Narrow Physician Networks, Switching Costs, and Product Variety in Employer Markets*

Nicholas Tilipman†

December 16, 2018

Abstract

I study optimal product variety for health insurance plans in a market primarily differentiated on provider networks. I endogenize employer health plan offerings with respect to hospital and physician networks to investigate whether observed plan menus reflect consumer preferences or switching frictions. I find that half of willingness-to-pay for provider networks stems from loyalty to a small number of previously-used physicians. However, high health plan switching costs leads consumers who would benefit from narrow-networks to disproportionately select broad-networks, resulting in suboptimal plan menus. Removing these switching costs increases product variety, decreases healthcare spending, and ultimately, increases social surplus.

Keywords: health insurance, narrow networks, inertia, switching costs
JEL Classification Codes: I11, I13, D83, G22

*I am deeply grateful to my dissertation committee, Sean Nicholson, Panle Jia Barwick, and Samuel Kleiner for their valuable advice and encouragement. I also thank Matthew Backus, Colleen Carey, Jean-Francois Houde, Pierre Leger, Anthony LoSasso, Claudio Lucarelli, Ross Milton, Yuta Toyama, Wesley Yin, and numerous seminar participants for their comments and insight. This project was also made possible by a Cornell University Human Ecology Alumni Association grant and an Institute for Social Sciences grant. All mistakes are my own.

†Division of Health Policy and Administration, School of Public Health, University of Illinois at Chicago. Contact: tilipman@uic.edu.

1 Introduction

Increasing health costs and the introduction of Health Insurance Exchanges (HIE) has ushered the introduction of new insurance plan designs, such as narrow networks, high-deductible health plans, and tiered networks. However, our understanding of optimal product variety in health insurance markets, as well as the drivers behind product offering and enrollment choices, has lagged behind these innovations. This is important both as states continue to manage the options available to consumers in the HIE and as employers increasingly move towards offering employees more choice of plans. Offering consumers more choice among differentiated products has the potential to both bring down costs and allow consumers to more efficiently select into plans that best suit their health care needs (Dafny et al., 2013). Conversely, offering increased choice may be costly for consumers and firms. Consumers may be forced to pay substantial search costs in order to evaluate the options or may exhibit switching costs that lead them to select into dominated, suboptimal plans (Handel, 2013; Liu and Sydnor, 2018; Abaluck and Gruber, 2016). Firms may also bear shopping costs or fixed costs of designing products and offering multiple plans to employees (Bundorf, 2002).

In this paper, I study the drivers of employer health plan menus and optimal product variety in a setting where plans are primarily differentiated on one dimension: provider networks. Indeed, 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 containing spending and offering consumers low-cost options.[1] Despite the increasing popularity of narrow-network plans on the HIE, however, employers have been slower to adopt, design, and offer such products. In 2016, only 7% of employers nationally offered a narrow network as part of their plan menu (Hall and Fronstin, 2016). Rather, most employers typically only offer one or two plans to their employees, which tend to have relatively large networks (Buchmueller et al., 2013; Dafny et al., 2013).

I tease out several potential explanations for the prevalence of broad networks and limited product choice along this dimension. The first explanation is that employers may be reacting to the preferences of their employees, who may either place high value on access to a broad array of hospitals and doctors or may otherwise not be very price sensitive.[2] Second, choice frictions may prevent employees and firms from switching to narrow-network products, even when consumers may benefit from the option to do so. In particular, I focus on frictions from three potential sources: consumer inertia to previously used physicians; consumer inertia to prior health plans; and firm fixed costs of product redesign.

To disentangle whether plan offering truly reflects consumer preferences or are suboptimal and driven by market frictions, I estimate a model of supply and demand for health insurance plans for

[1] 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. Approximately 70% of the plans available on the Affordable Care Act (ACA) 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 and Weiner, 2015).

[2] This may be in part due to the fact that employers subsidize a large portion of employee premiums–generally through a percentage of total costs–and that contributions to premiums are tax-deductible (Powell, 2016). These subsidies dampen the relative difference in premiums between narrower-network and broader-network products for consumers in employer markets, thereby making it cheaper to access a broader network.

a large-group purchaser (employer) in Massachusetts, where I endogenize the employer's choice of menu of products (in particular, the number of offered products and the networks of each product). 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 plan pricing and a set of products and provider networks to offer to its enrollees and customers. Using this model, I am able to decompose the extent to which employer persistence in offering broad-network products is attributable to: consumers valuing network *breadth* (a larger set of hospitals and physicians to choose from); consumer inertia in provider and plan choice; and firm fixed costs of product redesign. After estimating these model primitives, I simulate the equilibrium number of plans and networks offered by the employer under scenarios where consumers exhibit no inertia to their providers or health plans and firms face no fixed costs. I next test a potential policy intervention that moves the GIC to a fixed-dollar defined contribution pricing scheme.

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, including 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.[3] The GIC is an ideal setting for studying the welfare effects of narrow-network products. It has, in the last several years, been active in encouraging the creation and adoption of narrow-network products and offers plans with considerable variation in both hospitals and physicians covered. This variation allows me to estimate demand for insurance plans on a large section of the demand curve. Moreover, while I am not able to fully separate *physician* loyalty from switching costs, I am able to leverage a policy change in 2012 in order to separately identify *plan* switching costs from unobserved preference heterogeneity. Specifically, in 2012, the GIC offered a three-month "premium-holiday," both forcing all active state employees to re-enroll in a health plan, while simultaneously offering three months of free coverage if they switched from a broad-network to a narrow-network product (Gruber and McKnight, 2016). Such identification strategies in similar settings have been employed in prior literature on plan inertia, though usually in the context of variety on cost-sharing features, rather than variety on networks (Handel, 2013).

I offer two 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. In particular, 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] Much of the existing literature on networks has exclusively focused on hospitals (Ho, 2009; Shepard, 2016; Prager, 2016; Ho and Lee, 2017a; Ghili, 2017; Liebman, 2017) and has ignored the role

[3] 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.

[4] $https : //www.cdc.gov/nchs/data/ahcd/namcs_{s}ummary/2013_{n}amcs_{w}eb_{t}ables.pdf$

of physicians in determining consumer choice of insurance plans.[5] In my context, modeling the choice of physicians is important for several reasons. First, to the extent that consumer inertia to previously used providers is an important determinant of plan choice, this is likely to have a more significant effect on the physician dimension, particularly loyalty to primary care doctors. Second, estimates of premium elasticities and network valuations may be biased if the heterogeneity in physician preference is not incorporated. This may then lead to incorrect inferences about plan pricing, network structure, optimal plan menus, and ultimately consumer surplus and spending changes. Indeed, I show that demand for physicians is a more significant driver of health plan choice than demand for hospitals and that omitting them leads to significant overestimation of willingness-to-pay for hospitals and health plans.

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 would need to be compensated between \$15 and \$50 per month to move from a broad-network plan to a narrow-network plan from the same insurer (depending on the network and insurer), with approximately 85% of value coming specifically from loss in access to physicians. Of the total valuation of physicians, approximately 50% comes from loss of access to previously used physicians, particularly primary care doctors. These results have meaningful implications for demand patterns: I find that estimates of consumer price sensitivity are significantly biased downward if physician networks and provider inertia are ignored, raising the implied valuation of a plan by 25%.

This also has a significant impact on estimates of plan switching costs. I estimate that average switching costs are \$250 per household per month, representing about 63% of the largest family premium and about 95% of the smallest family premium. Without accounting for heterogeneity in physician networks, estimates of switching costs increase by 23% to \$308 per household per month. These switching costs explain a significant portion of enrollment into broad-network plans, even conditional on valuation of hospital and physician networks. In a sense, they lead to plans being "overpriced" relative to how much consumers value their networks.

My second main contribution is that I study the role that plan switching costs play in explaining the equilibrium products offered by endogenizing plan menus with respect to hospital and physician networks. I fully specify and estimate an employer objective function that balances value of the menu to the consumer, net spending on premiums, and fixed costs of offering additional plans. I leverage variation in networks over time and the number of plans offered over time to estimate two key parameters: the relative weight that the employer places on consumer surplus over net spending, and the fixed costs. I rely on using moment inequalities to bound parameter estimates in the employer objective function (Pakes et al., 2015; Pakes, 2010). This approach has been used in a variety of contexts and markets, including computers (Eizenberg, 2014; Nosko, 2014), pharmaceuticals (Mohapatra and Chatterjee, 2015), and smartphones (Fan and Yang, 2016). To

[5] 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, estimation of physician demand is complicated by dimensionality: 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.

my knowledge, this is the first paper that allows employers to reoptimize both its number of plans and specify the hospital and physician networks of those plans in response to the removal of consumer switching costs. Indeed, doing so significantly alters both the equilibrium plan menus and significantly alters the welfare and spending implications relative to a scenario when plans remained fixed.

I find that the employer places a significantly higher weight on consumer surplus relative to spending on premiums. In particular, the employer values a dollar of consumer preferences more than four times as much as it values a dollar of spending. This leads the employer to maintain access to broad-networks for its employees even when removing those networks may lead to substantial savings in costs. Further, I estimate the fixed costs (both monetary and non-monetary) of offering additional choice for consumers to be approximately $8 million per plan.[6] The presence of these costs inhibits the employer from offering additional choice of networks, even if doing so would see some benefit for consumers. Though quite large in an absolute sense, these estimates represent a small share (roughly 1%) of employer spending on claims. Moreover, as some of this estimate is due to the idiosyncratic error shock common to multinomial logit models, I provide additional estimates when the shock is set to zero. Unsurprisingly, this reduces the fixed cost estimates to around $1.5 million per plan, more in line with reported estimates.

I then use these estimates to conduct several policy-relevant counterfactual exercises where I remove various sources of consumer and employer frictions. First I simulate equilibrium plan menus, welfare, and premium spending when consumer loyalty to their previously used physicians is removed. I find that doing so results in significant declines in consumer surplus relative to baseline, but little declines in overall costs or premiums, resulting in the employer keeping the plan menu virtually identical. The implication is that, while patients derive much value from previously used physicians, these physicians are not necessarily the expensive providers in the market. This, in combination with the large value the employer places on consumer surplus, implies that the cost savings from narrowing networks would not be enough to compensate this consumer loss of surplus.[7]

I next model the removal of consumer health plan switching costs as well as scenarios where the switching costs are present, but that the employer does not consider them welfare-relevant when making its plan menu decision. Removing plan switching costs induces approximately 14% of households to switch from a broad to narrow plan. If plan menus are held fixed, this move results in a net premium cost decline of about $13 per household per month. Consumer welfare declines slightly if the switching costs are considered welfare-irrelevant (due to the presence of selection, as in Handel (2013)), but increase substantially if they are considered tangible costs.

However, I find these changes have significant impacts on equilibrium menus, which alter the welfare calculations. I find that in response to removing plan switching costs, the employer significantly increases the availability of narrow-network plans, both by increasing the *number* of plans available to employees and substantially increasing the *variety* of narrow network offerings. In

[6] These costs include not only the true monetary fixed costs of offering product variety, but non-monetary costs including firm concerns over confusing consumers, etc. It can therefore be thought of as a firm-level choice friction.

[7] One caveat is that, for the physician models, I cannot separate unobserved heterogeneity for physicians from physician switching costs. As a result, I take the most conservative approach and assume that the entirety of physician inertia is due to preference heterogeneity (or true loyalty).

addition, when the idiosyncratic logit error is set to zero (thereby implicitly removing consumer benefit from simply having a higher *number* of choices), the employer increases the number of narrow-network products, but decreases the number of broad-network products available. Notably, in this scenario, the network variety increases, but is limited to a lower-cost insurance carrier, while most plans from higher-cost carriers are dropped. This menu change alters costs dramatically: net spending on premiums declines by about $20 per household per month if the employer retains access to all broad-networks, and by about $40 per household per month if it removes the flagship broad-networks. Consumer surplus also increases from the new menu, particularly if the broad networks are retained. If the employer removes the broad networks, consumer surplus declines slightly, but by no more so than if plan menus remained fixed. Even so, this loss is more than compensated for by the substantial decline in costs.

Finally, I test how plan menus would change under an alternate pricing scheme, where the GIC subsidies premiums by a fixed-dollar amount equivalent to 90% of the lowest-cost plan available to consumers. I find that doing so significantly penalizes households wishing to purchase broad-network products. However, it also reduces premiums for narrow-network plans by such a significant extent that the GIC is able to removes access to most broad-network plans, while still compensated consumers for the switch. The result is a substantial increase in consumer surplus.

The implication of these results is that employers internalize consumer health plan switching costs when making product offer decisions, which lead to an underprovision of narrow networks and potential overprovision of broad networks. Removing switching costs or otherwise offering consumers incentives to switch would allow consumers to more efficiently sort into plans, which would lead to increased network variety in plan menus, and achieve significant social surplus gains.

This paper relates to several strands of literature. The first strand includes studies on switching frictions and health plan inertia (Handel, 2013; Polyakova, 2016; Abaluck and Gruber, 2016; Ho et al., 2017). I also contribute to the literature on network formation (Ho, 2006, 2009; Shepard, 2016; Lee, 2013; Liebman, 2017; Ghili, 2017) and valuation of narrow-network plans (Gruber and McKnight, 2016; LoSasso and Atwood, 2015; Dafny et al., 2015; Ericson and Starc, 2015a). A third strand focuses on the value of insurance plan choice, competition, and provision (Ericson and Starc, 2015b, 2016; Dafny, 2010; Dafny et al., 2012, 2013; Scheffler et al., 2016). Finally, I contribute to the literature on models of product entry, innovation, and variety that endogenize firm product quality choices (Nosko, 2014; Eizenberg, 2014; Mohapatra and Chatterjee, 2015). Of particular importance is Handel (2013), who uses a similar natural experiment in a different setting to identify switching costs and Shepard (2016), who uses a similar demand model to study whether adverse selection leads to the narrowing of networks on the individual market. In addition, Prager (2016) also studies network design on the GIC, focusing on whether tiered-network products lead to a reduction in negotiated prices between insurers and providers.

The paper proceeds as follows: Section 2 outlines the data and setting for my study and presents some empirical patterns. Section 3 details the model. Section 4 outlines my estimation and identification strategies. Section 5 presents the estimated parameters and results from the model. Section 6 presents the results of counterfactual policy simulations. Section 7 concludes.

2 Data and Empirical Patterns

2.1 Group Insurance Commission

The focus of this paper is the Group Insurance Commission (GIC) in Massachusetts, a large purchasing organization in Massachusetts that services the state's government employees–both employees of the state itself and local municipal governments. Though state employees constitute the bulk of GIC members, since 2007, municipalities have increasingly abandoned their existing insurance arrangements in favor of getting insurance through the GIC and, as such, there are a large number of municipal entrants in subsequent years. Therefore, the GIC has an interest in not only providing satisfactory health benefits for its existing members, but potentially competing for new members as well. In total, there are approximately 300,000 enrollees on the GIC, representing approximately 8% of the Massachusetts employer-sponsored-insurance market.

The GIC contracts with multiple health insurance carriers and provides multiple competing plans for enrollees. In particular, it contracted with six carriers throughout my sample period: Fallon Community Health Plan, Harvard Pilgrim Health Care, Health New England, Neighborhood Health Plan, Tufts Health Plan, and Unicare Health. Each carrier offers multiple plans at different premiums. The premiums do not vary by consumer risk type or geography, but rather only with whether the household is a single-member ("individual") or multi-member ("family") household. Specifically, all GIC family plans are 2.4 times the individual rate. Apart from premiums, these plans are entirely standardized with the exception of two dimensions. The first is that the GIC employs tiered copay arrangements, which generates variation in copays across providers (discussed at length in Prager (2016).

The second is the actual network of included providers on each plan. In 2009 and 2010, four of the carriers offered narrow network products with varying degrees of network breadth. In 2011, the GIC enacted a major change to the choice set by introducing narrow-network plans from both remaining insurers (Harvard Pilgrim and Tufts Health Plan, two dominant players in the state).[8] These plans are approximately 20% cheaper on average than their respective broad networks, though generally cover more providers than the narrow-network plans offered by the same insurers in other market segments.[9]

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. For households choosing to make the switch, the holiday entailed that they pay *no* premiums for three months of the fiscal year. Importantly, this holiday was not extended to municipal workers, but rather just active state employees. This served as the basis for prior work on evaluating the impact of narrow-network product introduction (Gruber and McKnight, 2016). The holiday was fairly successful, inducing approximately 10% of enrollees to switch and resulting in approximately 20% savings in spending for those enrollees, largely due to the use of lower-cost providers.

[8] These plans are called "Harvard Primary Choice" and "Tufts Spirit", but will hereafter be referred to as "Harvard Narrow" and "Tufts Narrow."

[9] For example, Harvard offers a narrow-network plan in the small-group market known as "Harvard Focus," which is considerably narrower than the "Primary Choice" plan offered on the GIC.

Table 1 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 (hereafter "Harvard Broad") and Harvard Primary Choice. 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. Interestingly, despite having lower out-of-pocket premiums, Tufts Spirit had a significantly lower market share than Harvard Primary Choice.[10] This is a point that I will return to below.

Table 1: GIC Summary Statistics, 2012

Insurer	Network Coverage	Market Share	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		

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

2.2 Data Sources

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.

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.

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

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.

Using the APCD as well as publicly available network and premium data from the GIC, I create samples for hospital admissions, physician visits, and insurance plan choice. I focus, in particular, on primary care physicians (PCPs), cardiologists, and orthopedists. A detailed description of the several different subsamples I create pertaining to different stages of my model are presented in Appendix A, along with summary statistics for the different specialties.

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. I also use the SK&A data to infer the physician practice networks of plans on the GIC. Details of the network construction are found in Appendix A.

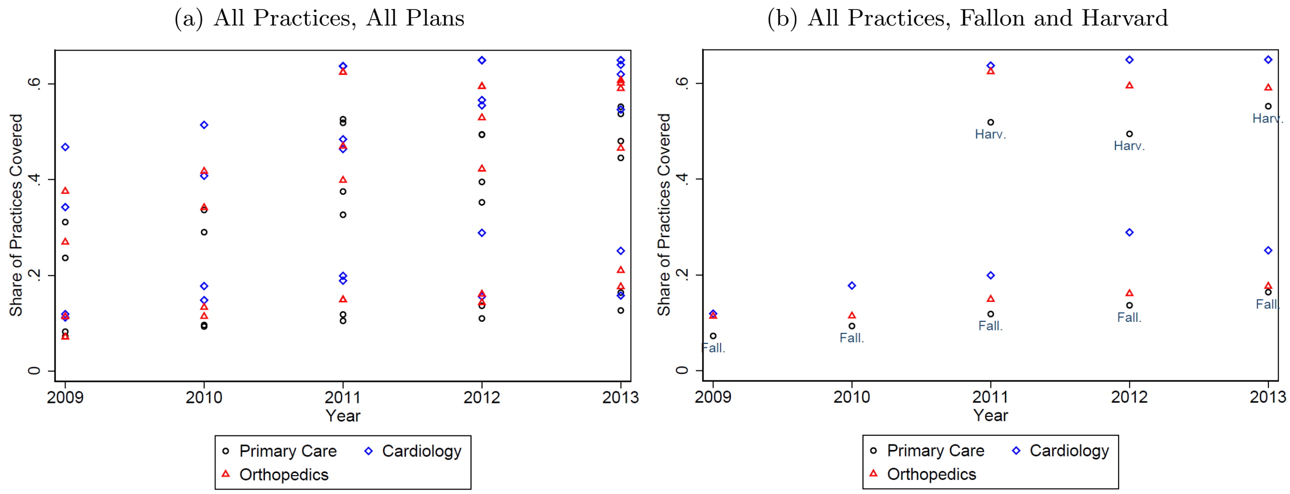
2.3 Empirical Patterns

Variation in Physician Networks: Figure 1 shows the variation in physician networks for each of the narrow-network products offered on the GIC between 2009 and 2013. For each plan, I report the share of physician practices in Massachusetts within a particular specialty group (PCP, cardiology, and orthopedics) that are covered by the plan. Panel (a) reports these network shares for all narrow-network plans. Within a given year, there is considerable variation in the physicians covered by these networks. In 2011, for instance, the share of primary care practices covered ranged from approximately 10% to 50%. The share of cardiology practices covered ranged from approximately 20% to 60%, and the share of orthopedic practices covered ranged from about 15% to 60%.

The network shares also vary over time in addition to across plan. In general, the networks of narrow plans seem to get slightly broader over time, covering somewhat larger shares of practices in 2013 compared with 2009. In 2013, the broadest cardiology network (among the narrow-network plans) covered over 60% of cardiology practices, compared with under 50% in 2009. The broadest orthopedist network rose from covering under 40% of practices in 2009 to just under 60% in 2013.

Much of this increase came from a noticeable spike in network breadth in 2011, which coincides with the introduction of Harvard Narrow and Tufts Narrow.

Figure 1: Share of Practices Covered by Year and Specialty



Notes: This figure plots the share of all physician practices covered by year and specialty for all narrow-network products on the GIC. Each point represents a particular insurance plan on the GIC. Panel (a) displays each insurance plan, while panel (b) displays only Fallon Direct (Fallon's narrow-network plan) and Harvard Primary Choice (Harvard's narrow-network plan)

However, some of this variation over time masks heterogeneity across insurer. Panel (b) reports the same network shares but only for Harvard Pilgrim's Primary Choice network and Fallon's Direct network. As Primary Choice only entered the market in 2011, networks are only reported for 2011-2013. Here it is clear that while Fallon's narrow network trended somewhat broader between 2009 and 2013, Harvard's network remained for the most part consistent. Its cardiology network might have even become somewhat *narrower* between 2011 and 2013.[11]

Figure 2 and Figure 3 show the hospital and primary care practice networks for a select group of products available on the GIC in 2011: Harvard Broad, Harvard Narrow, and Tufts Narrow. 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.[12] Panel (a) of Figure 3 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.[13].

