THREE ESSAYS ON HEALTH CARE AND INDUSTRIAL ORGANIZATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Nicholas Tilipman

May 2017

THREE ESSAYS ON HEALTH CARE AND INDUSTRIAL ORGANIZATION

Nicholas Tilipman, Ph.D.

Cornell University 2017

United States health spending grew dramatically over the last several decades, reaching approximately 18% of Gross Domestic Product in 2015. Much of these rising costs can be attributed to consolidation of health care providers, the provision of generous insurance plans, and innovations in medical technologies and pharmaceuticals. This dissertation studies how innovation in insurance plan design and pharmaceuticals affect health care prices, consumer behavior, consumer welfare, and the labor market. It also looks at whether certain policy interventions might lead to increased efficiency in these health care settings.

In the first chapter, I study employer incentives to offer narrow-network health plans to their enrollees, and the welfare effects of switching to these plans. To do so, I estimate a model of supply and demand for health insurance offered by a large benefits administrator in Massachusetts, where I endogenize the product menu offered to consumers with respect to hospital and physician networks. I then use these estimates to study how the employer's number of products, networks, premiums, and consumer welfare would respond to a hypothetical tax on expensive health plans, in the style of the ACA "Cadillac Tax." I find that consumers' plan choices are driven primarily by inertia rather than by the value of the plan's network, and that this inertia causes the employer to continue offering plans that add little value in a strict price-versus-provider-choice tradeoff. A 60% tax on health plans in excess of $6,000 annually would cause the employer to drop each of its broad plans in favor of more narrow-network

products, resulting in an approximately 21% reduction in health spending, or $76 per-member-per-month (pmpm). Consumer welfare from being moved to this new menu of products would decrease an average of $58 pmpm. I conclude that incentivizing employers to eliminate broad network products through a tax has the potential to increase social welfare.

While broad-network insurance plans are one source of rising health costs, spending on pharmaceuticals has also increased over the last thirty years, accompanying large gains in life expectancy. The second chapter of this dissertation addresses the question of whether the spending on medical care has been worth the cost. Using discrete choice methods from the industrial organization literature, I, along with coauthors Claudio Lucarelli and Sean Nicholson, construct a series of quality-adjusted price indices for colorectal cancer treatments, a condition for which the average price of treatment increased dramatically from 1993 to 2005, largely due to the approval and widespread use of five new drugs over that period. Of note, we estimate a price index using parameters from a pure characteristics model, a demand technique which drops the idiosyncratic error term that produces undesirable switching patterns in traditional logit models. We find that the naive price index for these treatments greatly overstates the true price increase. In contrast, indices that account for the fact that consumers value the quality gains from pharmaceutical innovation show much more modest price increases. We also find that the magnitude of the price increase varies with modeling assumptions. Traditional logit models tend to overstate the value of product innovation, whereas the pure characteristics model implies a more substantial price increase. These results suggest the importance of modeling assumptions when constructing quality-adjusted indices.

Chapter 3 turns its focus to the labor market, and examines the effects of

state mandated health benefits on job transitions and job separations. Specifically, it looks at variation in state mental health parity legislation throughout the 1990s and 2000s to assess whether the addition of a high-cost benefit mandate has led to any significant displacement effect or treatment effect. While prior studies on mental health parity have focused primarily on health outcomes and the probability of having insurance following a mandate, few have paid attention to labor market outcomes. Those that did primarily reported the effects on levels of employment, ignoring potential effects on job flows. I exploit state variation in mandate passage by using restricted-access data from the Medical Expenditures Panel Survey (MEPS). The state identifiers in this data allow me to estimate changes in both employment and job transitions for the population most likely impacted by mental health mandates: employees with previously held mental health diagnoses. Finally, I provide initial estimates of the effect of the Mental Health Parity and Addiction Equity Act of 2008 on these labor market outcomes. I find that mental health parity has had little effect on employment levels, consistent with prior literature, but has caused a decrease in job separations, primarily employment to employment transitions. I present evidence that at least a third of this decline comes from a reduction in involuntary transitions, implying that mental health parity has a positive effect for individuals with untreated illnesses. In addition, I find that mandates led to increased expenditures for mental health services, further suggesting that the involuntary decline was likely due to a treatment effect of the mandates.

BIOGRAPHICAL SKETCH

Nicholas Tilipman graduated from Cornell University in 2017 with a PhD in Economics. While at Cornell, Nick focused on a variety of topics in health economics, industrial organization, and applied microeconomics. These include the effects of competition and regulation on the behavior of consumers, employers, insurers, and health care providers; the effects of health insurance on the labor market; and the effects of innovation in the pharmaceutical industry. Nick also spent the academic year of 2012-2013 serving as a Staff Economist for the Council of Economic Advisers in Washington, DC. While there, he worked on a variety of policy-relevant issues in labor and health economics. Following graduation, Nick will be starting as an Assistant Professor of Health Policy and Administration in the School of Public Health at the University of Illinois at Chicago.

This dissertation is dedicated to my parents, without whom I would never have had the courage to pursue a PhD in the first place, and to my wife, who for some reason continues to love and support me.

ACKNOWLEDGEMENTS

I am deeply grateful to my PhD advisors, Sean Nicholson, Panle Jia Barwick, and Samuel Kleiner for their valuable advice and encouragement. This dissertation would not be possible without their support, guidance, and constant pushing for me to do better.

I also benefited from helpful discussions with Colleen Carey, Matt Backus, Sherry Glied, Chinhui Juhn, Ross Milton, Nicolas Ziebarth, Ankita Patnaik, Lee Tucker, Wes Yin, as well as members of the Cornell Industrial Organization seminar, the Industrial Organization working group, the Policy Analysis & Management seminar, and the Institute on Health Economics, Health Behaviors, and Disparities lunch.

This project was also made possible by a Human Ecology Alumni Association student grant, as well as a Cornell Institute for the Social Sciences small grant.

TABLE OF CONTENTS

[1]This chapter was coauthored with Claudio Lucarelli and Sean Nicholson

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CADILLAC TAX, NARROW NETWORKS, AND CONSUMER WELFARE

1.1 Introduction

As the market for health care becomes increasingly consolidated and medical care prices continue to soar, health insurers and employers have started offering so-called "narrow network" insurance plans as a means of controlling costs and product differentiating. These plans achieve lower costs, and lower premiums, by significantly limiting the set of hospitals and physicians that an insurer will cover to only those with lower negotiated reimbursement rates. Indeed, approximately 70% of the plans available on the Affordable Care Act Health Insurance Exchanges have been found to be "limited network" plans, covering fewer than 30% of the 20 largest hospitals in the market (McKinsey Center for U.S. Health System Reform, 2013) and about 40% of the plans cover less than 25% of the physicians in the market (Polsky & Weiner, 2015). Proponents have argued that limiting networks could be an important way to reign in health care cost growth. A recent study exploiting a natural experiment in Massachusetts showed that employees switching to narrow network plans spent approximately 40% less on medical care with no discernible adverse effects on health outcomes (Gruber & McKnight, 2016).

However, despite the increasing popularity of narrow network plans on the Exchanges, employers have been slower to adopt, design, and offer narrow network products. In 2016, only 7% of employers nationally offered a narrow network as part of their plan menu (Hall & Fronstin, 2016). Moreover, most employers typically only offer one or two plans to their employees, which tend

to be come from the same insurer and have relatively comprehensive benefits (T. Buchmueller, Carey, & Levy, 2013; Dafny, Ho, & Varela, 2013). One possible explanation for this is that consumers who have insurers through their employer may not be very price sensitive, in part because employers subsidize a large portion of employee premiums and in part because health benefits offered by employers are tax-deductible (Powell, 2016). Another explanation may be that consumers of employer-sponsored insurance may simply place high value on access to a broad array of providers. Indeed, narrow network products have generated significant controversy, with critics arguing that wider adoption would represent a significant loss to consumer choice and has the potential to adversely affect medical care.[1] A third explanation may be that search frictions and inertia prevent employees and firms from switching into narrow plans even when they might benefit from the reduced cost of the smaller network. Finally, offering a choice of plans may be costly for firms, who may face higher administrative fees and higher fixed costs from benefit design, negotiating with multiple insurers, educating consumers, and collecting premiums (K. Bundorf, 2002). As such, firms may be reluctant to offer narrow network products alongside broad network products unless they can ensure a substantial share of workers will take-up the plans.

In order to encourage larger firm take-up of lower cost insurance products, policymakers have proposed a "High Cost Health Plan" tax, otherwise known as the "Cadillac Tax." This policy, which was initially set to begin in 2018 but was delayed until 2020, imposes a 40% excise tax on all health plans with annual premiums exceeding $10,200 for individuals and $27,500 for families. Proponents argue that this tax would curb health care cost growth by inhibiting

[1] See for instance, commentary in Politico. http://www.politico.com/story/2013/12/doctors-the-new-gop-weapon-in-obamacare-fight-100617.html

employer and insurer ability to offer expensive insurance plans, and perhaps even incentivizing them to make investments in more efficient products, such as narrow networks. Critics argue, however, that the tax would instead result in significant welfare loss, with consumers either being moved into plans that they do not value or with firms passing on the additional costs of the tax in the form of even higher premiums.[2] Whether or not the welfare changes from being moved to a new menu of products exceeds the change in health spending depends critically on how sensitive consumers are to the price of their plans, how much consumers value the networks of providers they have access to, and how costly it is for firms to alter their product menu.

To study these issues, I estimate a model of supply and demand for health insurance plans for a large-group purchaser in Massachusetts, where I endogenize the employer's choice of menu of products (in particular, the insurers offered and the networks of those insurers). On the demand side, I model consumer demand for hospitals and physician practices, and then consumer demand for insurance plans. On the supply side, I model the employer and insurer decision on a set of products and provider networks to offer to its enrollees and customers, with employers incurring a fixed cost for each product combination offered. After estimating these model primitives, I run the counterfactual: how would the employer's product offerings, plan premiums, and consumer welfare change at various thresholds of a hypothetical "Cadillac Tax?"

To estimate the model, I use claims data from the Massachusetts All-Payer Claims Database (APCD). These data provide detailed information on the medical claims of each insurer licensed to operate in the state of Massachusetts, in-

[2] http://www.nytimes.com/2015/10/06/upshot/the-cadillac-tax-loved-by-economists-and-few-others.html

cluding diagnosis and procedure codes for each provider visit, individual identifiers, provider identifiers, and a wide variety of payment variables. I focus specifically on the claims and choices of one particular employer group: the Group Insurance Commission (GIC). The GIC is a large purchaser of health insurance in Massachusetts, offering coverage to approximately 300,000 enrollees a year, including active state government employees, as well as retirees and the employees of several municipalities. In this way, it acts as both a type of social planner and as a sort of employer Exchange, offering various products to all employees who participate in the group. The GIC is an ideal setting for studying the welfare effects of narrow network products for many reasons. First, the GIC has, in the last several years, been active in encouraging the creation and adoption of narrow network products, allowing me to estimate demand for insurance plans on a large section of the demand curve, rather than relying on minor differences between coverage of broad network products. In addition, in 2012, the GIC offered a three-month "premium-holiday" offering all state employees three months of free coverage if they switched from a broad-network to a narrow network product (Gruber & McKnight, 2016). This policy change, along with a five-year sample of choices of both new and existing enrollees on the GIC, allows me to separately identify consumer valuation of networks from plan inertia (i.e. wanting to remain on the same plan year-after-year), which has not previously been done in an employer market setting.[3]

This paper relates to several strands of literature. These include models of product entry, innovation, and variety that endogenize firm product quality choices (Nosko, 2014; Eizenberg, 2014; Mohapatra & Chatterjee, 2015), litera-

[3]The exception is Handel 2013, which studies the impact of plan inertia in an employer setting. However, Handel studied a market in which the variation in networks between plans was minimal, and thus was not considered explicitly in the analysis.

ture on the growing consolidation and bargaining power of physician groups (A. C. Dunn & Shapiro, 2014; Hausman & Lavetti, 2016; Kleiner, White, & Lyons, 2015; Baker, Bundorf, & Royalty, 2014), insurance plan choice, competition, and provision (Ericson & Starc, 2015b, 2016; Dafny, 2010; Dafny, Duggan, & Ramanarayanan, 2012; Dafny et al., 2013; Scheffler, Arnold, Fulton, & Glied, 2016), network formation (Ho, 2006, 2009; Shepard, 2016; Lee & Fong, 2013), and the effects of insurance plan networks (Gruber & McKnight, 2016; LoSasso & Atwood, 2015; Dafny, Hendel, & Wilson, 2015; Ericson & Starc, 2015a). Of particular importance is Shepard (2016), who uses a similar model to study whether adverse selection leads to the narrowing of networks on the individual market. Also of note is Prager (2016), who uses the Massachusetts APCD to estimate a bargaining model between insurers and hospitals and uses these estimates to simulate the effect of raising GIC tier levels on negotiated prices between insurers and providers (Prager, 2016).

I offer three main contributions to the existing literature. The first is that this is, to my knowledge, the first paper to model consumer demand for insurance plans incorporating valuations for physician practice networks in addition to hospital networks. Much of the existing literature on networks has exclusively focused on hospitals (Ho, 2009; Shepard, 2016; Prager, 2016) and has ignored the role of physicians in determining consumer choice of insurance plans. This is likely due to three factors. First, until recently, physician markets were often thought to be less interesting than hospital markets, as physicians had very little bargaining power to leverage high prices from insurance plans. Second is the sheer dimensionality of the problem: whereas there are typically a small number of hospitals in any given market, there are often thousands of physicians of various specialties, rendering the study of physician markets difficult

in structural IO models. Finally, there is the lack of available data allowing researchers to both link individual physicians to their respective medical groups and construct physician networks of insurance plans.

I circumvent these issues by merging the APCD with proprietary data from the SK&A in order to create linkages between physicians and group practices, hospitals, and health systems. Using these linkages as well as publicly available data on hospital and medical group networks from the GIC, I am able to create a dataset of physician networks on each insurance plan offered over time. Moreover, by affiliating each individual physician to a set of practice groups, I am able to significantly reduce the dimensionality problem by estimating demand on for larger physician entities rather than individual physicians themselves.

I model demand for physician practices of three different specialty groups: primary care physicians, cardiologists, and orthopedists. Together, these specialties comprise approximately 65% of all physician office visits.[4] The model follows the existing literature in estimating network "willingness-to-pay" for provider networks from a multinomial logit choice model (Capps & Dranove, 2004; Ho, 2006; Shepard, 2016; Prager, 2016; Gowrisankaran, Nevo, & Town, 2015). I construct such measures for each of these specialties separately, and then use these measures to estimate consumer demand for health plans on the GIC. I find that physician networks explain a considerable portion of consumer valuations of overall plan networks, with plans that have larger physician networks attracting more consumers, even conditional on hospital network. On average, single-member households are willing to pay between $19 and $50 per month to move from a narrow network plan to a broad network plan, where approximately 80% of value comes specifically from the physician network. More-

[4] $https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2013_namcs_web_tables.pdf$

over, consumers are extremely loyal to their primary care physicians, with about 67% of consumers preferring to seek care from a physician they used previously. These results imply that estimates of consumer price sensitivity may be biased downward if physician networks are ignored. In other words, consumers often select into broad network plans, not because they are insensitive to price, but because they wish to keep their physicians. This, in turn, gives employers an incentive to offer narrow network plans that cut high-cost hospitals, but not necessarily high-cost physicians.

A second contribution is that I use the differences in choices made between new cohorts entering the GIC and existing members from prior years, along with the changing plan choice set for consumers across years, in order to isolate the effect of plan inertia from the willingness-to-pay for provider networks. The method for doing so is similar in spirit to previous work in Medicare Part D (Polyakova, 2016), the employer insurance market (Handel, 2013), and the individual insurance market (Shepard, 2016). The main difference in my analysis is that I provide dollarized switching cost measure in an employer setting with varying provider networks, as opposed varying levels of deductibles and cost-sharing. Indeed, I find that new consumers entering the GIC for the time make very different choices from those who have been on the group previously, opting to select into narrower network insurance plans at much higher rates. My model estimates that households, on average, would be willing to pay approximately $272 per month to remain on the same plan they were in the previous year, a considerable increase from existing estimates.

Finally, this paper is the first to endogenize *employer* choice of plans and provider networks to offer to their enrollees, and the first that allows firms to

adjust both the *number* of insurance plans offered as well as the provider networks of existing plans. This allows me to simulate the effect of policies not only on premiums and consumer switching patterns, but also on plan choice sets and the quality of insurance products. To do so, I specify an objective function for both insurers and employers in choosing a set of plans, assuming that each plan offered incurs a fixed cost of plan administration. I estimate the fixed costs of offering multiple plans using a bounded estimation, revealed-preference approach (Pakes, Porter, Ho, & Ishii, 2015), and subsequently used in the industrial organization literature for markets including computers (Eizenberg, 2014; Nosko, 2014), pharmaceuticals (Mohapatra & Chatterjee, 2015), and smartphones (Fan & Yang, 2016). I find that fixed costs for the GIC range from approximately $1.15 to $6.64 million dollars per year. Though quite larger in an absolute sense, these estimates are a small share of the GIC's overall spending on medical claims, and are within the range of published estimates reported by insurers. They explain why firms may not offer narrow network products, even when consumers would benefit from their inclusion in their choice sets. Indeed, my model predicts that without fixed costs, firms would benefit by offering narrow network plans much earlier than they currently do. This is consistent with results from previous literature showing that firms might improve consumer surplus by expanding choice sets (Dafny et al., 2013).

I then use these estimates to conduct several policy-relevant counterfactual exercises designed to incentivize employer groups to reduce health care expenditures. The first is I simulate firm choices of insurance plans under various assumptions on enrollee inertia. I find that a model of supply that assumes the GIC considers enrollee inertia as welfare-relevant when choosing products explains the products offered in the data quite well. For the most part, this

model rationalizes the GIC's choice on which insurers to offer, whether or not to offer narrow network plans, and the relative size of those networks. However, a model in which the employer ignores enrollee inertia would lead the GIC to drop all broad network plans and offer only the narrowest plans that satisfy Massachusetts state law, as the cost savings associated with these products exceed the welfare loss that would be seen from a strict price-versus-choice tradeoff.

I next simulate the effects of the proposed Cadillac Tax on expensive health plans. The actual tax, as proposed in the Affordable Care Act (ACA), was set to begin with its thresholds for individuals' annual premiums at $10,200 in 2018. Pizer et al. found that, assuming a 6% annual medical care inflation rate (which is on the conservative side by historical standards), the 2009 value of this threshold was approximately $5,845 for individual premiums (Pizer, Frakt, & Iezzoni, 2011).[5] I therefore consider the effects of a tax levied on premiums exceeding $6,000 annually (or an enrollee contribution of $125 per month) at 40% (the rate proposed by the law), as well as 20%, 60%, and 80%. I find that, in the short-term, a particularly large tax of 60% above this threshold induces the GIC to move to the equilibrium described in the "no inertia" scenario above. In addition to dropping each of its broad network plans in favor of narrow network products, the group also drops its overall number of products offered as the fixed costs of offering multiple plans exceed the surplus gained by retaining them. This results in an approximately a 21% reduction in total health care spending, or $76 pmpm. Consumer welfare from a strict price-versus-choice tradeoff (i.e. ignoring inertia or passivity) implies a welfare loss of $58 pmpm on average.

[5] http://theincidentaleconomist.com/wordpress/how-a-cadillac-tax-becomes-a-chevy-tax/

The implication of these results is that inducing employers to offer expanded choice sets including narrow network products would require very large price incentives to counteract substantial consumer inertia. Moreover, large fixed costs of offering additional plans imply that firms are unlikely to offer narrow network plans without also eliminating existing broad network ones. However, from a social welfare perspective that views consumer inertia as inattentiveness, such a tax may actually have potential to achieve social welfare gains of approximately $18 per member per month, despite the elimination of broad network products.

The paper proceeds as follows: Section 2 provides some background on provider consolidation in Massachusetts, as well as the increasing prevalence of insurer innovation, narrow networks, and policy proposals to address rising health care costs. Section 3 outlines the data and sample selection for the empirical analysis. Section 4 presents my model, estimation, and identification. Section 5 details the parameter estimates, as well as the model fit. Section 6 presents the results of the effect of the Cadillac Tax on employer product choice, medical spending, and consumer welfare. Section 7 concludes.

1.2 Background

1.2.1 Provider Consolidation and the Rise in Medical Care Prices

The Massachusetts health care market is one of the most prolific provider markets in the U.S. Much of this is due to the presence of a large number of elite providers, academic medical center, and provider practice groups. In recent years, many of these groups have been consolidating into larger health systems (including Partners healthcare, which owns many of the state's academic medical centers such as Massachusetts General Hospital and Brigham and Women's Hospital), large provider groups (such Harvard Vanguard, Action Medical Associates, and Atrius Health), and hospital-owned practices (such the MGH Physicians Organization). This consolidation has resulted in a rapid increase in the unit-price of health care, and has made Massachusetts not only one of the most prestigious, but one of the most expensive markets in the country.

Figures 1 shows the variation in the average prices paid to physician practices and hospitals in 2011. The y axis in both figures represents the average negotiated price paid to the particular practice or hospital relative to the mean price paid across all providers. There is considerable variation in prices paid in both figures, with the top providers able to negotiate rates that are approximately 40-50% higher than the average in the market. It is also notable that of the top 6 hospitals able to exercise this negotiating leverage, 4 of them are owned by the Partners health system. The Mass General Physicians Organization is also similarly able to charge higher prices, in part due to its affiliation

with Mass General Hospital and ownership by Partners.

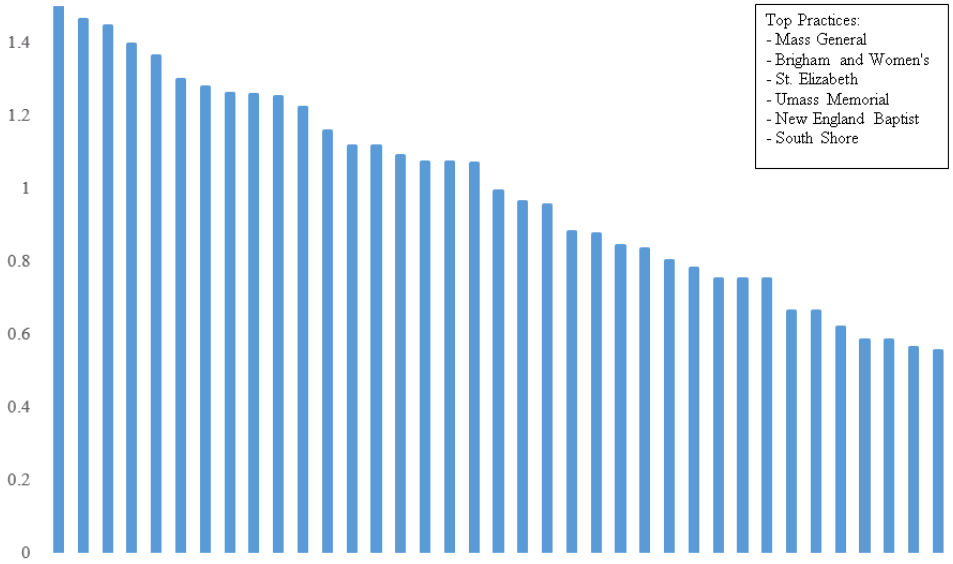
This consolidation is not just limited to Massachusetts, however. Across the country, hospital systems have been merging and have faced a litany of federal merger litigation (Gowrisankaran et al., 2015; Gaynor & Town, 2012; Town, Wholey, Feldman, & Burns, 2006). The effects of this hospital consolidation has been extensively studied, with the literature overwhelmingly pointing to notable increases in negotiated prices between hospital and insurers in markets with higher hospital concentration (Town, Wholey, Feldman, & Burns, 2008; Melnick & Keeler, 2007; Moriya, Vogt, & Gaynor, 2010) and following hospital mergers (Gowrisankaran et al., 2015; Capps & Dranove, 2004; Dafny, 2009; Town et al., 2006; Dafny, Ho, & Lee, 2016).

Similar to the hospital market, the physician market has also been consolidating in recent decades as solo-practitioners have either been abandoning fee-for-service models in order to become salaried employees of hospitals, or have otherwise been merging into large medical practice groups (Baker et al., 2014; Kleiner et al., 2015). However, despite the fact that physician services account for nearly the same proportion of national health expenditures as hospital spending,[6] there has been little research attempting to including physician networks into a model of demand for health insurance.[7]

Physician markets, however, differ in critical ways from hospital markets. While about 82% of Americans had contact with a health care professional in

[6]In 2014, hospital care accounted for approximately $972 billion (including outpatient care) while physician expenditures accounted for approximately $604 billion. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/highlights.pdf

[7]Some studies have looked, however, at the effect of physician consolidation on negotiated prices. See (Kleiner et al., 2015; Hausman & Lavetti, 2016; Baker et al., 2014; A. C. Dunn & Shapiro, 2014) for such evidence.



(a) Hospitals

(b) Medical Groups

Figure 1.1: Variation in Provider Prices per Visit

2012, only about 7.6% had an overnight inpatient stay at a hospital.[8] This likely makes physicians network an important component of consumer demand. Moreover, physicians interact with insurers in fundamentally different ways than do hospitals. While hospitals' reimbursements are determined based on the diagnosis of the patient, most physicians and practice groups are paid on a fee-for-service basis, and receive different negotiated amounts for each procedure performed. This gives physicians an incentive to overprescribe services, leading to overutilization of medical care, which is a significant driver of health care costs. When forming networks, insurers may therefore take into account not only the prices the practice groups would demand but also the average number of services per patient provided by the practice. Finally, insurers may have an interest in spanning a characteristic space of procedures and services within any given market. This means that oftentimes a smaller physician practice that performs a highly specialized service may be able to extract higher prices from insurers, despite attracting fewer patients than a larger practice that performs more general services.

1.2.2 Insurer Innovation and Narrow Networks

In response to these rapidly growing health care costs, insurance plans have increasingly turned to "value-based" insurance designs in an effort to steer patients towards lower cost providers. These designs include high-deductible health plans (HDHPs), narrow networks, tiered networks, reference pricing, accountable care organizations (ACOs), and others. The goal of such designs is generally to steer patients to lower cost health care providers by either forcing

[8] http://www.cdc.gov/nchs/fastats/physician-visits.htm

them to pay a higher out-of-pocket cost for care (HDHPs) or by actually prohibiting them from visiting the most expensive providers (narrow networks).

While the popularity of narrow network plans has been growing in the individual market and, in particular, in the health insurance exchanges (HIE), they have not grown as rapidly in the employer market. However, certain large employer groups and benefit managers have begun offering their employees increased choice of insurance plans and have pushed for the inclusion of more narrow network products. The Group Insurance Commission (GIC) in Massachusetts, the market I study in this paper, is one of those employer groups, and in recent years, it has aggressively encouraged participating insurers to design and offer narrow network products. Table 1 shows the plan offerings on the GIC by year. Throughout my sample period (2009 through 2013), six insurers participated on the GIC. In 2009 and 2010, four of those insurers offered narrow network products with varying degrees of network breadth. In 2011, both remaining insurers (Harvard Pilgrim and Tufts Health Plan) introduced narrow network products as well. These plans are approximately 20% cheaper on average than their respective broad network plans, though they are still fairly more broad than narrow network products offered in other market segments (such as Massachusetts' individual exchange, the Connector) by the same insurers.

Though the GIC has promoted the adoption of narrow network products, enrollment in these products was fairly limited in 2011 and health care spending continued to rise. As a result, in 2012, the GIC offered a three-month "premium holiday" for all active state employees who chose to switch to a narrow network plan. This premium holiday was fairly successful, resulting in approximately 10% of enrollees to switch (Gruber & McKnight, 2016). However, even with the

Table 1.1: GIC Products By Year

Insurer	Network	2009	2010	2011	2012	2013
Fallon	Narrow	x	x	x	x	x
Fallon	Broad	x	x	x	x	x
HPHC	Narrow			x	x	x
HPHC	Broad	x	x	x	x	x
HNE	Narrow	x	x	x	x	x
HNE	Broad					
NHP	Narrow	x	x	x	x	x
NHP	Broad					
Tufts	Narrow			x	x	x
Tufts	Broad	x	x	x	x	x
Unicare	Narrow	x	x	x	x	x
Unicare	Broad	x	x	x	x	x

Notes: GIC product offerings by year. "Broad" refers to whether the plan covers every provider in the state of Massachusetts and "narrow" refers to if the provider contracts selectively with providers.

switch, the GIC had to implement several additional changes in the following years to combat rising health care prices. In particular, in fiscal year 2015, the GIC converted Tufts Navigator and Harvard Pilgrim Independence (both broad network plans) from Preferred-Provider-Organization (PPO) plans to Point-of-Service (POS) plans, the latter requiring patients to select primary care providers and obtain referrals to specialists. They also raised deductibles for all plans by $50 per year, raised copays for high-tiered providers, and raised the enrollee share of premiums for employees hired prior to July 1, 2003 from 20% to 25%. In the subsequent year, the GIC completely closed Harvard Pilgrim's broad network plan to new enrollment, due to overwhelming increases in premiums from the plan. Executive Director Dolores Mitchell noted that "Harvard has been a disappointment for the past two years, not only on their own terms, but in comparison to other plans as well." She noted that over the past two years, Harvard

had requested a 9% increase in premiums each year.[9] This suggests that while narrow network products may be effective at reducing health spending, while offered alongside broad network plans, additional incentives might be necessary in order to stimulate cost reductions.

1.2.3 Public Policy and Cadillac Tax

In an effort to stimulate cost reductions among insurers and employers, a Cadillac Tax was passed as part of the ACA. The policy is an excise tax on high-cost health plans (plans with premiums in excess of $10,800 for individuals and $27,500 for families) aimed at the dual objective of raising government revenues and also encouraging plan sponsors to develop and switch to less costly health insurance. Although currently these thresholds are set quite high and would likely only affect unusually expensive plans offered primarily by unions and municipal governments (Claxton & Levitt, 2015), policymakers believe that over time the tax would apply to a greater variety of health insurance products, which would encourage insurers and employers to innovate more efficient, low-cost products (such as narrow network plans) (Gravelle, 2015).

According to proponents of the tax, the widespread use of expensive health insurance products by employers is in part due to the tax benefit received by employers for offering more expensive insurance plans. However, it is likely also due to the fact that purchasers of employer-sponsored tend to be older and less price-sensitive than purchasers in other markets, and that enrollee inertia in health insurance markets plans plays a critical role in driving demand. The tax could, in theory, correct for this by making it more expensive to offer

[9] http://www.capeplymouthbusiness.com/news/show/7420

such products, leading to a significant reduction in both health care spending (through the use of less expensive providers) and health care prices (through a bargaining effect that insurers would gain from enrollees switching to lower-cost plans).[10] Jason Furman, chair of the Council of Economic Advisers, wrote of the tax, "Many employers will probably focus on encouraging more efficient care delivery by deploying innovative payment models, directly complementing public-sector efforts, and finding creative ways to steer patients towards more efficient providers, investments that were often difficult to justify when the federal government was picking up much of the tab for inefficient care."[11]

Opponents, however, believe that the products that firms would switch to would lead to significant reductions in consumer welfare, with losses that exceed the reductions in health care spending. Narrow networks, for instance, are widely viewed as disruptive to patient care and as products that penalize primarily older and sicker individuals. In addition, many commentators believe that the tax would be primarily passed through to consumers in the form of higher premiums (through increased enrollee share contributions) or higher deductibles, rather than a move towards narrow networks or other forms of steering patients to more efficient providers. While certain forms of cost-sharing may indeed induce welfare gains through reduced negotiated prices (Prager, 2016), plans with higher deductibles often delay needed care rather than incentivizing consumers to be more effective shoppers of health care (Brot-Goldberg, Chandra, Handel, & Kolstad, 2015).

The GIC itself has commented on the impending Cadillac Tax: "There are many ways an employer can reduce exposure to the Cadillac tax, but there are

[10]The focus of this paper is on the former effect (i.e. steering to low-cost providers), though future work will consider the effect of the tax on negotiated rates between insurers and providers

[11]http://healthaffairs.org/blog/2016/04/25/about-that-cadillac-tax/

some troubling policy implications. Limited networks, which exclude higher cost facilities, are effective in reducing costs, but limited networks alone will not solve the problem of high health care costs. Increasing employee cost-sharing sharply reduces plan costs but leaves members exposed to financially ruinous bills."

The question, then, of how such a tax would affect social surplus is ambiguous. It depends on (a) whether a tax is likely to cause firms to switch to low-cost insurance plans rather than pass on the additional premiums to consumers and (b) how consumers value counterfactual plans and networks relative to the spending reductions they might achieve.

1.3 Data

I use two primary data sources to conduct the analyses in this paper: the Massachusetts All-Payer Claims Database (APCD) and the SK&A database of physicians. I now describe each of these datasets in detail.

1.3.1 Massachusetts All-Payer Claims Data

The APCD is a comprehensive database of medical claims from public and private payers in Massachusetts from 2009-2013. It contains detailed information on both hospital and physician visits, with variables indicating the patient's primary and secondary diagnoses (through ICD9 codes), procedures performed (CPT codes), patient demographics (including patient and provider 5-digit zip codes, which allow me to estimate the effect of distance on provider demand),

longitudinal patient identifiers, physician and facility identifiers, physician specialty, insurance and plan identifiers, and a wide variety of payment variables. Importantly, these payment variables contain not only the amount paid by the insurer and the out-of-pocket amounts paid by the patient for the medical service, but also the "allowed amount." This variable refers to the maximum allowable payment the insurer can make to a provider for any particular service. In other words, it is the negotiated rate between an insurance company and either a physician or a hospital. Most hospital admission data contain only variables depicting "charges," or what the hospital's list price is for a particular illness. However, these are rarely the prices that are actually paid, and therefore are an inaccurate representation of an insurer's marginal costs. By observing the allowed amounts, the APCD affords me the opportunity to more precisely depict what insurers pay each provider, and therefore how insurer costs might change under counterfactual networks.

I use the APCD to construct several different subsamples pertaining to different stages of the model that I then estimate.

Hospital Admissions

The first sample is the sample of hospital admissions, which I use to estimate the patient demand for hospitals, described in more detail in the next section. To construct this data, I limit the APCD to any facility claim flagged as an inpatient admission between the five-year sample period and to any hospital that is located within the state of Massachusetts. I therefore exclude any admission of patients receiving hospital care outside the state (regardless of whether the patient resides in Massachusetts or not). For each hospital, I used the organiza-

tion's National Provider Identification (NPI) number to match the hospital to a set of hospital characteristics from the American Hospital Association (AHA) Annual Survey. These characteristics include the type of hospital (teaching, critical-access, academic medical center, specialty, etc.) and hospital amenities (including number of beds and types of services offered). The data is aggregated to the hospital admission level, and the "allowed amounts" are summed over all service-lines for that particular admission, in order to construct a price-per-visit. For each admission, I link the primary diagnosis (ICD-9 code) to a set of Chronic Conditions Indicators (CCI) and Clinical Classifications Software (CCS) categories. These are indicators provided by the Agency for Healthcare Research and Quality (AHRQ) that allow me to aggregate diagnosis codes into a set of 18 distinct groups, and also to flag which patients suffer from chronic conditions.

Table 2 contains the hospital sample summary statistics for hospital admissions from 2009-2013. On average, patients admitted to Massachusetts hospitals are 52 years old, and about half of the patients suffer from a chronic condition. Approximately 16% of patients are admitted with a primary cardiac condition, while about 22% are admitted with an obstetrics-related diagnosis. Patients are, on average, willing to travel approximately 10 miles to visit a hospital, and visit teaching hospitals approximately 74% of the time, while visiting academic medical centers approximately one-quarter of the time.

Physician Visits

The second constructed sample from the APCD is used to estimate the physician demand portion of the model. I construct it by limiting the data to pro-

Table 1.2: Hospital Sample Summary Statistics

	Mean	Std Dev
Patient Characteristics		
Age	52.14	25.98
Female	0.58	0.49
Chronic	0.53	0.49
Neurological	0.02	0.15
Cardiac	0.16	0.37
Obstetrics	0.22	0.42
Imaging	0.27	0.44
Hospital Characteristics		
Distance	9.95	12.06
NICU	0.87	0.33
Neuro	0.96	0.19
MRI	0.90	0.30
Critcal Access	0.01	0.08
Teaching	0.74	0.44
Specialty	0.02	0.14
Academic Medical Center	0.25	0.43
Would Recommend	0.74	0.12

Notes: Hospital sample summary statistics.

fessional claims only. These capture reimbursements specifically to medical providers that are separate from reimbursements for facilities, even though the particular service may have been performed in a facility. This includes patient visits to independent offices, larger medical groups, or non-inpatient visits to hospitals, outpatient centers, or clinics within hospitals such that a separate claim is generated to pay individual physicians. The data is then merged with SK&A data on physician affiliations (described in more detail below), and each individual practitioner is assigned to their primary medical group. After constructing these practice groups, I then stratify the data into three different specialty groups: primary care physicians, cardiologists, and orthopedists. Primary care practices are defined as any medical group that contains at least one physician that is either an internist, general practitioner, family practice doctor,

pediatrician, or geriatric doctor. Similarly, cardiology practices and orthopedic practices are defined as any practice that employs at least one physician of the relevant specialty. I consider these three specialties in order to capture three different component of medical care: primary care, which is the most common type of visit to a health care provider (at about 55% of all office visits), medical specialty care primarily performed in office-based settings (exemplified by cardiology), and surgical care primarily performed in facility-based settings (exemplified by orthopedics).

For each service-line, I merge in Medicare Part B physician fee schedules from CMS. These data contain annual federal updates to each procedure code's "Relative-Value-Unit" (RVU) weight. These weights are constructed in order to assign each service an approximate measure capturing its relative intensity to other procedures. They are subsequently used to determine Medicare payment rates (described in more detail in the next section), and are also an input in determining private payer negotiated rates for physician care. I then finally aggregate the data to the patient-visit level, summing over all the RVU weights of each service provided during a visit and summing over all the "allowed amounts" for each service to determine a total payment per visit.

Table 3 shows summary statistics for the physician samples. On average patients going to see primary care physicians (PCPs) are younger and have a higher likelihood of being female than those going to specialists or surgeons. This is most likely due to the fact that OB/GYNs are considered primary care practitioners. Average RVUs for orthopedic services are higher than for PCPs and cardiologists, with significantly higher standard deviations. This reflects the fact that while orthopedists often perform routine office-based procedures,

they also perform surgeries which are more resource intensive and thus are assigned higher RVUs. About 67% of primary care patients saw a doctor between 2009 and 2013 that they also saw in the preceding year, while this number was about 38% for cardiologists and about 30% for orthopedists. Distance traveled to any of the specialty groups are all about 10 miles. When seeing a PCP or orthopedist, patients on average tend to go to sites with approximately 50 doctors, whereas cardiologists tend to form into larger groups with an average of 153 doctors at a site. This likely reflects the fact that patients tend to see cardiologists not in office-based settings, but in hospital settings, even if those cardiologists submit professional claims to insurers (i.e. non-facility claims).

Table 1.3: Physician Sample Summary Statistics

	PCPs	Cardiologists	Orthopedists
Age	34.32	52.81	45.85
	(22.80)	(15.59)	(17.52)
Female	0.58	0.41	0.53
	(0.49)	(0.49)	(0.50)
RVU	2.07	2.17	3.82
	(2.37)	(3.53)	(6.37)
Used Doc Last Year	0.67	0.38	0.30
	(0.47)	(0.49)	(0.46)
Distance	8.50	10.63	9.57
	(9.49)	(12.81)	(10.19)
Medical Group	0.84	0.67	0.62
	(0.37)	(0.47)	(0.48)
Doctors on Site	54.10	153.57	56.90
	(105.37)	(229.40)	(148.02)
Number of Locations	23.45	24.83	15.34
	(32.81)	(37.26)	(34.02)
Total Doctors	187.92	160.78	103.36
	(245.54)	(230.42)	(208.19)
Share Specialty	0.70	0.40	0.66
	(0.26)	(0.40)	(0.38)

Physician sample summary statistics for primary care physicians, cardiologists, and orthopedic surgeons.

GIC Member Sample

The final subsample constructed is a sample of GIC members by year, which is used to estimate the insurance demand portion of the model. In addition to claims data, the APCD contains an enrollment file, where each insurer provides a list of each of its enrollees by market, plan, and year. These files also come with a rich set of enrollee demographics, including 5-digit zip code, age, gender, employer industry code, employer zip code, monthly plan premium, annual plan individual and family deductible, enrollment start date, and enrollment end date. I limit this file to all enrollees who are part of the GIC between 2009 and 2013. The file also allows me to link individual enrollees to their family members when estimating insurance demand. Finally, I merge this list of GIC members to external data on GIC annual plan premiums and hospital networks. An advantage of studying this particular market is that plan premiums are the same for each member across the state, and only vary by family type ("Individual" versus "Family"). Each year, the GIC publishes these premium rates for each family type. It also publishes an annual list of the hospitals included in each plan's network for each of the commission's narrow-network plans. I merge this public information onto the enrollee dataset in order to obtain a full set of plan characteristics for each enrollee. For the year 2012, the year of the premium holiday, I assume that each active employee under the age of 65 pays only 9 of the 12 months of the annual premium if they switch to a narrow network plan in that year.

Table 4 shows the market shares and premiums for all the plans offered on the GIC in 2012, the year after Harvard and Tufts both introduced narrow network products. This also coincides with the first year of the premium holiday.

The most expensive plans on the market are Unicare's Indemnity plan, as well as Harvard Independence (broad network) and Harvard Primary Choice (narrow network). Each insurer's narrow network plans are priced an average of 20% below their own broad plans. The broad plans have the highest market shares, with Tufts and Harvard each making up about 25%-30% of the market. Their narrow plans, however, had much more limited enrollment in 2012, with about 5% for Harvard Primary Choice and 2% for Tufts Spirit. This is up from 2% and 1%, respectively, in 2011, due in large part to the premium holiday inducing members to switch to these narrow plans. One interesting note is that despite having lower out-of-pocket premiums, Tufts Spirit had a significantly lower market share than Harvard Primary Choice.[12] This is a point that I will return to below.

Table 1.4: GIC Summary Statistics, 2012

Insurer	Network Coverage	Market Share	Individual Premium ($PMPM)
Fallon Select	Broad	0.03	139.39
Fallon Direct	Narrow	0.02	112.97
Harvard Independence	Broad	0.21	163.98
Harvard Primary Choice	Narrow	0.05	131.50
Health New England	Narrow	0.06	110.34
Neighborhood Health Plan	Broad	0.02	113.02
Tufts Navigator	Broad	0.27	148.43
Tufts Spirit	Narrow	0.02	119.06
Unicare Indemnity	Broad	0.13	247.07
Unicare Plus	Broad	0.08	207.27
Unicare Community Choice	Narrow	0.10	111.61
Number of Enrollees in GIC	293,125		
Average Age	36.07		
Average Subscriber Age	48.04		

GIC plans for 2012. Average premiums refer to the enrollee share of the per-member-per-month premiums (25% of the overall premium).

There is also significant heterogeneity in terms of who is enrolling in narrow network plans. Figure 2 depicts the share of GIC consumers enrolling in nar-

[12]Though 5% versus 2% market share seems low, this represents a difference in almost 12,000 members

row network plans by year and by whether they were new to the GIC that year (i.e. "active choosers") or whether they were existing GIC members who are automatically re-enrolled in their current plan unless they take action (i.e. "passive choosers"). In 2010, the share of enrollment of active choosers and passive choosers in narrow network plans both hovered around 10%. In 2011, once the GIC introduced Harvard Primary Choice and Tufts Spirit narrow network products, enrollment in narrow network plans remained about 10% for those who had been enrolled in 2010, but spiked to about 30% among new GIC members. This suggests that the high enrollment in broad networks is in driven in part by inertia. In 2012, the share of active choosers on narrow networks remained the same as in 2011, but the share of passive choosers enrolling in narrow network plans rose due to the implementation of the premium holiday. As this was geared directly at passive choosers, it makes sense that the three-month discount incentivized these members to switch towards the narrow network plans. I use these sources of variation (within-year across-individual and across-year-within-individual) to help identify consumers' valuation of broad networks and sensitivity to price separately from that of inertia or inattentiveness.

In addition to observing the choices of new members versus existing members each year, the plan choice set for all enrollees changed considerably over time. Between 2009 and 2010, the GIC converted both the Tufts and Harvard broad network plans to "tiered network" systems, which increased the copayments for members choosing expensive physicians and hospitals. The effects of these tiering mechanisms on negotiated prices is explored in detail by Prager (2016). In 2011, two new narrow network plans entered the market. In 2012, the GIC introduced the premium holiday, which significantly discounted narrow network plans relative to broad network plans. Within narrow network

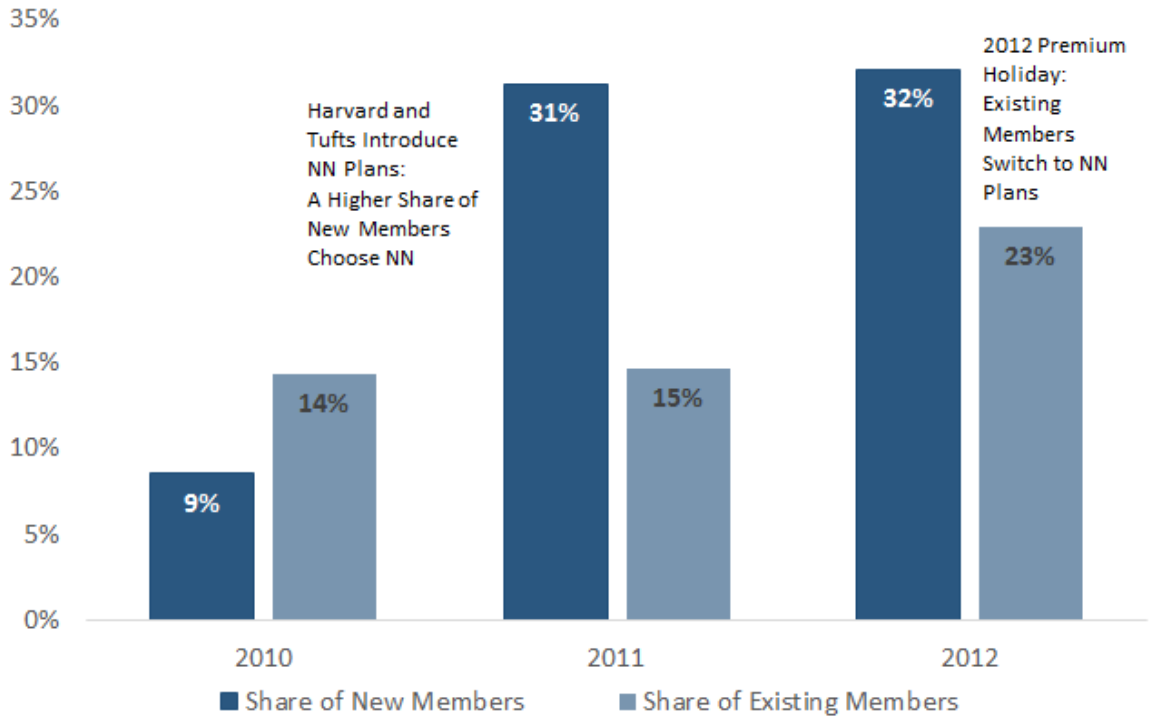


Figure 1.2: Share of People in Narrow Network Plans by Year and Whether New to GIC

products, the set of doctors and hospitals changed year over year.

