addition to a choice of technology choice, which could be interpreted as a choice whether to treat at all.

## 2.4 The Model

There is a continuum of identical consumers of measure 1, each endowed with income  $m$ . Consumers have the utility function  $U(H, y)$ , where  $H$  is health status and  $y$  is consumption, with strictly positive first and cross-partial derivatives<sup>2</sup>, as well as strictly negative second derivatives. Each consumer expects to become ill with probability  $q$  and remain healthy with probability  $1 - q$ . A healthy consumer enjoys a health status of  $H_1$  while an ill consumer enjoys  $H_2$ , where  $H_1$  is strictly greater than  $H_2$ . Consumers are willing to purchase health insurance as long as the expected utility ( $EU$ ) of owning insurance is at least as large as the expected utility of going uninsured. The parameters  $m$ ,  $q$ ,  $H_1$ ,  $H_2$ , and the function  $U(\cdot)$  are also common knowledge. Since consumers are identical, a strict preference toward health insurance for a single consumer results in the entire measure of consumers purchasing health insurance, with exactly  $q$  insured consumers expected to become ill and seek treatment.

If any consumer  $i$  with insurance falls ill, he is able to obtain treatment ( $t_i$ ) from a physician and will recover with probability  $1 - \rho(t_i)$ , but suffer an adverse outcome with probability  $\rho(t_i)$ , a function that is common knowledge. Ill consumers without insurance have zero probability of recovery. The function  $\rho(t_i)$  is positive, decreasing, strictly convex, continuous, and differentiable with at least the first, second, and third derivatives being finite for any  $t_i$ . Assume that  $\rho(t_i)$  is equal to 1 when  $t_i = 0$  and approaches zero as  $t_i$  approaches infinity.

---

<sup>2</sup>For an empirical investigation into these assumptions, see Finkelstein et al (2008).

Note that these assumptions imply that the first derivative of  $\rho(t_i)$  is negative and the second derivative is positive, while all derivatives approach zero as  $t$  increases. Since  $\rho(t_i)$  is always decreasing, the model does not allow treatment to become harmful to the patient at any level. For this reason, large increases in treatment are better considered as “gold-plating”; costly inputs that deliver marginal improvements in care, rather than the potentially harmful continued administration of any one particular medical treatment.

There is a continuum of identical physicians of measure  $D$ , which is common knowledge. There are two margins of physician behaviour of interest in this paper: the number of patients that each physician chooses to treat, as well as the amount of treatment the physician devotes to each patient. Given the significant barriers to entry into the medical profession, and the evidence that physicians are able to exercise some market power (Thurston 2001), the measure of physicians is determined exogenously rather than by a clearing condition in the physician labour market. A fixed measure of physicians prohibits this model from addressing questions of physician mobility or exit in response to changing malpractice pressure. This is another interesting margin of behaviour that will be explored in the next chapter<sup>3</sup>.

Physicians are risk-neutral income maximizers. Each one receives a revenue  $w$  from an insurer for each policyholder treated, as well as a stock of

---

<sup>3</sup>Also see Kessler et al 2005, Mello et al. 2005, Klick & Stratman 2007, and Matsa 2007.



resources ( $s$ ) for use in the treatment of policyholders. Resources are a composite of all rivalrous inputs to medical care that can be procured by an insurer and transferred to a physician. Examples include global budgets (Gaynor et al 2004), access to networks of other medical contractors (Arlen & MacLeod 2005), or the effective units of time a physician gains through improved support staff, technology, or software. The pair  $\{w, s\}$  should be considered a reduced form version of the revenues and resources from any of the main explicit methods of payment, such as prospective payment or fee-for-service. Intuitively, any explicit contract should leave a physician with a sense of the revenue they can expect to realize from taking on a patient, as well as the resources at their disposal for their patients' treatment. This allows the model to abstract away from the finer details of any one explicit contract and make the incentive compatibility constraints in the insurer's problem more tractable. Individual physicians are small relative to a health insurer, and therefore cannot behave strategically, taking both  $w$  and  $s$  as given.

Physicians incur an expected malpractice liability cost for each patient treated. Even though most real-world physicians hold malpractice insurance against legal costs and any settlements or payments, there remain considerable uninsured costs associated with each claim. These include the reputational and emotional cost due to an accusation of professional negligence, the time and inconvenience of participating in negotiations or trials, and the prospect that an award may exceed the limits of coverage. Any insured costs would enter

the physician's problem as a fixed cost in the form of a malpractice insurance premium. Since the measure of physicians is fixed exogenously, any fixed costs would have no effect in this model, thus insurable costs are normalized to zero.

Given the conditions necessary for a finding of negligence, liability costs are the result of a sequence of events. First, an injury from medical care must occur, which happens in the model with probability  $\rho(t_i)$ . Conditional on an injury, there is some probability that the patient initiates a malpractice claim. Conditional on the filing of a claim, potential outcomes include a dropping of the claim, a settlement, a judgement in favour of the defendant, or a judgement in favour of the plaintiff, each with an accompanying cost to the physician. Accounting for the effects of the physician's treatment decision on the relative probabilities and costs of each potential outcome adds several terms to the physician's problem, and greatly complicates the comparative statics. Instead, as in Currie & MacLeod (2008), liability costs are assumed to enter the physician's objective function in a reduced-form way. Let the uninsurable expected liability cost of treating patient  $i$  be  $g(t_i, P)$ ; a function of the resources dedicated to the patient by the physician ( $t_i$ ), and a parameter serving as a measure of malpractice pressure ( $P$ ), which is common knowledge. Assume that  $g(t_i, P)$  has the form:

$$g(t_i, P) = \rho(t_i) \cdot P$$

This form assumes that physicians understand the effects of their treatment decisions on the likelihood of an adverse outcome. However, both the probability of an accompanying malpractice claim, as well as the expected cost of the claim outcome, enter the problem through the parameter  $P$ , which is unchanging in the amount of treatment provided. While it is likely that these probabilities and costs would change with the amount of treatment, the relative infrequency of malpractice claims makes it unlikely that physicians completely understand these effects and incorporate them into their decisions. Published statistics and word of mouth are a more likely source of information allowing physicians to form expectations of the prevailing frequency of malpractice claims and the resulting costs, and both are independent of the amount of treatment provided by an individual physician.

Each physician chooses the size of his patient roll ( $n$ ) and the amount of treatment to provide to each patient ( $t_i$ ) as a share of available resources ( $s$ ). Due to the convexity of  $\rho(t_i)$ , expected liability costs are at their lowest for any number of patients when all patients receive an equal amount of treatment. This means that  $t_i = \frac{s}{n}$  for all  $i$  in equilibrium. Given values of  $w$  and  $s$ , each physician solves the problem:

$$\max_{n \geq 0} \left\{ wn - n \cdot g\left(\frac{s}{n}, P\right) \right\}$$

Let

$$n^* = n^*(w, s, P) \in \arg \max_{n \geq 0} \left\{ wn - n \cdot g\left(\frac{s}{n}, P\right) \right\}$$

$$\tilde{n} = \tilde{n}(w, s, P) = \min \left\{ n^*, \frac{q}{D} \right\}$$

$n^*(w, s, P)$  is the number of patients the physician would like to treat given any contract  $\{w, s\}$  and level of malpractice pressure  $P$ . Since the expected number of ill consumers is  $q$ , the maximum number of patients any of the  $D$  identically-behaving physicians could treat is  $\frac{q}{D}$ . Therefore, the number of patients the physician would *actually* treat given any  $\{w, s\}$  and  $P$  is  $\tilde{n}(w, s, P)$ . The physician's total liability costs are a convex function in  $n$ , so  $n^*$  and  $\tilde{n}$  are unique.

The final decision maker is a managed care organization (MCO). Consistent with the margins of physician behaviour of interest here, the assumption of a single firm avoids issues of physician mobility between competing MCOs. This firm offers consumers a health insurance policy delivering a probability of recovery equal to  $Q$  (hereafter referred to as “quality”) at a policy price of  $\tau$ . Recovery is the result of two events. First, the consumer must obtain a place on a physician's patient roll. In the absence of a complicated matching process or search frictions, pairings occur with probability equal to the total number of places available ( $Dn$ ) divided by the total number of ill consumers seeking placement ( $q$ ). As long as the former is not greater than the latter, this is a

probability serving as the level of “access” in the health care system. Second, conditional on success with the first event, the patient receives treatment and recovers with probability  $1 - \rho(t)$ . Quality,  $Q(n, t) = \frac{Dn}{q}(1 - \rho(t))$ , is thus the product of these two probabilities.

It is assumed that the MCO and physicians cannot contract directly on either  $n$  or  $t$ . Such a contract would amount to both a quota and state contingency, both of which are made prohibitively difficult in the real world by the vast heterogeneity across patients, illnesses, and treatment options. Instead, actual contracts require physicians to exercise professional judgement in determining whether and what kind of treatment is required. Despite the lack of heterogeneity in this model, the noncontractibility assumption is intended to preserve this aspect of real-world medical care.

A perfectly competitive market for health insurance (also assumed in Arlen & MacLeod 2005) determines the objective of the MCO, as well as its constraints. The MCO must offer an insurance policy that maximizes expected consumer utility subject to a zero profit constraint. Any other bundle, whether one that brought non-zero profits to the MCO or one delivering a lower  $EU$  to consumers without violating the zero profit constraint, would induce firm entry or exit and thus cannot hold in equilibrium. In order to be credible, the quality of the insurance policy must be incentive compatible with physicians’ behaviour under the MCO’s contract  $\{w, s\}$ , meaning  $n = \tilde{n}$  in the insurer’s problem. The firm acquires resources at a marginal cost of  $c$ , which is assumed

to be constant. The MCO's problem is thus:

$$\max_{w,s,\tau} \left\{ EU(\tilde{Q}, \tau) = \text{prob}(H_1) \cdot U(H_1, m - \tau) + \text{prob}(H_2) \cdot U(H_2, m - \tau) \right\}$$

$$\text{subject to } \tau = Dw\tilde{n} + Dcs$$

$$\text{where } \text{prob}(H_1) = (1 - q) + q\tilde{Q}$$

$$\text{prob}(H_2) = q - q\tilde{Q}$$

$$\tilde{Q} = Q\left(\tilde{n}, \frac{s}{\tilde{n}}\right) = \frac{D\tilde{n}}{q} \left[ 1 - \rho\left(\frac{s}{\tilde{n}}\right) \right]$$

Notice that, since the insurer's choice set contains  $\{w, s, \tau\} = \{0, 0, 0\}$ , a policy  $\{\tilde{Q}, \tau\} = \{0, 0\}$  can always be delivered that is equivalent to no insurance. Thus, the solution to the MCO's problem will always be such that consumers weakly prefer purchasing health insurance over going uninsured. Also note that the redistributive aspect of medical malpractice is ignored, since none of the physician's liability costs appear as increased consumption for those patients suffering adverse outcomes. This potential benefit of the malpractice system is not modelled due to the high loading charge of malpractice disputes as a method of social insurance (Danzon 2001), although other models have accounted for it (Gal-Or 1999).

This environment can be analyzed in the form of a decision problem from the perspective of the MCO. The firm maximizes expected consumer utility by

offering consumers a health insurance policy at a certain price and of a certain quality. In order to obtain such quality, the MCO must set a contract to procure physicians' services, while also accounting for how the structure of that contract affects physicians' treatment decisions. Given any set of parameters  $\{q, m, D, c, H_1, H_2, P\}$ , the solution to this problem is defined as a contract between the MCO and consumers  $\{\tau^*\}$ , a contract between the MCO and physicians  $\{w^*, s^*\}$ , and a choice of patient roll size given any contract with the MCO  $\{n^*(w, s, P)\}$  such that:

$$\begin{aligned}
 1. \quad & n^*(w, s, P) = \arg \max_{n \geq 0} \left\{ wn - n \cdot g\left(\frac{s}{n}, P\right) \right\} \\
 2. \quad & \{w^*, s^*, \tau^*\} \in \arg \max_{w, s, \tau} \left\{ EU(\tilde{Q}, \tau) \mid \tau = Dw\tilde{n} + Dcs \right\}
 \end{aligned}$$

where  $\tilde{n}^* = \min \left\{ n^*(w^*, s^*, P), \frac{q}{D} \right\}$  is the number of patients each physician treats in equilibrium. Also, let  $t^* = \frac{s^*}{\tilde{n}^*}$  and  $\tilde{Q}^* = \frac{D\tilde{n}^*}{q} [1 - \rho(\frac{s^*}{\tilde{n}^*})]$  be equilibrium values of treatment and system quality respectively.

## 2.5 Equilibrium: Physician

As this paper is focused on the changes in equilibrium outcomes induced by changes in malpractice pressure ( $P$ ), the effects of changes in the parameters  $q, m, D, c, H_1$ , and  $H_2$  are not investigated here, and are largely suppressed in the notation.

Solving backward, the first problem to be considered is the physician's. The solution to the physician's problem ( $n^*$ ) is the  $n$  satisfying:

$$\frac{w}{P} = \rho\left(\frac{s}{n}\right) - \left(\frac{s}{n}\right) \cdot \rho'\left(\frac{s}{n}\right)$$

where  $\rho'(\cdot)$  is the function's first derivative. Due to the properties of  $\rho(t)$  there will be a unique, non-negative, and finite  $n^*$  as long as  $0 \leq w < P$ . Also, since  $\frac{s}{n} = t$  and  $n$  only enters the right side through  $t$ , the physician's problem also characterizes a unique and non-negative level of treatment  $t(w, P) = \frac{s}{n^*}$  that is independent of  $s$  and determined entirely by  $w$  and  $P$ . The characterization of  $n^*$  yields the physician's responses to changes in the contract with the MCO and the level of malpractice pressure:

$$\begin{aligned} \frac{\partial n^*}{\partial P} &= -\left(\frac{w}{P}\right) \left(\frac{n^3}{s^2}\right) \left[P \cdot \left(\frac{\partial^2 \rho}{\partial t^2}\right)\right]^{-1} &< 0, \\ \frac{\partial t}{\partial P} &= \left(\frac{w}{P}\right) \left(\frac{n}{s}\right) \left[P \cdot \left(\frac{\partial^2 \rho}{\partial t^2}\right)\right]^{-1} &> 0, \\ \frac{\partial n^*}{\partial w} &= \left(\frac{n^3}{s^2}\right) \left[P \cdot \left(\frac{\partial^2 \rho}{\partial t^2}\right)\right]^{-1} &> 0, \\ \frac{\partial t}{\partial w} &= -\left(\frac{n}{s}\right) \left[P \cdot \left(\frac{\partial^2 \rho}{\partial t^2}\right)\right]^{-1} &< 0, \\ \frac{\partial n^*}{\partial s} &= \left(\frac{n}{s}\right) &> 0 \end{aligned}$$



These relationships together illustrate two points about physician behaviour under this model. First, physicians practice both positive and negative defensive medicine in the face of rising malpractice pressure. Given any contract with the MCO, rising pressure makes the marginal patient too risky to treat, causing the physician to remove some patients from his patient roll (negative) and also to increase the amount of treatment provided to each of the remaining patients (positive). Second, physicians respond to financial incentives in their contract with the MCO. A greater stock of resources lowers the risk of treating the marginal patient, and induces the physician to increase his patient roll size. Increases in compensation-per-patient increase the revenue from treating the marginal patient, leading the physician to take on more patients and accept greater expected liability costs. This behaviour is accounted for when the MCO chooses its contracts.

## 2.6 Equilibrium: MCO

From the physician's problem, for a given value of  $P$ , the revenue per patient necessary for the MCO to induce a physician to choose to treat a given roll size ( $\bar{n}$ ), as a function of resources, is given by  $\omega(s; \bar{n}, P)$ . Characteristics of this function are retrieved from the physician's problem:

$$\begin{aligned}\omega(s; \bar{n}, P) &= P \left[ \rho(t) - t \left( \frac{\partial \rho}{\partial t} \right) \right] > 0 \\ \frac{\partial \omega}{\partial s} &= -Pt \left( \frac{\partial^2 \rho}{\partial t^2} \right) \left( \frac{1}{\bar{n}} \right) < 0 \\ \frac{\partial^2 \omega}{\partial s^2} &= -P \left[ \frac{\partial^2 \rho}{\partial t^2} + t \left( \frac{\partial^3 \rho}{\partial t^3} \right) \right] \left( \frac{1}{\bar{n}} \right)^2\end{aligned}$$

Where  $t = \frac{s}{\bar{n}}$ . The function  $\omega(\cdot)$  is always decreasing in  $s$  and must have at least one inflection point. This is because  $\left( \frac{\partial^2 \rho}{\partial t^2} \right)_{s=0} > 0$  and  $t \left( \frac{\partial^3 \rho}{\partial t^3} \right)_{s=0} = 0$ , causing  $\omega(\cdot)$  to be concave at  $s = 0$ . It must eventually become and remain convex, however, since  $\omega(\cdot)$  is always decreasing in  $s$  but must remain positive.

The existence of an inflection point is counter-intuitive, as the convexity in  $\rho(t)$  implies that there are diminishing returns to treatment. That being the case, one might expect the impact of a marginal increase in the stock of resources available to physicians on  $\omega(\cdot)$  to decrease monotonically as more resources are provided. The reason for the existence of the inflection point is that  $\omega(\cdot)$  approaches  $P$  as  $s$  approaches zero. From the solution to the physician's problem, if  $w = P$ , the physician would want to see an infinite number of patients since the large payment would fully compensate for the liability cost of taking on a new patient, even as the probability of an adverse outcome approaches 1. In such a situation, small changes in  $w$  produce large changes in  $n$ , and so maintaining  $\bar{n}$  requires only small changes in  $\omega(\cdot)$  at levels of  $s$  close to 0. An inflection point in  $\omega(\cdot)$  should therefore be expected at low levels of  $s$ , and

thus  $t$ , when  $\omega(\cdot)$  is nearly as large as  $P$ . Without reason to expect more than one, Assumption 1 is made to rule out cases of multiple unconnected inflection points:

**Assumption 1:** The function  $\rho(t)$  is such that  $\exists! \underline{t} > 0$  where:

$$\frac{\partial^2 \rho}{\partial t^2} + t \left( \frac{\partial^3 \rho}{\partial t^3} \right) \begin{cases} \geq 0 \forall t \in [0, \underline{t}]; \\ < 0 \text{ otherwise.} \end{cases}$$

Let  $\theta(s; \bar{n}, c, P) = D\omega(s; \bar{n}, P)\bar{n} + Dcs$ . The function  $\theta(\cdot)$  represents the minimum policy price that the insurer can charge in order to induce a representative physician to treat  $\bar{n}$  patients, given that the physician is provided with  $s$  resources. As resources are removed, the total cost of inducing each physician to keep treating  $\bar{n}$  patients approaches  $D\bar{n}P$ . This cost would be high in environments where malpractice pressure is high. In such instances, it is entirely likely that the policy price necessary to get  $\bar{n}$  patients treated would be lower when physicians are provided with a positive amount of resources than when they are provided with no resources at all. For this to be case, treatment must be effective at lowering the probability of an adverse outcome, and the marginal cost of resources ( $c$ ) must be low relative to the prevailing level of malpractice pressure. It is to restrict the analysis to these cases that Assumption 2 is made:

**Assumption 2:** The function  $\rho(t)$  and parameters  $\{c, P\}$  are such that:

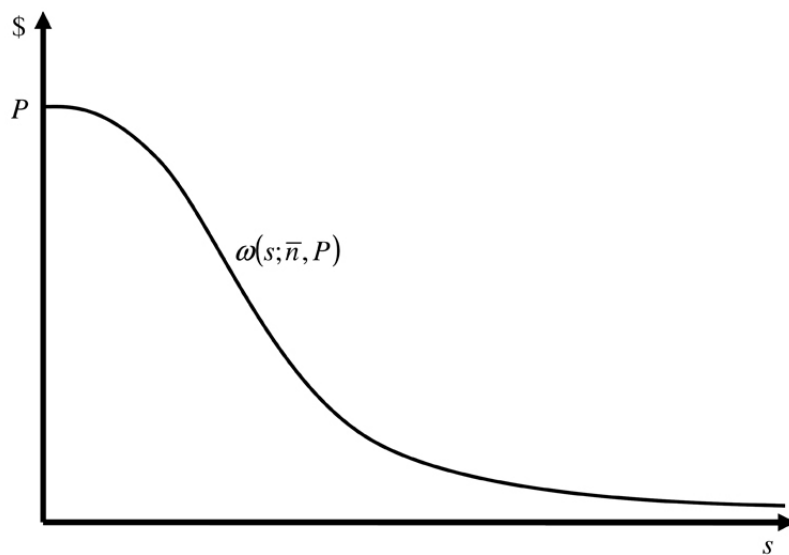
$$\exists \bar{t} > 0 \text{ where } c\bar{t} = P[1 - \rho(\bar{t}) + \bar{t}\rho'(\bar{t})]$$

Notice that the right side of the equation is actually  $\omega(0; \bar{n}, P) - \omega(\bar{s}; \bar{n}, P)$  where  $\frac{\bar{s}}{\bar{n}} = \bar{t}$ . Intuitively, for some level of treatment  $\bar{t}$ , the cost of providing a patient with that treatment ( $c\bar{t}$ ) is balanced by the reduction in the payment to a physician ( $\omega(0; \bar{n}, P) - \omega(\bar{s}; \bar{n}, P)$ ) that is required to get the physician to see that patient when the physician is provided with  $\bar{s} = \bar{n}\bar{t}$  in resources instead of zero. Essentially, it means that the least expensive way to get doctors to treat  $\bar{n}$  patients is not to provide them with zero resources for use in treatment.

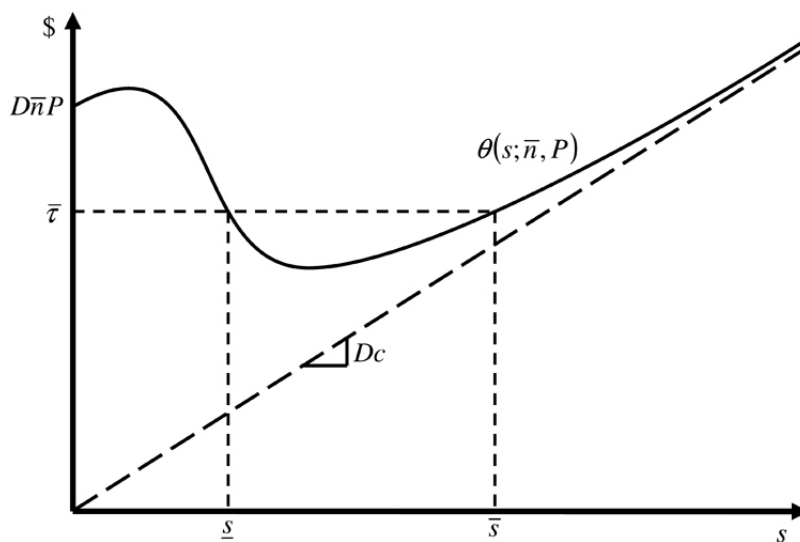
**Lemma 1:** If Assumptions 1 and 2 hold, then in any equilibrium,  $t^*$  must be such that:

$$\frac{\partial^2 \rho}{\partial t^2} + t \left( \frac{\partial^3 \rho}{\partial t^3} \right) < 0$$

**Proof:** Where Assumption 1 holds, the function  $\omega(\cdot)$  consists of a weakly concave segment followed by a strictly convex segment, as shown in Figure 2.1. The function  $\theta(\cdot)$  can be derived by scaling  $\omega(\cdot)$  by  $D\bar{n}$  and shifting it up by a vertical distance of  $Dc_s$ , as shown in Figure 2.2. The condition determining the



**Figure 2.1:** The revenue-per-patient necessary to induce a physician to treat  $\bar{n}$  patients, as a function of resources provided. By Assumption 1, it consists of a weakly concave segment followed by a strictly convex one.



**Figure 2.2:** The amount of spending per patient necessary to induce physicians to treat  $\bar{n}$  patients, as a function of resources provided. By Assumption 2, the global minimum cannot occur at  $s = 0$ .

concavity of  $\theta(\cdot)$  is identical to that of  $\omega(\cdot)$ , and so  $\theta(\cdot)$  also consists of a weakly concave segment followed by a strictly convex segment. Where Assumption 2 holds,  $\theta(\cdot)$  has an interior global minimum. Also note that this global minimum must be in the segment where  $\omega(\cdot)$  and  $\theta(\cdot)$  are convex. Any amount of resources in the domain of the concave segment ( $\underline{s}$ ) is dominated by another amount of resources ( $\bar{s}$ ), just as  $\theta(\underline{s}) = \theta(\bar{s}) = \bar{\tau}$  in Figure 2.2. This is because the two points induce the same patient roll size and entail the same policy price, but the greater amount of resources results in more treatment and higher quality at  $\bar{s}$ . Thus any equilibrium must be in the convex segment of  $\theta(\cdot)$ , where  $\frac{\partial^2 \rho}{\partial t^2} + t \left( \frac{\partial^3 \rho}{\partial t^3} \right) < 0$ .

Solving the MCO's problem requires working with the function  $\tilde{n}(w, s, P)$ . However, this function is not well-behaved, so the insurer's problem is modified using the function  $n^*(w, s, P)$  instead. This change requires the subsequent step of verifying whether or not the solution to the MCO's modified problem is feasible (ie. satisfies  $n^*(w, s, P) = \tilde{n}$ ) and if not, the optimal way for the MCO to restructure its contracts. As shown in the proof of Proposition 2.1, it turns out that the optimal restructured contract is such that  $n^* = \tilde{n} = \frac{q}{D}$ , and the solution to the insurer's problem can be determined in this case as well. It is convenient to define equilibria in terms of whether or not a restructuring is necessary, so let:

$$\left\{ \hat{w}(\tau, P), \hat{s}(\tau, P) \right\} = \arg \max_{w, s} \left\{ Q\left(\tilde{n}, \frac{s}{\tilde{n}}\right) \mid \tau \geq Dw\tilde{n} + Dcs \right\}$$

$$\bar{\tau}(P) = \min \left\{ \tau \mid \tilde{n}(\hat{w}, \hat{s}, P) = \frac{q}{D} \right\}$$

The insurance policy price  $\bar{\tau}(P)$  represents the lowest price which, if allocated optimally between physician payments and resources, would be sufficient to provide access for the expected number of ill policyholders, given the level of malpractice pressure. Define a “full-access equilibrium” as any equilibrium  $\{\tau^*, w^*, s^*, n^*(w, s, P)\}$  such that  $\tilde{n}^* = \frac{q}{D}$  and  $\tau^* > \bar{\tau}(P)$ . Also define a “limited-access equilibrium” as any equilibrium such that  $\tau^* \leq \bar{\tau}(P)$ . The first order conditions in the MCO’s modified problem are:

$$q \cdot \Delta U \cdot \frac{\partial Q^*}{\partial w} = WMU_y \cdot \frac{\partial \tau}{\partial w} \quad (2.1)$$

$$q \cdot \Delta U \cdot \frac{\partial Q^*}{\partial s} = WMU_y \cdot \frac{\partial \tau}{\partial s} \quad (2.2)$$

$$\tau = Dwn^* + Dcs$$

where,

$$\Delta U = U(H_1, m - \tau) - U(H_2, m - \tau) > 0,$$

and

$$WMU_y = (1 - q) \cdot \frac{\partial U(H_1, y)}{\partial y} + q \cdot \frac{\partial U(H_2, y)}{\partial y} + qQ^* \cdot \frac{\partial \Delta U}{\partial y} > 0$$

The term  $WMU_y$  stands for “weighted marginal utility of consumption”, and represents the marginal value of a unit of consumption that a policyholder faces ex ante. The  $\Delta U$  term represents difference in utility enjoyed by those who recover from the low health status to the high, or the value of recovery.