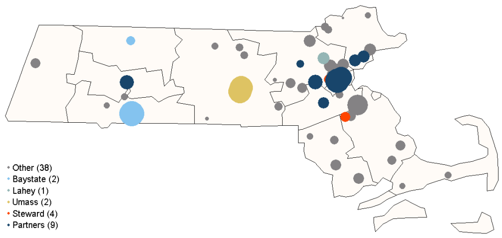
Panels (b) and (c) of Figure 2 and Figure 3 reveal that the Harvard Narrow and Tufts Narrow

[11]Some of these changes may be, in part, driven by measurement error on account of the way in which I constructed these networks. Appendix A describes details of the network measures. In addition, I describe robustness tests of these network trends in Appendix B. Overall, the trends appear similar regardless of the particular ways in which networks are constructed and inferred.

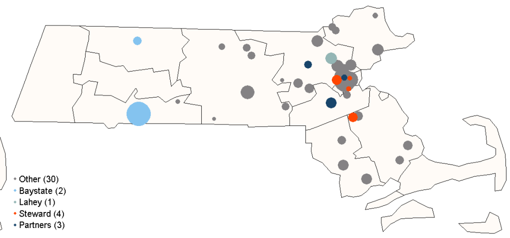
[12]In general, Partners is widely known to be one of the dominant players in the Massachusetts provider market, owning several large academic medical centers, including Mass. General Hospital and Brigham and Womens Hospital. Similarly, Atrius Health is one of the dominant players in the physician market, owning several key medical groups,

Figure 2: Hospital Networks by Plan, 2011

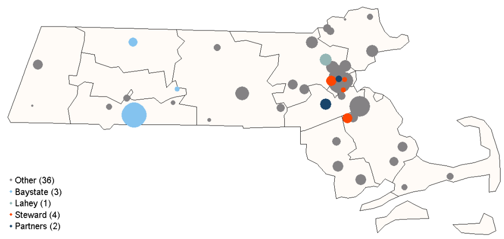
(a) Harvard / Tufts Broad



(b) Harvard Narrow



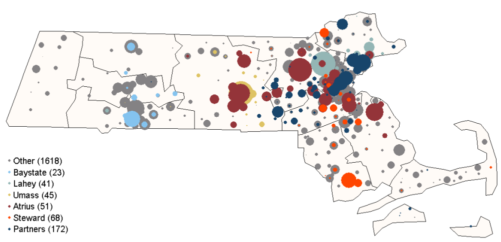
(c) Tufts Narrow



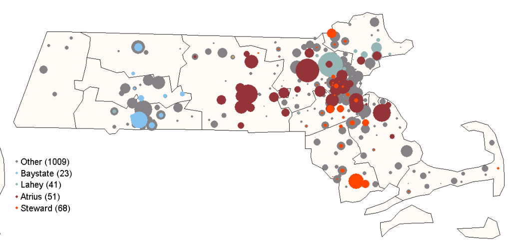
Notes: This figure plots the hospital networks of specified plans on the GIC in 2011. Sizes of the data points reflect relative market shares of the practices. Colors reflect ownership status (which health systems owns which practice).

Figure 3: Primary Care Practice Networks by Plan, 2011

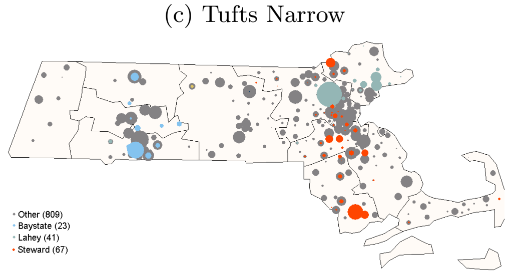
(a) Harvard / Tufts Broad



(b) Harvard Narrow



(c) Tufts Narrow



Notes: This figure plots the physician practice networks of specified plans on the GIC in 2011. Sizes of the data points reflect relative market shares of the practices. Colors reflect ownership status (which health systems owns which practice).

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 between the broad and narrow hospital networks is that most Partners hospitals were dropped from each narrow plan. However, as noted in Table 1, 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 3 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, but not Tufts, covers Atrius Health (noted by the red points in the map). This indicates 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 B 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.

Evidence of Consumer Switching Costs in Choice of Insurance Plans: There is significant heterogeneity in terms of who is enrolling in narrow network plans. Figure 4 depicts the share of GIC consumers enrolling in narrow network plans by year and by whether they were new to the GIC that year (i.e. "entering members" or "active choosers") or whether they were existing GIC members who were automatically re-enrolled in their current plan unless they took action (i.e. "existing members" or "passive choosers"). Panel (a) depicts the share of new members enrolling in narrow-network plans, whereas panel (b) depicts the share of existing members. In 2009 and 2010, the share of enrollment of active choosers and passive choosers in narrow network plans both hovered around 15%. However, there is a large spike in the share of new members enrolling in narrow-network plans in 2011 (to 30%), when the GIC introduced Harvard Narrow and Tufts Narrow. Conversely, once the GIC introduced the "premium holiday" in 2012, there is a significant spike in the share of *existing* members (for whom the policy applied, panel (b)) enrolling in narrow-network plans.

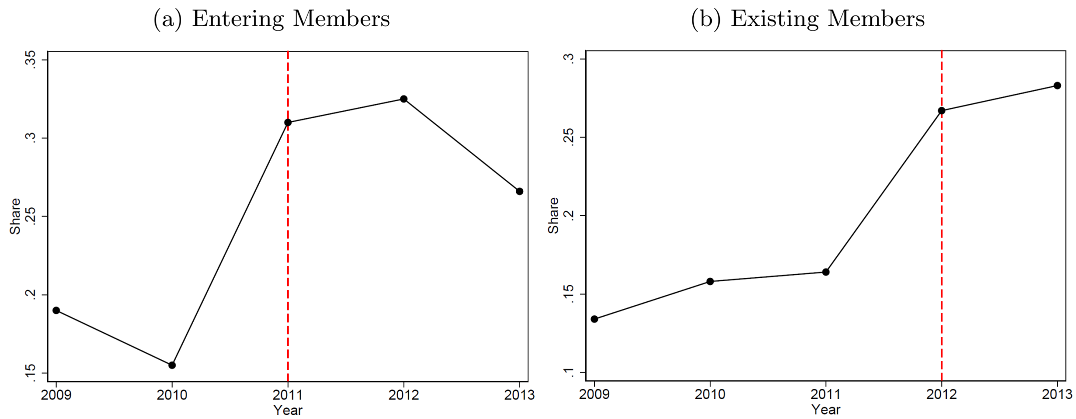
These differences in behavior between active and passive choosers suggest the presence of a high degree of consumer inertia in choosing health plans (Handel, 2013). One potential criticism of this conclusion is that active choosers might have different preferences for networks than passive choosers, or might otherwise be fundamentally different in ways that drive their choice of health plans. For instance, new employees of firms tend to be younger, and younger individuals tend to be more price sensitive in choosing health insurance plans than older families.

However, institutional details would suggest that these demographic differences between active and passive choosers is fairly minimal. First, most new members to the GIC are from new munic-

including Harvard Vanguard.

[13]This is due to its purchase of the Fallon Clinic, later renamed "Reliant Medical Group" in Worcester in 2011

Figure 4: Share of People in Narrow Network Plans by Year and Whether New to GIC



Notes: This figure plots the share of members selecting narrow network plans. Panel (a) plots the share of new members to the GIC enrolling in narrow-network plans. The dashed red line represents the year when the GIC introduced two new narrow-network plans from Harvard and Tufts. Panel (b) plots the share of existing members enrolling in narrow-network plans. The dashed red line represents the year of the "premium holiday."

ipalities in Massachusetts contracting with the GIC for health plans, rather than new employees entering the firm. These municipalities tend to be geographically disperse across the state, and the cohort of new workers entering from these municipalities tend to have similar observables to those already on the GIC. To see this, Table 2 shows a regression of enrollment in narrow-network plans on a set of household observables, as well as an indicator for whether the household was new to the GIC that year. Indeed, older households are less likely to enroll in a narrow network plan, as are households with at least one member with a chronic illness. Larger households are also less likely to enroll in a narrow-network plan. However, even controlling for these, as well as year and county fixed effects, existing members of the GIC are, on average, 11% less likely to be enrolled in a narrow-network plan than new members.

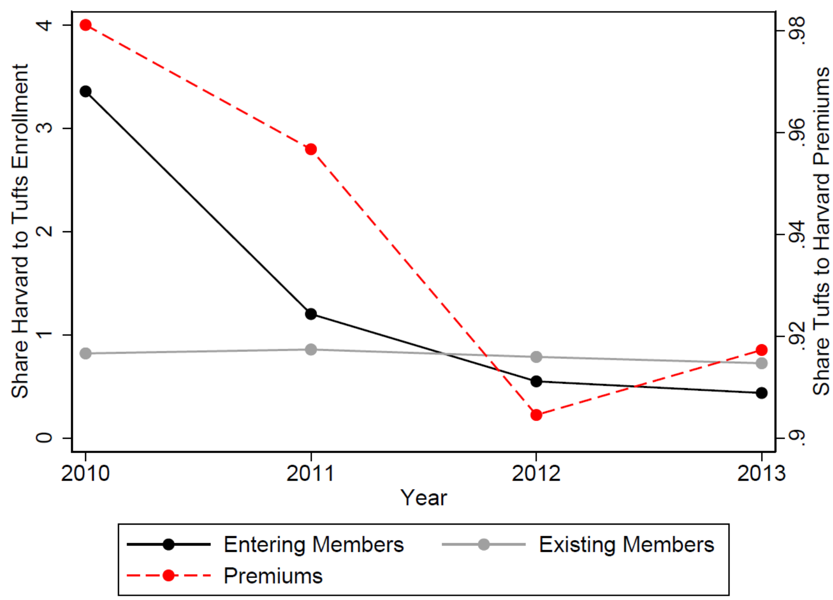
Table 2: Probability of Enrolling in a Narrow Plan

Variable	Coefficient	Standard Error
Existing GIC Member	-0.113***	0.003
Age	-0.003***	0.000
Female	0.009***	0.002
Chronic Condition	-0.021***	0.003
Members in HH	-0.009***	0.003
Constant	0.670***	0.006
Year FE	Yes	
County FE	Yes	
Obs.	151,331	
Adj R2	0.432	

Notes: Results from regression of enrollment in a narrow network plan on household characteristics. GIC sample 2009-2013.

To see more evidence that new cohorts behave differently than older cohorts (suggesting the presence of inertia in plan choice), one need not look only at enrollment in broad versus narrow network plans; but also at the stickiness of enrollment in broad-network plans as the characteristics of those plans change. To that end, I note that 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 thereafter, with Harvard's broad network growing significantly more expensive than Tufts.

Figure 5: Share of Members Enrolling in Tufts Broad Network Plan by Whether New to GIC



Notes: This figure plots the share of ratio of members selecting Harvard's broad-network plan over Tufts' broad-network plan as well as the ratio of the premium difference between Tufts and Harvard. The dark line plots the ratio of entering (new) members to the GIC that year. The light grey line plots the ratio of existing members on the GIC. The dashed red line plots the premium ratios.

Figure 5 shows the ratio of enrollment in Harvard Broad versus Tufts broad over time, along with the change in the ratio of Tufts Broad premiums to Harvard Broad premiums. The black line represents the Harvard-to-Tufts enrollment ratio for new members to the GIC, while the light grey line represents the Harvard-to-Tufts enrollment ratio for existing GIC members. First, it is notable that as Tuft's premiums fall relative to Harvard's, enrollment in Tufts relative to Harvard rises dramatically among new members to the GIC. By 2012, Tufts' premiums were about 10% less than Harvard's (representing about $30 per month for families). Enrollment in Harvard among new members, meanwhile, declined from more than three times that of Tufts in 2009 to about 90% that of Tufts in 2012. Second, existing members exhibit no such changes in enrollment patterns. Between 2010 and 2013, enrollment among existing members in Harvard relative to Tufts barely budged, even as the premium difference widened considerably.

Taken together, these two figures provide some suggestive evidence of two behaviors. The first is that consumers may exhibit a high degree of inertia after their initial choice of health plans, with existing members sticking to their choices, even as premiums grow relative to other similar plans on the market, or as new options appear that are considerably cheaper. Second, once taking this inertia into account, consumers may actually be quite price sensitive in their choices of insurance plans, which is a characteristic often not attributed to purchasers of employer-sponsored insurance.

These two stylized facts motivate my inclusion of inertia in the model I present in the next section.

3 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.

3.1 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^H_{jt}, 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^S_{jt}. Consumer utility for patient of type i, with either illness l or procedure r, from visiting a provider takes the following form:

$$u_{ilht} = \underbrace{T_{iht}\lambda_1 + T_{iht}v_{ilt}\lambda_2 + T_{iht}x_{ht}\lambda_3 + x_{ht}v_{ilt}\lambda_4 + \mathbb{1}(ih_t = ih_{t-1})\lambda_5 + \gamma_h}_{\phi_{ilht} \text{ (Hospitals)}} + \varepsilon_{ilht} \tag{1}$$

$$u^s_{irdt} = \underbrace{T^s_{idt}\lambda^s_1 + T^s_{idt}v_{irt}\lambda^s_2 + T^s_{idt}x^s_{dt}\lambda^s_3 + x^s_{dt}v_{irt}\lambda^s_4 + \mathbb{1}(id^s_t = id^s_{t-1})\lambda^s_5 + \gamma^s_d}_{\phi^s_{irdt} \text{ (Physician Specialty } s\text{)}} + \varepsilon^s_{irdt} \tag{2}$$

where x_{ht} is a vector of observed hospital characteristics, x^s_{dt} is a vector of observed physician practice characteristics for specialty type s, v_{itl} and v_{irt} are observed characteristics of patient i with diagnosis l or requiring procedure r, T^s_{idt} and T_{iht} is the distance in miles from patient i's location to provider d or h's location, γ^s_d 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 $\mathbb{1}(id^s_t = id^s_{t-1})$ refers to whether individual i saw physician practice d for specialty care s in any year prior to t. The latter parameter represents inertia to previously used physicians. Its specification is described in more detail in Appendix C.

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

$$\sigma_{ilht} = \frac{exp(\phi_{ilht})}{\sum_{k=1}^{N^H_{ijt}} exp(\phi_{ilkt})} \tag{3}$$

where N^H_{it} refers to the number of hospitals in individual i's network in time t. Similarly, the probability that patient i needing a procedure with RVU r from specialist group s will chose physician practice d is:

$$\sigma^s_{ird} = \frac{exp(\phi^s_{irdt})}{\sum_{k=1}^{N^S_{ijt}} exp(\phi^s_{irkt})} \tag{4}$$

where N^S_{ijt} is the network of practices of type s in individual i's network.

3.2 Consumer Demand for Insurance Plans

I assume that choice of health plan is done at the household level. Therefore, the utility of household I for plan j at time t is given by the following:

$$u_{Ijt} = \underbrace{-r_{Ijt}\alpha_I + EU^H_{Ijt}\beta_1 + \sum_s EU^s_{Ijt}\beta^s_2 + \mathbb{1}(Ij_t = Ij_{t-1})\beta_3 + \eta_j}_{\delta_{Ijt}} + \omega_{Ijt} \tag{5}$$

Here, r_{Ijt} refers to the plan rate, or premium, which varies only by whether 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^H_{Ijt} is the expected utility from the plan's hospital network and EU^s_{Ijt} is the expected utility from the plan's network of physician specialty s. They measure household I's willingness-to-pay for a particular insurance plan's provider network, incorporating not just network size, but relative quality of the providers in the network as determined by the provider demand stage.[14] These terms are defined as:

$$EU^H_{Ijt} = \sum_{i \in I} \sum_l f_{il} log \left(\sum_{h \in N^H_{jt}} exp(\phi_{ilht}) \right)$$

$$EU^s_{Ijt} = \sum_r f_{ir} log \left(\sum_{d \in N^S_{jt}} exp(\phi^s_{irdt}) \right)$$

where, f_{il} and f_{ir} are the ex-ante probabilities that individual i contracts diagnosis l (requiring hospital care) or requires procedure r (requiring physician care). Note that, as demand for

[14] A network may, for instance, have fewer providers, and yet still yield a higher value of EU^s_{Ijt} for specialty group s is the physicians included are of higher demand than the larger network.

insurance plans is at the *household* level, the expected utility variables are also aggregated to the household level by summing over each individual i's willingness-to-pay for the provider networks. The assumption is that a household's total utility for a particular hospital and physician network is a linear combination of all its individual household members. Both expected utility terms vary over time and across households. 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 switching costs are captured by $\mathbb{1}(Ij_t = Ij_{t-1})$, 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:

$$s_{Ijt} = \frac{exp(\delta_{Ijt})}{\sum_{k=1}^{J} exp(\delta_{Ikt})} \tag{6}$$

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

I assume that the GIC (the employer), in selecting products and setting prices, maximizes a weighted 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 care about satisfying the health care needs of their employees, not just cost. 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 expand 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 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 (Bundorf, 2002; Moran et al., 2001).[15]

In my setting, the GIC weights the benefits of offering these plans against these fixed costs. Formally the GIC objective is:

[15] Bundorf (2002) notes that firms report that these costs inhibit them from offering more choice and variety to their consumers. In a sense, these fixed costs can also be thought of as opportunity costs or firm switching costs. These costs do not include merely monetary costs of offering additional plans, but costs that firms internalize of informing consumers and perhaps contributing to consumer confusion. In particular, recent research has shown that consumers facing a large number of choices often feel overwhelmed, resulting in the choice of "dominated" plans that are financially inferior to other options (Liu and Sydnor, 2018).

$$W_t = \underbrace{\rho CS(\delta_{Jt}, \theta)}_{\text{Weighed Consumer Surplus}} \underbrace{- \sum_I \sum_j (1-\tau) s_{Ijt}(\delta_{Jt}, \theta) R_{Ijt}(\delta_{Jt}, \theta)}_{\text{Net Health Spending}} - \underbrace{\sum_j FC_j}_{\text{Fixed Costs}} \tag{7}$$

where:

$$CS(\delta_{Jt}, \theta) = \sum_I \frac{1}{\alpha_I} log \left(\sum_j^J exp(\delta_{Ijt}) \right)$$

Here, the term on the left-hand-side of the function, $CS(\delta_{Jt}, \theta)$ is the consumer surplus from the GIC offering J products to its employees. This consumer surplus is a function of estimated demand parameters, θ, and the GIC's chosen plan menu, δ_{Jt}. 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%. 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. The third term, FC_j represents the fixed cost to the GIC of offering plan j to its enrollees. ρ refers to the relative weight that the GIC places on consumer surplus over dollars spent on premiums (or medical claims) and fixed costs. I estimate both FC_j and ρ as part of the first-stage of the model.

Insurers, meanwhile, are assumed to set premiums to maximize profits. Consistent with prior literature, I assume that costs for a particular provider can be decomposed into an insurer-provider-specific base negotiated rate, p_{jht} and p_{jdt}, scaled by a disease or procedure weight. Let the marginal *hospital* cost for plan j therefore be given by:

$$c^o_{jtH}(N^H_{jt}) = \sum_{i \in I} \sum_l f_{il} w_{lt} \sum_{h \in N^H_{jt}} \sigma_{ilht}(N^H_{jt}) p_{jht} \tag{8}$$

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

$$c^o_{jtS}(N^S_{jt}) = \sum_{i \in I} \sum_s \sum_r f_{ir} RVU_{rt} \sum_{d \in N^s_{jt}} \sigma^s_{irdt}(N^s_{jt}) p^s_{jdt} \tag{9}$$

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

$$\pi_{mt} = \sum_{j \in J_m} \sum_I \left(s_{Ijt}(\delta_{Jt}, \theta) \left[R_{jt}(\delta_{Jt}, \theta) \theta^R_I - \underbrace{c^o_{jtH}(N^H_{jt}) - c^o_{jtS}(N^S_{jt})}_{c^o_{jt}(N_{jt})} - c^u_{jt}(N_{jt}) \theta^c_I \right] \right) \tag{10}$$

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^H_{jt} refers to the hospital network of plan j and N^S_{jt} refers to the physician network of plan j). R_{jt} denotes a "base premium" for each plan in each year. The θ^R_I next to the premium variable refers to the multiple of the base premium and depends only on household type (individual versus family). 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 8, w_{lt} refers to the weight assigned to a particular hospital diagnosis. These weights were also used in the hospital demand model in subsection 3.1 (a more thorough discussion of their construction follows in subsection 4.3. In Equation 9, 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. p_{jht} is the negotiated base price between plan j and hospital h in time t, while p^s_{jtd} 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 subsection 4.3. Finally, c^u_{jt} refers to "base" unobserved plan costs for plan j in time t.[16] I assume that these costs scale linearly across household type, i.e.:

$$c^u_{Ijt} = c^u_{jt}\theta^c_I$$

where θ^c_I is the parameter that scales these base unobserved costs across households. Assuming a multi-product Nash-Bertrand price-setting equation, the first-order condition for the insurer profit function (dropping the (δ_{Jt}, θ) for ease) is:

$$0 = \frac{\partial \pi_{mt}}{\partial R_{jt}} = \sum_I \left(s_{Ijt} + \sum_{n \in J_m} \frac{\partial s_{Int}}{\partial R_{jt}} \left(R_{jt}\theta^R_I - c^o_{jt}(N_{jt}) - c^u_{jt}(N_{jt})\theta^c_I \right) \right) \tag{11}$$

For single-product firms, this can be re-written in terms of the base premium rate as:

$$R_{jt} = \frac{1}{\sum_I s'_{Ijt}\theta^R_I} \sum_I s'_{Ijt}(c^o_{jt} + c^u_{jt}\theta^c_I) - \frac{\sum_I s_{Ijt}\theta^R_I}{\sum_I s'_{Ijt}\theta^R_I} \tag{12}$$

where $s'_{Ijt}(\delta_{Jt}, \theta) = \frac{\partial s_{Ijt}}{\partial R_{jt}}$. 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. In particular, the second term in Equation 12 is the markup term, depending on both the market share of a particular insurer and the price elasticity. 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 capture more market share and thus the markup term will be higher.