The final change in choice set occurred in premiums between broad network plans. In 2010, the premiums for Harvard and Tufts were fairly similar, while beginning in 2011, the premium difference between the two plans began to rise. Figure 3 shows the change in market share of enrollees in the Tufts broad network plan over time by whether the enrollee was a new member to the GIC or an existing member from a prior year. The lines report the family premiums for the Harvard and Tufts broad network plans. It is notable that as Harvard's premiums rise relative to Tufts' premiums, enrollment in Tufts rises dramatically among new members to the GIC. By 2013, as the premium difference between Harvard and Tufts reached about $30 per month, Tufts' market share among

new members increased to about 48%, significantly exceeding that of Harvard's 21%. Existing members, however, exhibit no such changes in enrollment patterns. Between 2010 and 2011, Tufts' market share among existing members falls, due primarily to the fact that Harvard saw a major spike in enrollment among new members in 2010 (when premiums for each plan were nearly identical). Thereafter, Tufts' market share among existing members barely budged, even as the premium differences widened.

Taken together, these two figures provide some suggestive evidence that new enrollees react more significantly to change in prices and choice sets than existing members, and that consumers do exhibit a significant degree of price sensitivity in these employer markets. These empirical facts motivate my inclusion of inertia in the model, described in the next section.

1.3.2 SK&A

In order to link physicians to their practices, I use proprietary data from the SK&A database for 2009 and 2013. The database includes information on each individual physician's name, location, specialty, NPI, affiliated medical group, affiliated hospital, and affiliated health system. It also contains characteristics for the site of the physician practice, including number of physicians on staff, the specialty of the practice, and the number of physicians on staff across all the locations of the particular medical group. The SK&A includes approximately 95% of all office-based physicians practicing in the United States, and the data is verified by the proprietors over the telephone.

Given the breadth of the data as well as the inconsistencies in reporting be-

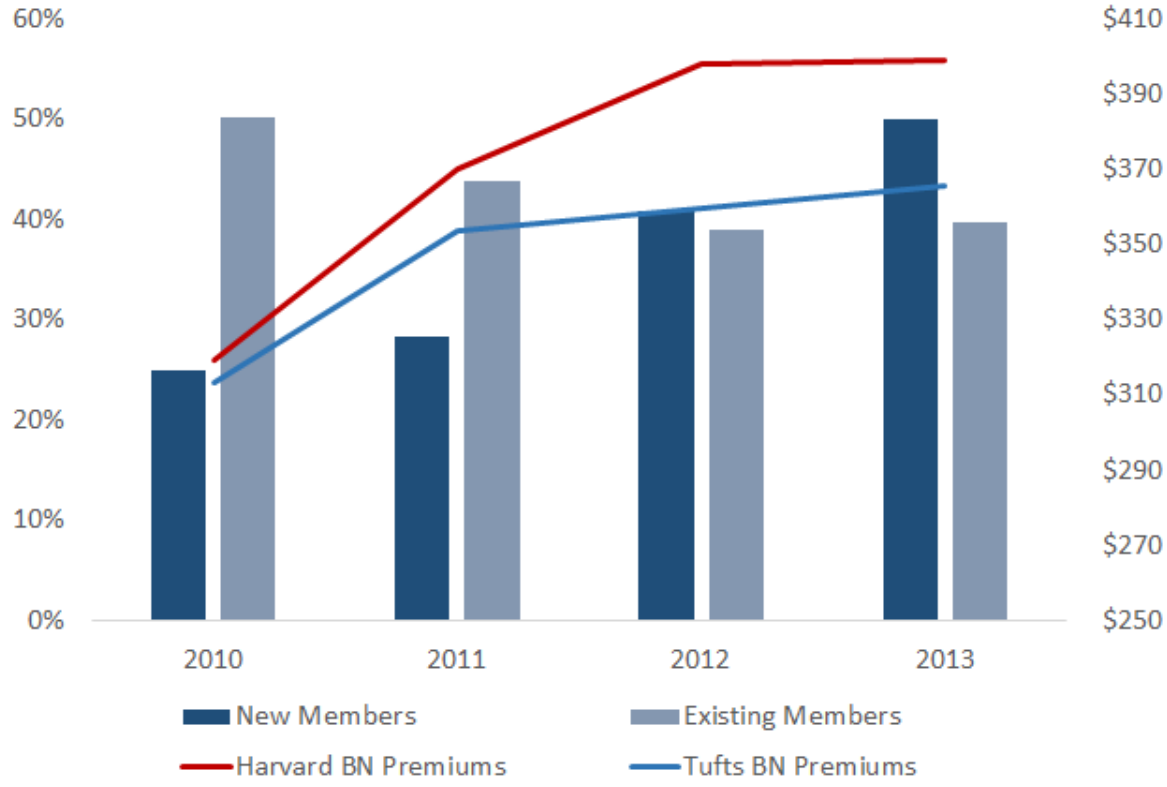


Figure 1.3: Share of Members Enrolling in Tufts Broad Network Plan by Whether New to GIC

tween the APCD and SK&A, linking the two datasets involved several steps. First, I matched every available physician in the SK&A to the APCD via the NPI variable and provider zip-code variables in each dataset. This ensures that all the matches were not only to the correct physician, but also to the correct practice location for each physician. In cases where this did not match, I then matched only by the NPI and assumed that the closest location in the SK&A to that where the service was rendered in the APCD was the correct practice.

However, not all insurers in the APCD report physician NPIs, opting instead to bill using the organizational NPI. For instance, Health New England only reports the NPI for the hospital or medical group when processing claims. Given

that the SK&A only contains individual doctors' NPIs, in instances where this occurs, I conduct an iterative string-matching algorithm to match the medical practice data. I use the first and last name fields in the APCD and match the provider's names and zip codes to the names and zip codes from the SK&A. For all records that did not match, I then match only by first and last name. Then I repeat this just for last name and zip code. These set of steps allowed me to match approximately 80% of the claims from the APCD to an appropriate physician from the SK&A.

After completing this procedure, I define two different variables. The first is a "practice" variable, which is the unit used in the demand analysis. This variable refers to any particular physician-medical group-location triple in the data that billed more than 50 claims in any particular year. If a medical group was not reported in the SK&A or if the particular physician did not work for a medical group, I consider the physician-hospital-location triple as the practice definition. These are physicians who are employed by hospitals but may be billed for physician services separately (for example if they take outpatient or office visits in the hospital clinic). If there is no medical group or hospital reported, I consider this variable to be just the physician-location double, and assume the physician is a solo-practitioner. I assume that when selecting a physician, individuals choose at this "practice" level.

The second variable I define is an "ownership" variable, which is used in defining networks. This refers to the highest level of vertical integration for the physician. If a particular physician's highest reported ownership in the SK&A is a medical group, then this variable is coded as the group. If the highest level of ownership, is a particular hospital (i.e. a hospital owned physician practice),

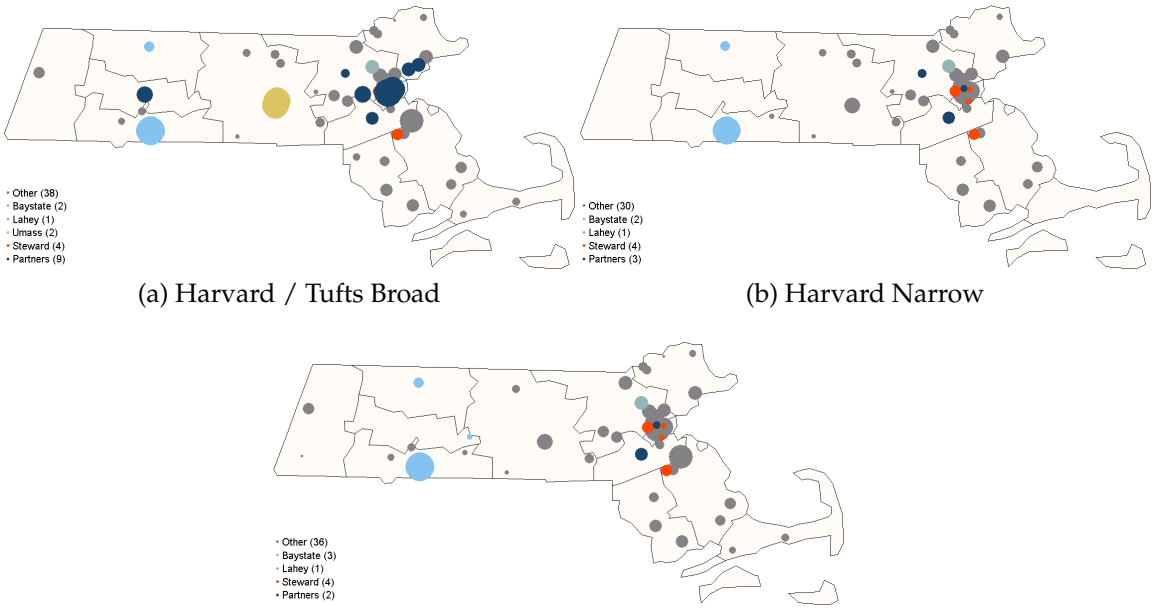
then this variable is coded as the hospital. Finally, if the highest level of ownership is reported as a health system (i.e. Partners Health Care, Steward Health System), then this variable is coded as the system. In considering counterfactual networks that the GIC could offer, I make the assumption that the insurers contract at the "ownership" level. Therefore, if the GIC chooses to eliminate a Partners physician, it must eliminate all physicians employed by the Partners health system.

I then assign each physician a specialty according to the specialty reported in either the APCD or the SK&A. For example, if a particular physician is reported as a cardiologist in either dataset, I flag that physician as a cardiologist. I consider any practice a cardiology practice if it employs at least one physician flagged as a cardiologist, or if the SK&A reports that the practice is a cardiology practice. I limit the sample to the 100 largest physician practices in each specialty (primary care, cardiology, and orthopedics) in terms of number of claims submitted in a given year. This reflects approximately 50% of primary care claims in a given year, and approximately 85% of both cardiology and orthopedic claims. The remaining practices are considered part of the outside option.

The final task involves determining which physician practices are in a particular insurance plan's network. While some GIC insurers actually report the medical groups that they cover in their narrow networks (i.e. Fallon), others only report the list of hospitals. I therefore assume for simplicity that if a particular hospital is excluded from a particular plan's network, then any physician, physician practice, or medical group that is owned by that particular hospital is also excluded from the network. Similarly, as bargaining between insurers and providers is typically done as the *system* level, I assume that if any particular

system is excluded from a plan's network in its entirety (e.g. if a particular plan excluded all Partners hospitals), I assume that any physicians or groups that are owned by Partners (even though they may not be affiliated with any particular hospital) are also excluded. For any large medical group that is not affiliated with a particular hospital or system (e.g. Atrius Health, Reliant Medical Group, etc.), I conduct manual checks on the insurers' websites to see whether these groups are covered by the plans. For all remaining practices, if they are not owned by any hospital or system, I assume they are in each plan's network unless a majority of claims that are processed for these practices by a particular plan is flagged as being "out of network."

Figures 4 and 5 shows the hospital and primary care practice networks for a select group of products available on the GIC in 2011: Harvard Independence (broad network), Harvard Primary Choice (narrow network), and Tufts Spirit (narrow network). The colors of the points on the maps refer to physician practices that are owned by the largest health systems in Massachusetts: Partners, Steward, Atrius, Umass, Lahey, Baystate, and all other practices. The sizes of the points are in proportion to total market share of the practice for the particular physician specialty. Looking at primary care practices, it is clear that Partners and Atrius Health dominate much of the primary care physicians in Massachusetts, which Partners owning 172 practices and Atrius owning approximately 51. Panel (a) of Figure 5 shows that these practices are largely concentrated in eastern Massachusetts, particularly around Boston and the surrounding suburbs. However, Atrius Health also owns practices in central Massachusetts. This is the result of Atrius having purchased Reliant Medical Group, a large physician practice, in 2011.



(a) Harvard / Tufts Broad

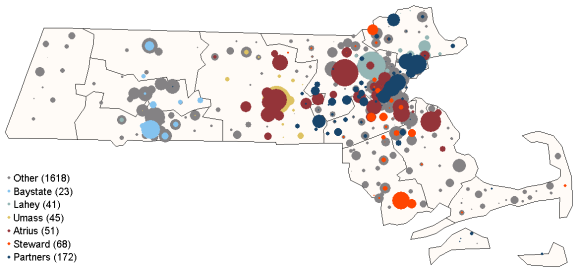
(b) Harvard Narrow

(c) Tufts Narrow

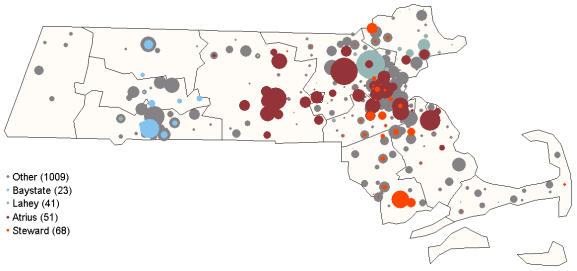
Figure 1.4: Hospital Networks by Plan, 2011

Panels (b) and (c) of Figures 4 and 5 reveals that the Harvard and Tufts narrow network plans still cover a large number of hospitals and physicians in Massachusetts. Interestingly, the hospital networks of both narrow plans are relatively similar. The only major difference in hospitals was the dropping of most Partners hospitals from each plan. However, as noted in Table 4, Harvard's narrow plan has a significantly higher market share than the Tufts narrow network, with almost three times the number of enrollees in 2012. Given that Tufts covers a larger number of hospitals, it is therefore unlikely that hospital networks explain this discrepancy in market shares.

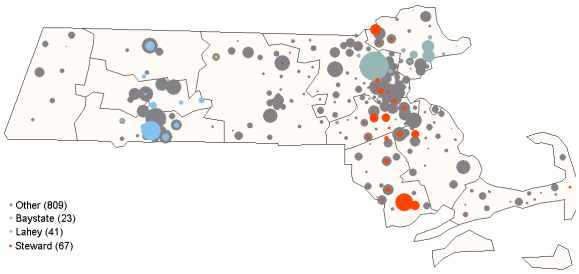
Turning to physician networks, however, provides more clues that might help to explain these plan choices. Figure 5 reveals clearly that the Harvard narrow physician network is considerably more comprehensive than the Tufts narrow physician network. This is largely due to the fact that Harvard cov-



(a) Harvard / Tufts Broad



(b) Harvard Narrow



(c) Tufts Narrow

Figure 1.5: Primary Care Practice Networks by Plan, 2011

ers Atrius health, which owns Harvard Vanguard Medical Group in Boston. This demonstrates that physician networks may be an important determinant of plan choice. Moreover, given that Partners physicians were primarily located in the Boston metro area, which also faces competition from Atrius, Lahey, Care Group, as well as many independent and solo practitioners, its removal from the network has minimal impact for choice of provider (as is shown later in the model).

Appendix A.1 shows additional networks from Fallon Direct and Health New England, both of which are considerably narrower than Harvard and Tufts. The Appendix also shows additional maps for cardiology and orthopedic practice networks.

1.4 Model

The model proceeds in four stages. A brief summary of these stages is as follows:

1. Employers select a number of products to offer to their enrollees and the network design of the plan. In selecting these plans, employers incur a fixed cost of adding each additional product.
2. Given the products selected, employers set premiums for self-insured products. Insurers set premiums for fully-insured products.
3. Consumers in each market select from the menu of insurance plans given their network breadth and composition, premiums, and various quality characteristics.
4. Consumers face some probability of contracting an illness, and based on that illness, along with individual and provider characteristics, patients select a hospital or doctor from one among their chosen insurance plan's network.

I now describe the model in detail from the latest stage through the earliest stage.

1.4.1 Patient Demand for Providers

The final stage of the model involves patient i enrolled in insurance plan j choosing a provider. The patient either has a condition that requires hospital care, l, in which case he or she chooses a hospital h from among the set of

hospitals in insurance network N_{jH}, or the patient requires procedure r from specialist type s, in which case he or she chooses physician practice d among a set of practices within that specialty within the plan's network N_{jS}. Consumer utility for patient of type i, with either illness l or procedure r, from visiting a provider takes the following form:

$$u_{ilh} = \underbrace{T_{ih}\lambda_1 + T_{ih}v_{il}\lambda_2 + T_{ih}x_h\lambda_3 + x_h v_{il}\lambda_4 + \mathbb{1}_{ih_t=ih_{t<1}}\lambda_5 + \gamma_h}_{\phi_{ilh} \text{ (Hospitals)}} + \epsilon_{ilh} \tag{1.1}$$

$$u_{irds} = \underbrace{T_{ids}\lambda_1 + T_{ids}v_{ir}\lambda_2 + T_{ids}x_d\lambda_3 + x_{ds}v_{ir}\lambda_4 + \mathbb{1}_{ids_t=ids_{t<1}}\lambda_5 + \gamma_{ds}}_{\phi_{irds} \text{ (Physician Specialty } s)} + \epsilon_{irds} \tag{1.2}$$

where x_h is a vector of observed hospital characteristics, x_{ds} is a vector of observed physician practice characteristics for specialty type s, v_{il} and v_{ir} are observed characteristics of patient i with diagnosis l or requiring procedure r, T_{ids} and T_{ih} is the distance in miles from patient i's location to provider d or h's location, γ_{ds} and γ_h are provider fixed effects, and ϵ are Type 1 Extreme Value error terms. Finally, $\mathbb{1}_{ih_t=ih_{t<1}}$ refers to whether patient i has used hospital h in any year prior to t, and $\mathbb{1}_{ids_t=ids_{t<1}}$ refers to whether individual i saw physician practice d for specialty care s in any year prior to t.

The patient characteristics include 5-digit zip code, age, an indicator for female, patient diagnosis (in the case of hospital care), patient procedure required (in case of physician care), and whether the patient has ever been treated for a chronic condition.

For hospital care, patient diagnoses, l, are grouped into 18 Clinical Classification Software (CCS) categories. Chronic conditions are grouped according to HCUP indicators mapping chronic conditions from ICD9 diagnosis codes.

Given that my data spans 2009-2013, I define patient i in time t as having a chronic condition if that patient has gone to see any provider at any time prior to t for a diagnosis that is considered to be "chronic." Each of the 18 diagnosis categories are further assigned numerical weights that proxy for the intensity of the particular diagnosis (the construction of these weights follow closely to work by Shepard; a discussion of their construction follow in Section 4.4.1). Hospital characteristics include location, location, number of beds, whether the hospital had a NICU, whether the hospital provided imaging services (including an MRI), and whether the hospital included a catheterization lab. I include indicators for whether the hospital is a critical access hospital, a teaching hospital, a specialty hospital (such as cancer center or children's hospital), or whether the hospital is an academic medical center. I further interact these hospital characteristics with each of the 18 disease categories. In addition, I include a full set of hospital fixed effects in the model to account for any unobserved quality components of hospitals not captured by the model. In order to capture additional heterogeneity, I interact these fixed effects with the numerical weights for the patient diagnoses, in effect allowing patients with different disease severities to prefer seeking care from different hospitals.

For patients requiring care from physician, I match procedure (CPT code) to a Medicare RVU weight, r, which serves as a proxy for procedure intensity. As described in Section 3.1.2, these weights are used by Medicare to calculate payments for physicians under the Part B schedule, and thus serve as an adequate measure of the intensity of service that a patient might require. For physician practice characteristics, I include a number of variables from the SK&A including the number of doctors at the particular practice's location, the number doctors across *all* the practice's locations, the share of the doctors at the prac-

tice who are specialists (relative to primary-care physicians), and the number of unique procedures performed at the practice. I interact each of these with patient characteristics, including the patient's RVU weight. I also include a full set of practice fixed effects within each specialty group, and interact those fixed effects with RVU weights.

The probability that patient i and diagnosis l will choose hospital h is thus given by:

$$\sigma_{ilh} = \frac{exp(\phi_{ilh})}{\sum_{k=1}^{N_{iH}} exp(\phi_{ilk})} \tag{1.3}$$

where N_{iH} refers to the number of hospitals in individual i's network. Similarly, the probability that patient i needing a procedure with RVU r from specialist group s will chose physician practice d is:

$$\sigma_{irds} = \frac{exp(\phi_{irds})}{\sum_{k=1}^{N_{iS}} exp(\phi_{irks})} \tag{1.4}$$

where N_{iS} is the network of practices of type s in individual i's network.

Estimation and Identification

The patient choice of providers is estimated using maximum likelihood. The model includes patient characteristics interacted with provider characteristics, travel time interacted with both patient and provider characteristics, and a full set of provider fixed effects (interacted with diagnosis/procedure intensity

weights) in order to account for unobserved heterogeneity across the providers in the data.

I run the model separately for hospitals and physician practices of various specialties. In particular, I run the model separately for primary care practitioners, cardiologists, and orthopedists. These all can be thought of as separate markets that do not compete with one another. For instance, patients who require a procedure for knee surgery would be unlikely to select a cardiology practice for that procedure. One limitation of this approach is that it abstracts away from referral networks across specialties and between physician groups and hospitals. Indeed, patients often seek care initially from their primary care physicians, who may subsequently refer them to a cardiologist or orthopedist. My model, by treating these specialty groups as independent, does not capture these behaviors. This may bias the parameter estimates, particularly in the hospital and specialist models (unlikely, however, in the primary care model) as choice may be driven not by, say, distance, but by the recommendation of a previously used provider. Future work aims to quantify these physician referral networks, and to see how these drive demand for different specialties.

Another problem is that there are thousands of physician practices within each specialty group in Massachusetts. To reduce the dimensionality of the problem, I narrow the scope of the demand model to the top 100 practices within each group (in terms of the market share). All other physician practices are considered part of the outside option. This not only has the benefit of making the model more easily estimable, but also is indicative of the fact that most physician practices outside the ones with the highest market shares are generally included in all plans' networks (including the narrow network plans).

Therefore, estimating how consumers will react to an insurance plan dropping a particular physician practice is the most relevant for the largest practices. For hospitals, I define the outside option to be any hospital outside of the state of Massachusetts. I normalize the value of the outside option in each demand model to be 0.

Each of the coefficients are identified through within-provider variation in patient characteristics. The parameter on distance, for example, is identified by differences in choice of a particular provider across patients who live in different zip-codes throughout Massachusetts. The identifying assumption is that patient choice of where to live is orthogonal to their preferences for providers.

1.4.2 Expected Utility from the Provider Network

The coefficients estimated from the provider demand model are then used to construct an expected utility variable, which is used as an input to the next stage in the model. This variable measures patient i's ex-ante willingness-to-pay for a particular insurance plan's provider network. The importance of this variable is that it provides a measure of network breadth that is not only based on the size of the plan's network, but on the relative quality of the providers in the network. A network may, for instance, be smaller in size, but still include many of the high-demand providers in the market in which the plan is operating.

As demand for insurance plans is at the *household* level (i.e., households choose one plan for all individuals in the family), I aggregate the expected utility variable to the household level by summing over each individual i's willingness-to-pay for the provider networks.

Specifically, household I's expected utility from the network offered by plan j in time t is given by:

$$EU_{Ijt} = \sum_{i \in I} \left(\underbrace{\sum_{l} f_{il} log \left(\sum_{h \in N_{jtH}} exp(\phi_{ilh}) \right)}_{EU_{ijtH}} + \underbrace{\sum_{s} \sum_{r} f_{ir} log \left(\sum_{d \in N_{jtS}} exp(\phi_{irds}) \right)}_{EU_{ijtS}} \right) \tag{1.5}$$

Despite the fact that provider demand was estimated separately for hospitals, primary care practices, cardiology practices, and orthopedic practices, the expected utility variable defined above aggregates the utilities from each of these different types of providers, with EU_{ijtH} representing consumer i's utility from the *hospital* network of plan j in time t, and EU_{ijtS} representing consumer i's utility from the *physician* network of plan j at time t, and this is summed over all specialists s. This allows individuals to have different valuations of different types of providers, depending on their probability of contracting various illnesses or requiring certain procedures.

Here, f_{il} is the probability that individual i contracts diagnosis l (requiring hospital care) and the frequency with which that individual needs to seek care for that illness. f_{ir} is the ex-ante probability that individual i requires procedure r (i.e. requires physician care). For the purposes of calculating these probabilities, individuals are grouped into distinct age-sex-chronic condition categories, with the following age bins: 0-19, 20-29, 30-39, 40-49, 50-64, 65+. f_{il} and f_{ir} are estimated directly from the claims data by averaging over the share of all GIC members of type i who sought medical treatment for diagnosis l or procedure r. For hospitalizations, diagnoses were grouped into the 18 CCS cate-

gories used in the demand estimation. For seeking physician care, diagnoses were grouped first into the probability of requiring care from a cardiology, orthopedist, and primary care practitioner, and were subsequently grouped into bins of RVU weights: 0-1; 1-2; 2-5; 5-10; 10-20; 20-40; 40+. This reflects the fact that individuals of different ages, genders, and medical histories have differing probabilities not only of needing to see certain specialists, but also of requiring treatment of varying levels of complexities.

A more robust model would specify the probability of requiring more specific procedures, rather than the probability of requiring a certain RVU-weight. Indeed, the probability of requiring knee surgery may be different than the probability of requiring shoulder surgery. However, given the number of procedures that any given specialists treats, this would present a significant computational burden. Grouping procedures into specialty-RVU categories is therefore a simplification towards computing ex-ante probabilities of valuing an insurer's provider network.[13]

1.4.3 Consumer Demand for Insurance Plans

The utility of household I for plan j at time t is given by the following:

$$u_{Ijt} = \underbrace{-r_{Ijt}\alpha_I + EU_{Ijt}\beta_1 + \mathbb{1}_{Ij_t=Ij_{t-1}}\beta_2 + \eta_j}_{\delta_{Ijt}} + \omega_{hjt} \tag{1.6}$$

Here, r_{Ijt} refers to the plan rate, or premium, which varies only by whether

[13] An alternative would be to explore models where patients choose their physicians based on their diagnoses, rather than procedures required. These models will be explored in future iterations of the paper.

the consumer has purchased individual coverage or family coverage. I allow the premium coefficient, α_I, to vary by age of the oldest member of the household. This is to reflect the fact that households with members of different age groups may react differently to insurance plan prices than other households. EU_{Ijt} is the expected utility from the plan's network as defined above, which also varies by household. η_j is the unobserved plan characteristics component, captured by a full set of plan fixed effects, reflecting the fact that plan demand may be driven by preferences for a particular plan unobserved by the econometrician, and ω_{Ijt} is the idiosyncratic, Type 1 Extreme Value error. Plan inertia is captured by $\mathbb{1}_{Ijt=Ij_{t-1}}$, which is an indicator function for whether household I was enrolled in plan j in year $t-1$.

The market share of households of type I for plan j in market t is derived as the familiar logit share:

$$s_{Ijt} = \frac{exp(\delta_{Ijt})}{\sum_{k=1}^{J} exp(\delta_{Ikt})} \tag{1.7}$$

Estimation and Identification

The model is estimated in a similar fashion to the provider demand model, using maximum likelihood through a multinomial logit model on the years 2009-2013. I do not observe Unicare products in my data, as the insurer does not contribute to the APCD. I therefore run the insurance demand model on the set of GIC enrollees who do not purchase Unicare products. A full set of plan fixed effects are included. One important note, as mentioned above, is that I include in the model an indicator variable for whether a particular plan matches

an enrollee's plan choice from the previous year. This follows prior literature on plan inertia (Handel, 2013; Polyakova, 2016; Shepard, 2016) and is designed to capture enrollee switching costs from moving to a different plan. This variable is extremely important towards matching observed choice behavior in the GIC. Without it, the model would attribute what is really plan inertia to a low value of α_I (premium sensitivity parameter) or a high value for β_1 (the network of the plan itself). This inertia coefficient becomes extremely important in determining employer choices of insurance plans as well.

For the year 2012 (the year in which the GIC began offering its premium holiday), I adjust premiums to reflect the fact that members choosing a narrow network plan would only pay for nine of the twelve months of the year. I do not apply this price adjustment to families in which the eldest member is over 65 years old (as the premium holiday was not extended to retirees). One caveat is that I cannot observe which members are active state employees and which members are municipal employees from years prior to 2012. Therefore, my estimates on the premium coefficient may be slightly biased due to this misassignment of the premium holiday.[14] However, running the model only on the set of new enrollees each year (i.e. those making an active choice) yields a similar premium coefficient and expected utility coefficient, indicating that any bias is likely small.

The expected utility coefficient is identified from within-plan variation in utility of provider networks across individuals. These differences in expected utility stem from differences in household ages, locations (i.e. households that live closer to more prestigious doctors and hospitals than others), and illness

[14]Future iterations of the paper will, however, include corrections for this by matching new enrollees' zip codes with publicly available data on which municipalities joined the GIC in which year

histories (i.e. individuals with a higher disease burden). The premium coefficient is identified through within-plan variation in premiums generated by differences in family type. For households with only one member, individuals pay a base premium, and for households with more than one member, the household pays a total of 2.4 times the base premium, a rate set exogenously by the GIC. Though there may be some concern that base premiums are set endogenously, which might bias my coefficient, premiums in Massachusetts adhere to medical loss ratio laws, which require that plan premiums be set no higher than prespecified amounts by the state government. The GIC is also quite active in negotiating lower premiums with insurers, and has traditionally upheld a medical loss ratio of approximately 90% on all plans (Prager, 2016). Therefore, I take the plan premiums as effectively exogenous conditional on utilization of health care services and expected plan costs, both of which are captured by EU_{Ijt}, and controlling for unobserved plan characteristics that might be correlated with ω_{Ijt}.

1.4.4 GIC Objective Function, Insurer Profit Function, and Premium Setting

I assume that the GIC (the employer), in selecting products and setting prices, maximizes a measure of consumer surplus from the chosen plans less the amount paid out in either medical expenditures (in the case of self-insured products) or premiums to insurers (in the case of fully-insured products). The consumer surplus measure is meant to capture the fact that employers, in part, care about satisfying the health care needs of their employees. A product menu

that can more closely match the needs of its employees would allow the employer to retain employees for longer periods of time, as well as attract new enrollees from other firms. This implies that the more heterogeneous a firm's employees are in terms of demographics, geography, and health preferences, the more employers should be willing to alter their product menu in order to accommodate the needs of the diverse employee preferences.

On the other hand, offering plans that are more generous (i.e. broader network) means that the firm pays out more in premiums, due to the presence of high-cost providers in the network. Moreover, offering multiple plans is costly for firms. I therefore assume that the GIC's plan choices are subject to a fixed cost for each additional product chosen. These costs reflect the fact that employers enjoy significant economies of scale in offering health plans, and that offering multiple plans means that employers need to bear the additional expenses of designing the products, informing consumers, collecting and setting premiums, and negotiating with insurers (K. Bundorf, 2002; Moran, Chernew, & Hirth, 2001). In my setting, the GIC weights the benefits of offering these plans against these fixed costs.

Formally the GIC objective is:

$$W_t = \underbrace{\sum_I \frac{1}{\alpha_I} log\left(\sum_j^J exp(\delta_{Ijt})\right)}_{\text{CS}(\delta_{Jt})} - \sum_I \sum_j (1-\tau) s_{Ijt} R_{Ijt} - \sum_j FC_j \qquad (1.8)$$

Here, the term on the left-hand-side of the function, $CS(\delta_{Jt})$ is the consumer surplus from the GIC offering J products to its employees. R_{Ijt} refers to the *full* premium (i.e. the enrollee plus the employer share). The term τ represents the

percentage of premium that is to be paid by the enrollee, set by the GIC. During the years of my sample period, the GIC set its enrollee share for employees hired prior to 2003 as 20%, while those hired after 2003 at 25%. In recent years, the GIC increased the enrollee share for employees hired prior to 2003 to 25% and is considering increasing all the enrollee shares to combat rising health care costs. The second term in the equation represents the payment in premiums to insurers the GIC contracts with. Note that for self-insured plans, this term would be the full cost of medical care expenses, rather than the premium cost to insurers. Finally, the third term, FC_j represents the fixed cost to the GIC of offering plan j to its enrollees.

Insurers, meanwhile, are assumed to set premiums to maximize profits. Let the marginal *hospital* cost for plan j be given by:

$$mc_{jtH}(N_{jtH}) = \sum_{i \in I} \sum_{l} f_{il} w_{lt} \sum_{h \in N_{jtH}} \sigma_{ilth}(N_{jtH}) p_{jth} \tag{1.9}$$

And let the marginal *physician* costs for plan j be given by:

$$mc_{jtS}(N_{jtS}) = \sum_{i \in I} \sum_{s} \sum_{r} f_{ir} RVU_{rt} \sum_{d \in N_{jtS}} \sigma_{irtds}(N_{jtS}) p_{jtds} \tag{1.10}$$

Then MCO (insurer) m's profits are given by:

$$\pi_{mt} = \sum_{j \in J_m} \sum_{I} \left(s_{Ijt}(\delta_{Jt}) \left[R_{Ijt}[1, 2.4] - \underbrace{mc_{jtH}(N_{jtH}) - mc_{jtS}(N_{jtS})}_{mc_{jt}(N_{jt})} \right] \right) \tag{1.11}$$

In the equation above, J_m refers to the set of products offered by MCO m

and N_{jt} refers to the overall network of plan j in time t (where N_{jtH} refers to the hospital network of plan j and N_{jtS} refers to the physician network of plan j). The $[1, 2.4]$ next to the premium variable, R_{Ijt} refers to the multiple of the premium depending on the household type. These are assumed to be set exogenously, where if the household type is "family" the premium is 2.4 times the base individual premium, regardless of family size. In equation (9), w_{lt} refers to the weight assigned to a particular hospital diagnosis. These weights were also used in the hospital demand model in section 4.1 (a more thorough discussion of their construction follows in section 4.4.1). In equation (10), RVU_{rt} refers to the RVU weight assigned to a particular physician procedure. Recall that f_{il} and f_{ir} are the probabilities that a type i individual contracts a particular diagnosis l or requires procedure r. Finally, p_{jth} is the negotiated base price between plan j and hospital h in time t, while p_{jtds} is the negotiated base price between plan j and physician practice d for specialty s in time t. I define each of these price and weight terms in the next section.

Assuming a multi-product Nash-Bertrand price-setting equation, the first-order condition for the insurer profit function is:

$$0 = \frac{\partial \pi_{mt}}{\partial R_{jt}} = \sum_{I} \left(s_{Ijt}(\delta_{jt}) + \sum_{n \in J_m} \frac{\partial s_{Int}(\delta_{nt})}{\partial R_{jt}} \left(R_{jt}[1, 2.4] - mc_{jt}(N_{jt}) \right) \right) \tag{1.12}$$

Equation (12) assumes that insurers have virtually full leeway to set premiums on the GIC, and that each insurer m competes with others for enrollees. Therefore, the more insurers the GIC chooses to contract with, the less any particular insurer will be able to mark up premiums over their marginal costs. If there are fewer plans, then insurers will be able to mark up premiums higher.

These assumptions are fairly strong for this setting, however, for several reasons. First, two of the largest plans offered by the GIC (Harvard Pilgrim Independence and Tufts Navigator, both broad network plans) are self-insured. For these plans, insurers offer the GIC a "suggested" premium, but the GIC is free to set them at whatever rate they want. This means not only that the GIC has incentives to keep premiums low (to increase the consumer surplus term, $CS(\delta_{Jt})$, in equation (8)), but also that insurers have incentives to keep their "suggested" premiums low. Self-insured plans, since they are merely paid for administrative services, care less about markups over marginal costs and prefer to attract the most enrollees possible. Second, the GIC, as a large employer group that covers about 8-9% of the state's employees, may have considerable bargaining leverage with insurers to reduce premiums, thereby inhibiting insurers from setting markups that are too high.[15] Finally, plans in Massachusetts are bound by state medical-loss-ratio (MLR) regulation requiring that plans spend no less than 85% of premium dollars on medical care expenses. For these reasons, plans on the GIC are observed to set premiums, on average, at about 10% over their medical expenditures.

Therefore, as an alternate pricing assumption, I allow the GIC/insurers to set premiums at a fixed 10% markup over marginal costs. The pricing equation then becomes:

$$\sum_{I} s_{Ijt}(\delta_{jt}) R_{Ijt}[1, 2.4] = 1.10 mc_{jt}(N_{jt}) \tag{1.13}$$

In my results section, I report the results of both assumptions: Nash-

[15] See Ho & Lee (2016) for a model that incorporates employer-insurer bargaining over premiums using data from CalPers (an employer group similar to the GIC) in California (Ho & Lee, 2016).

Bertrand and a fixed-cost markup. I show in section 5.5.1 that the latter assumption fits the observed price-cost margins in the data much more closely than the former.

I now describe the construction of the negotiated price variable for physicians and for hospitals.

Construction of the Negotiated Price, p_{jth} and p_{jtds}

In order to complete equation (11) above and in order to estimate insurer marginal costs for counterfactual plans and networks, I need to construct a measure for the negotiated price between insurers and providers, or the base reimbursement price. I leverage the fact that insurers and providers do not typically negotiate over a full menu of prices for different services, but rather negotiate over a base price and then use a series of weights to scale the base price in order to arrive at a payment for each diagnosis and procedure. These payments are observed in the APCD as "allowed amounts," or the maximum allowable payment that an insurer would be a particular provider for a particular episode of care. I use these observed amounts to back out a negotiated rate for each insurer-provider combination. Similar approaches have been taken by Gowrisankaran et al. (2015), Prager (2016), and Ho & Lee (2016).

For physicians, who are typically reimbursed on a fee-for-service basis for each procedure, r, I rely on observed RVU weights in addition to allowed amounts, as in Kleiner et al. (2015). I assume that price takes the following form:

$$A_{irjtds} = p_{jtds} * RVU_{rt} \tag{1.14}$$

$$ln(A_{irjtds}) = ln(p_{jtds}) + ln(RVU_{rt}) \tag{1.15}$$

A_{irjtds} refers to the allowed amount between plan j and physician practice d of specialty s, for patient i getting procedure r in time t. Here, the allowable amount is a function of the base negotiated price, p_{jtds} between plan j and practice d, multiplied by the RVU weight for the procedure, RVU_{rt}. The model I estimate is:

$$ln(A_{irjtds}) = ln(RVU_{rt})\rho + \gamma_{jtds} + \epsilon_{irjtds} \tag{1.16}$$

where γ_{jtds} refers to plan-practice-time fixed effects. After estimating this model, I fix the RVU to 1 (i.e. $ln(RVU_{lt}$=0). The resulting predicted payments yield a price for each insurer-practice-specialty combination for a *standardized* procedure, and these are used as p_{jtds}.

In the case of hospitals, I assume that the negotiated amount is multiplied by a weight related to the "Diagnosis-Related Group (DRG)" of the particular illness that is being treated, as hospitals are reimbursed by diagnosis. These weights are typically assigned annually by CMS. Unfortunately, the APCD does not have a variable organizing the ICD-9 diagnosis codes into DRGs. Therefore, I follow Shepard 2016 and take a reduced-form approach towards estimating the hospital base price, by running the following model:

$$ln(A_{irjth}) = \gamma_{jth} + \psi_{lt} + x_{ilt} + \epsilon_{iljth} \tag{1.17}$$

Here, A_{iljtd} refers to the observed allowed amount for patient i with diagnosis l on plan j in time t seeking care from hospital h. γ_{jth} are fixed effects for every plan-hospital-time combination. Rather than incorporating a numerical weight with an estimated linear parameter, as done in the physician model, I proxy for diagnoses by including ψ_l. These are a set of fixed effects for the 18 CCS diagnosis categories used in the demand model for hospitals. The model is therefore similar to the physician price construction model, except that by including these fixed effects, I estimate weights for each diagnosis rather than using observed weights. The model also includes Elixhauser comorbidity indexes for each of 12 secondary diagnoses, x_{ilt}. This is meant to capture nuances within diagnoses that may require heavier use of hospital resources than in generic cases (such as comas, hypertension, etc.). I use the model to predict prices for each insurer-hospital-year combination, $p_{jth} = exp(\gamma_{jth})$ and to predict the weights for each diagnosis group, $w_{lt} = exp(\psi_{lt})$. For each year, I then take the average predicted weight across admissions and consider this to be the "standardized diagnosis" for which base prices are negotiated between insurers and hospitals. I scale the predicted price by this factor in order to achieve the predicted base price for hospitals, p_{jth}.

Table 5 reports the average negotiated base prices and average weights by type of provider and facility type in 2013. Practices that are "office-based" are defined as practices in which more than 70% of the claims are conducted in an office-based setting. Any setting in which less than 70% of the claims are performed in an office is considered a "facility-based" setting. These include group practices in which services are primarily performed in outpatient settings of hospitals, or physicians performing services within hospital settings, but billing for professional services separately from inpatient admissions.

Table 1.5: Estimated Price and Weight Measures, 2013

Variable	PCPs	Cardiologists	Orthopedists	Hospitals
		Office-Based		
Average Base Price	61.70	53.08	58.42	–
	(17.07)	(15.76)	(24.56)	–
Average Weight	2.74	3.24	4.46	–
	(3.02)	(3.31)	(10.80)	–
		Facility-Based		
Average Base Price	69.87	56.91	55.07	10,303.73
	(28.56)	(19.54)	(16.82)	(3,177.89)
Average Weight	2.91	2.07	5.31	1.00
	(4.06)	(3.61)	(12.95)	(0.34)

Standard deviations in parentheses. "Average base price" refers to the negotiated price for a standardized unit of health care. In the case of physician practices, this refers to a case where $RVU_l = 1$. In the case of hospitals, this refers to the case where $w_l = 1$. Hospital weights are scaled so that the yearly average is one, meaning that hospital base prices refer to the price fo a procedure of average weight.

The table suggests that negotiated prices do not vary considerably across medical specialties in Massachusetts. Primary care practices do, somewhat surprisingly, receive higher reimbursements per RVU on average than do specialists. However, this is most likely due to the fact that both OB/GYN doctors and pediatricians are included in the PCP definition. Pediatricians, in particular, charge extremely high base rates.[16] Within specialty, there is considerable variation. The largest practices within a specialty receive more than $80 per RVU and the smallest receive as little as $30 per RVU. In the hospital market, the maximum base price in 2013 was $17,306 while the minimum was $3,545. Additionally, there are some notable differences in the average weights per procedure for physicians. Most notably, although the negotiated price for orthopedic practices appears lower per RVU, orthopedists tend to perform more labor-intensive procedures, and therefore have higher RVU weights per procedure performed. On

[16]The most expensive hospital in Massachusetts in terms of base price is Boston Children's Hospital. Both facility fees for this hospital and professional fees charged by physicians employed by it are considerably higher than the average.

average, PCPs in office-based settings receive average reimbursements of $169 per visit. Orthopedists, however, receive approximately $260 per visit. Going one standard deviation higher than the mean weight yields $355 per visits for PCPs and $891 for orthopedists.

1.4.5 Product and Network Choice

Having demand and cost estimates in hand, I proceed with the first stage of the model, where the GIC select a set of products to offer its enrollees and the networks of those products. Specifically, the GIC chooses δ_{Jt} to maximize:

$$\max_{\delta_{Jt}} \left[E \left(\underbrace{CS(\delta_{Jt}, \theta) - \sum_{I} (1-\tau) s_{Ijt} R_{Ijt}}_{\mathrm{S}(\delta_{Jt};\theta)} \right) - \sum_{j} FC_j \right] \tag{1.18}$$

Here, $S(\delta_{Jt})$ refers to the marginal social surplus from having product menu J (in other words, the consumer surplus, $CS(\delta_{Jt}, \theta)$, minus payments to insurers). The fixed costs in the equation refer to non-health-care or premium-related costs associated with purchasing and offering multiple insurance products for enrollees. These include administrative costs for maintaining multiple plans, negotiating costs (each separate plan needs to be negotiated with insurers and providers), informing consumers, printing costs, and any fees that the employer pays insurers that increase with the number of plans purchased.[17]

As an alternate specification, I also run the model assuming that, rather than

[17] Bundorf (2002) notes that employers cite the administrative burden as a constraint on the number of plans offered.

the GIC selecting the number of products and the networks of the products offered, that these decisions are made entirely by the insurers with whom the GIC contracts. In other words, under the alternate model, the GIC engages in long-term contracts with several insurers, but those insurers have full leeway to decide which products are offered and which provider networks are included in those products.

Under this alternate assumption, the fixed costs of designing and offering plans are borne by the insurers rather than the employer, and the maximization problem for insurer m in time t becomes:

$$\max_{\delta_{Jmt}} \left[E\left(\pi_{mt}(\delta_{Jt}, \theta)\right) - \sum_{j} FC_j \right] \tag{1.19}$$

where δ_{Jmt} is the "quality" of product menu J offered by MCO m. I report the fixed costs under both sets of assumptions in Section 5.

Estimating the Fixed Costs

To estimate the fixed costs associated with offering additional plans, I follow work by Pakes, Porter, Ho, and Ishii (2015) in constructing moment inequalities and bounding estimates on the fixed costs, rather than imposing an equilibrium through distributional assumptions on the fixed cost parameter. This procedure was subsequently adapted and used to estimate fixed and sunk costs of product introductions in markets such as computers, pharmaceuticals, and smartphones (Eizenberg, 2014; Nosko, 2014; Mohapatra & Chatterjee, 2015; Fan & Yang, 2016).

For this, I construct two counterfactual quality vectors associated with adding and removing products to the GIC. I define $\delta_{J+j,t}$ as the total product quality that would result from offering an additional product j that is not currently offered. I define $\delta_{J-j,t}$ as the total product quality that would result in the GIC removing one of its currently offered products, j.

The estimation follows from a revealed preference assumption that the products I observe in the data are chosen in equilibrium. That is, the GIC selects the products and networks of providers to offer its employees much as a social planner would; by maximizing consumer surplus less payments to insurers and fixed costs of offering multiple products. This establishes the necessary conditions that the GIC would not choose to add a product ($\delta_{J+j,t}$) or remove a product ($\delta_{J-j,t}$) unless these deviations increased its objective function, W_t. These necessary conditions allow me to estimate bounds on the fixed cost parameter.

One side of the bound comes from the assumption that any product the GIC chooses to offer must necessarily increase social surplus. Therefore, by removing a product currently offered and computing counterfactual surplus, I can infer that the fixed costs for offering an additional product must be less than the surplus gained by offering the product. Formally this upper bound on fixed costs is given by:

$$FC_j \leq E\left[S(\delta_{Jt};\theta) - S(\delta_{J-j,t};\theta)\right] \equiv \overline{FC}_j \tag{1.20}$$

where $\overline{FC}_j$ refers to the upper bound on fixed costs. Similarly, we can obtain the lower bound as follows:

$$FC_j \geq E\left[S(\delta_{J+j,t};\theta) - S(\delta_{Jt};\theta)\right] \equiv \underline{FC}_j \tag{1.21}$$

where $\underline{FC}_j$ is the lower bound on fixed costs. This side of the bound implies that if the GIC can offer a potential product, but is not observed to, then it must be the case that fixed costs are larger than the change in marginal social surplus from introducing it.