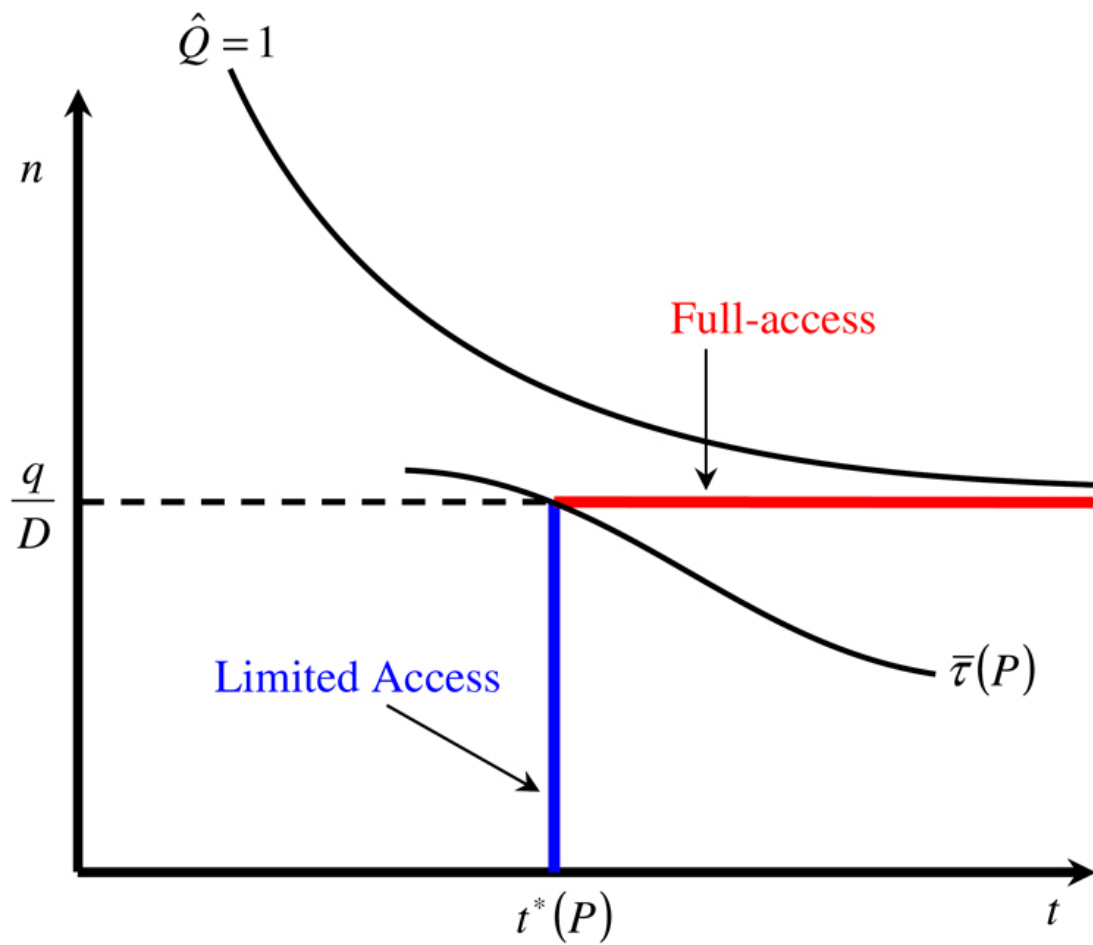
**Proposition 2.1:** If Assumptions 1 and 2 hold, then there exists a unique solution to the MCO’s problem that is either a full-access or a limited-access equilibrium.

**Proof:** In appendix.

Part of the insurer’s problem is to decide, given some level of malpractice pressure, how to divide up its revenues from selling insurance policies ( $\tau$ ) in order to procure inputs ( $t$  and  $n$ ) to maximize health care quality ( $Q$ ). This problem is shown in Figure 2.3. The proof to Proposition 2.1 shows that there is a unique level of treatment  $t^*(P)$  that solves the insurer’s modified problem, and that this level of treatment is independent of the amount of funds that the insurer has available for procurement. This means that, if the insurer were to charge higher policy prices, the optimal way to use the higher revenues would be to hold the level of treatment constant at  $t^*$  and use the additional funds to induce physicians to treat more patients. Therefore, as  $\tau$  increases, the expansion path is initially vertical in  $\{t, n\}$  space.

This is true until  $\tau = \bar{\tau}(P)$ , represented by its isocost line in Figure 2.3.





**Figure 2.3:** Expansion path showing the optimal allocation of insurer revenues in procuring health care inputs  $t$  and  $n$  as health insurance policies become more expensive.

The expansion path cannot continue along the vertical path from the modified problem, because the values of  $n$  in these allocations would be greater than  $\frac{q}{D}$ , and thus infeasible. In such a case, the insurer would need to choose the best alternative contract that brings about a feasible  $n$ . Since there is a unique solution to the insurer's first-order conditions,  $n^* > \frac{q}{D}$  implies that the choice  $n = \frac{q}{D}$  dominates all  $n < \frac{q}{D}$ , and so the alternative contract must be structured such that  $n = \frac{q}{D}$ . Thus, if  $\tau$  increases beyond  $\bar{\tau}(P)$ , the additional funds are used to increase  $t$  instead of  $n$ . This gives rise to the horizontal section of the expansion path in Figure 2.3. Any value of  $\tau$  results in an allocation somewhere along this expansion path, and so the unique equilibrium value must fall in either the full- or limited-access segments.

## 2.7 Full-Access Equilibrium

By definition, every ill policyholder receives treatment in a full-access equilibrium, so it must be that  $\tilde{n} = \frac{q}{D}$ . Analysis can therefore be confined to the values of  $w$  and  $s$  that induce a choice of  $n^*(w, s, P) \geq \frac{q}{D}$  in the physician's problem. Since any  $w$  and  $s$  such that  $n^*(w, s, P) > \frac{q}{D}$  would result in  $\frac{\partial \tilde{Q}}{\partial w} = 0$  and  $\frac{\partial \tau}{\partial w} > 0$ , which would violate (2.1), analysis can further be confined to those  $w$  and  $s$  inducing  $n^*(w, s, P) = \frac{q}{D}$ . From the previous section,  $w = \omega(s; \frac{q}{D}, P)$  is the unique value of  $w$  that would bring about such a choice for any value of  $s$ . Substituting  $w = \omega(s; \frac{q}{D}, P)$ ,  $\tilde{n} = \frac{q}{D}$ , and  $t = \frac{Ds}{q}$  into the MCO's problem and maximizing

with respect to  $s$  yields the first order conditions:

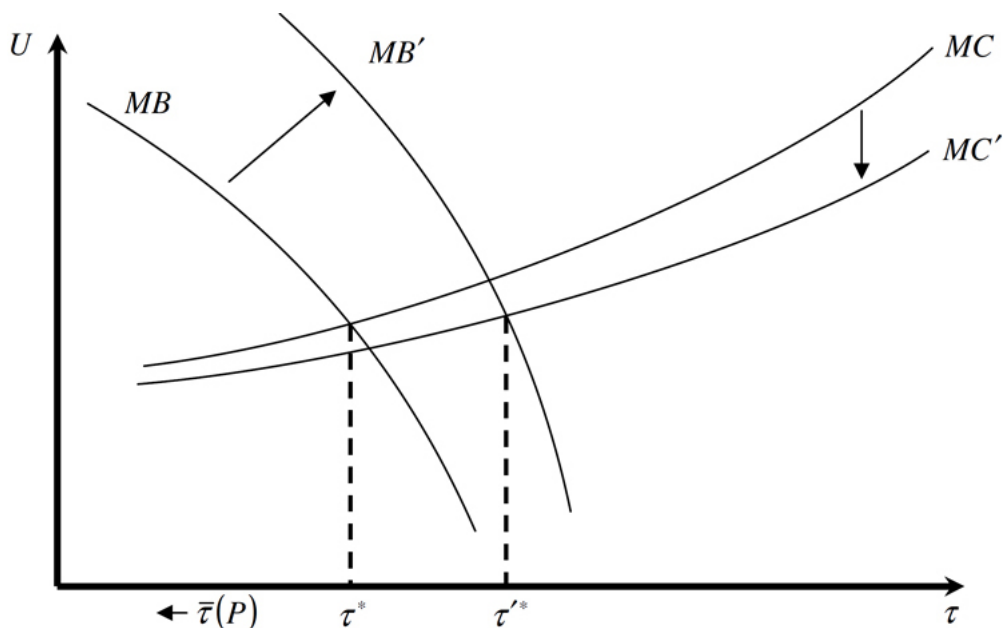
$$\begin{aligned} \Delta U \cdot \left( -\frac{\partial \rho}{\partial t} \right) &= WMU_y \cdot \left[ c - Pt \left( \frac{\partial^2 \rho}{\partial t^2} \right) \right] \\ \tau &= q \cdot \omega \left( s; \frac{q}{D}, P \right) + Dcs \end{aligned} \quad (2.3)$$

**Proposition 2.2:** If Assumptions 1 and 2 hold, then  $\frac{\partial \tau^*}{\partial P} > 0$ , and  $\frac{\partial \tilde{n}^*}{\partial P} = 0$  in any full-access equilibrium.

**Proof:** In appendix.

This proposition states that, starting in any full-access equilibrium, the MCO would respond to rising malpractice pressure by increasing prices in order to maintain the level of access to physicians enjoyed by their customers. Since the set of consumers is of measure 1, and all consumers purchase health insurance in equilibrium, the policy price is equal to total health care spending. Therefore, total health care spending is also increasing in malpractice pressure in any full-access equilibrium.

The intuition for these comparative statics can be found in Figure 2.4. The marginal benefit to consumers of an increase in the policy price (and thus increased funds for the procurement of resources and physicians' services) is  $MB = q \cdot \Delta U \frac{\partial \tilde{Q}}{\partial \tau}$ , which represents the value of a marginally higher probability of ending up with  $H_1$  instead of  $H_2$ .  $MB$  is monotonically decreasing in  $\tau$  for



**Figure 2.4:** The effect of an increase in malpractice pressure on equilibrium health care spending in a full-access equilibrium.

all  $\tau > \bar{\tau}(P)$ . The marginal cost is  $MC = WMU_y$ , which is increasing in  $\tau$  and represents the marginal value of forgone consumption.

An increase in  $P$  has three effects. First, given any amount of resources, the higher  $P$  means the MCO must provide physicians with a greater  $w$  in order to maintain  $n^* = \frac{q}{D}$ . At a given level of spending, this leaves less funds for the procurement of resources, causing ill consumers to receive less treatment and enjoy a lower-quality insurance policy. This increases the likelihood that the patient ends up with the lower health status instead of the higher. Since the marginal utility of consumption is lower for consumers enjoying  $H_2$  instead of  $H_1$ , a higher  $P$  causes consumption to be marginally less valuable ex ante to

consumers. This results in the downward shift in  $MC$  at every  $\tau$ . Second, since there are diminishing returns to treatment, the lower level of treatment means that a marginal increase in spending (and thus resources) delivers a greater marginal increase in quality. Finally, an increase in  $P$  causes the shadow price of resources  $\left(c - Pt \frac{\partial^2 \rho}{\partial t^2}\right)$  to decrease at every  $\tau$ . The second term in the shadow price, the amount by which the MCO can reduce  $\omega\left(s, \frac{q}{D}, P\right)$  as it increases resources, becomes greater at a lower  $t$  and higher  $P$ . The second and third terms create a bigger “bang for the buck” from increasing spending on resources, and thus an increase in  $P$  causes  $MB$  to shift up at every  $\tau$ . Taken together, these effects cause the new equilibrium to occur at an unambiguously higher level of spending. Since spending in the initial full-access equilibrium is greater than  $\bar{\tau}(P)$ , full access is preserved in the new equilibrium with the marginally higher  $P$ .

**Proposition 2.3:** If Assumptions 1 and 2 hold, then  $\frac{\partial t^*}{\partial P} > 0$  and  $\frac{\partial \tilde{Q}^*}{\partial P} > 0$  in a full-access equilibrium if and only if  $|\epsilon_{J,w}| < 1$ , where  $\epsilon_{x,y}$  is the percent change in  $x$  due to a one-percent change in  $y$  and  $J = \frac{WMU_y}{\Delta U}$ .

**Proof:** In appendix.

Proposition 2.3 describes the impact of increasing malpractice pressure on health care quality in a full-access equilibrium. From Proposition 2.2, the MCO raises the the policy price in response to a marginal increase in malpractice

pressure in any full-access equilibrium. Since all consumers purchase health insurance in equilibrium, and the MCO makes zero profits, the entire increase in revenues must be divided between increased per-patient payments and resources in the new equilibrium contract between the MCO and physicians. The condition in Proposition 2.3 is necessary and sufficient to determine whether or not the new equilibrium contract provides physicians with more resources than were available under the old contract. Also from Proposition 2.2, the number of patients that each physician would treat in equilibrium would remain unchanged. Therefore, the new equilibrium contract would allow physicians to divide a greater amount of resources among the same number of patients, leading to a greater amount of treatment, a lower chance of an adverse outcome, and greater health care quality.

There is reason to expect that  $|\epsilon_{J,w}| < 1$  in all full-access equilibria with reasonable parameter values and functional forms. Intuitively, the condition states that policy price increase necessary to raise physicians' revenues by one percent would result in less than a one percent change in the expected value of consumption ( $WMU_y$ ) relative to recovery ( $\Delta U$ ). As long as the share of income devoted to health insurance is relatively low and the value of recovery is high, small percentage changes in the prices of health care inputs would produce even smaller changes in a consumer's expected value of consumption relative to recovery. Given that the United States, the highest spender on health care both per-capita and as a percentage of GDP, spends less than 20% of GDP on

health care, the amount of spending necessary to violate the condition seems unrealistic, particularly since cheaper limited-access insurance policies could be provided if consumers so prefer. Taking Propositions 2.2 and 2.3 together, jurisdictions or medical specialties in full-access equilibria can expect rising malpractice pressure to cause increases in the cost of health insurance and total health care spending. Regarding quality, equilibrium health care quality would increase with malpractice pressure as long as the level of health care spending is not extremely high.

## 2.8 Limited-Access Equilibrium

An important finding in the proof of Proposition 2.1 is that the treatment  $t$  solving the modified insurer's problem is independent of the level of spending, as shown for  $t^*(P)$  in Figure 2.3. Since, for a given  $P$ , the level of treatment is uniquely determined in the physician's problem by  $w$ , then  $w^*(P)$  such that  $t(w^*(P), P) = t^*(P)$  is also independent of  $\tau$ . Since modified and unmodified solutions are equivalent in limited-access equilibria, both  $t^*(P)$  and  $w^*(P)$  from the physician's modified problem arise in a limited-access equilibrium for a given level of  $P$ .

**Proposition 2.4:** If Assumptions 1 and 2 hold, then  $\frac{\partial t^*}{\partial P} > 0$  in any limited-access equilibrium.

**Proof:** In appendix.

Since limited-access equilibria are interior solutions, the cost-effectiveness of a marginal unit of access and a marginal unit of treatment are the same in equilibrium. The intuition behind Proposition 2.4 is that malpractice pressure makes access more expensive relative to treatment, creating a substitution effect away from access toward treatment. Furthermore, since  $(1 - \rho)$  is solely a function of  $t$ , then the equilibrium probability of patient recovering from illness, conditional on gaining access to a physician, would increase in  $P$  as well.

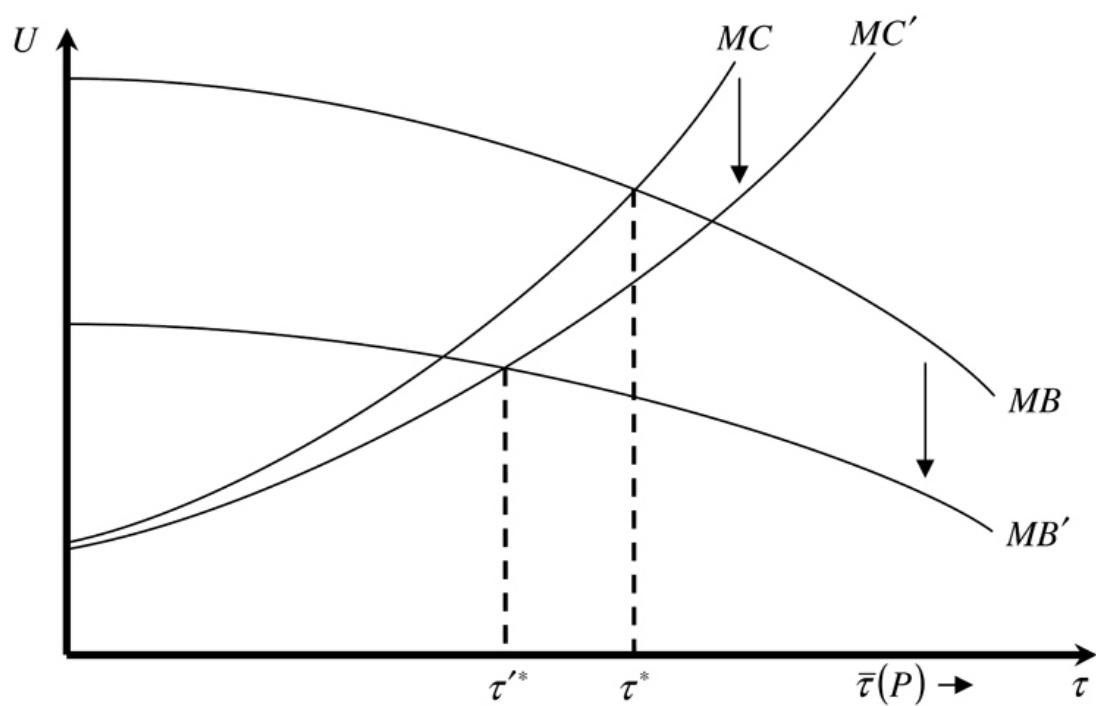
**Proposition 2.5:** If Assumptions 1 and 2 hold, then  $\frac{\partial \tau^*}{\partial P} < 0$  in a limited-access equilibrium if and only if:

$$\tau < \frac{\Delta U}{\frac{\partial \Delta U}{\partial y}} \tag{2.4}$$

**Proof:** In appendix.

Since  $\Delta U$  is concave in  $y$ , a level of health care spending greater than 50% of consumer income would be required to violate Condition (2.4). For this reason, and similarly to Proposition 2.3, Condition (2.4) can be expected to hold for all realistic parameter values. The reason that a qualifying condition is





**Figure 2.5:** The effect of an increase in malpractice pressure on equilibrium health care spending in a limited-access equilibrium satisfying (2.4).

required here is that increases in malpractice pressure have different implications for the marginal benefit and cost of health care spending in the two types of equilibrium. In comparing Figures 2.4 and 2.5, a rise in malpractice pressure has the same qualitative effect on  $MC$  in the two equilibrium types, but the opposite effect on  $MB$ . This is because marginal spending increases in the limited-access case are optimally devoted to procuring greater access instead of treatment. An increase in malpractice pressure makes it more expensive for the MCO to induce physicians to treat a given number of patients, which makes access more expensive. This increase in the marginal cost of access means that a marginal increase in spending is able to provide a relatively smaller increase in quality at the higher level of malpractice pressure. Essentially, unlike the full-access case, malpractice pressure lowers the “bang for the buck” from health care spending, and thus creates the downward shift in  $MB$ . Where (2.4) holds, the shift in  $MB$  dominates the shift in  $MC$ , leading to a lower equilibrium policy price.

**Corollary 2.5a:** In any limited-access equilibrium,  $\frac{\partial \tau^*}{\partial P} < 0 \Rightarrow \frac{\partial \tilde{n}^*}{\partial P} < 0$  and  $\frac{\partial \tilde{Q}^*}{\partial P} < 0$ .

**Corollary 2.5b:** If (2.4) holds in a limited-access equilibrium at  $P'$  then it holds for all  $P \in (P', \infty)$ .

The implication in Corollary 2.5a fairly straightforward. Once  $\{w, t\} =$

$\{w^*, t^*\}$ , the only endogenous variable left to affect health care quality  $\tilde{Q}$  is  $\tau$ . Holding other parameters constant,  $\frac{\partial \tilde{Q}^*}{\partial P} = \frac{\partial \tilde{Q}}{\partial \tau} \frac{\partial \tau^*}{\partial P} + \frac{\partial \tilde{Q}}{\partial P} \Big|_{\tau^*}$ . Since quality is increasing in spending for a given level of malpractice pressure, and decreasing in malpractice pressure for a given level of spending,  $\frac{\partial \tau^*}{\partial P} < 0$  makes the right side unambiguously negative. Essentially, rising malpractice pressure makes quality more expensive to provide, so if the funds available for spending on quality decrease then the amount of quality produced must also decrease. Since  $t^*$  increases with  $P$  by Proposition 2.4, the only way that  $\tilde{Q}^*$  can decrease in  $P$  is if  $\tilde{n}^*$  is also decreasing in equilibrium.

If spending is decreasing in malpractice pressure, then an increase in  $P$  causes the left side of (2.4) to decrease and the right side to increase, proving Corollary 2.5b. Taken together, the propositions and corollaries in this section imply there is a threshold level of malpractice pressure, after which any further increases in malpractice pressure would cause decreases in health care spending. Since  $\tilde{n}^*$  decreases in  $P$  in limited-access equilibria, this further implies that any full-access equilibria must occur before this threshold, at lower levels of malpractice pressure. Also, despite each patient receiving better treatment, the reduction in ill consumers' access to medical care causes the overall quality of the health care system to decrease in limited-access equilibria.

Notice in Table 2.1 that the effects of malpractice pressure on spending and quality in the limited-access equilibrium are the opposite of those in the full-access equilibrium. Intuitively, there are two inputs into the quality of health

	Full-Access Eqbm	Limited-Access Eqbm
Where it occurs	Low $P$	High $P$
Effect on access $\left(\frac{\partial \tilde{n}^*}{\partial P}\right)$	0	–
Effect on treatment $\left(\frac{\partial t^*}{\partial P}\right)$	+	+
Effect on spending $\left(\frac{\partial \tau^*}{\partial P}\right)$	+	–
Effect on system quality $\left(\frac{\partial \tilde{Q}^*}{\partial P}\right)$	+	–

Table 2.1: Summary of analytical results

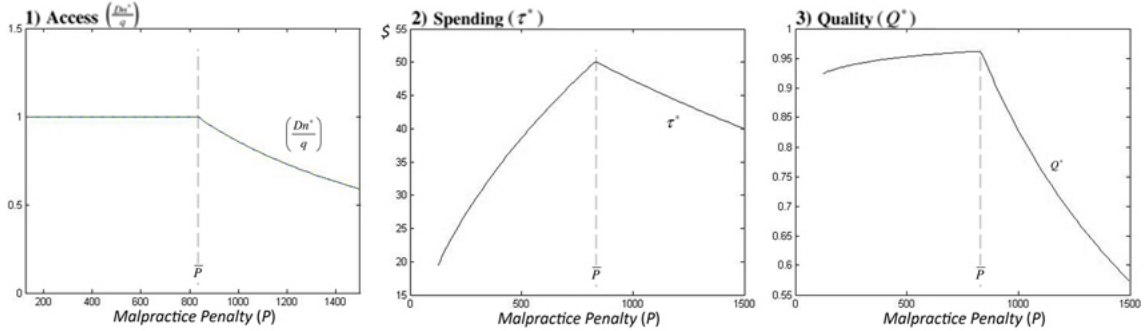
insurance: access and treatment. Where malpractice pressure is low, physicians do not require much incentive to treat a large number of patients. When physicians are so amenable to treating patients, access is cheap relative to treatment, making it optimal to provide as much access as possible. There is an upper bound on the amount of access consumers can be provided, that being “full access”. While full access is not costly at low levels of malpractice pressure, it becomes more costly as pressure increases. Therefore, at low levels of malpractice pressure, the price of insurance and total health care spending are increasing in pressure in order to maintain full access. Under the necessary and sufficient condition in Proposition 2.3, it is optimal to use enough of the increased revenue from higher policy prices to procure more resources for the physicians than they previously had, which causes quality to increase in malpractice pressure as well.

As the cost of full access increases, eventually consumers’ willingness to

pay becomes exhausted. Rather than continue to maintain full-access, the MCO instead offers a lower-priced insurance policy with more limited access to physicians. Should malpractice pressure continue to increase, each level of access becomes more costly. Since the marginal cost of resources remains constant, the MCO substitutes away from access toward treatment in the production of health care quality, causing access to decrease and treatment to increase in malpractice pressure. Finally, since there are decreasing returns to treatment, this substitution makes each dollar spent on health insurance less effective at producing quality. Therefore, consumers would prefer to substitute away from spending on health insurance and instead retain more income for consumption. These two substitution effects cause both spending and quality to decrease in malpractice pressure once pressure is high enough to make limited access preferable to full access. Altogether, given the different comparative statics in the two types of equilibrium, and the values of malpractice pressure for which we can expect each type of equilibrium to occur, the model predicts non-monotonic relationships between malpractice pressure and both health care quality and expenditure.

## **2.9 Discussion**

A numerical example provides a useful illustration of the analytical predictions from Table 2.1. A Cobb-Douglas utility function and the functional form



**Figure 2.6:** Numerical example illustrating non-monotonic relationships between malpractice pressure and 1) the level of access, 2) the price of health insurance, and 3) the quality of the health care system.

$\rho(t) = \frac{1}{1+\alpha t}$  where  $\alpha > 0$  produce the relationships in Figure 2.6. The non-monotonicities predicted by the model are clearly apparent. These predictions offer an explanation for the discrepancies in the empirical literature. The first two effects in Table 2.1 confirm the beliefs in the medical establishment on the effect of liability costs on defensive practices. Any effect that malpractice pressure has on access to care would be negative. In equilibrium, on average, physicians’ practices would thus become more restrictive to patients as malpractice pressure increases. Regardless of equilibrium type, the amount of resources used in the treatment of each patient is increasing in liability costs. Due to the diminishing returns from treatment and the constant marginal cost of resources, these additional treatments decline in cost-effectiveness, and may thus be interpreted as medically unnecessary or wasteful.

Despite consistent effects of malpractice pressure on defensive medicine in equilibrium, the links between defensive practices and the cost and quality of

care are more complicated than those espoused by the American Medical Association. Due to the non-monotonicity, regression analysis designed to uncover the monotonic relationship best approximating a set of observations would report coefficients biased by the sample's prevailing equilibrium type. Studies utilizing data predominantly to the left of the threshold would report positive coefficients while those from the right would report negative ones. Samples with observations from both sides of the threshold would report coefficients biased toward zero. In this way, even the widespread and consistent practice of defensive medicine could be hidden by inconsistent and seemingly conflicting empirical investigations into its effects on health care cost and quality.

The monotonic relationships found (or not found) in past empirical studies are consistent with the non-monotonic relationships predicted by the model presented here. Diagnostic and medical imaging procedures are useful examples of treatment intensity as defined here, and they tend to be positively affected by malpractice pressure (Kessler & McClellan 2002, Baicker et al 2007), as is predicted by the model. Regarding access, the incidence of late onset of prenatal care in a population from Dubay et al (2001) is a useful measure of the percentage of a sample in limited-access equilibrium. In regressing the late initiation of prenatal care on malpractice premiums, they find statistically significant negative effects that, while relatively small, were approximately twelve times greater in magnitude for unmarried black mothers with less than

high-school education than for white married mothers with a college education. In the former, 51.2% of mothers received late care versus 5.3% in the latter. These results could be capturing the relationship from the first diagram in Figure 2.6. The greater is the share of observations from the right side of the threshold, the more negative would be the coefficient reported by standard regression analysis. The model's predictions also match evidence from General Accounting Office (2003), which found effects on access were greater in rural populations than urban populations. Lower average incomes in rural populations could result in less of a willingness and ability to maintain full access, and thus a switch to limited-access equilibria at a lower threshold of malpractice pressure than would be observed in urban populations.

The studies that uncover no statistically or economically significant effects of malpractice pressure on the cost or quality of health care tend to be broad studies at the state-level, such as Baicker & Chandra (2005), Hellinger & Encinosa (2006), Congressional Budget Office (2006), and Baicker et al (2007). Broad data is more likely to incorporate observations from both types of equilibria, and the underlying monotonicity would bias their estimated coefficients toward zero. Studies focussed on a more narrowly defined populations exhibit a link between the prevailing liability risk of that population and the sign of the revealed coefficients, a pattern corresponding to the model's predictions. Kessler & McClellan (1996) and (2002) examine spending in data covering



heart-attack and heart disease patients on Medicare. Elderly patients are relatively unlikely sources of malpractice lawsuits (Kessler & McClellan 1996, Sloan & Shadle 2009) due to the lower lost wages and shorter pain and suffering horizons they could claim (Avraham & Shanzenbach 2010). Furthermore, cardiovascular physicians enjoy a percentage of claims that result in payment that is slightly under two thirds of the average (Dove et al 2010), and the specialty is not considered high-risk by the American Medical Association (Cotet 2012). Lakdawalla & Seabury (2009) also utilize data on elderly Medicare patients, and further draw many of their observations from California, which has experienced some of the lowest malpractice insurance rate increases (General Accounting Office 2003). Furthermore, the state's Medical Injury Compensation Reform Act (MICRA) of 1975 is credited by the medical industry with lowering the cost of health insurance, stemming the outflow of physicians, and improving patient access to care (Brenner & Smith 2004, Lumalcuri & Hale 2010). It seems plausible, therefore, that physicians serving the populations studied in the works of Kessler & McClellan and Lakdawalla & Seabury practice under relatively low malpractice risk, and both find the positive relationships that the model predicts would be associated with populations on the left side of the threshold.