These assumptions are fairly strong for this setting, however, for several reasons. First, two of the largest plans offered by the GIC (Harvard Broad and Tufts Broad) are self-insured, and as such as paid for administrative services only. As a result, the markup term will be less salient.[17] Second, the GIC, as a large employer group that covers about 8-9% of the state's employees, has considerable bargaining leverage with insurers to reduce premiums, thereby inhibiting insurers from setting markups that are too high.[18] Finally, plans in Massachusetts are bound by state medical-

[16] These costs include physician specialties not modeled in this paper, pharmaceutical spending, etc.

[17] Industry experts note that, for these plans, insurers offer the GIC a "suggested" premium based on anticipated costs, but that GIC is free to set rates for consumers at their discretion. As a result, the GIC has incentives to keep premiums low (to increase the consumer surplus term, $CS(\delta_{Jt}, \theta)$, in Equation 7).

[18] An industry expert noted that insurance plans gain considerably from contracting with the GIC and, as such, are largely willing to capitulate to the GIC's requests for premiums and plan designs. See Ho and Lee (2017b) for a model that incorporates employer-insurer bargaining over premiums using data from CalPers (an employer group similar to the GIC) in California.

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 (Prager, 2016).

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:

$$\sum_I s_{Ijt} R_{Ijt} \theta_I^R = 1.10(c_{jt}^o(N_{jt}) + c_{jt}^u(N_{jt})\theta_I^c) \tag{13}$$

My main specification in section 5 reports the results from the fixed-markup assumption. This assumption fits the observed price-cost margins in the data much more closely than the Nash-Bertrand assumption, which significantly overestimates predicted premiums.

3.4 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 plan menu δ_{Jt} to maximize:

$$\max_{\delta_{Jt}} \left[E \left(\underbrace{\rho CS(\delta_{Jt}, \theta) - \sum_I (1-\tau) s_{Ijt}(\delta_{Jt}, \theta) R_{Ijt}(\delta_{Jt}, \theta)}_{S_t(\delta_{Jt}, \theta)} \right) - \sum_j FC_j \right] \tag{14}$$

Here, $S_t(\delta_{Jt}, \theta)$ 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). As an alternate specification, I also consider a model where, rather than 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:

$$\max_{\delta_{Jmt}} \left[E\left(\pi_{mt}(\delta_{Jmt}, \delta_{J,-mt}\theta)\right) - \sum_j FC_j \right] \tag{15}$$

where δ_{Jmt} is the product menu J offered by MCO m and $\delta_{J,-mt}$ is the product menu offered by other mcos $-m$. I report the fixed costs under both sets of assumptions in section 5.

4 Estimation and Identification

4.1 Provider Choice Stage

Estimation: The patient choice of providers is estimated using maximum likelihood. Estimation of hospital demand follows techniques standard in the literature (Ho, 2006). For estimating the physician models, I make additional assumptions in order to reduce the dimensionality of the estimation. Further, I estimate the models separately by the seven Massachusetts health rating regions[19] and by specialty group (PCP, cardiology, and orthopedics). Details of the estimation are presented in Appendix C.

Identification: 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.

Identification of the inertia coefficient, λ_5^s relies on two conditions to be true. The first is that choices made by "active choosers" need to be different from choices made from "passive choosers." In my setting, this variation comes from differences in choices made between patients who have never sought care from *any* physician within a particular specialty group and patients who previously sought care from a physician, conditional on other observables included in the model. For instance, difference between a 30-year-old female patient's choice of physician practice who just moved to Boston, versus the choices made by a 30-year-old female patient with a similar medical history who lived in Boston for years helps identify λ_5^s. The second condition is that the choice set changes over the sample period. In this setting, there are many changes to physician choice sets over time. One comes from network changes, as depicted in Figure 1. Another is that physician practices frequently enter the market, exit the market, or change ownership. Indeed, over my sample period, a large wave of vertical integration resulted in large hospitals and health systems purchasing many of the individual practices. Finally, individual physicians frequently move across practices. As such, both the number of choices and the utility of choices changes considerably over time. The main identifying assumption is that, controlling for the considerable *observed* heterogeneity in the model, along with the changing choice set of providers, is sufficient towards identifying physician switching costs. This identification is similar to techniques employed by Handel (2013) and Polyakova (2016).

However, a few features of the data prevent me from fully separating true inertia from unobserved persistent preference heterogeneity. One is an initial conditions problem: patients who I observe never having used a particular specialist may have, in fact, done so prior to 2009 (the first year of data I have). Second, although the number of physicians moving across practices and locations is substantial, the number of new patients to Massachusetts is smaller. As such, I choose to focus on the most conservative interpretation of physician inertia possible and treat λ_5^s as entirely reflecting persistent preference heterogeneity.[20]

[19]The rating regions are detailed in https://www.cms.gov/CCIIO/Programs-and-Initiatives/Health-Insurance-Market-Reforms/ma-gra.html.

[20]Shepard (2016) discusses this issue in detail in his context of hospital inertia.

4.2 Plan Choice Stage

Estimation: 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. Details are presented in Appendix C.

Identification: The expected utility coefficients, β_1 and β_2, are 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 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, while expected utility from the provider network is linear in the number of household members.[21] 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}.

The inertia parameter, β_3, suffers the same potential identification problem as λ_5. However, unlike the provider demand stage of the model, features of the data and setting allow me to more cleanly separate plan switching costs from preference heterogeneity. First, I observe a substantial number of enrollees making choices for the first time, driven by households from municipalities entering the GIC for the first time between 2009 and 2013. As described in subsection 2.3, these municipal "active choosers," conditional on a rich set of observables, make extremely different choices in plans than members previously enrolled in a GIC. If inertia to previous plans was driven by preference heterogeneity, we would not expect such considerable differences between these two groups of enrollees. Second, the "premium holiday" in 2012 forced all active state employees to re-enroll in a plan at the same time the GIC both introduced new plans into the choice set and significantly decreased the premiums for a subset of those plans (the effect of this holiday on enrollment is shown in Figure 4). Similar in spirit to Handel (2013), the primary identifying assumption is that, controlling for detailed ex-ante health risk and preferences for networks from the claims, β_3 should identify "true" inertia (switching costs) rather than preference heterogeneity.

4.3 Cost Stage

Construction of p_{jht} and p^s_{jdt}: In order to complete Equation 10 above and to estimate insurer marginal costs for counterfactual menus, I construct a measure for the base reimbursement

[21] See Prager (2016) and Ho and Lee (2017b) for a discussion of this identification strategy.

price between insurers and providers. 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. I use observed "allowed amounts" to to specify a base rate for each insurer-provider combination.[22] Details are presented in Appendix C.

Estimating Unobserved Marginal Costs: To estimate $c^u_{Ijt}(N_{jt})$, I rely on standard inversion of the first-order condition specified in Equation 13. In traditional product markets, there are JT equations and JT unknowns, allowing for recovery of all necessary cost parameters. In health insurance markets, however, marginal costs do not merely vary by product, but also by consumer risk type. As a result, in my context, there are only JT equations but JTI unknowns, where I is household type. While the marginal costs for care from hospitals, PCPs, cardiologists, and orthopedists are observed in the my claims data, to recover unobserved marginal costs, I parameterize costs as $c^u_{Ijt}(N_{jt}) = c^u_{jt}(N_{jt})\theta^c_I$, where θ^c_I reflects a parameter that scales base plan-specific unobserved costs, $c_{jt}(N_{jt})$, across household type I. For simplicity, I assume that unobserved marginal costs only vary by whether the household is an individual or family, where similar to observed premium scaling in the GIC, I let $\theta^c_I = \theta^R_I = 2.4$ if the household is a family.[23] This reduces the number of unknowns to JT, allowing for full recovery of the base marginal costs, $c^u_{jt}(N_{jt})$.

To predict counterfactual $c^u_{jt}(N_{jt})$ with different networks of hospitals and physicians, I regress the recovered costs on a series of cost-shifters (and adding MCO subscript m back) such that:

$$c^u_{mjt}(N_{mjt}) = \kappa x_{mjt} + \gamma_m + \gamma_t + \varepsilon_{mjt} \tag{16}$$

In my estimation, these shifters include insurer fixed effects, year fixed effects, and an indicator, x_{mjt}, for whether or not the plan is a narrow-network plan.

4.4 Estimating the Employer Fixed Costs and Weight on Consumer Preferences

Estimation: The final task remains to estimate the fixed costs associated with offering additional plans, FC_j, and the employer weight on consumer preferences, ρ. To estimate these parameters, I closely follow work by Ho (2009), Pakes et al. (2015), and Pakes (2010) in constructing moment inequalities to bound the estimates of ρ and FC_j, rather than imposing an equilibrium through distributional assumptions on the parameters.[24] Such moment inequalities approaches were subsequently used to estimate fixed and sunk costs of product introductions in markets such as computers, pharmaceuticals, and smartphones (Eizenberg, 2014; Nosko, 2014; Mohapatra and Chatterjee, 2015; Fan and Yang, 2016).

The critical identifying assumption underlying the moment inequality approach is that the GIC's expected surplus in offering a particular set of plans with particular networks is greater than any

[22]Similar approaches have been taken by Gowrisankaran et al. (2015), Ho and Lee (2017b), and others.

[23]The critical assumption here is that all marginal costs that vary by risk type (e.g. age, gender, etc.) are captured through *observed* hospital and physician costs. While strong, this seems reasonable as a first-order approximation.

[24]An alternate estimator would be to specify a multinomial logit model, similar to the provider and plan demand models, with a logit error shock. However, this is not well-suited for the context of GIC network decisions. In particular, the choice set of possible plans to offer, hospital networks of those plans, and physician networks of those plans is so large that assigning a logit shock to each *potential* choice is likely to produce unreliable, biased estimates.

alternate set of plans and networks it could have chosen at a particular time. Let the expectation of the GIC from offering a particular plan menu, δ_{Jt}, conditional on information set, $\mathcal{J}$ be given by:

$$E\left[W_t(\delta_{Jt},\theta)|\mathcal{J}\right] = E\left[S_t(\delta_{Jt},\theta|\mathcal{J}) - \sum_j FC_j\right] \tag{17}$$

where S_t is the marginal social surplus at time t as defined in Equation 14. Let δ^a_{Jt} be alternate plan menu offered in time t and $E\left[\Delta W_t(\delta_{Jt},\delta^a_{Jt},\theta)|\mathcal{J}\right]$ be the expected change in surplus of the GIC from offering δ_{Jt} relative to δ^a_{Jt}. Then, to satisfy the identifying assumptions, it must follow that:

$$E\left[\Delta W_t(\delta_{Jt},\delta^a_{Jt},\theta)|\mathcal{J}\right] = E\left[W_t(\delta_{Jt},\theta)|\mathcal{J}\right] - E\left[W_t(\delta^a_{Jt},\theta)|\mathcal{J}\right] \geq 0 \tag{18}$$

Let $v_{1,\delta_{Jt}}$ be the difference between the GIC's realized surplus and expected surplus such that:

$$v_{1,\delta_{Jt}} = W_t(\delta_{Jt},\theta) - E\left[W_t(\delta_{Jt},\theta)|\mathcal{J}\right] \tag{19}$$

It follows that:

$$E\left[\Delta W_t(\delta_{Jt},\delta^a_{Jt},\theta)|\mathcal{J}\right] = W_t(\delta_{Jt},\theta) - W_t(\delta^a_{Jt},\theta)\underbrace{-v_{1,\delta_{Jt}} + v_{1,\delta^a_{Jt}}}_{v_{1,\delta_{Jt},\delta^a_{Jt}}} \geq 0 \tag{20}$$

Assuming that $E[v_{1,\delta^k_{Jt}}|\mathcal{J}] = 0 \ \forall \ k$, considering an instrument set $z \in \mathcal{J}$, and taking sample averages, this becomes:

$$\frac{1}{T}\sum_t \left[(W_t(\delta_{Jt},\theta) - W_t(\delta^a_{Jt},\theta)) \otimes g(z)\right] \geq 0 \tag{21}$$

where $g(z)$ is any positive function of instruments z.

I search for any values of ρ and FC_j that satisfy Equation 21. If no values satisfy all the inequalities, I find the values that minimize the squared deviations for all inequalities which were violated. More specifically, let:

$$\begin{aligned} Z_t &= -\Delta W_t(\delta_{Jt},\delta^a_{Jt},\theta) && \text{if } \Delta W_t(\delta_{Jt},\delta^a_{Jt},\theta) < 0 \\ Z_t &= 0 && \text{if } \Delta W_t(\delta_{Jt},\delta^a_{Jt},\theta) \geq 0 \end{aligned}$$

I then estimate the equivalent of a GMM model where:

$$\underset{\rho,FC_j}{min}\frac{1}{T}\sum_t Z_t * Z_t \tag{22}$$

One concern with the estimation procedure outlined above is that since I allow the GIC to add and remove plans from the market, some of the surplus estimates, $CS(\delta_{Jt},\theta)$, may be driven by the logit error shock, which may overestimates surplus gains or losses from product entry or exit.

For robustness, I also report estimates in Appendix H in which I calculate $CS(\delta_{Jt}, \theta)$, ρ, and FC_j assuming that the logit shock is zero.

Error Assumptions: I make several assumptions to proceed with the estimation of ρ and FC_j. First, I assume that the only disturbances to the expected surplus, $v_{1,\delta_{Jt}}$, are composed of two sources: $v^a_{1,\delta_{Jt}}$ and $v^b_{1,\delta_{Jt}}$. The former refers specifically to uncertainty about which municipalities will enter the GIC in the coming year and general uncertainty in demand for different networks. The latter refers to all other uncertainty in demand, including measurement error. Both disturbances are unknown to the GIC and the econometrician. I assume that $E[v^b_{1,\delta_{Jt}}] = 0$.

Rather than relying on instruments within the GIC's information set, I instead use observed data on municipal entrants by year to specify a distribution of household entrants over which the GIC has an expectation. I make a timing assumption that the GIC knows the number of municipalities that entered in the previous year and assumes the same number of municipalities enter the subsequent year, but does not known *which* municipalities, and therefore does not know the underlying risk and preferences (or location) of the households entering in any given year.[25,26] More formally:

$$E[v^a_{1,\delta_{Jt}}] = v^a_{1,\delta_{Jt-1}} + \omega_t$$

where $v^a_{1,\delta_{Jt-1}}$ is the realized disturbance from period $t-1$ and ω_t is a shock to the risk profile and location of entrants in year t. I assume $E[\omega_t] = 0$, or that the shocks to household risk in a given year, conditional on observing entrants in the prior year, are zero.

Translating to expectations, this implies:

$$v^a_{1,\delta_{Jt-1}} + \underset{K\to\infty}{plim} \frac{1}{K} \sum_{j}^{K} \omega_k = 0$$

In the estimation of Equation 21, I take the average of 100 disturbances of ω_k. That is, I estimate the moment inequalities assuming 100 different potential random sample of entrants in each year given the number of municipalities who entered the previous year.

The second assumption is that there is no presence of a structural error component that the GIC knows when making decisions, but the econometrician does not. Such structural errors would normally appear in the fixed cost term, FC_j, appearing as a potential disturbance such that, for instance,: $FC_j = FC + v_{2,j}$, where v_2 represents the structural shock to fixed costs. Eizenberg (2014) and Mohapatra and Chatterjee (2015) 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. In this setting, the GIC might choose to contract with certain insurers, offer certain products, or offer certain networks for which the fixed costs of doing so are lower. Without addition assumptions,

[25] Indeed, between 2009 and 2013, municipalities chose to enter the GIC during many different time-periods within a given year, leaving the GIC little room to incorporate those entrances into its menu decisions. As an example, if a municipality enters in April, it would be unreasonable to assume that the GIC could then reoptimize its product offerings to begin the following fiscal year in July.

[26] It would be more sophisticated to fully specify a model in which the GIC competes for municipal business as a function of the networks and products offered. This model is outside the scope of this paper. However, future work will consider this issue more explicitly.

this structural error would bias my estimates of both ρ and FC_j.

I circumvent this selection problem by assuming there is no structural error term and, namely, that the fixed costs do not vary by where the plan is in the quality space, i.e. $FC_j = FC$. 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. This is a single-agent problem, and 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.[27]

Alternate Estimators: I construct two additional estimators for comparison, setting $\rho = 1$ and constructing bounds on FC_j using merely exclusively one-step deviations in the *number* of products offered. Details are described in Appendix C.

Restricting the Potential Choice Set: The combination of number of products offered and networks of those products, given the number of hospital and physicians in Massachusetts, are nearly infinite, thereby making estimation largely infeasible. I therefore make several restrictions on the choice set of the GIC for estimation. First, I assume that the GIC cannot cease contracting with any insurer, but can adjust the number of plans offered by any insurer.[28] Second, because the smaller insurers typically offer fixed, non-adjustable designs, I assume the GIC can alter only the *number* of plans offered by Fallon, but cannot alter the networks of its plans. Similarly, the GIC must in each year offer the sole plan by Health New England and Neighborhood Health Plan[29] Third, I assume that the GIC may freely adjust both the number of plans offered and the networks of Harvard Pilgrim and Tufts plans, but must limit the number of plans offered by either to four.[30] Finally, I assume that the GIC can offer Harvard and Tufts networks equivalent to the Fallon Direct network (hereafter "Very Narrow"), the Tufts Spirit network ("Narrow"), the Harvard Focus network ("Narrow Small Group," sold primarily to employers outside the GIC), the Harvard Primary Choice network ("Medium"), and the Harvard Independence network ("Broad").[31] Therefore, the

[27] Similar assumptions were made by Nosko (2014). 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 for 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.

[28] The GIC engages in long-term, five-year contracts with insurers. During my sample period of 2009-2013, insurers were under their contract. Therefore, an assumption that the GIC could simply cease to offer any particular insurer would add a choice to the set that was not there in reality, thereby biasing my estimates of fixed costs and weight on consumer surplus.

[29] This is motivated by the fact that Fallon, Health New England, and Neighborhood Health Plan are all fully-insured products that also operate largely in the broader employer marketplace. Unlike Harvard Pilgrim and Tufts, which are self-insured products and offer GIC-specific network designs, the smaller insurers typically offer fixed, non-adjustable designs.

[30] Employers rarely offer more than two narrow-network designs from the same insurer.

[31] I restricted to these particular networks as (1) they were all observed to be offered in Massachusetts during my

GIC can offer most combinations of permutations of 14 potential products.

Identification: Identification of ρ comes from variation in the characteristics of the potential networks not chosen relative to the ones that were *conditional on the GIC offering the same number of plans.* Intuitively, suppose the GIC could broaden the network of one of the existing plans such that consumer surplus in Equation 14, $CS(\delta_{Jt}, \theta)$, increased. The fact that the GIC did not choose to offer this potential network would imply that its weight on consumer surplus was low relative to the added expenditures broadening that network would bring, thus dampening ρ. Conversely, if narrowing an existing network reduced $CS(\delta_{Jt}, \theta)$ while lowering spending, the fact that the GIC did not do this would raise the value of ρ. Since this holds the number of plans constant, FC_j is not affected by these scenarios.

Identification of FC_j relies on the assumption that $FC_j = FC$; over-time variation in the number of products offered; and finally the variation in the potential surplus that could be achieved from offering additional products or reducing the number of products within a time period. Within period, if the GIC could offer an additional product, but did not, then the fixed cost of offering it must outweigh the surplus gained from its introduction. If the GIC could have removed a product, but did not, it must be that the fixed costs are lower than the surplus gained from keeping that product. Over time, fixed costs are pinpointed by changes in the market driven by changes in the underlying provider costs, p_{jth} and p^s_{jtd}; changes in provider ownership structure (which ultimately change demand for providers); changes in the number of entering municipalities; and the risk profile of entering municipal households.

5 Model Results

5.1 Demand

Hospital and physician demand results are reported in Appendix D. To capture physician inertia, I include three separate indicators: whether a patient had sought care from a particular physician practice previously; whether a patient had sought care from any of the practice's locations previously; and whether a patient had previously sought care from any provider employed by the hospital or health system that owns the particular practice. All three of these inertia measures are highly important to predicting physician choice across all specialty groups, with having used the particular physician in the past being the biggest predictor and having used a provider owned by the same health system being the smallest. The estimates imply that a 35-year-old individual in average health living in Boston would be on average willing to travel an additional 11.3 miles to access the same PCP practice, 14.4 miles to access the same cardiology practices, and 20.4 miles to access the same orthopedic practices.[32]

Table 3 reports the results for the insurance plan demand model. Due to the high dimensionality of the data,[33] I only run the model on a subset of 5,000 households across the five years of data.

sample period and (2) they span a considerable range of network breadth, both in terms of hospitals and physicians.

[32] These estimates are, however, different in rating regions which are less dense and have fewer physicians to choose from. See Appendix D.

[33] There are approximately 200,000 GIC members per year multiplied by about 70 hospitals, 18 potential diagnoses, 50 practices in seven rating regions, and three different specialties groups.

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.

The first three columns all present estimates without accounting for plan switching costs, while columns 4-6 present estimates that account for switching costs by estimating a parameter on an indicator for whether a household was previously enrolled in a particular plan. Within each set, the first column (i.e. columns 1 and 4) present results only focusing on hospital utility (EU^H_{Ij}), 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^S_{Ij}), 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).

Panel A reports the estimated parameters. The monthly premium parameter, α_I, is negative and significant across all six specifications, suggesting that households are averse to paying higher premiums for health insurance. The coefficients on expected utility are also positive and significant across all of the models, with exception of EU^H_{Ijt} in model six, which loses significance. Overall, the results indicate that households have a positive valuation of plan networks, consistent with prior literature.

There is, however, significant heterogeneity in network preferences across households and provider types. Looking at the models with no plan switching costs (models 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, particularly premium elasticities. Moving from a model with only hospital networks to a model with physician networks included nearly triples the estimated premium disutility, while significantly reducing the estimated coefficient on hospital utility.