Assume that the GIC's expectation of its total surplus from adding or removing products follows the following form:

$$E[S(\delta_{Jt})] = O(\delta_{Jt}) + v_{fc} \tag{1.22}$$

Here, $O(\delta_{Jt})$ refers to the observed surplus from offering a certain combination of plans, and v_{fc} is a mean-zero idiosyncratic error term.

As long as the GIC has correct expectations on average, v_{fc} will go to zero as the number of potential products, K, increases, leading to the following estimation equation:

$$\underset{K\to\infty}{plim}\frac{1}{K}\sum_{j}^{K}\left(O(\delta_{Jt};\theta) - O(\delta_{J-j,t};\theta)\right) \geq FC \geq \underset{K\to\infty}{plim}\frac{1}{K}\sum_{j}^{K}\left(O(\delta_{J+j,t};\theta) - O(\delta_{Jt};\theta)\right) \tag{1.23}$$

Eizenberg (2014) and Mohapatra (2016) describe in detail a potential selection problem that would arise out of this formation if the error term varied by the type of product offered. Namely, the GIC may choose to contract with certain insurers, offer certain products, or offer certain networks for which the fixed

costs of doing so are lower. In my setting, I circumvent this selection problem by following Nosko (2014) and assuming the fixed costs of offering additional plans is the same regardless of where the plan is in the quality space.

While this may be a strong assumption in other settings that have wide variation in fixed or sunk costs of product introduction, it is a more reasonable approximation for this environment. For one thing, I am estimating the fixed costs associated with introducing additional plans under the umbrella of one large employer group. While such costs may differ across employers, it is unlikely that there are substantially different fixed costs *within* employer group, and therefore it is not likely that the GIC exhibits substantially different fixed costs for plan introduction to its own employees. Moreover, whereas the fixed costs of, for instance, introducing new pharmaceuticals into the market may highly differ depending on the nature of the drug, it is unlikely that the added administrative burden of offering an additional insurance plan depends significantly on the type of plan that is offered.

This assumption may be violated if, for instance, offering a product that was broader in network size than another product also meant an increase the cost of the negotiation process. However, this is unlikely to apply to the GIC for two reasons. First, I do not allow the GIC to offer any plans that in which the network is larger than the largest currently offered by the particular insurer anywhere in Massachusetts. In other words, insurers can only design plans that are narrower than what they currently offer, but not broader. This implies that there would be no additional contracting fixed costs for providers with whom any particular insurer does not currently negotiate with. Second, while employer groups negotiate premiums with different plans, they rarely ever negotiate base

prices with providers. This task falls largely onto the insurers, and it is therefore unlikely that the added negotiation cost of offering broader network plans would result in additional fixed costs for the GIC itself.

1.5 Results

I now describe the results of my model. I begin with the results of the physician and hospital demand section, followed by the results of the insurance plan demand, and finally the fixed costs estimates from the supply side of the model.

1.5.1 Demand for Hospitals

Table 6 displays the results for the hospital demand model. The results are displayed for a full sample of hospital admissions in Massachusetts for consumers on the GIC between 2009 and 2013. The model is run on a flexible set of interactions, including distance with patient characteristics, distance with provider characteristics, and patient characteristics with hospital characteristics. This is meant to capture heterogeneity in preferences for hospitals. In addition to the reported coefficients, the model also contains a set of fixed effects for the 18 CCS disease categories interacted with distance, a full set of hospital fixed effects as well as hospital fixed effects interacted with disease weights, w_{lt}. These latter fixed effects are meant to capture unobserved hospital quality, as well as allow patients with different disease severities to have differential preferences for different hospitals.

Consistent with prior literature on hospital demand, the distance coefficient

is negative and significant, implying that patients prefer to go to hospitals that are close to where they live. While this coefficient is difficult to interpret (the measure is in utils instead of a dollarized amount), comparing this coefficient with other parameter estimates shed some light on its practical magnitude. For instance, the estimates imply that hospital patients are on average willing to travel approximately 20 extra miles to reach the hospital with the highest unobserved quality parameter (i.e. the largest fixed effect estimate). This is indicative of the fact that patients are "willing-to-pay" in terms of extra miles traveled to access prestigious, academic medical centers, such as Mass. General and Brigham and Womens (both owned by Partners), Beth Israel, Lahey Medical Center, and others.

Table 1.6: Results of Hospital Demand

Variable	Utility Parameter	Standard Error
Distance	-0.2650***	0.0072
Used Hospital	3.5664***	0.0273
DistxFemale	-0.0034***	0.0011
DistxAge	-0.0004***	0.0000
DistxChronic	0.0190***	0.0014
DistxCritAccess	0.0127***	0.0038
DistxSpecialty	0.0631***	0.0024
DistxAcademic	0.0327***	0.0018
NeuroxNeuro	0.8089***	0.2382
CardiacxCathLab	0.4321***	0.0439
ObstetricsxNICU	2.2872***	0.0740
ImagingXMRI	0.2318***	0.0502
Hospital FE	Yes	
Obs.	2,815,140	
Pseudo R2	0.54	

Results from hospital demand model from years 2009-2013. "Chronic" refers to having a chronic condition, "Specialty" refers to being a specialty hospital, "NeuroxNeuro" refers to a patient with a neurological disorder interacted with an indicator for whether the hospital had a neurology unit. Omitted from the table are distance terms interacted with each of 18 CCS diagnosis categories and a full set of hospital fixed effects.

A second important finding concerns the large positive and significant coefficient on individuals who have used the hospital in the previous period. The coefficient implies that conditional on age, disease, and hospital characteristics, individuals would be willing to travel approximately 13 extra miles to be admitted to a hospital they have used previously.

Women are less likely to travel far to reach a hospital, and older individuals (conditional on diagnosis) also receive significant disutility from traveling. Conditional on age, however, patients with histories of chronic conditions (i.e. sicker patients) are willing to travel *more* to access a hospital of their choice. People are also on average more likely to travel to a hospital that has more beds, a specialty hospital (such as a children's hospital or a cancer center), or to travel for an academic medical center. This reinforces the point that prestigious academic medical centers in Massachusetts are able to generate high demand for their facilities.

Finally, I report the coefficients on a series of variables interacting patient diagnosis with hospital amenities. Each of these are, unsurprisingly, positive and significant. Patients with a neurological disorder significantly prefer hospitals that have neurology units. Patients with a cardiac CCS diagnosis significantly prefer hospitals with a catheterization laboratory, patients with obstetrics conditions significantly prefer hospitals with a neo-natal intensive care unit, and patients with a diagnosis requiring imaging (defined to be either a neurological, cardiac, or musculoskeletal diagnosis) prefer hospitals equipped with magnetic-resonance-imaging machines.

It is worth mentioning that this model omits copayments that plans charge to visit different hospitals. On the GIC, plans are differentiated in their premiums,

their networks, and the copays that patients pay for a hospital admission across *plans*, across *hospitals*, and over time (Prager, 2016). In practice, the demand effects of these copays is fairly minor and, for my purposes, does not alter the subsequent results in meaningful ways. However, in Appendix A.2, I report the results of alternate hospital demand models that include these copays.

Following previous literature, I also assume there is no selection on unobservables in this model (that is, providers are not horizontally differentiated in ways unobserved to the econometrician). Appendix A.3 addresses potential selection concerns in more detail.

1.5.2 Demand for Physicians

I next turn to the results of the physician demand models for primary care practices, cardiology practices, and orthopedic practices, which can be seen in Table 7. Due to the large number of physician visits during my time frame, I run the model on a random sample of 50,000 visits across the five years for each different specialty group. In order to further reduce the dimensionality given the large number of physician practices, I limit the sample to only the largest 100 practices in each specialty (in terms of number of claims). The model includes distance interacted with patient characteristics, physician practice characteristics, as well as patient characteristics interacted with provider characteristics. It also includes a full set of practice fixed effects for each specialty, as well as practice fixed effects interacted with RVU weights. Appendix A.2 shows additional results that look at only the top 50 practices in each specialty, as well as models assuming that patients choose only one physician per year (rather than make a

separate choice for each visit).

Table 1.7: Results of Physician Demand Models

Variable	PCP Practices	Cardiology Practices	Orthopedic Practices
Distance	-0.1154***	-0.1167***	-0.2046***
	(0.0037)	(0.0020)	(0.0024)
Used Doctors	5.3061***	4.0696***	4.4516***
	(0.0188)	(0.0179)	(0.0221)
DistxFemale	0.0014***	-0.0008	0.0002
	(0.0011)	(0.0008)	(0.0009)
DistxAge	-0.0002***	-0.0005***	-0.0001***
	(0.0000)	(0.000)	(0.0000)
DistxRVU	0.0007***	0.008*	0.0010***
	(0.0002)	(0.000)	(0.0001)
Practice FE	Yes	Yes	Yes
Obs.	5,379,725	5,187,693	5,448,055
Pseudo R2	0.75	0.53	0.53

Results of physician demand models are for years 2009-2013. Excluded from tables are estimates for distance interacted with number of doctors, distance interacted with the share of doctors who are of the specialty types being considered, and a full set of practice fixed effects.

Consistent with the results of the hospital demand model, distance plays an extremely important role in choosing physician practices. Across the three specialist groups, distance has a negative and significant effect on utility. While the magnitude of the coefficient is similar for primary care physicians and cardiologists, it is more negative for orthopedists, suggesting that consumers do not like traveling for orthopedic surgery.

Another important driver of choice, also similar to the hospital demand model, is whether the patient had gone to the physician practice (at the particular location) in the past. For all three specialist groups, the coefficient on "Used Doctor" is positive and significant. However, unlike the hospital demand model, the magnitudes on these coefficients are extremely large, implying that patients are willing to travel far greater distances to access the same physicians

than they are willing to access the same hospitals. Indeed, the magnitude of the coefficient is *particularly* high for primary care practices. The results imply that on average patients are willing to travel approximately 48 miles extra to see a primary care physician if they had seen that caregiver in the preceding year. Patients are willing to travel about 37 miles to see a cardiologist if they saw that same cardiologist in the previous year. However, patients in need or orthopedic surgery are only willing to travel about 22 miles to see an orthopedist if they saw that practitioner in the previous year.

While these magnitudes may seem high, particularly when compared to the hospital results (where patients are on average only willing to travel an extra 13 miles to see the same hospital), it is worth noting that "primary care" includes not only family practice physicians, but OG/GYN and pediatricians. Anecdotal evidence suggests patients are extremely loyal to these types of specialists, particularly since patients tend to build relationships with these types of providers from frequent use of routine care. Moreover, cardiologists tend to attract somewhat older and sicker patients (as shown in Table 3), who also require more frequent care and heart monitoring. In light of this, the *relative* magnitudes between the coefficients also makes sense. Patients seem less willing to travel to see the same orthopedists as they used previously, which could be due to the fact that patients do not typically (with the exception of physical therapy or rehabilitation) seek care from orthopedists for sustained periods of time.

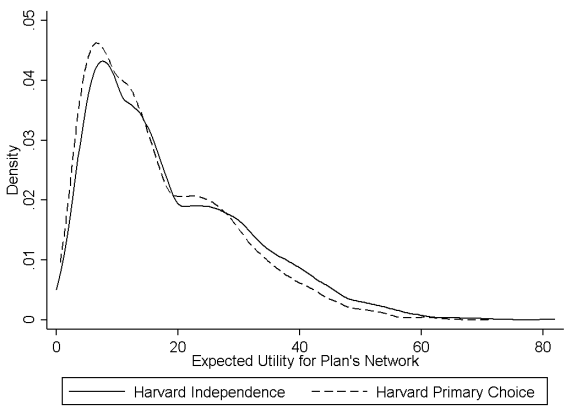
The coefficient on distance interacted with RVU for all three specialty types is positive, suggesting that patients are willing to travel further to see a physician if they require a more complicated procedure. In addition, for each of the three specialty types, older individuals are less likely to travel to see a particular

physician. This could be reflective of the fact that older individuals, conditional on severity of procedure (which is captured by the RVU coefficient), are less able to travel and therefore prefer physicians close to their homes. These results are broadly consistent with the results on hospital care suggesting similarly that older patients are less willing to travel for hospital care and that patients with histories of chronic conditions are more willing to travel. For primary care practitioners, women are more willing to travel, however there is no significant gender difference in physician preference for cardiologists and orthopedists. This is perhaps due to the fact OB/GYNs are included as part of the definition of "primary care."

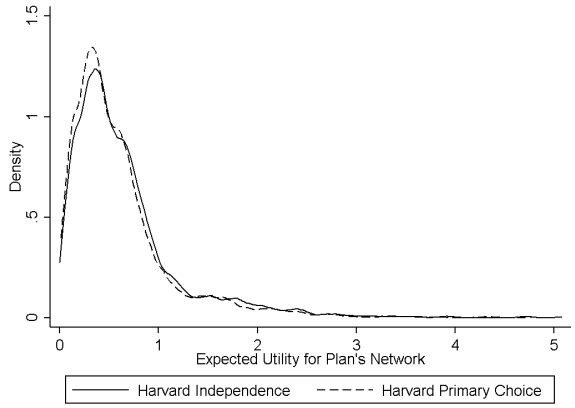
1.5.3 Relation Between Negotiated Price and Network Utility

Having the estimates of hospital and physician demand in hand, I use the estimated parameters from the model to construct the measure of expected utility for an insurance plan's network, EU_{Ijt}, which is subsequently used as an input in the plan demand model. Figures 6 and 7 plot the variation in expected utility across households for select plans in the GIC between 2010 and 2013. In Figure 6, I plot a kernel density function for each of Harvard's health plans: Harvard Independence (its broad network plan) and Harvard Primary Choice (its narrow network plan). In Figure 7, I do the same for Fallon's broad versus narrow plans.

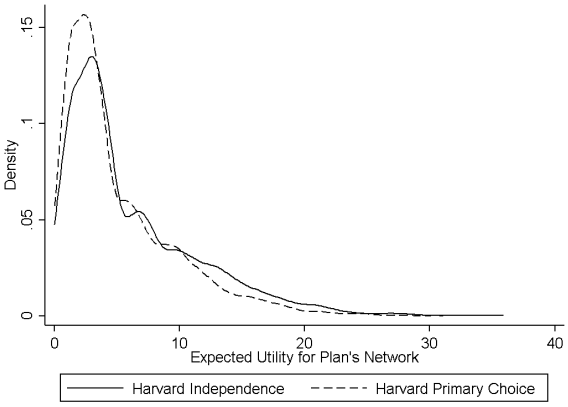
In each figure, the solid lines represent the insurer's broad network product and the dashed line represents the narrow network product. It is immediately clear from this series of charts depicted in Figure 6 that Harvard's broad



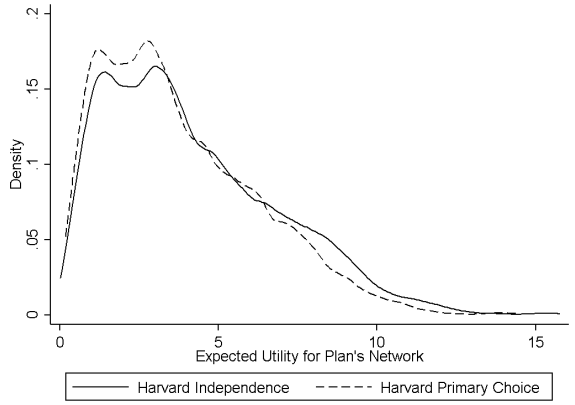
(a) Total



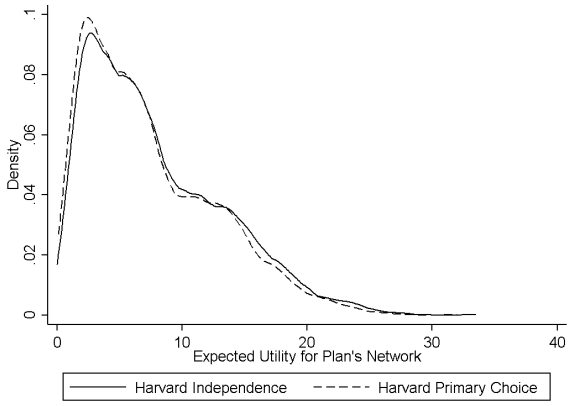
(b) Hospitals



(c) General Practices



(d) Cardiology Practices

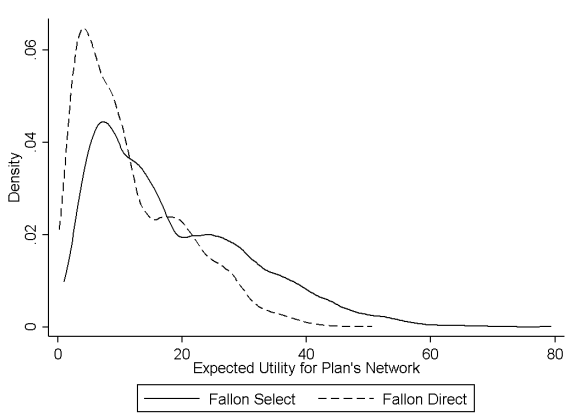


(e) Orthopedic Practices

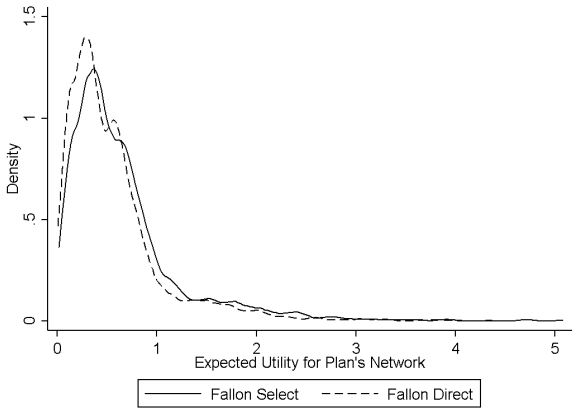
Figure 1.6: Expected Utility for Harvard Pilgrim's Broad and Narrow Plans

and narrow plans yield very similar expected utility values for consumers, even across all provider types (hospitals and physician specialties). Panel (a) shows the distribution of total utility for both hospitals and physicians. While the plot for the broad network does skew slightly to the right to that of the narrow network, the two graphs virtually overlay one another. Indeed, Harvard's narrow network, while narrower than its broad plan, still covers a large number of practices and hospitals, and is the largest narrow-network product on the GIC. Looking at panel (b), which shows the utility distribution only for hospitals, EU_{IjtH}, consumers appear to view both plans' hospital networks quite similarly, despite the fact that Harvard's narrow network excludes Partners hospitals. When translating to primary care practices (panel (c)) and cardiology practices (panel (d)), however, the distribution for Harvard's broad network does begin to skew more rightward than the hospital distribution, implying that most of the difference in valuation between the two plans comes from the physician network. However, these skews are still fairly minor. As depicted in Figures 4 and 5, Harvard's narrow plan is the only narrow network product to cover the physician practices affiliated with Atrius Health system, which draws a considerable share of primary care patients. Therefore, it follows that consumers would value Harvard's two networks at a fairly similar level, given the significant overlap of large-share providers in both plans.

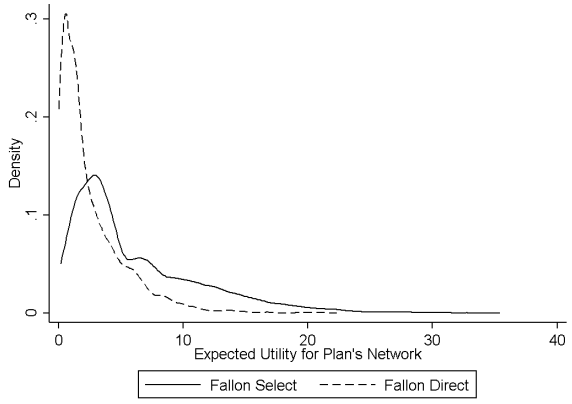
Figure 7, however, which depicts the difference in utility between Fallon's broad network and Fallon's narrow network tells a very different story. Looking at panel (a), it is clear that the utility distribution for Fallon's broad network skews considerably rightward of Fallon's narrow network, implying that consumers value Fallon's broad network significantly more than its narrow network. Disaggregating these utility measures by provider type sheds more light



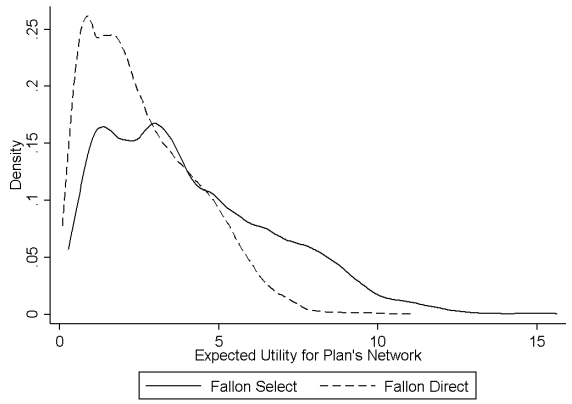
(a) Total



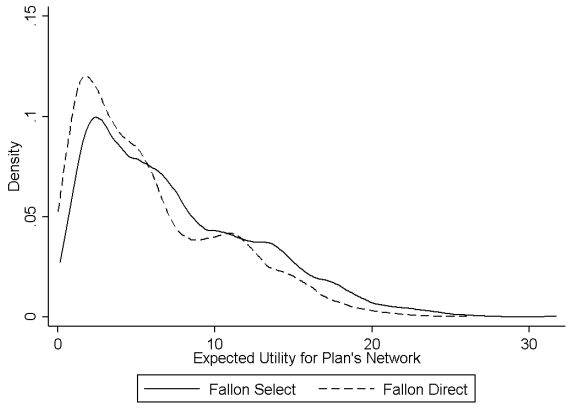
(b) Hospitals



(c) General Practices



(d) Cardiology Practices



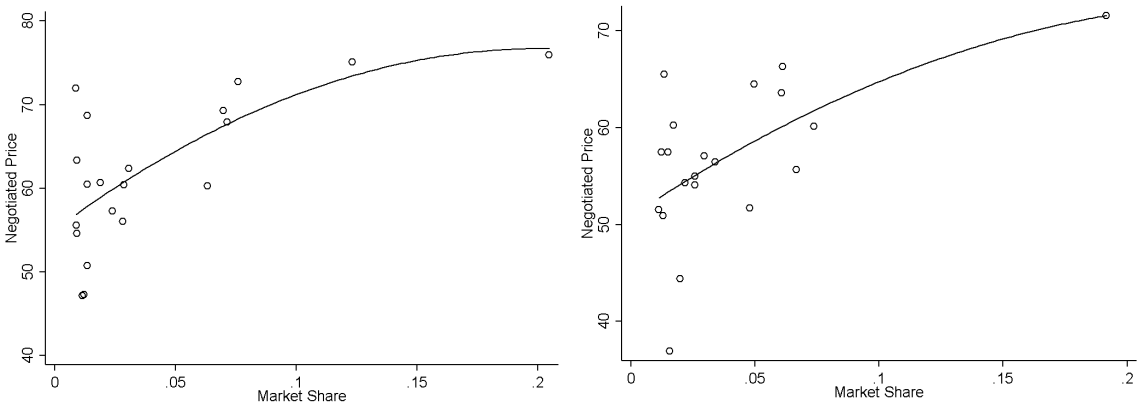
(e) Orthopedic Practices

Figure 1.7: Expected Utility for Fallon's Broad and Narrow Plans

on this. Panel (b) does reveal that consumers prefer the hospital network covered by Fallon's broad network more than they do covered on Fallon's narrow network, and the difference is more apparent than the hospital utility difference between the two Harvard plans. However, most of the difference similarly comes from variation in physician networks between the two plans. While the hospital distribution only skews somewhat to the right, the physician distribution skew very much to the right, particularly for primary care and cardiology practices. In fact, looking at panel (c), the maximum utility for Fallon's broad network is slightly above a value of 20, whereas the maximum for Fallon's narrow network is around 35. Panel (d) shows that consumers value the orthopedic practices in Fallon's broad network more than its narrow network as well, though not by as much as they value the cardiologists and primary care physicians. Again, this is likely due to the fact that PCPs and cardiologists develop more close relationships with patients through repeated use than do orthopedists.

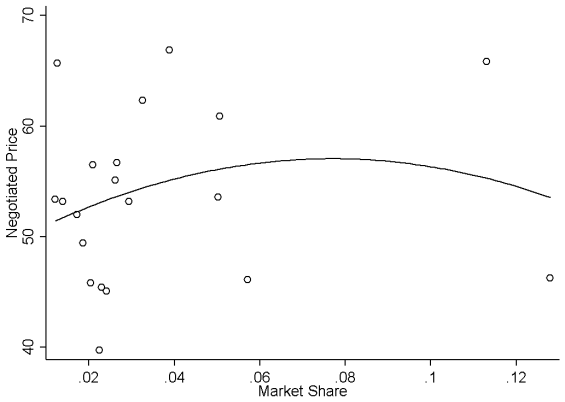
Taken together, these two figures clearly show that accounting for physician services is an important part of consumer valuation of networks. While hospital networks do play a role in consumer choice, preferences diverge more strongly when considering the variation in availability of physicians between narrow and broad network plans.

I next examine whether the preference for broad network plans translates into higher negotiated rates for those providers. Figure 8 depicts the relationship between demand and negotiated provider price for one of the insurers on the GIC. Due to the fact that actual negotiated prices are displayed, I do not display the names of the insurers in these figures. For the physician practices, I

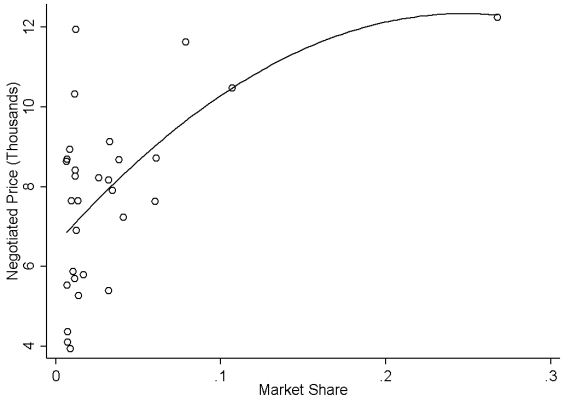


(a) General Practices

(b) Cardiology Practices



(c) Orthopedic Practices



(d) Hospitals

Figure 1.8: Insurer 1 Negotiated Price by Market Share, 2011

only display the negotiated prices for the largest 20 practices in Massachusetts. The y-axis depicts the negotiated base rate (as computed in Section 4.4.1), where the x-axis displays the predicted market share for those providers from the hospital and physician demand models. Appendix A.1 shows additional figures for a different insurer.

It is clear from the graphs that there is a distinct positive relationship between provider price and consumer valuation of a provider within the insurer's network. The relationship appears strongest for cardiologists, suggesting that

it is likely that cardiology practices have the most leverage to negotiate higher rates with insurers. However, primary care practices also exhibit a positive relationship, as does the relationship between hospital price and demand. Orthopedic practices do not appear to exhibit a positive relationship. These results suggest that within specialty groups, including high-demand providers indeed tends to translate into higher prices for medical care. These prices then, in turn, translate into higher premiums for consumers. The inherent tradeoff for insurers and employer in offering plan choice thus becomes clear: to offer a broad network plan to consumers would yield greater consumer surplus through the inclusion of high-valuation hospitals and doctors, but would also reduce surplus through higher premiums. This tradeoff is explored more in the next sections.

1.5.4 Demand for Insurance Plans

Table 8 reports the results for the insurance plan demand model. I run the model on the years 2010-2013. I omit 2009 as I cannot observe which members were enrolled in the GIC prior to that year, and would therefore mismeasure enrollee inertia. Further, due to the dimensionality of the data (approximately 200,000 GIC members per year multiplied by about 70 hospitals, 18 potential diagnoses, 100 practices, and three different specialties), I only run the model on a subset of 3,000 households across the four years of data. As I cannot observe Unicare products in the data, I run each model on the set of enrollees in all other GIC plans. Each of the models included a full set of plan fixed effects to capture unobserved plan quality, as well as an interaction between premium and age of the oldest household member in order to capture heterogeneity in

price sensitivity among consumers. Each model included the expected utility variable, EU_{Ijt}, computed in the previous section.

Table 1.8: Results of the Insurance Plan Demand Models

Variable	(1)	(2)	(3)
Premuim ($00s)	-1.04***	-2.32***	-2.50***
	(0.15)	(0.39)	(0.27)
PremiumxAge ($00s)	0.01***	0.02***	0.02***
	(0.00)	(0.01)	(0.01)
Expected Utility (EU_{Ijt})	0.45***	0.47***	0.30***
	(0.02)	(0.05)	(0.03)
Inertia Coefficient	–	–	4.57***
	–	–	(0.07)
Insurer FE	Yes	Yes	Yes
Obs.	46,719	7,436	46,719
Pseudo R^2	0.32	0.25	0.79

Results from insurance plan demand model for 2010-2013. For 2012, the GIC's "premium holiday" was incorporated by providing individuals under 65 years old with a three month discount. Column 1 includes results for the full sample but not including the inertia coefficient. Column 2 includes results for just the sample of new enrollees into the GIC. Column 3 includes the results for the full sample including the inertia coefficient. The premium variable is monthly and reported in hundreds.

Column 1 reports the results on the full sample across the four years, with no inertia coefficient. Column 2 reports the results on a subsample across the four years of only active enrollees who entered the GIC for the first time in that given year (this is similar to Prager 2016). Column 3 (the preferred model) reports the results on the full sample, including the inertia coefficient capturing the value of being on the same plan in year t as in year $t-1$. In all three models, the premium coefficient is negative and significant, implying that households dislike insurance plans with higher prices. Premium interacted with the age of the oldest household member is positive across the three models, suggesting that premium sensitivity *decreases* with age. This is an expected result, as older

individuals tend to be sicker and have higher incomes, and therefore prefer plans with the option value of broader networks. Finally, the expected utility, EU_{Ijt}, coefficient is positive and significant for all three models, implying that conditional on price and unobserved plan quality, households prefer plans with broader networks of hospitals and physicians.

Although the directions of the coefficients in all three models are the same, the magnitudes of the estimates are extremely informative. In the first column, average premium sensitivity is quite muted, whereas the value of a network is extremely high. In order to translate these utility parameters to willingness-to-pay estimates, I approximate the value of particular plans by noting the average differences in expected utility between their networks. As an example, the average difference between the expected utility of Fallon's narrow network product and its broad network for consumers currently enrolled in the broad network is 2.80 for individuals and 8.79 for families. The average age in the sample is 47. This implies that, on average, individuals currently enrolled in Fallon's broad plan would need to be paid approximately $135 per month and families would need to be paid $425 per month to move to the narrow plan. These estimates are quite high, although it should be noted that Fallon's narrow plan covers the fewest providers of any network on the GIC. In 2011, it covered almost a third of the hospitals in Massachusetts that were covered by Harvard's narrow network plan. Applying the same exercise for Harvard Pilgrim yields an average difference in utility of 0.99 for individuals and 2.90 for families. This implies individuals would need to be paid $45 to move from Harvard's broad plan to its narrow plan, and families would need to be paid $140 per month on average for the same switch. In 2011, the difference in premiums between the two Fallon plans was $20 per month for individuals and $50 per month for families,

whereas this difference was $30 and $75 for Harvard. These premium differences are considerably smaller than the differences in valuations for broad networks implied by the model results, suggesting the potential for bias in the form of an underestimated price sensitivity parameter or an overestimated network valuation.

In order to address this problem, I run the model in two alternate ways. The first is by running it only on the sample of enrollees who are first-time GIC members. By definition, inertia for this population is zero, as members had not previously chosen *any* plan and are therefore making active choices. Among these enrollees, the premium sensitivity coefficient drops considerably from -1.04 to -2.32. The expected utility coefficient remains roughly the same. These changes in magnitude suggest that enrollees choosing plans for the first time are considerably more price sensitive than the overall population. Using the same exercise as above, I find that individuals who enrolled in the GIC for the first time who chose Fallon's broad plan would need to be paid $142 per month to move to Fallon's narrow product, while families would need to be paid $242 per month for the same move. For Harvard Pilgrim, these estimates are $27 and $98 per month, respectively. These estimates are considerably improved from the first model, with the network valuation numbers matching Harvard Pilgrim's premium differences almost exactly, though still overestimating Fallon's price differences.

The third column estimates the model again on the full sample, but for all enrollees who were observed to be in the sample in year $t-1$, includes an indicator for whether the household was on the same plan that year. This "inertia" coefficient implies that, on average, a household would be willing to pay ap-

proximately $272 per month to be on the same plan as they were on last year. In other words, households would need to be paid $272 to switch from one plan to another, *conditional* on the network, brand, and unobserved quality components of those plans. Indeed, these switching costs are quite high, particularly when compared with Handel's estimate of approximately $170 per month and Shepard's estimate of $96 per month. However, it is worth noting that Shepard's estimate is for the *individual* market, where consumers tend to switch with greater frequency than in employer market settings such as the GIC, and where the out-of-pocket premiums for consumers is lower. Moreover, the estimate of $272 reflects a *household* switching costs, whereas both Handel and Shepard's estimates are for individual premiums. Indeed, $272 is less than the price of the average broad network family premium on the GIC. In Handel's setting, the average enrollee premium for the most generous PPO plan for an employee in a middle-income tier is about $2,000 annually, implying that switching costs in his setting *exceed* average premiums.[18] Therefore, $272 is a reasonable estimate for household switching costs in an employer setting.

The premium coefficient in this model is very similar to the model run on just active enrollees, suggesting that the population of new GIC enrollees and the full population (including those who switch plans) have similar premium sensitivities. Using this model, individuals on Fallon's broad plan would need to be paid $50 per month to move to the narrow plan, whereas families would need to be paid $166. For Harvard, individuals would need to be paid $19 to move to the narrow plan, whereas families would need to be paid $54. These estimates seem much more reasonable, and particularly for Harvard, much more closely match observed premium differences between broad and narrow plans.

[18]This is similar to results from Polyakova (2016), who finds switching frictions in Medicare Part D to be about twice to four times as large as premiums.

For these reasons, the preferred estimates used throughout the rest of the paper are the ones from Column 3.

Appendix A.4 reports the results of additional plan demand estimates that separate out the expected utility from *hospital* networks versus the expected utility of *physician* networks. The results demonstrate the important of including physician networks in a model of consumer demand. Indeed, it can be seen that when physician networks are ignored, the premium coefficient is significantly underestimated and the hospital network utility is significantly overestimated. This is due largely to the absence of important heterogeneity in physician networks from the model, which drives consumer choice of plans. The implication is that consumers may select into broad network plans not because they are insensitive to price, as traditionally thought, but rather because they value retaining access to certain primary care physicians and specialists.

1.5.5 Fixed Cost Estimates

I next proceed to reporting the estimates of firm (employer/GIC and insurer) fixed costs for introducing additional products to the market. I report estimates of lower and upper bounds on GIC fixed costs over the years 2009-2013. In order to do so I make several assumptions. First, I bound the list of potential products the GIC can offer to any network that was offered by any insurer at any point during the five-year sample. There are two reasons for making this assumption. The first is that Harvard and Tufts both created new products in 2011 to offer on the GIC that were not offered anywhere else in Massachusetts. This implies that the GIC has enough leverage to induce insurers to create brand new networks.

The second reason is that Massachusetts state law contains network adequacy requirements that prevent networks from being too narrow, or from not serving members in particular regions. As I do not observe these specific legislative requirements, I instead assume that any network observed in the data must have satisfied such laws and are therefore feasible potential networks. As an example, if Fallon offered a particular network, I assume that Harvard Pilgrim could have offered the exact same network and mimicked Fallon's strategy. This creates a possible 12 different products that any insurer could offer at any given time, corresponding to 12 different hospital and physician practice networks.

A second assumption is that Neighborhood Health Plan and Health New England could only offer their own particular networks. I make this assumption to accommodate the fact that both insurers are only observed to have one network in the entire sample and throughout all of Massachusetts. In addition, Health New England is a regional insurer, operating only in Western Massachusetts. Therefore, assuming that it could offer a broader network plan that includes providers in Eastern Massachusetts would conflate fixed costs of offering new plans with fixed costs of contracting with an entirely new set of providers or operating in an entirely new region, which I want to avoid.

A third assumption is that at any given time, the GIC can offer as many products as it wants (subject to the potential products available), however it cannot offer two products with the same network.

Finally, I exclude the broad network plans offered by Harvard, Tufts, and Health New England from the upper bound estimates. The reason for this is that these are the three largest insurers in the GIC, each having considerably higher market shares than the rest of the plans. Since any product the GIC may

choose to *add* are narrow-network products, each attracting fairly low numbers of enrollees, the best comparison for the upper bound are similar network products that have similar market shares on the GIC.

The fixed cost estimates are presented in Table 9. I present two sets of fixed cost estimates: one assuming that all the costs of offering new products are borne by the employer/GIC (i.e. equation 18), depicted in panel A, and the other assuming that all the costs of offering new products are borne by the insurer (i.e. equation 19), depicted in panel B. The first set of estimates assumes that if a product is currently not observed, it is due to the fact that the GIC decided it would not increase net marginal social surplus, $S(\delta_{Jt}; \theta)$ to offer the product, whereas if a product *was* offered, social surplus would be harmed by removing it. The second set assumes that an insurer has full leeway to design and offer products on the GIC; therefore if a product is not observed, it is due to the fact that the insurer did not find it *profitable* to design this product, whereas if a product was observed, the insurer would be less profitable by removing it. The true estimate is likely to lie somewhere between the two set of assumptions.

The fixed cost estimates suggest that the GIC spends between $1.15 and $6.64 million a year on fixed costs for each plan offered. Though this estimate seems high, it is actually quite a small fraction of the GIC's overall net spending. When net spending is defined as either premium revenue less medical spending (in the case of self-insured plans) or 75% of premiums paid out to insurers (in the case of fully-insured plans), the estimates of fixed costs fall between 0.12% and 0.70% of total spending. Similarly, these bounds are approximately between 0.26% and 1.50% of the net social surplus (again, defined as total consumer welfare less net spending). Therefore, relative to the overall budget that the GIC allocates

towards health expenditures (nearly $1 billion per year), fixed costs associated with managing multiple plans remains a small, but important component of its objective function.

Table 1.9: Results of Fixed Cost Estimation

	Lower Bound FC_j	Upper Bound FC_j
	Panel A: GIC Chooses Products	
GIC/Employer ($Millions)	1.15	6.64
Share of Net Spending	0.12	0.70
Share of Net Surplus	0.26	1.50
	Panel B: Insurers Choose Products	
Insurer ($Millions)	1.99	3.00
Share of Fallon Profits	1.97	2.97
Share of Harvard Profits	1.23	1.86
Share of HNE Profits	4.73	7.14
Share of NHP Profits	2.07	3.12
Share of Tufts Profits	0.83	1.26
Obs.	124	21

Results from fixed cost estimation for 2009-2013. The "GIC/Employer" row reports the results of fixed costs estimated for the GIC, while the "Insurer" row reports estimates obtained by assuming the fixed costs are borne by the insurers. The corresponding shares are also reported. For the upper bound, Harvard Independence, Tufts Navigator, and Health New England are excluded due to their large size.

The fixed cost bounds for insurers are slightly smaller than those for the GIC, but within the same general range. The estimates range from approximately $2 to $3 million per product offered. Although this represents a fairly substantial share of any individual insurer's variable profits within the GIC, it is a fairly small component relative to their overall statewide profits. Harvard Pilgrim and Tufts, in particular, which are two of the largest insurers in Massachusetts, see fixed costs at less than 2% of total profits for each, and a lower bound of less

than 1% in the case of Tufts. These numbers are plausible relative to reported administrative costs estimates by insurers in Massachusetts.[19]

Two caveats should be noted regarding these estimates. The first is that although the lower bound has a fairly large sample due to the wide availability of various product networks in Massachusetts, fairly few of these networks were offered during my sample in the GIC. Therefore, the upper bound estimates have a sample size of 21, which may affect the standard error estimates. Second, and related, the low rate of offered products in the GIC be driving up the estimates. Since there are only 8 products offered in a given year in the GIC, any particular product removed, if it has a large enough market share, will cause a large decrease in profits (or consumer surplus), which when averaged over a small sample, may bias the estimates upward. I try to correct for this by omitting products with large shares from the upper bound, but the range may still be upwardly biased. While the upper bound is estimated less precisely than the lower bound, its closeness in proximity to costs reported by insurers is cause to believe that these estimates are reliable and useful for conducting counterfactual exercises regarding network choice and product development.[20]

[19] In a 2010 hearing held in Massachusetts with the state's major health insurers, at least one plan identified its expenditure of costs and resources associated with implementing new products as varying between $1 and $3 million in total costs, which is nearly identical to the range of estimates I am finding (*Small Group Health Premiums in Massachusetts*, 2010).

[20] A natural extension of this approach is to consider more employers and more market segments in Massachusetts (including the individual Exchange, CommChoice). If the fixed costs of offering new products are similar across these segments and similar across employers, then including such segments could increase my sample size and therefore more precisely estimate an upper bound. Ongoing work investigates such markets and approaches.

1.5.6 Model Fit

Costs, Premiums, Shares

I now report the fit of my model in predicting insurer costs, premiums, market shares, and products offered. Figure 9 reports the observed total hospital costs versus the predicted hospital costs for the year 2013, in millions of dollars.

The model fit for total hospital costs is very good, predicting total spending almost exactly for each insurer. Tufts and Harvard's broad network plans each spend approximately $25 - $30 million dollars per year, whereas the model predicts they spend $21 million (for Harvard) and $29 million (for Tufts). The model under-predicts each insurer's spending by a small amount. This is likely due to the fact that the prediction smooths over outliers that may drive up observed insurer costs, such as patients in long-term care, patients in comas, etc. Since the model aggregates each hospital diagnosis into 18 possible categories, some of these extreme outliers will not be predicted accurately.

Figure 10 reports the model fit for premiums per-member-per-month for the year 2013. The dark green, leftmost bar displays the observed premiums for each insurer offered in the GIC, where second bar presents estimates for predicted PMPM premiums, assuming a fixed 10% markup over marginal costs (the predicted premiums from equation 13). Here, the model predicts the GIC plan premiums quite well. All of the broad network plans are predicted within a few dollars of error. In addition, Harvard and Tufts' narrow networks are predicted quite accurately as well. The largest differences are with Fallon's narrow network and Neighborhood Health Plan, which are under-predicted by about $15. This could be due to the fact that while Harvard and Tufts are of-

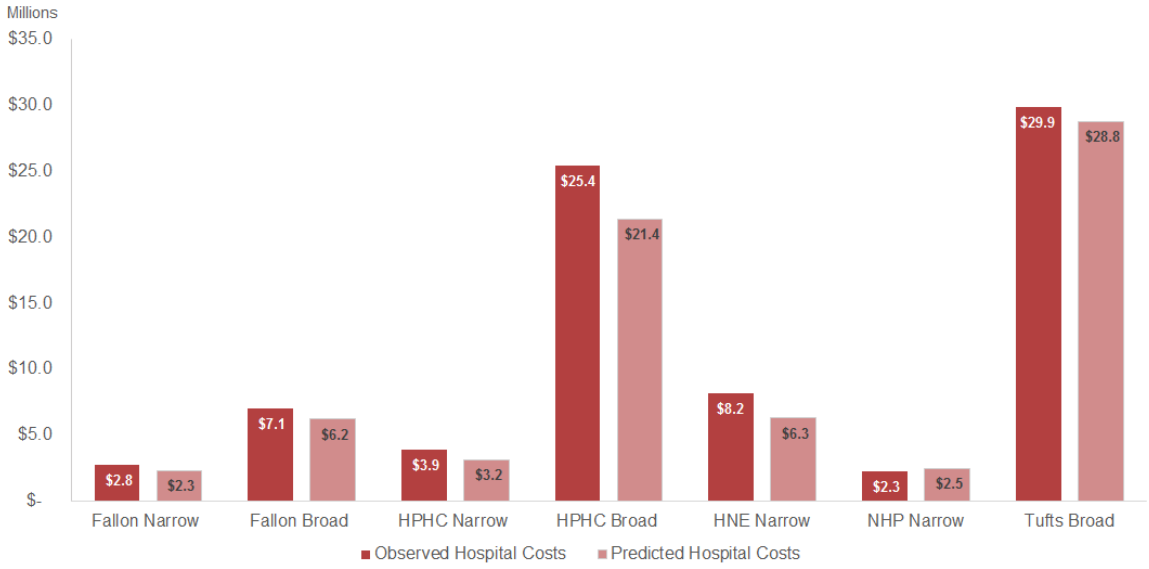


Figure 1.9: Model Fit for Hospital Costs, 2013

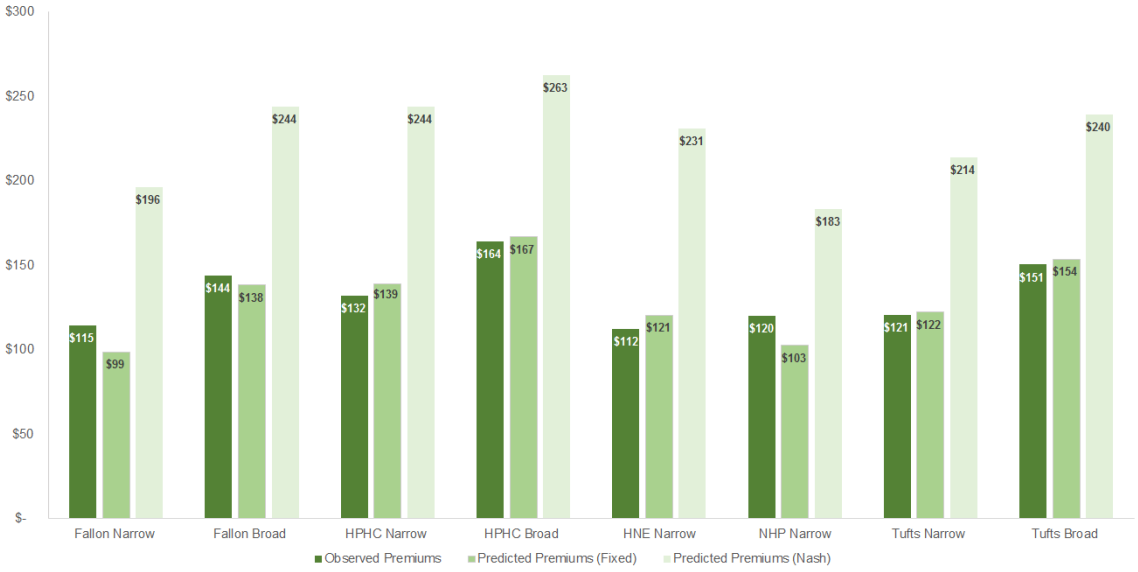


Figure 1.10: Model Fit for Premiums ($PMPM), 2013

fered as self-insured plans–meaning the GIC sets its own premiums–Fallon and NHP are both contracted on a fully-insured basis, meaning the insurers fully set their own premiums, and therefore could potentially price their narrow network products at a higher markup above marginal costs.