On the other hand, the studies reporting negative relationships use data on

infant health and obstetric care, which is one of the riskiest medical specialties for the frequency and severity of malpractice lawsuits. The cost of malpractice insurance for obstetricians increased 180 percent between 1977 and 1984, versus 109 percent for lower risk specialties (Danzon et al 1990) while growth rates in the 1990s and early 2000s have been erratic (Reyes 2010). Obstetricians face a high variance in malpractice payments (Kravitz et al 1991), the highest payment rate conditional on a claim, and the greatest likelihood that payments will exceed the limits of malpractice insurance coverage (Jena et al 2011). If liability risk is great enough to push certain groups of mothers into limited-access equilibria, then negative relationships between malpractice pressure and the cost and quality of care would be expected. This was the case regarding quality in Dubay et al (2001), since newborns' health indicators were unaffected while mothers' access suffered.

The use of cesarean section in childbirth as a measure for spending or defensive behaviour is challenging because it doesn't easily map into either treatment or access as modelled here. It is similar to treatment intensity since it requires more resources per patient to conduct each birth, but to access as well in that it cannot be conducted more than once per birth. This is problematic as the model predicts that the use of cesarean section should increase with malpractice pressure in its role as treatment intensity, but decrease in its role as access. Perhaps reflecting this, as well as the variation in patients' distribution across full- and limited-access equilibria derived from Dubay et al (2001),

empirical studies examining the use of cesarean section have uncovered mixed effects. Currie & MacLeod (2008) stands out in finding that liability-reducing reforms increase the incidence of cesarean section. While this paper's predictions are consistent with their empirical findings if much of their population is in limited-access equilibrium, the model presented here provides an alternative intuition for the result. They propose that reductions in the fear of liability induce physicians to perform more expensive cesarean sections in order to pull in higher fees. While this incentive to gain revenue surely exists, it is unlikely to be held in check by the fear of liability. Many more obstetrical claims (31% vs 3%) are associated with nonperformance or delay in performing a cesarean section than the unnecessary performance of one (Kravitz et al 1991) and there is some evidence that cesarean section is not susceptible to supplier-induced demand (Tussing & Wojtowycz 1992). This model alternatively suggests that reductions in malpractice pressure make health insurance policies with greater access to cesarean section more affordable, thus increasing the incidence. Determining which mechanism is behind the negative relationship would be worthwhile.

## **2.10 Summary and Implications**

Medical malpractice reform has been proposed as a policy option for reducing health care spending and improving health care quality. This is due to

consistent responses in surveys of physicians the malpractice liability costs encourage the practice of defensive medicine. Discrepancies between these surveys and the lack of consistent findings from the empirical literature have confounded efforts to determine whether tort reforms produce the desired effects. This presents a model of the interactions between patients, physicians, and health insurers that provides an explanation for these discrepancies. It implies that conflicting findings in the empirical literature are due to the existence of two equilibrium types that exhibit the opposite effects of malpractice pressure on health care spending and quality. These opposite effects, plus the switching of equilibrium types once a threshold of malpractice pressure is reached, create non-monotonic relationships between health care spending and quality, despite the practice of defensive medicine at all levels of pressure.

Any normative implications that should be taken away from this model are limited. By ignoring the compensation of injured patients, and the identical treatment of each patient in equilibrium, the model assumes away potential benefits of medical malpractice law. This produces the fragile implication that malpractice pressure should be reduced to zero. This implication arises, however, due to assumptions made for tractability rather than realism. The intuition behind the model's positive findings, however, is robust to the assumptions that produce more persuasive normative implications found in the existing theoretical work on medical malpractice. Malpractice pressure makes access to health care more expensive relative to treatment. Consumers may have

some willingness to bear the cost of maintaining full access to care as malpractice pressure rises, but this willingness to pay becomes exhausted if their consumption falls to low. Thereafter, consumers may prefer less expensive insurance with some limits to access, such as congestion, in order to preserve their consumption.

There are several implications of these findings for public policy. First, tort reform is not a “silver bullet” policy capable of raising health care quality while also lowering the cost of care for a given homogenous population. Even if a given reform was successful in changing the prevailing level of malpractice pressure, quality and spending would move together. Policymakers, therefore, face a tradeoff and must decide whether quality improvements or cost reductions would be of greater benefit to the affected population. Second, sweeping tort reforms would produce the opposite effects in distinct geographic regions, medical specialties, and patient demographics according to the group’s dominant equilibrium type. This suggests that a targeted approach to tort reform would produce consistent effects better than sweeping changes. Finally, the results provide clues for developing alternative empirical model specifications for investigating defensive medicine. Non-linear model specifications would be useful in approximating the peaked relationships predicted here, notwithstanding certain concerns raised in the literature with some continuous measures of malpractice pressure (Kessler & McClellan 1996). In the case of state-level tort reform, which indicates a change in malpractice pressure rather than

a continuous measure of it, the model predicts that the qualitative effect is determined by the current level of access that a population of ill consumers enjoy with their physicians. Excessive waiting times, a high incidence of late treatment, or other significant difficulty in securing a physician's services can be considered examples of poor access. Where these are at practically low levels, tort reforms lowering malpractice pressure should lower insurance policy prices and total health care expenditure, while causing some reductions in health care quality. Where they are unnaturally high, the same reforms would have the opposite effect. This particular role for access measures, that of distinguishing observations expected to experience different qualitative effects, has not been investigated in the empirical literature thus far. Data linking measures of access alongside measures of malpractice pressure, health care spending, and quality would be required for such an investigation to occur.

## 2.11 Bibliography

- Arlen, J., and W.B. MacLeod. 2005. Torts, Expertise, and Authority: Liability of Physicians and Managed Care Organizations. *The RAND J. of Economics* **36**(3) 494–519.
- Avraham, R. and M. Schanzenbach. 2010. The Impact of Tort Reform on Private Health Insurance Coverage. *American Law and Economics Review* 1–37.
- Baicker, K., and A. Chandra. 2005. The Consequences of the Growth of Health Insurance Premiums. *AEA Papers and Proceedings* **95**(2) 214–218.

- Baicker, K., E.S. Fisher, and A. Chandra. 2007. Malpractice Liability Costs and the Practice of Defensive Medicine in the Medicare Program. *Health Affairs* **26**(3) 841–852.
- Bovbjerg, R.R., L.C. Dubay, G.M. Kenney, and S.A. Norton. 1996. Defensive Medicine and Tort Reform: New Evidence in an Old Bottle. *Journal of Health Politics, Policy and Law* **21**(2) 267–288.
- Brenner, R.J., and J.J. Smith. 2004. The Malpractice Liability Crisis. *J. of the American College of Radiology* **1**(1) 18–22.
- Congressional Budget Office. 2004. Limiting Tort Liability for Medical Malpractice. Economic and Budget Issue Brief, Congressional Budget Office.
- Congressional Budget Office. 2006. Medical Malpractice Tort Limits and Health Care Spending. Background Paper, Congressional Budget Office.
- Cotet, A.M. 2012. The Impact of Noneconomic Damage Cap on Health Care Delivery in Hospitals. *American Law and Economics Review* 1–43.
- Currie, J., and W.B. MacLeod. 2008. First Do No Harm? Tort Reforms and Birth Outcomes. *The Quarterly J. of Economics* May 795–830.
- Danzon, P.M. 1985. Liability and Liability Insurance for Medical Malpractice. *The J. of Health Economics* **4** 309–331.
- Danzon, P.M., M.V. Pauly, and R.S. Kington. 1990. The Effects of Malpractice Litigation on Physicians' Fees and Incomes. *AEA Papers and Proceedings* **80**(2) 122–127.
- Dove, J.T., J.E. Brush Jr, R.A. Chazdal, and W.J. Oetgen. 2010. Medical Professional Liability and Health Care System Reform. *J. of the American College of Cardiology* **55**(25) 2801–2803.
- Dubay, L., R. Kaestner, and T. Waidmann. 2001. Medical malpractice liability and its effect on prenatal care utilization and infant health. *J. of Health Economics* **20** 591–611.
- Feess, E. 2012. Malpractice liability, technology choice and negative defensive medicine. *European J. of Health Economics* **13** 157–167.
- Finkelstein, A., E.F.P. Luttmer, and M.J. Notowidigdo 2008. What Good Is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption. Working Paper.

- Gal-Or, E. 1999. Optimal Reimbursement and Malpractice Sharing Rules in Health Care Markets. *J. of Regulatory Economics* **16** 237–265.
- Gaynor, M., J.B. Rebitzer, and L.J. Taylor. 2004. Physician Incentives in Health Maintenance Organizations. *The J. of Political Economy* **112**(4) 915–931.
- General Accounting Office. 2003. Medical Malpractice Insurance: Multiple Factors Have Contributed to Increased Premium Rates. GAO-03-702 Report to Congressional Requesters, General Accounting Office.
- Helland, E. and M.H. Showalter. 2009. The Impact of Liability on the Physician Labor Market. *J. of Law & Economics* **52**(November 2009) 635–663.
- Hellinger, F.J. and W.E. Encinosa. 2006. The Impact of State Laws Limiting Malpractice Damage Awards on Health Care Expenditures. *American J. of Public Health* **96**(8) 6–12.
- Jena, A.B., S. Seabury, D. Lakdawalla, and A. Chandra. 2011. Malpractice Risk According to Physician Specialty. *The New England J. of Medicine* **365**(7) 629–636.
- Keren-Paz, T. 2010. Liability Regimes, Reputation Loss, and Defensive Medicine. *Medical Law Review* **18** Autumn 363–388.
- Kessler, D.P. 2011. Evaluating the Medical Malpractice System and Options for Reform. *J. of Economic Perspectives* **25**(2) 93–110.
- Kessler, D., and M. McClellan. 1996. Do Doctors Practice Defensive Medicine?. *The Quarterly J. of Economics* **111**(2) 353–390.
- Kessler, D.P., and M.B. McClellan. 2002. How liability law affects medical productivity. *J. of Health Economics* **21** 931–955.
- Kessler, D.P., W.M. Sage, and D.J. Becker. 2005. Impact of Malpractice Reforms on the Supply of Physician Services. *J. of the American Medical Association* **203**(21) 2618–2625.
- Klick, J., and T. Stratmann. 2007. Medical Malpractice Reform and Physicians in High-Risk Specialties. *J. of Legal Studies* **36**(2) S121–S142.
- Kpelitse, K.A. 2010. Physician Payments and Medical Malpractice Mechanisms. Working Paper.



- Kravitz, R.L., J.E. Rolph, K. McGuigan. 1991. Malpractice Claims Data as a Quality Improvement Tool 1: Epidemiology of Error in Four Specialties. *J. of the American Medical Association* **266**(15) 2087–2092.
- Lakdawalla, D.A., and S.A. Seabury. 2009. The Welfare Effects of Medical Malpractice Liability. NBER Working Paper 15383.
- Leger, P.T. 2000. Quality control mechanisms under capitation payment for medical services. *Canadian J. of Economics* **33**(2) 564–586.
- Lumalcuri, J. and R.W. Hale. 2010. Medical Liability: An Ongoing Nemesis. *Obstetrics & Gynecology* **115**(2.1) 223–228.
- Malani, A. and J. Reif. 2010. Accounting for Anticipation Effects: an Application to Medical Malpractice Tort Reform. NBER Working Paper 16593.
- Matsa, D.A. 2007. Does Malpractice Liability Keep the Doctor Away? Evidence from Tort Reform Damage Caps. *The J. of Legal Studies* **36**(S2) S143–S182.
- Mello, M.M. and C.N. Kelly. 2005. Effects of a Professional Liability Crisis on Residents' Practice Decisions. *Obstetrics & Gynecology* **105**(6) 1287–1295.
- Mello, M.M., D.M. Studdert, C.M. DesRoches, J. Peugh, K. Zapert, T.A. Brennan, and W.M. Sage. 2005. Effects of a Malpractice Crisis on Specialist Supply and Patient Access to Care. *Annals of Surgery* **242**(5) 621–628.
- Moffett, P. and G. Moore. 2011. The Standard of Care: Legal History and Definitions: The Bad and Good News. *Western J. of Emergency Medicine* **12**(1) 109–112.
- Moriya, A.S. 2010. Consolidation, Insurance Coverage, and Tort Reform in Health Care Markets. Doctoral Dissertation. Carnegie Mellon University.
- Olbrich, A. 2008. Heterogeneous physicians, lawsuit costs, and the negligence rule. *International Review of Law and Economics* **28** 78–88.
- Paik, M., B.S. Black, D.A. Hyman, and C.M. Silver. 2011. Will Tort Reform Bend the Cost Curve? Evidence from Texas. Working Paper.
- Reyes, J.W., 2010. The Effect of Malpractice Liability on the Specialty of Obstetrics and Gynecology. NBER Working Paper 15841.

- Sethi, M.K., W.T. Ombremsky, H. Natividad, H.R. Mir, and A.A. Jahangir. 2012. Incidence and Costs of Defensive Medicine Among Orthopedic Surgeons in the United States: A National Survey Study. *American Journal of Orthopedics* **41**(2) 69–73.
- Shurtz, I. 2010. The Impact of Malpractice Litigation on Physician Behavior: The Case of Childbirth. Working Paper.
- Sloan, F.A. and J.H. Shadle. 2009. Is there empirical evidence for Defensive Medicine? A reassessment. *J. of Health Economics* **28** 481–491.
- Studdert, D.M., M.M. Mello, and T.A. Brennan. 2004. Medical Malpractice. *The New England J. of Medicine* **350**(3) 283–292.
- Studdert, D.M., M.M. Mello, W.M. Sage, C.M. DesRoches, J. Peugh, K. Zapert, and T.A. Brennan. 2005. Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment. *J. of the American Medical Association* **293**(21) 2609–2617.
- Thurston, N.K. 2001. Physician Market Power - Evidence from the Allocation of Malpractice Premiums. *Economic Inquiry* **39**(3) 487–498.
- Tussing, A.D. and M.A. Wojtowycz. 1992. The Cesarean Decision in New York State, 1986: Economic and Noneconomic Aspects. *Medical Care* **30**(6) 529–540.
- Yang, Y.T., M.M. Mello, S.V. Subramanian, and D.M. Studdert. 2009. Relationship Between Malpractice Litigation Pressure and Rates of Cesarean Section and Vaginal Birth After Cesarean Section. *Medical Care* **47**(2) 234–242.
- Zeiler, K. 2004. Medical Malpractice and Contract Disclosure: An Equilibrium Model of the Effects of Legal Rules on Behaviour in Health Care Markets. *American Law & Economics Association Annual Meetings* **60**. 1–75.

## **Chapter 3**

# **Physician Mobility and the Differential Effects of Defensive Medicine**

### **3.1 Introduction**

Defensive medicine refers to the treatment decisions made by physicians primarily to avoid medical malpractice liability rather than benefit patients. Examples include the ordering of unnecessary and costly diagnostic procedures to assure against any claims of negligence, or the avoidance of patients or procedures thought to be particularly at risk of resulting in a malpractice claim. Numerous surveys of physicians and the advocacy efforts of the American Medical

Association report that defensive medicine is both widely practiced and a primary reason for the rising cost of consumer health insurance and/or restricted access to care (Bovbjerg et al 1996, Studdert et al 2005, Dove et al 2010, Lumalcuri & Hale 2010, Reyes 2010, Sethi et al 2012). This has created advocacy for legislated malpractice reforms, mostly at the state-level. Researchers, too, have paid attention to the subject of defensive medicine. These empirical studies most often involve investigations into a possible relationship between various measures of medical malpractice liability costs, or “malpractice pressure”, and health care spending and/or quality. The results of these studies have been mixed, and have led some researchers to conclude that defensive medicine has not played a policy-relevant role in rising US health care expenditure, and that malpractice reform is thus unwarranted (Helland & Showalter 2009, Lakdawalla & Seabury 2009, Sloan & Shadle 2009, Avraham & Shanzenbach 2010, Reyes 2010, Cotet 2012).

A significant share of the empirical investigations into defensive medicine utilize data based on geographic jurisdictions, usually at the state or county level (Baicker & Chandra 2005, Baicker et al 2007, Hellinger & Encinosa 2006, Lakdawalla & Seabury 2009, Paik et al 2011). The goal is to determine whether health care spending or quality is systematically different in jurisdictions with high malpractice pressure. The empirical models used in these studies, however, assume that the malpractice pressure in a given state affects health care

spending only within that state. This specification neglects any potential effects that malpractice pressure in one state may have on the cost or quality of health care in other states.

One channel through which these cross-jurisdictional effects might occur is physician mobility. Both practicing physicians and newly graduated physicians have some discretion over the geographic region in which to locate their practices. Given the survey results, it appears that malpractice insurance premiums and the prospect of a malpractice lawsuit are a source of concern for physicians. For this reason, if all other considerations were equal, physicians would prefer to locate in jurisdictions where they would face low malpractice pressure. It then follows that rising malpractice pressure in one jurisdiction could trigger physician relocation, and thus affect the cost or quality of health care in other jurisdictions.

The potential reactions of health insurers to physician mobility makes the ultimate effects on health care spending and quality ambiguous. If physicians carry with them the skills, attention, and resources that are useful in the treatment of patients, then the quality of a health care system could increase with the number of practicing physicians. If these physicians were to begin departing due to an increase in malpractice pressure, one possible response from insurers is to raise physicians' compensation to offset the increased malpractice liability costs and induce them to remain. An insurer could take this action a step further by raising compensation enough to attract physicians from other

jurisdictions in order to spread the malpractice risk of treating the insurer's policyholders. This would leave insurers in other jurisdictions facing outflows of physicians, and could induce them to take like action. In this way, rising malpractice pressure in one jurisdiction could cause increases in health care spending in multiple jurisdictions. On the other hand, if the initial insurer chose not to raise physician compensation, the outflow could create more competition among physicians in other jurisdictions and thus lower the cost of care. For these reasons, it is not obvious how mobile physicians' reactions to changes in malpractice pressure would affect health care spending and quality across other jurisdictions in equilibrium.

The existence of cross-jurisdictional effects would introduce important considerations into the subject of defensive medicine. First, it would drive a wedge between the direct, or jurisdictional, effects of rising malpractice pressure and the aggregate effects. This is relevant to the idea of tort reform as a strategy to decrease national health care spending, because estimating the direct effect of state reforms on state variables and extrapolating to the national level would not capture the aggregate effect. Second, failing to account for cross-jurisdictional effects could introduce bias into estimates of the direct effects of changes in malpractice pressure. For example, if rising malpractice pressure in jurisdiction  $i$  caused increases in health care spending in most other jurisdictions as well as  $i$ , the estimated effect on jurisdiction  $i$  would be biased toward zero. Alternatively, if the effects were predominantly negative, the estimated

effect would be positive and inflated.

One goal of this paper is to investigate whether physician mobility creates cross-jurisdictional effects from rising malpractice pressure, and if so, what kind of bias this would create in an empirical study that didn't account for physician mobility. This is done using a theoretical model of the interactions between consumers, physicians, and health insurers. These decision makers reside in one of two jurisdictions. It is assumed that consumers are immobile, and must therefore make all of their decisions within their jurisdiction of origin. Through an assumption on the competitiveness of the health insurance market described below, health insurers are also rendered effectively immobile. Physicians, on the other hand, are able to locate in either jurisdiction. In equilibrium, therefore, physicians would only occupy a jurisdiction in which they have a weak preference to practice.

While the most literal interpretation of these jurisdictions is a geographical one, there is also a more abstract interpretation. They can represent mutually exclusive specialties, subspecialties, or other "options" that a physician could pursue. For example, a location decision for obstetrician/gynaecologists would consist of whether to focus exclusively on obstetrics or gynaecology, or the share of their practice to devote to one or the other. This allows the model to serve a second function: determining whether and how changes in a common level of malpractice pressure would differentially affect two distinct populations of consumers. This exercise sheds some light on the equity considerations of rising

malpractice pressure.

To determine the qualitative effects on health care spending and quality, the model is solved numerically for various cases of rising malpractice pressure. Cross-jurisdictional effects were confirmed in every specification of the two-jurisdiction case investigated. Changes in one jurisdiction's malpractice pressure affected the other jurisdiction's health care spending and quality, as well as access to care. Based on the signs of these effects, results indicate that cross-jurisdictional effects due to physician mobility in a  $k$ -jurisdiction case would inflate estimates of the effect of malpractice pressure on measures of health care system quality. On the other hand, the effects on health care spending can be biased either toward or away from zero, depending on other jurisdictions' approach to competition for physicians. Finally, in investigating the differential effects of defensive medicine, the model predicts that physicians would exit poorer jurisdictions for more wealthy ones. This matches some existing empirical findings, which showed that physicians tend to depart rural areas for urban areas as malpractice pressure rises (General Accounting Office 2003).

## 3.2 The Model

The model environment is made up of two jurisdictions. Each jurisdiction  $i \in \{1, 2\}$  contains a population of identical consumers of measure 1, although



consumers need not be identical across jurisdictions. Each consumer is endowed with income  $m_i$  and is immobile, and thus confined to his jurisdiction of origin. Similar to Chapter 2, a consumer's preferences are represented by the utility function  $U_i(y, H)$ , where  $y$  is consumption and  $H$  is the consumer's health status. The utility function is continuous, differentiable, and strictly concave. In the absence of health insurance, the health status of a consumer in jurisdiction  $i$  is a binary random variable, taking the value  $H_{i1}$  (healthy) with probability  $1 - q_i$  and  $H_{i2}$  (ill) with probability  $q_i$ . Before the value of a consumer's health status is revealed, he can purchase a health insurance policy at a price of  $\tau_i$ . Health insurance allows consumers who become ill to recover their healthy status with probability  $Q_i$ , which is labelled the "quality" of the health insurance available to consumers in jurisdiction  $i$ . Therefore, the expected utility of an insured consumer in jurisdiction  $i$  is:

$$EU_i = (1 - q_i + q_i Q_i) U_i(m_i - \tau_i, H_{i1}) + (q_i - q_i Q_i) U_i(m_i - \tau_i, H_{i2})$$

There is a continuum of physicians of measure  $D$ . Each physician is endowed with  $s$  rivalrous units of resources for use in the treatment of ill consumers, otherwise known as patients. Examples of these kinds of resources include the physician's time and attention. By expending resources  $t$  in the

treatment of a given patient, a physician increases the probability that the patient recovers from 0 to  $1 - \rho_i(t)$ . The function  $\rho_i(t)$  can be considered the probability of an adverse outcome, where the patient remains in the poor health status despite receiving treatment. It is assumed to be positive, decreasing in  $t$ , and strictly convex, where  $\rho_i(0) = 1$  and  $\lim_{t \rightarrow \infty} \rho_i(t) = 0$ . These assumptions are designed to impose diminishing returns of physicians' endowed resources on the treatment of patients.

Each adverse outcome in jurisdiction  $i$  brings an uninsurable expected malpractice liability cost of  $P_i$  upon the treating physician. This parameter, which serves as a measure of “malpractice pressure” in the model, is a composite of several factors contributing to malpractice liability costs. These include the likelihood that adverse outcome leads to a lawsuit, the reputational and psychic costs of participating in a trial, the prospect of malpractice awards exceeding the limits of malpractice insurance, etc. This means that the expected liability cost from treating a patient in jurisdiction  $i$  with  $t$  units of resources is  $\rho_i(t) \cdot P_i$ .

Each physician  $j$  must choose the jurisdiction in which to locate her medical practice ( $c_j \in \{1, 2\}$ ), and given such choice, must also choose the number of cases to take on ( $n_i$ ). Each patient actually treated in jurisdiction  $i$  brings in a payment of  $w_i$ . Also, since all consumers in a given jurisdiction are identical and  $\rho_i(\cdot)$  is convex and decreasing, then for any choice  $n_i$ , total liability costs will be minimized where each patient receives an equal amount of resources

in their treatment. Therefore, given the choice to locate in jurisdiction  $i$ , each physician's profit-maximizing caseload size ( $n_i^*$ ) would solve:

$$\max_{n_i \geq 0} \left\{ w_i n_i - n_i \cdot \rho_i \left( \frac{s}{n_i} \right) P_i \right\}$$

In making a location decision, physicians' payoffs are made up of two components. The first is the net returns from practicing medicine in jurisdiction  $i$ , labelled  $\pi_i$ . Let  $D_1$  and  $D_2$  be the measures of physicians practicing in jurisdictions 1 and 2 respectively. Since the total measure of ill consumers in jurisdiction  $i$  is  $q_i$ , the maximum number of patients that each physician could actually treat is  $\frac{q_i}{D_i}$ . Let  $\tilde{n}_i = \min \left\{ n_i, \frac{q_i}{D_i} \right\}$  be the actual number of patients the physician treats. Therefore, net returns are total revenues ( $w_i \tilde{n}_i$ ) minus total expected malpractice liability costs  $\tilde{n}_i \rho_i(\cdot) P_i$ . The second component is an idiosyncratic locational preference of physician  $j$  for practicing in jurisdiction  $i$  ( $\epsilon_{ij}$ ). Let  $\epsilon_{1j} = 0$  for all  $j$  and  $\epsilon_{2j}$  be distributed according to the cumulative distribution function  $F(\epsilon_2)$  over the support  $(-\infty, \infty)$ . This means that  $\epsilon_{2j}$  is physician  $j$ 's relative preference for practicing in jurisdiction 2 instead of jurisdiction 1. Let  $dF(\cdot)$  be symmetric around the origin to abstract away from any systematic preference for one jurisdiction over another. Therefore, physician  $j$ 's location choice  $c_j^*$  is jurisdiction  $i$  if and only if:

$$\pi_i + \epsilon_{ij} \geq \pi_k + \epsilon_{kj} \quad i, k \in \{1, 2\}, i \neq k$$

$$\text{where} \quad \pi_i = \pi_i(w_i, P_i) = \max_{\tilde{n}_i \geq 0} \left\{ w_i \tilde{n}_i - \tilde{n}_i \cdot \rho_i \left( \frac{s}{\tilde{n}_i} \right) P_i \right\}$$

Since  $\pi_i(w_i, P_i)$  is a function of  $\tilde{n}_i$  instead of  $n_i^*$ , location decisions are based on the actual number of patients each physician would treat in jurisdiction  $i$  rather than that which the physician would like to treat. Since  $n_i^*$  is unaffected by  $D_i$ , while  $\tilde{n}_i$  may be affected by  $D_i$ , the net returns from practicing that each physician  $j$  uses to make her location decision must be consistent with the location decisions of every other physician ( $c_{l \neq j}$ ).

Assume that the market for health insurance is perfectly competitive in each jurisdiction. This means that, in equilibrium, the insurance policies offered to consumers in jurisdiction  $i$  must be that which maximizes expected consumer utility subject to a zero-profit constraint. All results are identical whether there is a single insurer in each jurisdiction, or alternatively, a single insurer across jurisdictions facing potential jurisdiction-based entrants. All insurers operating in this environment are managed care organizations (MCO), and as such, sign contracts with both consumers (for  $Q_i$  and  $\tau_i$ ) as well as a contract with physicians. As a simplification of the model in Chapter 2, the contract between an MCO in jurisdiction  $i$  and a physician consists of a payment ( $w_i$ ) for each policyholder the physician treats. It is assumed that MCOs

and physicians cannot contract directly on  $n_i$ . In the real world, this type of contract would be extremely costly to enforce due to the vast heterogeneity across patients and illnesses and the difficulty in verifying illness and proper treatment for the purposes of a contract. Instead, contracts require physicians to exercise judgement in determining whether and what kind of treatment is provided. Even though this model abstracts away from heterogeneity across patients within a given jurisdiction, it assumes the contracting problem to exist without explicitly modelling it. Therefore, the number of patients treated enters each MCO's problem as an incentive compatibility constraint rather than a choice variable.