The effect of including physician networks can be more clearly seen in panel B, which translates the estimated parameters to dollarized "willingness-to-pay" (WTP) amounts for networks. These estimates report what single-member (individual) households on Harvard Broad would need to be paid to have their network reduced to that of Harvard narrow. In a model with only hospital networks, individuals would need to be paid approximately \$270 per month to have their network reduced. Note that this figure is considerably higher than the actual premium differential between Harvard Broad and Harvard Narrow, which averaged approximately \$30 across the five-year period. Moving to a model that includes physician networks, however, drastically reduces the implied WTP for hospital networks to merely \$33 per month, while yielding an implied WTP for physician networks of \$88 per month. Decomposing this valuation of physician networks, approximately \$30 per month comes from WTP for PCPs, \$20 per month from WTP for cardiologists, and \$30 from WTP for orthopedists. In all, approximately \$44 (50%) of the valuation of physician networks comes from retaining access to previously used physicians.

Turning to models with plan switching costs included (models 4-6) again significantly increases the magnitudes of the premium elasticities, with further increases seen when including physician networks in addition to hospital networks. When only hospital networks are included in the model, the estimates imply that individuals would need to be paid an average of \$19 per month to move from Harvard broad to Harvard narrow. The estimated switching cost in this model is \$308 per

Table 3: Results of Plan Demand Models

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Estimated Parameters					
	No Switching Costs			Switching Costs		
Prem (PM)	-0.0017***	-0.0042***	-0.0046***	-0.0159***	-0.0185***	-0.0192***
	(0.0008)	(0.0008)	(0.0008)	(0.0017)	(0.0017)	(0.0017)
EU_{Ijt}^{H}	7.4645***	2.2465***	1.5413***	4.8361***	1.4852**	0.6029
	(0.6594)	(0.5410)	(0.5116)	(0.8808)	(0.8538)	(0.8078)
EU_{Ijt}^{S}		0.3281***			0.1929***	
		(0.0134)			(0.0203)	
EU_{Ijt}^{PCP}			0.1598***			0.0885***
			(0.0223)			(0.0275)
EU_{Ijt}^{CAR}			0.5060***			0.5529***
			(0.0973)			(0.1371)
EU_{Ijt}^{ORS}			1.4711***			0.6232***
			(0.1578)			(0.2141)
Prior Plan				4.8914***	4.8520***	4.8480***
				(0.0940)	(0.0969)	(0.0965)
Plan FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	36,830	36,830	36,830	36,830	36,830	36,830
Pseudo R^2	0.30	0.33	0.34	0.79	0.80	0.80
	Panel B: Willingness-to-Pay for Harvard Broad v. Harvard Narrow					
	No Switching Costs			Switching Costs		
WTP Hosp	$270	$33	$21	$19	$5	$2
WTP Phys		$88			$12	
WTP Phys (No Inertia)		$44			$6	
WTP PCP			$29			$4
WTP CAR			$23			$6
WTP ORS			$28			$3
Switching Cost				$308	$263	$251

Notes: Columns 1-3 are results for models without the plan inertia coefficient. Columns 4-6 include these coefficients. EU_{Ijt}^{H} refers to the household's expected utility for the hospital network, EU_{Ijt}^{S} refers to the total utility of the physician network, EU_{Ijt}^{PCP} refers to the utility of the primary care network, EU_{Ijt}^{CAR} refers to the utility of the cardiology network, and EU_{Ijt}^{ORS} 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). "WTP (No Inertia)" refers to willingness-to-pay for the physician network *without* inertia to previously used physicians. "Switching cost" refers to the estimated dollarized plan switching cost. The premium variable is reported in monthly terms. Omitted are interactions between age and premium.

month. When physicians are included in the model, the WTP for hospital utility decreases to a mere \$5 per month, while the estimated average WTP for the physician network differential is \$12 per month. These estimates come much closer to approximating the observed premium differential between Harvard Broad and Harvard Narrow. Decomposing this further, approximately \$4 per month comes from WTP for primary care physicians, \$6 per month comes from WTP for cardiologists, and \$3 per month comes from WTP for orthopedists. Similar to the model with no plan switching costs, approximately 50% of the physician valuation comes from losing access to previously used doctors. Including physicians in the model decreases the estimated switching cost declines by about \$60 to \$251 per month.[34]

The WTP estimates in Table 3 are averages and only reported for the network differential of Harvard Broad and Harvard Narrow. To show the heterogeneity in WTP across consumers and different network types, I plot the distributions of WTP across households for Harvard Broad versus Harvard Narrow and for Fallon Broad versus Fallon Narrow derived from model 6 (the preferred specification). These distributions are reported in Figure 6. Three conclusions emerge from this figure. First, although the mean reported values for Harvard reported in Table 3 were around \$15 per month, there is clearly significant heterogeneity, with certain households willing to pay nearly \$100 per month for access to the broader network. Second, overall the WTP for Fallon's broad versus narrow network is larger than Harvard. This makes sense given that the difference in the networks is more substantial between Fallon plans. Finally, households are on willing to pay significantly more for networks overall than networks when physician inertia is removed, consistent with the averages previously reported.

Figure 6: Willingness-to-Pay for Broad Versus Narrow Networks

Notes: This figure plots the distribution of willingness-to-pay across households for various networks. Panel (a) reports willingness-to-pay for Harvard Broad versus Harvard Narrow. Panel (b) reports willingness-to-pay for Fallon Broad versus Fallon Narrow. Estimates are in per-household-per-month dollars.

These results reveal several important findings. First, ignoring physician networks in models of insurance demand leaves out important consumer heterogeneity, which yield premium elasticities

[34]Though high, these estimates are in range of prior work. In particular, Polyakova (2016) finds switching frictions in Medicare part D to be about twice to four times as large as premiums. A switching friction of \$250 per month is approximately 1.67 times the individual premium for a broad network and approximately 70% of the family premium for the same network.

that are likely underestimated. This has the effect of making households on employer markets appear less price sensitive than they are. In fact, the results suggest that households tend to select into broad-network plans not necessarily because they are price-insensitive, but because they fairly high valuations for physician networks, particularly those that include physicians with whom they formed relationships. This conforms with patterns shown in subsection 2.3. In addition, ignoring the role of physician networks drives up estimates of willingness-to-pay for hospitals and drives up estimated plan switching costs.

The second important finding is that, even conditional on heterogeneity in preferences for hospital and physician networks, plan switching costs are substantial. The results imply that even if a consumer were offered an identical network for a lower premium, they may still be unlikely to switch to that network.

5.2 Supply

Estimates of unobserved marginal costs are reported in Appendix F. I next proceed to reporting the estimates of firm (employer/GIC and insurer) fixed costs for introducing additional products to the market as well as the weight the GIC places on consumer preferences relative to net payments to insurers. These are presented in Table 4. Panel A reports the main specification: estimates of both ρ and FC_j from the full moment inequality approach. Panel B reports the approach that sets $\rho = 1$ and uses one-step deviations in plans to estimate bounds on FC_j. Panel C does the same as Panel B, but assuming that insurers make the product decisions and using their profits as objective functions.

The estimates from Panel A are presented as point estimates rather than bounds, as no set of parameters, ρ and FC_j, satisfied each of the inequalities presented in Equation 21.[35] As such, the values reported in Panel A are the same in the "lower bound FC_j" column and the "upper bound FC_j" column. The estimate of ρ is 4.35, suggesting that the GIC places considerable weight on consumer surplus relative to net spending.[36] The point estimate of fixed costs is \$8.40 million for each plan. Though this estimate appears quite high, it is actually a very small fraction of the GIC's overall net spending.[37] Hereafter, these will be the main estimates I employ in my counterfactuals.

Panel B, using merely one-step deviations in plans and setting $\rho = 1$, suggests that the GIC spends between \$1.15 and \$6.64 million a year on fixed costs for each plan offered. The fixed cost bounds for insurers (Panel C) are slightly tighter than those for the GIC, but within the same general range. The estimates range from approximately \$2 to \$3 million per product offered. Although

[35]This is to be expected, given the large number of inequalities (the large number of potential plans and networks the GIC could offer in a given year).

[36]This is, in particular, a larger estimate relative to prior work in Gowrisankaran et al. (2015), which estimates a value of 2.79 for the employer weight on consumer surplus. This may be partly explained by the fact that their employers were private. The GIC, in contrast, is not a traditional employer, but rather a government insurance purchasing organization. This implies it likely needs to cater more explicitly to consumer preferences, particularly since it also competes for municipal business. Further, most municipal health insurance plans tend to be quite generous.

[37]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 represent about 0.87% of net spending. Similarly, these estimates represent about 1.90% 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 4: Results of Employer Objective Function Estimation

	ρ	Lower Bound FC_j	Upper Bound FC_j
Panel A: Estimating ρ and FC_j			
GIC/Employer ($Millions)	4.35	8.40	8.40
Percentage of Net Spending		0.87	0.87
Percentage of Net Surplus		1.90	1.90
Panel B: One-Step Deviations, GIC			
GIC/Employer ($Millions)	1	1.15	6.64
Percentage of Net Spending	1	0.12	0.70
Percentage of Net Surplus	1	0.26	1.50
Panel C: One-Step Deviations, Insurer			
Insurer ($Millions)		1.99	3.00
Percentage of Fallon Profits		1.97	2.97
Percentage of Harvard Profits		1.23	1.86
Percentage of HNE Profits		4.73	7.14
Percentage of NHP Profits		2.07	3.12
Percentage of Tufts Profits		0.83	1.26

Results from ρ and FC_j estimation for 2009-2013. Panel A reports the main specification, estimating both parameters using moment inequalities. ρ and FC_j are point estimates in this panel, therefore the "lower bound" and "upper bound" on FC_j are identical. Panel B reports the results only estimating FC_j using one-step deviations in plans. Panel C reports estimates of FC_j using one-step deviations and assuming insurers make the decisions on which products and networks to offer.The corresponding percentages of fixed costs relative to net GIC health spending, net GIC surplus (consumer surplus minus health spending), and relative to insurer profits are also reported. FC_j reported in millions of dollars.

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.[38]

Three caveats should be noted regarding these estimates, particularly in Panels B and C. 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 very small sample size. Second, and related, the low rate of offered products in the GIC may 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. Combined these two issues produce pretty wide bounds, particularly in Panel B. 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. Nevertheless, I continue

[38]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 (Murray, 2010).

by using estimates from the much more precise and larger sample approach in Panel A.[39]

The third caveat is that some of the estimates of $CS(\delta_{Jt}, \theta)$ may be driven by the adding and removal of logit error shocks. In Appendix H, I report estimates of ρ and FC_j assuming the shock is zero. Indeed, doing so significantly reduces the estimates of FC_j to approximately \$1.6 million per plan, which is more in line with reported estimates. It also further increases the weight that the GIC places on consumer surplus. I report counterfactual estimates in the appendix using these alternate estimates of ρ and FC_j. Though this does change the equilibrium menus, the qualitative results remain similar.

6 Equilibrium Plan Menus

The three primary goals of this paper are to: (1) understand the drivers of continued prevalence of broad-network products in the employer market, (2) see the role that both consumer inertia and employer fixed costs play in explaining product variety, and (3) assess the welfare implications of alternate plan menus. Motivated by these goals, I first consider a series of counterfactuals where I remove various sources of choice frictions:

1. **Baseline:** I assess the model fit by predicting the number of products and networks offered on the GIC using the estimated model parameters. For all subsequent counterfactuals, I compare consumer surplus and spending estimates to these baselines.

2. **Removing Physician Inertia:** I remove from the physician demand models consumer loyalty to their previously used physicians, practices, and health systems for primary care, cardiology, and orthopedics.[40]

3. **Removing Plan Switching Costs:** I remove all health plan inertia from the insurance demand model.

4. **Removing Employer Fixed Costs:** I set $FC_j = 0$, thereby making offering additional plans costless for firms.

5. **Price-versus-Provider-Choice:** In this counterfactual, I assume that both physician inertia and plan switching costs *do* exist (so that consumers choose their plans internalizing these frictions), but that the GIC, in making its product decisions, does not care about the plan switching frictions. Rather, the GIC only considers as welfare-relevant the premium-versus-provider-choice tradeoff.

[39] 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.

[40] Note that for these counterfactuals, I report consumer surplus estimates assuming that the full source of this inertia is due to unobserved preference heterogeneity, rather than true switching costs. These should be taken as conservative estimates. The goal of this exercise is to show how much of the existence of broad networks can be explained by simple physician loyalty as opposed to option value of larger networks, and the potential spending impacts of removing this loyalty. If some of this loyalty is due to switching costs, then these should be taken as overestimates of consumer surplus changes.

6. **Fixed-Dollar Defined Contribution:** In this scenario, I alter the pricing subsidy so that, rather than paying 75% of the full premium, the GIC subsidies 90% of the lowest-price available plan in any zip code.

The simulation procedure used to evaluate these counterfactuals is described in more detail in Appendix G.

For each counterfactual, I report two different measures of consumer surplus changes:

$$\Delta CS_1(\delta_{Jt}, \delta_{Jt}^a, \theta) = \sum_I \frac{1}{\alpha_I} ln\left(\sum_j^J exp(\delta_{Ijt}^{1,a})\right) - \sum_I \frac{1}{\alpha_I} ln\left(\sum_j^J exp(\delta_{Ijt}^{1})\right) \tag{23}$$

$$\Delta CS_2(\delta_{Jt}, \delta_{Jt}^a, \theta) = \sum_I \frac{1}{\alpha_I} ln\left(\sum_j^J exp(\delta_{Ijt}^{2,a})\right) - \sum_I \frac{1}{\alpha_I} ln\left(\sum_j^J exp(\delta_{Ijt}^{2})\right) \tag{24}$$

where $\delta_{Ijt}^1 = \delta_{Ijt}$ from Equation 3.3, $\delta_{Ijt}^{1,a}$ is the counterfactual plan menu and:

$$\delta_{Ijt}^2 = -r_{Ijt}\alpha_I + EU_{Ijt}^H\beta_1 + \sum_s EU_{Ijt}^s\beta_2^s + \eta_j$$

ΔCS_1 is therefore the change of consumer surplus where plan inertia is treated as a tangible, relevant switching cost. If switching costs are removed and are considered welfare-relevant, consumers may benefit from the reduction in these tangible costs. ΔCS_2, conversely, is the change in consumer surplus if the plan switching costs are considered welfare-irrelevant, i.e. they represent intangible costs that may not represent a clear benefit should they be removed, apart from the potential benefit of consumer reallocation to more optimal plans.[41]

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 5 reports the equilibrium predicted products/networks offered under the counterfactuals outlined above. The columns report the observed plan menus, predicted plan menus, and then predicted plan menus when inertia at various stages are removed. For all scenarios, I report changes in both consumer surplus measures and change in overall health spending assuming that the plan menu remained fixed, but that insurers were allowed to reprice and consumers were allowed to reallocate across plans. The consumer surplus change, ΔCS (Fixed), then comes entirely from the removal of the switching friction/inertia and the new prices. I then report these same changes, but allowing the GIC to reoptimize their product menu. These consumer surplus changes, ΔCS

[41] Note that in all the estimates presented, physician inertia is not only considered welfare-relevant, but is treated as unobserved preference heterogeneity, rather than a switching cost. This is to present the most conservative estimates of their removal: if physician inertia were removed, consumers would be losing some measure of loyalty that is considered a true preference.

Table 5: Counterfactuals: Equilibrium Networks Chosen Under Varying Inertia Assumptions, 2013

Insurer	Network	Observed	Pred.	No Phys. Inert.	No Plan Inert.	No FC	Price-v-Prov.	Defined 10%
Fallon	VN	x	x	x	x	x	x	
Fallon	Broad	x	x	x		x	x	x
HPHC	VN							x
HPHC	N				x	x	x	x
HPHC	N(SG)		x	x	x	x	x	
HPHC	Med	x	x	x	x	x	x	x
HPHC	Broad	x	x	x	x	x	x	
HNE	N	x	x	x	x	x	x	x
NHP	N	x	x	x	x	x	x	x
Tufts	VN							
Tufts	N	x			x	x	x	
Tufts	N(SG)				x	x	x	
Tufts	Med		x	x	x	x	x	x
Tufts	Broad	x	x	x	x	x	x	
Total Plans		8	9	9	11	12	12	7
Welfare and Spending Holding Plan Menu Fixed								
ΔCS_1 (Fixed)				-\$92.42	\$54.28	-	-	-\$20.23
ΔCS_2 (Fixed)				-\$90.85	-\$2.92	-	-	\$29.20
$\Delta Costs$ (Fixed)				-\$2.06	-\$12.97	-	-	-\$26.83
ΔFC (Fixed)				-	-	-\$60.50	-	-
Welfare and Spending Allowing Plan Menu to Change								
ΔCS_1 (Change)				-\$92.42	\$59.28	\$1.87	\$1.87	\$19.88
ΔCS_2 (Change)				-\$90.85	\$2.08	\$7.86	\$7.86	\$51.99
$\Delta Costs$ (Change)				-\$2.06	-\$21.38	-\$2.08	-\$2.08	\$0.36
ΔFC (Change)				\$0.00	\$13.41	-\$60.50	\$20.17	-\$13.45

Notes: GIC observed and predicted products offered under various counterfactual inertia assumptions. "No Phys. Inertia." refers to predicted networks when all physician loyalty is removed. "No Plan Inert." refers to predicted networks when all plan switching costs are removed. "No FC" refers to predicted networks when employer fixed costs are removed. "Price-v-Prov." refers to predicted networks when plan switching costs are presented, but when the GIC considers them welfare-irrelevant, i.e. considers only the price-versus-provider-choice tradeoff when making product variety decisions. "Defined 10%" refers to a counterfactual policy of imposing a fixed premium defined contribution set to 10% of the lowest priced plan. "ΔCS_1 refers to change in consumer surplus per-household-per-month when plan switching costs are considered tangible welfare-relevant costs. "ΔCS_2" are consumer surplus estimates when switching frictions are considered welfare-irrelevant. $\Delta Costs$ (fixed)" refer to the change in consumer surplus per-household-per-month and change in total GIC costs per-household-per-month assuming the plan menu remains fixed. "$\Delta FC'$" refer to changes in firm fixed costs.

(Change), would therefore reflect the added or lost utility from the new choice set consumers have access to.

The predicted networks match the observed plan menu very well. The model correctly predicts that the GIC offers the exact broad and narrow-network plans from Fallon, Harvard, and Tufts that are, in fact, offered. There are a few differences in the predictions, however. The first is that the model predicts that the GIC offers a total of 9 plans (not counting Unicare products), as opposed to the actual observed 8 plans. In particular, the model predicts that the GIC, in addition to the two observed Harvard plans, also offers an additional narrow-network plan from Harvard (labeled "N(SG)"), equivalent to the Harvard "Focus" network observed on the small-group market. Moreover, the model predicts the narrow-network offered by Tufts to be broader than that observed, equivalent to the Harvard Primary Choice network. Despite these differences, the model clearly performs very well both in predicting nearly the exact number of products the GIC offers, and in predicting close to the exact networks.[42]

Removing Physician Inertia: The first counterfactual considered is the removal of all physician inertia/loyalty. Under the assumption that inertia reflects underlying unobserved persistent preferences, both measures of consumer surplus decrease significantly when plan menus are held fixed, with a decline of approximately $92 per household per month. This reflects entirely the loss of valuation that consumers had in having formed relationships with their previously used physicians. Moreover, the change in predicted GIC costs under this scenario is quite small, declining by only $2 per household per month. This implies, somewhat surprisingly, that while removing physician inertia may steer patients to different providers, it does not necessarily steer them towards substantially *cheaper* providers. Even removing this loyalty, patients still appear to prefer the more expensive providers, and derive considerable value out of having networks of large *size*. As such, allowing the GIC to alter its product menu has no effect, as the potential cost reductions do not outweigh the loss that consumers would face from losing access to those providers. Because physician loyalty does not appear to drive cost increases and is a key factor in the utility consumers derive from their plans, policies aimed at severing these relationships are likely to result in large social surplus declines.[43]

Removing Plan Switching Costs: I next consider the removal of all plan switching costs. Holding the plan menu fixed, the GIC costs decrease by a sizeable amount: approximately $13 per household per month. Overall, this change represents a reduction in approximately $16.2

[42]One concern may be that the networks offered are driven by the GIC's reluctance to adjust a network if they believe the consumer will pay a switching cost even if the number of products remains the same. However, all of the models reported above assume that if the GIC simply changes the network of an existing plan, consumers would be defaulted into it and pay no switching cost. For example, I assume that if the GIC decides to switch from offering "Harvard Broad" and "Harvard Medium" to offering "Harvard Very Narrow" and "Harvard Narrow," this would be equivalent to the GIC changing the *existing* networks rather than removing two products and adding two new ones. In this instance, consumers would be defaulted into the network closest in terms of provider size.

[43]Indeed, this rests on a heavy assumption that *all* of the physician inertia estimates are due to preference heterogeneity and none due to switching costs. However, the difference in consumer surplus and spending is so substantial that physician switching costs would need to play an extremely large role to offset the decline in surplus. A back of the envelope calculation reveals that, at minimum, physician switching costs would need to be about 40% of total physician inertia for social surplus to increase as a result of the removal of these frictions.

million in annual costs. In addition, there is considerable heterogeneity in who experiences the cost declines. Table 6 reports the changes in market shares and premiums when switching costs are removed, holding baseline plan menus fixed. Among the 14% of individuals who switch from a broad to narrow plan following removal of switching costs, these individuals face sizable premium reductions of up to $30 per household per month ($72 for multi-member households). For the non-switchers, those remaining in Tufts Broad see some premium increases, suggesting the presence of adverse selection against Tufts plans. However, surprisingly, the non-switchers remaining on Harvard Broad see premium *declines*, suggesting that households who switch away from Harvard Broad towards narrower networks appear to have been higher-cost enrollees. Indeed, non-switchers on the narrow-network plans see some premium increases as they households from Harvard Broad switch into the narrow plans. The overall welfare implications of this combination of adverse selection and advantageous selection is a slight *decline* in consumer surplus when switching costs are considered welfare-irrelevant. These declines in overall welfare are similar to Handel (2013), though the magnitudes are much smaller. In fact, when switching costs are considered tangible, welfare-relevant costs, consumer surplus increases by a substantial $54 per household per month. Overall the results indicate that, even at baseline plan menus, social surplus significantly increases by removal of plan switching frictions, although consumer surplus may slightly decline if those costs are entirely intangible.