In order to investigate this further, I also report predicted premiums, but assuming a Nash-Bertrand pricing model (equation 12) rather than a fixed markup over marginal costs. In this scenario, represented by the third bar in Figure 10, the model over-predicts the PMPM premiums for each insurer by a considerable amount. Indeed, some premiums that were perfectly predicted in the fixed-markup model are now almost twice as high. This suggests that either the GIC has significant bargaining power against insurers such that they can negotiate premiums down considerably from what would be predicted by a Nash-Bertrand model, or that the MLR regulations bind for this market. Moreover, the relative differences between the plans within-firm have changed. Under the observed prices, each insurer prices its narrow network approximately 20% under its broad network. Under Nash-Bertrand, both Harvard and Tufts price their narrow network plans at 10% or less than its broad network. However, Fallon now prices its narrow network plan at about 20% less than its broad network plan, matching the relative difference between products better than the fixed markup model. This lends credence to idea that fully-insured products might be pricing at closer to Nash-Bertrand, whereas self-insured products at a fixed-markup over marginal costs. It is therefore possible that the true pricing equation lands somewhere between fixed markups and Nash-Bertrand.[21]

Figure 11 reports the model fit for market shares in 2013. The bars represent

[21] Future iterations of this paper will explore alternative pricing models that differentiate between how the premiums of self-insured products and fully-insured products are set.

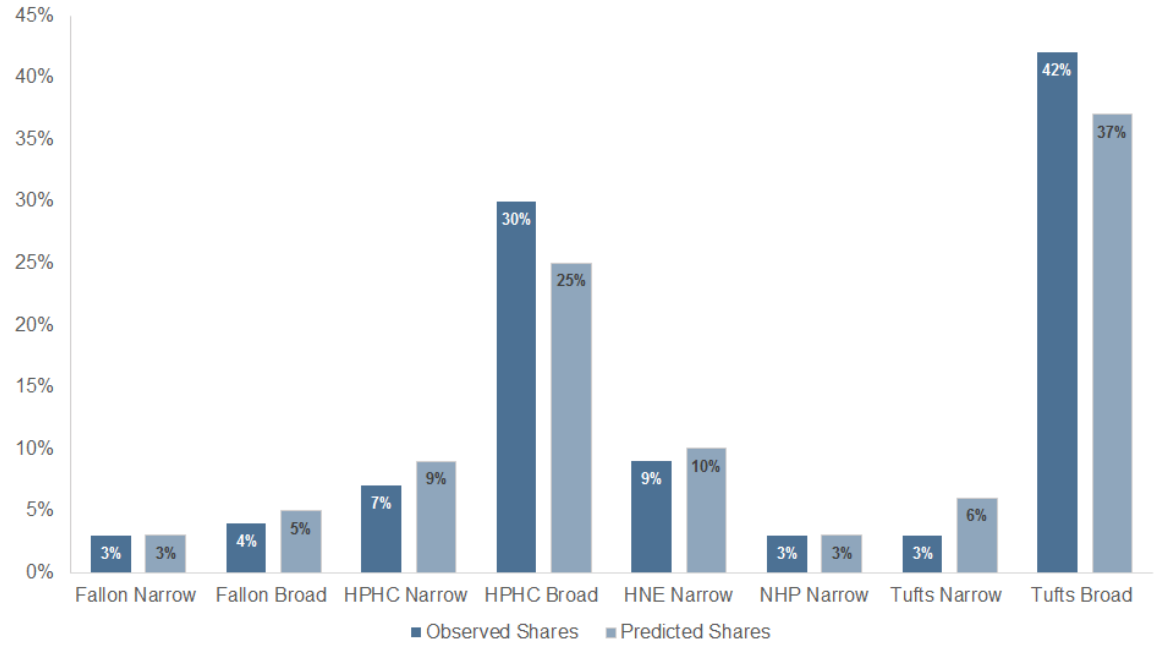


Figure 1.11: Model Fit for Plan Market Shares, 2013

the observed market shares versus predicted market shares, assuming insurers fixed markups over marginal costs. For the most part, the model predicts market shares really well. The major difference comes from the fact that the model slightly under-predicts the market share of Harvard broad and Tufts broad and slightly over-predicts the market share of Tufts narrow. Harvards narrow plan has an observed market share of 30% in 2013, while the predicted share is 25%. Similarly, Tufts' broad plan has an observed market share of 42% and a predicted share of 37%. Meanwhile, Tufts' narrow plan has an observed share of only 3%, but a predicted share of 6%.

Product Choice

I now describe the procedure used to implement the first stage of the model: the endogenous product and network choice. In order to reduce the dimension-

ality of the computation, rather than having the GIC choose between a set of 12 potential narrow network products for every insurer, I restrict the GIC to choose between offering 4 different networks for Fallon, Harvard, and Tufts. These are a "very narrow" network, which is equivalent to the observed network of Fallon Direct, a "narrow" network, which is equivalent to the observed network of Tufts Spirit, a "medium network" equivalent to Harvard Primary Choice, or a "broad" network, which covers each hospital and physician practice in the market. For Health New England and Neighborhood Health Plan, I only allow the GIC to choose to offer their observed network or not to offer these insurers at all. This leaves a possible set of 14 products for the GIC to offer. I proceed computing the equilibrium networks offered in a series of steps:

1. Construct a vector of $2^{14} = 16,384$ possible equilibria combinations of products offers
2. For each vector, compute the expected utility of the hospital and physician networks for each member, EU_{ijtH} and EU_{ijtS}, and then a household utility, EU_{Ijt}, for each offered product's network using the estimates from the provider demand model
3. Compute the predicted medical costs, mc_{Ijt} for each household if they enrolled in any of the offered products, using the negotiated price construction
4. Compute the expected market shares and premiums, s_{Ijt} and r_{Ijt}, for each household in each offered product, using the results from the insurance plan demand model and assuming a fixed markup over marginal costs pricing equation

5. Compute the estimated consumer surplus, $CS(\delta_{Jt})$, and total outlays for the GIC under the current product offered
6. Subtract the midpoint of the fixed cost bounds from the GIC's objective function, W_t, for each product offered.
7. Repeat this procedure for each vector of possible equilibria, and take the max of all the computed welfare functions.

I report the estimates of GIC's network choices under various scenarios. The first (the base scenario) assumes that if the GIC removes a particular plan from its menu, enrollees would pay a switching cost and forced be forced to re-enroll in a new plan. The second assumes that the GIC can modify the networks of any product within-insurer without removing that product (and therefore without consumers having to incur a switching cost). In other words, if the GIC removes Harvard's narrow network plan but replaces it with a network equivalent to Tufts' narrow plan, I assume this is equivalent to simply changing the Harvard narrow network. Under this scenario, I assume if the GIC removes one narrow product, but replaces it with two different narrow products (offered by the same insurer), enrollees will be automatically re-enrolled to the product with the closest network size. The third scenario assumes that there are no fixed costs in offering additional plans. Finally, the fourth scenario assumes that the GIC does not consider enrollee inertia/switching costs in its objective function.

In all scenarios, I make a "no uninsurance assumption." That is, between all the plans offered on the GIC, every single member must have access to a plan. To do so, I leverage data on which available networks are currently offered in which Massachusetts counties. I then assume that any counterfactual network must be offered in the same counties as where those networks are currently

observed. For example, if the GIC chooses to reduce Harvard's narrow network plan to be the size of Fallon's narrow network plan, then the new network can only be offered in counties in which Fallon's plan is offered. Between all the offered plans, individuals in all counties must be offered insurance.

Table 10 reports the equilibrium predicted products/networks offered from the first stage of the model, where the GIC's product choice is endogenized. Column 3 reports the observed products offered in 2013, Column 4 reports the predicted products offered under the base scenario, Column 5 reports equilibrium products assuming enrollees get re-enrolled in the closest narrow plan of the same insurer, Column 6 reports product choices assuming that the GIC has no fixed costs for offering any additional plans, and Column 6 reports the equilibrium offered products assuming that the GIC does not consider enrollee inertia as part of the total welfare function when making product choice decisions.

The predicted networks offered in Column 4 match the observed products very well. It correctly predicts the fact that Harvard Pilgrim and Fallon offer both a "narrow" and "broad" product, and also correctly predicts that HNE and NHP continue to offer their respective networks. The only difference in the predicted plans from the observed plans is the fact that the GIC chooses not to offer Tufts' narrow network plan. This is due to the fact that Tufts' narrow plan only had about a 3% market share, even after the premium holiday was instituted in 2012. The model therefore has difficulty in rationalizing the GIC's choice of offering such a product, and instead predicts that the fixed costs of offering the plan outweigh the possible price reductions and marginal social welfare gains of offering it.[22]

[22]It may be the case that the GIC is more forward-looking than the static model I have developed predicts. If the GIC, in considering plan design decisions, weighs future social welfare changes, this might explain its decision to offer Tufts' narrow network plan, despite its initial

Table 1.10: Model Fit: Equilibrium Products and Networks 2013

Insurer	Network	Observed	Pred.	Pred(NN Ch)	Pred. (No FC)	Pred. (No Inert)
Fallon	Very Narrow	x	x		x	x
Fallon	Narrow				x	x
Fallon	Medium			x	x	
Fallon	Broad	x	x	x	x	
HPHC	Very Narrow				x	
HPHC	Narrow				x	
HPHC	Medium	x	x	x	x	
HPHC	Broad	x	x	x	x	
HNE	Narrow	x	x	x	x	x
NHP	Narrow	x	x	x	x	x
Tufts	Very Narrow				x	
Tufts	Narrow	x			x	
Tufts	Medium				x	
Tufts	Broad	x	x	x	x	

Notes: GIC observed and predicted products offered under various model assumptions. "NN CH" refers to the assumption that if a narrow network product is removed, the GIC would automatically enroll consumers in the closest network plan by the same insurer (i.e. these consumers would not exhibit any switching costs so long as another insurer was present). "No FC" refers to a model without any fixed costs. "No Inert" refers to a model where the GIC does not consider enrolle inertia when making product choice decisions.

The model likewise predicts plan choices pretty well when enrollees are allowed to stay in narrow network plans that are altered. Interestingly, the GIC continues to offer Harvard's exact network under this scenario (the "medium" network), but induces Fallon to switch from a "very narrow" to the same "medium" network. The stems directly from the willingness-to-pay estimates from the plan demand model, showing that most Fallon enrollees would be willing to pay a significant premium in order to move away from its narrow network (while the same is not true of Harvard).

To further investigate the role that fixed costs play for employer product offerings, Column 5 reports the equilibrium products offered assuming that there are no fixed costs. Unsurprisingly, this predicts that the GIC would offer every

poor performance. Indeed, this would also explain the very reason the GIC offered the premium holiday as well. In the future, I will look at alternative models where I allow the GIC to have a more dynamic objective function.

available product on the market, as each would generate some welfare gains and reduce costs. This suggests that fixed costs do play an important role in keeping the GIC from offering multiple products.

Finally, to demonstrate the role that inertia plays in determining employer product choices, I report the results of the equilibrium model assuming that in choosing plans and networks, the GIC does not consider inertia in its overall objective function. In other words, the GIC accurately predicts consumer enrollment patterns given the extent to which inertia exists in the market, but when measuring consumer surplus, it considers inertia as irrelevant. In effect, this would be similar to an assumption that if a plan is removed from the GIC product line, the GIC would automatically reenroll all members of the removed product into their preferred choice without having the consumer make an active choice. In this model, the GIC only cares about enrollee welfare as it pertains to the tradeoff between premiums and networks, not whether the individual remains on the same plan as the previous year.

In this scenario, the model predicts that the GIC drops every single plan with the exception of two Fallon plans (a "very narrow" network and a "narrow" network), HNE, and NHP. The reason for this is that Fallon is able to secure better rates for Reliant Medical Group, the dominant medical practice on the very narrow network, than both Harvard and Tufts if they offered the same network. Fallon, in other words, has a comparative advantage, particularly in Central Massachusetts due to its favorable rates with the medical groups in that area. Therefore the largest cost savings come from moving consumers onto Fallon's plan. However, offering this plan by itself would leave a majority of consumers at a huge welfare loss, since most do not live in Central Massachusetts.

To counteract these welfare losses, the GIC would need to ensure that there are providers located near a majority of its customers. NHP has a relatively broad network compared to the other "narrow network" plans, while also having lower prices than its competitors, particularly for Partners-affiliated hospitals and medical groups.[23] While its brand effects are lower than that of Tufts and Harvard, its presence would still generate enough welfare gains through its broad network and low prices to warrant it being offered as part of the GIC's product menu. Finally, the GIC continues to offer HNE, largely to satisfy the "no uninsurance" assumption. This is due to the fact that neither Fallon nor NHP cover the Baystate health system, the dominant system in Western Massachusetts. HNE, meanwhile, does offer comprehensive coverage in Western Massachusetts and is able to negotiate favorable rates with Baystate.

Therefore, with no inertia, the GIC would only keep the minimum number of plans necessary in order to ensure full coverage across Massachusetts, while reducing costs by favoring the insurers able to negotiate low rates in each region. NHP, through being vertically integrated with Partners, is able to keep prices low in Eastern Massachusetts; Fallon is able to keep rates low in Central Massachusetts; and HNE is able to keep rates low in Western Massachusetts. Under this scenario, approximately 70% of the market would enroll in NHP compared to about 30% enrolling in Fallon or HNE (reflecting the fact that most consumers live in Eastern Massachusetts). This suggests that while Harvard and Tufts may be the dominant insurers on the GIC, they are likely rewarded due primarily to inertia in plan choices, rather than an actual valuation of their networks or brands, and that the high costs they charge may not be worth their

[23]This is largely due to the fact that in 2013, Partners Health system actually purchased NHP. This explains the lower negotiated rate that NHP has with Partners physicians and hospitals and its ability to reign in premiums.

presence as part of the GIC's menu. Indeed, assuming a world where this is no welfare loss due to switching costs, the GIC would have little reason to offer these products.

1.6 Counterfactuals

In the previous section, I showed that dominant plans are rewarded largely due to plan inertia, implying that employers may be offering plans that may be overpriced relative to how consumers value the networks of those plans. I now use the results from the estimates above to conduct several policy counterfactual analyses aimed at reducing these inefficiencies. In particular, I ask what the effects on GIC plan choices would be following implementation of a "Cadillac Tax" on expensive health plans. I also ask how high a tax would need to be in order to induce the GIC to make changes in their product menu that may bring about social surplus. The original tax, as written in the ACA, was scheduled to begin in 2018, with a 40% levy on thresholds of $10,200 for individual plans and $27,500 for family coverage. The tax has been delayed to 2020. Projecting this back to 2013, and assuming an annual health care inflation of approximately 6%, leaves a 2013 value for single-premiums at approximately $6,600 per year. I conduct a slightly modified counterfactual from the actual proposed tax, setting the single-premium threshold at an even $6,000 per year, or combined with the 25% enrollee share of premiums on the GIC, $125 per-member-per-month.[24] I then run the equilibrium model as I did above through several different rates charged beyond this threshold: 20%, 40%, 60%, and 80%. I then use the model to predict counterfactual product offerings, networks, medical spending, pre-

[24] In practice, running the analysis assuming a threshold of $6,600 changes very little.

miums, consumer welfare, and total welfare that result from the tax. The preferred specification used for all of these analyses is the scenario in which the GIC does take into consideration enrollee inertia in its welfare calculus when setting plans. However, I also allow the GIC to re-enroll consumers into narrow network plans within the same insurer (Column 5 from Table 10).

Three assumptions are necessary to proceed with the analysis. First, I assume that the tax is paid entirely by the employer and not by the insurer. The rules of the excise tax are written such that plan sponsors are responsible for the tax. As the GIC offers a mix of self-insured and fully-insured products, this implies that insurers of the fully-insured plans bear the tax, while the GIC bears the tax for the self-insured products. Since each of the products likely to be subject to the tax are the large, broad-network plans, and the GIC self-insures these, I assume that tax is levied on the GIC rather than the insurer. Second, I assume that additional costs from tax cannot be passed on to the employees through higher premiums, higher enrollee-shares of premiums, or higher cost-sharing, but mainly through product and network choice. The assumption about premiums is likely valid, as most employers are prohibited from charging premiums that are too high in excess of medical payments due to MLR rules. Simulating employer choice of offering higher cost-sharing (in the form of deductibles, for instance) would require a model that incorporates enrollee response to deductibles, which is not possible in my setting, since the GIC deductibles are the same across plans, meaning there is no variation with which to estimate demand.[25] Finally, modeling enrollee-share of premiums involves estimating the relative weights that the employer places on consumer welfare versus outlays

[25]Future iterations of this paper will look at other employers who offer plans with multiple deductibles in order to model employee response to cost-sharing, and thus employer choice of not only networks, but plan deductibles.

through medical spending and premiums. Future analysis will conduct this counterfactual as well, and treat the enrollee-share of premiums as an equilibrium object.

The final assumption is for the consumer surplus measure. In the reported estimates, I assume that when considering which plans consumers will enroll in and which products to offer, the GIC takes inertia into account to accurately predict market shares and to provide value to consumers that include their preference to stay on similar plans. However, from a government or public policy perspective, it is unclear that inertia should be considered in measures of social welfare or whether it reflects passivity or inattentiveness. Therefore, when evaluating the tradeoff between decreases in spending and consumer surplus, I report surplus measures that set inertia to zero. These are similar assumptions that have been made in the literature many times (Polyakova, 2016; Shepard, 2016; Handel & Kolstad, 2015; Abaluck & Gruber, 2016).

Formally, consumer surplus in counterfactuals will be given by the modified equation:

$$CS'(\delta_{Jt}) = \sum_I \frac{1}{\alpha_I} log\left(\sum_j^J exp(-r_{Ijt}\alpha_I + EU_{Ijt}\beta_1 + \eta_j)\right) \tag{1.24}$$

Overall social welfare then is giving by:

$$W' = CS'(\delta_{Jt}) - \sum_I \sum_j (1-\tau) s_{Ijt} R_{Ijt} - \sum_j FC_j \tag{1.25}$$

Table 11 reports the results of the counterfactual exercises. A 20% and 40% tax above the $6,000, while binding for Harvard and Tufts' broad plans, are not

Table 1.11: Counterfactuals: Cadillac Tax 2013

Insurer	Network	Observed	20%	40%	60%	80%
Fallon	Very Narrow	x			x	x
Fallon	Narrow					
Fallon	Medium		x	x		
Fallon	Broad	x	x	x		
HPHC	Very Narrow					
HPHC	Narrow					
HPHC	Medium	x	x	x		
HPHC	Broad	x	x			
HNE	Narrow	x	x	x	x	x
NHP	Narrow	x	x	x	x	x
Tufts	Very Narrow					
Tufts	Narrow	x				
Tufts	Medium					
Tufts	Broad	x	x	x		
Δ Spending (PMPM)			14.61	29.20	-76.05	-76.05
Δ Spending (%)			0.04	0.08	-0.21	-0.21
$\Delta CS(\delta_{jt})$ (PMPM)			0.00	0.00	-58.13	-58.13
$\Delta CS(\delta_{jt})$ (%)			0.00	0.00	-0.47	-0.47

Notes: Counterfactuals of a proposed "Cadillac tax" on expensive health plans at various rates past a $6,000 annual threshhold.

substantial enough to cause the GIC to alter its product line. Instead, the tax merely causes a slight decrease in social welfare, as GIC spending goes up per member as a result of the tax. A 60% tax, however, causes the GIC to alter its existing networks considerably. The first thing to note is that under a tax, the GIC drops each of its broad network plans by every insurer. The reasoning behind this is selection: if the GIC responded to the tax by dropping Harvard's broad plan but not Tufts, Harvard's enrollees who lived closer to expensive physicians and hospitals would simply switch to Tufts. While this strategy would allow the GIC to save on annual fixed costs, the marginal costs (premiums paid to insurers) would barely be affected and consumer welfare would decrease substantially by the loss of Harvard's broad network plan. Therefore in order to avoid the tax, the GIC would have to drop *every* broad plan.

The second result is that the GIC moves to an equilibrium in which only three products are offered. The new equilibrium products are the Fallon "very narrow" network, the HNE network, and the NHP narrow network. These products remain the same for any tax level set about 60%. It is interesting to note why the GIC moves to this particular equilibrium rather than one in which Harvard or Tufts offers narrow network plans instead of the Fallon, HNE, or NHP. The reason is similar to that of the "no inertia" model stated above. Indeed, the optimal strategy for the GIC under a scenario where the tax both binds and is large enough to overcome substantial enrollee plan persistence would be to offer the minimum number of plans necessary that just span all the counties in Massachusetts. The HNE network offers access to Baystate Health in Western Massachusetts, Fallon offers the best negotiated rates with Reliant Medical group in Central Massachusetts, and NHP offers the best rates with Partners in Eastern Massachusetts. Between these three networks alone, the GIC is able to achieve the most savings in premiums. Although welfare gains could be achieved by offering additional narrow network plans, these gains are not substantial enough to offset the fixed costs of doing so.

At the new equilibrium, I find that any tax rate above 60% decreases net GIC spending by approximately $76 per member per month, or a reduction in spending of approximately 21%. Consumer surplus (less inertia) also decreases under these scenarios, both from having a smaller choice set of plans (and less favorable insurer brands) and from having a smaller scope of providers to choose from within these plans. However, the surplus loss is about $58 per member per month (approximately a 47% decrease), less than the decrease in overall spending that would result from the product switch. This suggests that a tax can achieve social welfare gains of approximately $18 per member per month,

though the tax rate would need to be quite high in order to achieve these gains.

1.7 Conclusion

In this paper, I estimated a model of supply and demand for health insurance plans in order to assess employer incentives to design and offer "narrow-network" insurance plans to their enrollees. I find that despite the fact that consumers in employer markets are commonly thought to be less price-sensitive and more willing to pay for broad networks than consumers shopping among individual Exchanges, much of the demand for broad-network plans derives from enrollee inertia, or preference to remain on the same plan as the previous year. This inertia not only explains why enrollees select into broad network plans (which have been offered on the market for longer periods of time), but employer choice of products offered as well, leading to increases in health care costs in both hospital and physician markets. I find that a hypothetical "Cadillac Tax" on plans in excess of the thresholds set by the Affordable Care Act will offset this inertia and induce employers to drop their broad network products in favor of limited network designs.

Overall, this paper contributes towards our understanding of what consumers in employer markets value in their choice of plan, and how employers select plan characteristics. While narrow networks grow more prevalent in the individual exchanges, they have yet to gain a strong hold in employer markets, where most people obtain their health insurance. Policymakers have been debating whether such plans are viable in the employer market or whether their widespread adoption would lead to substantial losses. Using detailed data on

provider networks, the results of this paper suggest that moving towards narrow networks beyond the individual market would have the potential towards bringing about social welfare gains, by achieving considerable reductions in health spending that exceed the welfare loss of choice.

This leaves the question of how policymakers might incentivize the move towards narrow networks. The Cadillac Tax is a step in the direction of incentivizing insurers to adopt more value-based design, but knowing how employers aggregate consumer preferences and how they select products for their enrollees is of paramount importance to evaluating the effect of policy. By modeling the employer choice under various scenarios and by estimating the fixed costs of offering multiple plans, I am able to evaluate the likely impact of the tax. I am also able to demonstrate how high a tax would need to be in order to bring about changes in product portfolios, medical spending, and consumer welfare.

There are some extensions that can be applied to this work. I only model the plan offer decisions of one large employer. The effect of the "Cadillac Tax" would differ across firms depending on the enrollee composition of those firms, the size of the firms, and the bargaining power that the firms have over insurers to set premiums. If, for instance, insurers have greater leeway to set high premiums in other firms, then offering more choice may be an effective way to combat the tax, through inducing insurers to compete for lower premiums. Incorporating multiple employers with firm heterogeneity is a natural next step for this paper

My model also abstracts from insurer-provider bargaining. One effect of the tax might be to induce hospitals and physician practices to reduce their

negotiated prices with insurers for fear of no longer being offered by the employer. Future work will incorporate a stage in the model where insurers and physicians/hospitals bargain over negotiated rates, while simultaneously the employer chooses a set of products endogenously. Finally, future work can also incorporate non-network measures that employers might take in response to a tax (including increasing enrollee premium contributions or cost-sharing).

CHAPTER 2

HAS MEDICAL SPENDING BEEN WORTH IT? ESTIMATING QUALITY ADJUSTED PRICE INDICES FOR COLORECTAL CANCER TREATMENTS[1]

2.1 Introduction

In 2015, United States health care spending reached $3.2 trillion, or about 18% of GDP, about twice the share it was in 1980.[2] Medical technology, broadly defined as the products and services patients receive when treated, is widely accepted as the engine behind such increases in medical spending. For example, Newhouse (1992) estimated that approximately one-quarter to one-half of the growth in medical costs in the United States between 1960 and 2007 was due to changing medical technologies (Newhouse, 1992).[3] A more recent study attributed approximately a quarter to a half of the increase in health spending since 1960 to technological change in the health care (Smith, Newhouse, & Freeland, 2009).

Despite these increases, there is also evidence that medical care in the US

[1]THIS CHAPTER WAS COAUTHORED WITH CLAUDIO LUCARELLI AND SEAN NICHOLSON

[2]See the National Health Expenditure Accounts from CMS: https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html

[3]Newhouse estimates the proportion of the growth in medical spending accounted for by the aging of the population, improved health insurance based upon the RAND health insurance experiment price elasticity, and rising income based on the RAND income elasticity. Medical technology is the residual once the factors above have been accounted for. Finkelstein (2007) argues that the aggregate, market-wide effects of health insurance on spending are larger than those derived from individual choices in the RAND experiment. She estimates that about one-half of the growth in US hospital spending between 1950 and 1990 was due to the spread of health insurance, which may indicate that medical technology accounts for less than one-half of the growth in medical spending

has improved substantially over the past 30 years for certain health conditions such as cardiovascular disease. For example, the life expectancy of a 45-year old today is 4.4 years longer than it was in 1950 (Cutler, 2004).[4] Murphy and Topel (2006) estimate that the total rise in life expectancy in the US between 1970 and 2000 increased national wealth by $3.2 trillion per year, which is approximately equal in value to one-half of GDP (Murphy & Topel, 2006).

An important policy issue, therefore, is whether the value of new medical technology exceeds its cost. More generally, if new technologies are priced higher than the technologies they replace and consumers value the superior health outcomes that can now be produced, are medical prices rising or falling once one correctly accounts for quality? If the government and private health insurers believe new medical technologies are causing quality-adjusted prices to increase, on average, they may make it more difficult for new products to reach the market and/or reduce payment for new products, which would dampen the financial incentives to innovate. Conversely, if public and private payers believe technology is helping drive down quality-adjusted prices, they are likely to maintain or enhance incentives to innovate.

There are several important empirical challenges in measuring a medical price index (Berndt et al., 2000). One challenge is that quantities change as new goods displace old goods in the market. Therefore, a price index needs to allow the bundle of goods to change over time as well as the weight on each component in the bundle. Another challenge is that often transaction prices are not observed. Most publicly available data on health care prices rely on "list prices," i.e. the prices that pharmaceutical firms, medical firms, physicians, or hospitals

[4]Cutler (2004) estimates that two-thirds of this increase was due to changing medical technologies, primarily for the treatment of cardiovascular disease. The remaining one-third was due to behavioral changes.

charge for a product or service. Health insurers usually receive discounts off of list prices. In addition, due to health insurance, most consumers do not face the full price of medical care and seek to consume beyond the point where the marginal value of care equals the full price. Patients also rely on physicians to provide information regarding the value of medical goods and services. The implication is that consumer purchases in the medical market will not necessarily reveal their marginal valuation of a good or service. Finally, prices increases are often accompanied by quality improvements. Therefore, perhaps the greatest empirical challenge for constructing meaningful health care price indices is how to account for the changing quality of medical products and services. For example, life expectancy and mortality rates are often used as proxies for quality, but in some medical settings improvements in the quality of life are more important than the length of life. Information on such measures is often difficult to obtain.

Due to these considerable empirical challenges, there have been relatively few studies of whether medical prices are rising or falling once one takes into consideration the attributes of the new products and consumers valuations of those attributes. Existing studies can be placed into three groups based on the method used. One method to account for changing quality is to measure actual changes or expected changes in health outcomes (e.g., life expectancy or remission of a depressive episode) due to the adoption of new medical technologies, monetizing these improvements based on separate estimates from the literature (e.g., $100,000 per year of life), and subtracting the value of the health gains from the rising costs. Cutler et al. (1998) show that the life expectancy of heart attack patients increased by eight months between 1984 and 1991. The value of per-patient expected longevity ($11,100) increased three times more than treat-

ment costs ($3,600) during this time period in real terms, which implies that the quality-adjusted price index fell by about one percent annually (Cutler, McLellan, Newouse, & Remler, 1998). Berndt et al. (2002) conclude that the real cost of treating major depression decreased by about two percent per year between 1991 and 1996 once one takes into account the probability that a patients depression goes into remission (Berndt, Bir, Busch, & Frank, 2002). Eggleston et al. (2011) conclude that the cost of treating diabetes declined between 1999 and 2009 once one accounts for the value of reduced coronary heart disease mortality (Eggleston, Shah, Smith, & Berndt, 2011).

Two studies estimate hedonic price indices by regressing prices on objective measures of product attributes and time indicator variables. Cockburn and Anis (2001) find that rheumatoid arthritis prices rose over time, even after accounting for each drugs expected efficacy and toxicity (Cockburn & Anis, 2001). Dunn (2012) reports that the prices of cholesterol drugs fell by 30 percent between 1996 and 2007 once one accounts for each drug's expected change in a patients bad and good cholesterol levels (A. Dunn, 2012). The two studies closest to our approach also construct price indices, or the components of an index, based on discrete choice demand models. Estimating a nested multinomial logit model, Trajtenberg (1989) shows that the social returns to CT scanner innovation exceeded the R&D costs, especially when the first products were introduced (1972-1978) (Trajtenberg, 1989). Dunn (2012) uses a logit model to estimate patients demand for cholesterol drugs and concludes that quality-adjusted prices fell by about 25 percent between 1996 and 2007, similar to the results from the hedonic model.

The pharmaceutical industry is the source of considerable innovation in

medical care. The pharmaceutical industry invests over $40 billion per year in research and development, which represents about 16 percent of the industrys revenue. Most new pharmaceuticals are priced higher than the treatment methods they replace, and this has certainly been the case with colorectal cancer drugs. The average price of providing a colorectal cancer patient with a 24-week drug regimen increased from $127 in 1993 to $36,300 in 2005.[5] There has been substantial innovation in the treatment of colorectal cancer patients over the last decade. Five new drugs were approved by the Food and Drug Administration (FDA) for the treatment of colorectal cancer between 1996 and 2004, and these drugs collectively had an 86 percent market share by the third quarter of 2005.[6]

We examine whether the substantial increase in spending associated with pharmaceutical innovation in the treatment of colorectal cancer has been worth it. Specifically, we estimate a price index for colorectal cancer drugs for each quarter between 1993 and the first half of 2005 that takes into consideration the quality (i.e., the efficacy and side effects as reported in clinical trials) of each drug on the market and the value that oncologists attach to drug quality. We estimate a quality-adjusted price index using methods developed by (Berry, 1994), (Berry, Levinsohn, & Pakes, 1995)and (Nevo, 2001, 2003). These techniques have been used to estimate the welfare effects of new automobiles (Petrin, 2002), computers (Song, 2007, 2008),breakfast cereal (Nevo, 2003), and a host of other markets, but we are aware of only two studies applying such discrete choice methods towards the construction of price indices in the pharmaceutical market, as mentioned above. We improve on these studies both by applying newer methods for demand estimation not used in prior work and

[5]Most colorectal cancer patients are treated with a combination of two or more drugs, which we refer to as a regimen

[6]Market share data are from IntrinsiQ.

in focusing on transaction-level data for physician purchases of pharmaceuticals (A. Dunn, 2012). In addition to constructing price indices using hedonic price regressions and logit demand models, we also estimate a full random-coefficients model with consumer (physician) heterogeneity, and a pure characteristics model of demand. The pure characteristics model, to our knowledge, has only been applied empirically in a select few studies on computer markets (Song, 2007, 2008; Nosko, 2014). We show that by shedding the idiosyncratic logit error term that is a workhouse of traditional demand models, this model produces much more realistic substitution patterns as new colorectal cancer treatments enter the market and replace existing regimens.

The first step in constructing the price index is to estimate oncologists demand for colon cancer drug treatment regimens, which we argue is a function of the observed and unobserved quality of each regimen, as well as the price a physician must pay to acquire the regimen. An observation in this estimation is the market share of each regimen for each quarter. The second step is to calculate the equivalent variation (EV) between each adjacent pair of quarters. The EV between period $t-1$ and t is the change in spending required to achieve physicians utility in period t relative to $t-1$, taking into consideration that both the quality of the drug regimens on the market and the prices of those regimens may have changed between these two periods. Positive values imply that the value of the drugs is increasing over time by more than their prices, whereas negative values imply the opposite. The third step is to translate the EV into a quality-adjusted price index by calculating the change in drug prices that is consistent with the welfare effect captured by the EV.

We generate several indices for purposes of comparison. First, we estimate a

"nave" price index that merely reports the mean price of colon cancer regimens in each quarter, relative to the first quarter of 1993, without any adjustments for regimen attributes but allowing the market shares (or bundle weights) to change. The naive index will be based on the prices physicians pay to acquire each regimen and the market share of each regimen. Comparing the quality-adjusted price indices to the naive index illustrates the importance of accounting for the changing quality of pharmaceutical products. Second, we estimate a hedonic price index by regressing prices on product characteristics and quarter indicator variables. This regression, which controls for changing product attributes by means of a reduced form projection of markups on the characteristic space, is the traditional method used to account for changing attributes and the introduction of new goods. Third, we estimate a logit model of physician utility maximization that takes into account the equivalent variation described above. Fourth, we extend the logit model to incorporate physician heterogeneity in preferences for attributes of the treatment regimens by estimated a full random-coefficients model. Finally, we estimate a pure characteristics model of demand in order to test whether the price index changes when removing the idiosyncratic error term implicit in the logit and full random-coefficients models and allowing for more realistic substitution patterns between products.

We find that the naive price index greatly overestimates the price increase. While the quality-adjusted price indices show increases over the 13-year period we study, these increases are far more modest than the naive index would suggest. Specifically, while the (log) naive price index increases about 100% during our sample period, the logit index increases by only 4.5% and the pure characteristics model increases by about 18%. All of the new drugs in our sample period are still under patent protection. Presumably consumers would capture

more of the surplus once generic drugs enter the market.

2.2 Background

2.2.1 Medical Treatment of Colorectal Cancer

Colorectal cancer is a good health condition for studying the welfare effects of medical innovation because it is a common health condition, the majority of patients today are treated with drugs that did not exist two decades ago, and treatment costs are rising rapidly. According to the National Cancer Institute (NCI), approximately 133,000 patients are diagnosed with colorectal cancer in the United States per year, resulting in about 50,000 deaths annually from the disease. This places colorectal cancer as the fourth most common cancer based on number of new patients, after breast, prostate, and lung. It is estimated that people born today will have a 4.5 percent chance of being diagnosed with colorectal cancer over their lifetime. According to the NCI, between 2005 and 2011 colorectal cancer patients had a 65 percent chance of surviving for five years. The probability a patient will survive for five years ranges from 90 percent for those diagnosed with Stage I cancer to 13 percent for those diagnosed with Stage IV (or metastatic) cancer.[7]

Colorectal cancer is treated differently depending on the stage of the disease at diagnosis. Most patients with a Stage I, II, or III disease at diagnosis will have the tumor removed surgically, or resected, in a hospital. The National Com-

[7]Cancers are classified into four stages, with higher numbers indicating that the cancer has spread to the lymph nodes (Stage III) or beyond its initial location (Stage IV).

prehensive Cancer Network (NCCN) recommends that patients with Stage III disease receive six months of drug treatment following the resection; they do not recommend drug treatment for Stage I patients; and they recommend that Stage II patients consider drug treatment after discussing the potential benefits and side effects with their physician. Most Stage IV patients are treated with drugs, either to shrink the tumor so that it can be resected, following a resection, or when the tumor is unresectable in order to lengthen and improve the quality of life. In this paper we do not model the process by which a physician and her patient decide whether or not to receive drug treatment; we examine prescription drug choices conditional on the decision to receive drug treatment.

Treating colorectal cancer is expensive. Paramore et al. (2006) report that between 1998 and 2004, which is part of the time period we study in this paper, newly-diagnosed metastatic colorectal cancer patients in the United States received $97,000 of additional medical treatment in the 13 months following their diagnosis relative to a similar (based on age, gender, and geography) group of people (Paramore, Thomas, Knopf, .s Cragin, & Fraeman, 2006). Most of the spending occurred for hospital care ($37,400, on average), including the surgical resection, and physician visits ($34,600), including drugs and their administration.

In this paper we present a price index for drug treatment only, not for the total cost of treating colorectal cancer. We focus on drug costs because this is where the treatment innovation has occurred starting in 1996, and drug innovation has been the main driver of the subsequent increase in treatment costs and improvements in health. In our data set, the average cost of providing six-months of drug treatment to a colorectal cancer patient increased from about

$100 in 1993 to over $36,000 in 2005 (see Figure 2). Hospital costs, on the other hand, did not increase substantially during this time period for colorectal cancer patients nor were there any surgical innovations that would have lengthened life or improved the quality of life substantially.

About 95 percent of colorectal cancer patients who were treated with drugs between 1993 and 1996 received a chemotherapy regimen called 5-FU/leucovorin. This regimen was inexpensive because the patents on the two component drugs (fluorouracil and leucovorin) had expired, so there were many companies offering low-cost generic versions. Between 1996 and 2004, five new drugs were approved to treat colorectal cancer. By 2005, these drugs collectively captured about 80 percent of the market. Most of the new drugs are biologics that target the inner workings of cancer cells, whereas 5-FU/leucovorin is a standard chemotherapy treatment that targets all fast-growing cells. Although the new drugs extended life in randomized clinical trials relative to chemotherapy and sometimes have less severe side effects due to their targeted nature, they are also priced much higher than standard chemotherapy drugs. Most colorectal cancer patients who are treated with pharmaceuticals receive multiple drugs in the form of a regimen rather than a single drug, similar to anti-retroviral "cocktail treatments for AIDS patients. For example, the regimen with the greatest market share in 2005 contained four separate drugs: bevacizumab, oxaliplatin, fluorouracil, and leucovorin. Most of our analysis in this paper, therefore, is conducted at the level of a regimen rather than a drug. We describe the characteristics of the various drug regimens in greater detail in the Data section.

2.2.2 Physician Behavior and Modeling Approach

In most markets one can use prices to infer consumers willingness to pay for products and services because consumers are well informed and face the full price. Several features of the health care market create challenges for estimating and interpreting a quality-adjusted price index. Physicians (or sometimes hospitals on behalf of their physicians) purchase oncology drugs from wholesalers (who previously purchased them from drug manufacturers), administer them to patients in their offices, and then bill the patients insurance company for the drugs.[8] Health insurers reimburse physicians for cancer drugs and the time required to administer the drugs to patients, and control moral hazard by requiring patients to pay a portion of the treatment cost and by restricting (expensive) treatments to patients who are likely to benefit from them (e.g., requiring a physician to receive authorization from the insurer before treating a patient).[9] Finally, patients rely on physicians to recommend a treatment. Thus, colorectal cancer patients do not face the full price of drugs and rely on physicians to articulate both the choice set and the attributes of the drugs in the choice set. And physicians have an opportunity to exploit their informational advantage and recommend particular drugs when the health insurers reimbursement amount deviates substantially from physicians drug acquisition costs.

We model the choice of a colorectal cancer drug regimen from a physicians perspective by positing that physicians choose treatments to maximize their own utility. Our key assumption in interpreting the price indices is that physi-

[8] Based on data from IMS Health, 59 percent of colorectal cancer drugs in the third quarter of 2005 were purchased by physician offices/clinics and 28 percent by hospitals. The remainder was purchased by retail and mail order pharmacies, health maintenance organizations, and long-term care facilities.

[9] Medicare patients, for example, pay 20 percent of the price of drugs administered by physicians if they have Part B coverage and no Medigap supplemental insurance.

cians act as the social planner, trading off the full price of oncology drugs against the financial value of the health benefits patients receive from those drugs. We focus on a physicians choice because they are arguably the most knowledgeable party. Physicians observe the full price of oncology drugs when they purchase them from wholesalers and store them in their offices. Physicians should be aware of the efficacy and safety of oncology drugs based on how the drugs performed in randomized controlled trials during the approval process, as well as by observing how their own patients respond to the drugs.

One criticism of our assumption is that physicians may, in fact, be insulated from the price of oncology drugs because health insurers ultimately reimburse them for drug acquisition costs. Although physicians are eventually reimbursed, they do take temporary ownership of oncology drugs. As such, physicians face the possible risk of not being reimbursed by health insurers and may incur substantial carrying costs. For example, a physician who pays $50,000 for the drugs in one patients regimen and experiences a three-month delay between when she acquires the drugs and when her practice is reimbursed by a health insurer would incur an inventory carrying cost of $625 at an interest rate of five percent. Furthermore, if physicians care about patients out-of-pocket costs they will internalize part of the price. Both carrying costs and patients out-of-pocket costs, which are often a percentage of the price the insurer pays, increase with a drugs price. Another criticism of our assumption is that if physicians earn profits on oncology drugs and profits influence prescribing decisions, this might bias the price and attribute estimates in our model. In the early 2000s, the federal government concluded that oncologists were earning profits on most oncology drugs by acquiring them for less than the Medicare reimbursement amount (General Accounting Office, 2001). At the time Medicare reimbursed oncologists

95 percent of a drugs listed average wholesale price (AWP), whereas physicians could acquire many drugs from wholesalers for substantially less than AWP. Most private health insurance companies reimbursed physicians using a similar formula, so these profits occurred for all patient types [10] Most of these profits were eliminated in 2005 when Medicare began reimbursing oncologists based on a drugs actual average selling price (ASP) rather than its list price (MedPAC, 2006).

There is evidence that the profits oncologists earned on drugs did affect treatment decisions, especially prior to 2004. These papers identify effects by examining variations in profits across oncology drug treatments due to geographic reimbursement rules (Jacobson & Newhouse, 2006),or variation within treatments over time due to the 2005 Medicare policy change (Jacobson, Earle, Price, & Newhouse, 2010; andTom Y. Chang, Newhouse, & Earle, 2013), or the entry of generic drugs (Conti, Rosenthal, Polite, Bach, & Shih, 2012). Our interpretation of these studies is that the magnitudes of the effects are small. In Jacobson et al (2006), for example, a one-standard deviation increase in reimbursement generosity is associated with an increase of about five percent in the cost of drug treatment prescribed to colorectal cancer patients. .

One factor supporting our assumption that physicians act as the social planner is that health insurers have the incentive to encourage physicians to tradeoff the full price of drugs against patients health benefits, and the tools to do so. When physicians use expensive drugs it forces a health insurer to charge its customers higher premiums, but when patients receive treatments that improve their health, the insurance plan becomes more attractive to prospective

[10]In the IntrinsiQ data set we use, Medicare patients account for about one-half of all colorectal cancer patients who receive drug treatment

enrollees. Therefore, insurers should set a patients drug co-insurance rate, physicians drug profits, and its oversight policies (e.g., the criteria for pre-authorizing a proposed oncology treatment) to encourage optimal drug treatment decisions. Dunn (2012) and Cockburn and Anis (2001) make similar assumptions that physicians and insurers act in the best interest of patients when interpreting their medical price indices.

2.3 Models

2.3.1 Hedonic Price Regression

Hedonic price regressions were introduced by (Court, 1939) and formalized by (Griliches, 1961) as a way to account for the new goods problem. In essence, newer goods usually contain more desirable characteristics, and therefore, failing to account for the value of these characteristics will overstate the true change in prices. The hedonic price regression is motivated by the "hedonic hypothesis, which states that goods can be viewed as aggregations or bundles of lower-order variables that the literature calls characteristics or attributes. These characteristics are what consumers care about and are present in utility functions. The hedonic function disaggregates the observed transaction prices into variables that affect the economic agents behavior. Court and Griliches proposed estimating a surface that relates prices to product characteristics and time, and then using the results to estimate price changes conditional on constant characteristics.

$$ln(p_{it}) = \beta X_{jt} + \sum_t \gamma_t d_t + \epsilon_{jt} \tag{2.1}$$

where p_{jt} is the price of regimen j in quarter t, x_{jt} contains the attributes of each regimen, including effectiveness attributes such as expected length of survival, time-to- progression, and response rate, as well as grade 3 and grade 4 side effects such as diarrhea, nausea, abdominal pain, vomiting, and neutropenia, and an indicator of whether the regimen is a pill versus an infused drug. The time indicators denote quarters, and the base quarter is the first quarter of 1993.

The price changes, which are obtained from the parameters of the above model to form a price index, are:

$$\frac{PI_t - PI_{t-1}}{PI_{t-1}} = exp(\gamma_t - \gamma_{t-1}) - 1$$

and therefore,

$$PI_t = exp(\gamma_t - \gamma_{t-1}) PI_{t-1}$$

Although the parameters β in equation (1) have been referred to as "implicit prices, the theoretical foundations of the hedonic surface are not clear. (Pakes, 2005) argues that in oligopolistic markets prices depend on both marginal costs and markups, and therefore the hedonic price regression constitutes a reduced form or projection of these markups on the characteristic space. In the industry we study, given the high costs of R&D relative to marginal production costs, markups should be large and Pakess insights are applicable. In other words,

it is unclear ex ante whether a particular sign on β or γ should reflect demand or supply factors. Suppose we observe prices rising from one quarter to the next (i.e., the γ for that quarter is positive). This could either reflect increased demand for colorectal cancer drugs or increased marginal costs, both of which would generate higher markups.

The parameters, therefore, do not necessarily follow any restriction based on economic theory; the β parameters may have the "wrong sign relative to what one would expect. For example, if there is more entry of products into a particular region of the attribute space that has more desirable characteristics for the average consumer (e.g., survival), this will reduce the markups of these products (so that prices might be lower due to competitive pressure rather than higher due to higher demand), and allow products with less desirable characteristics for the average consumer (e.g., worse side effects) to exploit monopoly power among the consumers who do not experience the strong side effects. We discuss the results of the hedonic price regression in the results section of the paper.

2.3.2 Logit Model

Our second price index explicitly calculates welfare changes due to the introduction of new products. This index is derived by estimating an equilibrium model of colon cancer drug pricing. (Trajtenberg, 1990), who introduced the idea, proposed a two-stage method to construct a quality-adjusted price index. In the first stage one obtains the welfare changes from product innovation, and in the second stage one builds the price index upon these welfare changes.

In this paper we specify a logit model, where physician demand for oncology drugs has the following form:

$$u_{ijt} = -\alpha p_{jt} + \beta x_{jt} + \xi_j + \Delta\xi_j + \epsilon_{ijt} \tag{2.2}$$

where p_{jt} is the price of regimen j at time t, x_{jt} are the observable attributes of the regimen, ξ_j is the mean of the unobserved characteristics, and $\Delta\xi_j$ is a time-specific deviation from this mean. ϵ_{ijt}, which is an idiosyncratic shock to preferences for regimen j, is assumed to follow a Type I Extreme Value distribution.[11] The outside option ($j = 0$) in this paper includes off-label colon cancer drug treatments, regimens with very small market shares, and regimens with missing efficacy or side effect attributes.