Given the structure of the contracts, the competition in the health insurance market, and the need for incentive compatibility, the  $\{Q_i, \tau_i\}$  offered by an MCO to consumers depends on the choice of  $w_i$ , and the resulting behaviour of physicians. A patient's recovery in this model is the result of two events. First, the patient must gain access to a physician in order to receive medical services. If patients are allocated randomly across physicians in a given jurisdiction, and the number of patients treated does not exceed the number of ill consumers, then the probability that any patient gains access to care is  $\frac{D_i n_i}{q_i}$ . Second, conditional on gaining access to a physician and receiving treatment, the probability that the treatment is successful and leads to recovery is  $1 - \rho_i(\cdot)$ . Thus, the probability of recovery ( $Q_i$ ) is the product of these two probabilities:

$$Q_i\left(n_i, \frac{s}{n_i}\right) = \left(\frac{D_i n_i}{q_i}\right) \left(1 - \rho_i\left(\frac{s}{n_i}\right)\right)$$

Due to perfect competition, the MCO in jurisdiction  $i$  must set the revenues from the sale of health insurance policies to each of the measure 1 of consumers ( $\tau_i$ ) equal to the costs of treating patients. With  $D_i$  physicians each treating  $\tilde{n}_i$  patients at a cost to the insurer of  $w_i$  per patient, these total costs equal  $D_i \tilde{n}_i w_i$ . Therefore, the MCO in jurisdiction  $i$  solves the problem:

$$\max_{w_i} \left\{ (1 - q_i)U_i(m_i - \tau_i, H_{i1}) + q_i U_i(m_i - \tau_i, H_{i2}) + q_i \tilde{Q}_i \cdot \Delta U_i \right\}$$

where  $\tau_i = D_i \tilde{n}_i w_i$

$$\tilde{Q}_i = Q_i\left(\tilde{n}_i, \frac{s}{\tilde{n}_i}\right) = \left(\frac{D_i \tilde{n}_i}{q_i}\right) \left(1 - \rho_i\left(\frac{s}{\tilde{n}_i}\right)\right)$$

$$\Delta U_i = U_i(m_i - \tau_i, H_{i1}) - U_i(m_i - \tau_i, H_{i2})$$

Given  $\{s, D, \{q_i, m_i, H_{i1}, H_{i2}, P_i\}_{i=1,2}\}$ , equilibrium is each physician  $j$ 's location choice  $c_j^*$ , a physician's choice of caseload size in jurisdiction  $i$ ,  $\{n_i^*(w_i, s, P_i)\}_{i=1,2}$ , and each MCO's choice of a payment per patient,  $\{w_i^*\}_{i=1,2}$  such that:

- 1) given  $c_{l \neq j}$ ,  $c_j^* = \begin{cases} 1 & \text{if } \pi_1 \geq \pi_2 + \epsilon_{2j}; \\ 2 & \text{otherwise.} \end{cases}$
- 2)  $n_i^*(w_i, s, P_i) = \arg \max_{n_i \geq 0} \left\{ w_i n_i - n_i \cdot \rho_i \left( \frac{s}{n_i} \right) P_i \right\}$
- 3)  $w_i^* = \arg \max_{w_i \geq 0} \left\{ EU_i \mid \tau_i = D_i \tilde{n}_i w_i \right\}$

### 3.3 Equilibrium: Analytical

The condition characterizing a physician's optimal caseload size  $n_i^*(w_i, s, P_i)$  is the same as in Chapter 2. A physician practicing in jurisdiction  $i$ , and thus facing  $w_i$  and  $P_i$ , would like to set  $n_i$  such that:

$$\frac{w_i}{P_i} = \rho \left( \frac{s}{n_i} \right) - \left( \frac{s}{n_i} \right) \cdot \rho' \left( \frac{s}{n_i} \right)$$

where  $\rho'(\cdot)$  is the first derivative of  $\rho(\cdot)$ . A unique finite solution exists for all  $w_i \in [0, P_i)$ . Changes in  $w_i$  and  $P_i$  produce the opposite effects on  $n_i^*$ . An increase in  $w_i$  increases the profitability of the marginal patient and induces physicians to increase their caseloads, thus both increasing total revenues as well as taking on greater liability exposure. An increase in  $P_i$  makes the marginal patient too risky to treat, causing physicians to reduce liability exposure by taking on fewer patients.

Physicians' location decisions depend on the actual number of patients they would treat in each jurisdiction  $\tilde{n}_i$  rather than the number they would like to treat. This is important in determining whether or not  $n_i^*$  is a best response of a physician locating in jurisdiction  $i$ . If  $n_i^* < \frac{q_i}{D_i}$ , then  $n_i^*$  is the only element of the argmax in the physician's problem. The case of  $n_i^* \geq \frac{q_i}{D_i}$  is slightly more complicated since  $\tilde{n}_i$  is unchanging as  $n_i^*$  increases. In this case, therefore, the argmax consists of the set  $\left[\frac{q_i}{D_i}, \infty\right)$ . However, since  $n_i^* \in \left[\frac{q_i}{D_i}, \infty\right)$  in this case, it is therefore always a best response for every physician practicing in jurisdiction  $i$  to choose the caseload size  $n_i^*$ .

Physician  $j$  must compare the net returns from practicing in the two jurisdictions. Since there is a continuum of physicians, the measure of physician  $j$  is infinitesimal. Therefore, the location decisions of every other physician ( $c_{l \neq j}$ ) result in  $D_1$  and  $D_2$  physicians practicing in jurisdictions 1 and 2 respectively. Given ( $c_{l \neq j}$ ), and thus  $D_1$  and  $D_2$ , it is optimal for physician  $j$  to locate in jurisdiction 1 if  $\epsilon_{2j} \leq \pi_1 - \pi_2$ , and in jurisdiction 2 otherwise. Given the distributional assumptions on  $\epsilon_{2j}$ , and that this decision rule must hold for all  $j$ , it must be true in equilibrium that  $D_1 = D \cdot F(\pi_1 - \pi_2)$  and  $D_2 = D \cdot (1 - F(\pi_1 - \pi_2))$ .

The first-order condition from MCO  $i$ 's problem is:

$$q_i \Delta U_i \frac{\partial \tilde{Q}_i}{\partial w_i} = W MU_{y^i} \frac{\partial \tau_i}{\partial w_i} \quad (3.1)$$



where  $WMU_{yi}$  is the “weighted marginal utility of consumption” in jurisdiction  $i$ , and can be considered the expected marginal utility of consumption for a consumer after purchasing health insurance.

The left side of Equation 3.1 is the marginal benefit of increasing physician payments in jurisdiction  $i$  ( $MB_i$ ). It is the increase in the likelihood  $\left(\frac{\partial \bar{Q}_i}{\partial w_i}\right)$  that the ill consumers in jurisdiction  $i$  (with measure  $q_i$ ) will receive the increase in utility that arises due to recovery ( $\Delta U_i$ ). The right side is the marginal cost ( $MC_i$ ). It is the value of the consumption forgone as the price of insurance increases to fund the increased physician payments.

If the Inada conditions hold, then  $w_i^* \in \left[0, \frac{m_i}{q_i}\right)$ . As  $w_i$  approaches the upper bound of this range, consumption would be driven to zero, causing  $WMU_{yi}$  to approach infinity and the first-order condition to be violated. The MCOs’ problems are complicated because neither  $\frac{\partial \bar{Q}_i}{\partial w_i}$  nor  $\frac{\partial \tau_i}{\partial w_i}$  is necessarily continuous in  $w_i$ . The principal effects of increasing  $w_i$  are first, an increase in physicians practicing in jurisdiction  $i$   $\left(\frac{\partial D_i}{\partial w_i} \geq 0$  since  $\frac{\partial \pi_i}{\partial w_i} \geq 0\right)$  and second, each physician desiring a greater caseload  $\left(\frac{\partial n_i^*}{\partial w_i} \geq 0\right)$ . As long as  $n_i^* < \frac{q_i}{D_i}$ , the product  $D_i \tilde{n}_i$  increases in  $w_i$ . Once  $n_i^* \geq \frac{q_i}{D_i}$ , the product equals  $q_i$ ; a constant. This is because, even though the higher payments induce more physicians to locate in jurisdiction  $i$ , the scarcity of ill consumers results in each of those physicians actually treating fewer patients, even though they would like to treat more. This results in a discontinuous negative shift in both  $\frac{\partial \bar{Q}_i}{\partial w_i}$  and  $\frac{\partial \tau_i}{\partial w_i}$  once  $w_i$  is such that

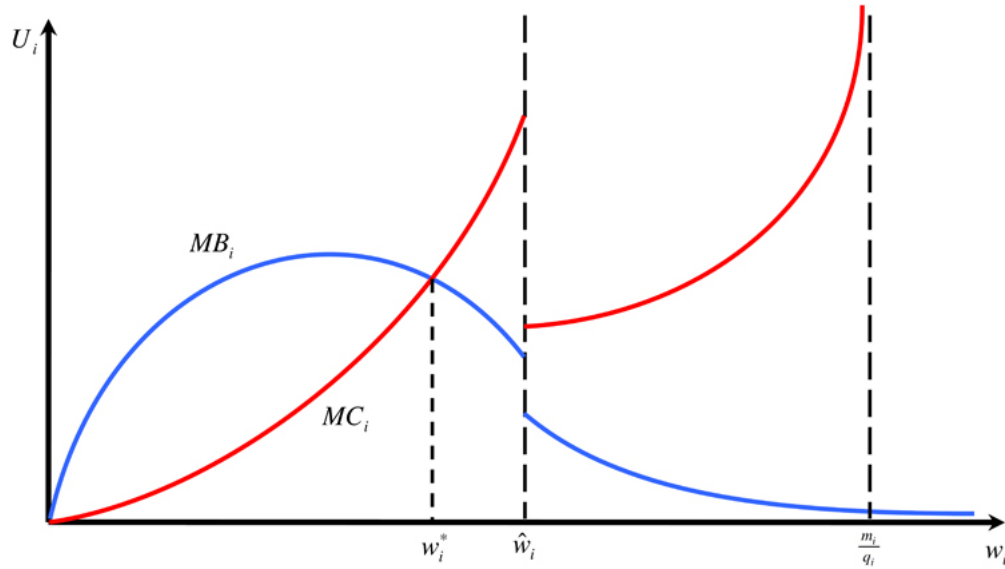


Figure 3.1: Limited-Access Solution

$n_i^* = \frac{q_i}{D_i}$ . This value of  $w_i$  is labelled “ $\hat{w}_i$ ” in Figures 3.1 to 3.3. The discontinuities mean that an equilibrium could exist where the first-order conditions do not hold, as in Figure 3.3. That is, it could be optimal for an MCO to set  $w_i = \hat{w}_i$ . A jurisdiction  $i$  in this kind of equilibrium would be characterized by a condition other than its first-order condition, namely  $n_i^* = \frac{q_i}{D_i}$ .

Altogether, there are three types of solution for each MCO’s problem: limited-access ( $n_i^* < \frac{q_i}{D_i}$ ), full-access-corner ( $n_i^* = \frac{q_i}{D_i}$ ), and full-access-interior ( $n_i^* > \frac{q_i}{D_i}$ ). These different types of solution are shown in Figures 3.1 to 3.3. Since there are two MCOs modelled here, the possible combinations of the three potential solution types result in six potential equilibrium types. Even though there is the potential for a no-insurance equilibrium ( $w_i^* = 0$ ), this case will not be

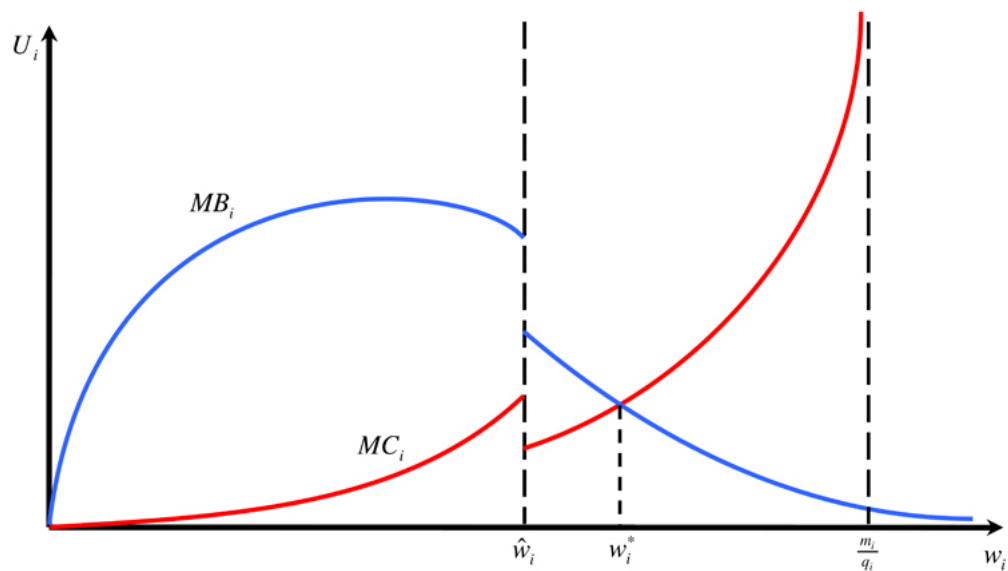


Figure 3.2: Full-Access-Interior Solution

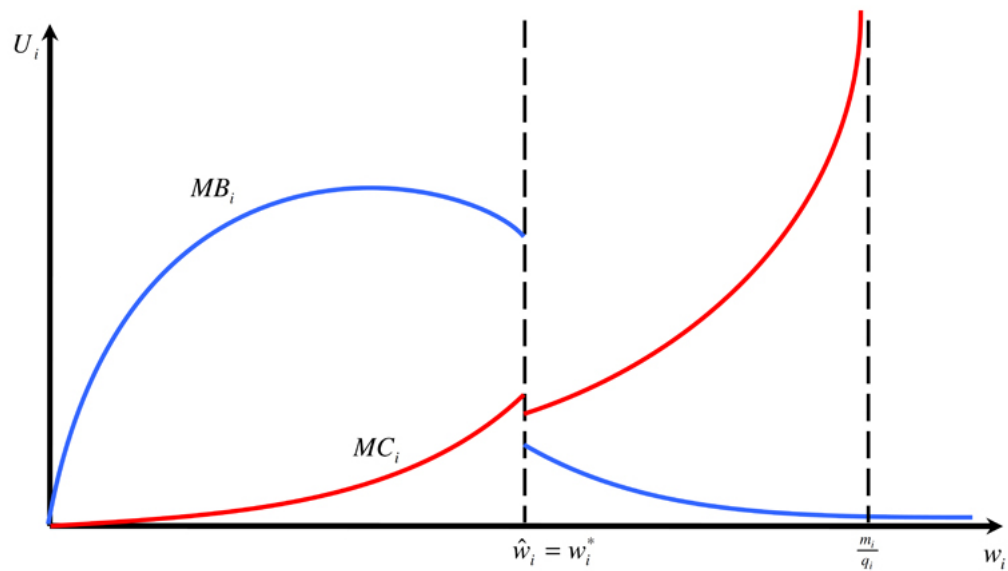


Figure 3.3: Full-Access-Corner Solution

investigated here since it is easily distinguished from the other potential equilibrium types. Each set of these equilibrium choices  $\{w_i^*\}_{i=1,2}$  is characterized by a pair of conditions for  $i = 1, 2$ :

$$\begin{aligned} \text{if } n_i^*(w_i^*, s, P_i) &\neq \frac{q_i}{D_i}, & q_i \Delta U_i \frac{\partial \tilde{Q}_i}{\partial w_i} &= WMU_{y_i} \frac{\partial \tau_i}{\partial w_i} \\ \text{else} & & n_i^*(w_i^*, s, P_i) &= \frac{q_i}{D_i} \end{aligned}$$

### 3.4 Equilibrium: Numerical

In order to investigate these equilibria numerically, assume the following functional forms:

$$\rho_i(t) = \frac{1}{1 + \alpha_i t}, \quad U_i(y_i, H_i) = H_i^{\beta_i} y_i^{1-\beta_i}, \quad \epsilon_{2j} \sim \mathcal{N}(0, \sigma^2)$$

where  $\alpha_i > 0$ ,  $0 < \beta_i < 1$ , and  $\sigma^2 > 0$ . Besides complying with the conditions posed earlier, the choice of functional form for  $\rho_i(\cdot)$  is convenient as it yields a closed form solution for  $n_i^*$ , which is non-negative and finite for all  $w_i \in [0, P)$ :

$$n_i^* = \alpha_i s \left[ \left( \frac{P_i}{P_i - w_i} \right)^{\frac{1}{2}} - 1 \right]$$

The MCO's problem is complicated by the relationships between  $w_i$ ,  $\pi_i$ , and  $D_i$ . Given a choice of  $w_1$  and  $w_2$ ;  $\pi_1$ ,  $\pi_2$ ,  $D_1$ , and  $D_2$  would be determined simultaneously by the system of equations:

$$\begin{aligned}\pi_1 &= w_1 \tilde{n}_1 - \tilde{n}_1 \rho_1 \left( \frac{s}{\tilde{n}_1} \right) P_1 \\ \pi_2 &= w_2 \tilde{n}_2 - \tilde{n}_2 \rho_2 \left( \frac{s}{\tilde{n}_2} \right) P_2 \\ D_1 &= D \cdot F(\pi_1 - \pi_2) \\ D_2 &= D \cdot (1 - F(\pi_1 - \pi_2))\end{aligned}$$

Since  $\tilde{n}_i$  is case-specific, and these cases depend on  $D_i$ , this system is computationally difficult to work with. Alternatively, since  $D_i$  is monotonically increasing in  $\pi_i$ , and  $\pi_i$  is monotonically increasing in  $w_i$ , any pair  $\{\pi_1, \pi_2\}$  produces the unique pairs  $\{D_1(\pi_1, \pi_2), D_2(\pi_1, \pi_2)\}$  and  $\{w_1(\pi_1, \pi_2), w_2(\pi_1, \pi_2)\}$  such that:

$$D_1(\pi_1, \pi_2) = D \cdot F(\pi_1 - \pi_2)$$

$$D_2(\pi_1, \pi_2) = D \cdot (1 - F(\pi_1 - \pi_2))$$

$$w_1(\pi_1, \pi_2) = \begin{cases} w \text{ such that } \pi_1 = wn_1^* - n_1^*\rho_1\left(\frac{s}{n_1^*}\right)P_1 & \text{if } n_1^*(w, s, P_1) < \frac{q_1}{D_1(\pi_1, \pi_2)}; \\ \left(\frac{D_1(\pi_1, \pi_2)}{q_1}\right)\pi_1 + \rho_1\left(\frac{D_1(\pi_1, \pi_2)s}{q_1}\right)P_1 & \text{otherwise.} \end{cases}$$

$$w_2(\pi_1, \pi_2) = \begin{cases} w \text{ such that } \pi_2 = wn_2^* - n_2^*\rho_2\left(\frac{s}{n_2^*}\right)P_2 & \text{if } n_2^*(w, s, P_2) < \frac{q_2}{D_2(\pi_1, \pi_2)}; \\ \left(\frac{D_2(\pi_1, \pi_2)}{q_2}\right)\pi_2 + \rho_2\left(\frac{D_2(\pi_1, \pi_2)s}{q_2}\right)P_2 & \text{otherwise.} \end{cases}$$

without having to simultaneously solve a system of equations. Therefore, instead of choosing  $w_i$ , each MCO will do the equivalent of choosing  $\pi_i$  to solve its problem given the other MCO's choice of  $\pi_k$ .

The program solves MCO  $i$ 's problem given a parameterized value of the other MCO's net return from practicing ( $\bar{\pi}_{k0}$ ). It then compares the solution to that problem  $\pi_i^*(\bar{\pi}_{k0})$  with the initial parameterized value used in the other MCO's problem ( $\bar{\pi}_{i0}$ ). If  $|\pi_i^*(\bar{\pi}_{k0}) - \bar{\pi}_{i0}|$  for either MCO is greater than some tolerance parameter, the program replaces both  $\bar{\pi}_{i0}$  with  $\bar{\pi}_{i1} = \pi_i^*(\bar{\pi}_{k0})$  and then resolves the two MCO's problems until the solutions converge to the updated parameterized values.

## 3.5 Results

The purpose of the numerical simulations is to examine the effects of rising malpractice pressure on variables of interest in the two jurisdictions. This is done in three parts. Part 1 examines the effect of rising malpractice pressure in jurisdiction 1 while the malpractice pressure in jurisdiction 2 is held constant at a level such that MCO 2 offers full access to its policyholders. Other than the difference in malpractice pressure, the two jurisdictions are identical in every parameter. Part 2 does the same, with the exception that malpractice pressure in jurisdiction 2 is held constant and is high enough to induce MCO 2 to provide only limited access. These exercises allow for investigation into possible systematic biases in the coefficients uncovered by studies using data at the jurisdiction-level (state or county) without accounting for physician mobility. Part 3 has a different purpose from Parts 1 and 2. It is designed to investigate the effects on two distinct populations of increases in a common level of malpractice pressure ( $P_1 = P_2 = P$ ) when physicians can decide whether and how much to focus on each population. In this context, a population is equivalent to a jurisdiction, where one group has higher income than the other. This allows for investigation into the differential effects of rising malpractice pressure on rich versus poor consumers.

As parameter values in all three parts, assume that the measure of the entire set of physicians ( $D$ ) is 0.2 and that each physician possesses resources ( $s$ ) equal to 50. The variance on physicians' relative locational preference ( $\sigma^2$ )

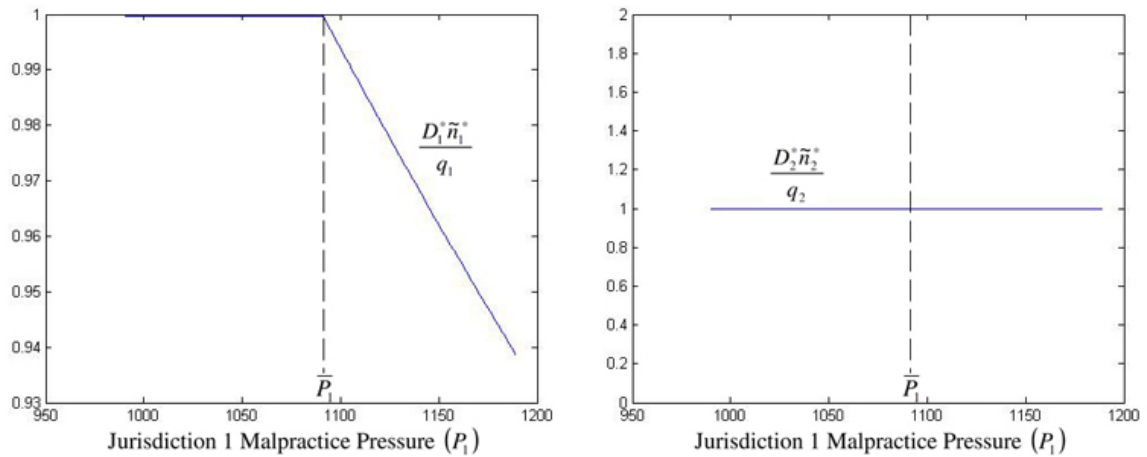
is 50. The technology parameter of converting resources into successful outcomes from treatment ( $\alpha_i$ ) is set at 2. The probability of becoming ill in either jurisdiction ( $q_i$ ) is 0.5, and health statuses  $H_{i1}$  and  $H_{i2}$  are 1 and 0.5 respectively. For Parts 1 and 2, consumer income in either jurisdiction ( $m_i$ ) is set at 100. In Part 3, while  $m_1$  remains at 100,  $m_2$  is reduced to 90. As a tolerance parameter used in the numerical optimization, equilibrium  $\{\pi_1^*, \pi_2^*\}$  is considered found on the  $z^{th}$  iteration if and only if  $|\pi_i^*(\bar{\pi}_{kz}) - \bar{\pi}_{iz}| \leq 0.1$  for both jurisdictions.<sup>1</sup> Let the equilibrium values for the number of doctors emerging in jurisdiction  $i$  in equilibrium be  $D_i^* = D_i(\pi_1^*, \pi_2^*)$ , each physician's caseload size be  $\tilde{n}_i^* = \min\{n_i^*(w_i(\pi_1^*, \pi_2^*), s, P_i), \frac{q_i}{D_i^*}\}$ , and equilibrium health care system quality be  $\tilde{Q}_i^* = Q_i\left(\tilde{n}_i^*, \frac{s}{\tilde{n}_i^*}\right)$ .

### 3.5.1 Part 1: Jurisdiction 2 at Full-Access

The first numerical exercise examines the set of  $P_1$  values [990, 1189] and a single  $P_2$  value of 200. As shown in Figure 3.4, these values are chosen to cover the transition in jurisdiction 1 from full-access to limited-access solutions while jurisdiction 2 only provides full access to its policyholders. Jurisdiction 1 exhibits the same relationship between malpractice pressure and access found in Chapter 2. MCO 1 provides consumers with a health insurance policy with full access to physicians as long as consumers are willing to pay for it. As malpractice pressure rises, this willingness to pay holds initially, but eventually the cost of

<sup>1</sup>The value of  $\pi_i^*$  never fell below 130 for any parameter values investigated, making this a relatively narrow tolerance.

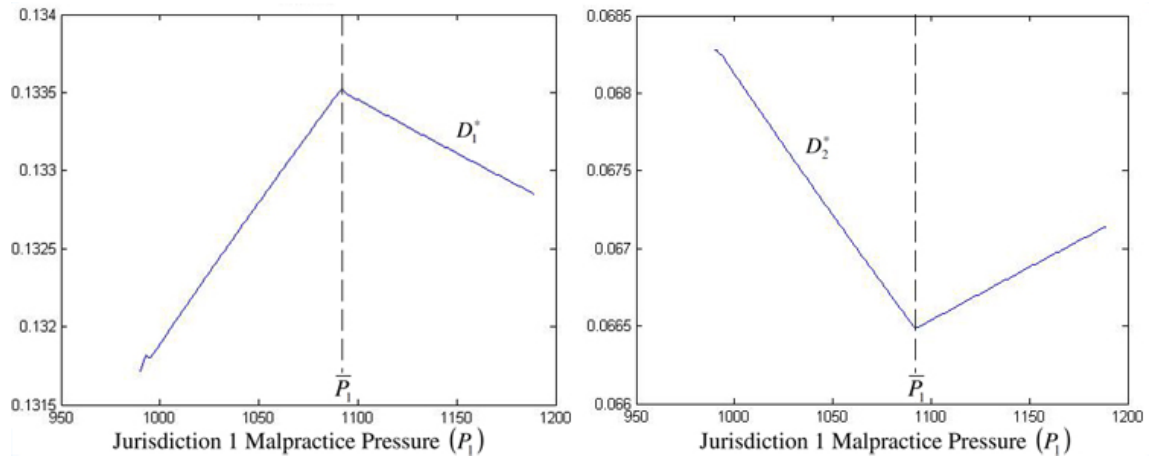




**Figure 3.4:** The effects of rising malpractice pressure in jurisdiction 1 on patient access to physicians in both jurisdictions. Malpractice pressure in jurisdiction 2 is such that MCO 2 provides full access to its policyholders.

maintaining full access becomes so high that consumers would rather keep more of their income for consumption and instead purchase cheaper health insurance with imperfect access to physicians.

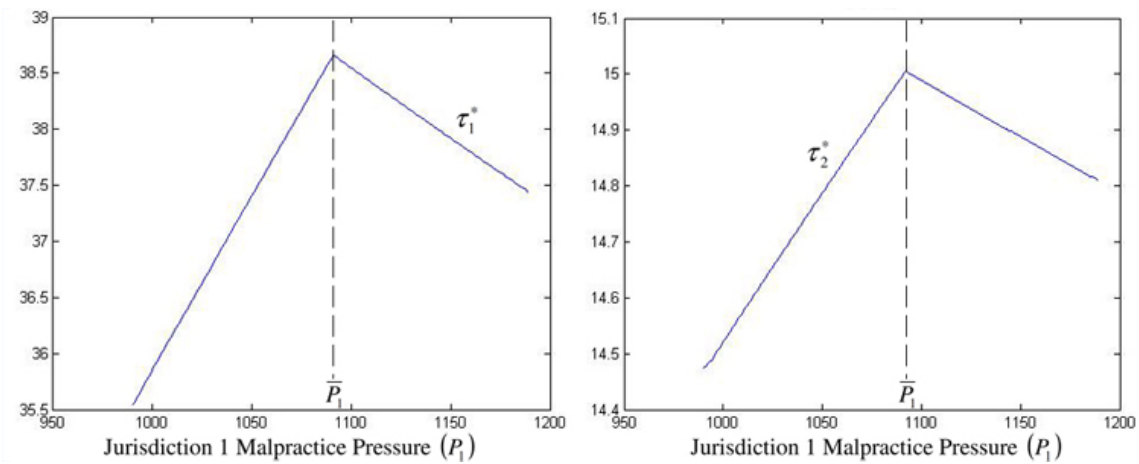
As seen in Figure 3.5, physicians flow into jurisdiction 1 as long as MCO 1 maintains full access in the face of rising malpractice pressure. This means that, even though malpractice pressure is rising in jurisdiction 1, consumers in jurisdiction 1 are willing to pay for increases to physician payments that are sufficient to attract physicians willing to relocate. There are two reasons why an MCO facing rising malpractice pressure might adjust its contracts to draw in more physicians. First, the additional resources brought by the new physicians make each patient less costly to treat. Also, full-access health insurance is cheaper to provide when there are many physicians instead of a few. This



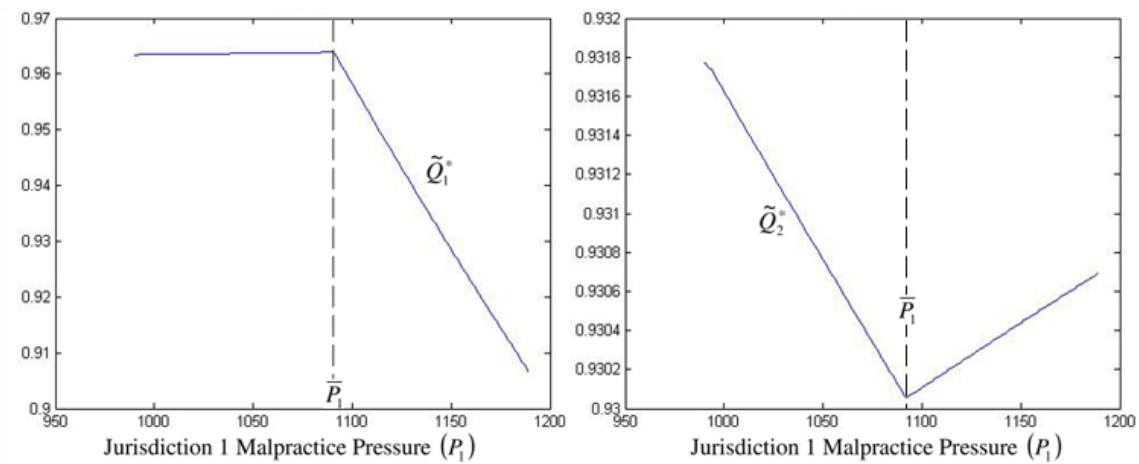
**Figure 3.5:** The effects of rising malpractice pressure in jurisdiction 1 on the movement of physicians when jurisdiction 2 provides full access.

is because each physician's total malpractice liability costs are convex and increasing in  $n_i$ . This means that two physicians could treat a given number of patients at a lower cost than could one physician. Therefore, rising malpractice pressure induces MCOs to attract more physicians as additional resources and cost savings they bring become more significant.