Table 6: Counterfactuals: Shares and Premiums, Plan Menu Fixed, 2013

Insurer	Network	Baseline	Counterfactual	Baseline	Counterfactual
		Market Shares		Individual Premiums	
Fallon	VN	0.04	0.04	$114	$118
Fallon	Broad	0.04	0.04	$144	$151
HPHC	N(SG)	0.06	0.12	$131	$134
HPHC	Med	0.05	0.09	$135	$140
HPHC	Broad	0.28	0.23	$164	$158
HNE	N	0.12	0.11	$112	$112
NHP	N	0.03	0.04	$120	$124
Tufts	Med	0.02	0.05	$127	$131
Tufts	Broad	0.36	0.27	$151	$154

Notes: Market shares and individual monthly premiums for baseline and counterfactual predictions, holding the GIC's product menu fixed.

However, allowing the GIC to reoptimize its plans results in some major changes to the plan menu and additional welfare gains. In particular, the GIC increases in number of products offered from 9 to 11, adding 2 new Tufts narrow-networks, a new Harvard narrow-network, and removing Fallon Broad. The implications are clear: removing switching frictions makes it more worthwhile for the GIC to pay the fixed costs of offering additional network variety, while making it less worthwhile to offer expensive, broad networks with small market share (i.e. Fallon Broad). If these are considered tangible, welfare-relevant costs, then consumer surplus (CS_1) increases now by approximately $59 per month, an additional $5 per household per month relative to the baseline plan menu. Even if these costs are considered welfare-irrelevant (CS_2), however, consumer surplus still increases by $2 per month, suggesting that the gains from the increase in network variety outweigh the loss due to selection. Moreover, costs decrease by an additional $8 per household per month relative to the scenario where menus were fixed, though this is offset by an increase in fixed costs. As a result, regardless of which definition of CS is employed, social surplus increases on net, and consumer welfare increases in any scenario.[44] These results imply that plan inertia is a key driver of current product offerings, and that efforts to reduce consumer switching costs would not only result in consumers self-sorting more optimally, but it would actually induce employers to offer them increased product variety, thereby further enhancing sorting efficiency.

Removing Fixed Costs: Next, I consider the scenario where consumer switching costs are still present, but the *employer* fixed costs of offering additional products are removed. In such a scenario, unsurprisingly, the GIC offers every plan available at its disposal, increasing variety in a similar way as the scenario without consumer frictions. However, this scenario has different welfare implications due to the continued presence of those switching costs which inhibit consumer movement across plans. In particular, despite the fact that offering additional products is now costless, consumers largely do not shift plans as they remain inertial to their previous choices. As a result, the decline in costs is quite modest at about \$2 per household per month. However, consumer surplus increases by between \$2 and \$10 per month depending on whether switching costs are considered welfare-relevant.[45] This implies that efforts to reduce employer fixed costs will only have modest impacts without a corresponding effort to reduce plan switching costs.

Price-v-Provider Choice: I present results of a scenario where switching costs are present, but the GIC considers them welfare-irrelevant *when making product choice decisions*. The goal of this exercise is to understand how plan menus may be affected by employers' *perceptions* of consumer preferences rather than the preferences themselves. Under this scenario, the GIC offers 12 total products, similar to the previous two scenarios, implying that by simply removing the disutility of switching from the GIC's objective function yields substantially more products narrow-network products offered. Even if consumers do not switch and costs barely decline, offering additional product choice becomes worth it to the GIC.[46],[47]

Fixed Dollar Defined Contribution: Finally, I present the results of an alternate price subsidy scheme. In this scenario, the GIC offers a fixed-dollar subsidy equivalent to 90% of the total premium of the lowest-priced plan in any zip code. If a consumer wants to purchase a more

[44] It should be noted that some these results are driven by the presence of the logit error shock. In Appendix H, I re-estimate these results simulating the removal of the logit shock. The results do change, with the GIC opting to instead cease offering most Harvard and Fallon products. However, the network *variety* is preserved, and the welfare implications are largely similar.

[45] It is particular noteworthy CS_2 shows greater gains in surplus than in the scenario with no plan switching costs. Two factors drive this result. First, the "no fixed cost" scenario contains an additional product, Fallon Broad, which is responsible for part of the welfare gains. Second, the selection issues that were present in the "no switching cost" scenario are absent here. Due to the low rate of switching, the gains come entirely from expanded number of products in the choice set. Indeed, in Appendix H, the results when the logit error is dropped suggest that the "no fixed cost" scenario bring virtually no welfare gains whatsoever.

[46] I do not report estimates for the scenario in which plan menus remained fixed, as the only change made in this counterfactual was to the GIC objective function. All consumer welfare and costs change would occur exclusively as a result of plan redesign.

[47] More so than the previous counterfactuals, however, these results change when the logit error is removed. In particular, the GIC reduces access to several broad-network plans, resulting in sizable cost decreases as well as some welfare declines. See Appendix H.

expensive plan, the subsidy still remains fixed and the consumer would therefore need to pay the rest out-of-pocket.

If plan menus remain constant, net costs decrease while consumer surplus increases. This is not surprising, since the fixed-dollar contribution, by definition, minimizes the employer's part of the premium subsidy to be smaller than the baseline. As a result, households who want to purchase broader network plans now face substantially higher premiums. However, those households all now face substantially *lower* premiums for narrow-network products.[48] Consumers therefore see a decline in surplus from the new pricing scheme of about $20 per household per month (if switching costs are considered relevant), but costs decrease by $27 per household per month.

If the GIC is allowed to adjust its plan menu, however, the equilibrium choices change significantly. In particular, the GIC actually *decreases* its total products offer to seven, and only keeps one broad-network plan (from Fallon), while adding additional narrower-network plans from Harvard. In effect, the GIC is selecting a plan menu such that the subsidy becomes generous enough to induce consumers to switch to narrow networks, but not so generous that GIC costs balloon. Specifically, it removes Fallon Narrow, thereby making NHP the lowest-cost option for most consumers, while removing access to the rest of the broad-network products, forcing consumers to switch. Although consumers are harmed by the removal of the broad networks, they are more then compensated by the firm in the form of the higher subsidies. As such, consumer surplus increases dramatically, by between $20 and $50 per household per month. Costs, however, increase by less than $1 per household per month, leading to high social surplus gains.

Summary: In each of the scenarios described above (with exception of removing physician inertia), switching frictions inhibited the employer from offering additional product variety, not merely in terms of number of plans, but in the number of distinct networks offered to consumers. In nearly all of these scenarios, these switching frictions benefited dominant, broad-network products, while driving up premiums for consumers. Removing these frictions, particularly plan switching costs, therefore would result in the employer re-optimizing its menu to increase choice, decrease costs, and increasing overall social surplus.

7 Conclusion

The rollout of the Affordable Care Act has brought a renewed focus on managed competition in health insurance markets. Particularly as new types of insurance innovations emergence (including narrow networks, tiered networks, health savings account, and high-deductible health plans), policymakers and employers have struggled with balancing offering consumers a wide variety of choice options, while keeping premiums low, keeping consumers well-informed, and preventing confusion. With regards to the HIE, states vary dramatically in how many plans they offer consumers and the level of plan standardization. California, for instance, is very active about limiting the choice set to consumers, while aggressively standardizing plans to prevent confusion. New York, on the other hand, contracts far less selectively (Scheffler et al., 2016). As a result, these states have very different experiences in terms of consumer enrollment, premiums, and spending.

[48] The fixed-dollar contribution essentially widens the gap between expensive and lower-cost plans in the menu.

Employers similarly struggle to balance affordability and choice. As companies grow and cater to employees with much more heterogeneous preferences, firms have increasingly turned to offering a choice of plans. So far, most of this choice has been among financial dimensions of health plans: copayments, coinsurance, and deductibles. However, as narrow-network insurance plans grow more popular on the HIE, employers are increasingly, though slowly, considering offering employees a choice on the network dimension as well.

In this paper, I show that moving towards offering employees more variety in terms of narrow-network products has the potential to significantly decrease costs and increase surplus. In particular, I show that, while valuations of hospitals and physicians are important to explain consumer selection into broad network plans, they alone are not sufficient to explain the persistence of broad networks offered among employers, nor are they sufficient to explaining the lack of choice of plans with network variety. Contrary to prior estimates, I find that consumers in employer markets are actually quite price-sensitive. However, high switching costs across plans, combined with high employer fixed costs of offering multiple plans, explain the reluctance of employers to offer these new types of products. As I demonstrate, if these costs were removed, consumers would find more value in enrolling in narrower products, and employers would pay the cost of offering additional network variety. In certain cases, I show that the optimal product menu may be one *without* access to multiple broad-network plans, but with the availability of several narrower-network plans.

This analysis has some limitations. First, I only consider valuations for a limited set of physician specialties. This leaves out some heterogeneity that may drive choice into broad networks that I am not picking up. For instance, consumers may have extremely high valuations of certain high-cost provider types, such as oncologists, that may drive their preferences for health plans, and subsequently employer offerings. Second, while I am able to leverage a natural experiment to separate plan switching costs from unobserved preference heterogeneity, I am not able to separate this for physician inertia. The results of this paper suggest that removing physician loyalty would result in substantial surplus losses. However, if part of this loyalty comes from a true "physician switching cost," rather than preference, the welfare losses may be tempered. In fact, providing consumers with additional information to better help them choose physicians may actually, in the long-run, improve welfare. Third, the model does not consider bargaining effects of altering plan menus. Indeed, if employers decide to place additional emphasis on narrow-network plan designs, this may impact the negotiations between insurance carriers, hospitals, and physicians.

Overall, this paper contributes to our understanding of what consumers in employer markets value in their choice of plan, and how employers aggregate those preferences to design insurance choices for those employees. The policy implications suggest that a first-best option would be to invest in information infrastructure so as to reduce consumer switching frictions. This would allow employers to take advantage of newly designed insurance products by allowing price-sensitive consumers to more easily switch into them. In the absence of this, a second-best option may be to offer significant financial incentives to consumers to induce them to switch towards these plans. In certain cases, a blunt option may be to remove access to certain broad network products whose prices far exceed consumer valuation of the providers, while compensating consumers directly for the switching costs they would incur. A fixed-dollar pricing scheme would achieve this result. Each of these scenarios has the potential to improve sorting and reduce costs.

References

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

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

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

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

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

Dafny, L., Duggan, M., and 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., and 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., and 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).

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

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

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

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

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

Ghili, S. (2017). Network formation and bargaining in vertical markets: The case of narrow networks in health insurance. *Working Paper*.

Gowrisankaran, G., Nevo, A., and Town, R. (2015). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review*, 105(1):172–203.

Gruber, J. and 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. and Fronstin, P. (2016). Narrow provider networks for employer plans. Technical report, Employee Benefit Research Institute.

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

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., and 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. and Lee, R. (2017a). Equilibrium provider networks: Bargaining and exclusion in health care markets. *Working Paper.*

Ho, K. and Lee, R. (2017b). Insurer competition in health care markets. *Econometrica*, 85(2):379–417.

Kleiner, S., White, W. D., and Lyons, S. (2015). Market power and provider consolidation in physician markets. *International Journal of Health Economics and Management*, 15(1):99–126.

Lee, R. (2013). Markov-perfect network formation: An applied framework for bilateral oligopoy and bargaining in buyer-seller networks. *Working Paper*, Forthcoming.

Liebman, E. (2017). Bargaining in markets with exclusion: An analysis of health insurance networks. *Working Paper.*

Liu, C. and Sydnor, J. R. (2018). Dominated options in health-insurance plans. *NBER Working Paper.*

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

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

Mohapatra, D. P. and Chatterjee, C. (2015). Price control and access to drugs: The case of india's malarial market. *Working Paper.*

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

Murray, J. G. (2010). Small group health premiums in massachusetts. Technical report, Commissioner of Insurance.

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

Pakes, A. (2010). Alternative models for moment inequalities. *Econometrica*, 78(6).

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

Polsky, D. and Weiner, J. (2015). The skinny on narrow networks in health insurance marketplace plans. Technical report, 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., and 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., and 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. *NBER Working Paper Series, 22600*.

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.

For Online Publication

A Data Descriptions

A.1 APCD Sample Creation

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 subsection 3.1. 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 organization'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 A.1 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 professional 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, or geriatric doctor. Similarly, cardiology practices and

Table A.1: 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 2009-2013. Diagnosis characteristics (e.g. "Neurological," "Cardiac," etc.) are derived from AHRQ's Clinical Classification Software categoreis and Chronic conditions Indicators indicators.

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 (exemplified by cardiology), and surgical care (exemplified by orthopedics).

For each service-line, I merge in Medicare Part B physician fee schedules from Center for Medicare and Medicaid Services (CMS).[49] These data contain annual federal updates to each procedure (CPT) code's "Relative-Value-Unit" (RVU) weight, which are constructed in order to assign each service an approximate measure capturing its relative intensity to other procedures. These weights are then used to determine Medicare payment rates. Specifically, each year CMS releases updates to its Medicare "conversion' factor" and to its RVUs. The "conversion factor" reflects the base Medicare payments per RVU that it pays to physicians in a given year. This factor is then scaled by the RVU for a particular procedure to determine the physician reimbursement.[50] I 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 and total RVUs performed per visit. I also use these RVUs in construction of insurers' negotiated rates with physician practices, described further in subsection C.3.

[49]https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/PFS-Relative-Value-Files.html.

[50]As an example, if the Medicare Conversion Factor for a given year is $36, a procedure performed in that year with an RVU of 2 will receive a total reimbursement of $72.

Table A.2 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 cardiologists, though patients seeing orthopedists tend to be the youngest on average. 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 57% of primary care patients saw a doctor between 2009 and 2013 that they also have seen previously, while this number was about 63% for cardiologists and about 60% for orthopedists. Distance traveled to any of the specialty groups are all about 10 miles, with the distance to see PCPs somewhat shorter. When seeing a PCP, patients on average visit practices with 26 doctors on site, whereas this number is significantly higher for orthopedic practices and, especially, for cardiology practices. Moreover, patients tend to visit cardiology practices with a greater number of locations and that disproportionately tend to be part of medical groups, owned by hospitals, or owned by health systems.

Table A.2: Physician Sample Summary Statistics

	PCPs	Cardiologists	Orthopedists
Age	47.64	54.09	44.34
	(15.59)	(13.89)	(18.51)
Female	0.56	0.43	0.52
	(0.50)	(0.49)	(0.50)
RVU	2.64	2.95	5.55
	(1.79)	(4.89)	(12.54)
Used Doc Previously	0.57	0.63	0.60
	(0.49)	(0.48)	(0.49)
Used Med Group Previously	0.58	0.70	0.64
	(0.49)	(0.46)	(0.48)
Used System Previously	0.59	0.73	0.66
	(0.49)	(0.44)	(0.47)
Distance	7.21	9.67	9.83
	(9.18)	(11.13)	(10.59)
Doctors on Site	26.44	97.51	49.66
	(75.80)	(159.11)	(109.73)
Number of Locations	6.93	7.79	4.18
	(7.56)	(7.94)	(6.27)
Part of Medical Group	0.60	0.67	0.59
	(0.49)	(0.47)	(0.49)
Owned by Hospital	0.22	0.32	0.16
	(0.22)	(0.46)	(0.37)
Owned by System	0.40	0.45	0.24
	(0.49)	(0.50)	(0.43)

Notes: Physician sample summary statistics for select variables for primary care physicians, cardiologists, and orthopedic surgeons 2009-2013. For practice characteristics (e.g. "doctors on site", "number of locations," etc.), these estimates reflect means and standard deviations weighted by patient visits. In other words, "Doctors on Site" reflects the number of doctors at a particular practice location weighted by patient visits to that practice.

GIC Member Data: 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.

A.2 SK&A Sample Creation

Matching Physicians to Practices: Given the breadth of the data as well as the inconsistencies in reporting between 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-practice-location triple in the data that billed more than 50 claims in any particular year. If a physician was not reported as being employed by a medical group in the SK&A, 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 (e.g. 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.

Constructing Physician Practice Networks: 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.[51] For any large medical group that is not affiliated with a particular hospital or system, 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 use the claims to infer whether the practices are in a particular plan's network. In particular, I assume that any practice that has more than 10 in-network claims from a particular plan is considered in-network. For robustness, I also construct networks that default each each of these small practices as being in-network unless a majority of claims that are processed for these practices by a particular plan is flagged as being "out of network."

[51] In practice, this is a close approximation of contracts observed on the GIC. Harvard Primary Choice and Tufts Spirit, for instance, cease contracting with all Partners-owned medical groups as well as Partners hospitals. The exception appears to be for Fallon Direct, which does contract with certain Partners-affiliated Medical groups (e.g. Charles River Medical Associates) and certain Atrius-affiliated groups (e.g. Reliant Medical Group). Fortunately, Fallon reports these covered groups on its website and, as such, I was able to incorporate them into the network structure.

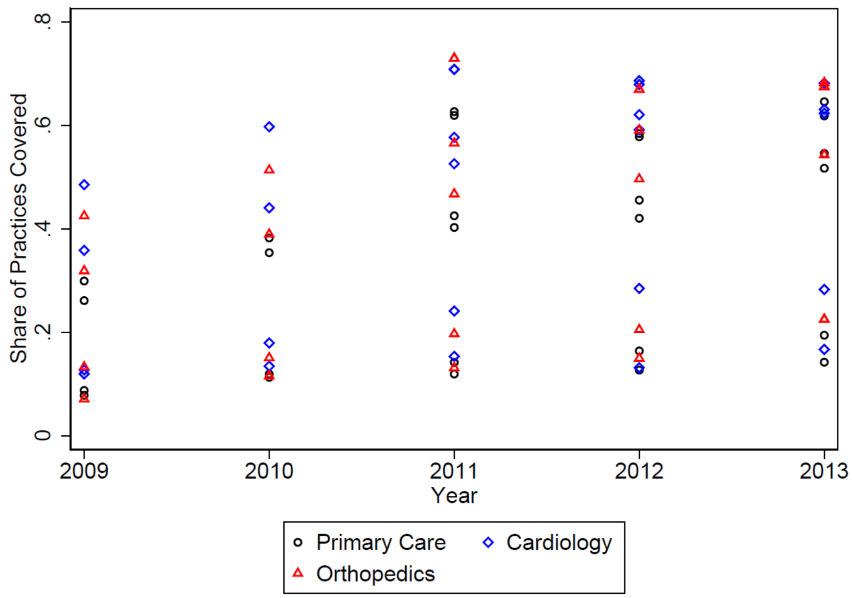
B Additional Network Figures

Given the nature of how these networks were constructed, there may be some measurement error in the network variable. This error should be small for the reasons outlined in Appendix A–namely that many practices were matched to owned hospitals and health systems for which there is observable network data, while the rest were aggregated to large enough group sizes such that inferring from claims ought to be more precise. As robustness, I reproduce Figure 1 reporting only network shares among the top 50 practices of each specialty type in each of the seven rating regions in Massachusetts. The results show that the trends in networks–both over time and across plans–remains very similar to the trends observed the figures with *all* practices. I also produce the variation in network shares across rating region.[52]

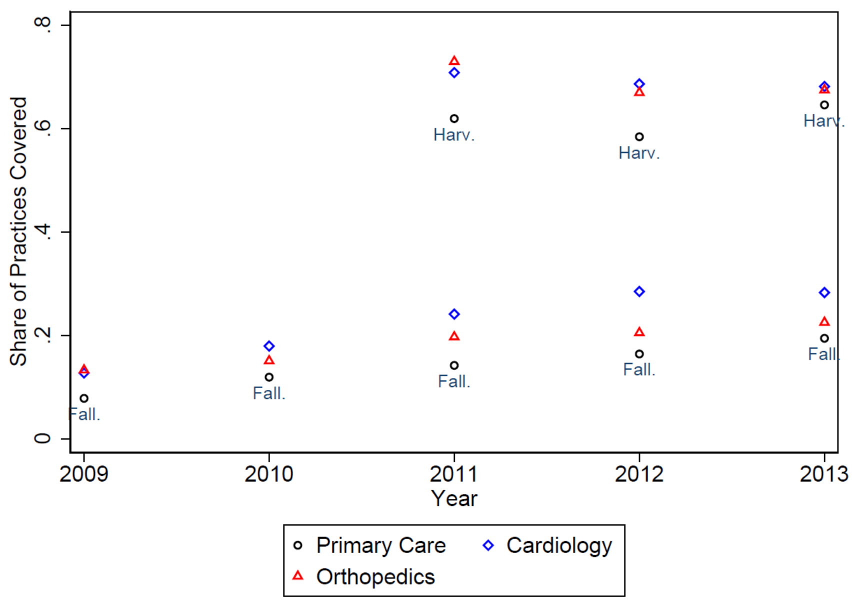
[52] Rating region 7 is omitted due to small sample sizes. It should be noted that there is considerable variation in network coverage within plans across regions. For instance, Fallon Direct has fairly high coverage in rating region 2 (containing Worcester, where Fallon is quite prominent), but fairly low coverage in rating region 5 (containing Boston).