Patients are assumed to be administered one regimen at a time, which maximizes their utility. This implicitly defines a region of the unobserved term for which alternative j yields a higher utility than any other alternative k

$$A_{jt} = \{\epsilon_{it} | u_{ijt} > u_{ikt} \forall k \neq j\}$$

We obtain the market shares for each regimen j by aggregating individual preferences over the region A_{jt}:

$$S_{jt} = \int_{A_{jt}} dP(\epsilon)$$

[11] In this model all the individual-specific heterogeneity is contained in the idiosyncratic shock to preferences, and therefore, it suffers from the well-known independence of irrelevant alternatives criticism (see Nevo (2000) for a complete discussion of the limitations of this approach (Nevo, 2000)). In addition, Petrin (2002) points out that the welfare calculations based on these models depend heavily on the error term

In this case, where ϵ is assumed to be Type I Extreme Value, the integral can be computed analytically as:

$$S_{jt} = \frac{exp(-\alpha p_{jt} + \beta x_{jt} + \xi_j + \Delta\xi_j)}{1 + \sum_{k=1}^{J} exp(-\alpha p_{kt} + \beta x_{kt} + \xi_k + \Delta\xi_k)} \tag{2.3}$$

We define the mean utility level at time t as: $\delta_{jt} = -\alpha p_{jt} + \beta x_{jt} + \xi_j + \Delta\xi_j$. Thus, market shares can be written as the function of δ_{jt}:

$$S_{jt} = \frac{exp(\delta_{jt})}{1 + \sum_{k=1}^{J} exp(\delta_{jt})}$$

The market shares predicted by the model are then matched with the observed market shares s_{jt}. Berry (1994) shows that δ_j can be uniquely identified by inverting the market share function $S^{-1}(s_t) = \delta_t$. For the logit model, the inversion yields

$$lns_{jt} - lns_{0t} = -\alpha p_{jt} + \beta x_{jt} + \xi_j + \Delta\xi_j \tag{2.4}$$

Because the unobserved drug regimen characteristics are likely to be correlated with price, estimating the equation above requires instrumental variables. We obtain our instruments by using the supply side market equilibrium conditions. Because price is a function of marginal costs and markups, any exogenous variable that shifts marginal costs or markups should be a valid instrument. Bresnahan et al. (1997) use as instruments the plausibly exogenous number of products in the market and the sum of observed characteristics of the competitors, which measure how crowded and competitive the product space is

(Bresnahan, Stern, & Trajtenberg, 1997). These sets of instruments should affect markups via changes in the competitive environment, and therefore should be correlated with price but uncorrelated with unobserved characteristics.

However, product attributes do not vary much over time in our sample due to infrequent product entry and exit. Therefore, our first-stage regressions using these "BLP-style" instruments are generally weak, and the estimation results do not differ measurably from standard OLS models. Therefore, we follow Nevo (2000) and use "Hausman-style" instruments. Specifically, we construct two instruments with the lagged prices of other regimens. For the price of regimen j in period t, one instrument is the average price in period $t-1$ of all other regimens other than j. A second instrument is the average price in period $t-1$ of regimens produced by firms whose drugs are not used in regimen j. The key identifying assumption is that these instruments are uncorrelated with the current-period demand shock, but are correlated with price of regimen j. For this to hold, we require that the demand shock for regimen j in period t is uncorrelated with the demand shock for regimen k in period $t-1$, a condition likely to hold true unless there exists a time-persistent market-level demand shock.

The logit price index is built based on changes in compensating variation derived from the above model. The compensating variation provides a measure of how much income could be taken away from (or given to) an individual to leave her indifferent between facing the old choice set and the new improved (inferior) choice set. Given the logit functional form, the compensating variation is calculated as:

$$CV_t = \frac{u_t - u_{t-1}}{\alpha}$$

where u_t is the unconditional indirect utility $u_t = max_j u_{jt}$ and α is the marginal utility of income. (Small & Rosen, 1981) show that u_t can be computed as:

$$u_t = ln \sum_j exp(\delta_j)$$

Trajtenberg (1990) shows that if the price change takes the form of a shift by a factor of $(1 - \mu_t)$ in the distribution of prices but the variance remains the same, then the price index can be obtained as:

$$PI_t = (1 - \mu_t) PI_{t-1}$$

where:

$$\mu_t = \frac{CV_t}{CV_t + \bar{p_t}}$$

and $\bar{p_t}$ is the average price in period t.

2.3.3 Full Random-Coefficients Model

The logit model, while easily computed and parsimonious, has some well-known problems related to substitution patterns of new goods. In particular, there is the famous "red bus, blue bus problem," wherein when the price of one good rises, consumers will tend to switch to other goods, regardless of their attributes. The reason for this is embedded in the assumption of logit models that

market shares are driven primarily by the mean utilities, δ_{jt}. This implies that a new good entering the market will generate the same substitution patterns towards it, regardless of the characteristics of existing products.

Because it is important to address the substitution pattern between a new product and existing products when constructing the price index, we also estimate a full random-coefficients model. The appeal of this model is that physicians can differ in their tastes for colorectal cancer drug treatment characteristics. Specifically, physicians can have different preferences regarding the efficacy and side effects measures reported from randomized clinical trials. Allowing this realistic feature comes at the cost of increasing the computational burden and introducing some simulation error. The benefit, though, is much more reasonable substitution patterns that may affect welfare estimates.

Specifically, the utility function of the full random-coefficients model takes the following form:

$$u_{ijt} = -\alpha p_{jt} + \beta x_{jt} - \alpha^u v_i p_{jt} + \beta^u v_i x_{jt} + \xi_j + \Delta\xi_j + \epsilon_{ijt} \tag{2.5}$$

where all of the variables and parameters are the same as in equation (2) above, but now v_i is an unobserved physician preference component, α^u is the parameter associated with individual physician price sensitivity, and β^u is the parameter associated with individual physician preference for drug attributes.

Using the Type I Extreme Value assumption for ϵ_{ijt}, the predicted market share equation now collapses to:

$$S_{jti} = \frac{exp(-\alpha p_{jt} + \beta x_{jt} - \alpha^u v_i p_{jt} + \beta^u v_i x_{jt} + \xi_j + \Delta\xi_j)}{1 + \sum_{k=1}^{J} exp(-\alpha p_{kt} + \beta x_{kt} - \alpha^u v_i p_{kt} + \beta^u v_i x_{kt} + \xi_k + \Delta\xi_k)} \quad (2.6)$$

$$= \frac{exp(\delta_{jt} - \alpha^u v_i p_{jt} + \beta^u v_i x_{jt})}{1 + \sum_{k=1}^{J} exp(\delta_{kt} - \alpha^u v_i p_{kt} + \beta^u v_i x_{kt})} \quad (2.7)$$

where $\delta_{jt} = -\alpha p_{jt} + \beta x_{jt} + \xi_j + \Delta\xi_j$.

Because equation (7) is a function of mean utilities, δ_j, and the random components of the utility function , we cannot invert it to be a function solely of observed market shares as in equation (3) above to yield the analytical form of equation (4).

The parameters we need to estimate are $\theta = (\theta_1, \theta_2)$, where $\theta_1 = (\alpha, \beta)$ and $\theta_2 = (\alpha^u, \beta^u)$.

To do so, we first aggregate over individual preferences by drawing physician preferences, v_i , from a standard normal distribution and approximating equation (7) conditional on a starting value for δ_{jt} and the unobserved preference parameters:

$$S_{jt}(\delta_{jt}, \alpha^u, \beta^u) = \frac{1}{ns} \sum_{i=1}^{ns} \frac{exp(\delta_{jt} - \alpha^u v_i p_{jt} + \beta^u v_i x_{jt})}{1 + \sum_{k=1}^{J} exp(\delta_{kt} - \alpha^u v_i p_{kt} + \beta^u v_i x_{kt})}$$

Berry, Levinsohn, and Pakes (1995) show that there exists a contraction mapping on the following equation:

$$\delta_{jt}^{h+1}(\theta_2) = \delta_{jt}^{h}(\theta_2) + (ln(s_{jt}) - ln(S_{jt}(\delta_{jt}^{h}, \theta_2))) \quad (2.8)$$

where h is the iteration of the contraction and s_{jt} are the observed market shares. Once we find δ_{jt} conditional on θ_2, we can back out the unobserved product quality, ξ_j , as follows:

$$\xi_j(\theta_2, s_{jt}, P^{ns}) = \delta_{jt}(\theta_2, s_{jt}, P^{ns}) - \beta x_{jt} + \alpha p_{jt}$$

We interact this unobserved product quality with our instruments Z and estimate the model using the Generalized Method of Moments (GMM). The identifying assumption, which is similar to that of the logit model described above, is that our instruments are correlated with drug prices but not with any unobserved period-specific demand shocks. We then perform a non-linear search for the values of θ that minimize the objective function, and these values represent our estimates of the structural parameters of the model.

Specifically, our estimating equation is:

$$E[\xi(\theta) * Z] = 0$$

and our objective function is:

$$\xi(\theta)' Z \Phi^{-1} Z' \xi(\theta)$$

where Φ^{-1} is a consistent estimate of $E[\xi * Z]$.].

Finally, we calculate the price index in a similar manner as with the logit model above, with exception being that the index needs to be summed across individuals in the model (i.e., the 100 simulated physicians). Therefore, using

the parameters estimated above, we compute a compensating variation for each individual i that takes the following form:

$$CV_{it} = \frac{u_{it} - u_{it-1}}{\alpha_i}$$

where $\alpha_i = \alpha + \alpha^u v_i$ is the marginal utility of income for physician i and $u_{it} = max_j u_{ijt}$. Finally, the price index is computed as:

$$PI_t = (1 - \mu_t) PI_{t-1}$$

where:

$$\mu_t = \frac{\bar{CV}_t}{\bar{CV}_t + \bar{p}_t}$$

Here $\bar{p}_t$ is the average price in period t and:

$$\bar{CV}_t = \frac{1}{ns} \sum_{i}^{ns} CV_{it}$$

2.3.4 Pure Characteristics Model

While the full random-coefficients model mitigates some of the substitution problems inherent in a logit model, it still relies on the assumption of the Type I Extreme Value idiosyncratic error term, ϵ_{ijt}. The presence of this term allows the model to be solved and the mean utility levels, δ_{jt}, to be derived as a function of observed market shares. However, the error term is problematic when

calculating welfare because it guarantees that regardless of product characteristics, there exist certain consumers who would purchase *any* product (i.e., market shares cannot be zero for any product regardless of its price and attributes). This implies that product entry necessarily increases the utility of some consumer, and therefore consumer utility will continue to increase as more products are added to the product space.

Berry and Pakes (2007) show how this shortcoming can lead to biased welfare estimates. They propose eliminating the idiosyncratic error term from the model altogether, which alters the model so that it becomes purely a model of "tastes for characteristics." The randomness of choices comes from different consumer sensitivity to price and heterogeneous product attribute preferences, as opposed to heterogeneity in the "tastes for products" models described above. The benefit of estimating this model is that it generates more plausible substitution patterns by eliminating the Type I error assumption.

The utility function is identical to that in equation (5) except the error term is now dropped. Therefore, utility in the pure characteristics model becomes:

$$u_{ijt} = \bar{\delta_{jt}} - \alpha_i p_{jt} + \beta_i x_{jt} + \xi_j + \Delta\xi_j \tag{2.9}$$

where $\bar{\delta_{jt}}$ is the combination of all the attributes for which the parameters are non-random (i.e., physicians agree on their valuations), and we call the parameters β. Note that all the heterogeneity now comes from the α_i term, the price sensitivity of physician i, and β_i, the physician taste for attributes. For tractability, we allow physicians to have different sensitivity to price and heterogeneous preferences for just one non-price attribute: a patients expected response rate

on a specific drug regimen, which is arguably the most important attribute for most patients.

Following Berry and Pakes (2007) and Song (2008), we derive the market share equation first by ordering the products such that product 1 has the lowest price and product J has the highest price. That is:

$$p_1 < p_2 < ... < p_J$$

Given this ordering by price, consumer i will buy product j if and only if:

$$\begin{aligned} u_{ijt} &> u_{ikt} \\ \delta_{ijt} - \alpha_i p_{jt} &> \delta_{ikt} - \alpha_i p_{kt} \end{aligned}$$

where $\delta_{ijt} = \bar{\delta_{jt}} + \beta_i x_{jt} + \xi_j + \Delta\xi_j$. Thus we have the cutoff points:

$$\begin{aligned} \alpha_i &< \frac{\delta_{ijt} - \delta_{ikt}}{p_{jt} - p_{kt}} \quad if \quad p_{jt} > p_{kt} \\ \alpha_i &> \frac{\delta_{ikt} - \delta_{ij}}{p_{kt} - p_{jt}} \quad if \quad p_{kt} > p_{jt} \end{aligned}$$

Therefore consumers i will buy good j if:

$$\alpha_i < \min_{k<j} \frac{\delta_{ijt} - \delta_{ikt}}{p_{jt} - p_{kt}} = \bar{\Delta}_{ijt}$$

and

$$\alpha_i > \max_{k>j} \frac{\delta_{ikt} - \delta_{ijt}}{p_{kt} - p_{jt}} = \underline{\Delta}_{ijt}$$

and so the share for good j becomes:

$$s_j = \int F(\bar{\Delta}_{ijt}|\beta_i) - F(\underline{\Delta}_{ijt}|\beta_i)1[\bar{\Delta}_{ijt} > \underline{\Delta}_{ijt}]dG(\beta_i)$$

where $F(.|\beta_i)$ is the cumulative distribution function (CDF) of α_i conditional on β_i, and $dG(\beta_i)$ is the CDF of β_i. For our model, we assume a lognormal distribution on α_i, so that $log(\alpha_i) \sim N(0, \theta_1)$ and a normal distribution on β_i, so that $\beta_i \sim N(\beta_2, \theta_2)$.

The parameters of the model to be estimated are $\theta = (\beta, \beta_2, \theta_1, \theta_2)$. Note that if there were only one random coefficient, for instance on price, then this model would collapse to a vertical model (see Bresnahan (1987) and Lucarelli & Nicholson (2009)). In a vertical model, the δ_{jt} can be derived analytically, and thus estimating the model is relatively straightforward. However, with the presence of two random coefficients, one can no longer back out the mean utilities.

In addition, this model imposes a computational burden; dropping the idiosyncratic error term means that one is no longer guaranteed the existence of a contraction mapping that can be used to estimate the full random-coefficients model, as demonstrated by Berry, Levinsohn and Pakes (1995). We therefore follow Berry and Pakes (2007) and Song (2008) and use a three-step procedure to estimate the model. We first re-estimate the full random-coefficients model and rescale the logit error term by a factor that drives it toward zero. This brings the mean utility closer to its true value in the absence of an error term. We then use a combination of a fixed-point homotopy method and a Newton-Raphson search. Details of the search procedure are available in Appendix B.1.

2.4 Data

We use a number of different data sources to collect four types of information: drug prices, regimen market shares, the recommended quantity/dose of each drug used in each regimen, and regimen attributes from randomized controlled clinical trials. Wholesalers purchase drugs from manufacturers and then sell them to retail pharmacies, hospitals, physician practices, and other customers. IMS Health records transactions between wholesalers and its customers. Specifically, IMS Health collects information on the sales in dollars and the quantity of drugs purchased by 10 different types of customers (e.g., hospitals, physician offices, retail pharmacies) from wholesalers in each quarter from 1993 through the third quarter of 2005. Prices and quantities are reported separately by National Drug Classification (NDC) code, which are unique for each firm-product-strength/dosage-package size. We calculate the average price paid per milligram of active ingredient of a drug by averaging across the different NDC codes for that particular drug. IMS Health reports the invoice price a customer actually pays to a wholesaler, not the average wholesale price (AWP), which often differs substantially from the true transaction price. We use nominal rather than real prices because any deflator would itself be a price index, and we do not want to build one index on top of another.

The price we calculate includes on-invoice discounts (e.g., for paying the wholesaler promptly) but does not include any discounts or rebates a customer may receive from a manufacturer after purchasing the product from the wholesaler. Based on interviews with oncologists and an analysis reported in Lucarelli, Nicholson, and Town (2010), we do not believe that manufacturers offered substantial rebates during this period (Lucarelli, Nicholson, & Town,

2010).[12] In a recent paper Howard et al. (2015) agree that we do not believe that rebates refunds from manufacturers to hospitals, physicians, pharmacies and third party insurers are large in the market for new anticancer drugs (Howard, Bach, Berndt, & Conti, 2015). Although we have information on 10 different types of customers, we focus on the prices paid by the two largest customers - hospitals and physician offices – because most oncology drugs are infused in a physicians office or hospital clinic.

Most colon cancer patients who receive drug treatment are treated with a regimen that contains two or more component drugs. The IMS Health data contain information on market share by drug, but not market share for the combinations of drugs (regimens) actually used on patients. We rely, therefore, on two different sources for regimen-specific market shares. IntrinsiQ is a company that provides information systems to oncologists to help them determine the proper drug dosing for their cancer patients. As a result, IntrinsiQ collects monthly data from its oncology clients on the types of drugs used for patients. IntrinsiQ provided data on the proportion of colorectal cancer patients (of all ages) treated with drugs who are treated with each regimen for each month between January 2002 and September 2005.[13]

[12]For the five patent-protected colorectal cancer drugs in our study, these authors compared prices that include discounts and rebates to the IMS prices that we use in this paper. They found that prices from the two data sources were within two to four percent of one another, which is consistent with no or small rebates/discounts. Although pharmacy benefit managers, or PBMs, are able to negotiate price discounts for health insurers for many self-administered oral drugs, PBMs are less effective at negotiating discounts on the physician-administered drugs that we examine in this paper

[13]Because we observe the market shares of regimens among patients with colorectal cancer, we do not need to worry about off-label use. Off-label use occurs when a physician treats a colorectal cancer patient with a drug that has not been approved by the FDA to treat colorectal cancer, or when a physician uses a drug approved for colorectal cancer on a patient with a different type of cancer. In October 2005, seventy-six percent of patients being treated with the four drugs approved solely for the treatment of colorectal cancer (irinotecan, oxaliplatin, cetuximab, and bevacizumab) actually had colorectal cancer. That is, off-label use accounted for approximately 24 percent of the quantities of these drugs.

We derive market shares for the 1993 to 2001 period from the Surveillance Epidemiology and End Results (SEER) data set, which tracks the health and treatment of cancer patients over the age of 64 in states and cities covering 26 percent of the United States population.[14] We calculate the proportion of colorectal cancer patients who are treated with each drug regimen in each quarter based on Medicare claims data available in SEER. In October 2003, approximately 48 percent of all colorectal cancer patients treated with cancer therapy drugs were 65 years or older.[15] In the 1993 to 2001 period, when there were relatively few treatment options for colorectal cancer, we include all regimens that contain drugs that were explicitly approved by the Food and Drug Administration (FDA) for colorectal cancer and had a market share greater than two percent. Market shares of all other drugs are combined into an outside option, which in this early period will consist primarily of off-label drugs - drugs approved for conditions other than colorectal cancer that are used on colorectal cancer patients.[16] In the 2002 to 2005 period, the outside option includes off-label drugs, regimens with less than one percent market share in the third quarter of 2005 (the end of the sample period), and regimens with missing attribute data.

Market shares for the 12 regimens in our sample and the outside option are plotted in Figure 1. The regimens are also described more fully in Tables 1 and 2. Between 1993 and 1996, about 95 percent of colorectal cancer patients were treated with 5-FU/leucovorin, which at that time was generic, with the remainder treated with off-label drugs or regimens with very small mar-

[14]SEER contains data on the incidence rate of cancer among the non-elderly, but only has medical claims available for Medicare patients.

[15]Data from IntrinsiQ.

[16]Off-label use is more likely to occur if a patients initial treatment has been unsuccessful.

ket share.[17] Irinotecan (brand name Camptosar) was approved by the FDA for treating colorectal cancer in 1996, and over the next several years the market share of irinotecan (approved as a second-line treatment for metastatic colorectal cancer patients who had already been treated with a different drug regimen) and irinotecan combined with 5-FU/LV grew at the expense of 5-FU/LV.[18] Capecitabine (Xeloda), a tablet that produces the same chemical response as 5-FU/LV, was approved for treatment of colorectal cancer in April of 2001 and was administered as a standalone therapy or combined with irinotecan. All other drugs for treating colorectal cancer in our sample are delivered intravenously under the supervision of a physician or nurse.

Oxaliplatin (Eloxatin) was introduced in August of 2002, followed by cetuximab (Erbitux) and bevacizumab (Avastin) in February of 2004. By the third quarter of 2005, two of the regimens created by these three new drugs (oxaliplatin + 5-FU/LV; and bevacizumab + oxaliplatin + 5-FU/LV) surpassed the market share of 5-FU/LV, whose share had fallen to about 14 percent. The market shares of several regimens change sharply in the first quarter of 2002 when we use market share data from IntrinsiQ rather than SEER. One explanation for these changes is that Medicare patients may be treated with different regimens than non-Medicare patients. Another possible explanation is that the samples used by IntrinsiQ and/or SEER may not be consistent.[19] In order to smooth market shares between the pre- and post-2002 periods, we apply a regimen-specific factor to adjust the pre-2002 market shares based on the ratio of total (from IntrinsiQ) to Medicare-only (from SEER) market shares for the four quarters of 2002, when the two data sets overlap.

[17] 5-FU contains the drug fluorororacil.

[18] Because it takes Medicare a while to code new drugs into their proper NDC code, for several quarters a new drug will appear in the outside option.

[19] The SEER sample is drawn from locations representing 26 percent of the U.S. population.

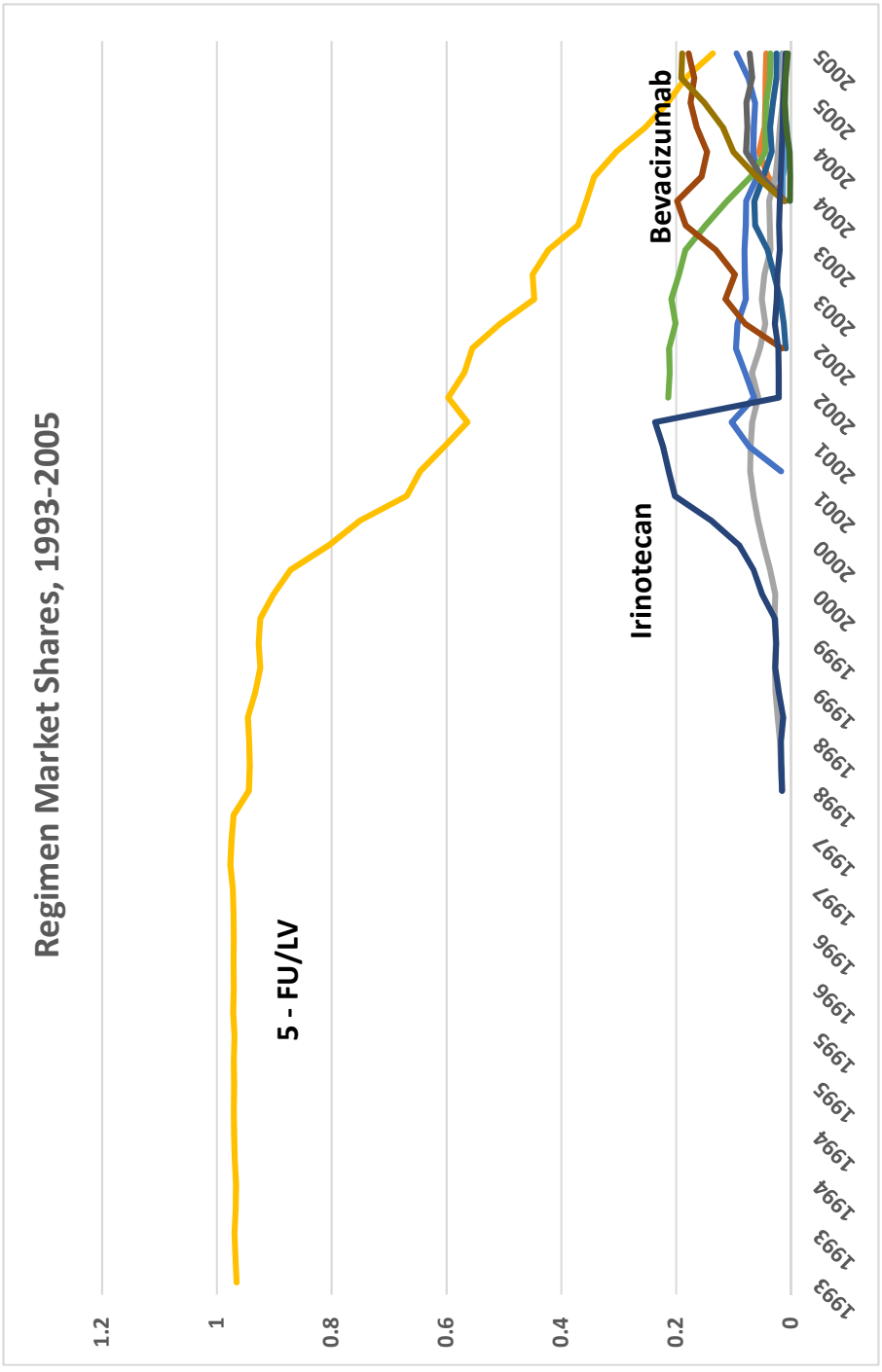


Figure 2.1: Regimen Market Shares

In order to calculate the price per regimen, we require information on the quantity of each drug in a regimen. The National Comprehensive Cancer Network (NCCN) reports the typical amount of active ingredient used by physicians for the major regimens. We supplement this where necessary with dosage information from drug package inserts, conference abstracts, and journal articles. Dosage information is reported in Appendix B.2. For example, the standard dosage schedule for the regimen with the second largest market share in 2005 is 85 milligrams (mg) of oxaliplatin per meter squared of a patients surface area infused by IV on the first day of treatment, followed by a 1,000 mg infusion of 5-FU per meter squared of surface area on the first and second treatment days, and a 200 mg infusion of leucovorin per meter squared on the first and second treatment days. This process is repeated every two weeks. We price the regimens for a representative patient who has 1.7 meters squared of surface area (Jacobson et al., 2006) weighs 80 kilograms, and is treated for 24 weeks. Regimen prices are derived by multiplying the average price a customer paid per milligram of active ingredient in a quarter by the recommended dosage amounts for each drug in the regimen over a 24-week period.[20]

We obtain most of the attribute information for each regimen from the FDA-approved package inserts that accompany each drug. These inserts describe the phase 3 clinical trials that were conducted, including the number and types of patients enrolled in the trials, the health outcomes for patients in the treatment and control groups, and the side effects experienced by those patients. Often there are multiple observations for a regimen, either because a manufacturer conducted separate trials of the same regimen, or because a regimen may have been the treatment group in one clinical trial and the control group in a

[20]The regimens are priced using data for the contemporaneous quarter only.

subsequent trial run by a different firm. In these cases we calculate the mean attributes across the separate observations. Where necessary, we supplement the package insert information with abstracts presented at oncology conferences and journal articles.

The attribute information is summarized in Tables 1 and 2, organized according to the year when each regimen was introduced. Table 1 shows three measures of a regimens efficacy: the median number of months patients survive after initiating therapy; the percentage of patients who experience a complete or partial reduction in the size of their tumor (i.e., the response rate); and the mean number of months (across patients in the trial) before their cancer advanced to a more serious state.[21] For all three of these measures, higher values are associated with superior health outcomes. We also record whether a regimen contains the capecitabine tablet, which should make the administration of the regimen more convenient for a patient, and whether the regimen is approved (and was tested) as a second-line treatment. Efficacy measures for second-line regimens will generally be worse than those for first-line regimens because the patients cancer is likely to be more advanced at the beginning of the clinical trial and the first treatment was not completely successful.

We also collected data on the percentage of patients in phase 3 trials who experienced either a grade 3 or a grade 4 side effect for five separate conditions: abdominal pain, diarrhea, nausea, vomiting, and neutropenia. These are displayed in Table 2. Although many more side effects are recorded for most regimens, these five were consistently recorded across the 12 regimens in the sample. Side effects are classified on a standard one to four scale, with four

[21]Cancers are classified into four stages, with higher numbers indicating that the cancer has metastasized beyond its initial location.

Table 2.1: Drug Regimens and Efficacy Measures

Regimen	Launch Year	Price (2005)	Survival Months	Efficacy Measures	
				Response Rate	Time to Progression
5-FU + Leucovorin	1991	75	12.5	20.8	4.7
Irinotecan (Second Line)	1996	23,478	9.5	15.0	4.2
Irinotecan + 5-FU/LV	1996	20,124	15.6	35.4	6.7
Capecitabine	2001	9,223	13.1	21.0	4.4
Ironetecan + capecitabine	2001	21,385	15.6	35.4	6.7
Oxaliplatin + 5-FU/LV	2002	25,426	19.4	46.1	9.1
Cetuximab (Second Line)	2004	53,859	N/A	10.8	1.5
Cetuximab + irinotecan (Second Line)	2004	73,519	6.1	22.9	4.1
Bevacizumab + oxaliplatin + 5-FU/LV	2004	76,636	23.2	41.0	9.9
Bevacizumab + oxaliplatin + capecitabine	2004	57,541	23.2	41.0	9.9
Bevacizumab + irinotecan + 5-FU/LV	2004	46,991	20.3	45.0	10.6

Notes: All attribute information is based on results of patients in Phase 3 Clinical Trials. Median survival is given in months. Time to progression is measured as the mean number of months for a tumor to advance to a more severe stage. Side Effects are measured as percentage of patients who experienced a grade 3 or 4 side effect of the given variety.

Table 2.2: Drug Regimens and Side Effects Measures

Regimen	Launch Year	Price (2005)	Side Effects Measures				
			Abdominal Pain	Diarrhea	Nausea	Vomiting	Neutropenia
5-FU + Leucovorin	1991	75	5.5	10.4	4.8	4.4	33.7
Irinotecan (Second Line)	1996	23,478	16.0	31.0	17.0	12.0	26.0
Irinotecan + 5-FU/LV	1996	20,124	5.3	24.0	11.9	8.0	39.5
Capecitabine	2001	9,223	9.5	15.0	4.0	4.5	3.0
Ironetecan + capecitabine	2001	21,385	5.3	24.0	11.9	8.0	39.5
Oxaliplatin + 5-FU/LV	2002	25,426	6.0	15.4	4.4	5.5	38.8
Cetuximab (Second Line)	2004	53,859	14.0	2.0	2.0	6.0	5.0
Cetuximab + irinotecan (Second Line)	2004	73,519	8.0	22.0	6.0	7.0	5.0
Bevacizumab + oxaliplatin + 5-FU/LV	2004	76,636	8.0	23.1	7.9	8.6	12.2
Bevacizumab + oxaliplatin + capecitabine	2004	57,541	8.0	23.1	7.9	19.0	12.0
Bevacizumab + irinotecan + 5-FU/LV	2004	46,991	8.0	34.0	1.0	1.0	21.0

Notes: All attribute information is based on results of patients in Phase 3 Clinical Trials. Median survival is given in months. Time to progression is measured as the mean number of months for a tumor to advance to a more severe stage. Side Effects are measured as percentage of patients who experienced a grade 3 or 4 side effect of the given variety.

being the most severe. Higher values for the side effect attributes should be associated with worse health outcomes although, as we will show later, regimens that are more toxic are likely to be both more effective and have more severe side effects.

New colorectal cancer regimens tend to be more efficacious than the existing regimens, with side effect profiles that are sometimes more and sometimes less severe than earlier regimens. Consider the new entrant in 1996, irinotecan + 5-FU/LV (third row of Table 1). Relative to patients who received 5-FU/LV in a clinical trial (first row of Table 1), patients in clinical trials who received irinotecan + 5-FU/LV lived 3.1 months longer, on average, had a 14.6 percentage point higher probability of experiencing a reduction in the size of their tumor, and experienced a two month delay in the time it took for the cancer to advance to a more severe state. However, patients taking the new regimen were more likely to experience four of the five side effects listed in Table 2

Oxaliplatin + 5-FU/LV, which was launched in 2002 (sixth row of Table 1 and Table 2), is more efficacious and has fewer severe side effects than irinotecan + 5-FU/LV. Patients in clinical trials of the former regimen lived an average of 3.8 months longer, had a 10.7 percentage point higher probability of experiencing a reduction in the size of their tumor, and experienced a 2.4 month delay in the time it took for the cancer to advance to a more severe stage relative to the latter regimen. Oxaliplatin + 5-FU/LV patients are also less likely to experience a grade 3 or 4 side effect for four of the five measures relative to irinotecan + 5-FU/LV. Finally, the arrival of bevacizumab + oxaliplatin + 5-FU in 2004 increased the median survival time by about four months relative to oxaliplatin + 5-FU/LV, with substantial improvements with one side effect measure and

worse performance on the other four measures.

Two new second-line regimens entered the market in 2004 to compete against the first second-line regimen (irinotecan) that was launched in 1996.[22] Cetuximab + irinotecan has a substantially better response rate than irinotecan administered by itself, although median survival is shorter. The new regimen also is superior to irinotecan on all five of the side effect measures.

The log of the naive price index is plotted in Figure 2. Note that prices are relatively stable in the pre-1998 period where 5-FU/LV dominated the market and all other brands were considered into the outside option. Then share-weighted prices began to rise in 1998 when irinotecan was introduced, and then continued to increase until the 2005 period. This is further documented in Table 1, where it can be seen that the 2005 price of irinotecan was considerably higher than that of 5-FU/LV, and that this continued to be the case again in 2001 when capacitabine was introduced, and again in 2004 when Avastin came on the market.

2.5 Results

Table 3 reports the results of the hedonic price regression. The dependent variable is the logarithm of the price a customer paid for regimen j in quarter t. Regimen attributes are included as well as a full set of quarter indicator variables. The coefficient on one of the efficacy measures, the tumor response rate, is positive and significant, while the coefficients on the other two measures, survival months and time to progression, are negative. For the response rate, a

[22] Regimens that include the tablet, capecitabine, are chemically equivalent to regimens that include 5- FU/LV.

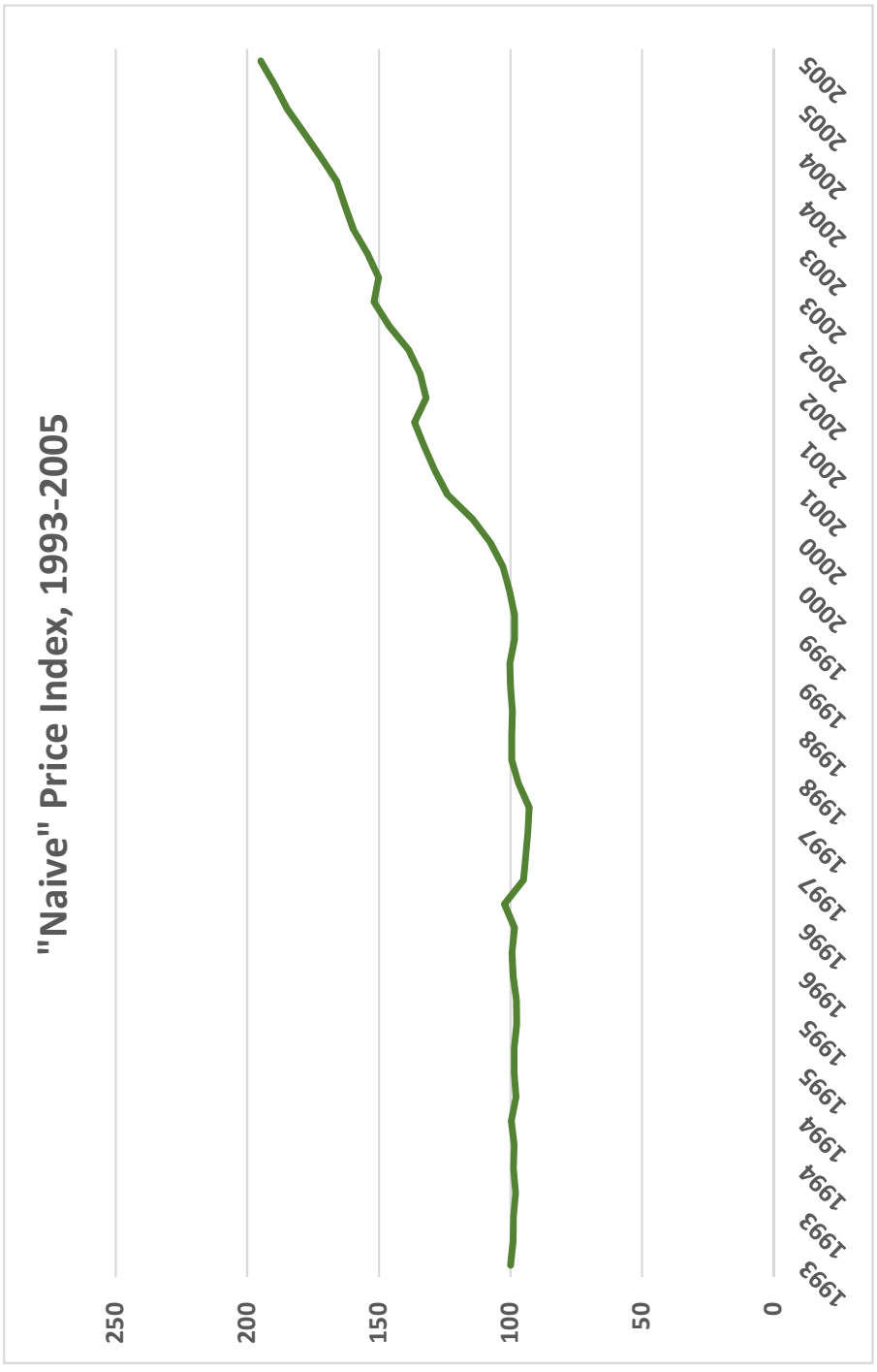


Figure 2.2: Naive Price Index

one percentage point increase in the tumor response rate is associated with a 51.6% increase in the price of a regimen. Evaluated at the mean regimen price in the sample ($21,113), this implies an increase of $10,894. Physicians are thus implicitly valuing gains in response rates quite highly.

Table 2.3: Results of the Hedonic Model

Variable	(Parameter)	(Standard Error)
	Efficacy Measures	
Survival Months	-0.216	0.040
Response Rate	0.516	0.023
Time to Progression	-1.231	0.148
	Side Effects Measures	
Diarrhea	0.032	0.015
Nausea	0.060	0.054
Abdominal Pain	0.461	0.033
Vomiting	0.040	0.081
Neutropenia	-0.072	0.006
Tablet	-0.094	0.169
Constant	1.171	0.677
Obs.	208	

Notes: Robust standard errors in parentheses. All models include quarter fixed effects. Dependent variable is the log of price physicians paid for chemo treatments in a quarter.

The negative coefficients on the other two measures seem to indicate that physicians assign negative valuation to those attributes. However, as discussed above, coefficients in hedonic models do not necessarily hold their expected signs. The degree of competition in this case could differ across the attribute space. For instance, there could be greater differentiation and less competition with the response rates attribute relative to survival months and time to pro-

gression. Indeed, it is quite plausible that at any given point in time, there could be more differentiation on certain attributes than others, as pharmaceutical and biotech firms design their products and construct clinical trials that will define attributes years before the products are approved and marketed.

Similarly, four of the five side effects measures are positive, implying that physicians place higher value on drugs that cause side effects. This could simply indicate that physicians place less value on the side effects of drugs than on the efficacy measures, and therefore their signs should hold no economic meaning. For instance, if a physician prescribes anti-vomiting or anti-pain medication along with colorectal cancer treatments, that physician would likely downgrade the importance of these attributes, even potentially yielding a positive sign (note that the estimates are much more insignificant than the efficacy measures). In addition, the sign on neutropenia is negative and significant. This makes sense, given that neutropenia is a much more severe side effect than the other four measures, and this would likely imply that physicians place a greater importance on mitigating it relative to the other side effects. A third potential explanation is that physicians simply place more value on efficacy than is suggested by the Phase 3 clinical trials. This could happen if physicians use the drugs differently in practice due to learning about patient-drug matching. Since newer drugs tend to have greater side effects, this would explain the positive coefficients.

Figure 3 plots the log hedonic price index using the methods described above (the quarter 1993 is omitted). The pattern of the hedonic index is dramatically different from the naive price index, which emphasizes the importance of controlling for product attributes. Prices were relatively stable between 1993 and

1996 when the only regimen approved for colorectal cancer treatment was 5-FU/LV and the components of this regimen were no longer patent protected. After this, the hedonic index drops by about 20 log points with the introduction of the two irinotecan-based therapies, which were priced at about $16,000 at the time of introduction. Over the next six years, prices remained relatively stable, then spiked by about 30 points with the introduction of bevacizumab (Avastin) based regimens. Since then, prices remained stable towards the end of the sample period.

In the first column of Table 4 we report estimated coefficients from the OLS logit model without instrumenting for price; the second column reports results of the IV logit, which addresses the endogeneity of prices; the third column reports the coefficients from the full random coefficients model ("BLP") that both instruments for price and incorporates physician heterogeneity in drug preferences. These results follow closely with Song, Lucarelli, and Nicholson (forthcoming), with the exception of the "full model." (Song, Lucarelli, & Nicholson, 2016). Rather than add a random coefficient for only the logprice, we include an additional random coefficient on the response time variable in order to allow heterogeneity in physician preferences for efficacy measures. This also allows for a more direct comparison to the pure characteristics model, where we also place a random coefficient both on price and response rate. Comparing the price coefficients between the column 1 and columns 2 and 3 confirms that there is a positive correlation between a drugs price and demand shocks, and that the instrumental variables mitigate the problem of price endogeneity. The price coefficient changes from -0.690 in the OLS logit to -2.15 in the IV logit, and -2.78 in the full random-coefficients model. The coefficient is significant in all three models.

Table 2.4: Results of the Logit and Full Models

Variable	(OLS Logit)	(IV Logit)	(BLP)
Price	-0.690	-2.150	-2.783
	(0.124)	(0.461)	(0.739)
Constant	2.196	7.537	9.478
	(0.606)	(1.621)	(4.229)
	Efficacy Measures		
Survival Months	0.086	-0.421	-0.447
	(0.056)	(0.124)	(0.145)
Response Rate	0.166	0.913	1.149
	(0.071)	(0.240)	(0.345)
Time to Progression	-0.335	-2.070	-2.839
	(0.242)	(0.610)	(0.926)
	Side Effects Measures		
Diarrhea	0.024	0.072	0.123
	(0.022)	(0.034)	(0.065)
Nausea	-0.137	-0.065	-0.085
	(0.077)	(0.109)	(0.147)
Abdominal Pain	0.135	0.806	0.962
	(0.076)	(0.223)	(0.294)
Vomiting	0.166	0.245	0.337
	(0.117)	(0.164)	(0.236)
Neutropenia	-0.008	-0.109	-0.152
	(0.010)	(0.033)	(0.051)
	Random Coefficients		
α^u (logprice)			0.342
			(0.135)
β^u (Response Rate)			-0.094
			(0.064)
Obs.	208	208	208
Adj R2	0.771	0.684	
GMM Obj.			$6.88x10^{-9}$

Notes: Robust standard errors in parentheses. All models include quarter fixed effects. Dependent variable for the logit models are the logarithm of the market share of a regimen minus the logarithm of the market share of the outside option.

Again, the signs of the efficacy measures look counterintuitive in all of these models. In the IV Logit and full models, only the response rate coefficient is positive, which is the same pattern observed in the hedonic model. This might suggest that physicians place a greater demand on response rate than the other efficacy measures. It is also possible that the three variables are correlated with one another, which would dampen the effects of some of the measures.

Regarding the side effects measures, in all models the coefficients are positive for diarrhea, abdominal pain, and vomiting, while negative for nausea and neutropenia. This is an improvement relative to the hedonic model, where 4 of the 5 measures were positive. In particular, it seems that physicians place a significant negative weight on the neutropenia side effect measure, which again, is one of the most severe side effects of colorectal cancer treatments.

Turning to the random coefficients from the BLP, there is a positive and significant value for the standard deviation on price, α^u, and a negative but insignificant coefficient on the standard deviation on response rate, β^u. This is suggestive of the fact that colorectal cancer treatments that are similar prices are closer substitutes than ones that have significantly different prices. Moreover, the average price coefficient is higher in magnitude than in the IV logit model, suggesting that physician heterogeneity for price sensitivity does play a role in selecting the regimen for a particular patient.

Although the full model incorporates some degree of heterogeneity in preferences, the model is still plagued by the inclusion of the idiosyncratic error term, which ensures that each regimen that entered the market would generate non-zero market shares regardless of whether it was in the quality space. This tends to produce inflated estimates of welfare gains due to product innovation.

Table 5 presents the results of the pure characteristics model of demand, which drops this error term. In this model, consumers agree on the rankings of product quality but differ in their willingness-to-pay for different quality levels.

Table 2.5: Results of the Pure Characteristics Model

Variable	(Parameter)	(Standard Error)
Constant	4.012	1.737
	Efficacy Measures	
Survival Months	-0.155	0.042
Response Rate	0.418	0.084
Time to Progression	-1.066	0.285
	Side Effects Measures	
Diarrhea	0.060	0.025
Nausea	-0.014	0.145
Abdominal Pain	0.331	0.174
Vomiting	0.135	0.188
Neutropenia	-0.067	0.008
	Random Coefficients (θ)	
$\log(\alpha_i) \sim$ N(0 ,θ_1)	0.067	0.017
$\beta_i \sim$ N(β_2, θ_2)	0.057	0.011
Obs.	208	
GMM Obj.	3.709	

Notes: Robust standard errors in parentheses. All models include quarter fixed effects. Dependent variable is the log of price physicians paid for chemo treatments in a quarter.

The standard deviation of the price coefficient, θ_1, is 0.067 and statistically significant, implying that there is heterogeneity in physician preferences for regimens of different prices. Unlike the BLP model, however, the variance of the response rate coefficient, θ_2, is also positive and significant, with a value 0.057, implying that physicians also value response rate differently, even conditional on their α_i. Indeed, it appears that what was previously captured by the idiosyncratic error term is now being captured by these random coefficients in the

pure characteristics model, ensuring much more flexible substitution patterns. The coefficients on the efficacy and side effects measures are the same sign as in the OLS, IV Logit, and BLP models. In addition, the efficacy measures all remain statistically significant, as do the abdominal pain and neutropenia side effects measures.

The log price indices for the four models are plotted in Figure 3. During the initial period before irinotecan was introduced, the indices are relatively flat and look fairly similar to the naive and hedonic indices, as would be expected. The index from the logit model increases slightly, by about 10 percentage points in 1997, and gradually declines back to its original value throughout. In general, it is a pretty stable index. This is likely due to the fact that substitution patterns in the logit model guarantee that elasticities are increasing in prices, so that when the price of a more expensive good increases, there will be more substitution away from it than when the price of a cheaper good increases. The full model corrects for this by allowing individuals to substitute towards products they consider similar if they have a strong taste for those products. Therefore, physicians might be more likely to stay with Avastin, one of the more expensive drugs on the market, if they are more sensitive to its efficacy improvements and less to price. Indeed, looking at the BLP index, we can see that it is also relatively stable during the initial period, then drops a bit after irinotecan is introduced, but finally steadily increases beginning in around 2002 when the oxaplatin-based regimens are introduced, followed by the bevacizumab regimens.