Even though some physicians flow between jurisdictions in equilibrium, it is clear in Figure 3.6 that rising malpractice pressure causes two MCOs intent on providing full access to compete for physicians. Even though malpractice pressure in jurisdiction 2 remains constant, MCO 2 must raise the compensation it provides. This reduces the outflow of physicians and induces the remaining physicians to take on those patients who would have been treated by their departing colleagues. This shows the different ways that the two jurisdictions are affected by rising jurisdiction 1 malpractice pressure. First, as a

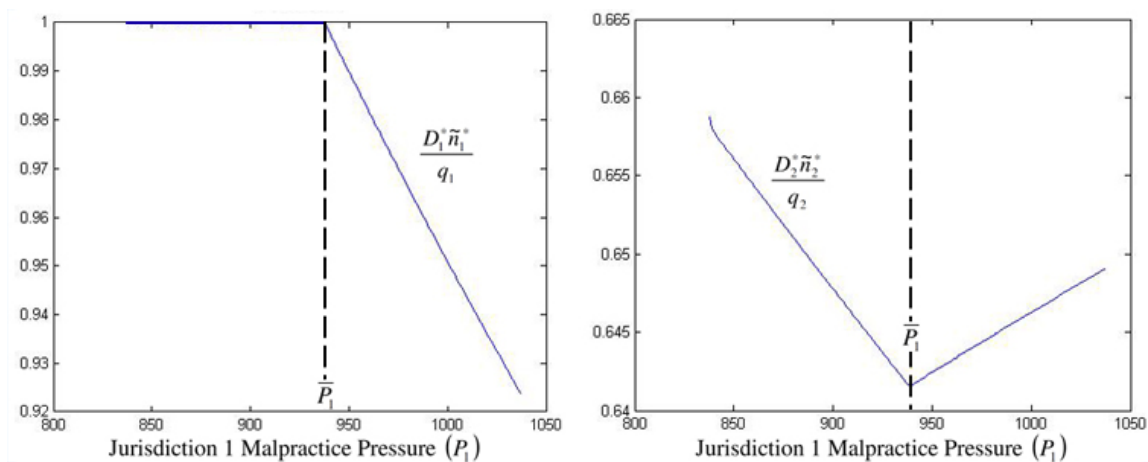


**Figure 3.6:** The effects of rising malpractice pressure in jurisdiction 1 on health care spending in both jurisdictions when jurisdiction 2 provides full access.



**Figure 3.7:** The effect of rising malpractice pressure in jurisdiction 1 on health care quality in both jurisdictions when jurisdiction 2 provides full access.

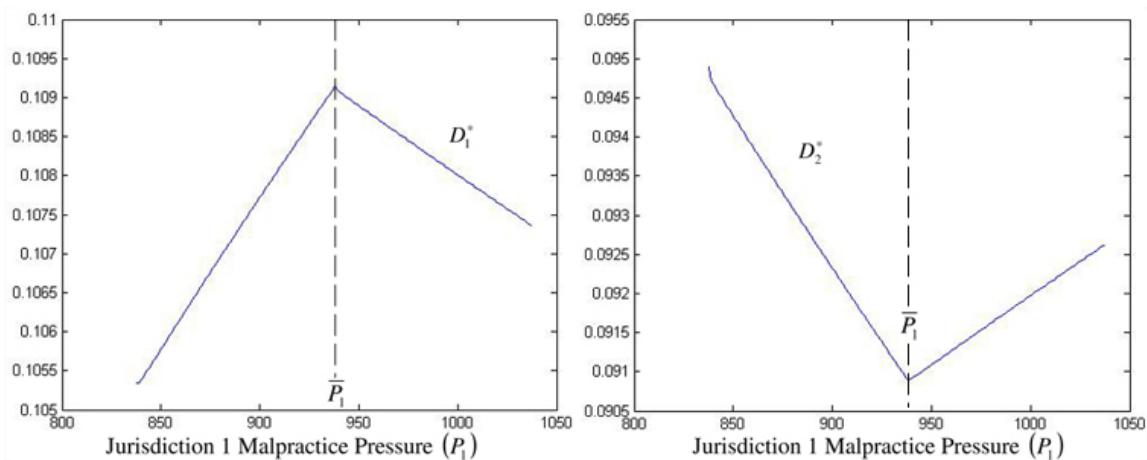
jurisdiction's malpractice pressure rises, physicians become more valuable in a jurisdiction where consumers are willing to pay for full access. This is because malpractice pressure makes a given  $n_i^*$  more costly to induce, and thus  $D_i$  relatively less costly in achieving full access. This jurisdiction would shift toward more physicians, each treating fewer patients, in order to spread out the increased malpractice liability costs. Second, due to the efforts of consumers and insurers in jurisdiction 1 to raise physician payments, physicians become more costly for jurisdiction 2 to retain. Since malpractice pressure is unchanged in jurisdiction 2, the cost of inducing any  $n_2^*$  is unchanged. The best response is therefore to retain fewer physicians but have each one treat more patients in order to maintain full access. Once malpractice pressure in jurisdiction 1 is high enough to cause MCO 1 to forgo full access, jurisdiction 2 becomes a more favourable option to physicians and this competition decreases. The pattern of effects on health care system quality shown in Figure 3.7 reflect the flow of physicians. Given the signs of the direct and cross-jurisdictional effects in Figures 3.4 to 3.7, and extrapolating to a  $k$ -jurisdiction case, the presence of cross-jurisdictional effects would bias the estimated effects of malpractice pressure on health care spending biased toward zero, while also inflating the effect on health care quality.



**Figure 3.8:** The effects of rising malpractice pressure in jurisdiction 1 on patient access to physicians in both jurisdictions. Malpractice pressure in jurisdiction 2 is such that MCO 2 provides limited access to its policyholders.

### 3.5.2 Part 2: Jurisdiction 2 at Limited-Access

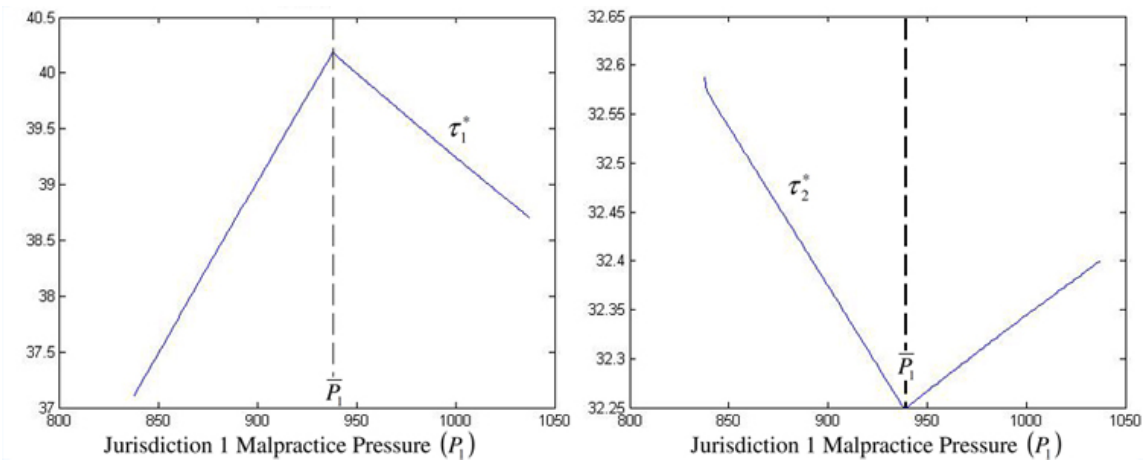
Part 2 performs the same numerical exercise as Part 1, with the exception that  $P_2$  is held constant at 1500 instead of 200. This increase makes access in jurisdiction 2 costly enough to push jurisdiction 2 into a limited-access solution. As shown in Figure 3.8, the same pattern of behaviour from MCO 1 has different effects on access in jurisdiction 2 where consumers in jurisdiction 2 are unwilling to purchase full-access health insurance policies. As malpractice pressure rises and MCO 1 maintains full access, it raises physicians' compensation in order to draw in more physicians. This makes it more costly for MCO 2 to keep physicians in jurisdiction 2. The fact that consumers in jurisdiction 2 prefer limited access to full access shows an unwillingness to pay for better access. As MCO 1 competes for physicians more aggressively, access in jurisdiction 2



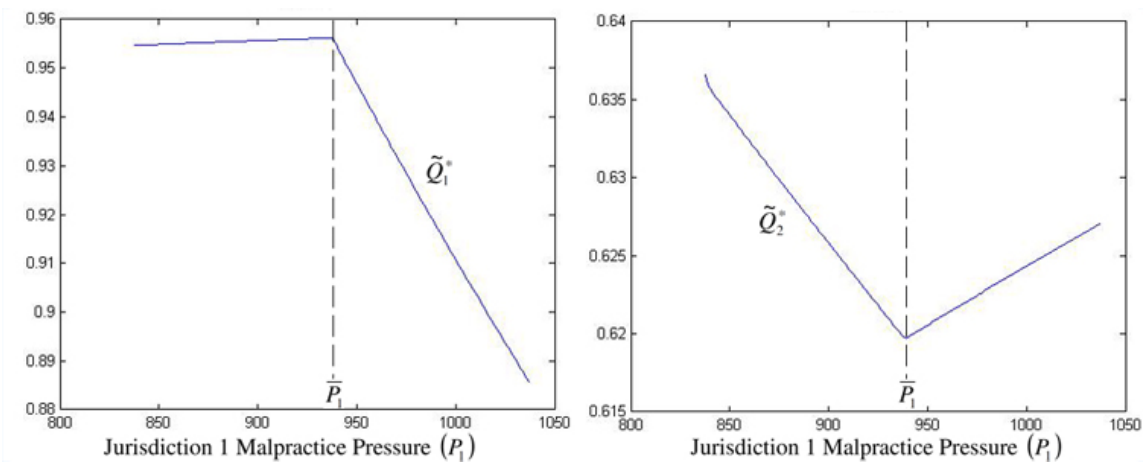
**Figure 3.9:** The effects of rising malpractice pressure in jurisdiction 1 on the movement of physicians when jurisdiction 2 provides limited access.

becomes more costly to provide, and so MCO 2 substitutes away from health insurance.

The MCOs' behaviour creates the same movement of physicians as in Part 1. Figure 3.9 shows that physicians are drawn to the jurisdiction willing to maintain full access as malpractice pressure rises, and then leave the jurisdiction once this willingness is exhausted. As shown in Figure 3.10, however, the effect of increasing jurisdiction 1 malpractice pressure on health care spending in jurisdiction 2 is the opposite of that seen in Part 1. This is because MCO 2, while trying to maintain full access for its consumers, is willing to raise the price of health insurance in order to secure the funds necessary to slow the outflow of physicians. Where consumers in jurisdiction 2 are unwilling to pay for full access, their MCO cannot raise the necessary funds to compete with the aggressive behaviour from MCO 1. Alternatively, it substitutes away



**Figure 3.10:** The effects of rising malpractice pressure in jurisdiction 1 on health care spending in both jurisdictions when jurisdiction 2 provides limited access.



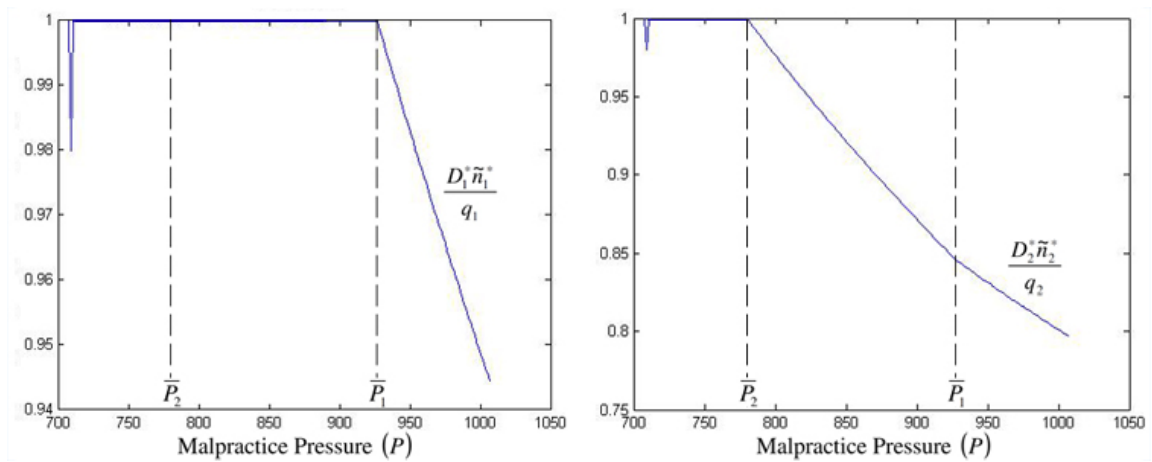
**Figure 3.11:** The effect of rising malpractice pressure in jurisdiction 1 on health care quality in both jurisdictions when jurisdiction 2 provides limited access.

from health insurance by lowering the price and quality of an insurance policy. Once jurisdiction 1 malpractice pressure reaches the point that MCO 1 chooses to provide limited access, the reduction in competitive behaviour makes health insurance less costly to provide in jurisdiction 2. This causes MCO 2 to substitute toward health insurance, thus raising health care spending and quality, as malpractice pressure in jurisdiction 1 increases. Figure 3.11 shows the effects of this behaviour in health care quality in the two jurisdictions. Since it shows the same pattern as Figure 3.7, cross-jurisdictional effects in the  $k$ -jurisdiction case would create the same inflated estimates of malpractice pressure's effect on quality, regardless of other jurisdictions' behaviour surrounding access. Figure 3.10 shows that malpractice pressure's effect on health care spending would be inflated when most other jurisdictions choose to provide their policy holders with limited access; the opposite of the full-access case examined in Part 1.

### **3.5.3 Part 3: Heterogeneous Jurisdictions, Common Malpractice Pressure**

The final numerical exercise examines two jurisdictions that, while facing the same rising level of malpractice pressure ( $P_1 = P_2 = P$ ), contain consumer populations that are different from one another. The purpose is to examine how rising malpractice pressure would differentially affect two distinct populations, between which physicians have some mobility. An obvious application is urban



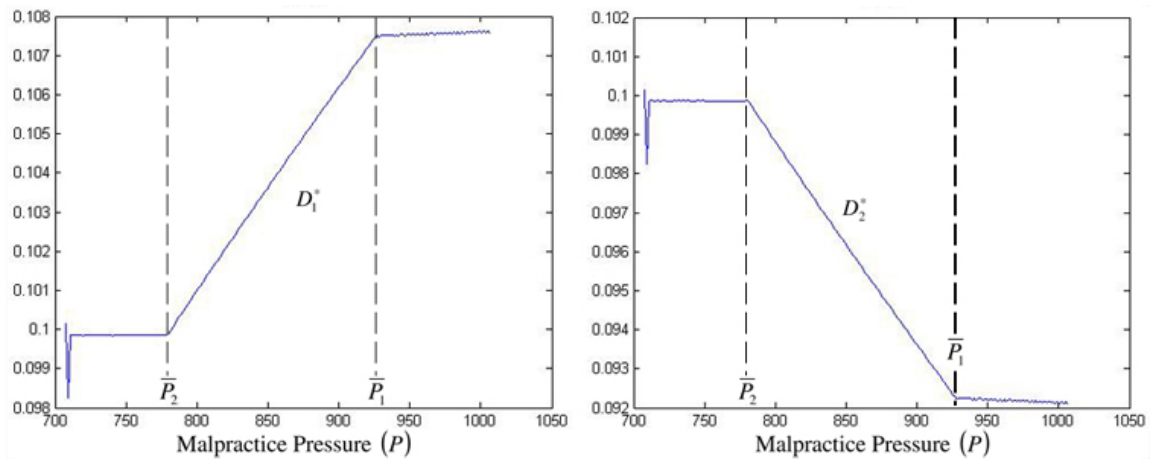


**Figure 3.12:** The effects on patient access to physicians in both jurisdictions when they face the same rising level of malpractice pressure. Consumers in jurisdiction 1 have higher incomes than those in jurisdiction 2.

versus rural consumers, and whether physicians' mobility between urban and rural areas of a state causes the two populations to experience a general rise in malpractice pressure differently from one another<sup>2</sup>. In the exercise, jurisdiction 1 is more wealthy ( $m_1 = 100$ ) than jurisdiction 2 ( $m_2 = 90$ ).

Consumers with relatively high income have a greater willingness to pay for health insurance than those with low incomes. This is why, in Figure 3.12, MCO 1 is willing to bear the cost of maintaining full access up until  $P = \bar{P}_1$ , while MCO 2 must abandon full access at the lower level of  $\bar{P}_2$ . This yields three distinct ranges of malpractice pressure. As long as  $P < \bar{P}_2$ , both MCOs choose to provide full-access health insurance, while for all  $P > \bar{P}_1$ , they offer limited access. Finally, where  $\bar{P}_2 \leq P \leq \bar{P}_1$ , MCO 1 maintains full access while

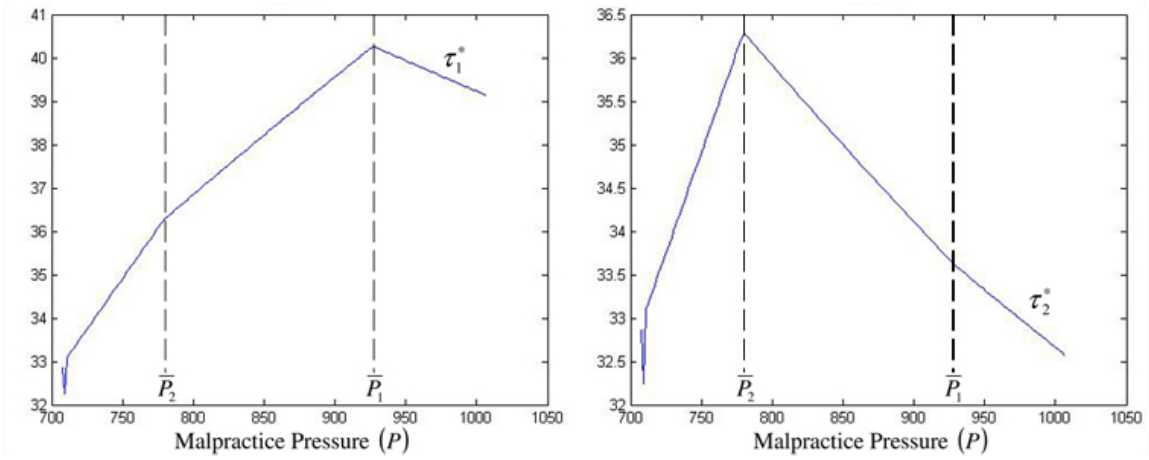
<sup>2</sup>While these populations may not necessarily rely on separate health insurers, the same bundles would be offered in the case of a single insurer facing potential entrants targeting specific populations



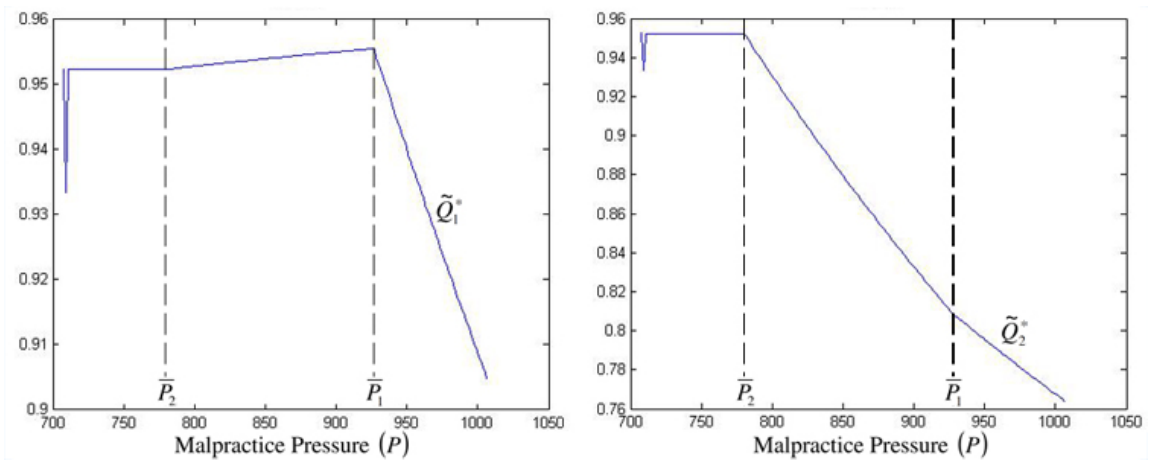
**Figure 3.13:** The effects of rising malpractice pressure in the two jurisdictions on the movement of physicians.

MCO 2 chooses to provide limited access instead.

Figure 3.13 illustrates an important point about physician mobility. As long as the solutions to the two MCOs problems exhibit the same kind of access (full or limited) there is little movement of physicians in equilibrium. For this reason, a lack of observed movement of physicians between jurisdictions does not necessarily indicate that physicians are immobile or that location decisions are unaffected by malpractice pressure. Instead, the lack of movement could indicate a calm surface; where physicians are sensitive to malpractice pressure, but due to escalating or abating competition for physicians among MCOs, their equilibrium numbers in each jurisdiction are unaffected. When physicians do exhibit mobility in equilibrium, it is from areas or populations that are unwilling or unable to pay the cost of full access to those that are. This offers an explanation for the empirical finding that rural populations are particularly



**Figure 3.14:** The effects of rising malpractice pressure the two jurisdictions on health care spending in both jurisdictions.



**Figure 3.15:** The effect of rising malpractice pressure in the two jurisdictions on health care quality in both jurisdictions.

subject to outflows of physicians when malpractice pressure increases (General Accounting Office 2003). This would be the expected outcome since rural areas generally enjoy lower access to health care than do urban areas (Chan et al 2007). Figures 3.14 and 3.15 show the same relationships as in Parts 1 and 2. Rising malpractice pressure induces an MCO willing to pay for full access to compete for physicians more aggressively. Those unwilling to pay for full access respond by substituting away from health insurance, instead providing less expensive and lower quality insurance policies and leaving consumers with more income for consumption.

### **3.6 Summary and Implications**

Several empirical investigations into the existence and extent of the practice of defensive medicine utilize data at the jurisdiction level. The model specifications used by these studies assume that the extent of any effect of rising malpractice pressure on health care cost or quality is confined to that jurisdiction. This ignores the potential cross-jurisdictional effects of changes in malpractice pressure, which could introduce bias into estimates of the effects of changing malpractice pressure on health care spending and quality.

This study finds that physician mobility can result in one jurisdiction's malpractice pressure affecting health care spending and quality in another jurisdiction. These cross-jurisdictional effects drive a wedge between the direct effects of tort reform with the aggregate effects, which is important to the discussion of tort reform as a policy tool for reducing national health care spending. If cross-jurisdictional effects are present in a  $k$ -jurisdiction case, the signs of these effects uncovered in this paper suggest that physician mobility would inflate estimates of the effect of health care spending on health care system quality. The effect on health care spending can be biased toward or away from zero, depending on the prevailing level of access among these jurisdictions. Regarding the differential effects of defensive medicine, the model predicts that for certain ranges of malpractice pressure, competing MCOs will induce mobile physicians to remain immobile in equilibrium. Also, where physicians do relocate, they leave poorer jurisdictions for more wealthy ones, which matches the findings in the empirical literature surrounding the departure of physicians from rural areas to urban areas.

Beyond these results, significant cross-jurisdictional effects of changing malpractice pressure indicate the presence of externalities from jurisdiction-level tort reform. Currently, all malpractice reform has been undertaken at the state level. By changing competition among states for a scarce supply of mobile physicians, malpractice reform in one state could create external costs and/or

benefits for other states. Since these are not internalized by the state considering the malpractice reforms, the malpractice reform passed in equilibrium is almost certain to deviate from the socially efficient amount. This adds to the argument for a federal role in malpractice reform, which could manage tort reform across states and internalize cross-jurisdictional effects in the effort to limit the growth in national health care spending.

### 3.7 Bibliography

- Avraham, R. and M. Schanzenbach. 2010. The Impact of Tort Reform on Private Health Insurance Coverage. *American Law and Economics Review* 1–37.
- Baicker, K., and A. Chandra. 2005. The Consequences of the Growth of Health Insurance Premiums. *AEA Papers and Proceedings* **95**(2) 214–218.
- Baicker, K., E.S. Fisher, and A. Chandra. 2007. Malpractice Liability Costs and the Practice of Defensive Medicine in the Medicare Program. *Health Affairs* **26**(3) 841–852.
- Bovbjerg, R.R., L.C. Dubay, G.M. Kenney, and S.A. Norton. 1996. Defensive Medicine and Tort Reform: New Evidence in an Old Bottle. *J. of Health Politics, Policy and Law* **21**(2) 267–288.
- Chan, L., L.G. Hart, and D.C. Goodman. 2007. Geographic Access to Health Care for Rural Medicare Beneficiaries. *The J. of Rural Health* **22**(2) 140–146.
- Cotet, A.M. 2012. The Impact of Noneconomic Damage Cap on Health Care Delivery in Hospitals. *American Law and Economics Review* 1–43.
- Dove, J.T., J.E. Brush Jr, R.A. Chazdal, and W.J. Oetgen. 2010. Medical Professional Liability and Health Care System Reform. *J. of the American College of Cardiology* **55**(25) 2801–2803.

- General Accounting Office. 2003. Medical Malpractice Insurance: Multiple Factors Have Contributed to Increased Premium Rates. GAO-03-702 Report to Congressional Requesters, General Accounting Office.
- Helland, E. and M.H. Showalter. 2009. The Impact of Liability on the Physician Labor Market. *J. of Law & Economics* **52**(November 2009) 635–663.
- Hellinger, F.J. and W.E. Encinosa. 2006. The Impact of State Laws Limiting Malpractice Damage Awards on Health Care Expenditures. *American J. of Public Health* **96**(8) 6–12.
- Kessler, D.P., W.M. Sage, and D.J. Becker. 2005. Impact of Malpractice Reforms on the Supply of Physician Services. *J. of the American Medical Association* **293**(21) 2618–2625.
- Lakdawalla, D.A., and S.A. Seabury. 2009. The Welfare Effects of Medical Malpractice Liability. NBER Working Paper.
- Lumalcuri, J. and R.W. Hale. 2010. Medical Liability: An Ongoing Nemesis. *Obstetrics & Gynecology* **115**(2.1) 223–228.
- Paik, M., B.S. Black, D.A. Hyman, and C.M. Silver. 2011. Will Tort Reform Bend the Cost Curve? Evidence from Texas. Working Paper.
- Reyes, J.W., 2010. The Effect of Malpractice Liability on the Specialty of Obstetrics and Gynecology. NBER Working Paper.
- Sethi, M.K., W.T. Ombremskey, H. Natividad, H.R. Mir, and A.A. Jahangir. 2012. Incidence and Costs of Defensive Medicine Among Orthopedic Surgeons in the United States: A National Survey Study. *American Journal of Orthopedics* **41**(2) 69–73.
- Sloan, F.A. and J.H. Shadle. 2009. Is there empirical evidence for Defensive Medicine? A reassessment. *J. of Health Economics* **28** 481–491.
- Studdert, D.M., M.M. Mello, W.M. Sage, C.M. DesRoches, J. Peugh, K. Zapert, and T.A. Brennan. 2005. Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment. *J. of the American Medical Association* **293**(21) 2609–2617.