Figure B.1: Share of Practices Covered by Year and Specialty

(a) Top 50 Practices, All Plans



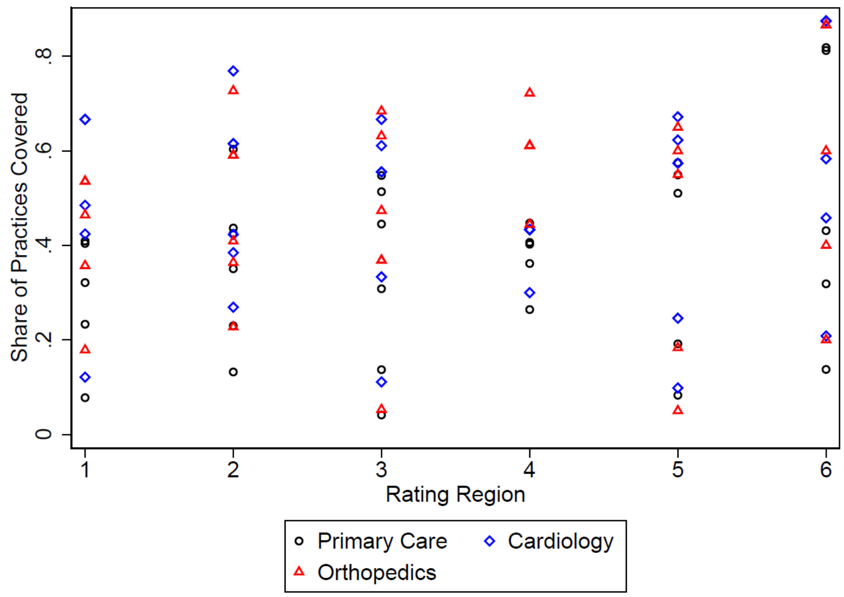
(b) Top 50 Practices, Fallon and Harvard



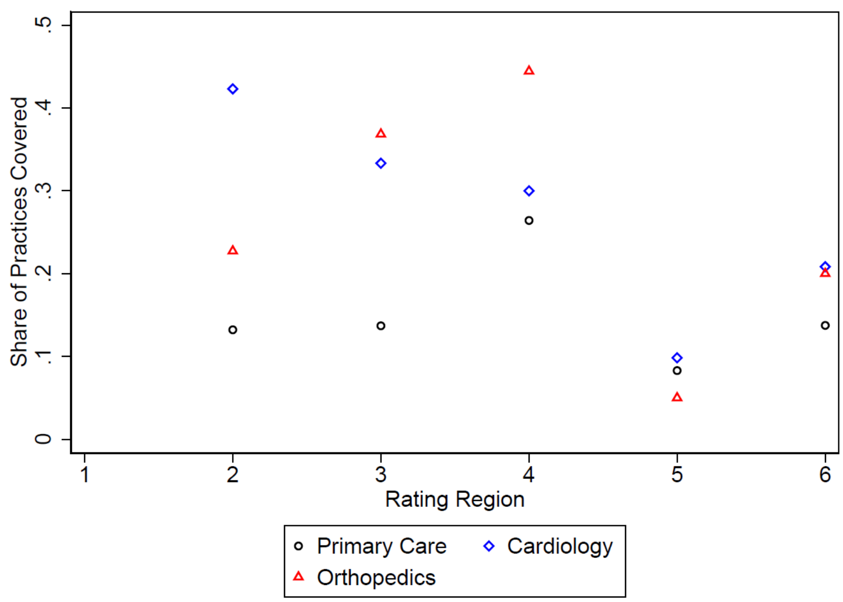
Notes: This figure plots the share of the top 50 physician practices (by market share) in a given rating region covered by year and specialty for all narrow-network products on the GIC. Each point represents a particular insurance plan on the GIC. Panel (a) displays each insurance plan, while panel (b) displays only Fallon Direct (Fallon's narrow-network plan) and Harvard Primary Choice (Harvard's narrow-network plan).

Figure B.2: Share of Practices Covered by Specialty and Rating Region, 2011

(a) All Practices, All Plans



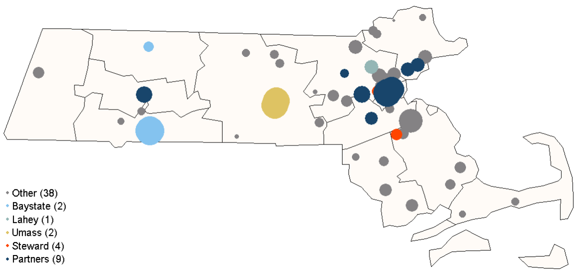
(b) All Practices, Fallon



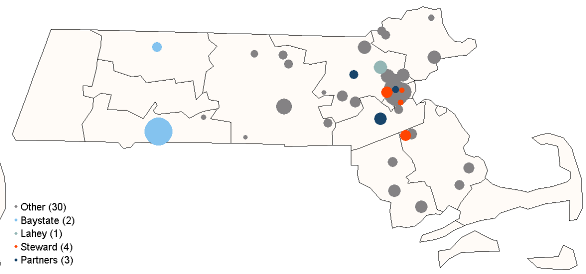
Notes: This figure plots the share of the top 50 physician practices (by market share) in a given rating region covered rating region and specialty for all narrow-network products on the GIC in year 2011. Each point represents a particular insurance plan on the GIC. Panel (a) displays each insurance plan, while panel (b) displays only Fallon Direct (Fallon's narrow-network plan)

Figure B.3: Hospital Networks by Plan, 2011

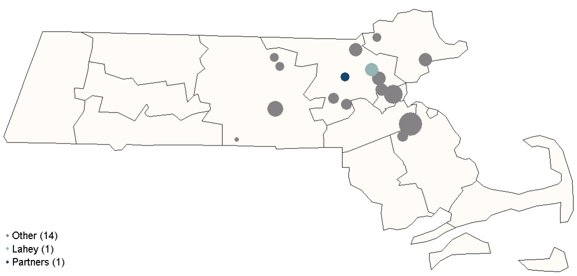
(a) Harvard / Tufts Broad



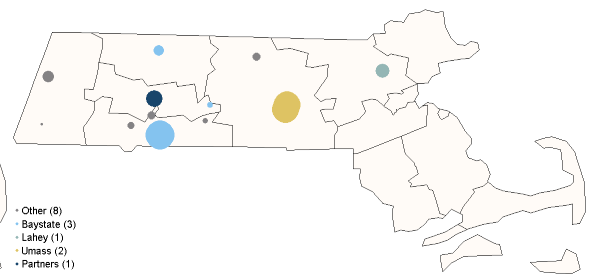
(b) Harvard Narrow



(c) Fallon Narrow



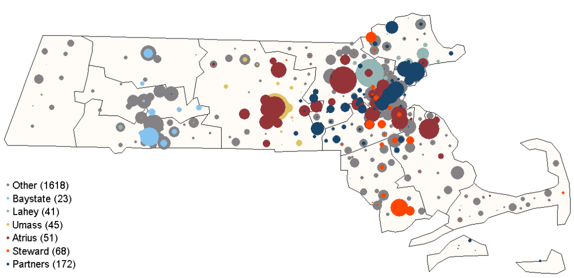
(d) Health New England Narrow



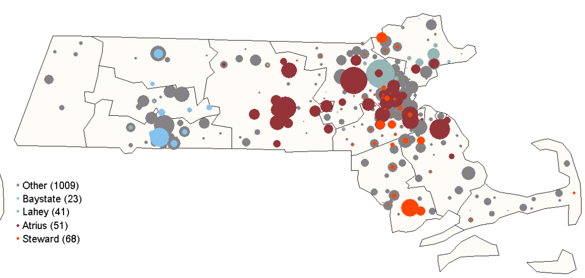
Notes: This figure plots the hospital networks of specified plans on the GIC in 2011. Sizes of the data points reflect relative market shares of the practices. Colors reflect ownership status (which health systems owns which practice).

Figure B.4: Primary Care Practice Networks by Plan, 2011

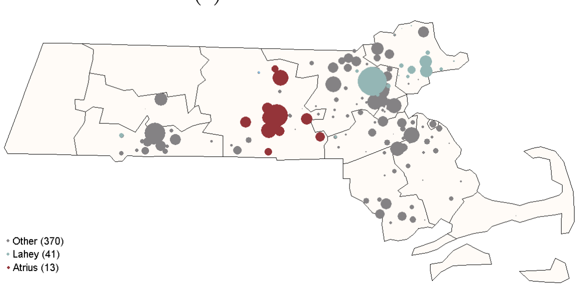
(a) Harvard / Tufts Broad



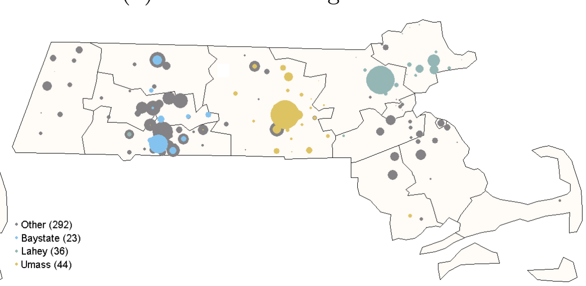
(b) Harvard Narrow



(c) Fallon Narrow



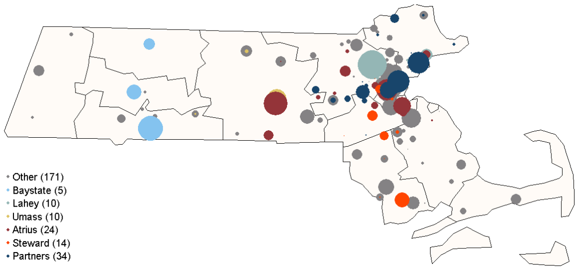
(d) Health New England Narrow



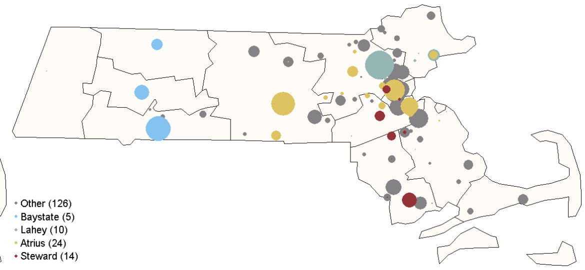
Notes: This figure plots the primary care practice networks of specified plans on the GIC in 2011. Sizes of the data points reflect relative market shares of the practices. Colors reflect ownership status (which health systems owns which practice).

Figure B.5: Cardiology Networks by Plan, 2011

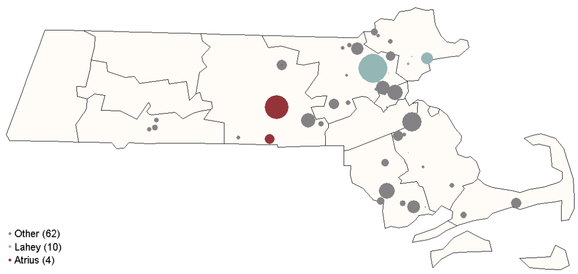
(a) Harvard / Tufts Broad



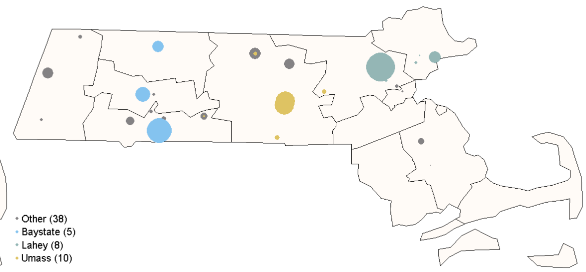
(b) Harvard Narrow



(c) Fallon Narrow



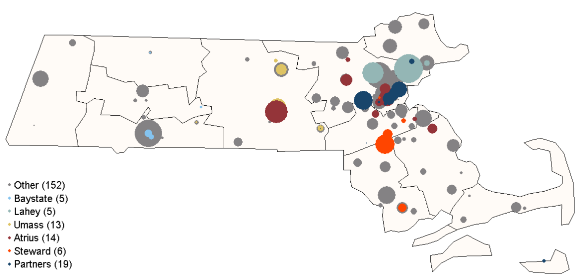
(d) Health New England Narrow



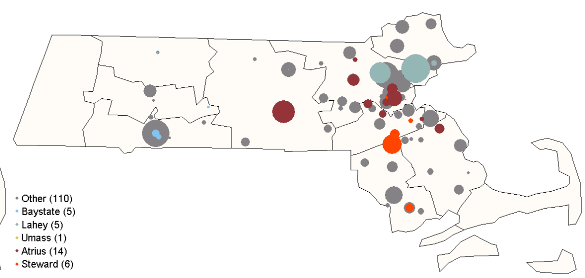
Notes: This figure plots the cardiology practice networks of specified plans on the GIC in 2011. Sizes of the data points reflect relative market shares of the practices. Colors reflect ownership status (which health systems owns which practice).

Figure B.6: Orthopedic Networks by Plan, 2011

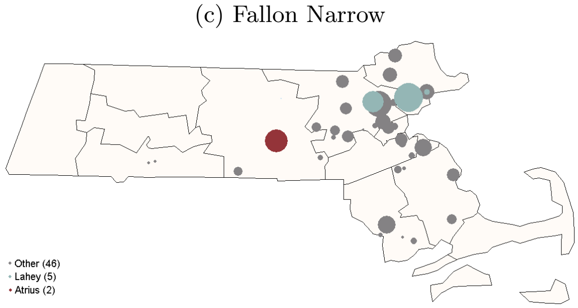
(a) Harvard / Tufts Broad



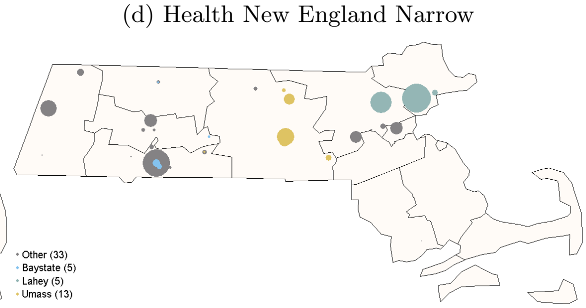
(b) Harvard Narrow



(c) Fallon Narrow



(d) Health New England Narrow



Notes: This figure plots the orthopedic practice networks of specified plans on the GIC in 2011. Sizes of the data points reflect relative market shares of the practices. Colors reflect ownership status (which health systems owns which practice).

C Model Estimation Details

C.1 Physician Demand

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. The patient characteristics include five-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 ICD-9 diagnosis codes. Given that my data span 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 (2016); a discussion of their construction can be found in subsection C.3). Hospital characteristics include 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 physicians, I match procedures performed (CPT codes) to Medicare RVU weights, r, which serves as a proxy for procedure intensity. 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 practice who are specialists (relative to primary-care physicians), whether the practice is part of a medical group, whether the practice is owned by a hospital or health system, and the number of total locations of the medical group. 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.

To capture physician inertia, I include three separate indicators: whether a patient had sought care from this particular physician practice previously; whether a patient had sought care from any of the practice's locations previously; and whether a patient had previously sought care from any provider employed by the hospital or health system that owns the particular practice.

I run the model separately for hospitals, PCPs, cardiologists, and orthopedists. I assume 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.[53]

Dimensionality Reduction Perhaps the most salient issue in estimation of the physician models is the presence of tends of thousands of physicians within each specialty group in Massachusetts, making estimation of parameters through a multinomial logit framework difficult. I take three primary approaches to reduce the dimensionality problem. The first is that, as previously described, I estimate the provider demand model at the physician *practice*-zip-code level rather than the individual physician level. This reduces the patient choice set considerably. Second, I estimate the model separately by the seven rating regions in Massachusetts, as defined by CMS.[54] As individual practices are location-specific, this allows me to include a larger span of the full Massachusetts physician practice space in my estimation. In addition, it allows for estimation of flexible parameters that vary by region.[55]

Finally, I assume that only the top 50 practices (by market share) within each region and specialty group have an individual mean utility. All practices outside the top 50 are assumed to have identical mean utilities and only be differentiated on distance to the patient. In order to further narrow the choice set, I assume that practices outside the top 50 in a region can be grouped into a set of 7 discrete distance bands, b, where $b = 0$ to 5 miles, 5 to 10 miles, 10 to 15 miles, 15 to 30 miles, 30 to 50 miles, 50 to 100 miles, and over 100 miles. I assume that the distance between any given patient and physician practice, T_{id}, is constant within each of these bands and takes the value of the midpoint of the distance band, i.e. $\{T_{id} \in b\} = b^{mid}$.[56] Given these assumptions, and dropping the region and time subscripts for convenience, the model in Equation 1 becomes:

$$u^s_{ird} = \underbrace{\phi^s_{ird} + \varepsilon^s_{ird}}_{\text{Utility for Top 50 Practices}} \tag{25}$$

$$u^s_{ird} = \underbrace{\sum_b \mathbb{1}\{T^s_{id} \in b\}(T^s_{ib}\lambda^s_1 + T^s_{ib}v_{ir}\lambda^s_2 + N^s_{ib}\gamma^s_b) + \varepsilon^s_{irb}}_{\text{Utility for Practices Outside Top 50}} \tag{26}$$

where N^s_{ib} is the number of physicians of specialty s in individual i's network in distance band b. This specification can be thought of as adding a single option to the choice set for each distance band b, rather than an individual option for each physician practice in those distance bands. γ^s_b, then, rather than estimating a fixed effect for each individual practice $d \in b$, simply estimates a fixed effect for each distance band b and scales it by the number of physicians in that band. This allows patient value of these options to vary by the number of doctors in those groups. As an

[53] 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.

[54] https://www.cms.gov/CCIIO/Programs-and-Initiatives/Health-Insurance-Market-Reforms/ma-gra.html

[55] For instance, it is likely that individuals in Boston would be more averse to traveling for physician care than individual in Worcester, due to the density of patients and providers in the former relative to the latter.

[56] As an example, $b^{mid} = 2.5$ for distance band b =0 to 5 miles.

example, if patient i's physician network removed a physician practice in distance band b, patient i's utility would decrease by $(N^s_{ib} - 1)\gamma^s_b$.

The assumption that practices outside the top 50 have the same mean utility conditional on distance bands may seem like a strong one. However, it makes sense given two empirical facts. First, the top 50 practices by market share in a given region account for most of patient claims.[57] Second, most practices outside the top 50 are included in all plans' networks, even narrow-network products. As a result, most of the variation in networks across plans comes from network choice among these top practices. Therefore, treating these smaller practices as essentially undifferentiated in quality (but for distance) not only has the benefit of making the model more easily estimable, but also likely to hold true given observed networks.

Outside Option: For the hospital choice model, I define the outside option to be any hospital outside the state of Massachusetts. For the physician models, I assign any physician practice in distance band $b = 7$ (i.e. outside of 100 miles from the patient's location) to be the outside option. I normalize these goods to be 0 in the utility models.

C.2 Plan Demand

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. As with the provider demand model, I include 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 and β_2 (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. 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, as a first-approximation, I match enrollee zip codes to public data on municipalities entering the GIC by year and do not extend the premium holiday to members with zip codes in the corresponding municipalities who joined during the corresponding years.[58]

[57] In Boston, for instance, where there is the highest density of physicians, the top 50 PCP practices account for approximately 70% of all claims, while the top 50 cardiology and orthopedic practices account for nearly 90% of all claims.

[58] This is likely to produce some amount of measurement error, but sensitivity checks on the specific zip codes used revealed very minor fluctuations of the core coefficients. Moreover, 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.

Construction of EU_{Ijt}: Recall the main specification is the one given by Equation 5. However, I also consider additional specifications at different levels of aggregation of the EU_{Ijt} terms. The first is a specification where I aggregate the expected utility terms for each physician specialty is reported as one linear combination such that:

$$EU_{Ijt}^{S} = \sum_{s} \sum_{r} f_{ir} log \left(\sum_{d \in N_{jtS}} exp(\phi_{ird}^{s}) \right) \tag{27}$$

Thus, the utility function function collapses to:

$$u_{Ijt} = \underbrace{-r_{Ijt}\alpha_I + EU_{Ijt}^{H}\beta_1 + EU_{Ijt}^{S}\beta_2 + \mathbb{1}_{Ijt=Ijt-1}\beta_3 + \eta_j}_{\delta_{Ijt}^{agg}} + \omega_{Ijt} \tag{28}$$

In addition, I consider specifications where I interact the physician utility terms with rating region (the level at which they were estimated).

For the ex-ante illness probabilities, f_{il} and f_{ir}, 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 categories used in the demand estimation. For those 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.[59]

C.3 Estimates of p_{jht} and p_{jdt}^{s}

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 observed allowed amounts, as in Kleiner et al. (2015). I assume that price takes the following form:

$$A_{irjdt}^{s} = p_{jdt}^{s} * RVU_{rt} \tag{29}$$

$$ln(A_{irjd}^{s}) = ln(p_{jdt}^{s}) + ln(RVU_{rt}) \tag{30}$$

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

[59] A perhaps 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

$$ln(A^s_{irjdt}) = ln(RVU_{rt})\rho + \gamma^s_{jdt} + \epsilon^s_{irjdt} \tag{31}$$

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

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:

$$ln(A_{iljht}) = \gamma_{jht} + \psi_{lt} + x_{ilt} + \epsilon_{iljh} \tag{32}$$

Here, A_{iljht} refers to the observed allowed amount for patient i with diagnosis l on plan j seeking care from hospital h. γ_{jht} are fixed effects for every plan-hospital-year 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_{lt}. 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, $p_{jht} = exp(\gamma_{jht})$ 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_{jht}.

C.4 Additional Estimators for FC_j

As two additional estimators, I assume that ρ is zero (i.e. the GIC values consumer surplus equally to its spending on health care and fixed costs) and construct moments solely based on one-step deviations of the *number* of products, rather than altering networks within-product. In other words, I estimate bounds on FC_j from the addition and removal of products from the menu.