Finally, looking at the pure characteristics model, it is obvious that the effect of dropping the idiosyncratic error term is to essentially "load" all of welfare changes onto the observable characteristics of the model. Notably, while the

logit and BLP models are essentially identical until irinotecan was introduced in 1996, the pure characteristics model sits *below* these price indices. This is due to the fact that during this initial period, only one regimen (5-FU + LV) dominated the marketplace, and its price gradually declined during those initial years. Given that no other regimens were present and that all other quality measures of this regimen remain unchanged during this period, the pure characteristics model interprets this as a large welfare increase, thereby decreasing the quality-adjusted price. Note, however, that when oxaplatin was introduced in 2002, the pure characteristics price index spiked above the other quality-adjusted indices, and continued to expand higher than the other indices through 2005. This is reflective of the fact that the logit and BLP indices attributed a significant portion of the welfare gains of oxaplatin and bevacizumab regimens to the error term. As the pure characteristics model shows, however, this had the effect of dampening the quality-adjusted prices due to an inflated estimate of welfare gains due to their introduction. This effect is a significant one: between 1993 and 2005, the logit index implies that *log* prices rose by about 5 percent, while the pure characteristics price index rose by about 18 percent.

The four price indexes depicted in Figure 3 (in addition to the nave index) assume that physicians act as the social planner, and trade off the full price of drug treatments against the financial value of patients health benefits from those treatments. If, however, health insurers do not create the optimal incentives for patients and physicians or if physicians exploit their information advantage to capture profits on drug treatments at the expense of patients health outcomes, these indexes may not accurately measure changes in social welfare. In a separate paper, Lakdawalla et al. (2015) use the same data to present a price index for colorectal cancer drug treatment that does not rely on physicians valuations

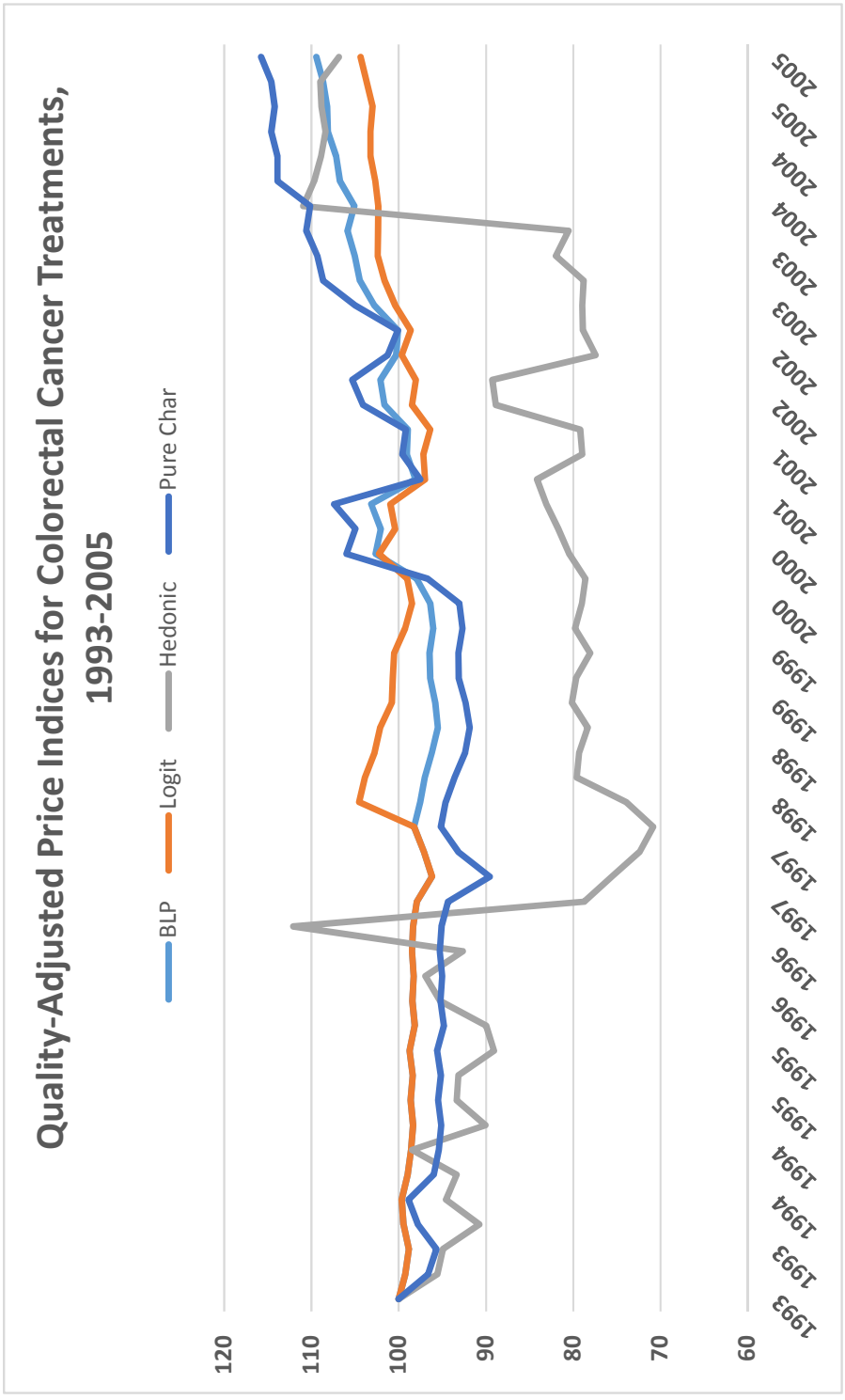


Figure 2.3: Quality Adjusted Price Indices

of the drug attributes (Lakdawalla et al., 2015). Specifically, they use the performance of the drugs in randomized clinical trials, reported here in Table 1 and Table 2, and the observed market shares to calculate a patients expected quality adjusted life years (QALYs) between 1998 and 2005. Applying a value of $100,000 per QALY based on the published literature, they then subtract the change in monetized expected health gains from the change in the drug treatment costs, which is similar to the method used by Cutler et al. (1998) for heart attack treatment and Eggleston et al. (2011) for diabetes treatment. Lakdawalla et al. (2015) find that this quality-adjusted price index, which imposes a value per QALY and forecasts QALY changes based on differences between drugs in survival and side effect incidence, increased by $1,400 over this time period, versus the nave increase of $36,000 in the per-patient average drug treatment cost. That is, a price index that does not rely on physicians subjective valuations is essentially flat, much as the four indexes depicted in Figure 3.

2.6 Conclusion

In this paper, we estimate a series of quality-adjusted price indices for colorectal cancer treatments. In particular, we compute a naive price index that does not adjust for improving attributes, and compare it with a hedonic price index, a logit index, a full model index, and a forthcoming pure characteristics index. Indeed, we do see that adjusting for quality and substitution between new goods matters for properly tracking prices, as each price index generates substantially smaller price increases than the naive index. Moreover, properly accounting for substitution patterns matters towards generating accurate welfare estimates, implying that the method of constructing price indices matters

significantly for assessing the value of new pharmaceuticals. While the naive index shows a dramatic increase over the 13-year period, the quality-adjusted indices increase only slightly, with the largest increase being approximately almost 20 points over the sample period (compared with an about 100 point increase for the naive index).

CHAPTER 3

MENTAL HEALTH MANDATES AND JOB TRANSITIONS

3.1 Introduction

A majority of Americans receive their health insurance through an employer. 55.1% of the U.S. population had employer-sponsored insurance (ESI) in 2011, compared with a total of 63.9% enrolled in private insurance plans overall.[1] These plans are primarily regulated at the state level. For instance, states implement their own community rating laws, which prohibit insurance companies from varying premiums for eligible groups by health status or other factors. As of January 2012, 18 states and the District of Columbia have implemented some form of community rating restriction.[2] Other states provide premium subsidies to certain groups. The Health NY program, for example, provides subsidies to small businesses (50 employees or less) within New York State in which 30% of employees earn wages of less than $40,000.[3]

Many states now have also implemented benefit and provider mandates, which require that group health insurance plans include minimum levels of certain insurance benefits. These include maternity coverage, mental health parity, eating disorders, substance abuse, infertility treatment coverage, and others. The first of such mandates was passed in 1956 and required dependent coverage for handicapped children in Massachusetts. By the late 1990s, there were over 1,000 active state benefit mandates (Jensen & Gabel, 1992). Today, the number of mandates varies by source. The National Conference of State Legislatures

[1] http://www.census.gov/prod/2012pubs/p60-243.pdf
[2] http://www.statehealthfacts.org
[3] http://www.dfs.ny.gov/website2/hny/english/hnyecsm.htm

(NCSL) estimates more than 1,900 mandates, while the Council on Affordable Health Insurance (CAHI) estimates approximately 2,271 mandates (Council for Affordable Health Insurance, 2012; National Conference on State Legislatures, 2011). Figure 1 shows the breakdown of benefit mandates by state, estimated by CAHI, for 2012 (Council for Affordable Health Insurance, 2012). It is evident from the figure that benefit mandates are prevalent today, but to varying degrees. While Idaho only had 13 benefit mandates passed as of 2012, Rhode Island had approximately 69. The degree to which certain states pass these policies in response to macroeconomic conditions (policy endogeneity) will be explored later in this paper (see section 4).

Figure 2 shows CAHI estimates of benefit mandates by year from 2004 through 2010 (2004 was chosen as it was the earliest the data was made publicly available). Indeed, the number of total mandates in the United States has risen considerably, from 1,823 in 2004 to 2,271 in 2012. Note in particular that the number of mandates continued to rise after the 2007 recession and in after the implementation of 2010's Affordable Care Act. Regardless of which source is used, it is clear that the number of mandates continues to rise in the United States.

The implementation of these mandates is an important national issue. Supporters argue that mandates provide much-needed benefits to individuals who otherwise would not have access through the private market. Their costs would be partially subsidized by individuals within the group who purchase these plans yet do not necessarily require the benefit. In other words, mandates are a form of preventing adverse selection. These added benefits might result in increased utilization for specific health benefits, and therefore increased health

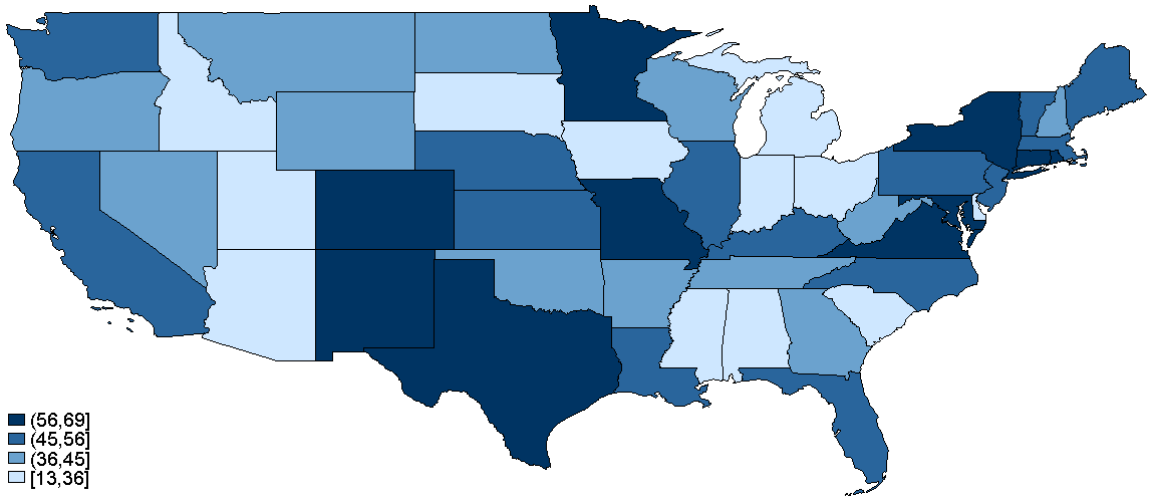


Figure 3.1: Number of Benefit Mandates By State, 2012

outcomes. However, this benefit might also impact the labor market. A mandate for a specific type of health insurance might lead to an increase in overall labor supply among individuals who value the benefit, as their search costs would be decreased. Therefore, after passage of particular mandates, we might expect to see diminished job turnover and increased job duration. In addition, a healthier workforce might result in added productivity for the firm, which would increase profits.

Opponents, however, argue that mandates could generate moral hazard problems, where consumers over-utilize newly offered coverage, leading to market inefficiency. In addition, this could then further raise costs for insurers, which would induce higher premiums for employers and employees. In turn, mandates could result in large displacement effects, where employers drop coverage altogether, hence reducing the probability that an individual become insured at all (Jensen & Gabel, 1992; Sloan & Conover, 1998; M. K. Bundorf, Henne, & Baker, 2007). According to CAHI, "Mandated benefits currently increase the cost of basic health coverage from slightly less than 10% to more

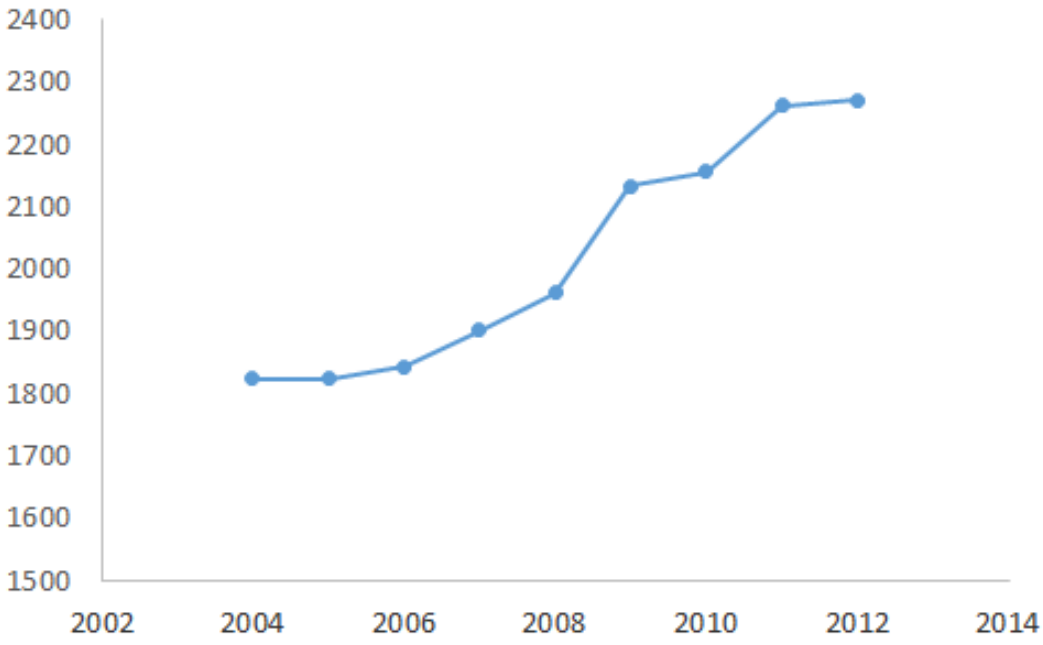


Figure 3.2: Number of Benefit Mandates by Year, 2004-2012

than 50%, depending on the state, specific language, and type of health insurance policy" (Council for Affordable Health Insurance, 2012). This might also lead employers to reduce their labor demand, hence hiring fewer employees, or may result in employers reducing the hours of their employees so as to not be forced to offer health insurance coverage. The theory behind benefit mandates and labor market outcomes will be discussed further in the next section (see section 2).

In this paper, I investigate the labor market effects of a particular kind of mandate: mental health parity. Mental health parity–the requirement that insurers in a particular state provide benefits for mental illness equivalent to that of comparable medical and surgical benefits–has been growing in popularity since the late 1990s. Indeed, states continue to pass legislation requiring comprehensive coverage of mental illness today, despite an already passed federal

legislation in 1996 and subsequently in 2008. New York States, for instance, passed legislation in 2006, known as "Timothy's Law", designed to offer mental health parity.[4] As recent as July 2013, New Jersey held hearings to discuss adding more comprehensive benefits to its already existing state parity law.[5]

An analysis of mental health parity is of special interest to policymakers and various stakeholders, especially considering the growing number of people with mental health disorders. A recent report from the Utah Department of Human Services, for example, found that in 2012, less than 31% of those in need of mental health treatment actually received services (Utah Department of Human Services, 2012). Part of this is due to lack of mental health coverage, as Utah lacks mental health parity (and is still debating whether to accept the Medicaid expansion offered by the Affordable Care Act).[6] Many argue that mental health parity is a step forward in providing needed benefits to those with serious mental health conditions (McGuire & Montgomery, 1982; Barry & Busch, 2007; Busch & Barry, 2008; Lang, 2013; Cseh, 2008), while critics note the potential cost increase on employers (Council for Affordable Health Insurance, 2012; Matthews, 1999; American Benefits Council, 2007). However, while there have been many papers citing the effects of mental health parity on insurance coverage and outcomes, literature on the effects on the labor market have been sparse. Most recently, Attila Cseh analyzed the effects of state parity mandates on a variety of labor market outcomes, including levels of employment, hours, and wages (Cseh, 2008). However, this analysis is limited in that it only looks at *levels* of employment, masking any effects on labor market *flows*. In addition, it can only identify effects of parity on employment for all individuals,

[4] http://www.timothyslaw.org/

[5] http://www.nj.com/politics/index.ssf/2013/07/nj_lawmakers_discuss_retooling_mental_health_coverage_la

[6] http://www.nytimes.com/2013/07/25/us/mental-health-cuts-in-utah-leave-some-feeling-adrift.html?partnerr̄ss&emcr̄ss&_r0̄

where there might be amplified effects for individuals seeking mental health treatment. Finally, its analysis concludes with states that have passed parity in 2002, whereas numerous states (including large ones, such as New York and Oregon) have passed mental health parity since.

I contribute to the literature by analyzing the effect of mental health parity legislation on both levels of employment and worker flows. In particular, I use restricted-access longitudinal data from the Medical Expenditures Panel Survey (MEPS) to look at the effects of parity on transitions between jobs and transitions from employment to unemployment among people who have sought mental health treatment or have been diagnosed with a mental health disorder. The MEPS is a public-use dataset that contains national employment, health insurance, and health expenditure data. The restricted-access data also contains state identifiers, allowing me to exploit the state variation in mandate passages.

An analysis of worker flows in addition to levels of employment is important in that simply analyzing the levels could be masking the transition effects imposed by health coverage mandates. In particular, many have argued that health coverage offered through an employer could contribute significantly to a phenomenon known as "job lock," where employees do not leave their current occupations for other jobs they might otherwise take for fear of losing their comprehensive health benefits.

I find that the passage of state mental health parity laws did cause a significant *decrease* in job separations for the employed population diagnosed with a mental health disorder. The effect is strongest for job to job transitions, and weaker for employment to unemployment transitions. However, I argue that the increased length-of-stay in employment is not indicative of "job lock," but

rather a decrease in involuntary separations due to a treatment effect, and a decrease in voluntary separations due to the added value of having the benefit. In other words, after the passage of mental health mandates, individuals with mental health disorders are able to work at their jobs for longer periods of time, presumably without a mental health incident that would result in job loss or without seeking other jobs with more comprehensive coverage. To investigate this further, I look at the effects of parity on mental health utilization, which I proxy by looking at total and out-of-pocket mental health expenditures. I find an increase in mean total expenditures, but no increase in out-of-pocket expenditures, further suggesting a treatment effect. I find similar effects following the passage of the Mental Health Parity and Addiction Equity Act, a federally mandated mental health parity law enacted in 2008 and implemented in 2009. This suggests that mental health parity may be a good policy in promoting increased health access, and ultimately improving worker health and productivity at their jobs.

The remainder of this paper proceeds as follows: Section 2 describes the background on mental health parity legislation, the theory and literature on mandates, mental health parity, and employment, and offers a simple model of job transitions following the passage of a mandate. Section 3 describes the data and identification of parity legislation dates. Section 4 describes the effects of mandates on job flows. Section 5 discusses the results and possible extensions. Section 6 concludes.

3.2 Background

3.2.1 Mental Health Parity Legislation

Benefits for mental health illness have been provided by insurers since the 1970s. However, the passage of laws requiring insurers to provide benefits for mental illness were primarily known as "minimum mandated benefits," laws, requiring some coverage for mental illness, though often not comprehensive. These laws also had lower lifetime expenditure limits, higher deductibles, and higher co-payments or coinsurance rates. Mental health "parity" legislation, conversely, requires that all insurers within a given state provide coverage for mental health equivalent to that of medical and surgical coverage. This implies a significant increase in the level of benefits mandated.

The first major push towards widespread parity was the 1996 Mental Health Parity Act (MHPA).[7] This act equated aggregate lifetime limits and annual limits for mental health benefits to that of physical conditions. However, the effect of this legislation on increasing access to care has been dubious. One limitation is that it only provided parity for annual and lifetime limits, but not for cost-sharing, coinsurance and other benefits. It also did not mandate that health insurance plans provide mental health benefits, but rather that those that already did provide some level of coverage expand to match those annual and lifetime limits. There were also many exempted groups, including small-firms with less than 25 employers. Hence, the effect of the legislation was widely considered to be minimal (Gitterman, Sturm, Pacula, & Scheffler, 2001).

[7] http://www.dol.gov/ebsa/mentalhealthparity/

Since 1996, many states have begun individually passing more comprehensive legislation. These laws did not necessarily offer mental health "parity," however the popularity of parity mandates did grow, particularly in the late 1990s and early 2000s. There are four major groups of mental health laws, with varying degrees of benefits. These are the following:

Minimum Mandated Benefits. Minimum mandated benefits require that insurers within a state provide *some* mental health coverage in their plans, particularly in the group and individual insurance markets. However, these benefits need not be very comprehensive, nor do they need to be equal to that of comparable medical and surgical benefits. For instance, Illinois currently has a minimum mandated benefit that requires a minimum of 45 inpatient days and 35 outpatient days only for serious mental illnesses, and more limited coverage of other mental health disorders.

Mandated if Offered Benefits. Mandated if Offered benefits require parity in coverage, but only if an insurance plan already offers mental health coverage. Unlike Mental Health Parity, these types of laws do not actually require that insurers provide any coverage for mental health coverage at all. An example is Arizona, which currently has a "Mandated If Offered" law in place for broad-based mental health disorders.

Mandated Offering Benefits. Mandated offering benefits require each insurance plan to offer the *option* of mental health coverage that is equal to that of medical and surgical benefits (i.e. parity), though it does not require the purchaser to take such insurance. Alabama currently has a mandated offering ben-

efit for serious mental illnesses.

Mental Health Parity. Mental health parity is the more comprehensive form of mental health benefit, and the focus of the analysis of this paper. It requires that each insurance plan with in a state provide benefits for mental health coverage that is equal (in terms of generosity, co-payments, coinsurance, visits, etc.) to that of medical and surgical coverage offered by the plan.

Figure 3 shows the number of states passing mental health parity, as well as other forms of mental health benefits between 1999 and 2010. It is clear that between 1990 and 1996, only three states had passed a parity mandate (though more states had passed other types of mental health benefits). The number of states then dramatically increases, with 7 states enacting parity mandates in 2000 alone. After 2002, there was a brief break in the passage of parity mandates, though this resumed as states like New York, Oregon, and Ohio began passing parity mandates in the mid-2000s. Table 1 further shows the specific states that passed parity mandates, as well as other kinds of mental health benefits, and the years of legislative activity.

Though many policymakers and advocates welcome the passage of this type of legislation as improving access to necessary benefits, other policymakers, state officials, employers and insurance plans have cited the extremely high cost of parity (Council for Affordable Health Insurance, 2012; Matthews, 1999; American Benefits Council, 2007). The National Center for Policy Analysis, in an open letter, wrote, "While it could help some patients, it would drive up the cost of health insurance and force more people into the ranks of the uninsured" (Matthews, 1999). CAHI mentioned, "mental health parity mandates,

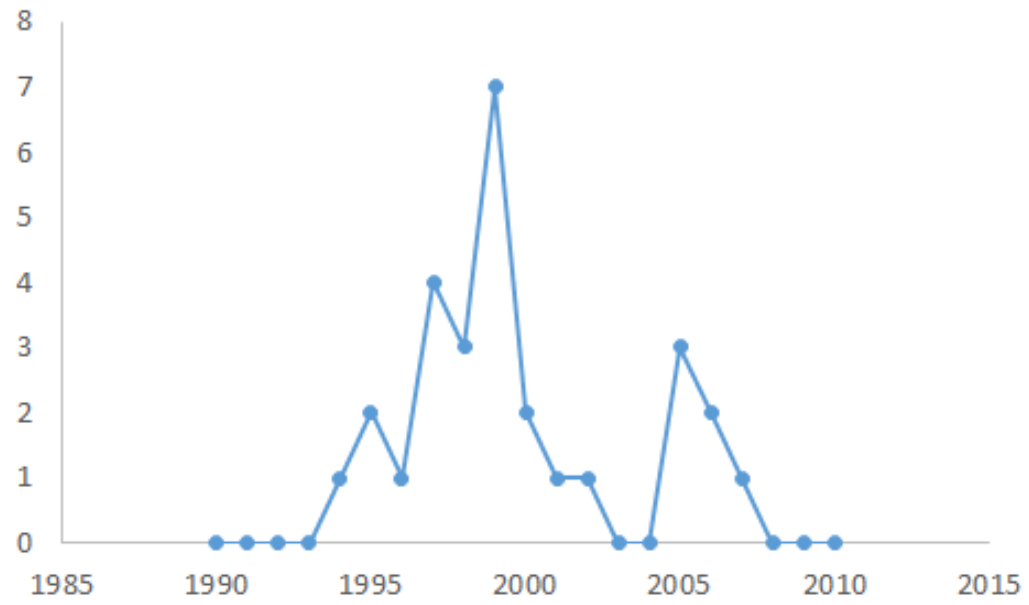


Figure 3.3: Number of States Passing Mental Health Parity Laws, 1996-2010

which require insurers to cover mental health care at the same level as physical health care, have a much greater impact on the cost of premiums than would mandates for inexpensive procedures which few people need" (Council for Affordable Health Insurance, 2012).

These concerns became particularly important as the United States government prepared to pass the Mental Health Parity and Addiction Equity Act of 2008. This was a much more comprehensive act than its predecessor in 1996, requiring parity in deductibles, copayment, coinsurance, out-of-pocket expenses, limits on visits, limits on treatment, and limits on days of coverage.[8] However, it still exempts many smaller employers and still does not mandate the coverage of a full spectrum of illnesses (the illnesses covered is still left to the discretion of the states).

[8]http://www.dol.gov/ebsa/mentalhealthparity/

Table 3.1: States Passing Mental Health Benefit Mandates

MH Parity	Min. Benefit	Mandated Offering	Mandated if Offered
Arkansas (1997)	California (1974)	Arizona (1998)	Alabama (2001)
California (1999)	Colorado (1992)	Indiana (2000)	Georgia (1998)
Colorado (1997)	District of Columbia (1999)	Kentucky (2000)	New York (1999)
Connecticut (1999)	Florida (1992)	Nebraska (2000)	Utah (2001)
Delaware (1998)	Hawaii (1998)	Ohio (1985)	
Hawaii (1999)	Illinois (1991)		
Iowa (2005)	Kansas (1998)		
Illinois (2001)	Massachusetts (1996)		
Louisiana (1998)	Michigan (2000)		
Maine (1995)	Mississippi (2002)		
Maryland (1996)	Nevada (2000)		
Massachusetts (2000)	Oregon (2000)		
Missouri (1999)	Pennsylvania (1999)		
Montana (1999)	South Carolina (1995)		
New Hampshire (1995)	Tennessee (1999)		
New Jersey (1999)	Wisconsin (1998)		
New Mexico (2000)			
New York (2006)			
North Carolina (2007)			
Ohio (2006)			
Oklahoma (1999)			
Oregon (2005)			
Rhode Island (1994)			
South Dakota (1998)			
Vermont (1997)			
Virginia (1999)			
Washington (2005)			
West Virginia (2002)			
Texas (1997)			

Notes: Dates of passage are taken from a variety of sources. Primarily, I use the dates reported by the National Alliance for Mental Illness (NAMI), National Conference for State Legislatures (NCSL), and published literature, including Lang (2013), Cseh (2008), and Buchmueller (2007). Where discrepancies exist, I refer to primary source, legislative documents directly from states

3.2.2 Theory and Literature on Mandates, Parity and Employment

The theory of benefit mandates and the labor market extends from an analysis conducted by Larry Summers in 1989 (Summers, 1989). In it, he argues that mandated benefits for employers, be they health benefits, unemployment

insurance, or others, increase the cost of labor to employers, thus shifting the labor demand curve down. However, mandated benefits also increase the value of a job for workers, and thus simultaneously shift out the supply curve. The amount of the shifts depend on the cost of the mandate to employers and the employee valuation of the benefit he or she receives from the mandate. For instance, if the value of the benefit to the worker is exactly equal to the cost of the mandate to the employer, then in equilibrium employment will remain constant. However, the new equilibrium wage will be lower than in the equilibrium prior to the passage of the mandate. In fact, it will be lower by the exact cost of the benefit. This phenomenon is depicted in Figure 4.

As is seen in the figure, the new equilibrium labor supply (at the point E') remains unchanged (from point E), however the wage has declined. In the case where employees value benefits less than the cost to the employer, the supply curve would shift out to a lesser degree than the shift in labor demand, thus resulting in decreased labor supply and diminished wages. Conversely, if employees valued the benefit more than the cost to employers, we would expect to see an increase in employment.

Regardless of employee valuation of the benefit, however, employers can achieve a new equilibrium by reducing the wages of their employees who take up the benefit. Being that this is the case, then it is necessary to ask why governments would *mandate* benefits at all. That is, employers should be indifferent between voluntarily providing these benefits with reduced wages and not.

In actuality, firms do not voluntarily offer benefits for several reasons. One of these reasons is the possibility of adverse selection. For example, firms that offer benefits for mental health services might be concerned that this policy will only

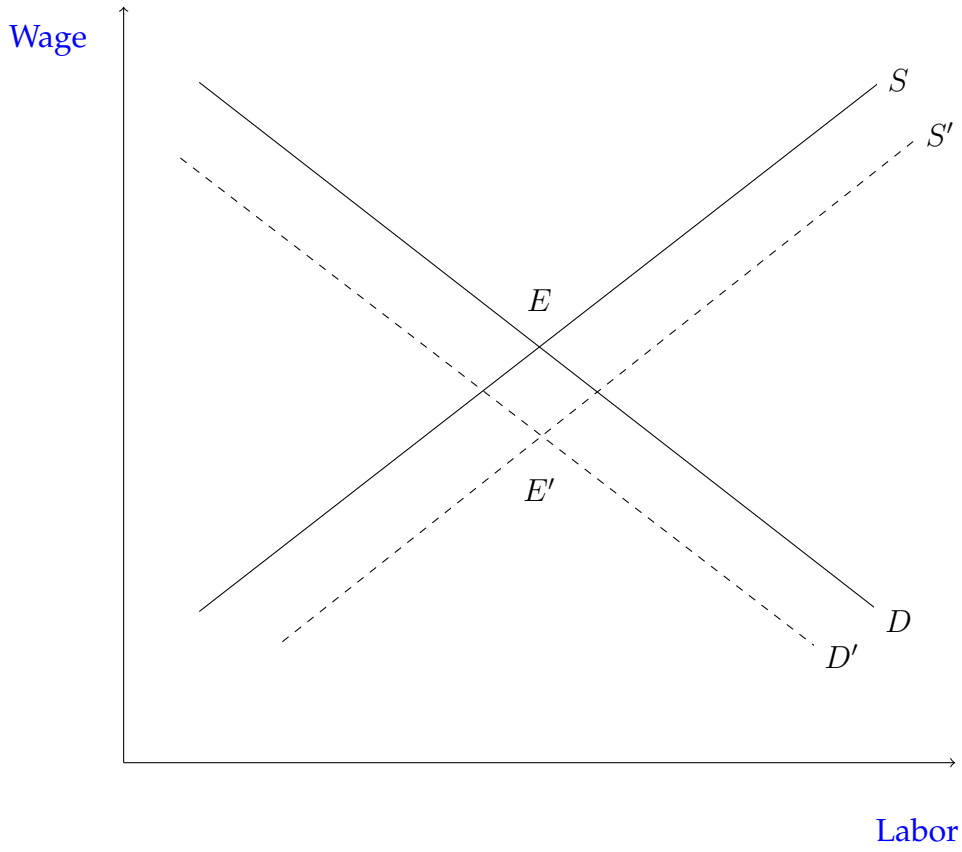


Figure 3.4: The Effect of Mandated Benefits When Cost to Employers is Equal to Value of Consumers

attract employees with mental health disorders, which may decrease overall productivity and therefore lead to reduced profits. Another possibility is moral hazard. Generally, firms are uncertain about a priori employee valuation of benefits. Therefore mandating mental health services, for example, might engender increased utilization of mental health services, which would then lead to higher expenditures by their employees and a greater cost burden for firms. This has led many critics (see introduction above) to speculate that some firms, particularly small firms, would drop insurance coverage altogether, leading more individuals into the ranks of the uninsured. (Matthews, 1999; Goodman & Mitchell, 2002a, 2002c, 2002b).

If firms are reluctant to offer these benefits on their own, then there are many reasons why governments might want to mandate certain health benefits. One is paternalism: there is a sense of fairness that comes with ensuring that individuals are provided with a certain minimum level of benefits. Many, for instance, consider health insurance to be a basic human right that should be afforded to each individual. Moreover, the scenario depicted in Figure 4 works under the assumption that individual employees are rational and value benefits accordingly. However, governments may believe that many individuals undervalue certain benefits–such as mental health services–and would therefore need to be protected against unexpected losses. Finally, there could be positive externalities associated with many of these mandates. For example mandating health insurance benefits might, in the long-run, decrease overall health expenditures in the United States as less individuals would rely on emergency room use. They might also actually *increase* worker productivity.

There has been a considerable amount of literature on the effects of benefit mandates on a wide variety of outcomes, particularly aiming to test whether the labor market consequences predicted by theory match the evidence. Most of this literature has focused on the effect of benefit mandates on the incidence of insurance coverage. This literature set out to test the theory posed by mandate critics that state mandates cause rises in uninsurance rates. In a seminal 1992 paper, Jonathan Gruber investigated the claim that state-mandated benefits caused employers to drop health insurance. He found that that inclusion of five high-cost mandates at the time (continuation coverage, alcohol, drug abuse, mental health minimum benefits, and chiropractic services) had little effect on employer-sponsored insurance coverage (Gruber, 1992). Kaestner and Simon (2002) similarly found no effect of mandates on the probability of having

employer-sponsored insurance, whether including all mandates or just high-cost ones.

Other studies, however, have shown different–and more negative–effects of mandates on insurance coverage. Sloan and Conover found that higher numbers of mandated benefits increased the probability that an individual was uninsured within the state of the mandate (Sloan & Conover, 1998). Jensen and Gabel also used data on firms in the late 1980s to estimate the decision of small firms to offer health insurance. They observed that a large percent of uninsurance among sample businesses in the late 80s was attributable to state-mandated benefits (Jensen & Gabel, 1992).

However, employers may respond to benefit mandates in ways that do not involve dropping coverage. Some of these involve either reducing employment or dropping wages, as predicted by the Summers model discussed above (Baicker & Chandra, 2006). In a 1994 paper, Gruber investigated the labor market effects of a particular set of federally and state-mandated benefits: those that comprehensively covered childbirth. He found that the inclusion of such mandates caused employers to shift costs to women of childbearing age in the form of reduced wages, but consequently had little effect on employment at the extensive margin (i.e. there was little change in labor demand) (Gruber, 1994). That is, those who were benefiting (or most likely to benefit from) the particular mandate were paying a premium for that benefit. This perfectly matched the theory depicted in Figure 4. Similarly, Thurston (1997) found that a mandatory employer provision of health insurance in Hawaii caused the Hawaiian industries most affected by the mandate to experience slower wage growth to other Hawaiian industries (Thurston, 1997). Kaestner and Simon, however, found

no significant wage-shifting in their analysis of a wide variety of mandates, in contrast with Gruber and Thurston (Kaestner & Simon, 2002).

There are other labor market margins on which employer could react to the added cost of a mandate as well. For instance, instead of reducing employment or wages, employers could simply reduce the hours of their employees (or only hire part-time workers) so as to exclusively attract employers that would not have to be covered by health insurance. Indeed, Baicker and Chandra find evidence that rising health insurance premiums significantly decrease working hours, and raise the probability of being employed part-time (Baicker & Chandra, 2006, 2005). This is a particular concern with the Affordable Care Act, as many critics have postulated that the added costs of the comprehensive reform to employers would result in a reduction in full-time employees.[9] However, recent analysis has pointed out that this has not yet been the case. Dubay et al. in particular, use the Massachusetts health reform in 2006 to study employment growth, finding that there were no significant negative trends to either full-time or part-time employment after the reform took place (Dubay, Long, & Lawton, 2012).

There have also been studies done on the effects of various benefit mandates on utilization and outcomes. Bundorf, Henne and Baker, for example, have estimates the effects of infertility treatment regulations on utilization and outcomes of infertility treatments and found that use is significantly greater in states that have implemented these mandates (M. K. Bundorf et al., 2007). That is, they found evidence of significant moral hazard in response to these mandates.

[9] http://online.wsj.com/article/SB10001424127887324694904578601922653718606.html and http://economix.blogs.nytimes.com/2013/07/03/the-new-economics-of-part-time-employment/

The evidence on the effects of mental health parity, in particular, on various outcomes has been mixed. McGuire and Montgomery found that initial "minimum mandated benefits" in the 1970s resulted in increased utilization, approximating a 10-20% increase in psychiatric services. (McGuire & Montgomery, 1982). McGuire also found that early mental health mandates resulted in the reduction of stigma associated with seeking mental health coverage. In estimating the effects of mental health parity legislation, Buchmueller et al. found that despite increases in states with mandated benefits, parity legislation only affects about 20% of workers due to the self-insurance restriction due to 1974's ERISA laws (T. C. Buchmueller, Cooper, Jacobson, & Zuvekas, 2007). Barry and Busch found that parity reduces the financial burden on families of children with mental health disorders (Barry & Busch, 2007). Barry, Frank and McGuire concluded that comprehensive parity would have little impact on total spending (Barry, Frank, & McGuire, 2006). Despite this positive evidence, however, Bao and Strum (2004) have found that parity has no effect on perceived generosity or quality of coverage (uhua Bao & Strum, 2004). However, Shinogle and Salkever (2005) found the opposite using cross-sectional firm data: they found that mental health parity legislation increased mental health generosity offered by firms (Shinogle & Salkever, 2005).

Some recent literature re-examined the older evidence on the effects of mental health mandates. In particular, Busch and Barry (2008) found that, once controlling for firm sizes and mental health status, mental health parity mandates were found to have a significant effect on utilization of mental health services, especially among people with mental health disorders (Busch & Barry, 2008). In a very recent study, Lang (2013) found that mental health parity mandates significantly decreased suicide rates in states (Lang, 2013).

Despite all this evidence, the effect of mental health parity on the labor market has gone relatively understudied. Cseh recently used the Annual Supplement to the Current Population Survey (CPS) to study the effects of parity on employment levels, wages, and hours, finding no significant effects in any of the labor market categories (Cseh, 2008). However, this analysis is limited in several ways. First, there may be masked labor market effects not picked up by employment levels. For instance, a mandate may have an effect on job flows, such as workers transitioning from old jobs to new jobs, which would have insignificant effects on overall employment levels. Second, the use of the CPS does not allow researchers to focus on the population most affected by mental health parity: those with mental illnesses. Third the CPS can only identify whether individuals transitioned from employment to unemployment (or vice versa), but not whether individuals transferred from one job to another. Fourth, Cseh's analysis stops at states that have passed parity laws in 2002, though many have since done so (including New York State, Oregon and Ohio).

Therefore, it is natural to look at the effects of mental health parity on job flows. In particular, there is a wealth of literature on the effects of health insurance and mandates on a phenomenon known as "job lock"–where employees remain in jobs longer than they would otherwise like due to a fear of losing comprehensive benefits. Madrian (1994) found that access to employer-sponsored health insurance decreased the voluntary turnover rate by 25%. (Madrian, 1994). Gruber and Madrian (1997) found that mandates providing continuation coverage after job separation led to an increase in the number of individuals who separate and the total time spent jobless, though at least some of this time is spent in productive search (Gruber & Madrian, 1997). Most recently, Garthwaite et al. (2013) found an immediate increase in job search behavior

following the disenrollment of approximately 170,000 Tennessee residents from public health insurance in 2005 (Garthwaite, Gross, & Notowidigdo, 2013).

3.3 Data

3.3.1 Individual-Level Data

Individual-level data comes from the Household Component of the 1996-2010 Medical Expenditures Panel Survey (MEPS). The MEPS full-year consolidated data files contains monthly information on detailed health insurance coverage (including source of coverage, i.e. private, Medicare, Medicaid, employer, etc.), as well as periodic (5 rounds over 2 years) data on employment status, job switching, health expenditures by source (hospital, inpatient, outpatient, etc.) and payer (out-of-pocket, Medicare, Medicare, etc.), and individual demographics (firm size, age, education, race, etc.). There are approximately 30,000 individuals sampled on a nationwide basis each year. It is a semi-rolling-panel, in that it follows individuals for two years on a rolling basis. Each individual is observed for 5 rounds over the course of two years, and those individuals are asked the same employment and health insurance questions each round. Therefore, I am able to see job turnover for a particular individual within this two-year period by comparing individual employment data between rounds. In addition, the MEPS contains several component and event files each year that can be merged onto the full-year consolidated data file: inpatient, outpatient, prescription drugs, office-based visits, and emergency room. These event files contain information on individual medical care utilization and expenditures for

particular sources and particular conditions. The conditions are identified by three-digit ICD-9 codes within each event file.

Using these ICD-9 codes, I am able to identify individuals who sought medical care particularly for mental health services, either as an inpatient, outpatient, office-based visit, emergency room, or whether an individual got a prescription for mental health drugs. I use three-digit codes ranging from 290-319, which are consistent with the mental health disorder definitions used by other datasets such as the National Ambulatory Medical Care Survey (NAMCS), National Hospital Ambulatory Medical Care Survey (NHAMCS), and National Hospital Discharge Survey (NHDS). For the purposes of this paper I classify individuals as having sought medical care for mental illness if they were associated with any of the aforementioned condition codes. Using the merged files, I am thus able to estimate not only the proportion of individuals nationally that sought medical care services and can be identified as having a mental health disorder, but also expenditures by payer (private insurance, public insurance, out-of-pocket, etc.) specifically for these mental illnesses.

In this way, the MEPS presents a distinct advantage for the analysis of the effects of mental health parity mandates. Though most analyses of the effect of mandates on the labor market use the Current Population Survey (CPS), the CPS is limited in its ability to identify utilization, expenditures, and the population most affected by mandates (i.e. mental health disorders). Furthermore, a distinct advantage of the MEPS is that I can not only identify when an individual transitioned from employment to unemployment, but I can actually identify whether an individual *switched* jobs, i.e. when an individual's primary job changes from one reported in the previous round.

In addition, I am able to identify the reason an individual changed a job, which is how I define "voluntary" versus "involuntary" transitions. The MEPS provides a variable which asks why an individual changed jobs between rounds of the panel, and the respondent is allowed to provide a variety of reasons. Here, I define "voluntary" transitions as those transitions involving "quitting," most of which come from the category "quitting to take another job." I define involuntary transitions as responses including having the 'job end," being "laid off," or having an "illness or injury."

A limitation of the MEPS that has prevented this sort of analysis in the past is that the public-use files, though a precise source of household medical expenditures, are nationally representative and do not contain information about expenditures or utilization at the state level. The restricted-access MEPS, however, allows me to merge in state identifiers and state-level weights to the public-use files for the 29 most densely populated states. I currently have access to this restricted-level data from 1996 through 2010, therefore I am able to match individuals to their respective states. Hence, I am able to see state variation in insurance and employment and expenditures.

For the purposes of my analysis, I restrict the sample to individuals in either full-time or part-time employment and have been flagged as previously having a mental health disorder. In subsequent specifications, I also limit the data to individuals who have employer-sponsored-insurance and stratify the data by firm size, as this allows me to see the effects on individuals affected by the mandate. I only limit the analysis at 2007, so as to not pick up any potential effects of the Great Recession, which began in December 2007, which may bias the results.

Table 2 shows the summary statistics for my sample for both the entire pop-

ulation in the data and for the "mental health" population. It is of note that my main dependent variables of interest (job separation and Employment to Employment transitions) are both higher for the mental health population than they are for the population as a whole. In addition, employment is slightly lower among the mental health population than the population as a whole. This lends legitimacy to the argument that the mental health population might have costs that inhibit them from staying in a particular job for a long time and that the addition of mandates might reduce those costs, thereby reducing both voluntary and involuntary turnover rates. It is also of note that the mental health population tends to be disproportionately female, to be non-Hispanic white, and to have fewer advanced degrees than the population as a whole.

3.3.2 Mental Health Parity

The variable of interest in this paper is whether a state has passed mental health parity legislation. Mental health parity refers to a state-level mandate that insurance companies provide mental health benefits to group plans equal to the level of coverage provided for medical and surgical inpatient and outpatient plans (i.e. if a particular plan provides 30 days of inpatient care, then so must the plan provide 30 days of inpatient care for mental health disorders, etc.). I obtained dates of passage of mental health legislation primarily from several different sources, including The National Alliance on Mental Illness (NAMI), the National Conference on State Legislatures (NCSL), and published sources (National Alliance on Mental Illness, 2009; National Conference on State Legislatures, 2011; Barry & Busch, 2007; T. C. Buchmueller et al., 2007; Gitterman et al., 2001; Lang, 2013; Cseh, 2008). The NAMI keeps detailed records of state

Table 3.2: Summary Statistics

	Whole Population		Mental Health Population	
Variable	Mean	Std. Dev.	Mean	Std. Dev
Total MH Expend.	1497.140	4507.152	1497.140	4507.152
OOP MH Expend.	312.689	907.760	312.689	907.760
Age	34.038	22.358	41.549	20.446
Separated from Job	0.366	0.481	0.429	0.494
E to E Transition	0.203	0.402	0.227	0.419
Mental Health Disorder	0.133	0.339	x	x
Employed	0.663	0.472	0.592	0.496
ESI	0.487	0.499	0.538	0.499
Any Ins	0.842	0.365	0.889	0.313
Married	0.391	0.488	0.397	0.489
Female	0.523	0.500	0.616	0.486
Less than HS	0.386	0.487	0.367	0.482
HS Grad	0.226	0.418	0.186	0.389
Some College	0.145	0.352	0.185	0.389
Bachelors or Higher	0.243	0.429	0.185	0.389
Non Hisp White	0.561	0.496	0.701	0.458
Non Hisp Black	0.167	0.373	0.124	0.329
Hispanic	0.273	0.445	0.175	0.380
	N=331,853		N = 53,394	

Notes: Columns 2 and 3 are the means and standard deviations of pertinent variables for the entire dataset from 1996 through 2007, excluding encrypted states (those my data cannot identify). Columns 4 and 5 are the means and standard deviations of pertinent variables for the mental health population from 1996 through 2007, excluding encrypted states. For Total MH Expenditures and OOP MH Expenditures, Columns 2 and 3 are identical to 4 and 5, as these expenditures only exist for the "mental health" population.

mental health parity laws based on state statutes, House and Senate bills. They also distinguish between the type of mandate passed (minimum mandated benefit, comprehensive parity, etc.) as well as provide estimates years of passage. The NCSL similarly provides information on years of passage of mental health legislation by type of mandate.