## **Chapter 4**

# **Mitigating Risk Selection with Patient-Level Competitive Bidding**

### **4.1 Introduction**

The field of health economics has struggled with methods of provider compensation for decades. The reason why this struggle is more pronounced in health economics than in other sub-disciplines is the nature of information in the relationships between payers, providers, and patients. The asymmetry of information between payer and provider, in particular, creates contracting problems,



and thus tradeoffs in every known method of provider compensation. Any investigation into the pros and cons of provider compensation methods must begin with the “selection-efficiency” tradeoff coined by Joseph Newhouse (1996). At the heart of the tradeoff is heterogeneity and uncertainty in the cost of treatment across patients and illnesses, and how different compensation methods distribute risk between payer and provider. Payers; which include government, private health insurers, and large employers; tend to be the largest among the players in markets for health care, and so are the most obvious candidate to bear risk. Problems arise, however, because the cost of treatment is not completely exogenous, and is instead partly determined by the choices of providers and patients. Under compensation arrangements like cost-based reimbursement and fee-for-service, both providers and patients can benefit from greater quantity and quality while leaving the marginal cost of their decisions to be borne by the payer. The result is incentives for over-utilization of health care services in the treatment of all patients, and thus excessive health care expenditure (Bovbjerg et al 1987, Hoerger & Waters 1993, Newhouse 1996, Ellis 1998).

In order to mitigate these incentives, payers can reach cost sharing arrangements with the parties making treatment decisions. Examples of cost sharing on the demand side include high deductibles, copayments, and coinsurance. These mechanisms are verifiable among payers, providers, patients and monitors; making regulation easy (Frank et al 2000); but leave patients exposed to

risk and “compromise the purpose of insurance” (Eggleston 2000). Cost sharing can take place on the supply side as well. Alternatives to cost-based or fee-for-service arrangements include capitation, where providers receive a lump-sum per patient enrolled or treated, and prospective payment, which pays based on predetermined rates independent of patient-specific costs. The benefit of supply-side cost sharing is the placement of both revenue certainty and cost risk with providers, who are larger than patients and better able to bear uncertainty (Ellis & McGuire 1986, van de Ven et al 2003). With providers bearing the full marginal cost of treatment decisions, there is the incentive to reduce the utilization of services, and thus expenditures (Newhouse 1996, Eggleston 2000). On the other hand, this incentive to reduce utilization can lead to under-provision of services and the avoidance of patients likely to require expensive treatments. While the prospect of medical malpractice lawsuits and published measures of quality can reduce the incentive to under-treat patients, these considerations can exacerbate the avoidance problem, since these encourage providers to accept the patients most likely to recover and be pleased with service (Frank et al 2000).

Risk selection is the efforts by health plans or providers to enroll patients believed to be low-risk or low-cost (cream-skimming), and to disenroll (or dump) high-risk or high-cost patients (Pauly 1984, Eggleston 2000). Such efforts would only be temporary in a perfectly competitive market for health services since separate prices would emerge for different risk types and equalize the

relative profitability of patients across risk type (Pauly 1984, Diamond 1992). The problem occurs when separate prices cannot emerge because payers either cannot distinguish between risk types, or are constrained by regulation from pricing along certain dimensions associated with risk type. This includes regulated community rating on the basis of gender and age (Pauly et al 1991), characteristics that explain only a small part of the variance in health care expenditure and thus contain within-community heterogeneity in risk (Beck 2000). Given pricing based on community averages, managed care organizations have incentives to enroll only those patients with expected treatment costs that fall below these prices, and avoid all others (Beck 2000). Where successful, risk selection results in broken pooling arrangements, where payers pay too much for patients accepted by providers, and must make alternative (often inferior) arrangements for those rejected (van de Ven et al 2003). Risk selection is often performed as part the provider's day-to-day operations, making it difficult for payers and policymakers to regulate selection practices (Chalkley & Malcolmson 2000, Frank et al 2000). The profitability of risk selection for one provider can result in risk selection as a dominant strategy across all providers, since greater concentrations of rejected high-risk patients seeking care would reduce the profitability of non-selecting providers (Frank et al 2000, Eggleston 2000, van de Ven et al 2003).

Adverse selection is distinct from risk selection, although the practice of the

two in health care is often indistinguishable. Unlike risk selection, adverse selection is behaviour on the demand side, where patients and policyholders have information superior to payers and providers regarding underlying health risks or preferences for consuming health services (Pauly 1984, Eggleston 2000). The distinction becomes blurred in the various methods of indirect risk selection, where providers design plans to be attractive to certain risk groups, and thus induce adverse selection. While mandatory coverage for basic services would eliminate adverse selection, at least in basic services (Pauly et al 1991), risk selection can persist where providers can attract the risks they deem favourable.

The rise in capitation and prospective payment in the United States has coincided with the rise in various forms of managed care since the 1960s. Managed Care Organizations (MCOs) are contractual arrangements between insurers and providers that coordinate the financing and delivery of health care. By 1983, Medicare developed a prospective payment system based on diagnosis-related groups (DRGs) to compensate hospitals, and for physicians' fees in 1984 (Hoerger & Waters 1993). By the end of the 20th century, three quarters of private health insurance was under managed care and over 14% of Medicare and 40% of Medicaid patients were enrolled in health plans using capitation (Frank et al 2000). Medicare Advantage, which pays private managed care organizations by capitation, currently serves roughly one quarter of Medicare beneficiaries (Brown et al 2011). The large number and percentage of government-sponsored enrollees under capitated arrangements indicates the potential for

risk selection among vulnerable populations.

The most commonly suggested solution to risk selection is risk adjustment. It requires that payers classify patients according to verifiable characteristics, and then set different payments for each classification that increase with the expected cost of treatment. Typical dimensions of risk adjustment include age, gender, and geographic area of residence. The optimal set of payments would be high enough for high cost patients to make providers indifferent among all patient types, eliminating the incentive to select low risks. There are criticisms of risk adjustment as a solution to risk selection. The first concerns the practicality of a classification system so fine that it captures all within-group cost variation observable to providers (Chalkley & Malcomson 2000). Improvements in risk adjustment may require costly investments in research, and could start an “information race” between payers and providers interested in gaining an informational advantage (Barros 2003). Furthermore, it may be inappropriate to adjust on the basis of some relevant characteristics, particularly those under the control of providers, in order to preserve incentives for cost-effective innovation (Schokkaert & Van de Voorde 2004). Therefore, in order to completely eliminate risk selection, risk adjustment must be perfect both over time and across all cost-relevant patient characteristics, which is practically impossible.

A second potential solution to risk selection is open enrollment, which allows consumers to switch between health plans, with guaranteed acceptance

and renewal. While this eliminates the most direct form of risk selection (outright refusal), there are several methods of indirect risk selection. These include underservice to motivate selective disenrollment (Newhouse 1996, Brown et al 2011), reducing capacity in high cost service areas (Frank et al 2000), becoming ill-equipped to accommodate the most severe cases (Chalkley & Malcolmson 2000), selective marketing, designing supplemental coverage to appeal to healthy types (Eggleston 2000, van de Ven et al 2003, Brown et al 2011), greater online presence to attract tech-savvy types, helping unprofitable patients switch, using screening software, and ignoring phone calls from unprofitable consumers (van de Ven et al 2003). Such internal business decisions are difficult to regulate, and so health plans can and do engage in risk selection despite open enrollment rules (Glazer & McGuire 2000, Barros 2003, van de Ven et al 2003, Beck et al 2010, van de Ven et al 2007).

In an abstract sense, risk selection occurs when a payer, seeking a service on behalf of a set of patients, offers a uniform price to a set of potential providers, each of which has private information regarding the costliness or riskiness of agreeing to treat each patient. In cases of asymmetric information such as these, a procurement auction is superior to uniform pricing, since it both motivates the revelation of private information and results in an equilibrium within the set of core (and thus efficient) allocations (Milgrom 1985). In mapping risk selection into an auction environment, the payer is a principal,

the providers are bidders, and a given patient is a contract. The association assumes that providers wish to reject patients whenever they determine that the expected cost of accepting said patient is greater than the capitation fee offered by the payer. These expected treatment costs are private information that is discarded upon rejection under a uniform pricing scheme like capitation, while it is aggregated when providers can bid competitively to serve each patient. Furthermore, competitive bids over patients can adjust rapidly to provider- or industry-specific changes production costs, and reward cost-reducing innovations.

Competitive bidding arrangements have been proposed both in the past (Hogan 1983, Christianson & Smith 1984, Pauly et al 1991, Keijser & Kirkman-Liff 1992) and in more recent health policy discussions (Berwick & Hackbarth 2012, Feldman et al 2012). The main drawbacks of competitive bidding in the context of health care concern limitations on consumer choice and quality assurance (Bovbjerg et al 1987, Hoerger & Waters 1993, Newhouse 1996), which are also concerns with capitation and prospective payment. Thus the opportunity cost of incorporating competitive bidding into capitation systems is minimal. Existing attempts and proposals all involve the use of competitive bidding to price services rather than combat risk selection by pricing individual patients. This paper investigates the second type of bidding system by evaluating two compensation regimes. The first is a traditional capitation system, where providers are offered a fixed fee per consumer they take on.

Given this capitation, consumers and providers interact, and providers accept favourable risks and reject the rest. Consumers deemed profitable by at least one provider are assumed to enjoy superior health to those who are universally rejected and do not receive dedicated care. The second regime starts similarly to the first. Providers are offered a capitation fee, consumers and providers interact, and providers take on those they find acceptable. The difference in the second regime is, upon rejecting a consumer, a provider must submit a minimum acceptable payment at which the consumer *would* be acceptable. The final step, for those consumers rejected by all providers, the payer awards the service contract to the provider submitting the lowest amount, and pays this provider the second-lowest submitted amount.

The model shows that, regardless of the degree of information asymmetry between payer and providers and the variation in treatment costs, risk selection always occurs in equilibrium in a pure capitation system and never occurs once competitive bidding is incorporated. It also shows that a system with a mix of allocations, by both capitation and competitive bids, is superior to a system where all consumers are allocated by competitive bids. Finally, the paper shows sufficient conditions under which the mixed system dominates the pure capitation system. When this dominance holds, the mixed allocation system carries all the benefits of capitation and prospective payments without the drawback of risk selection, and is not subject to a selection-efficiency tradeoff.



## 4.2 Literature Review

The extent to which risk selection can take place depends on several factors. First, there must be substantial variation in health care expenditures across patients within payment groups. Of all DRGs in Medicare's initial classification system, one sixth exhibited cost distributions with standard deviations exceeding means (Dranove 1987). Substantial variation alone, however, is not enough for risk selection to occur. In order to engage in risk selection, providers must be able to use observable characteristics to predict within-group variation in patient costs. This would be indicated where providers are able to predict a greater share of the variation in health care spending than that predicted by the risk adjustment models used by payers. Medicare's initial risk adjustment on age, gender, disability, and Medicaid status could explain approximately 1% of the variation (Brown et al 2011) while hospitals in the same time frame could predict between 10% and 20% of the variation in hospital costs based on characteristics observable upon admission (Dranove 1987). Payers have made advances in risk adjustment models, but the predictive power of payment categories has remained relatively low. Medicare's current risk adjustment scheme, the Hierarchical Conditions Categories model, constructs 70 disease categories out of 15,000 disease codes, and can explain 11% of the variation in health care expenditure, although it systematically underpredicts above-average expenditures (Brown et al 2011). Adjusting based on prior health care utilization, such as hospitalization in the previous year, was able to explain less than 10% of the

variance in the US (Glazer & McGuire 2000) and close to 14% in Switzerland (Beck 2000), although including health care utilization introduces incentives to over-utilize in health plans based on gatekeeper models.

Much of the analysis of whether favourable selection takes place for managed care organizations concerns comparisons between the time  $t$  health care expenditures of “stayers”, who are patients electing to remain in a given health plan, and those of “switchers”, who are those patients choosing to switch plans in time  $t + 1$ . Several studies have found that patients switching from fee-for-service plans to capitated plans exhibited 11% to 37% lower health care expenditure than their staying counterparts, while switches in the opposite direction incurred 18% to 60% higher expenditure (Beck 2000, Frank et al 2000, Glazer & McGuire 2000, Nicholson et al 2004, Brown et al 2011). While this indicates that low-risk patients select into capitated arrangements and high-risk select into fee-for-service, it does not show that this selection is specifically risk selection instead of adverse selection. The most compelling evidence that risk selection occurs is the recent working paper by Brown et al (2011) showing changes in the margins of selection before and after comprehensive changes to Medicare’s risk adjustment formula in 2003. The evidence shows that selection decreased along the dimensions included in the risk adjustment formula, but increased along those dimensions excluded. Since the reforms would not have affected patients’ choice of plan, this indicates that risk adjustment made selection along certain dimensions unprofitable, inducing providers to shift selection

efforts into alternate patient characteristics. As the study showed, the persistence of risk selection resulted in the comprehensive risk adjustment scheme having no statistically significant impact on selection overall.

Competitive bidding has been utilized on a limited basis in health services markets since the 1970s. Almost exclusively, the goal of these bidding systems was pricing services. These included laboratory services in New York City and mental health services in Massachusetts (Schlesinger 1986, McCombs & Christianson 1987), as well as per-diem hospital care for Medicaid patients in California and Wisconsin (Bovbjerg et al 1987, Paringer & McCall 1990). The Arizona Health Care Cost Containment System (AHCCCS) began in 1982 with a competitive bidding component. Providers were required to submit per-capita bids in four categories of service for each of five patient categories, as well as the maximum number patients they could accommodate (Christianson & Smith 1984). At the federal level in the United States, The Centers for Medicare and Medicaid Services launched demonstration projects with competitive bidding for durable medical equipment, prosthetics, orthotics, and supplies in 1997 (Katzman & McGeary 2008) and laboratory services in 2003 (Waters 2006). The closest system utilizing bidding at the patient level is CareAuction.nl in the Netherlands for maternity care. Consumers report the amount of maternity care hours needed to insurers, who put these needs on a web site facilitating bids by providers. The system's allocation rule is based primarily on

consumer preference rather than lowest price, and given the homogenous nature of the product, the system is designed to induce high quality care rather than allocate heterogeneous patients at the lowest cost (Smits & Jansen 2008).

The effectiveness of these competitive bidding systems has been mixed, but generally positive. Organized industry opposition in New York City and Massachusetts and the resulting lack of competition caused those two systems to be unsuccessful (McCombs & Christianson 1987). Despite cost savings not being the primary goal, CareAuction.nl has achieved 2% to 4% in spending reductions (Smits & Jansen 2008). In Arizona, AHCCCS costs rose 34.2% over eight years, compared to traditional Medicaid cost increases of 60.7% (Paringer & McCall 1990), resulting in savings of 11% of medical costs and 7% of total costs relative to traditional Medicaid (Iglehart 1995). In the first year of California's system, per-diem payments fell 15% to 16% where they had previously been rising at 7% per year (Bovbjerg et al 1987). Bidding reduced Medicare expenditure on durable medical equipment by 19% (Mechanic & Altman 2010), but met significant opposition from the clinical laboratory industry that resulted in motions to cease the demonstration project (Garrott 2007). Besides industry opposition, these competitive bidding systems encountered several problems. First, allocation rules based on weighted averages of submitted bid prices can produce inefficient allocations and result in winners selectively avoiding unprofitable

procedures (Katzman & McGeary 2008). A lack of transparency between bidding authority and bidders, and not allowing bidders enough time to gather information and secure reinsurance, created uncertainty and inflated bids (Hillman & Christianson 1984, Waters 2006). Further concerns include political controversy, uncertainty over system costs (Christianson & Smith 1984), and concerns about both not enough competition (Hoerger & Waters 1993) and “too much” competition where small and rural providers are priced out (Waters 2006, Garrott 2007).

### 4.3 The Model

There is a population of consumers of measure 1. Consumers are identical in endowed income  $m$  and preferences represented by the quasi-concave utility function  $U(y, H)$ , where  $y$  is consumption and  $H$  is health status. Health status is binary and determined by whether or not the consumer has a dedicated care provider in the event of illness, which for simplicity is assumed to occur with probability 1. Those consumers with a dedicated provider enjoy health status  $H_1$  and those without enjoy  $H_2$ , where  $H_1 > H_2$ . Like Kifmann & Lorenz (2011), consumer interactions with providers post-match are not modelled here, though incentives would be the same in any of the payment mechanisms investigated. Consumers are heterogenous in ex ante risk type  $k$ , which is unobservable to consumers. There are  $K \in \mathbb{Z}^{++}$  different types and the share

of consumers of risk type  $k$  is  $\alpha_k$ , where  $\sum_{k=1}^K \alpha_k = 1$ .

Consumers match randomly and sequentially with each of  $n \in \mathbb{Z}^{++}$  identical health care providers, where  $n \geq 2$ . Each provider is unaware of its place in the order of these matches. The simplest interpretation of  $n$  is the total number of providers available to meet with a consumer. Alternatively, even though the steps are not modelled here,  $n$  could be the number of providers consumers visit before either falling ill or giving up the search. Providers are risk-neutral income maximizers. The cost to provider  $j$  of taking on consumer  $i$  of risk type  $k$  is  $c_{ijk}$ . These costs are drawn independently from the distribution  $F_k(c)$ , with corresponding probability density function  $f_k(c)$  over the support  $[\underline{c}, \bar{c}]$ . The assumption of a common support across distributions ensures that other players cannot directly infer consumer type based on any reports of  $c_{ijk}$ . Assume that these distributions are smooth, continuous, atomless, and that distribution  $F_{k-1}(c)$  is first-order stochastically dominated by distribution  $F_k(c)$  for all  $k \geq 2$ . This implies and that the average cost of treating a consumer is increasing in ex ante risk type. Assume that consumer risk type is verifiable among providers, while  $c_{ijk}$  is private information held by provider  $j$ . Any provider is able to engage in risk selection, and thus can reject excessively costly consumers. Consumer utility is unaffected by the number of rejections as long as it is less than  $n$ . This means that a consumer accepted by the first provider visited would be just as well off as one rejected by the first  $n - 1$  providers and accepted by the  $n^{\text{th}}$  provider.

There is a payer operating in a perfectly competitive market for health insurance. The payer takes in revenue of  $\tau$  per insurance policy sold to consumers, and compensates providers for taking on consumers. The payer cannot condition compensation on risk type. This is either due to superiority of information among providers, or a conscious effort by the payer (or a regulator of payers) to disregard a set of variables affecting risk type when pursuing risk adjustment. Therefore, the payer chooses the capitation rate  $w$  at which providers are compensated per consumer taken on, regardless of risk type. Competition motivates the payer to choose the capitation rate that maximizes expected consumer utility. For those policyholders that fail to be taken on by a dedicated care provider, the payer bears a cost  $w_{ER}$ . This is the cost of having the policyholder's ailment treated outside of a primary care setting, such as an emergency room or acute care setting for a preventable illness.

#### 4.4 Equilibrium: Capitation

The first compensation method evaluated is a simple capitation system, where payers charge  $\tau_c$  per health insurance policy and offer providers a fixed fee per consumer taken on. Given this capitation, and since they are able to select risks, providers take on those consumers they find profitable and reject those who are not. Provider  $j$ 's problem when confronted with a consumer with cost

of treatment  $c_{ijk}$  is whether to accept or reject that consumer. For each consumer accepted, the provider would receive  $w$  in revenue and bear cost  $c_{ijk}$ . In rejecting a consumer, the provider takes in no revenue and bears no cost. Therefore, in this subgame, each provider  $j$  has a simple optimal strategy:

accept if  $w - c_{ijk} \geq 0$   
 reject otherwise.

Given the assumptions on consumer costs, the probability that a given consumer of risk type  $k$  will be unacceptable to a provider is  $1 - F_k(w)$ . Since these costs are independent across providers, the probability that all  $n$  providers will find this consumer unacceptable is  $[1 - F_k(w)]^n$ . The payer's problem thus becomes:

$$\max_w \left\{ EU_c(w) = U(m - \tau_c, H_1) - \sum_{k=1}^K \alpha_k [1 - F_k(w)]^n \cdot \Delta U \right\}$$

$$\text{where } \tau_c = w + \sum_{k=1}^K \alpha_k [1 - F_k(w)]^n \cdot (w_{ER} - w)$$

$$\Delta U = U(m - \tau_c, H_1) - U(m - \tau_c, H_2)$$

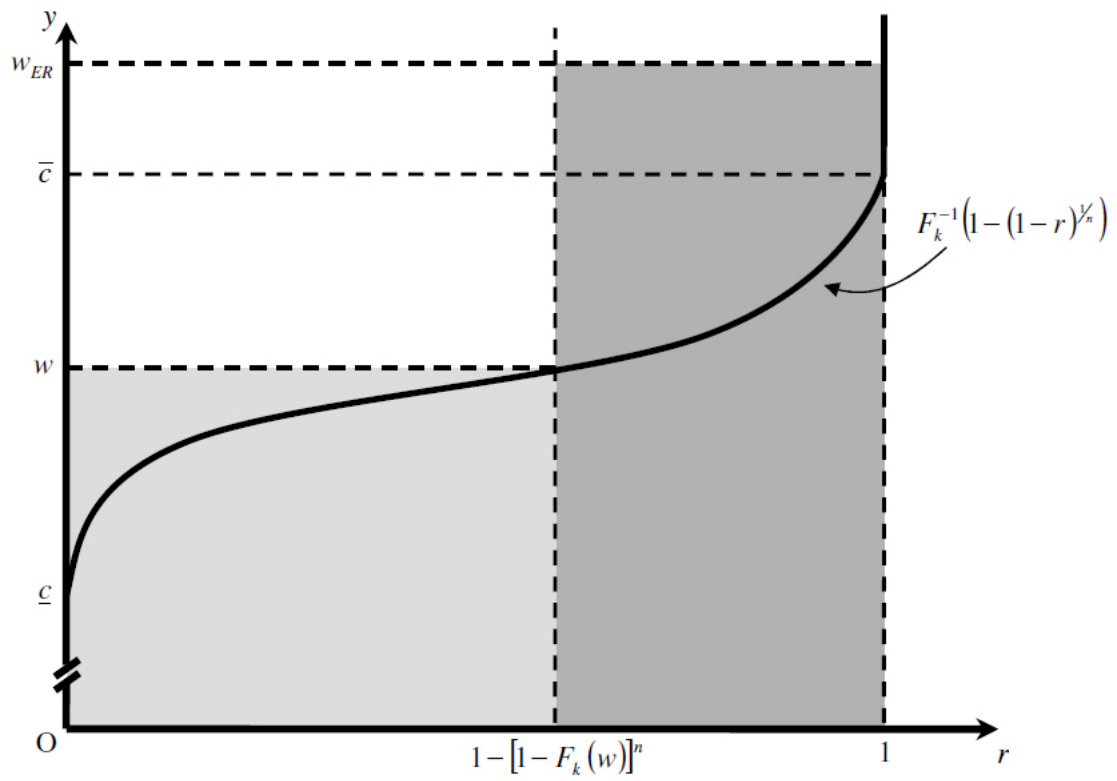


While a higher capitation payment increases the chance that any consumer secures a provider, it also increases the pure profit earned by providers treating consumers with low  $c_{ijk}$  values. The optimal  $w$  recognizes this trade-off. The optimal capitation payment under the traditional capitation system ( $w_c^*$ ) satisfies:

$$n \left( \frac{\Delta U}{WMU_y} + w_{ER} - w \right) \sum_{k=1}^K \alpha_k [1 - F_k(w)]^{n-1} f_k(w) = 1 - \sum_{k=1}^K \alpha_k [1 - F_k(w)]^n \quad (4.1)$$

where  $WMU_y$  stands for “weighted marginal utility of consumption” and is equal to  $\frac{\partial U(m-\tau_c, H_1)}{\partial y} - \sum_{k=1}^K \alpha_k [1 - F_k(w)]^n \frac{\partial \Delta U}{\partial y}$ . The term on the left is the benefit of a marginal increase in the capitation rate. More specifically, it is the increase in the measure of consumers both forgoing emergency room care (and thus costing  $w$  instead of  $w_{ER}$ ) and also gaining the boost in utility from having a primary care provider ( $\Delta U$ ), the value of which is measured in units of consumption once divided by  $WMU_y$ . The increase in costs due to the marginally higher capitation rate on the measure of consumers already receiving care in expectation is on the right.

**Proposition 1:** *For any equilibrium capitation payment under the capitation system ( $w_c^*$ ), i)  $[1 - F_k(w_c^*)]^n > 0 \forall k$ , and ii)  $[1 - F_{k-1}(w_c^*)]^n < [1 - F_k(w_c^*)]^n \forall k \geq 2$ .*



**Figure 4.1:** Total spending under the capitation method for a given capitation payment. The light rectangle is expenditure on consumers receiving primary care while the dark rectangle is expenditure on consumers rejected by all  $n$  physicians

*Proof:* As  $w_c$  approaches  $\underline{c}$  from above, the left side of Condition 4.1 remains positive while the right side approaches zero. As  $w_c$  approaches  $\bar{c}$  from below, the left side is driven to zero while the right side approaches 1. Since both sides of Condition 4.1 are continuous in  $w$ , by the Intermediate Value Theorem, there exists  $w \in (\underline{c}, \bar{c})$  such that Condition 4.1 holds. Since  $w_c^* < \bar{c}$  and the cost of treating patients of any risk type is distributed over the support  $[\underline{c}, \bar{c}]$ , then  $[1 - F_k(w_c^*)]^n > 0 \forall k$ . Stochastic dominance assumptions imply that  $F_1(w) > \dots > F_K(w)$ , which in turn implies that  $[1 - F_1(w)]^n < \dots < [1 - F_K(w)]^n$ . ■

Intuitively, as  $w_c$  approaches  $\bar{c}$ , the slope of  $F_k^{-1} \left[ 1 - (1 - r)^{\frac{1}{n}} \right]$  approaches infinity. This means that value of the infinitesimally small average gain in access to primary care for the most costly consumers across risk type is outweighed by the value of consumption forgone in order to cover the increase in payments for those already receiving care. Therefore, in equilibrium under the pure capitation system, a positive measure of consumers will not find a dedicated care provider, and the shares of higher risk types in this universally rejected group are greater than in the consumer population.

## 4.5 Equilibrium: Capitation-Plus

The second compensation mechanism begins similarly to the first. Providers are offered a capitation fee, consumers and providers interact to draw  $c_{ijk}$  values, and providers take on those they find acceptable. The difference is that under the second mechanism, upon rejecting a consumer, a provider must report to the payer an amount ( $\beta_j$ ) at which the consumer *would* be acceptable. These amounts are the providers' "bids". For those consumers rejected by all  $n$  providers, the payer awards the consumer to the provider submitting the lowest bid, and pays this provider the second-lowest bid. Let this mechanism be the "capitation-plus system", in which the payer charges consumers  $\tau_{cp}$ .

Working backward, the payer need allocate consumer  $i$  of risk type  $k$  on the basis of bids if and only if the consumer has been rejected by all  $n$  providers. Under the rules described above, this "all-reject" state is equivalent to a second-price reverse or procurement auction, with the payer as the principal, all  $n$  rejecting providers as bidders, and the universally rejected consumer as the contract. It is well established in the literature that bidder  $j$ 's optimal bid in such an auction ( $\beta_j^*$ ) is the opportunity cost of the object for auction ( $c_{ijk}$ ). From a rejecting provider's perspective, any bid delivers a payoff of 0 in any state where at least one provider found that consumer acceptable, and so the optimal bid in the all-reject state is also optimal in the  $2^n - 1$  other states.