For this, I construct two counterfactual quality vectors. 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 similar revealed preference assumption as the previous estimator, namely that the products I observe in the data are chosen in equilibrium. 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 variable social surplus, $S(\delta_{Jt}, \theta)$. 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:

$$FC_j \leq E\left[S(\delta_{Jt}, \theta) - S(\delta_{J-j,t}, \theta)\right] \equiv \overline{FC}_j \tag{33}$$

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

$$FC_j \geq E\left[S(\delta_{J+j,t}, \theta) - S(\delta_{Jt}, \theta)\right] \equiv \underline{FC}_j \tag{34}$$

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, where $v_{3,\delta_{Jt}}$ is a disturbance such that $E[v_{3,\delta_{Jt}}] = 0$:

$$E[S(\delta_{Jt}, \theta)] = S(\delta_{Jt}, \theta) + v_{3,\delta_{Jt}} \tag{35}$$

As long as the GIC has correct expectations on average, the estimation equation becomes:

$$\underset{K\to\infty}{plim} \frac{1}{K} \sum_{j}^{K} \left(S(\delta_{Jt}, \theta) - S(\delta_{J-j,t}, \theta)\right) \geq FC \geq \underset{K\to\infty}{plim} \frac{1}{K} \sum_{j}^{K} \left(S(\delta_{J+j,t}; \theta) - S(\delta_{Jt}; \theta)\right) \tag{36}$$

I construct an additional estimator for comparison under the strong assumption that insurers have the freedom to design and offer products without input from the GIC. In other words, Equation 36 becomes:

$$\underset{K\to\infty}{plim} \frac{1}{K} \sum_{j}^{K} \left(\pi_m(\delta_{Jt}, \theta) - \pi_m(\delta_{J-j,t}, \theta)\right) \geq FC \geq \underset{K\to\infty}{plim} \frac{1}{K} \sum_{j}^{K} \left(\pi_m(\delta_{J+j,t}, \theta) - \pi_m(\delta_{Jt}, \theta)\right) \tag{37}$$

D Provider Demand Specifications

D.1 Hospital Demand

Table D.1 reports 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 Women's (both owned by Partners), Beth Israel, Lahey Medical Center, and others.

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.[60]

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

[60] As discussed, this may be due to true "inertia," such as a tangible switching cost of moving hospitals, or it may be due to unobserved preference heterogeneity.

Table D.1: 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	

Notes: 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.

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). Table D.2 presents the results for alternate hospital demand models that include these copays. 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 patient *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 copays 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). 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.

In addition, these models 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

Table D.2: 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

Notes: 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.

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.

D.2 Physician Demand

Table D.3 reports the results of the physician demand models for PCP practices, cardiology practices, and orthopedic practices for the Boston rating region. 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 four years for each different specialty group.[61] 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.[62] As the model was estimated separately for each of the seven Massachusetts health rating regions, I only report here select coefficients for the Boston rating region and the Worcester rating region. 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 E addresses potential selection concerns in more detail.

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 quite large for primary care physicians, it is about half the size for cardiology and orthopedic practices.[63] Across all three specialty groups, patients, on average, prefer visiting practices owned by hospitals or health systems, though the effect is considerably stronger for cardiology practices.[64]

Somewhat surprisingly, distance interacted with female and distance interacted with age are small and insignificant across most of the models, in contrast to the results in the hospital demand model. The only exceptions are a significant negative coefficient for distance interacted with female in the orthopedic model, and a significant negative coefficient for distance interacted with age in the PCP model. The former may be driven by the large number of sports injuries that orthopedists treat, which tend to be among patients who are disproportionately male. The latter is consistent with the result from hospital demand, namely that conditional on risk, older individuals prefer to

[61] I omit year 2009, the earliest year of data in the claims, as I cannot observe patients' prior-use of physicians in that year.

[62] RVU weights are a proxy for intensity of procedure needed. They are used by Medicare to determine payment rates for physicians paid fee-for-service on Medicare Part B . See Appendix A for a more detailed description.

[63] However, these coefficients should be interpreted with caution on their own. As these models are estimated separately, these coefficients are not directly comparable, as their magnitudes are driven in part by relation to practice fixed effects as well as well as scaling of the logit error.

[64] This is consistent with descriptive statistics showing that patient-weighted visits to cardiologists tend to be among larger practices. See Appendix A.

Table D.3: Results of Physician Demand Models (Boston)

Variable	PCP Practices	Cardiology Practices	Orthopedic Practices
Distance	-0.5110***	-0.2875***	-0.2306***
	(0.0192)	(0.0146)	(0.0162)
Owned by Hosp. or System	0.3406***	1.3385***	0.6836***
	(0.0971)	(0.0899)	(0.0867)
Used Prac Previously	4.2067***	1.5982***	2.0279***
	(0.0305)	(0.0255)	(0.0319)
Used Med Grp Previously	1.7844***	1.8032***	1.7478***
	(0.0394)	(0.0325)	(0.0418)
Used System Previously	0.2065***	0.7691***	0.9249***
	(0.0341)	(0.0290)	(0.0351)
	Interactions with Patient Characteristics		
DistxFemale	0.0008	-0.0041	-0.0062*
	(0.0014)	(0.0029)	(0.0034)
DistxAge	-.0009***	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
DistxChronic	0.0026	0.0289***	0.0296***
	(0.0019)	(0.0079)	(0.0057)
	Interactions with Provider Characteristics		
DistxNumDocs	0.0003***	-0.0004***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
DistxNumLocs	0.0036***	-0.0035***	-0.0049***
	(0.0008)	(0.0006)	(0.0007)
DistxMedGrp	0.0375***	-0.0386***	-0.0548***
	(0.0079)	(0.0077)	(0.0073)
AgexNumDocs (00s)	-0.0019**	0.0045***	0.0000
	(0.0000)	(0.0000)	(0.0000)
AgexNumLocs (00s)	0.0318**	0.1467***	0.0722***
	(0.0166)	(0.0132)	(0.0149)
AgexMedGrp	-0.0015	0.0106***	0.0240***
	(0.0014)	(0.0015)	(0.0014)
Practice FE	Yes	Yes	Yes
Obs.	3,062,854	1,853,668	1,635,091
Pseudo R2	0.63	0.59	0.56

Notes: Results of physician demand models are for years 2010-2013 for Boston rating region only. Excluded from the table are distance, RVU weights, and gender interacted with additional practice characteristics: number of unique services as the practice, share of physicians at the practice who are specialists, and number of doctors across the entire system. Model contains a full set of practice fixed effects. Note that AgexNumDocs and AgexNumLocs are reported in hundreds (00*s*).

travel smaller distances to seek care, particularly for routine primary care treatment.[65] Across all three specialty groups, the presence of a chronic condition is associated with increased travel time, though this coefficient is insignificant in the PCP demand model. This is suggestive that sicker patients tend to have stronger preferences for specialists than primary care physicians.

Patients seeking primary care are willing to travel further to access practices with more physicians on site. In addition, they are willing to travel further for practices with more locations and practices that are affiliated with medical groups. This result makes sense, particularly in the Boston rating area, as many physician practices are owned by larger groups, such as Partners and Atrius. However, this result is reversed for cardiologists and orthopedists. Patients are less willing to travel for larger practices, practices with multiple locations, and practices that are part of larger medical groups. While somewhat surprising, this is tempered by the age interactions, which show that older individuals significantly prefer visiting physicians from larger practice sites, physicians who are part of medical groups, and groups with multiple locations.[66] This is particularly pronounced for cardiologists, where the age effect on visiting larger practices is considerably larger than the other specialty groups.This result holds on Worcester as well.

To capture physician inertia, I include three separate indicators: whether a patient had sought care from a particular physician practice previously; whether a patient had sought care from any of the practice's locations previously; and whether a patient had previously sought care from any provider employed by the hospital or health system that owns the particular practice. All three of these inertia measures are highly important to predicting physician choice across all specialty groups, with having used the particular physician in the past being the biggest predictor and having used a provider owned by the same health system being the smallest. The estimates imply that a 35-year-old individual in average health would be on average willing to travel an additional 11.3 miles to access the same PCP practice, 14.4 miles to access the same cardiology practices, and 20.4 miles to access the same orthopedic practices. This implies that individuals are, conditional on age, RVU, presence of a chronic condition, and practice characteristics willing to travel somewhat more for specialty care than primary care. The magnitudes are quite similar to the magnitudes in the hospital demand model. Altogether, these results imply that inertia to previously used physicians play a significant role in provider choice.

For comparison to the models from Boston reported in subsection D.2, Table D.4 reports the results of the physician demand model for the Worcester rating region. The results are qualitatively similar to the results from the Boston rating region, however there are some notable exceptions. First, physician inertia, particularly to PCPs, plays a much larger role in Worcester than in Boston in terms of distance traveled. While in Boston, patients were on average willing to travel an additional 11.3 miles to access the same PCP practice, this figure is approximately 42 miles in Worcester. This may be, in part, due to high volume of PCPs in Boston relative to Worcester, or may be due to the fact that Worcester is an area that requires driving more so than walking.[67] Moreover, seeking care from a physician owned by a hospital or health system seems to have less of an effect in Worcester and is, in fact, *negative* for orthopedic practices. This may be

[65] In addition, the model includes distance interacted with RVU weight (omitted), which likely proxies for age.

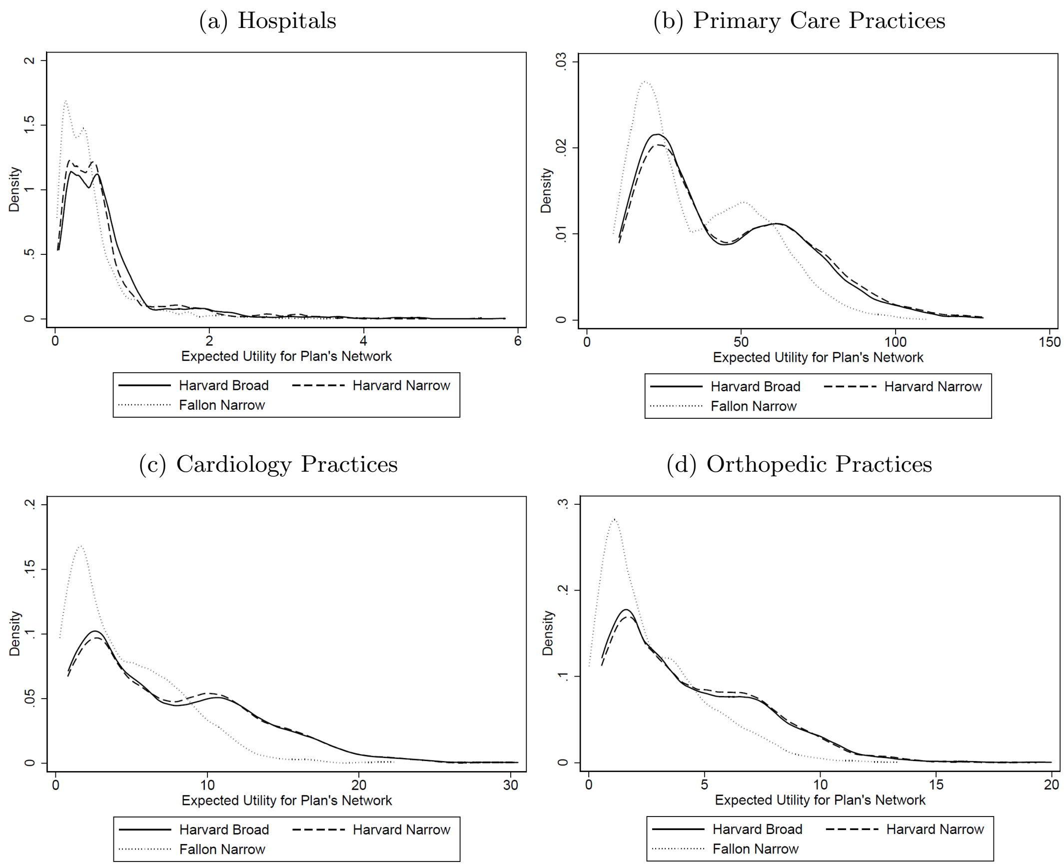
[66] The exception is PCPs, which shows older individuals preferring smaller practice locations.

[67] The average distance traveled for PCPs in Boston is about half that of Worcester.

reflective of the fact that, unlike Boston, Worcester has fewer prestigious academic medical centers. However, Worcester does contain a prominent medical group: the Fallon Clinic (later renamed Reliant Medical Group). Much like in Boston, older patients therefore significantly prefer seeking care from doctors that are part of medical groups and that work for practices which have multiple locations.

Figure D.1 plot the density of each household's expected utility, EU_{Ijt}, for a hospital and physician specialty for three plan's networks in the Boston rating region: Harvard Broad, Harvard Narrow, and Fallon Narrow. It is immediately clear from this series of charts that Harvard's narrow plans yield lower utility than its broad plans, and that Fallon's narrow plan yields even lower utility. This pattern is consistent across provider types. This makes sense given that Harvard's narrow network covers a fairly large number of providers–almost all excluding those owned by Partners–whereas, Fallon covers significantly fewer providers in Boston.

Figure D.1: Expected Utility for Various Networks, Boston Rating Region



Notes: This figure plots the distribution of EU^{H}_{Ijt} and EU^{s}_{Ijt} for each physician specialty. Figures are plotted for households in the Boston rating region. Each figure plots the density of expected utility for three plans: Harvard Broad, Harvard Narrow, and Fallon Narrow.

However, the differences across provider types tells a more illuminating story. Panel (a) shows the distribution of total utility for hospitals, EU^{H}_{Ijt}. While the plot for the broad network does skew slightly to the right to that of the narrow network, the three network utilities virtually overlap one another for a significant portion of the density plot. Looking at panel (b), which shows the utility

Table D.4: Results of Physician Demand Models (Worcester)

Variable	PCP Practices	Cardiology Practices	Orthopedic Practices
Distance	-0.1459***	-0.1682***	-0.2046***
	(0.0123)	(0.0106)	(0.0107)
Owned by Hosp. or System	-0.1753**	0.1083	-0.3115***
	(0.0692)	(0.0969)	(0.0900)
Used Prac Previously	5.0031***	1.6125***	2.6716***
	(0.0273)	(0.0381)	(0.0458)
Used Med Grp Previously	0.4997***	1.2256***	1.152***
	(0.0445)	(0.0561)	(0.0667)
Used System previously	0.7585***	0.8114***	0.9833***
	(0.0449)	(0.0535)	(0.0622)
	Interactions with Patient Characteristics		
DistxFemale	0.0015	-0.0002	-0.0032
	(0.0012)	(0.0025)	(0.0027)
DistxAge	-0.0003***	-0.0002**	0.0001
	(0.0000)	(0.0001)	(0.0000)
DistxChronic	0.0064***	0.0190***	0.0496***
	(0.0019)	(0.0057)	(0.0046)
	Interactions with Provider Characteristics		
DistxNumDocs	-0.0001*	0.0000	-0.0002***
	(0.0000)	(0.0000)	(0.0000)
DistxNumLocs	-0.0106***	-0.0021***	0.0000
	(0.0013)	(0.0006)	(0.0000)
DistxMedGrp	-0.3889***	-0.0860	-0.4409***
	(0.0704)	(0.1015)	(0.0973)
AgexNumDocs (00s)	0.0010	-0.0024	0.0027***
	(0.0000)	(0.0000)	(0.0016)
AgexNumLocs (00s)	0.3278***	0.1889***	-0.0000
	(0.0378)	(0.0403)	(0.0000)
AgexMedGrp	-0.0017	0.0108***	0.0098***
	(0.0022)	(0.0032)	(0.0022)
Practice FE	Yes	Yes	Yes
Obs.	2,540,635	657,160	552,351
Pseudo R2	0.59	0.61	0.59

Notes: Results of physician demand models are for years 2010-2013 for Worcester rating region only. Excluded from the table are distance, RVU weights, and gender interacted with additional practice characteristics: number of unique services as the practice, share of physicians at the practice who are specialists, and number of doctors across the entire system. Model contains a full set of practice fixed effects. Note that AgexNumDocs and AgexNumLocs are reported in hundreds (00s).

distribution for PCPs, EU_{Ijt}^{PCP}, consumers appear to view both Harvard plans quite similarly, whereas the Fallon Narrow plan noticeably skews left, suggesting that there is considerably more variation in the *physician* utilities across these networks than the hospital utilities. This becomes even more pronounced in panel (c) and panel (d), where the utility for cardiologists and orthopedists in Fallon's plans skews even further to the left.

Taken together, these figures 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.

E Selection on Unobservables in the Provider Demand Models

A concern with multinomial logit demand models of the type presented in section 3 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^H_{ijt} and N^S_{ijt}, the expected utility of a particular hospital and physician network, EU^H_{Ijt} and EU^S_{ijt}, 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 unobserved 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 et al., 2017).

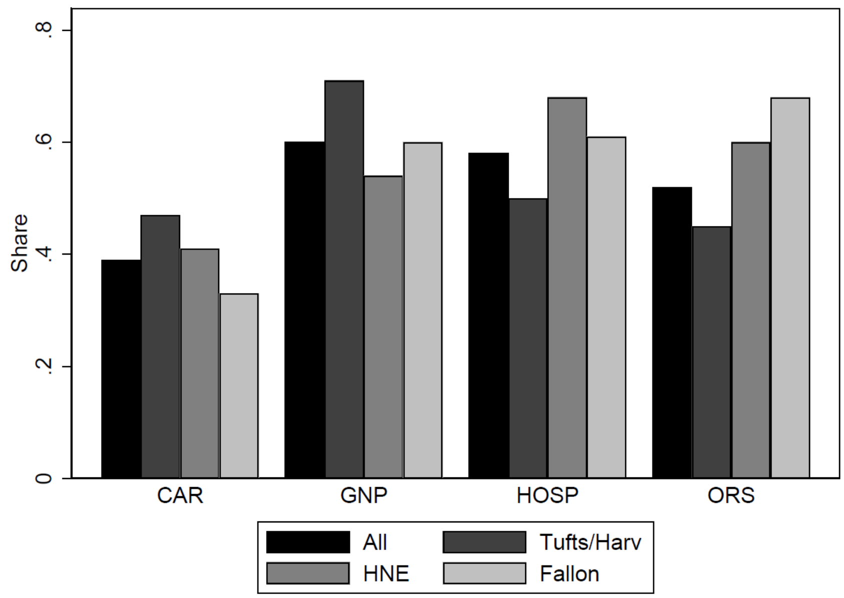
I present here some reduced form evidence suggesting that such selection is not a major concern in my setting. Figure E.1 displays the share of individual choices of hospitals and physicians for individuals only in *narrow-network* plans that are accurately predicted by a model of provider demand run 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, however, the logit model predicts the choices of narrow-network patients quite well. For PCPs, 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 E.2 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.[68] 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.

[68] 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).

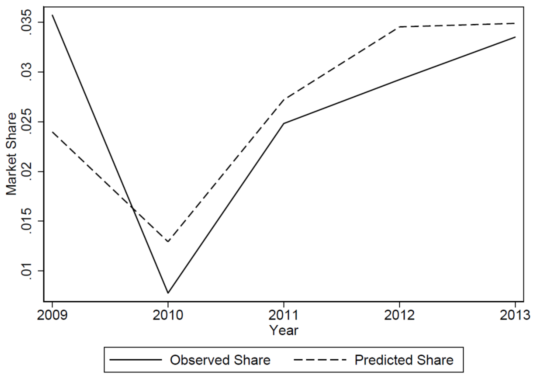
Figure E.1: Share of Actual Choices Accurately Predicted, by Specialty



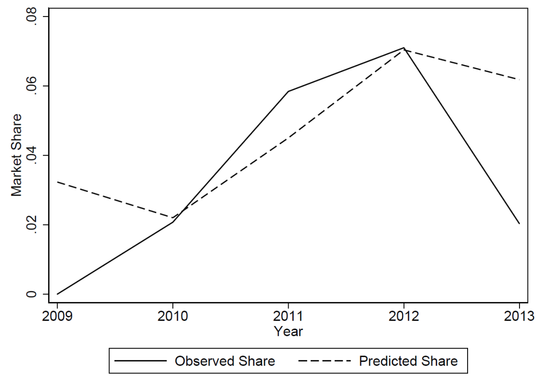
Notes: This figure plots the share of choices of providers made by individuals in narrow-network plans that are accurately predicted. Parameters used for prediction were estimated from a demand model among only individuals in broad-network plans.

Figure E.2: Observed versus Predicted Hospital Shares for Narrow Network Patients

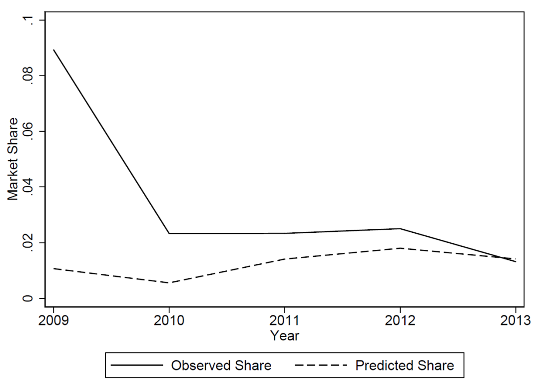
(a) Tufts Medical Center Share



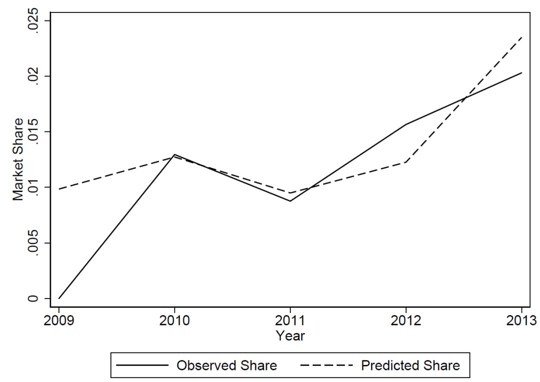
(b) Beth Israel Hospital Share



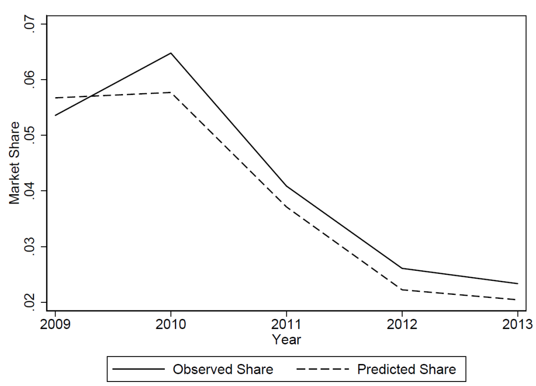
(c) Boston Medical Center Share



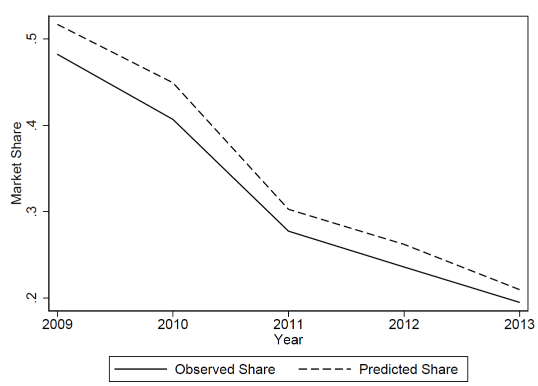
(d) Lahey Hospital Share



(e) Berkshire Medical Center Share



(f) Baystate Medical Center Share



Notes: This figure plots actual market shares of select medical centers against the predicted market shares of those medical centers among consumers in narrow-network plans. Parameters used for prediction were estimated from a demand model among only individuals in broad-network plans.