There are, however, many challenges to obtaining precise dates of passages for mental health legislation. One challenge is that mandates legislation is amended frequently at the state level, either by changing the included conditions covered by the mandate, altering the copayment and coinsurance requirement, or changing the groups affected by the legislation. Hence, there is considerable heterogeneity making it difficult to characterize mental health parity. In particular, some states have passed comprehensive parity legislation, meaning all group plans within the state are required to provide equal coverage of a broad range of mental health disorders, without exemption of any significant policy group. Other states have passed more broad-based parity, which include equal coverage of a broad range of diseases, but many include some limitations or exemptions, in particular restrictions on co-payments or costs. Most states have passed some form of limited parity, which provides coverage equal to that of medical benefits, but for a limited set of diseases. These are generally restricted to "serious" mental illnesses, including schizophrenia, schizoaffective disorder, psychotic disorders, bipolar disorder, major depression, panic disorder, and obsessive compulsive disorder. Unlike most broad-based and comprehensive parity mandates, these do not require coverage for substance abuse. In addition, these laws may impose limits on this coverage based on time or finances (i.e. they may impose cost-sharing requirements).

There is also heterogeneity in whether the legislation mandates coverage or mandates an offer of coverage, as described in section 2. Any legislation mandating coverage requires that employees of the group plan purchase mental health insurance. However, some parity legislation, i.e. "mandated if offered" laws, require equal coverage of mental health disorders only if the insurance plan already provides some level of mental health coverage. Other legislation, i.e. "mandated offering," requires only the offer of coverage, but does not require employees to actually purchase.

Finally, there are many state mandates which do not offer mental health parity in the sense of providing equal coverage, but do require certain minimum mandated benefits. These plans require the purchase of some level of mental health coverage, but this coverage is usually limited and not equal in benefit levels to that of the provided medical care.

For the purposes of this paper, l only classify a state having passed mental health parity legislation if it had enacted legislation requiring *equal* coverage of mental health benefits before the year 2007 (the strongest mandates available). That is, I include states that provided comprehensive, broad-based or limited parity. I do not include states that had "mandated offering" or "mandated if offered" laws, nor do I include states that provided "minimum mandated benefits" as opposed to parity. Where differences in classifications existed between the sources mentioned above, I use primary-sources (legislative documents) to settle the discrepancies.

3.4 Methods and Results

3.4.1 The Effect of State Mental Health Parity on Employment and Job Transitions

The basic model is a difference-in-differences framework:

$$y_{ist} = \beta_0 + \beta_1 M_{st-1} + \beta_2 X_{ist} + \tau_t + \delta_s + \epsilon_{ist} \tag{3.1}$$

Here, the variables are defined as follows:

y_{ist} is an indicator for whether individual i is employed in state s at time t OR individual i switched jobs in state s in between time t and time $t+1$. For the transition analysis, my dependent variable is whether an individual *ever* transitioned between t and $t+1$ in his or her time in the panel.

M_{st-1} is an indicator for whether state s had passed mental health parity legislation by time $t-1$. I use time $t-1$ in the model, as these refer to the *passage* dates of the mandates. Most parity laws, however, take a year to go into effect. Therefore mandates passed in time $t-1$ are likely to take effect in time t, where the effects would be seen.

τ_t is a set of year dummies and

δ_s is a set of state fixed effects.

X_{ist} refers to a set of observable covariates. These include age and its square, indicators for education groups, indicators for race, an indicator for female, an

indicator for married, an interaction term of female and married, an indicator for whether the employee was part-time worker (working less than 35 hours per week).

I estimate the model above using a linear probability model to determine the effect of having passed mental health parity legislation on the probability of being employed and having switched jobs. The estimator, β_1, is the difference-in-difference estimator that is identifying this effect.

A negative estimate on β_1 would imply that individuals with mental health conditions are less likely to separate from their jobs or be employed in states that pass mental health parity legislation than in states that do not pass legislation. Looking at the model in section 2.3, this would identify the scenario in case 1 (the individuals who, following parity, obtain more comprehensive mental health coverage, thus reducing transitions). A positive coefficient would identify the scenario in case 2 (individuals who, following parity, move to more attractive jobs, thus increasing transitions). Table 3 shows the results of this model on both employment and job separations. Column 1 in Panel A shows the effect of employment conditional on the mental health population (i.e. those flagged as having a mental health disorder). Consistent with prior literature, mental health parity appears to have a very tiny and insignificant effect on employment, even limiting to the mental health population.

Column 2 shows the effects on job separations conditional on mental health and employed populations. Panel A shows the effect on total job separations (i.e. transitioning from employment to either employment at another current-main-job or unemployment or out-of-labor-force). The effect of mental health parity here is negative, though insignificant. However, Panel B, which shows

Table 3.3: The Effect of State Mental Health Parity on Job Transitions

	(1)	(2)	(3)	(4)
	Employed	Separation	Separation (ESI)	Separation (Unins)
	A: Difference-in-Difference Estimates: Total Separations			
MH Parity Passed	0.004	-0.019	-0.036**	0.038
	(0.008)	(0.015)	(0.017)	(0.052)
Obs.	42468	25553	13632	2273
	B: Difference-in-Difference Estimates: E to E Transitions			
MH Parity Passed	x	-0.037***	-0.037***	-0.026
	x	(0.011)	(0.017)	(0.052)
Obs.	x	25553	8046	2273
	C: Difference-in-Difference Estimates: E to U Transitions			
MH Parity Passed	x	0.004	0.003	0.030
	x	(0.012)	(0.013)	(0.042)
Obs.	x	25553	13632	2273

Notes: Robust standard errors clustered at the state year level shown in parentheses. All models include state and year fixed effects. Column 1 is conditional on population of everyone diagnosed with a mental health disorder. Column 2 is conditional on the mental health population and anyone who was employed. Column 3 is conditional on the mental health, employed, and population of individuals with employer-sponsored insurance. In panel B, this is further reduced to those with full-year ESI. Column 4 is conditional on the mental health, employed, and full-year uninsured population.
*Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level

the effects exclusively on employment to employment transitions shows significant effects of mental health on job switching. The implication here is that the passage of mental health parity legislation reduces job switching by about 3.7 percentage points. Panel C shows the effect of parity exclusively on employment to unemployment/out-of-labor-force transitions. Here, I see a very small and insignificant effect.

The negative effect of mental health parity on job transitions is consistent

with case 1. In order to explore this further, column 3 of Table 3 looks at the effects of parity mandates conditioned on the mental health employed population who also previously held employer-sponsored insurance (ESI) at their job. For panel B, since I are looking exclusively at employment to employment transitions, I limit this further to only look at individuals who had ESI for the duration of their time in the panel (two years). If the observed reduction comes primarily from individuals gaining job security or a possible treatment effect of the mandate, then this should then be more pronounced in the insured population.

Indeed, this is what is shown in column 3. A mental health parity mandate caused a significant decline in overall job separations (panel A), that is now significant and more negative than job separations in column 2. For employment-to-employment transitions (Panel B), there is also a highly negative and significant effect of parity mandates. For employment-to-unemployment transitions (Panel C), the effect is still small and insignificant.

As a further specification check, column 4 of Table 3 looks at the full-year uninsured population. For overall separations (Panel A), the sign is now reversed, implying that mandates cause an approximate 3.8 percentage point *increase* in job separations. However, the estimate is insignificant. For employment-to-employment transitions among the mental health, employed, and uninsured population (Panel B), the sign is still negative, but considerably smaller than for the insured population, and also insignificant. Finally, for employment-to-unemployment transitions, the sign is more positive than the insured population, yet still insignificant.

Columns 3 and 4 confirm that the effects of parity mandates are much larger

for the insured population, lending legitimacy to the claim that mental health mandates might reduce unnecessary job turnover. We would expect the uninsured population to have either positive or small effects following the passage of mandates, since they are mostly unaffected by the benefits unless they obtain insurance. The one mechanism through which the uninsured could be affected by mandates is if they switch to a job that offers benefits following the passage of the mandate. This, however, does not seem to be the case given the estimates in Table 3. It is also useful to decompose the effects of mandates as being either voluntary or involuntary transitions. Our model predicts that effect of mandates on involuntary transitions should be negative, however mental health parity could theoretically either increase or decrease voluntary transitions. A decrease, as observed in Table 3, could occur if mandates reduce voluntary turnover for those individuals who do not already have comprehensive mental health coverage, and that would have taken other jobs previously to get these benefits. However, it could also theoretically increase voluntary turnover for individuals who previously held mental health coverage if the widespread prevalence of mental health benefits following a mandate allowed individuals to take new jobs they otherwise would not have (a reduction in "job lock"). Unfortunately, I am unable to identify the type of benefits individuals have and thus whether individuals previously had mental health coverage. However, an analysis on whether individuals simply have ESI coverage is revealing.

To see the effects on voluntary and involuntary transitions, I turn to Table 4. Column 1 shows the effect of mental health parity on involuntary transitions for the mental health employed population. The effects of parity on total job separation, employment to employment transitions, and employment to unemployment transitions are all small and insignificant. However, looking at Col-

umn 2, which limits the population to those who previously held ESI, shows different results. The effect of mental health parity on total involuntary job separations (Panel A) and employment to unemployment transitions (Panel C) for this population is also small and insignificant (Panel A). However, the effect on employment to employment transitions (Panel B), shows a negative and significant decline in involuntary transitions of about 1.2 percentage points due to the passage of mandates.

Table 3.4: The Effect of State Mental Health Parity on Job Transitions

	(1)	(2)	(3)	(4)
	Involunt. Sep.	Involunt. Sep. (ESI)	Volunt. Sep.	Volunt. Sep. (ESI)
	A: Difference-in-Difference Estimates: Total Separations			
MH Parity Passed	-0.004	-0.000	-0.012	-0.020
	(0.010)	(0.011)	(0.011)	(0.014)
Obs.	25553	13362	25553	13632
	B: Difference-in-Difference Estimates: E to E Transitions			
MH Parity Passed	-0.009	-0.012**	-0.019**	-0.023**
	(0.007)	(0.005)	(0.010)	(0.011)
Obs.	25553	8046	25553	8046
	C: Difference-in-Difference Estimates: E to U Transitions			
MH Parity Passed	0.005	0.011	0.003	0.005
	(0.006)	(0.009)	(0.007)	(0.007)
Obs.	25553	13362	25553	13362

Notes: Robust standard errors clustered at the state year level shown in parentheses. All models include state and year fixed effects. Columns 1 and 3 are conditional on population of everyone diagnosed with a mental health disorder and employed. Columns 2 and 4 are conditional on the mental health and employed population as well as those with employer-sponsored insurance. In Panel B, this is further reduced to those with full-year ESI.
*Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level

Similarly, looking at Column 3, there is a negative and insignificant effect

of parity on total voluntary separations (about 1.2 percentage points, Panel A). I once again see a small, positive and insignificant effect on employment to unemployment transitions (Panel C). However, I observe a decline in voluntary employment to employment transitions of about 1.9 percentage points, which is a significant result (Panel B). Note that this is among the mental health employed population. Looking at Column 4, which again conditions in the full-year ESI population, I see a stronger negative effect of employment to employment transitions (mandates cause a 2.3 percentage point decline in job switching). This is consistent with the theory that both voluntary and involuntary transitions would decline among the insured population (case 1 in section 2.3). Indeed, especially among the employment to employment transitions, I observe that about one third of the decline in total separations comes from involuntary transitions and the two thirds from voluntary transitions.

The fact that job separations, particularly involuntary separations, decline following the passage of a mandate suggest the possibility that mandates are having a treatment effect. That is, individuals who previously did not have mental health coverage but now do increase utilization of mental health services, become healthier, and become more productive in the workforce. Although I so far have not been able to test the effects of mental health parity on health outcomes, I use mean health care expenditures among the mental health population as a proxy for utilization and test the model above on this dependent variable.

Table 5 shows these results. Columns 1 and 2 show the effects of mental health parity mandates on mean real total mental health expenditures for the mental health employed population and mental health employed population

with ESI coverage. These estimates reveal that mean expenditures for mental health services among workers in states that have passed parity mandates have been between $250 and $300 higher than those in states without mandates. As a falsification test, I also included the effect of parity in mean real out-of-pocket (OOP) expenditures by the same population, as the goal of mental health parity is to reduce such expenditures among the employed and insured, having private insurance take up more of the burden. Indeed, the rise in OOP expenditures in states that have passed parity mandates has been negligible and insignificant, confirming that the increase in expenditures is likely due to private insurance–an anticipated effect of the mandates. This also further supports the theory that the reduction in involuntary turnover among this population is due to increased use of mental health services.

Table 3.5: The Effect of State Mental Health Parity on Mental Health Expenditures

	(1)	(2)	(3)	(4)
	Total Exp.	Total Exp. (ESI)	OOP Exp.	OOP Exp. (ESI)
	Difference-in-Difference Estimates: Total Separations			
MH Parity Passed	257.698*	300.847**	3.358	4.126
	(135.101)	(137.926)	(32.943)	(41.587)
Obs.	19012	10163	19012	10163

Notes: Robust standard errors clustered at the state year level shown in parentheses. All models include state and year fixed effects. Columns 1 and 3 are conditional on population of everyone diagnosed with a mental health disorder and employed. Columns 2 and 4 are conditional on the mental health and employed population as well as those with employer-sponsored insurance.
*Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level

3.4.2 The Effect of the Mental Health Parity and Addiction Equity Act on Employment and Job Transitions

To test the effects of the federal parity law in 2008, I restrict the sample to the years 2006 through 2010, and employ a similar framework to that of equation (6):

$$y_{ist} = \beta_0 + \beta_1 NM_{st-1} + \beta_2 I\{t >= 2009\} + \beta_3 NM_{st-1} I\{t >= 2009\} + \beta_4 X_{ist} + \tau_t + \delta_s + \epsilon_{ist} \tag{3.2}$$

y_{ist} is an indicator for whether individual i is employed in state s at time t OR individual i ever switched jobs in state s in between time t and time $t+1$.

NM_{st-1} is an indicator for whether state s did *not* pass mental health parity laws in time $t-1$

$I\{t >= 2009\}$ is an indicator for whether it is year 2009 or 2010

τ_t is a set of year dummies and

δ_s is a set of state fixed effects.

X_{ist} are the same covariates as the state model.

In this mode, β_3 is our parameter of interest. Since the 2008 law mandates mental health parity, it would primarily apply to those states that had not yet passed mental health parity legislation. Further, the law went into effect in 2009, which is why our variable of interest interacts the years 2009 or 2010 with these states that had not yet passed parity legislation. Similar to the specification in

section 4.1, a negative sign on β_3 would imply a reduction in job transitions, consistent with case 1 of the model in section 2.3. A positive sign would identify a positive transition effect for individuals with mental health conditions in states that had not yet passed parity legislation, consistent with case 2 of the model.

Table 6 shows the results of the effect of the federal law. Overall, the estimates are similar to those in Table 3, however much of the significance is lost (likely due to small samples of job separations in the 5 year window). Column 1 shows that the job separations for the mental health, employed population. It implies that the federal legislation caused a decline in job separations by about 2.1 percentage points, though this estimate is insignificant (Panel A). Job transitions, however, declined by about 2.6 percentage points (Panel B), and is a significant estimate. Employment to unemployment transitions (Panel C) is again small and insignificant.

Column 2 again limits the population to the population of those that had ESI. Here, total job separations was small and insignificant (Panel A), as was the estimate for employment to unemployment transitions (Panel C), however transitions from employment to employment went down by approximately 3.8 percentage points, a significant estimate (Panel B). Column 4 provides the effect on the uninsured population for comparison. All the estimates are insignificant, though this could be due to the small sample size of the employed, uninsured population in this subset of years.

The negative estimates as a result of the Federal law are encouraging in light of the state estimates. In particular, I am looking at an entirely different group of states for these estimates (the treatment group is the group of states that had *not*

Table 3.6: The Effect of Federal Mental Health Parity on Job Transitions

	(1)	(2)	(3)
	Separation	Separation (ESI)	Separation (Unins.)
	A: Difference-in-Difference Estimates: Total Separations		
MH Parity Passed	-0.021	-0.008	0.022
	(0.017)	(0.022)	(0.070)
Obs.	11230	5760	1227
	B: Difference-in-Difference Estimates: E to E Transitions		
MH Parity Passed	-0.026**	-0.038**	0.073
	(0.013)	(0.015)	(0.067)
Obs.	11230	5760	1227
	C: Difference-in-Difference Estimates: E to U Transitions		
MH Parity Passed	-0.010	0.018	-0.027
	(0.016)	(0.019)	(0.056)
Obs.	11230	5760	1227

Notes: Robust standard errors clustered at the state year level shown in parentheses. All models include state and year fixed effects. Column 1 is conditional on the mental health population and anyone who was employed. Column 2 is conditional on the mental health, employed, and population of individuals with employer-sponsored insurance. In panel B, this is further reduced to those with full-year ESI. Column 3 is conditional on the mental health, employed, and full-year uninsured population.
*Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level

passed mandates previously. Therefore, if we believed the negative estimates in Table 3 to be associated primarily with characteristics of the states that had passed mandates, then the fact that the effects of the Federal law are similarly signed as the state mandates on an entirely different group of states suggest that the mandates are having an effect. The next section further explores some additional robustness checks.

3.4.3 Additional Robustness Checks

One possibility often unexplored in the mandates literature is to use firm size as a potential control. The reason for this is that 1974's Employee Retirement Income Security Act (ERISA) exempts firms that self-insure from state-level regulation.[10] Though I cannot perfectly identify firms who self-insure in the individual-level MEPS, it is well-known that very large firms predominantly self-insure. In 2010, firms with 1,000 employees or more had 83.6 percent of their enrollees in very small plans, contrasted with 12.5 percent of enrollees being in self-insured plans in small firms (25 or less) (Crimmel, 2010). Thus, large firms serve may serve as an adequate control sample for the effect of mandates within a particular state.

In addition, many mental health parity laws (and other benefit mandates) exempt very small firms for fear of imposing too large a cost burden and reducing the firms' probability of offering insurance at all to their employees. Unfortunately, state mandate laws are heterogeneous in specifically the small-groups that are exempt. Nevertheless, looking at small firms might also serve as an adequate control on the effects of mental health parity.

To that effect I ran model (6) on individuals who in the panel worked for small firms, which I defined as having between 1 and 24 employees, medium firms, which I defined as having 25 to 99 employees, and large firms, which I defined as having 100 or more employees. I expect to see the largest results occur in the medium sized firms. The results of this are detailed in Table 7.

Column 1 of Table 7 shows the effect on job separations for small firms for the

[10] http://www.dol.gov/compliance/laws/comp-erisa.htm

mental health employed population. The effect on total separations (Panel A) is extremely small and insignificant, as is the effect on employment to unemployment transitions (Panel C). However, the effect on employment to employment transitions (Panel B) is negative and significant. Column 2 shows the effect of parity on medium-sized firms. The effect on total separations here is extremely significant and negative. In fact, it appears as though parity mandates here lead to a decline in job separations of approximately 5.2 percentage points. Note that this is a larger effect than the one observed in Table 3, which confirms our theory. The effect is also negative and insignificant for employment to employment transitions (Panel B). The effect on employment to unemployment transitions (Panel C) is still small and insignificant. Finally, Column 3 of Table 7 shows the effect on separation for large firms. There are negative and insignificant results in total separations (Panel A) and employment to employment transitions (Panel B), though as expected these are smaller in magnitude than the effects on the medium sized firms.

Indeed, I do see the largest effect for medium-sized firms, as expected, suggesting that these parity laws do have an effect. Although I do see some significant results in the small firm and large firm categories, this could be due to imperfect identification of the small firms that mental health parity laws exempt in certain states and possibly picking up large firms that do not self-insure.

Another potential concern with the estimates presented in Table 3 is the possibility of policy endogeneity. While state fixed effects should pick up most state-specific characteristics, there could still be characteristics of states or macroeconomic conditions that correlate heavily with the passage of parity laws, biasing the estimates. To investigate the possibility of such policy endo-

Table 3.7: The Effect of State Mental Health Parity on Job Transitions by Firm Size

	(1)	(2)	(3)
	Separation (1-24 Employees)	Separation (25-99 Employees)	Separation (100+ Employees)
	A: Difference-in-Difference Estimates: Total Separations		
MH Parity Passed	-0.001	-0.052**	-0.030*
	(0.019)	(0.024)	(0.018)
Obs.	12023	7176	9580
	B: Difference-in-Difference Estimates: E to E Transitions		
MH Parity Passed	-0.048***	-0.048***	-0.039**
	(0.018)	(0.021)	(0.017)
Obs.	12023	7176	9580
	C: Difference-in-Difference Estimates: E to U Transitions		
MH Parity Passed	0.013	-0.021	-0.008
	(0.017)	(0.022)	(0.017)
Obs.	12023	7176	9580

Notes: Robust standard errors clustered at the state year level shown in parentheses. All models include state and year fixed effects and are conditional on the population of individuals with a mental health disorder and employed. Column 1 is also conditional on those working for small firms (1-24 employees). Column 2 is also conditional on those working for medium firms (25-99 employees). Column 3 is conditional on those working in large firms (100 or more employees).
*Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level

geneity, I test whether the results in Table 3 are sensitive to the inclusion of state by year unemployment rates. I obtain these official unemployment from the Bureau of Labor Statistics (BLS).[11] In addition, I add the state by year proportion of individuals with a mental health diagnosis as controls in the model to test for the possibility that parity mandates are legislated due to a high presence of

[11]http://www.bls.gov/lau/

people with mental health disorders in the region. The results are presented in Table 8.

Column 1 of Table 8 shows job separations for individuals with a mental health condition who were employed, controlling for state by year proportions of people with mental health conditions. This control appeared to have very insignificant effects on the estimates, as the coefficients in all three panels are nearly identical to those in Table 3, column 2. The same is true of Column 2, which limits the population to those with ESI (and those with full-year ESI in the case of panel B).

Columns 3 and 4 control for state by year unemployment rates. Column 3 presents the results for the mental health, employed population (analogous to Column 2 of Table 3). Controlling for unemployment rates appeared to have reduced the estimate of total job separations (Panel A) from -0.019 to -0.010 (Panel A). However, it appears as though most of this comes from employment to unemployment transitions (Panel C), as this estimate increased from 0.004 to 0.009. Employment to employment transitions (Panel B) only changed slightly, remaining largely negative and significant, implying mandates caused about a 3.6 percentage point decline in job switching.

Column 4 limits this again to the ESI population. Again, it appears as though including unemployment rates reduced the estimate of overall job separations (Panel A) and that this mostly came from the employment to unemployment transitions (Panel C). Employment to employment transfers remained relatively unchanged, still with a negative and significant estimate.

This demonstrates that, to the extent that broader macroeconomic trends had

Table 3.8: Endogeneity Checks: The Effect of State Mental Health Parity on Job Transitions

	(1)	(2)	(3)	(4)
	Separation	Separation (ESI)	Separation	Separation (ESI)
	A: Difference-in-Difference Estimates: Total Separations			
MH Parity Passed	-0.020	-0.037**	-0.010	-0.025
	(0.015)	(0.017)	(0.015)	(0.017)
Mean MH	Yes	Yes	No	No
Unemp. Rate	No	No	Yes	Yes
Obs.	25553	13362	25553	13632
	B: Difference-in-Difference Estimates: E to E Transitions			
MH Parity Passed	-0.037***	-0.039***	-0.036***	-0.034***
	(0.011)	(0.011)	(0.011)	(0.012)
Mean MH	Yes	Yes	No	No
Unemp. Rate	No	No	Yes	Yes
Obs.	25553	8046	25553	8046
	C: Difference-in-Difference Estimates: E to U Transitions			
MH Parity Passed	0.003	0.003	0.009	0.011
	(0.013)	(0.013)	(0.013)	(0.014)
Mean MH	Yes	Yes	No	No
Unemp. Rate	No	No	Yes	Yes
Obs.	25553	13362	25553	13362

Notes: Robust standard errors clustered at the state year level shown in parentheses. All models include state and year fixed effects. Columns 1 and 3 are conditional on population of everyone diagnosed with a mental health disorder and employed. Columns 2 and 4 are conditional on the mental health and employed population as well as those with employer-sponsored insurance. In Panel B, this is further reduced to those with full-year ESI. Columns 1 and 2 also include the state-year proportion of individuals with a mental health condition as a control. Columns 3 and 4 include the state-year unemployment rates as a control.
*Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level

an effect in parity states, this was mostly seen in employment to unemployment transitions, which were largely insignificant in Table 3. The effects on job switching, however, remained relatively unchanged, suggesting the validity of the results in Table 3, and implying a direct effect of mental health parity legislation.

3.5 Discussion

The results on employment levels presented in this paper (Table 3) are broadly consistent with prior literature that has shown no significant effect of mental health parity on employment. This paper extends the literature by presenting evidence on the effect of parity in job flows instead of levels. This has not been explored in prior literature, likely due to the lack of available longitudinal data with employment, health insurance and expenditure variables, and with state identifiers.

The results above suggest that mental health parity mandates have caused a decline in job separations, particularly employment to employment transitions among the population of workers with mental health diagnoses. These effects are more pronounced among the insured population, suggesting that the decline is due to either a treatment effect of the mandates that result in a decline in involuntary transitions or that the presence of new mental health coverage prevents job transitions to other employers who would have offered the benefit. The effects are also strongest among individuals who worked for medium-sized firms, also suggesting an effect of the mandates, as state laws and ERISA legislation would have exempted many small and very large firms.

The positive effect on mental health expenditures has not been seen much before in the literature. Part of the reason for this could be that most literature on the effect of parity on utilization was conducted in the early 2000s, before many of the newest mandates were passed. However, the more likely reasons include the fact that most literature on the subject has not limited the population to those with a mental health diagnosis (and therefore focused primarily on overall expenditures–not just on mental health services). However, some recent papers have found this result before. In particular, Busch and Barry recently found that by conditioning on the population of individuals with a mental health condition, there is indeed a significant utilization effect of parity mandates (Busch & Barry, 2008).

There are several potential extensions for this paper. First, I am interested in the utilization effects of mental health parity. One way that I proxy for utilization is by using mental health expenditures, but one could also simply look at the number of visits rather than expenditures. In addition, I cannot currently identify whether individuals currently hold mental health coverage or not. Obtaining data that shows a breakdown of the benefits individuals previously held would provide for cleaner identification of the decline in job transitions and separations.

Another issue is the definition of mental health legislation, described in sections 2 and 3. There is considerable heterogeneity in the types of mandates passed, as described above. They could vary in comprehensiveness of the benefits, in populations affected, and in the "type" of mandate (i.e. parity, minimum mandated benefit, etc.). In this paper, I considered only the effect of the most comprehensive and costly mandate, i.e. mental health "parity." However, a fur-

ther step would be to distinguish between the types of mandates and services covered in particular states. An even more detailed analysis would also involve distinguishing between different types of mental health disorders among individuals in the MEPS, and matching those to the legislative requirements of the mandates. For example, if one state mandates coverage for schizophrenia, one could only flag individuals with a mental health disorder in that state if those individuals had an ICD-9 code for schizophrenia.

Another issue is the lack of clean identification of firms who self-insure (and hence are exempt from mandates due to ERISA laws). As mentioned above, I am so far controlling for these firms using large-firms (which mostly self-insure) as a proxy. However, this is an imperfect measure, and using employer data to assess whether firms self-insure would be another step for a future draft.

The policy endogeneity of the mandates themselves is another issue that could be further explored. Controlling for state unemployment rates and the proportion of individuals with a mental health diagnoses are good tests for this. However, in the event that there are still unobserved characteristics influencing the passage of mental health mandates, then it is unlikely that mental health parity in particular would be singled out as the exclusive mandate passed by the state, especially considering its perceived cost to employers. It is more likely then that these factors would engender the passage of groups of mandates in addition to mental health parity. Therefore, another extension of this paper would be to control for the number of mandates each state has passed in order to minimize this potential source of endogeneity. Another option would be to instrument for mental health parity passage with the political affiliation within a state (perhaps the fraction Democrat within a state, as Democrats are

traditionally pro-mandate).

Another potential extension would be to look at the effect of mental health mandates on other outcomes, such as overall health outcomes and wage effects. This would provide a useful analysis of not only whether mandates are inducing a utilization effect that results in increased job duration, but also whether these effects can be seen in worker productivity. Wages would also serve as a useful means of separation voluntary and involuntary transitions: if workers, pre-mandate, change jobs for lower wages, it would be further evidence that they are giving up preferred jobs for those that offer benefits.

A final extension for future work would be to use employer data (such as the MEPS Insurance Component) to check whether mandates are binding (by identifying which firms already offered mental health benefits prior to the passage of mandates) and by assessing whether employers restructure their benefits in response, rather than shifting wages. This employer-level analysis has never been performed with regards to any mandate in the United States before, and would likely be an interesting topic.

3.6 Conclusion

The analysis presented in this paper suggest that mental health parity reduces job separations, particularly employment to employment transitions. Typically, declines in job to job transitions are considered as evidence of increased "job lock," though that is not necessarily the case here. About a third of the decline seen in this paper is due to a fall in involuntary separations, suggesting a treatment effect of mandates. Individuals suffering from mental health

disorders receive comprehensive mental health coverage, which allows them to treat prior illnesses, and remain in their jobs longer. This is further supported by an increase in expenditures for mental health services following the passage of mandates, which do not occur in out-of-pocket expenses. There is also an observed decrease in voluntary transitions. However, this is likely due to individuals receiving more comprehensive benefits and no longer needing to seek jobs that offer mental health coverage. All of this suggests that mental health parity mandates are good policies for aiding the labor market outcomes of those with mental health conditions. Treating prior illnesses could decrease the costs of having a mental illness, increase worker productivity and aid in the overall welfare of this population.

REFERENCES

Abaluck, J., & Gruber, J. (2016). Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review*, *106*(8).

American Benefits Council. (2007). Group letter to energy and commerce committee on mental health parity expansion. Retrieved July 2013, from `http://www.americanbenefitscouncil.org`

andTom Y. Chang, M. J., Newhouse, J. P., & Earle, C. C. (2013). Physician agency and competition: Evidence from a major change to medicare chemotherapy reimbursement policy. *NBER Working Paper 19247*.

Baicker, K., & Chandra, A. (2005). The consequences of growth in health insurance premiums. *American Economic Review*, *95*(2), 214-218.

Baicker, K., & Chandra, A. (2006). The labor market effects of rising health insurance premiums. *Journal of Labor Economics*, *24*(3).

Baker, L. C., Bundorf, M. K., & Royalty, A. B. (2014). Physician competition and prices paid by private insurers for office visits. *Journal of the American Medical Association*, *312*(16).

Barry, C. L., & Busch, S. H. (2007). Do state parity laws reduce the financial burden on families of children with mental needs? *Health Services Research*, *42*(3), 1061-1084.

Barry, C. L., Frank, R. G., & McGuire, T. G. (2006). The costs of mental health parity: Still an impediment? *Health Affairs*, *25*(3), 623-634.

Berndt, E. R., Bir, A., Busch, S. H., & Frank, R. G. (2002). The medical treatment of depression, 1991-1996: Productive inefficiency, expected outcome variations, and price indices. *Journal of Health Economics*, *3*(21), 373-396.

Berndt, E. R., Cutler, D., Frank, R. G., Griliches, Z., Newhouse, J., & Triplett, J. E. (2000). *Handbook of health economics*. Elsevier Science.

Berry, S. T. (1994). Estimating discrete choice models of product differentiation. *Rand Journal of Economics*, *25*(2), 242-262.

Berry, S. T., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841-890.

Berry, S. T., & Pakes, A. (2007). The pure characteristics demand model. *International Economic Review*, *48*(4), 1993-1225.

Bresnahan, T. F., Stern, S., & Trajtenberg, M. (1997). Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics*, *28*(0), S17-S44.

Brot-Goldberg, Z. C., Chandra, A., Handel, B. R., & Kolstad, J. T. (2015). What does a deductible do? the impact of cost-sharing on health care prices, quantities, and spending dynamics. *NBER Working Paper Series*, *21632*.

Buchmueller, T., Carey, C., & Levy, H. G. (2013). Will employers drop health insurance coverage because of the affordable care act? *Health Affairs*, *32*(9).

Buchmueller, T. C., Cooper, P. F., Jacobson, M., & Zuvekas, S. H. (2007). Parity for whom? exemptions and the extent of state mental health parity legislation. *Health Affairs*, *26*(4), w483-w487.

Bundorf, K. (2002). Employee demand for health insurance and employer plan choices. *Journal of Health Economics*, *21*.

Bundorf, M. K., Henne, M., & Baker, L. (2007). Mandated health insurance benefits and the utilization and outcomes of infertility treatments. *NBER Working Paper Series*, *12820*.

Busch, S. H., & Barry, C. L. (2008). New evidence on the effects of state mental health mandates. *Inquiry*, *45*(3), 308-322.

Capps, C., & Dranove, D. (2004). Hospital consolidation and negotiated ppo prices. *Health Affairs*, *23*(2).

Claxton, G., & Levitt, L. (2015). *How many employers could be affected by the cadillac plan tax?* (Tech. Rep.). Kaiser Family Foundation.

Cockburn, I. M., & Anis, A. H. (2001). Hedonic analysis of arthritis drugs. In D. M. Cutler & E. R. Berndt (Eds.), *Medical care output and productivity* (p. 439-462). Chicago: University of Chicago Press.

Conti, R. M., Rosenthal, M. B., Polite, B. N., Bach, P. B., & Shih, Y.-C. T. (2012). Infused chemotherapy use in the elderly after patent expiration. *American Journal of Managed Care*, *18*(5).

Council for Affordable Health Insurance. (2012). *Health insurance mandates in the states 2012* (Tech. Rep.). Council for Affordable Health Insurance.

Court, A. (1939). The dynamics of automobile demand. In (p. 99-117). General Motors Corporation.

Crimmel, B. L. (2010). *Self-insured coverage in employer-sponsored health insurance for the private sector, 2000 and 2010* (Tech. Rep.). Agency for Healthcare Research and Quality.

Cseh, A. (2008). Labor market consequences of mental health parity mandates. *Forum for Health Economics and Policy*, *11*(2).

Cutler, D. (2004). *Your money or your life: Strong medicine for americans health care system*. Oxford University Press.

Cutler, D., McLellan, M., Newouse, J., & Remler, D. (1998). Are medical prices declining? evidence from heart attack treatments. *Quarterly Journal of Economics*, *63*(4), 991-1024.

Dafny, L. (2009). Estimation and identification of merger effects: An application to hospital mergers. *Journal of Law and Economics*, *52*(3).

Dafny, L. (2010). Are health insurance markets competitive? *American Economic Review*, *100*(4), 1399-1431.

Dafny, L., Duggan, M., & Ramanarayanan, S. (2012). Paying a premium on your premium? consolidation in the us health insurance industry. *American Economic Review*, *102*(2).

Dafny, L., Hendel, I., & Wilson, N. (2015). Narrow networks on the health insurance exchanges: What do they look like and how do they affect pricing? a case study of texas. *American Economic Review: Papers and Proceedings*, *105*(5).

Dafny, L., Ho, K., & Lee, R. (2016). The price effects of cross-market hospital mergers. *NBER Working Paper Series*, *22106*.

Dafny, L., Ho, K., & Varela, M. (2013). Let them have choice: Gains from shifting from employer-sponsored health insurance and toward an individual exchange. *American Economic Journal: Economic Policy*, *5*(1).

Dubay, L., Long, S. K., & Lawton, E. (2012). *Will health reform lead to job loss? evidence from massachusetts says no* (Timely Analysis of Immediate Health Policy Issues). Urban Institute.

Dunn, A. (2012). Drug innovations and welfare measures computed from market demand: The case for anti-cholesterol drugs. *American Economic Journal: Applied Economics*, *4*(3), 167-189.

Dunn, A. C., & Shapiro, A. (2014). Do physicians possess market power? *Journal of Law and Economics*, *54*.

Eggleston, K. N., Shah, N. D., Smith, S. A., & Berndt, E. R. (2011). Quality adjustment for health care spending on chronic disease: Evidence from diabetes treatment, 1999-2009. *American Economic Review: Papers and Proceedings*, *101*(3).

Eizenberg, A. (2014). Upstream innovation and product variety in the u.s. home pc market. *Review of Economic Studies*, *81*.

Ericson, K. M., & Starc, A. (2015a). Measuring consumer valuation of limited provider networks. *American Economic Review: Papers and Proceedings*, *105*(5).

Ericson, K. M., & Starc, A. (2015b). Pricing regulation and imperfect competition on the massachusetts health insurance exchange. *Review of Economics and Statistics*, *97*(3).

Ericson, K. M., & Starc, A. (2016). How product standardization affects choice: Evidence from the massachusetts health insurance exchange. *Journal of Health Economics*, *Forthcoming*.

Fan, Y., & Yang, C. (2016). Competition, product proliferation and welfare: A study of the u.s. smartphone market. *Working Paper*.

Garthwaite, C. L., Gross, T., & Notowidigdo, M. (2013). Public health insurance, labor supply, and employment lock. *NBER Working Paper Series*(19220).

Gaynor, M., & Town, R. (2012). Handbook of health economics, volume 2. In (chap. Provider Competition). Elsevier.

Gitterman, D. P., Sturm, R., Pacula, R. L., & Scheffler, R. M. (2001). Does the sunset of mental health parity really matter? *Administration and Policy in Mental Health*, *28*(5), 353-369.

Goodman, J. C., & Mitchell, W. (2002a). *The case against mental health parity part i: Faulty assumptions* (Brief Analysis No. 411). National Center for Policy Analysis.

Goodman, J. C., & Mitchell, W. (2002b). *The case against mental health parity part iii: Does the care really work?* (Brief Analysis No. 412). National Center for Policy Analysis.

Goodman, J. C., & Mitchell, W. (2002c). *The case against mental health parity part ii: Predictable harms* (Brief Analysis No. 411). National Center for Policy

Analysis.

Gowrisankaran, G., Nevo, A., & Town, R. (2015). Merges when prices are negotiated: Evidence from the hospital industry. *American Economic Review*, *105*(1).

Gravelle, J. G. (2015). *The excise tax on high-cost employer-sponsored health insurance: Estimated economic and market effects* (Tech. Rep.). Congressional Research Service.

Griliches, Z. (1961). The price statistics of the federal government. In (chap. Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change). National Bureau of Economic Research.

Gruber, J. (1992). State mandated benefits and employer provided health insurance. *NBER Working Paper Series*, *4239*.

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, *84*(3), 622-641.

Gruber, J., & Madrian, B. (1997). Employment separation and health insurance coverage. *Journal of Public Economics*, *66*(3), 349-382.

Gruber, J., & McKnight, R. (2016). Controlling health care costs through limited network insurance plans: Evidence from massachusetts state employees. *American Economic Journal: Economic Policy*, *8*(2).

Hall, M. A., & Fronstin, P. (2016). *Narrow provider networks for employer plans* (Tech. Rep.). Employee Benefit Research Institute.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, *103*(7).

Handel, B. R., & Kolstad, J. T. (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, *105*(8).

Hausman, N., & Lavetti, K. (2016). Physician concentration and negotiated prices: Evidence from state law changes. *Working Paper*.

Ho, K. (2006). The welfare effects of restricted hospital choice in the us medical care market. *Journal of Applied Econometrics*, *21*, 1039-1079.

Ho, K. (2009). Insurer-provider networks in the medical care market. *American Economic Review*, *99*(1), 393-430.

Ho, K., Hogan, J., & Morton, F. S. (2017). The impact of consumer inattention on insurer pricing in the medicare part d program. *RAND Journal of Economics, Forthcoming*.

Ho, K., & Lee, R. (2016). Insurer competition in health care markets. *Econometrica*, *Forthcoming*.

Howard, D. H., Bach, P. B., Berndt, E. R., & Conti, R. M. (2015). Pricing in the market for anticancer drugs. *Journal of Economic Perspectives*, *29*(1), 139-162.

Jacobson, M., Earle, C. C., Price, M., & Newhouse, J. P. (2010). How medicare's payment cuts for cancer chemotherapy drugs changed patterns of treatment. *Health Affairs*, *29*(7).

Jacobson, M., & Newhouse, J. P. (2006). Does reimbursement influence chemotherapy treatment for cancer patients? *Health Affairs*, *25*.

Jensen, G. A., & Gabel, J. R. (1992). State-mandated benefits and the small firm's decision to offer insurance. *Journal of Regulatory Economics*, *4*, 379-404.

Kaestner, R., & Simon, K. I. (2002). Labor market consequences of state health insurance regulation. *Industrial and Labor Relations Review*, *56*(1), 136-159.

Kleiner, S., White, W. D., & Lyons, S. (2015). Market power and provider consolidation in physician markets. *International Journal of Health Economics and Management*, *Forthcoming*.

Lakdawalla, D., Shafrin, J., Lucarelli, C., Nicholson, S., Khan, Z. M., & Philipson, T. J. (2015). Quality-adjusted cost of care: A meaningful way to measure growth in innovation versus the value of health gains. *Health Affairs*, *34*(4), 555-561.

Lang, M. (2013). The impact of mental health insurance laws on state suicide rates. *Health Economics*, *22*, 73-88.

Lee, R. S., & Fong, K. (2013). Markov-perfect network formation: An applied framework for bilateral oligopoly and bargaining in buyer-seller networks. *Working Paper*.

LoSasso, A. T., & Atwood, A. (2015). The effect of narrow preovider networks on health care use. *Working Paper*.

Lucarelli, C., Nicholson, S., & Town, R. J. (2010). The effect of physician reimbursement on chemotherapy treatment decisions and patient outcomes. *Unpublished Manuscript, Cornell University*.

Madrian, B. (1994). Employment-based health insurance and job mobility: Is there evidence of job-lock? *Quarterly Journal of Economics*, *109*(1), 27-54.

Matthews, M. (1999). Do we need mental health parity? Retrieved July 2013, from `www.ncpa.org`

McGuire, T. G., & Montgomery, J. (1982). Mandated mental health benefits in private health insurance. *Journal of Health Policy, Politics and Law*, *7*(2), 380-406.

McKinsey Center for U.S. Health System Reform. (2013). *Hospital networks: Configuration on the exchanges and their impact on premiums.*

Melnick, G., & Keeler, E. (2007). The effect of competition and multi-hospital systems on hospital prices. *Journal of Health Economics*, *26*(2).

Mohapatra, D. P., & Chatterjee, C. (2015). Price control and access to drugs: The

case of india's malarial market. *Working Paper*.

Moran, J. R., Chernew, M. E., & Hirth, R. A. (2001). Preference diversity and the breadth of employee health insurance options. *Health Services Research*, *36*(5).

Moriya, A., Vogt, W., & Gaynor, M. (2010). Hospital prices and market structure in the hospital and insurance industries. *Health Economics, Policy, and Law*, *5*(4).

Murphy, K. M., & Topel, R. H. (2006). The value of health and longevity. *Journal of Political Economy*, *114*(5), 871-904.

National Alliance on Mental Illness. (2009). State mental health parity laws. Retrieved July 2013, from `www.nami.org`

National Conference on State Legislatures. (2011). Mandated health insurance benefits and state laws. Retrieved November 2011, from `http://www.ncls.org`

Nevo, A. (2000). A practitioner's guide to estimation of random coefficients logit models of demand. *Journal of Economics & Management Strategy*, *9*(4), 513-548.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, *69*(2), 307-342.

Nevo, A. (2003). New products, quality changes and welfare meausres computed from estimated demand systems. *The Review of Economics and Statistics*, *85*(2), 266-275.

Newhouse, J. (1992). Medical care costs: How much welfare loss? *Journal of Economic Perspectives*, *6*(3), 3-21.

Nosko, C. (2014). Competition and quality choice in the cpu market. *Working Paper*, *Forthcoming*.

Pakes, A. (2005). Hedonics and the consumer price index. *Annales d'Economie et de Statistique*, *79-80*, 729-749.

Pakes, A., Porter, J., Ho, K., & Ishii, J. (2015). Moment inequalities and their applications. *Econometrica*, *83*(1).

Paramore, C. L., Thomas, S. K., Knopf, K. B., .s Cragin, L., & Fraeman, K. H. (2006). Estimating cost of care for patients with newly diagnosed metastatic colorectal cancer. *Clinical Colorectal Cancer*, *6*(1).

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, *110*(4), 705-729.

Pizer, S. D., Frakt, A. B., & Iezzoni, L. I. (2011). The effect of health reform on public and private insurance in the long run. *Health Care Financing & Economics Working Paper*, *2011*(3).

Polsky, D., & Weiner, J. (2015). *The skinny on narrow networks in health insurance marketplace plans* (Tech. Rep.). Leonard Davis Institute of Health Economics and Robert Wood Johnson Foundation.

Polyakova, M. (2016). Regulation of insurance with adverse selection and switching costs: Evidence from medicare part d. *American Economic Journal: Applied Economics*, *21*(3).

Powell, D. (2016). Optimal health insurance and the distortionary effects of the tax subsidy. *Working Paper*.

Prager, E. (2016). Tiered hospital networks, health care demand, and prices. *Working Paper*.

Raval, D., Rosenbaum, T., & Wilson, N. E. (2017). Using disaster induced closures to evaluate discrete choice models of hospital demand. *Working Paper*.

Scheffler, R. M., Arnold, D. R., Fulton, B. D., & Glied, S. A. (2016). Differing

impacts of market concentration on affordable care act marketplace premiums. *Health Affairs*, *35*(5).

Shepard, M. (2016). Hospital network competition and adverse selection. *Working Paper*.

Shinogle, J., & Salkever, D. (2005). Firms' demand for employment-based mental health benefits. *NBER Working Paper Series*, *No. 11436*.

Sloan, F. A., & Conover, C. J. (1998). Effects of state reforms on health insurance coverage of adults. *Inquiry*, *35*(280-293).

Small, K., & Rosen, H. (1981). Applied welfare economics with discrete choice models. *Econometrica*, *49*(1), 105-130.

Small group health premiums in massachusetts (Tech. Rep.). (2010). Commissioner of Insurance.

Smith, S., Newhouse, J., & Freeland, M. S. (2009). Income, insurance, and technology: Why does health spending outpace economic growth? *Health Affairs*, *28*(5), 1276-1284.

Song, M. (2007). Measuring consumer welfare in the cpu market: An application of the pure characteristics demand model. *Rand Journal of Economics*, *38*(2), 429-446.

Song, M. (2008). The quality-adjusted price index in the pure characteristics demand model. *Journal of Business and Economic Statistics*, *28*(1), 190-199.

Song, M., Lucarelli, C., & Nicholson, S. (2016). Merger with inter-firm bundling: A case of pharmaceutical cocktails. *RAND Journal of Economics, Forthcoming*.

Summers, L. H. (1989). Some simple economics of mandated benefits. *American Economic Review*, *79*(2), 177-183.