Before submitting a bid, the provider must first decide whether to accept

or reject. It is assumed, following Kirkman-Liff et al (1985), that a provider can construct a subjective estimate of the probability of losing the contract for a given set of bid prices. Let  $h_k$  be the lowest treatment cost at which it is a best response for any provider to reject a consumer of risk type  $k$ . Given the capitation payment offered by the payer, the value of accepting is the same as in the pure capitation system ( $w - c_{ijk}$ ). For any  $w < \bar{c}$ , the value to the provider of rejecting is strictly positive since there is a positive probability that the  $n - 1$  other providers also reject and the patient is allocated based on bids. In such a case, each rejecting provider has a non-negative expected payoff. The winning bidder submitting  $c_{ijk}$  in a second-price auction would receive the lowest of the set of bids higher than  $c_{ijk}$ . This payment ( $\hat{c}$ ) is a random variable distributed according to the cumulative distribution function  $1 - G_k(\hat{c}) = 1 - [1 - F_k(\hat{c})]^{n-1}$ . The expected payment conditional on  $c_{ijk}$  being the lowest bid is:

$$E(\hat{c} | \hat{c} > c_{ijk}) = c_{ijk} + \int_{c_{ijk}}^{\bar{c}} \frac{G_k(x)}{G_k(c_{ijk})} dx. \quad (4.2)$$

This would be the bid of a provider with cost  $c_{ijk}$  in a first-price reverse auction (Milgrom & Weber 1982, Huh & Roundy 2002), and also the expected payment to a winning provider bidding  $c_{ijk}$  in a second-price reverse auction due to the revenue equivalence theorem (Myerson 1981). Therefore, the expected value of rejecting a consumer is the product of the probability that the  $n - 1$  other providers also reject ( $G_k(h_k)$ ), the probability that  $c_{ijk}$  represents the

lowest bid  $\left(\frac{G_k(c_{ijk})}{G_k(h_k)}\right)$ , and the expected payoff conditional on  $c_{ijk}$  being the lowest bid ( $E(\hat{c} | \hat{c} > c_{ijk}) - c_{ijk}$ ). Thus, unlike the capitation system, a provider's accept or reject decision is:

$$\text{accept if } w \geq c_{ijk} + G(h_k) \cdot \frac{G_k(c_{ijk})}{G(h_k)} \cdot \int_{c_{ijk}}^{\bar{c}} \frac{G_k(x)}{G_k(c_{ijk})} dx \quad (4.3)$$

reject otherwise.

Condition 4.3 defines  $h_k$ :

$$h_k = h_k(w, n) \text{ such that } h_k + \int_{h_k}^{\bar{c}} G_k(x) dx = w \quad (4.4)$$

The right side of (4.3) is increasing in  $c_{ijk}$ , so the set of consumer costs that would be rejected by a provider is  $[h_k, \bar{c}]$ . Therefore, when facing a consumer of type  $k$ , a capitation payment  $w$ , and  $n - 1$  other providers, provider  $j$  would accept consumer  $i$  if  $c_{ijk} < h_k$ . Since  $\int_{h_k}^{\bar{c}} G_k(x) dx \geq 0$ , it is clear that;  $\forall n, w \in [\underline{c}, \bar{c}]; h_k(w, n) < w$ . Intuitively, given some  $c_{ijk}$ , the prospect of realizing a positive payoff from the auction stage makes rejecting more attractive in the capitation-plus system than in the capitation system. This means that any choice of capitation payment would induce more rejections in the capitation-plus system than in the capitation system. Let  $R_k(w, n) = [1 - F_k(h_k)]^n - [1 - F_k(w)]^n$  be the increased likelihood that a patient of type  $k$  is

rejected ex ante in the capitation-plus system relative to the capitation system, given payment  $w$ .

Rather than the actual costs taken on by the providers, the expected payment to winning providers (ie. the expected second-lowest bid) is relevant to the payer's problem. When allocating a consumer of type  $k$  according to bids, the probability that the second-lowest bid is less than  $\pi$  is the probability that any two providers (of which there are  $n(n-1)$  combinations) both bid below  $\pi$  while the remaining  $n-2$  other providers bid higher. This means that the second lowest payment for a type  $k$  consumer is distributed according to the probability density function  $j_k(\pi | h_k) = n(n-1)f_k(\pi)[F_k(\pi) - F_k(h_k)][1 - F_k(\pi)]^{n-2}$ . Therefore, the expected payment to providers per consumer of type  $k$  allocated by bids is:

$$E[\pi_k | w, n] = \int_{h_k}^{\bar{c}} \pi j_k(\pi | h_k) d\pi$$

Under the capitation-plus system, each consumer is either accepted by a provider during an initial match, or allocated to the lowest bidding provider in the auction stage. This means that all consumers ultimately receive primary care and enjoy health status  $H_1$ . Consumers enjoy expected utility  $EU_{cp}(w_{cp}) = U(m - \tau_{cp}, H_1)$ , and so the payer's problem becomes the minimization of health care spending  $\tau_{cp}$ :

$$\min_w \left\{ \tau_{cp} = w + \sum_{k=1}^K \alpha_k \left\{ [1 - F_k(h_k)]^n \cdot (E[\pi_k | w, n] - w) \right\} \right\}$$

with the optimal capitation payment ( $w_{cp}^*$ ) characterized by:

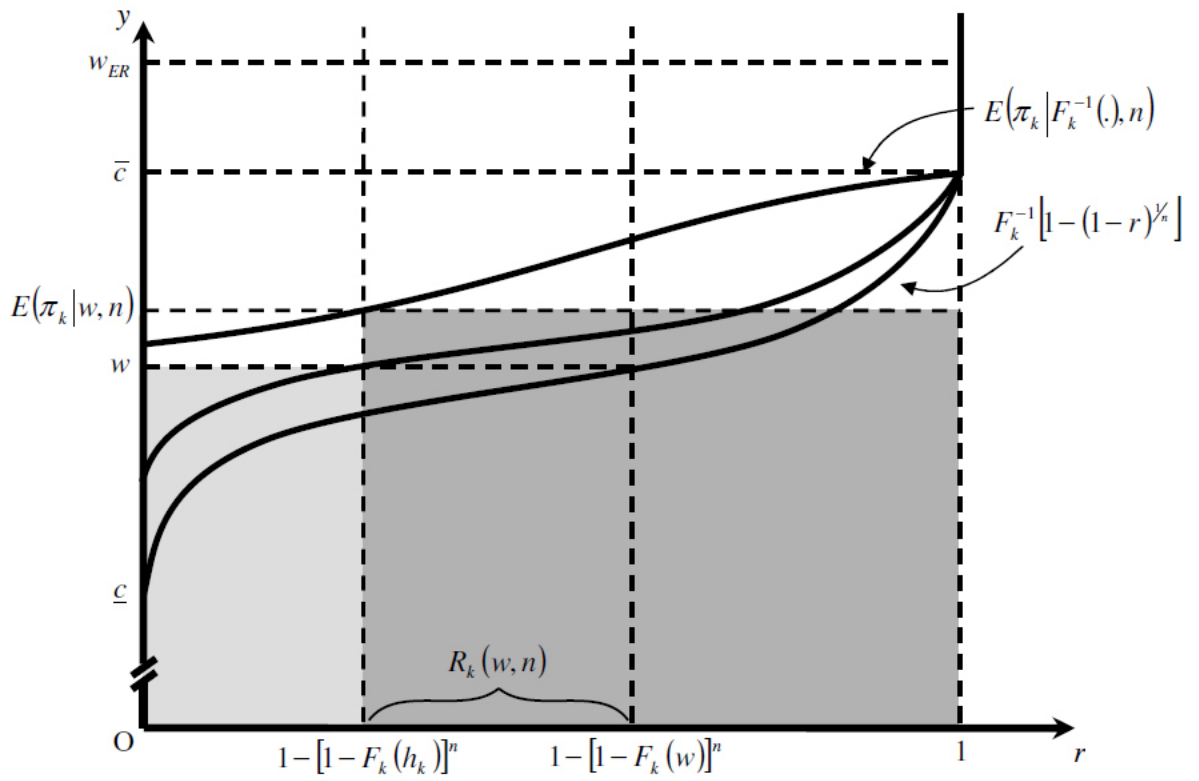
$$\begin{aligned} & n \sum_{k=1}^K \alpha_k \left\{ [1 - F_k(h_k)]^{n-1} f_k(h_k) \frac{\partial h_k}{\partial w} (E(\pi_k | w, n) - w) \right\} \\ & = 1 - \sum_{k=1}^K \alpha_k [1 - F_k(h_k)]^n + \sum_{k=1}^K \alpha_k \left\{ [1 - F_k(h_k)]^n \frac{\partial E(\cdot)}{\partial w} \right\} \end{aligned} \quad (4.5)$$

The left side of the equation represents the marginal benefit of increasing the capitation payment in the capitation-plus system. The top term is the reduction in the measure of consumers expected to be allocated in the auction stage, each of whom brings the additional cost ( $E(\pi_k | w, n) - w$ ) above the capitation fee. The right side of the equation is the marginal cost, which includes the increased expense of both higher capitation fees on the measure of consumers avoiding the auction stage as well as higher expected payments in the auction stage for those who do not.

**Proposition 2:**  $\sum_{k=1}^K \alpha_k [1 - F_k(h_k(w_{cp}^*, n))]^n < 1$  in any equilibrium under the capitation-plus system.

*Proof:* Define  $\underline{w}_k = \underline{w}_k(n)$  such that  $h_k(\underline{w}_k, n) = \underline{c}$ . This is the highest capitation





**Figure 4.2:** Total spending under the capitation-plus method for a given capitation payment. The light rectangle is expenditure on consumers receiving primary care while the dark rectangle is expenditure on consumers allocated in the auction stage

payment such that providers still reject all consumers of type  $k$ , regardless of  $c_{ijk}$ . The value of  $\underline{w}_k$  increases in  $k$ . Obtaining the first derivatives of  $h_k(w)$  and  $E(\pi_k | w, n)$ :

$$\frac{\partial h_k}{\partial w} = \begin{cases} \frac{1}{1-G_k(h_k)} & \text{if } h_k(w, n) \geq \underline{c}; \\ 0 & \text{otherwise.} \end{cases}$$

$$\frac{\partial E(\pi_k | w, n)}{\partial w} = \begin{cases} \frac{nf_k(h_k)}{1-F_k(h_k)} \frac{\partial h_k}{\partial w} \left\{ E(\pi_k | w, n) - E(\hat{c} | \hat{c} > h_k) \right\} & \text{if } h_k(w, n) \geq \underline{c}; \\ 0 & \text{otherwise.} \end{cases}$$

and using (4.2) and (4.4) allows Condition 4.5 to be rearranged into:

$$\sum_{k=1}^K \alpha_k n f_k(h_k) \int_{h_k}^{\bar{c}} G_k(\pi) d\pi = 1 - \sum_{k=1}^K \alpha_k [1 - F_k(h_k)]^n.$$

As  $w_{cp}$  approaches  $\underline{w}_1$  from above, the right side of the equation approaches 0 while the left side approaches a value greater than 0. Therefore, costs could be lowered by setting  $w_{cp}$  strictly above  $\underline{w}_1$ . Since  $w_{cp}^* > w_1 \Rightarrow [1 - F_1(h_1(w_{cp}^*, n))]^n < 1$ , so  $\sum_{k=1}^K \alpha_k [1 - F_k(h_k(w_{cp}^*, n))]^n < 1$ . ■

Proposition 2 states that it is never optimal in the capitation-plus system to set the capitation payment so low that every consumer is allocated by auction. Costs can be reduced by having at least an infinitesimally small positive measure of type 1 consumers accepted before the auction stage because the newly

accepted measure would cost  $\underline{w}_1$  instead of  $E(\pi_1 | \underline{w}_1, n)$  where  $\underline{w}_1 < E(\pi_1 | \underline{w}_1, n)$ . Intuitively, a provider drawing  $c_{ijk} = h_k$  is guaranteed to have the lowest bid and win any auction for this patient, and can expect to realize a large payoff. When deciding whether or not to accept, however, such a provider realizes that this payoff is not guaranteed. It is contingent on every other provider also rejecting, which is the only case in which an auction occurs. By accepting, however, the provider assures itself of taking in revenues equal to the capitation payment. Providers thus face a tradeoff between a high potential payoff from rejecting versus a lower certain payoff from accepting, which the payer can exploit by raising the capitation payment. The result is that at least a small measure of patients are treated at a price equal to the capitation payment, which is less than that measure would have cost if allocated by auction.

**Proposition 3:** *If  $\sum_{k=1}^K \alpha_k R_k(w_c^*, n) \left\{ E[\pi_k | w_c^*, n] - w_c^* \right\} \leq \sum_{k=1}^K \alpha_k [1 - F_k(h_k)]^n \left\{ w_{ER} - E[\pi_k | w_c^*, n] \right\}$ , then  $EU_{cp}(w_{cp}^*) > EU_c(w_c^*)$ .*

*Proof:* By Proposition 1,  $[1 - F_k(w_c^*)]^n > 0 \forall k$ , meaning that in equilibrium, a positive measure of consumers will not receive primary care under the capitation system. Due to the rules of the auction stage under the capitation-plus system, all consumers eventually receive primary care. This means that, for a given price of health insurance, the capitation-plus system delivers a

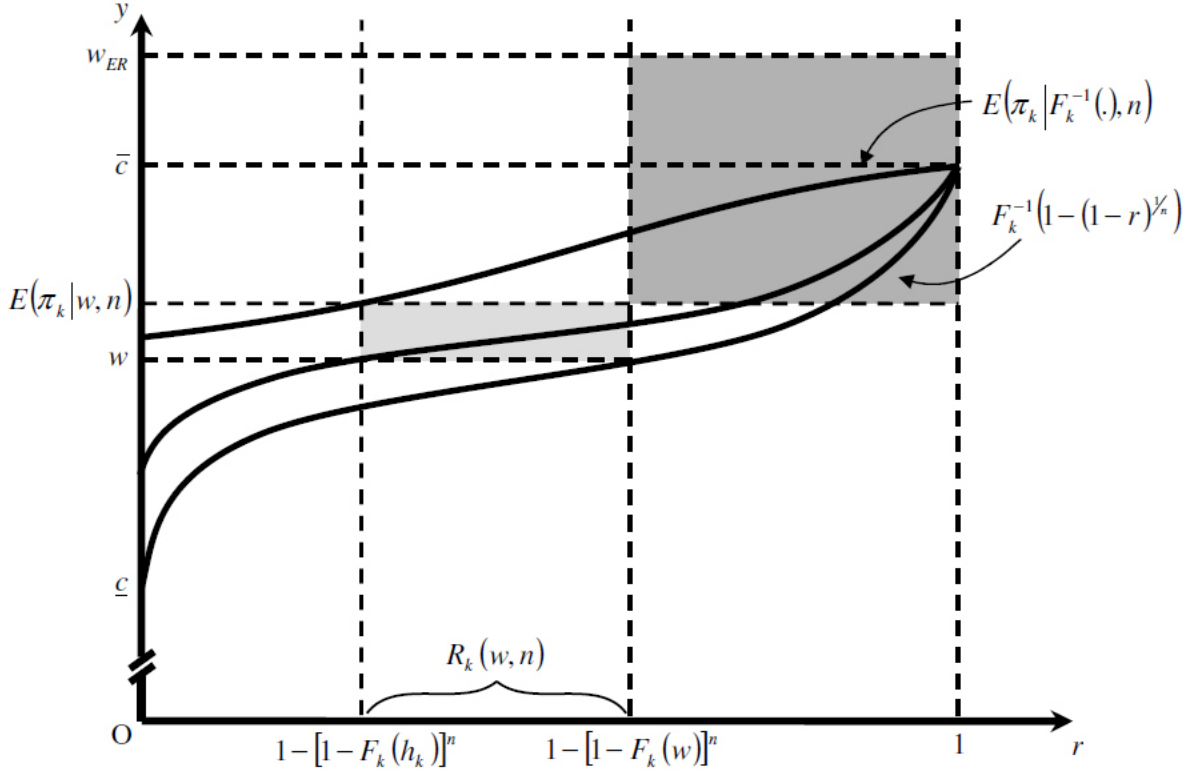


Figure 4.3: Spending differences under the two reimbursement methods for a fixed capitation payment

higher expected utility than the capitation system. The above condition implies  $\tau_{cp}(w_c^*) \leq \tau_c(w_c^*)$ . Since  $w_{cp}^*$  is the solution to the cost-minimization problem, it must be that  $\tau_{cp}(w_{cp}^*) \leq \tau_{cp}(w_c^*)$ . The condition is thus sufficient for  $\tau_{cp}(w_{cp}^*) \leq \tau(w_c^*)$ , and therefore  $EU_{cp}(w_{cp}^*) > EU_c(w_c^*)$ . ■

For a given capitation payment, there are two effects on the price of insurance to consider when switching from a pure capitation system to a capitation-plus-auction system. First, under competitive bidding, those patients allocated

by auction cost less than those who would have gone unallocated in the pure capitation system. These are the cost savings from introducing competitive bidding. On the other hand, the prospect of positive economic profits from winning in the auction stage means that providers reject more often for a given capitation payment. Since patients cost more when allocated by auction than by capitation, this second effect is a cost increase from adopting competitive bidding. If the first effect (the darker rectangle in Figure 4.3) is greater than the second effect (the lighter rectangle), the competitive bidding system dominates the capitation system.

## **4.6 Summary and Implications**

This chapter presents a novel method of compensation intended to mitigate the problem of risk selection. This compensation system works in two stages, where providers can first agree to accept and treat patients in exchange for a capitation payment. Providers can reject those patients they find unprofitable, but for each rejection, the provider must submit to the payer a minimum amount at which they would agree to treat the rejected patient. After these initial acceptances, those patients who did not find a provider willing to treat them are allocated to the lowest bidding provider, which is compensated at a rate equal to the second lowest submitted amount. The chapter also models a compensation system based purely on capitation, and compares the two

systems on the amount of risk selection and expected cost. Findings include, first, that a mixed system dominates a pure competitive bidding system. Second, that risk selection occurs in equilibrium in a pure capitation system, and never occurs in the mixed system. Finally, the model reveals sufficient conditions under which the mixed system dominates the pure capitation system.

A number of significant implications emerge from these findings. First, the addition of a competitive bidding component to widely used capitation systems can eliminate risk selection while preserving provider incentives to contain costs. This does not mean that competitive bidding solves providers' incentives to undertreat patients, only that it leaves them unchanged from those under capitation or prospective payment. In this way, selection is reduced without increasing inefficiency and without introducing demand side cost sharing, thus keeping consumers fully insured. Second, a competitive bidding component eliminates risk selection regardless of the number of unverifiable ex ante consumer risk types ( $K$ ) and the variance in match-specific treatment costs. This suggests that costly risk adjustment schemes; where payers take steps measure patient characteristics, estimate their effects on treatment costs, and condition payments on the findings; would be unnecessary for overcoming risk selection once competitive bidding is introduced. This is not to say that risk adjustment is entirely unwarranted, as this model does not investigate whether partitioning the set of consumer risk types can reduce the price of health insurance overall. Third, capitation and competitive bidding are best utilized in

a mixed allocation system. This is a departure from past efforts at designing provider payment methods, where capitation and competitive bidding systems have been mutually exclusive options.

## 4.7 Bibliography

- Barros, P.P. 2003. Cream-skimming, incentives for efficiency and payment system. *J. of Health Economics* **22** 419–443.
- Beck, K. 2000. Growing importance of capitation in Switzerland. *Health Care Management Science* **3** 111–119.
- Beck, K., M. Trottmann, and P. Zweifel. 2010. Risk adjustment in health insurance and its long-term effectiveness. *J. of Health Economics* **29** 489–498.
- Berwick, D.M., and A.D. Hackbarth. 2012. Eliminating Waste in US Health Care. *J. of the American Medical Association* **307**(14) 1513–1516.
- Bovbjerg, R.R., P.J. Held, and M.V. Pauly. 1987. Privatization and Bidding in the Health-Care Sector. *J. of Policy Analysis and Management* **6**(4) 648–666.
- Brown, J., M. Duggan, I. Kuziemko, and W. Woolston. 2011. Does Risk-selection Respond to Risk-adjustment? Evidence from the Medicare Advantage Program. NBER Working Paper 16977.
- Chalkley, M. and J.M. Malcomson. 2000. Government Purchasing of Health Services, ch. 15 p. 847–890 in Culyer, A.J. and J.P. Newhouse eds, *Handbook of Health Economics* **1**, Elsevier.
- Christianson, J.B. and K.R. Smith. 1984. Options in the design of competitive-bidding processes for indigent medical care. *Contemporary Economic Policy* **3**(2) 55–68.
- Diamond, P. 1992. Organizing the Health Insurance Market *Econometrica* **60**(6) 1233–1254.

- Dranove, D. 1987. Rate-setting by diagnosis related groups and hospital specialization. *The RAND J. of Economics* **18**(3) 417–427.
- Eggleston, K. 2000. Risk Selection and Optimal Health Insurance-Provider Payment Systems *The J. of Risk and Insurance* **67**(2) 173–196.
- Ellis, R.P. 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *J. of Health Economics* **17** 537–555.
- Ellis, R.P. and T.G. McGuire. 1986. Provider behavior under prospective reimbursement. *J. of Health Economics* **5** 129–151.
- Feldman, R., R. Coulam and B. Dowd. 2012. Competitive Bidding Can Help Solve Medicare's Fiscal Crisis.  
[http://www.aei.org/files/2012/02/16/-competitive-bidding-can-helpsolve-medicares-fiscal-crisis\\_081704430956.pdf](http://www.aei.org/files/2012/02/16/-competitive-bidding-can-helpsolve-medicares-fiscal-crisis_081704430956.pdf). Accessed: July 4, 2012.
- Frank, R.G., J. Glazer, and T.G. McGuire. 2000. Measuring adverse selection in managed health care. *J. of Health Economics* **19** 829–854.
- Garrott, P. 2007. Competitive Bidding - the Continuing Saga. *Clinical Laboratory Science* **20**(4) 194–195.
- Glazer, J. and T.G. McGuire. 2000. Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care. *The American Economic Review* **90**(4) 1055–1071.
- Hillman, D.G., and J.B. Christianson. 1984. Competitive Bidding as a Cost-Containment Strategy for Indigent Medical Care: The Implementation Experience in Arizona. *J. of Health Politics, Policy, and Law* **9**(3) 427–451.
- Hoerger, T.J., and T.M. Waters. 1993. Competitive Bidding for Medicare Services. *Medical Care* **31**(10) 879–897.
- Hogan, A.J. 1983. Financing long term care under Medicaid: competitive bidding, vouchers, and expenditure matching in an HMO framework. *Socio-Economic Planning Sciences* **17**(3) 131–140.
- Huh, W.T. and R.O. Roundy. 2002. Using auctions for procurement. *Cornell University ORIE Technical Reports* (1355). <http://hdl.handle.net/1813/9233>.
- Iglehart, J.K. 1995. Health Policy Report: Medicaid and Managed Care. *The New England J. of Medicine* **332**(25) 1727–1731.



- Katzman, B. and K.A. McGeary. 2008. Will Competitive Bidding Decrease Medicare Expenditures? *Southern Economic J.* **74**(3) 839–856.
- Keijser, G.M.W.M., and B.L. Kirkman-Liff. 1992. Competitive bidding for health insurance contracts. *Health Policy* **21** 35–46.
- Kifmann, M. and N. Lorenz. 2011. Optimal Cost Reimbursement of Health Insurers to Reduce Risk Selection. *Health Economics* **20** 532–552.
- Kirkman-Liff, B.L., J.B. Christianson, and D.G. Hillman. 1985. An Analysis of Competitive Bidding by Providers for Indigent Medical Care Contracts. *HSR: Health Services Research* **20**(5) 549–577.
- McCombs, J.S., and J.B. Christianson. 1987. Applying Competitive Bidding to Health Care. *J. of Health Politics, Policy, and Law* **12**(4) 703–722.
- Mechanic, R. and S. Altman. 2010. Medicare's Opportunity to Encourage Innovation in Health Care Delivery. *The New England J. of Medicine* **362**(9) 772–774.
- Milgrom, P.R. 1985. Auction Theory. Cowles Foundation Discussion Paper no. 799.
- Milgrom, P.R. and R.J. Weber. 1982. A Theory of Auctions and Competitive Bidding *Econometrica* **50**(5) 1089–1122.
- Myerson, R.B. 1981. Optimal Auction Design. *Mathematics of Operations Research* **6**(1) 58–73.
- Newhouse, J.P. 1996. Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection. *J. of Economic Literature* **34**(3) 1236–1263.
- Nicholson, S., K. Bundorf, R.M. Stein, and D. Polsky. 2004. The Magnitude and Nature of Risk Selection in Employer-Sponsored Health Plans. *HSR: Health Services Research* **39**(6) 1817–1838.
- Paringer, L. and N. McCall. 1991. How competitive is competitive bidding? *Health Affairs* **10**(4) 220–230.
- Pauly, M.V. 1984. Is cream-skimming a problem for the competitive medical market?. *J. of Health Economics* **3** 87–95.
- Pauly, M.V., P. Danzon, P. Feldstein, J. Hoff. 1991. A plan for 'responsible national health insurance'. *Health Affairs* **10**(1) 5–25.

- Schlesinger, M, R.A. Dorwart, and R.T. Pulice. 1986. Competitive Bidding and States' Purchase of Services: The Case of Mental Health Care in Massachusetts. *J. of Policy Analysis and Management* **5**(2) 245–263.
- Schokkaert, E. and C. Van de Voorde. 2004. Risk selection and the specification of the conventional risk adjustment formula. *J. of Health Economics* **23** 1237–1259.
- Smits, M. and R. Janssen. 2008. Impact of Electronic Auctions on Health Care Markets. *Electronic Markets* **18**(1) 19-29.
- van de Ven, W.P.M.M., K. Beck, F. Buchner, D. Chernichovsky, L. Gardiol, A. Holly, L.M. Lamers, E. Schokkaert, A. Shmueli, S. Spycher, C. Van de Voorde, R.C.J.A. van Vliet, J. Wasem, I. Zmora. 2003. Risk adjustment and risk selection on the sickness fund insurance market in five European countries. *Health Policy* **65** 75–98.
- van de Ven, W.P.M.M., K. Beck, C. Van de Voorde, J. Wasem, I. Zmora. 2007. Risk adjustment and risk selection in Europe: 6 years later. *Health Policy* **83** 162–179.
- Waters, R.J. 2006. The outlook for CMS competitive bidding: Part II. *MLO: Medical Laboratory Observer* **38**(10) 56.

## **Chapter 5**

### **Conclusion**

The production of health care and health insurance involves a set of complex interactions among payers, health insurers, and providers. These interactions are affected by the environment in which they take place, which includes a variety of external factors that are subject to change. Many valuable contributions have been made by empirical studies attempting to estimate the effects of changes in these external factors on the cost and quality of health care. The subtleties in the behaviour that makes up these interactions can produce unexpected results when external factors change, and so rigorous theoretical analysis can offer insight when observed outcomes are difficult to explain.

This is the intended contribution of Chapters 2 and 3 to investigations

studying defensive medicine. They provide explanations for seemingly inconsistent results across empirical studies, and reveal the potential for externalities and omitted variable bias. Both find that the relationships between malpractice pressure and health care spending and quality are non-monotonic, both rising initially up to a threshold and declining thereafter. This can leave defensive medicine hidden from empirical methods designed to investigate for monotonic relationships. The fourth chapter utilizes auction theory in designing a mechanism for payers to compensate providers that mitigates risk selection while holding onto provider incentives to operate efficiently. It shows sufficient conditions for a mixed system of capitation and competitive bidding to dominate systems utilizing pure capitation or bidding.

These three chapters are the beginning of a research agenda including both empirical and theoretical components. The empirical implications emerging from the first two chapters suggest alternate specifications for investigating the degree of defensive medicine practiced in today's health care markets. Important to these new investigations is the use of access measures in a novel way: to determine an observation's equilibrium type or the prevailing equilibrium type of a given population. Regarding the theoretical component, the use of auction theory in allocating patients across providers would benefit from further extensions. These include the use of common value auctions and two sided auctions, where consumers have some input over the set of providers bidding on their contract. This has the potential to reveal more information to payer,

and thus overcome more problems created by information asymmetry.

While the debate surrounding the best way to structure health care markets may never be resolved, it is hoped that the findings in these three chapters provide researchers and policymakers with new insight into payer-provider interactions and a sophisticated intuition for use in predicting the impacts of malpractice reform and innovations in provider compensation.