F Cost Estimates

F.1 Estimates of p_{jt}

Table F.1 reports the average negotiated base prices for hospitals and physicians and average weights by type of provider and facility type in 2011.[69] The table suggests that negotiated prices do not vary considerably across medical specialties in Massachusetts, on average. Within specialty, however, there is considerable variation. Facility-based cardiology practices, for instance, receive an average price-per-rvu of $56, but with a standard deviation of $20. Certain practices, therefore, receive more than $80 per RVU. In the hospital market, the maximum base price in 2011 was $17,306 while the minimum was $3,545. Additionally, there are some notable differences in the average weights per procedure for physicians. Office-based PCPs, for instance, submit an average of 2.19 RVUs per visit, yielding an average of $122 per visit. Orthopedists, however, perform an average of 4 RVUs per visit, implying an average payment of $220 per visit.

Table F.1: Estimated Price and Weight Measures, 2011

Variable	PCPs	Cardiologists	Orthopedists	Hospitals
		Office-Based		
Average Base Price	56.55	56.29	55.37	–
	(12.43)	(14.79)	(16.94)	–
Average Weight	2.19	2.74	3.99	–
	(0.60)	(1.25)	(2.45)	–
		Facility-Based		
Average Base Price	56.71	56.59	52.51	10,303.73
	(14.48)	(19.61)	(16.43)	(3,177.89)
Average Weight	2.35	2.07	5.38	1.00
	(1.13)	(1.95)	(5.22)	(0.34)

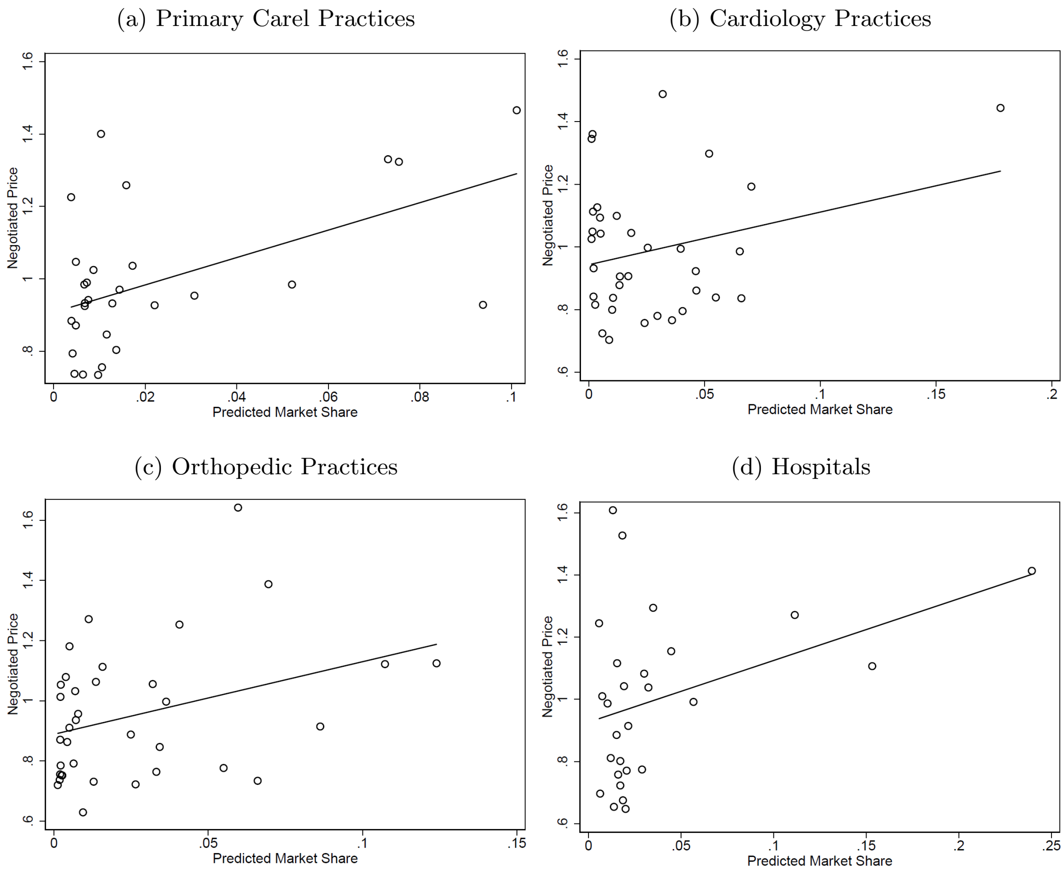
Notes: Standard deviations in parentheses. "Average base price" refers to the negotiated price for a standardized unit of health care. In the case of physician practices, this refers to a case where $RVU_l = 1$. In the case of hospitals, this refers to the case where $w_l = 1$. Hospital weights are scaled so that the yearly average is one, meaning that hospital base prices refer to the price fo a procedure of average weight. "Office-based" settings are defined are practices where more than 70% of claims are flagged as in an office-based setting.

I next examine whether the preference for broad network plans translates into higher negotiated rates for those providers. Figure F.1 depicts the relationship between demand and negotiated provider price for one of the insurers on the GIC in the Boston rating region. Due to confidentiality concerns, I omit both the identity of the insurer and the actual negotiated rate. Instead, I report the negotiated rate relative to the insurer-specific average. The y-axis depicts this standardized rate, where the x-axis depicts the predicted market share from the provider demand models.

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

[69] I define 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.

Figure F.1: Insurer Negotiated Price by Market Share, Boston Rating Region 2011



Notes: This figure plots the the negotiated price for hospitals, p_{jht}, and for physician practices, p^s_{jtd}, against predicted market share from the provider demand models. Prices are reported for a single insurer and relative to the insurer-specific mean. Data is for year 2011.

for hospitals and, surprisingly, primary care providers, though there is still a positive relationship for cardiologists and orthopedists as well.[70] 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.

F.2 Unobserved Marginal Cost Estimates

Table F.2 reports the results Equation 16, regressing the log of unobserved marginal costs from the supply side on insurer fixed effects, year fixed effects, and an indicator for whether the plan is narrow or not. Year 2012 is omitted due to potential bias in estimates from it being the year of

[70] This may be explained by the presence of Harvard Vanguard in the Boston rating region, which has considerable bargaining power. Other large primary care practices in the area likely hold similar bargaining power. Though modeling the full bargaining game between physician practices and insurers is outside the scope of this paper, it is an interesting subject for future work.

the premium holiday.

Table F.2: Unobserved Marginal Cost Estimates

Variable	Coefficient	Standard Error
Narrow Network	-0.174***	0.022
Harvard Pilgrim	0.056**	0.025
Health New England	-0.053*	0.030
Neighborhood Health Plan	-0.046	0.030
Tufts Health Plan	0.054**	0.025
2010	-0.006	0.026
2011	0.037	0.025
2013	0.082***	0.025
Constant	5.904***	0.024
Obs.	28	
Adjusted R2	0.87	

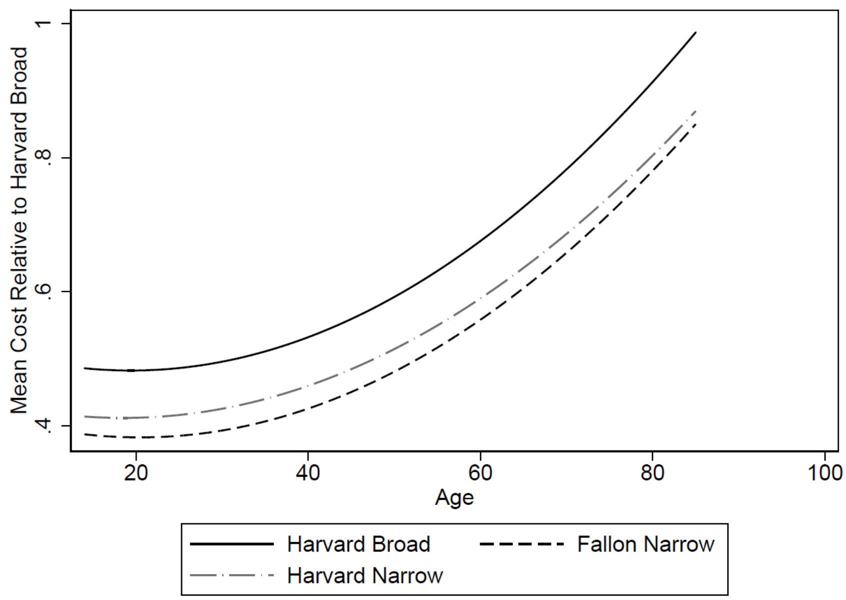
Notes: Results from marginal cost estimation. Dependent variable is the log of unobserved marginal costs. Omitted insurer is Fallon Health Plan. Omitted year is 2009. Year 2012 is left out of analysis due to concern about bias in estimates from it being the year of the premium holiday.

The results indicate that being a narrow network plan reduces unobserved marginal costs by approximately 17%. Among insurers, Harvard and Tufts each have higher relative unobserved costs, compared with Health New England, Neighborhood Health Plan, and Fallon. This indicates that Harvard and Tufts may have non-hospital, PCP, cardiology, and orthopedic expenditures that are higher, potentially due to contracting with larger set of providers unaccounted for by the chosen specialties.[71] Unobserved costs increase steadily over time, likely reflecting increases in negotiated prices with providers over time as well as general medical inflation. In particular, unobserved costs in 2013 are estimated to be approximately 8% higher than in 2009.

Figure F.2 plots the total estimated insurer marginal costs (hospital + PCP + cardiology + orthopedics + unobserved) against age for single-member households. I report estimated cost-curves for Harvard Broad, Harvard Narrow, and Fallon Narrow. As expected, predicted insurer costs rise rapidly with age. Moreover, the broad-network plan has consistently higher predicted costs than the narrow-network plans at all age levels. Further, the cost-curves do slope upward as similar rates, although Harvard Broad does have a slight uptick in the rate at which it rises after age 60 relative to the narrow products. This suggests the potential for selection on expensive providers, particularly among older individuals, conforming to the results of the hospital and physician demand models.

[71] An alternate explanation is that these costs reflect higher administrative costs or more generous drug formularies.

Figure F.2: Estimated Insurer Marginal Costs



Notes: This figure plots estimated marginal cost curves for select plans in 2013. Note that the y axis reflects costs relative to the average cost of Harvard Broad.

G Simulation Procedure

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 dimensionality of the computation, as with the employer objective function estimation, I restrict the GIC offer set to that outlined in subsection 4.4. 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^H_{ijt} and EU^s_{ijt}, for each offered product's network using the estimates from the provider demand model
3. Compute the predicted medical costs, c^H_{Ijt} and c^S_{Ijt} for each household if they enrolled in any of the offered products, using the negotiated price construction
4. Compute the base unobserved insurer marginal costs, c^u_{Ijt}, using the parameters estimated from Equation 16.
5. Compute the expected market shares and premiums, $s_{Ijt}(\delta_{Jt}, \theta)$ and $R_{Ijt}(\delta_{Jt}, \theta)$, for each household in each offered product, using the results from the insurance plan demand model and the pricing equation in Equation 13.
6. Compute the estimated consumer surplus, $CS(\delta_{Jt}, \theta)$, and total outlays for the GIC under the current product offered
7. Compute the GIC's objective function using estimated $CS(\delta_{Jt}, \theta)$, total expenditures, and estimated FC_j.
8. Repeat this procedure for each vector of possible equilibria, and take the max of all the computed welfare functions.

H Counterfactual Results Assuming No Logit Error

Estimates of ρ and FC_j: The results presented in subsection 5.2 assume that consumer surplus, $CS(\delta_{Ijt}, \theta)$ is calculated in the traditional "logsum" way, which implicitly assigns consumers a positive valuation of any counterfactual product added to a choice set regardless of where that product lies in the quality space. This valuation is assigned through the idiosyncratic logit error. As a result, even if products of "low quality" are introduced to the market, consumers may be made better off in a way that may bias the number of equilibrium products upward.[72] To remove this potential bias, I re-estimate Equation 21, allowing consumers to select products as they would in a logit would, but setting the logit error to zero for the purposes of computing consumer surplus.[73]

Table H.1: Results of Employer Objective Function Estimation

	ρ	FC_j
GIC/Employer ($Millions)	5.27	1.58
Percentage of Net Spending		0.16
Percentage of Net Surplus		0.36

Results from ρ and FC_j estimation for 2009-2013. For these estimates, $CS(\delta_{Ijt}, \theta)$ is computed assuming the logit error shock is zero. The corresponding percentages of fixed costs relative to net GIC health spending, net GIC surplus (consumer surplus minus health spending), and relative to insurer profits are also reported. FC_j is reported in millions of dollars.

The results here, unsurprisingly, differ than those presented in the main specification. The estimate of ρ increased to 5.27, implying that the GIC places a relatively higher emphasis on consumer surplus relative to premium spending and fixed costs. The estimates of fixed costs themselves has decreased dramatically from $8.4 million to approximately $1.6 million. Though a large decline, this is quite intuitive: each additional product that could have been offered but was not brings substantially less utility to consumers without the presence of the logit error. The GIC is therefore sacrificing less in terms of utility by *not* offering additional product variety, thereby reducing the estimate of FC_j.

Counterfactual Product Offerings: I use these new estimates to re-estimate the counterfactuals presented in section 6. Table H.2 reports these results. As with the main results, the model without the logit error predicts observed products very well. In fact, it predicts the products slightly better, as it matches the exact number of products observed.[74] The counterfactual results for removing physician inertia result in similar changes as the main specification with largely consistent welfare implications, albeit with one slight difference: the GIC reduces Tufts' "Medium"

[72] This is particularly true in my setting, where the employer explicitly has $CS(\delta_{Ijt}, \theta)$ in its objective function.

[73] A more sophisticated approach would be to estimate a pure characteristics model of demand for insurance, assuming away the logit error, as in Berry and Pakes (2007), Nosko (2014), and Song (2007). However, due to the complexity of this task, I take a more simplified approach and simply preserve the estimated parameter estimated from a logit model with shocks, but remove those shocks for the purposes of computing surplus. Though this approach is a simplification, it is meant to be an approximation as to what reasonable bounds on product offerings might be due to the removal of switching costs.

[74] The only discrepancy is that Tufts' narrow-network plan is predicted to be wider than its observed network.

network to a “Narrow” network. However, the models removing plan switching costs, firm fixed costs, and where the GIC chooses under the assumption that switching costs are welfare-irrelevant yield some notable changes in equilibrium plan menus.

Table H.2: Counterfactuals: Equilibrium Networks Chosen Under Varying Inertia Assumptions, 2013

Insurer	Network	Observed	Pred.	No Phys. Inert.	No Plan Inert.	No FC	Price-v-Prov.
Fallon	VN	x	x	x	x	x	x
Fallon	Broad	x	x	x		x	
HPHC	VN						
HPHC	N						x
HPHC	N(SG)				x	x	
HPHC	Med	x	x	x		x	x
HPHC	Broad	x	x	x		x	
HNE	N	x	x	x	x	x	x
NHP	N	x	x	x	x	x	x
Tufts	VN						
Tufts	N	x		x	x	x	
Tufts	N(SG)				x	x	
Tufts	Med		x		x	x	
Tufts	Broad	x	x	x	x	x	x
Total Plans		8	8	8	8	11	6
Welfare and Spending Holding Plan Menu Fixed							
ΔCS_1 (Fixed)				-\$92.39	\$11.24	-	-
ΔCS_2 (Fixed)				-\$92.38	-\$4.94	-	-
$\Delta Costs$ (Fixed)				-\$1.93	-\$10.64	-	-
ΔFC (Fixed)				-	-	-\$10.14	-
Welfare and Spending Allowing Plan Menu to Change							
ΔCS_1 (Change)				-\$92.50	\$10.31	\$0.06	-\$17.58
ΔCS_2 (Change)				-\$92.41	-\$5.86	\$0.08	-\$0.95
$\Delta Costs$ (Change)				-\$2.30	-\$38.29	-\$1.93	-\$32.97
ΔFC (Change)				\$0.00	\$0.00	-\$10.14	-\$2.54

Notes: GIC observed and predicted products offered under various counterfactual inertia assumptions. All estimates compute CS with no logit shock. “No Phys. Inertia.” refers to predicted networks when all physician loyalty is removed. “No Plan Inert.” refers to predicted networks when all plan switching costs are removed. “No FC” refers to predicted networks when employer fixed costs are removed. “Price-v-Prov.” refers to predicted networks when plan switching costs are presented, but when the GIC considers them welfare-irrelevant, i.e. considers only the price-versus-provider-choice tradeoff when making product variety decisions. “ΔCS_1 refers to change in consumer surplus per-household-per-month when plan switching costs are considered tangible welfare-relevant costs. “ΔCS_2” are consumer surplus estimates when switching frictions are considered welfare-irrelevant. $\Delta Costs$ (fixed)” refer to the change in consumer surplus per-household-per-month and change in total GIC costs per-household-per-month assuming the plan menu remains fixed. “$\Delta FC'$” refer to changes in firm fixed costs.

In particular, under all the scenarios, the GIC offers fewer products than they did in Table 5 and in some cases fewer than baseline. However, the logic to how the GIC responds is similar to that in the main specification. The main difference with the removal of the logit shock is that the GIC continues to preserve network *variety*, but does so among fewer *insurers*. In particular, among all scenarios, the GIC no longer finds it worth it to offer broad Harvard products. Consider, for instance, the case of removing plan switching costs. In the main specification, this resulted in the GIC adding a significant number of products to its menu, as it became worthwhile to offer consumers increase choice among a variety of network sizes. In this specification, however, consumers see less benefit from having that variety offered among multiple competing insurers, as consumers no longer benefit from simply having an increased number of plans. As a result, the GIC continues to offer more differentiated networks, but only among Tufts' plans. In addition, the GIC removes nearly all of Harvard's plans except one narrow plan, and removes Fallon Broad as well.[75] Removal of switching costs in this specification yields to an \$10 increase in CS_1 per household per month, far smaller than in the main specification. CS_2 actually *declines* by about \$5 per household per month, suggesting that if switching costs are considered welfare-irrelevant, the GIC's new plan menu actually decreases well-being.[76] Nevertheless, costs decrease by a substantial \$38 per household per month. The implication is that social welfare increases whether or not switching costs are considered welfare-relevant and despite the loss of Harvard plans.

Similar patterns can be seen in the removal of GIC fixed costs. Under this scenario, it offers more products than at baseline. Unlike the main specification, however, the welfare change under this scenario is is essentially a wash. Since consumers no longer benefit simply from increased *numbers* of plans and since consumers retain their switching costs, the added number of plans does very little to benefit consumers.

Finally, the largest difference from the main specification occurs when the GIC considers switching costs welfare-irrelevant when making plan choice decisions. Here, rather than increasing the number of plans, the GIC actually decreases the number of plans from eight to six. Similar to the "no plan switching cost" scenario, it removes Fallon Broad and Harvard Broad. In addition, it removes a Tufts narrow network and adds a Harvard narrow network. When switching costs are present, but the GIC does not consider them welfare-relevant in its objective function, this explicitly makes the cost of offering a variety of broad-network products not worth it. The GIC optimizes by removing those products and forcing consumers into narrower networks. Moreover, since those switching frictions remain in place, as do the fixed cost, offering additional variety of Tufts plans no longer becomes worth the cost. This move results in a decrease in consumer surplus of \$18 per month if switching costs are, in fact, tangible. If, on the other hand, switching costs are welfare-irrelevant, surplus only decreases by \$1 per household per month. Net costs, however, decline considerably by \$33 per month due to the removal of the broad networks, leading to sizable social surplus gains, even as consumers largely do not see the benefit.

Overall, the implications of the results with no logit error, despite some notable differences, are very similar to the main specification. Primarily, it is clear that the GIC internalizes consumer switching costs when making product choice decisions, leading to an overprovision of broad-networks and and underprovision narrow-network plans. The result of this friction is to keep consumers enrolled in plans they otherwise might not be, leading to significant increases in premiums and reductions in welfare. Removing these switching costs forces the GIC to largely abandon its broad-network designs. However, it still offers network *variety*, especially among Tufts plans. This variety partly compensates consumers for the loss of flagship broad-networks and, moreover, leads to significant cost decreases that achieve net increases in social surplus.

[75] The GIC expands its variety for Tufts for two primary reasons. The first is that Tufts is the dominant-share insurer in 2013, and as such yields a larger brand effect over Harvard and Fallon. Second, Tufts' negotiated rates with insurers is more favorable than Harvard, allowing it offer lower premiums for similar networks.

[76] This comes from two sources. The first, similar to the main specification, is the resulting increase in premiums for members who don't switch and the adverse selection into Tufts Broad. The second is from the loss of Harvard plans. Note that consumers, under the new product menu, have access to a wide variety of networks from Tufts. Therefore, the loss attributed to losing access to Harvard is due entirely to the brand effect.