Thurston, N. K. (1997). Labor market effects of hawaii's mandatory employer-

provided health insurance. *Industrial and Labor Relations Review*, *51*(1), 117-135.

Town, R., Wholey, D., Feldman, R., & Burns, L. R. (2006). The welfare consequences of hospital mergers. *NBER Working Paper Series*, *12244*.

Town, R., Wholey, D., Feldman, R., & Burns, L. R. (2008). Is the impact of managed care on hospital prices decreasing? *Journal of Health Economics*, *27*(2).

Trajtenberg, M. (1989). The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of Political Economy*, *97*(2).

Trajtenberg, M. (1990). Product innovations, price indices, and the (mis)measurement of economic performance. *NBER Working Paper No. 3261*.

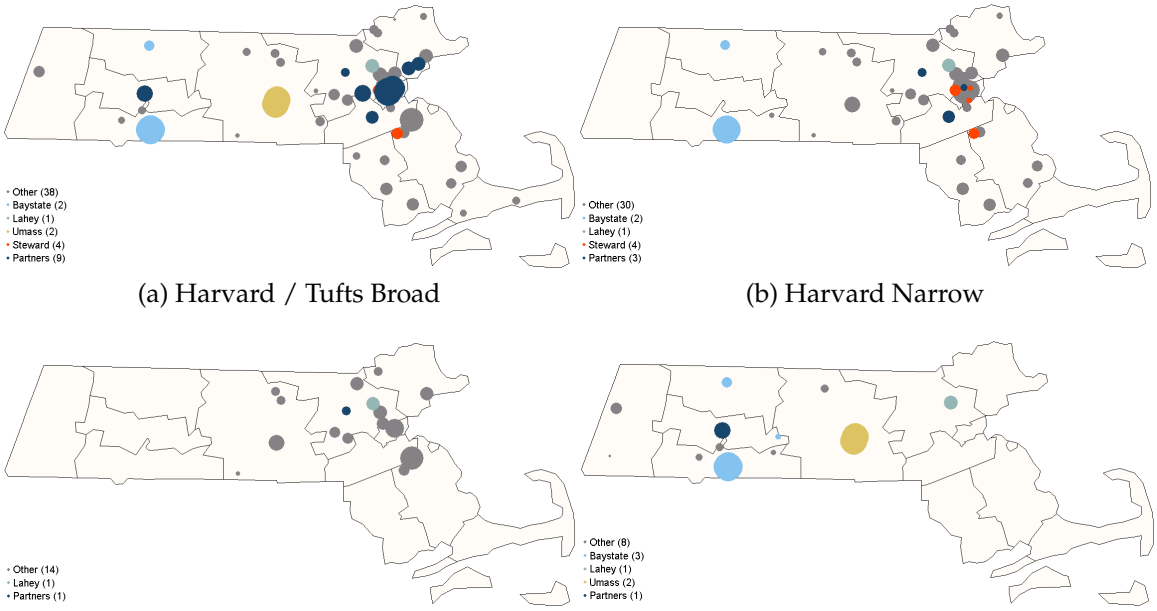
uhua Bao, & Strum, R. (2004). The effects of state mental health parity legislation on perceived quality of insurance coverage, perceived access to care, and use of mental health specialty care. *Health Services Research*, *38*(5), 1361-1378.

Utah Department of Human Services. (2012). *Treating illness, promoting health, supporting recovery* (Annual Report). Division of Substance Abuse and Mental Health.

APPENDIX A

APPENDIX FOR CHAPTER 1

A.1 Additional Figures



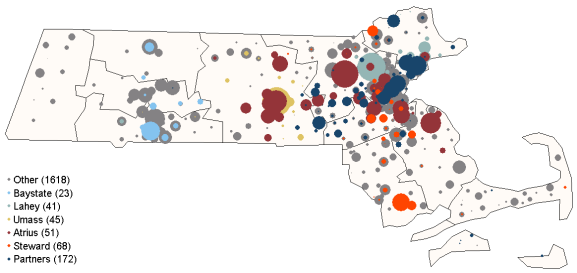
(a) Harvard / Tufts Broad

(b) Harvard Narrow

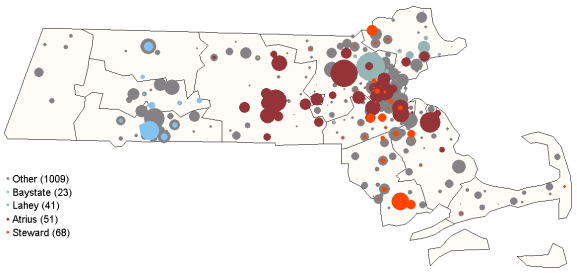
(c) Fallon Narrow

(d) Health New England Narrow

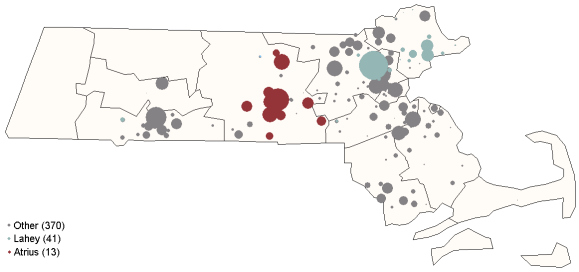
Figure A.1: Hospital Networks by Plan, 2011



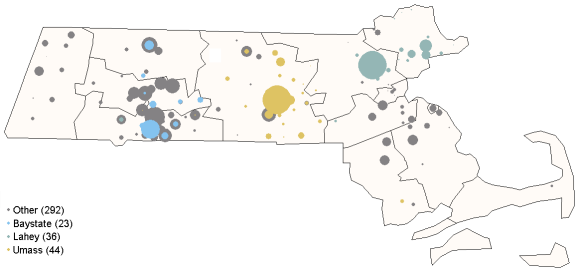
(a) Harvard / Tufts Broad



(b) Harvard Narrow

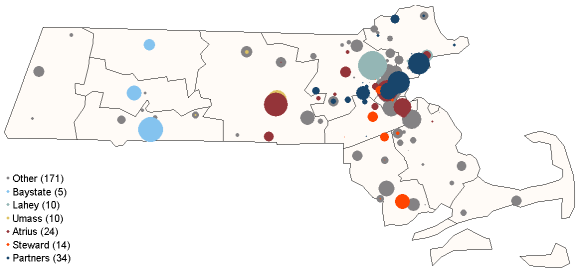


(c) Fallon Narrow

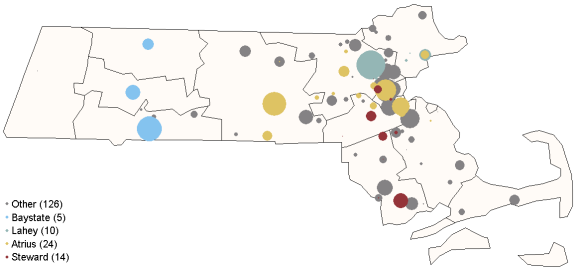


(d) Health New England Narrow

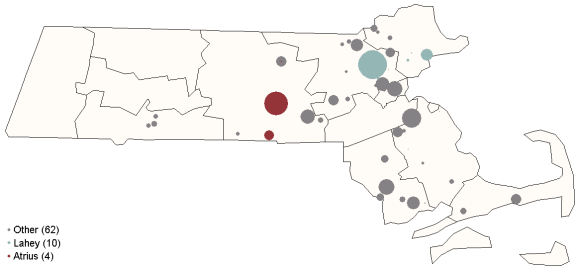
Figure A.2: Primary Care Practice Networks by Plan, 2011



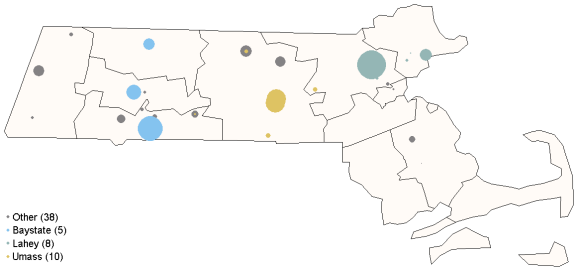
(a) Harvard / Tufts Broad



(b) Harvard Narrow

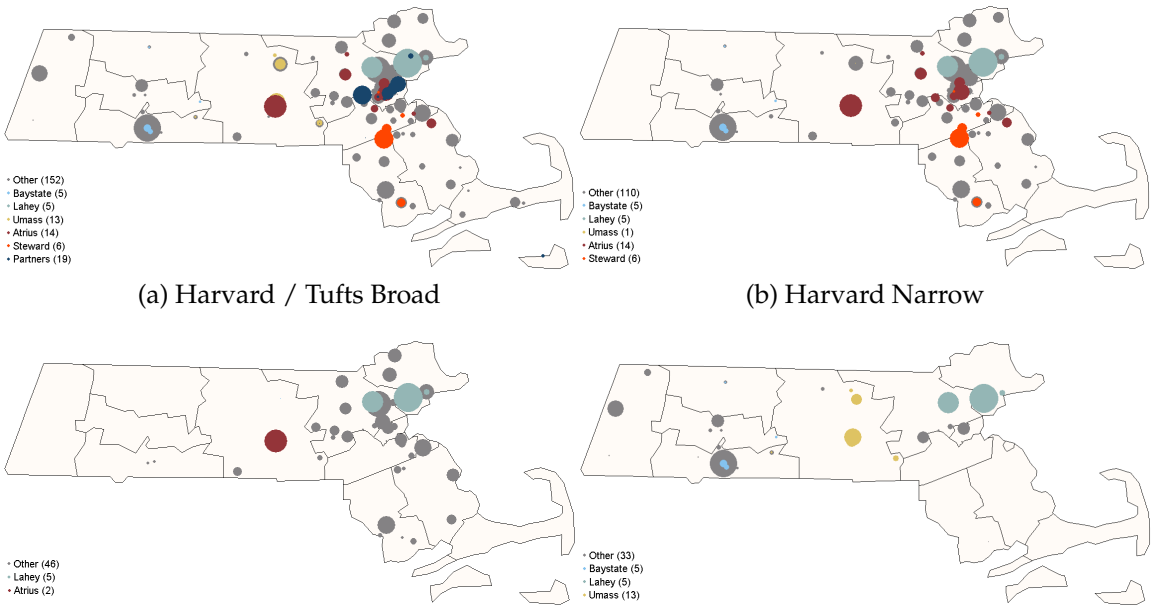


(c) Fallon Narrow



(d) Health New England Narrow

Figure A.3: Cardiology Networks by Plan, 2011



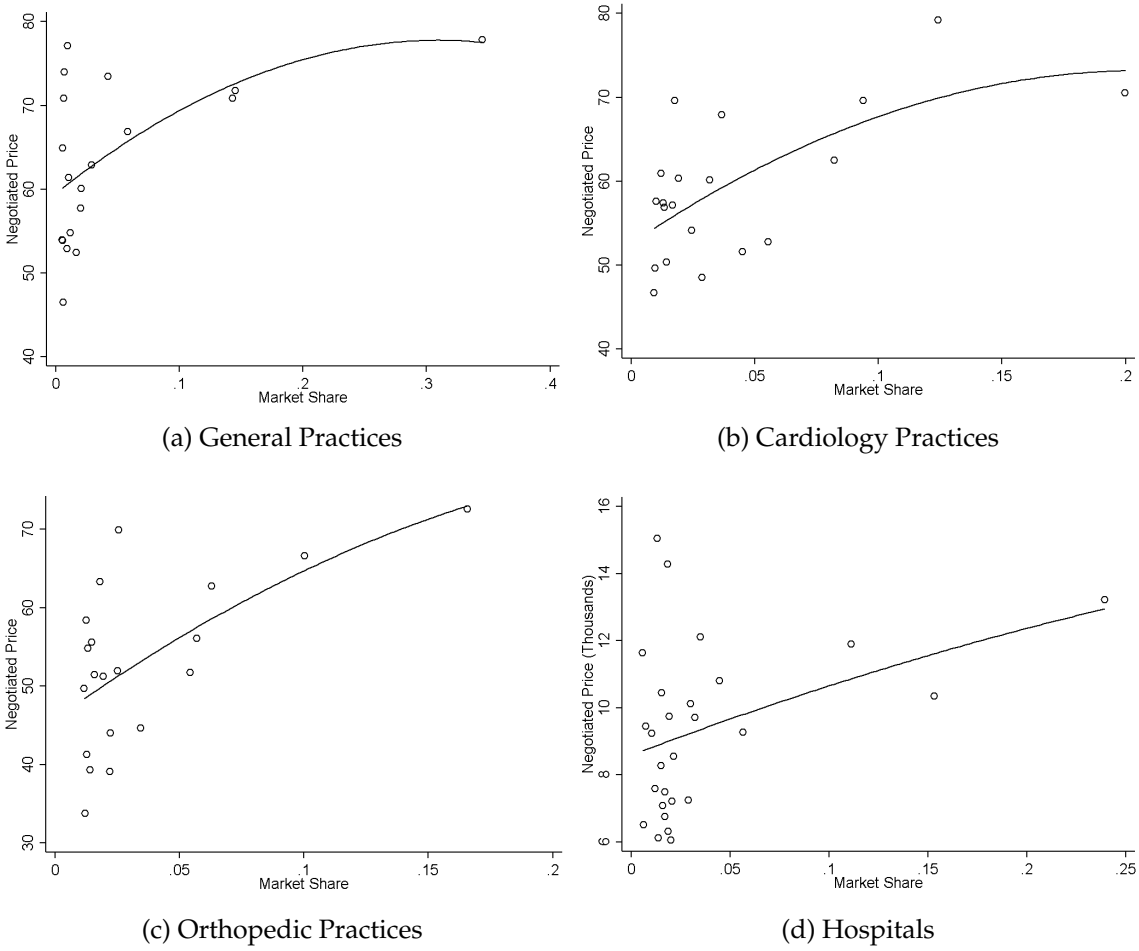
(a) Harvard / Tufts Broad

(b) Harvard Narrow

(c) Fallon Narrow

(d) Health New England Narrow

Figure A.4: Orthopedic Networks by Plan, 2011



(a) General Practices

(b) Cardiology Practices

(c) Orthopedic Practices

(d) Hospitals

Figure A.5: Insurer 2 Negotiated Price by Market Share, 2011

A.2 Additional Provider Demand Specifications

Table A.1 presents the results for alternate hospital demand models. Specifically, each model incorporates more flexible distance coefficients interacted with county identifiers in Massachusetts. This is done in order to allow patients to react differently to distance traveled to a particular hospital depending on where in Massachusetts they reside. Coefficients are for Barnstable county (the omitted variable), Worcester (Central Massachusetts), Hampden (Western Massachusetts), and Suffolk (Eastern Massachusetts). The distance coefficients are negative and significant in all reported counties. Notably, patients are far less reactive to distance in Barnstable, Hampden, and Worcester than they are in Suffolk. The parameter estimate ranges from -0.204 in Hampden County to -0.250 in Worcester. However, the coefficient surges to -0.350 in Suffolk county, indicating that patients are far less willing to travel in metropolitan Boston (which is part of Suffolk county) than they are in other regions of Massachusetts, where they are more likely to drive by car in order to find a hospital. The estimates imply that consumers in Barnstable are willing to travel an additional 13 miles on average in Barnstable in order to access a hospital they have used before. In Suffolk, however, they would only be willing to travel an additional 8 miles to access a previously used hospital.

Both models reported in Table A.1 also include a "copay" variable, which refers to the out-of-pocket amount that a patient has to pay when they are admitted as an inpatient to a particular hospital. In column (1), I exclude all observations where patients are either admitted through the hospital's emergency room or admissions resulting from a hospital transfer. This is done for two reasons. The first is that ER and transfer admissions may not necessarily reflect pa-

Table A.1: Results of Alternate Hospital Models

Variable	(1)	(2)
Distance	-0.2171***	-0.2379***
	(0.0122)	(0.0079)
DistancexWorcester	-0.0334***	-0.0287***
	(0.0054)	(0.0041)
DistancexHampden	0.0135***	0.0091**
	(0.0048)	(0.0037)
DistancexSuffolk	-0.1346***	-0.1612***
	(0.0146)	(0.0109)
Used Hospital	2.8474***	2.8324***
	(0.0438)	(0.0299)
Copay	-0.0001*	-0.0000
	(0.0000)	(0.0001)
DistxFemale	-0.0048***	-0.0021
	(0.0017)	(0.0013)
DistxAge	-0.0003***	-0.0004***
	(0.0001)	(0.0000)
DistxChronic	0.0234***	0.0247***
	(0.0026)	(0.018)
DistxSpecialty	0.0326***	0.0454***
	(0.0026)	(0.0023)
DistxAcademic	0.0186***	0.0259***
	(0.0023)	(0.0018)
CardiacxCathLab	0.6072***	0.2523***
	(0.1180)	(0.0603)
ObstetricsxNICU	3.9403***	3.6289***
	(0.2797)	(0.2200)
ImagingxMRI	0.0832	0.1268
	(0.1242)	(0.0790)
Hospital FE	Yes	Yes
ER & Transfers	No	Yes
Obs.	1,021,481	1,949,285
Pseudo R2	0.52	0.54

Results from hospital demand model from years 2009-2013. Omitted distance category is for the Barnstable county. "Copay" refers to the plan-specific copayment amount in dollars for a particular hospital visit. "Chronic" refers to having a chronic condition, "Specialty" refers to being a specialty hospital. Omitted from the table are distance terms interacted with each of 18 CCS diagnosis categories, a full set of hospital fixed effects, hospital fixed effects interacted with disease weights, as well as other patientxhospital interaction variables.

tient *choice* of a hospital. Faced with an emergency, a patient may be taken to the closest hospital rather than the hospital of his or her choice. The second reason is that the copayments are typically different for hospital admissions through the ER and transfers rather than voluntary admissions. Therefore, observations that pick up transfers might register a copay amount that is not reflective of the full amount. Indeed, column (1) shows that the coefficient on copay is negative and somewhat significant. The result is similar in magnitude to Prager (2016), though somewhat smaller. In column (2), where I include the full sample of admissions (including ER and transfers), the coefficient on copay reduces effectively to zero and becomes insignificant.

Table A.2 presents the results of alternate physician demand models for primary care practices, cardiology practices, and orthopedic practices. There are two main differences in the specifications here than the specification laid out in Table 7. First, much like with Table A.1, I allow the distance coefficient to vary with the county in Massachusetts. The pattern of the coefficients across the three specialties follows very closely with the pattern observed in the alternate hospital demand models. Indeed, consumers seem much less willing to travel in Suffolk County than in other counties in Central and Western Massachusetts. This difference, however, is far less pronounced among primary care physicians (where the distance coefficient is very similar regardless of whether the patient is in Worcester, Hampden, or Suffolk) than it is for cardiology practices and orthopedic practices.

The other difference in these specifications is that, rather than having the observations be at the patient-visit level, I assume that each patient makes only one choice of physician each year and can only visit that physician. In order to

Table A.2: Results of Physician Demand Models with Single Provider Choice

Variable	PCP Practices	Cardiology Practices	Orthopedic Practices
Distance	-0.1031***	-0.0928***	-0.1353***
	(0.0046)	(0.0029)	(0.0029)
DistancexWorcester	-0.0237***	-0.0239***	-0.0248***
	(0.0029)	(0.0021)	(0.0022)
DistancexHampden	-0.0436***	-0.0031	0.0044*
	(0.0044)	(0.0020)	(0.0026)
DistanceXSuffolk	-0.0519***	-0.0932***	-0.0634***
	(0.0049)	(0.0053)	(0.0046)
Used Doctors	4.9617***	3.7307***	4.0130***
	(0.0189)	(0.0163)	(0.0189)
DistxFemale	-0.0008	0.0005	0.0014
	(0.0012)	(0.0008)	(0.0010)
DistxAge	0.0002***	-0.0003***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
DistxRVU	0.0002	0.0004***	0.0003***
	(0.0002)	(0.0000)	(0.0000)
Practice FE	Yes	Yes	Yes
Obs.	4,603,357	4,595,697	4,311,426
Pseudo R2	0.74	0.56	0.53

Results of physician demand models are for years 2009-2013, where patients are assumed to make one selection of provider each year. Excluded from tables are estimates for distance interacted with number of doctors, distance interacted with the share of doctors who are of the specialty types being considered, and a full set of practice fixed effects as well as practice fixed effects interacted with RVU weights.

do so, for each patient i in each year t, I only include the physician d that the patient visits the most frequently throughout that year (within each specialty group). I then aggregate the RVU weights, r, of *all* the procedures that patient i seeks from physician d in year t. In essence, this is meant to reduce the impact of the logit shocks across physician visits. Since patients are highly likely to visit the same physicians that they have seen before (Table 3), it is not necessarily the case that each physician *visit* is a separate choice that patients are making. Rather, it may be more appropriate to think of patients as choosing a *physician* every year given their anticipated health requirements (RVUs) for the year.

The results in A.2 shed light on the effect these specifications have on the parameter estimates. The most notable difference is that the coefficients on "Used Doctors" is reduced across all three specialties. This indeed implies that the logit shocks across visits played a role in blowing up the physician inertia coefficients. In Table 7, the estimates implied that, on average, people would be willing to travel approximately 48 miles to see the same primary care physician they have seen in the past (and approximately 37 miles for a cardiologist and 22 miles to see an orthopedist). Here, patients in Worcester would on average be willing to travel 39 miles to see the same primary care physician, but only 32 miles in Suffolk. Similarly, cardiology patients would be willing to travel 31 extra miles in Worcester, and only 20 extra miles in Suffolk, whereas orthopedic patients would be willing to travel 25 extra miles in Worcester and 20 extra miles in Suffolk.

While these inertia coefficients seem somewhat more dampened than in Table 7, some of the other coefficients have changed in these alternate specifications as well. For instance, distance interacted with the indicator for female loses its significance in this model, while the distance interacted with RVU coefficient loses its significance in the PCP model. Moreover, the distance interacted with age coefficient flips signs in the PCP model.

Table A.3 presents the results of the same models, however run only on the top 50 physician practices of each specialty instead of the top 100. The results are largely similar to those in the model with the top 100 practices. The distance coefficients and physician inertia coefficients are somewhat dampened across all three specialties, but broadly consistent. For primary care physicians, an average patient in Worcester would be willing to travel an additional 42 miles to

see the same provider they have seen in prior years, whereas patients in Suffolk would be willing to travel an additional 35 miles. Indeed, these are similar to the implied estimates when including the top 100 practices. However, in this model, the distance interacted with female coefficient returns to being negative and significant, as in Table 7. Moreover, the distance interacted with RVU coefficient also regains its positive and significant value.

Table A.3: Results of Physician Demand Models with 50 Practices

Variable	PCP Practices	Cardiology Practices	Orthopedic Practices
Distance	-0.0776***	-0.0870***	-0.1153***
	(0.0056)	(0.0028)	(0.0029)
DistancexWorcester	-0.0332***	-0.0231***	-0.0274***
	(0.0035)	(0.0019)	(0.0021)
DistancexHampden	-0.0450***	-0.0033*	0.0051*
	(0.0049)	(0.0018)	(0.0027)
DistanceXSuffolk	-0.0572***	-0.0823***	-0.0903***
	(0.0054)	(0.0051)	(0.0043)
Used Doctors	4.7412***	3.5708***	3.8179***
	(0.0215)	(0.0169)	(0.0200)
DistxFemale	-0.0028**	0.0009	0.0012
	(0.0014)	(0.0008)	(0.0009)
DistxAge	0.0003***	-0.0004***	-0.0000
	(0.0000)	(0.0000)	(0.0000)
DistxRVU	0.0003*	0.0005***	0.0002***
	(0.0002)	(0.0001)	(0.0000)
Practice FE	Yes	Yes	Yes
Obs.	2,344,464	2,367,615	2,285,786
Pseudo R2	0.78	0.55	0.54

Results of physician demand models are for years 2009-2013, where patients are assumed to make one selection of provider each year. Model includes top 50 practices, as opposed to top 100. Excluded from tables are estimates for distance interacted with number of doctors, distance interacted with the share of doctors who are of the specialty types being considered, and a full set of practice fixed effects as well as practice fixed effects interacted with RVU weights.

In all, these checks suggest that the model reported in Tables 6 and 7 are fairly robust to alternate specifications, and that the results of the insurance demand models and supply side do not change by a significant amount due to differing provider specifications.

A.3 Selection on Unobservables in the Provider Demand Models

A concern with multinomial logit demand models of the type presented in Section 4.1 is that they may suffer from a problem with selection on unobservables. Due to the fact that the models condition on the hospital and physician networks of each patient i at time t, N_{ijtH} and N_{ijtS}, the expected utility of a particular provider network, EU_{IjtH} and EU_{ijtS}, is calculated assuming that there is no selection in the plan choice stage. This assumption may be violated, however, if individuals select into narrow network plans differentially from broad network plans for reasons unobserved by the econometrician (such as an aversion to high-cost providers, including Partners hospitals and Atrius physicians). If such selection were a major concern, this would bias EU_{Ijt}, and therefore subsequently bias the parameter estimates from the plan demand stage. Indeed, there is literature that such discrete choice models are prone to incorrect predictions when hospitals are exogenously removed from a patient's choice set (Raval, Rosenbaum, & Wilson, 2017).

I present here some reduced form evidence suggesting that such selection is not a major concern in my setting. Figure A.6 displays the share of individual choices of hospitals and doctors for individuals only in *narrow network* plans that are accurately predicted by a model of provider demand run only on individuals in *broad network* plans. The logic is that if unobserved selection into narrow network plans were a big concern, we would expect a model of choice only run on patients in broad network plans to significantly misrepresent the choices of patients with reduced choice sets. According to the figure,

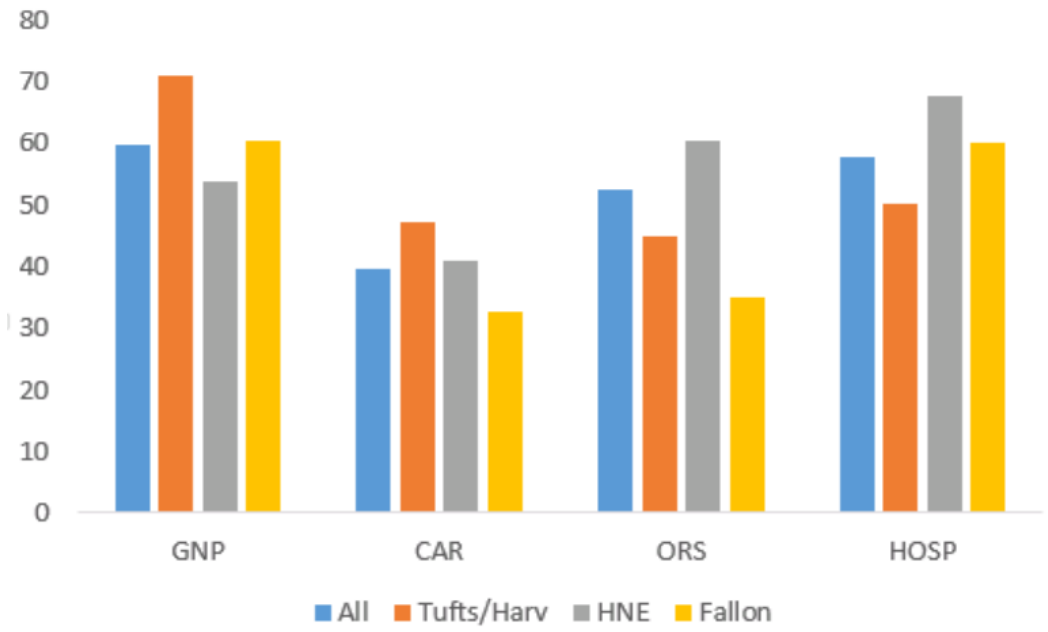


Figure A.6: Share of Actual Choices Accurately Predicted, by Specialty

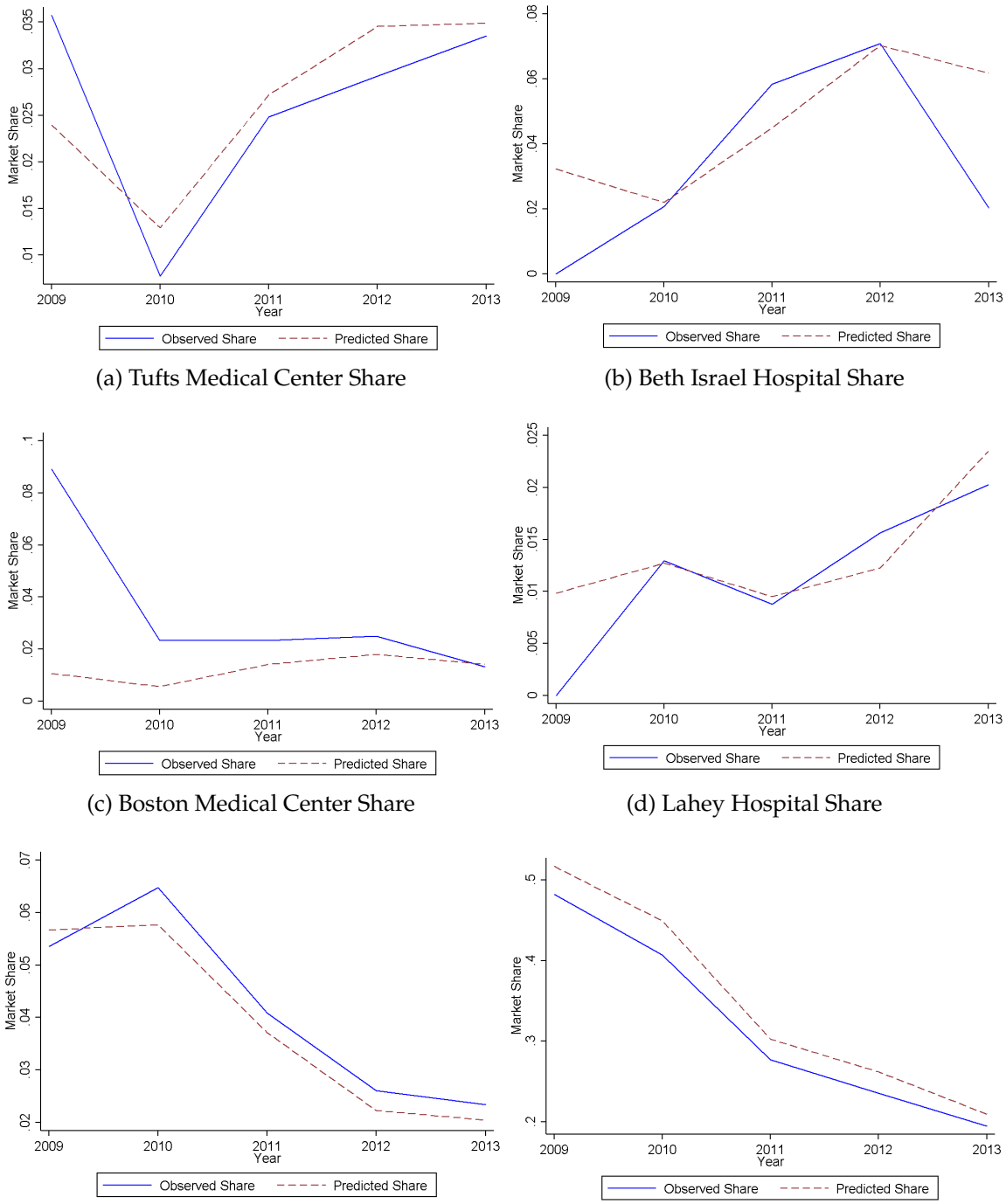
however, the logit model predicts the choices of narrow network patients quite well. For general practitioners (primary care physicians), the model accurately predicts about 60% of individual choices, and over 70% of the choices in the Tufts and Harvard narrow networks, in particular. The model also predicts hospital choices quite well, with a particularly good fit for patients in Health New England. The model does slightly worse for orthopedic surgeons, predicting about 55% of choices overall, and does worse still for cardiologists, with about 40% of choices predicted.

In addition, Figure A.7 plots the actual market share of selected medical centers versus the predicted market share among only narrow network patients. For the most part, the model predicts these market shares very well. For the hospitals in the metropolitan Boston area (Tufts, Beth Israel, and Boston Medical Center), the model seems to have some trouble predicted accurate market

shares in 2009, but then converges for every year after 2010. This is likely due to small sample sizes of hospital admissions among narrow network patients, which is particularly true in 2009 (prior to the introduction of the Tufts and Harvard narrow plans). Despite this, the model seems to predict the market share patterns across time very well, although it predicts a less steep decline in 2013 for Beth Israel (panel b) than the observed share. Finally, the model does extremely well in predicting the market shares of the Berkshire and Baystate medical centers, both of which are located in Eastern Massachusetts.

Taken together, these figures imply that selection is likely not a major concern in my model. Indeed, the predicted market shares for hospitals in the Boston area (which contains the highest number of academic medical centers and high-cost physicians excluded in narrow network plans) for the most part track nicely with the observed shares, despite some difficulty in 2009. The hospitals in Eastern Massachusetts are predicted with much better accuracy. Future work will explore the sensitivity of the provider demand results to alternate specifications, including more flexible semiparametric and nonparametric models.[1]

[1] An additional option I am exploring is to model a random coefficient with an assumed distribution directly in the choice model, which may capture additional unobserved heterogeneity. This would be in the spirit of Polyakova (2016) and Ho et al. (2017) (Ho, Hogan, & Morton, 2017).



(a) Tufts Medical Center Share

(b) Beth Israel Hospital Share

(c) Boston Medical Center Share

(d) Lahey Hospital Share

(e) Berkshire Medical Center Share

(f) Baystate Medical Center Share

Figure A.7: Observed versus Predicted Hospital Shares for Narrow Network Patients

A.4 Additional Plan Demand Estimates

Table A.4 shows the results of additional plan demand specifications that demonstrate the importance of including physician networks in the model. The first three columns all present estimates without the inertia coefficient, while columns 4-6 present estimates that include this coefficient. Within each set, the first column (i.e. columns 1 and 4) present results only focusing on hospital utility (EU_{IjH}), the second column (i.e. columns 2 and 5) present the results including both hospital utility and the utility for the combined physician network (EU_{IjS}), and the third column (i.e. columns 3 and 6) present results separating the physician networks into the three distinct specialties (the most granular specification).

Looking at columns 1-3, it is clear that moving from a model with only hospital networks to a model that includes physician networks has a significant effect on the estimated parameters. Between columns 1 and 2, the premium sensitivity parameter nearly triples in magnitude, from -0.198 to -0.582. Between columns 2 and 3, this parameter remains basically unchanged. Moreover, the household expected utility for the plan's hospital network decreases significantly, from 9.544 to 1.892 between columns 1 and 2, and then to 1.394 in column 3. Turning to columns 4-6, the magnitudes of the premium coefficient again increases considerably, moving from -0.198 in column 1 to -1.355 in column 4 as the inertia coefficient is included. Between columns 4 and 5, the premium coefficient increases in magnitude once again as the physician networks are included, and the inertia coefficient drops slightly from 4.620 to 4.576. These results demonstrate not only the importance of including inertia to accurately estimate consumer price sensitivity, but also that responsiveness to premiums is

Table A.4: Results of the Alternate Plan Demand Models

Variable	(1)	(2)	(3)	(4)	(5)	(6)
		No Inertia			Inertia	
Prem ($00s)	-0.198***	-0.582***	-0.573***	-1.355***	-1.713***	-1.763***
	(0.065)	(0.069)	(0.069)	(0.123)	(0.126)	(0.126)
EU_{IjH}	9.544***	1.892***	1.364**	7.076***	2.522**	1.266
	(0.816)	(0.621)	(0.590)	(1.006)	(1.010)	(0.967)
EU_{IjS}		0.444***			0.293***	
		(0.023)			(0.027)	
EU_{IjPCP}			0.177***			0.136**
			(0.028)			(0.031)
EU_{IjCAR}			0.243***			0.381***
			(0.086)			(0.131)
EU_{IjORS}			0.949***			0.653***
			(0.060)			(0.085)
Inertia Coefficient				4.620***	4.576***	4.572***
				(0.072)	(0.073)	(0.073)
Insurer FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	46,719	46,719	46,719	46,719	46,719	46,719
Pseudo R^2	0.29	0.32	0.32	0.78	0.79	0.79
WTP Hosp	$182	$12	$9	$20	$6	$3
WTP Phys		$91			$20	
WTP PCP			$19			$5
WTP CAR			$12			$6
WTP ORS			$51			$11
FE	$711	$240	$242	$80	$64	$61
WTP Tot	$888	$341	$331	$99	$89	$86
Inertia				$341	$267	$260

Notes: Columns 1-3 are results for models without the inertia coefficient. Columns 4-6 include these coefficients. EU_{IjH} refers to the household's expected utility for the hospital network, EU_{IjPCP} refers to the utility of the primary care network, EU_{IjCAR} refers to the utility of the cardiology network, and EU_{IjORS} refers to the utility of the orthopedic network. "WTP" refers to "willingness-to-pay" for Harvard Pilgrim's broad hospital and physician networks over relative to its narrow network (that is, the willingness to pay for the additional expected utility). "FE" refers to the dollarized willingness-to-pay for Harvard's broad plan over its narrow plan not captued by the expected utility variable. "WTP Tot" refers to the sum of the willingness to pay for the extra utility and the fixed effect. The premium variable is in hundreds of dollars.

also significantly affected by including physician networks in the model. In particular, this suggests that by ignoring physician networks, previous models may have been significantly *underestimating* the extent to which consumers respond to prices in the market. Taken together with the significant, positive estimates on physician network utility, the results imply that consumers may select into broad network plans for two primary reasons. First, they place an extraordinary value on being in the same plan they were in previously. Second, consumers are not necessarily insensitive to price when choosing their plans, as is widely believed about consumers in employer-sponsored-plans, but rather are extremely inertial to their *physicians*, and are therefore willing to pay significant amounts to retain access to physicians they have used in the past.

Translating these parameter estimates into "willingness-to-pay" (WTP) estimates helps shed light on this phenomenon as well. The bottom panel of Table A.4 reports estimates for implied willingness-to-pay for Harvard Pilgrim's broad network relative to its narrow network, as well as dollarized values of the inertia coefficient (switching costs) and the portion of WTP for Harvard's broad plan not captured by the network utility (this latter amount is given by the difference in Harvard's broad plan fixed effect over Harvard's narrow plan, "FE"). In column 1, the estimates imply that individuals, on average, would be willing to pay about $182 per month to move from Harvard's narrow network to its broad network for hospitals alone. Not only is this amount extremely high relative to the actual premium differences between Harvard's broad and narrow plans (about $30 per month on average for individuals), but the difference in fixed effect estimates amounts to about $711. Therefore, in a model without physician networks, consumers are on average willing to pay about $888 to move from Harvard's narrow to broad plan.

Moving to column 2, consumers are now on average only willing to pay about $12 per month to move from Harvard's narrow hospital network to its broad hospital network. They are, however, willing to pay about $91 per month to move from its narrow *physician* network to its broad physician network. Moreover, the fixed effect estimates drops significantly from $711 to $240, rendering a total WTP of $341. It is clear that this is much improved from the estimates implied by column 1, as both the total willingness-to-pay and the amount unexplained by networks is considerably reduced.

In column 4 (the model with the inertia coefficient included, but excluding physician networks), consumers are willing to pay an average of $20 per month to move from Harvard's narrow network to its broad network. They are also willing to pay about $80 to move plans for reasons unexplained by the hospital networks, yielding a total of about $100 per month. In column 5 (which includes physician networks), the WTP for hospital networks decreases to a mere $6 per month, with the WTP for physician networks equaling about $64 per month, for a total of about $90 per month. It is clear that these estimates are much more reasonable than in columns 1-3, and much more closely approximate the difference in actual premiums between the two plans. Moreover, the switching cost estimate decreases from $341 in column 4 (when physician are not included) to $267 in column 5 (when physicians are included). This suggests that not only does including physicians capture some of what was previously attributed to the fixed effect, but also by increasing the magnitude of the premium sensitivity parameter and decreasing the magnitude of the inertia coefficient, the model with physicians clearly captures some significant heterogeneity driving patient choice of plans that was previously interpreted as a high switching cost.

APPENDIX B

APPENDIX FOR CHAPTER 2

B.1 Estimating the Pure Characteristics Model

The pure characteristics model described in section 3.4 is notoriously difficult to estimate. The challenge stems from the absence of the idiosyncratic error term, ϵ_{ijt}, present in both the logit and random-coefficients models. In the previous models, the error term ensured that the market share function was a smooth function of the regimen characteristics, which allowed for easy integration. Without the error term, there is no longer a guaranteed contraction mapping that can be used to equate the predicted and observed market shares, and hence back out the mean utility, δ_j. We, therefore, resort to alternate methods of estimating the mean product quality as described by Berry and Pakes (Berry & Pakes, 2007). Our methods also follow closely that of Minjae Song (Song, 2007, 2008).

As in Berry & Pakes 2007 and Song 2007, the estimation procedure proceeds in three steps. The first step is to bring the mean product quality closer to the true product quality by use of a contraction mapping, as in BLP. Recall the utility function in BLP is:

$$u_{ijt} = -\alpha p_{jt} + \beta x_{jt} - \alpha^u v_i p_{jt} + \beta^u v_i x_{jt} + \xi_j + \Delta\xi_j + \epsilon_{ijt} \tag{B.1}$$

In principle, then, the utility function from the pure characteristics model is a limiting case of the utility function from the random coefficients model as ϵ_{ijt} approaches zero. We thus begin by implementing the BLP contracting mapping

with a scaling factor applied to the error term that gradually proceeds to zero. The market share in this model collapses to:

$$S_{jt}(\delta_{jt}, \alpha^u, \beta^u) = \frac{1}{ns} \sum_{i=1}^{ns} \frac{exp[(\delta_{jt} - \alpha^u v_i p_{jt} + \beta^u v_i x_{jt})\mu]}{1 + \sum_{k=1}^{J} exp[(\delta_{kt} - \alpha^u v_i p_{kt} + \beta^u v_i x_{kt})\mu]} \tag{B.2}$$

where μ is the scaling factor and gradually grows larger in the estimation routine. In practice, as μ grows larger, the exponential function rapidly blows up, resulting in incalculable or missing mean values. We iterate on μ and proceed slowly until the point that we can no longer compute a δ_j. The closes δ_j we obtain using this scaled share function brings us closer to the true mean value from the pure characteristics model, and we hence use this value as the starting point for the next part of the procedure.

The second step is an element-by-element fixed-point homotopy method. This goal of this inversion is to find a mean utility value, δ_j, to satisfy the following equation:

$$|S_{jt}(\delta_j, \delta_{-j}, \theta) - s_{jt}| < tol \tag{B.3}$$

The difficult is that this inversion is not guaranteed to be a weak contraction mapping. Therefore, both Berry & Pakes, and Song, combine this element-by-element inversion with a homotopy method using the following:

$$\delta'_j(t) = (1-t)\delta_{0j} + t\delta_j, j = 1, ..., J \tag{B.4}$$

where δ_{0j} is an initial guess for the mean value, δ_j is the current iteration

of the mean value, and t is between 0 and 1. As Song points out, when $t = 1$, this collapses back to the strict element-by-element inverse, which is not guaranteed to contract. However, when $t < 1$, there is a strict contraction mapping guaranteed, albeit the fixed point the model converges to may not necessarily be the true value of the pure characteristics model, δ_j. Therefore, much like the first step of the procedure in which the BLP contraction is implemented with a scaling factor to bring the value of δ_j closer to the true value, the homotopy method is implemented with a value of t that is strictly less than 1, but repeated while approaching one very slowly. Following Berry & Pakes, and Song, we begin with $t = 0.99$ and increase by 0.0025, with the element-by-element mapping repeating at least 50 times before altering the value of t.

The final step is to use the Newton-Rhapson search method. This occurs when the element-by-element inversion in equation (B.3) is not satisfied after an iteration of the homotopy search, yet all the predicted market shares are non-zero. This implies the true value of δ_j has not been found yet, but since the current predicted market shares are non-zero, an alternate, more rapid search method for smooth functions can be implemented. While both Berry & Pakes, and Song had to rely on the Newton method in their simulations of the pure characteristics model, our implementation rarely relied on this method, with the homotopy method approaching the true δ_j and satisfying equation (B.3) most of the time.

B.2 Additional Tables

Table B.1: Composition and Dosages of Chemotherapy Drugs

Regimen	1^{st} Drug	2^{nd} Drug	3^{rd} Drug	4^{th} Drug
5-FU + LV	425 mg of 5-Fu/m^2/day for days 1-5, every 4 weeks	20 mg of LV/m^2/day for days 1-5, every 4 weeks		
Irinotecan (Pfizer)	125 mg of irinotecan per week/m^2 for 4 weeks, every 6 weeks			
Irinotecan + 5-FU/LV	180 mg of irinotecan/m^2 on day 1, every 2 weeks	1,000 mg of 5-FU/m^2 on day 1 and 2, every 2 weeks	200 mg of LV/m^2 on day 1 and 2, every 2 weeks	
Capecitabine (Roche)	2,500 mg of capecitabine per m2/day for days 1-14, every 3 weeks			
Capecitabine + Irinotecan	70 mg of irinotecan/m^2/week, every 6 weeks	2,000 mg of capecitabine per m^2/day for days 1-14, every 3 weeks		
Oxaliplatin (Sanofi) + 5-FU/LV	85 mg of oxaliplatin per m^2 on day 1, every 2 weeks	1,000 mg of 5-FU/m^2 on day 1 and day 2, every 2 weeks	200 mg of LV/m^2 on day 1 and day 2, every 2 weeks	
Oxaliplatin + Capecitabine	130 mg of oxaliplatin per m^2 on day 1, every 3 weeks	1,700 mg of capecitabine per m^2/day for days 1-14, every 3 weeks		
Cetuximab (ImClone)	400 mg of cetuximab per m^2 on day 1; then 250 mg/m^2 once a week, every 6 weeks			
Cetuximab + Irinotecan	400 mg of cetuximab per m^2 on day 1; then 250 mg/m^2 once a week, every 6 weeks	125 mg of irinotecan per week/m^2 for 4 weeks, every 6 weeks		
Bevacizumab (Genentech) + Oxaliplatin + 5-FU/LV	5 mg of bevacizumab per kg, every 2 weeks	85 mg of oxaliplatin per m2 on day 1, every 2 weeks	1,000 mg of 5-FU/m^2 on day 1 and day 2, every 2 weeks	200 mg of LV/m^2 on day 1 and day 2, every 2 weeks
Bevacizumab + Irinotecan + 5-FU/LV	5 mg of bevacizumab per kg, every 2 weeks	180 mg of irinotecan per m^2 on day 1, every 2 weeks	1,000 mg of 5-FU/m^2 on day 1 and day 2, every 2 weeks	200 mg of LV/m^2 on day 1 and day 2, every 2 weeks
Bevacizumab + Oxaliplatin + Capecitabine	7.5 mg of bevacizumab per kg, every 3 weeks	130 mg of oxaliplatin per m^2 on day 1, every 3 weeks	1,700 mg of capecitabine/m^2/day for days 1-14, every 3 weeks	

mg=miligram of active ingredient; m^2=meter squared of a patient's surface area; kg=kilogram of a patient's weight.
Source: National Comprehensive Cancer Network, Colon Cancer, Version 2.2006; package inserts.