# Appendix A

## Chapter 2 Appendix

### A.1 Proof of Proposition 2.1

Combining FOCs 2.1 and 2.2:

$$\frac{\left(\frac{\partial \bar{Q}}{\partial w}\right)}{\left(\frac{\partial \bar{Q}}{\partial s}\right)} = \frac{\left(\frac{\partial \tau}{\partial w}\right)}{\left(\frac{\partial \tau}{\partial s}\right)}$$

and

$$\frac{w}{P} = \rho - t \frac{\partial \rho}{\partial t}$$

FOCs become:

$$\begin{aligned}
& \frac{\left(\frac{\partial \bar{Q}}{\partial w}\right)}{\left(\frac{\partial \bar{Q}}{\partial s}\right)} = \frac{\left(\frac{\partial \tau}{\partial w}\right)}{\left(\frac{\partial \tau}{\partial s}\right)} \\
& \dots \\
& \frac{1 - \rho + t \frac{\partial \rho}{\partial t}}{1 - \rho} = \frac{tP \frac{\partial^2 \rho}{\partial t^2} + \frac{w}{t}}{c + \frac{w}{t}} \\
& \dots \\
& \frac{1 - \rho}{w + ct} = \frac{-\frac{\partial \rho}{\partial t}}{c - Pt \frac{\partial^2 \rho}{\partial t^2}} \tag{A.1}
\end{aligned}$$

Condition (A.1) is the same tangency condition necessary to solve the problem:

$$\begin{aligned}
& \max_{n,t} \left\{ Q(n,t) = \left(\frac{D}{q}\right) n(1 - \rho) \right\} \\
& \text{subject to } \tau = Dn(\omega(t;P) + ct) \\
& \text{where } \omega(t;P) = P\left(\rho - t \frac{\partial \rho}{\partial t}\right)
\end{aligned}$$

The isoquants in this problem are convex, but isocosts are not. It is thus necessary to investigate whether there is a  $t$  in the choice set that can satisfy Condition (A.1) when Assumptions 1 and 2 hold, and if it does exist, whether or not it is unique.

### Existence of tangency point

Not only is it necessary to show that a  $t$  satisfying (A.1) exists, but it must exist given that Assumptions 1 and 2 hold, that is where  $\frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} < 0$  in order for that  $t$  to be an argmax.

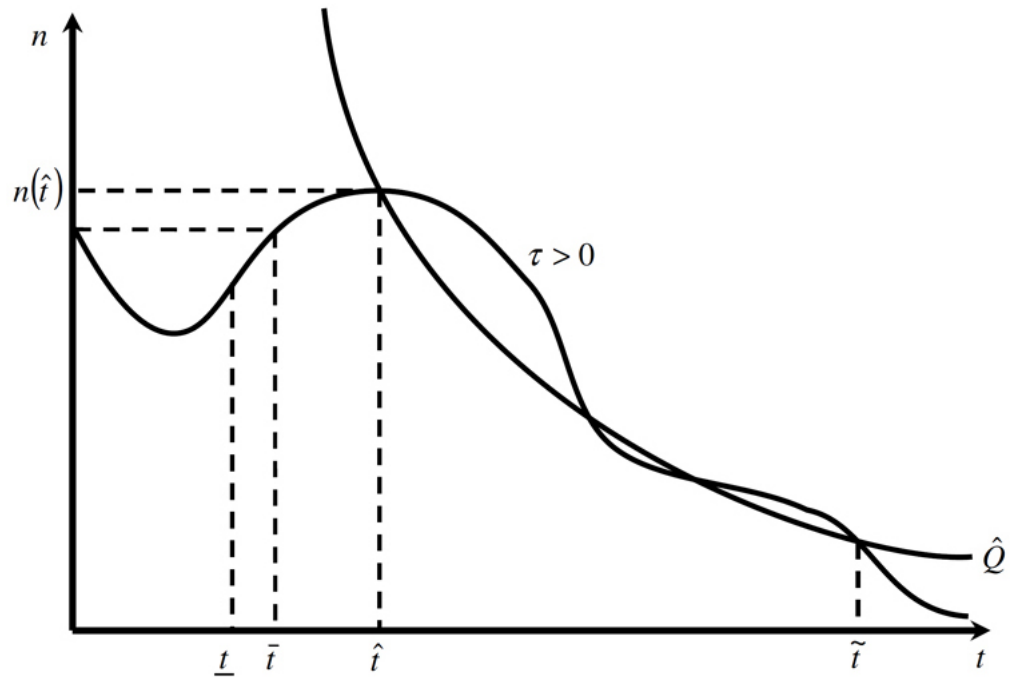


Figure A.1: Existence of a point satisfying (A.1) in the range of  $t$  where Assumptions 1 and 2 hold.



An illustration of all points described can be found in Figure A.1. By Assumption 2, on the same isocost curve, there exists an interior level of treatment  $\bar{t}$  such that  $n$  is as high as where  $t = 0$ . Let  $\hat{Q}$  represent the level of quality described by the isoquant that passes through the maximum  $n$  achievable on a given isocost curve, defined as  $n(\hat{t})$  where  $\hat{t}$  is the level of treatment at which the maximum  $n$  is reached. Note that  $\hat{Q} > 0$ ,  $\hat{t} > 0$ , and by Assumption 1,  $\frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} < 0$  holds for all  $t \geq \hat{t}$ . At  $\{\hat{t}, n(\hat{t})\}$ , the slope of the isocost curve is zero while the slope of the isoquant is negative. Also, since  $\hat{Q} > 0$  and for a given  $\tau > 0$ ,  $\lim_{t \rightarrow 0} Q = 0$ , there is a level of treatment  $\tilde{t}$  where the isoquant and isocost intersect and the isoquant remains above the isocost for all  $t > \tilde{t}$ . Thus, there is a negative difference between isocost and isoquant for all  $t \notin [\hat{t}, \tilde{t}]$ , at least a subset of  $[\hat{t}, \tilde{t}]$  such that there is a positive distance, and zero distance at  $\hat{t}$  and  $\tilde{t}$ . Note that all  $t \notin [\hat{t}, \tilde{t}]$  are irrelevant alternatives, and thus the set of relevant choices is compact.

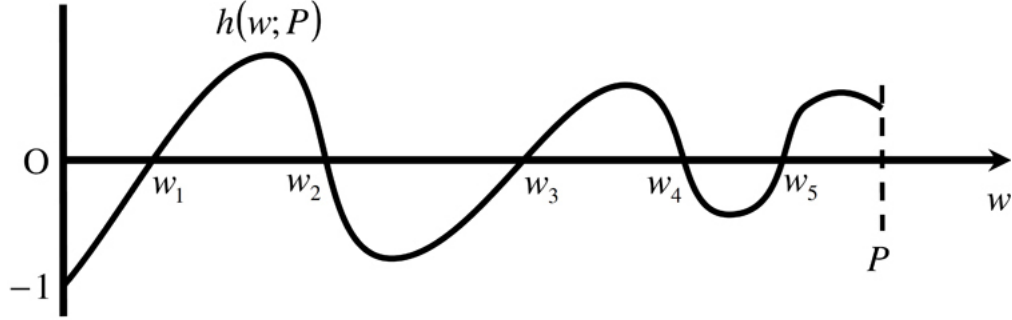
At  $\hat{t}$ , the values of  $n$  in the isoquant describing  $\hat{Q}$  and the isocost describing  $\tau$  are equal, while the slope of the isocost is greater than that of the isoquant. This implies that, at  $\hat{t}$ , the left side of (A.1) is less than the right side. At  $\tilde{t}$ , the values of  $n$  in the isoquant describing  $\hat{Q}$  and the isocost describing  $\tau$  are equal, while the slope of the isocost is less than that of the isoquant. This implies that, at  $\tilde{t}$ , the left side of (A.1) is greater than the right side. Since both isoquant and isocost are continuous, by intermediate value theorem there exists at least one point  $t^* \in [\hat{t}, \tilde{t}]$  such that (A.1) holds. Since (A.1) is solely a function of  $t$ , tangency will hold at  $t^*$  for any  $n$ . Therefore, there always exists a point of tangency in the region where Assumptions 1 and 2 hold.

### Unique maximum

By subbing in the solution to the physician's problem, (A.1) can be rearranged into:

$$\frac{w}{P} = \frac{c - PA(t)}{c - PB(t)} \quad (\text{A.2})$$

where  $A(t) = (1 - \rho)t \frac{\partial^2 \rho}{\partial t^2} > 0$  and  $B(t) = \frac{\partial \rho}{\partial t} < 0$ . Note from the physician's problem that for a given  $P$ , there is a unique finite  $t$  for any  $w \in (0, P)$ , where  $t$  is monotonically decreasing in  $w$ . Let:



**Figure A.2:** Illustration of uniqueness result. Even though (A.1) holds at  $w_1$  to  $w_5$ , only  $w_1$  can hold in equilibrium where Assumptions 1 and 2 hold.

$$h(w; P) = \frac{w}{P} - \frac{c - PA(t)}{c - PB(t)}$$

For an illustration, see Figure A.2. Conditions (A.1) and (A.2) are satisfied where  $h(w; P) = 0$ . From the physician's problem,  $\lim_{w \rightarrow 0} t = \infty$  and  $\lim_{w \rightarrow P} t = 0$ . This means that  $\lim_{w \rightarrow 0} h(w; P) = -1$  and  $\lim_{w \rightarrow P} h(w; P) > 0$ . Since  $\rho$  is continuous, by intermediate value theorem there exists at least one  $w$  such that  $h(w; P) = 0$ . The first derivative of  $h(\cdot)$  is:

$$\frac{\partial h(\cdot)}{\partial w} = \frac{1}{P} + \frac{P \frac{\partial t}{\partial w} \left[ \frac{\partial A}{\partial t} (c - PB) - \frac{\partial B}{\partial t} (c - PA) \right]}{(c - PB)^2}$$

where the sign of  $\frac{\partial h(\cdot)}{\partial w}$  depends on the condition:

$$\begin{aligned} \frac{\partial h(\cdot)}{\partial w} &\begin{cases} \geq 0 \\ < 0 \end{cases} \\ \left( \frac{\rho}{1 - \rho} \right) \left( \frac{\partial^2 \rho}{\partial t^2} \right) + \frac{A(c - PB)}{P(1 - \rho)^2} - \left( \frac{\partial^2 \rho}{\partial t^2} \right) \cdot h(\cdot) &\begin{cases} \geq \frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} \\ < \frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} \end{cases} \end{aligned}$$

Where the left side is greater than zero whenever  $h(w; P) = 0$ . Let  $\{w_1, w_2, \dots\}$  be values of  $w$  such that  $h(w; P) = 0$ , and let  $\{t_1, t_2, \dots\} = \{t(w_1), t(w_2), \dots\}$ . Let  $w_i$  be increasing in  $i$ , which implies that  $t_i$  is decreasing in  $i$ . If any  $t_i$  is such that  $\frac{\partial h(\cdot)}{\partial w} \leq 0$ , then  $\frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} \geq 0$  at  $t_i$  and, by Assumption 1, at all  $t_j$  such that  $j > i$ . Since we've already shown that a point of tangency exists where  $\frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} < 0$ , and that  $h(w; P)$  is continuous, it must be true that that  $t_1$  is the only element of the set  $\{t_1, t_2, \dots\}$  that can hold in equilibrium where Assumptions 1 and 2 hold. Therefore, the level of treatment  $t^*(c, P)$  and revenue-per-patient  $w^*(c, P)$  satisfying (1) and (2) are unique and depend only on  $c$  and  $P$ . The  $c$  is suppressed in the remainder of this proof.

Multiplying both sides of FOC (1) by  $-P \left(\frac{s}{n^2}\right) \left(\frac{\partial^2 \rho}{\partial t^2}\right)$  and adding each side to each side of FOC (2) yields the necessary condition:

$$\begin{aligned} \Delta U \cdot \left(-\frac{\partial \rho}{\partial t}\right) &= WMU_y \cdot \left[c - Pt \left(\frac{\partial^2 \rho}{\partial t^2}\right)\right] \\ \frac{\left(-\frac{\partial \rho}{\partial t}\right)}{c - Pt \left(\frac{\partial^2 \rho}{\partial t^2}\right)} &= \frac{WMU_y}{\Delta U} \end{aligned}$$

Substituting  $w^*(P)$  and  $t^*(P)$  into this condition leaves the left side unchanging in  $s$  while the right side is monotonically increasing in  $s$  through its effect on  $n$ . There is thus a unique  $s^*$  solving this necessary condition whenever  $\tau^* > 0$ , and therefore a unique  $n^*(w^*, s^*, P)$  for any given  $P$  solving the insurer's first-order conditions. If this  $n^*(w^*, s^*, P)$  is less than  $\frac{q}{D}$ , then the unique limited-access equilibrium is  $\{w^*, s^*\}$ . If it is greater than  $\frac{q}{D}$ , then the  $n$  solving the insurer's first-order conditions is infeasible and setting  $n = \frac{q}{D}$  on the boundary dominates any  $n < \frac{q}{D}$  due to the uniqueness of  $n^*(w^*, s^*, P)$ . This would give rise to the full-access equilibrium, where  $t = \frac{Ds}{q}$ . FOC (3) is the same as the necessary condition above, with the exception that the left side is also decreasing in  $s$  over all values that can hold in equilibrium. This means that there is a unique  $s^*$  in the full-access equilibrium, and that the full-access equilibrium  $\{\omega(s^*; \frac{q}{D}, P), s^*\}$  is also unique. ■

## A.2 Proof of Proposition 2.2

Assumptions 1 and 2 ensure that any equilibrium  $s^*$  must occur to the right of the global minimum of  $\tau(s; \bar{n}, P)$ , where  $\tau(s; \bar{n}, P)$  is monotonically increasing in  $s$ . The function can therefore be inverted, the inverse  $\sigma(\tau; \bar{n}, P)$  being resources as a function of spending. Using the inverse, the insurer's problem can be rewritten with  $\tau$  as the only choice variable:

$$\max_{\tau} \left\{ (1-q)U(H_1, m-\tau) + qU(H_2, m-\tau) + q\tilde{Q}(\tau, P)\Delta U \right\}$$

$$\text{where } \tilde{Q}(\tau, P) = 1 - \rho\left(\frac{D \cdot \sigma(\tau; \frac{q}{D}, P)}{q}\right)$$

This problem yields the first-order condition:

$$q\left(\frac{\partial Q}{\partial \tau}\right)\Delta U = WMU_y$$

and the comparative static:

$$\frac{d\tau}{dP} = \frac{\left(\frac{\partial WMU_y}{\partial P}\right) - q\Delta U\left(\frac{\partial^2 Q}{\partial \tau \partial P}\right)}{q\left(\frac{\partial \Delta U}{\partial \tau}\right)\left(\frac{\partial Q}{\partial \tau}\right) + q\Delta U\left(\frac{\partial^2 Q}{\partial \tau^2}\right) - \left(\frac{\partial WMU_y}{\partial \tau}\right)}$$

Since  $WMU_y > 0$ ,  $\frac{\partial WMU_y}{\partial P} < 0$ ,  $\frac{\partial \Delta U}{\partial \tau} < 0$ ,  $\frac{\partial Q}{\partial \tau} > 0$ ,  $\frac{\partial WMU_y}{\partial \tau} > 0$ , the sign of  $\frac{d\tau}{dP}$  depends on the signs of  $\frac{\partial^2 Q}{\partial \tau \partial P}$  and  $\frac{\partial^2 Q}{\partial \tau^2}$ .

$$\frac{\partial^2 Q}{\partial \tau \partial P} = -\left(\frac{D}{q}\right)\left[\frac{\partial^2 \rho}{\partial t^2}\left(\frac{D}{q}\right)\frac{\partial \sigma}{\partial P}\frac{\partial \sigma}{\partial \tau} + \frac{\partial \rho}{\partial t}\left(\frac{\partial^2 \sigma}{\partial \tau \partial P}\right)\right]$$

where,

$$\frac{\partial^2 \sigma}{\partial \tau \partial P} = D^{-1}\left[c - Pt\frac{\partial^2 \rho}{\partial t^2}\right]^{-2}\left[t\frac{\partial^2 \rho}{\partial t^2} + P\left(\frac{D}{q}\right)\frac{\partial \sigma}{\partial P}\left[\frac{\partial^2 \rho}{\partial t^2} + t\frac{\partial^3 \rho}{\partial t^3}\right]\right]$$

Since Assumptions 1 and 2 imply that  $\frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} < 0$  in any equilibrium,  $\frac{\partial^2 \sigma}{\partial \tau \partial P}$  is positive, and therefore  $\frac{\partial^2 Q}{\partial \tau \partial P}$  is positive as well.

$$\frac{\partial^2 Q}{\partial \tau^2} = - \left( \frac{D}{q} \right) \left[ \frac{\partial^2 \rho}{\partial t^2} \left( \frac{D}{q} \right) \left( \frac{\partial \sigma}{\partial \tau} \right)^2 + \frac{\partial \rho}{\partial t} \left( \frac{\partial^2 \sigma}{\partial \tau^2} \right) \right]$$

where,

$$\frac{\partial^2 \sigma}{\partial \tau^2} = P \left( \frac{D^2}{q} \right) \left( \frac{\partial s}{\partial \tau} \right)^3 \left[ \frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} \right]$$

Once again, Assumptions 1 and 2 make  $\frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} < 0$  in any equilibrium. This makes  $\frac{\partial^2 \sigma}{\partial \tau^2}$  negative and thus  $\frac{\partial^2 Q}{\partial \tau^2}$  negative. Along with the previously mentioned signed expressions,  $\frac{\partial^2 Q}{\partial \tau \partial P} > 0$  and  $\frac{\partial^2 Q}{\partial \tau^2} < 0$  make  $\frac{d\tau^*}{dP}$  unambiguously positive.

Since full-access equilibria are defined such that  $\tau^*(P) > \bar{\tau}(P)$ , the marginal change in spending produced by a marginal change in malpractice pressure  $P$  to  $P'$  still leaves  $\tau^*(P') > \bar{\tau}(P')$ . Therefore, a marginal change in  $P$  does not precipitate a departure from the full-access equilibrium, and  $\tilde{n}$  remains unchanged. Thus,  $\frac{\partial \tilde{n}^*}{\partial P} = 0$ . ■

### A.3 Proof of Proposition 2.3

In the full-access equilibrium, the FOC with respect to  $s$  is:

$$q \left( \frac{\partial Q}{\partial s} \right) \Delta U = W M U_y \left( \frac{\partial \tau}{\partial s} \right)$$

Substituting for the first derivatives and rearranging yields:

$$\frac{\left(-\frac{\partial \rho}{\partial t}\right)}{c - Pt \left(\frac{\partial^2 \rho}{\partial t^2}\right)} = \frac{WMU_y}{\Delta U}$$

$$H\left(\frac{Ds}{q}, P\right) = J\left(\frac{Ds}{q}, \tau\right)$$

Since  $\tau = \tau\left(s; \frac{q}{D}, P\right)$  in the full access equilibrium, this is an implicit function describing equilibrium resources ( $s^*$ ). By the Implicit Function Theorem:

$$\frac{ds^*}{dP} = \frac{-\left(\frac{\partial H}{\partial P} - \frac{\partial J}{\partial \tau} \frac{\partial \tau}{\partial P}\right)\Big|_s}{\frac{\partial H}{\partial t} \frac{\partial t}{\partial s} - \frac{\partial J}{\partial t} \frac{\partial t}{\partial s} - \frac{\partial J}{\partial \tau} \frac{\partial \tau}{\partial s}}$$

Where Assumptions 1 and 2 hold, the denominator is unambiguously negative. Therefore,  $\frac{ds^*}{dP}$  will be positive if and only if  $\left(\frac{\partial H}{\partial P} - \frac{\partial J}{\partial \tau} \frac{\partial \tau}{\partial P}\right)\Big|_s$  is positive for a given amount of resources. The expression for  $\left(\frac{\partial H}{\partial P} - \frac{\partial J}{\partial \tau} \frac{\partial \tau}{\partial P}\right)\Big|_s$  can be rearranged into:

$$\left(\frac{\frac{\partial \Delta U}{\partial y}}{\Delta U}\right) WMU_y + \frac{\partial WMU_y}{\partial \tau} < \left[ \frac{t \cdot \Delta U \left(\frac{\partial^2 \rho}{\partial t^2}\right) \left(-\frac{\partial \rho}{\partial t}\right)}{q \left(\rho - t \frac{\partial \rho}{\partial t}\right) \left(c - Pt \frac{\partial^2 \rho}{\partial t^2}\right)^2} \right]$$

Since:

$$\frac{\partial \omega}{\partial t} = -Pt \frac{\partial^2 \rho}{\partial t^2}$$

$$\frac{\partial \tau}{\partial t} = q \left[ c - Pt \frac{\partial^2 \rho}{\partial t^2} \right]$$

$$\omega = P \left[ \rho - t \frac{\partial \rho}{\partial t} \right]$$

$$\frac{\left(-\frac{\partial \rho}{\partial t}\right)}{c - Pt \left(\frac{\partial^2 \rho}{\partial t^2}\right)} = \frac{WMU_y}{\Delta U}$$

The expression for  $\left(\frac{\partial H}{\partial P} > \frac{\partial J}{\partial \tau} \frac{\partial \tau}{\partial P}\right)\bigg|_s$  can further be rearranged into  $|\epsilon_{J,w}| < 1$ , where  $\epsilon_{x,y}$  is the percent change in  $x$  due to a one-percent change in  $y$  and  $J = \frac{WMU_y}{\Delta U}$ . This expression implies  $\frac{ds^*}{dP} > 0$ , and since  $\tilde{n}^*$  is constant in the full-access equilibrium,  $\frac{ds^*}{dP} > 0$  implies that  $\frac{\partial t^*}{\partial P} > 0$  and  $\frac{\partial \tilde{Q}^*}{\partial P} > 0$ . ■

## A.4 Proof of Proposition 2.4

From the physician's problem, there is a unique level of treatment for any combination of  $w$  and  $P$ , defined as  $t(w, P)$ . From proposition 2.1, there is a unique equilibrium  $w$  ( $w^*$ ) and thus a unique equilibrium level of treatment for any value of  $P$ , or  $t^*(P) = t(w^*(P), P)$ . The effect of a change in  $P$  on  $t^*(P)$  is thus:

$$\frac{\partial t^*}{\partial P} = \frac{\partial t}{\partial w} \frac{\partial w^*}{\partial P} + \left(\frac{\partial t}{\partial P}\right)\bigg|_w$$

Using comparative statics from the physician's problem, this can be rearranged into:

$$\frac{\partial t^*}{\partial P} = \frac{\partial t}{\partial w} \left(\frac{\partial w^*}{\partial P} - \frac{w}{P}\right)$$

If  $\frac{\partial w^*}{\partial P}$  is negative, then  $\frac{\partial t^*}{\partial P}$  is unambiguously positive. If  $\frac{\partial w^*}{\partial P}$  is positive, then  $\frac{\partial t^*}{\partial P}$  will be positive as long as  $\frac{w}{P} > \frac{\partial w^*}{\partial P}$ . Using (A.2) where  $t = t^*(P)$ , The first derivative of  $w^*(P)$  can be written as:

$$\frac{\partial w^*}{\partial P} = \frac{(c - PA)^2 - P^2 A(A - B) - \left(\frac{w}{P}\right) P^2 G(t) \frac{\partial t}{\partial w}}{(c - PB)^2 - P^2 G(t) \frac{\partial t}{\partial w}}$$

where  $G(t) = (c - PA) \frac{\partial B}{\partial t} - (c - PB) \frac{\partial A}{\partial t}$ . The denominator will be positive as long as:

$$\frac{\partial^2 \rho}{\partial t^2} + t \frac{\partial^3 \rho}{\partial t^3} < \left(\frac{\rho}{1 - \rho}\right) \left(\frac{\partial^2 \rho}{\partial t^2}\right) + \frac{A(c - PB)}{P(1 - \rho)^2}$$

The right side is positive, so Assumptions 1 and 2 ensure that this inequality holds. The denominator is therefore positive. Equilibrium treatment is increasing in  $P$  if and only if:

$$\frac{w}{P} > \frac{\partial w^*}{\partial P}$$

...

$$B < A$$

■ Since  $B$  is negative and  $A$  is positive, this condition always holds, so  $\frac{\partial t^*}{\partial P} > 0$ .

## A.5 Proof of Proposition 2.5

Since the equilibrium values  $\{w^*, t^*\} = \{w^*(c, P), t^*(c, P)\}$  depend only on  $c$  and  $P$ , resources can be expressed as a function of  $\tau$  and  $P$  alone. Let this function be  $s = \psi(\tau, P)$ :

$$\begin{aligned} \tau &= Dw^*n^*[w^*, \psi(\tau, P), P] + Dc\psi(\tau, P) \\ \Rightarrow \frac{\partial \psi}{\partial \tau} &= \frac{t}{D(w + ct)} > 0, \\ \frac{\partial \psi}{\partial P} &= \left( \frac{nt}{w + ct} \right) \left[ \left( \frac{w + ct}{t} - \left( c - t \frac{\partial^2 \rho}{\partial t^2} \right) \right) \frac{\partial t^*}{\partial P} - \frac{w}{P} \right] \end{aligned}$$

Since (A.1) must hold in equilibrium,  $\frac{\partial \psi}{\partial P}$  can be rearranged into:

$$\frac{\partial \psi}{\partial P} = \left( \frac{n}{1 - \rho} \right) \left[ \left( 1 - \frac{w}{P} \right) \frac{\partial t^*}{\partial P} - t \left( \frac{1 - \rho}{w + ct} \right) \frac{w}{P} \right]$$



Similar to Proposition 2.2, the insurer's problem can be written in terms of  $w^*$  and  $\psi(\tau, P)$ , with  $\tau$  being the choice variable:

$$\begin{aligned} \max_{\tau} \quad & \left\{ (1-q)U(H_1, m-\tau) + qU(H_2, m-\tau) + qQ(\tau, P)\Delta U \right\} \\ \text{where} \quad & Q(\tau, P) = \left( \frac{Dn^*(w^*, \psi, P)}{q} \right) (1 - \rho(t^*)) \end{aligned}$$

Solving yields the same first order condition as in Proposition 2.2, but where  $\frac{\partial Q}{\partial \tau} = \frac{1}{q} \left( \frac{1-\rho}{w+ct} \right) > 0$  and  $\frac{\partial^2 Q}{\partial \tau^2} = 0$ . The FOC can thus be simplified into:

$$\frac{1-\rho}{w+ct} = \frac{WMU_y}{\Delta U}$$

Applying the implicit function theorem yields the same comparative static as in Proposition 2.2, with the exception that  $\frac{\partial^2 Q}{\partial \tau^2} = 0$ , the signs of the terms in the denominator are the same as in Proposition 2.1, making the denominator unambiguously negative. Therefore, the sign of  $\frac{\partial \tau^*}{\partial P}$  will be determined by the sign of the numerator. Breaking down each component of the numerator. Using  $\tilde{Q}(\tau, P) = \left( \frac{Dn^*(w^*, \psi, P)}{q} \right) (1 - \rho(t^*))$  and (A.1), we find  $\frac{\partial WMU_y}{\partial P} = -q \frac{\partial \Delta U}{\partial y} \frac{Dn}{q} \left( \frac{1-\rho}{w+ct} \right) \frac{w}{P}$ , and  $\frac{\partial^2 Q}{\partial \tau \partial P} = -\left( \frac{1}{q} \right) \left[ \frac{(1-\rho) \frac{w}{P}}{(w+ct)^2} \right]$ . Substituting the simplified forms of  $\frac{\partial WMU_y}{\partial P}$  and  $\frac{\partial^2 Q}{\partial \tau \partial P}$  into the numerator of  $\frac{\partial \tau^*}{\partial P}$ :

$$\begin{aligned} \frac{d\tau}{dP} &< 0 \\ \Leftrightarrow \left( \frac{\partial WMU_y}{\partial P} \right) &> q\Delta U \left( \frac{\partial^2 Q}{\partial \tau \partial P} \right) \\ &\dots \\ \tau &< \frac{\Delta U}{\frac{\partial \Delta U}{\partial y}} \end{aligned}$$

Therefore, if and only if  $\tau < \frac{\Delta U}{\frac{\partial \Delta U}{\partial y}}$ , then  $\frac{d\tau^*}{dP} < 0$ . ■

## Curriculum Vitae

**Name:** Daniel Montanera

**Place of Birth:** Toronto, Canada

**Year of Birth:** 1983

**Post-Secondary Education and Degrees:** King's University College at The University of Western Ontario  
London, Ontario  
2002–2006 Combined Honours B.A. Economics and Political Science.

Queen's University  
Kingston, Ontario  
2006–2007 M.A. Economics

The University of Western Ontario  
London, Ontario  
2007–2012 Ph.D. Economics

**Honors and Awards:** Western Graduate Research Scholarship  
The University of Western Ontario  
2007-2011

Ontario Graduate Scholarship  
The University of Western Ontario  
2008 - 2011

Queen's University  
2006 - 2007

**Related Work Experience:** *Teaching Assistant*  
The University of Western Ontario  
2007–2011

*Instructor*  
The University of Western Ontario  
2009–2011

*Limited-term Lecturer*  
Georgia State University  
2012-