ONLINE APPENDIX

Employer Incentives and Distortions in Health Insurance Design: Implications for Welfare and Costs

Nicholas Tilipman[1]

Data Descriptions

A1. APCD Sample Creation

Hospital Admissions: I first create a sample of hospital admissions, which I use to estimate patient demand for hospitals. To do so, I limit the APCD to any facility claim flagged as an inpatient admission within my 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 use the organization's National Provider Identification (NPI) number to match the hospital to a set of hospital characteristics from the American Hospital Directory (AHD) database (American Hospital Directory, 2013). 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 are aggregated to the hospital admission level and the "allowed amounts" are summed over all service lines for that particular admission 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 (Healthcare Cost and Utilization Project, 2015).

Table A.1 contains the hospital sample summary statistics for hospital admissions from 2009-2013. On average, patients admitted to Massachusetts hospitals are 45 years old, and about half of the patients suffer from a chronic condition. Approximately 7% of patients are admitted with a primary cardiac condition, while about 31% are admitted with an obstetrics-related diagnosis. Patients are, on average, willing to travel approximately 13 miles to visit a hospital, and visit teaching hospitals approximately 80% of the time, while visiting academic medical centers approximately 37% of the time.

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

Table A.1—: Hospital Sample Summary Statistics

	Mean	Std Dev
Patient Characteristics		
Age	45.08	22.94
Female	0.67	0.47
Chronic	0.53	0.50
Neurological	0.02	0.12
Cardiac	0.07	0.26
Obstetrics	0.31	0.46
Imaging	0.26	0.44
Hospital Characteristics		
Distance	12.62	14.07
NICU	0.89	0.31
Neuro	0.95	0.22
MRI	0.94	0.24
Critical Access	0.01	0.10
Teaching	0.80	0.40
Specialty	0.05	0.22
Academic Medical Center	0.37	0.48

Notes: Hospital sample summary statistics 2009-2013. Diagnosis characteristics (e.g. Neurological, Cardiac, etc.) are derived from AHRQ's Clinical Classifications Software categories and Chronic Conditions Indicators

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 (PCPs), cardiologists, and orthopedists. PCP practices are defined as any medical group that contains at least one physician who is either an internist, general practitioner, family practice doctor, or geriatric doctor. Similarly, cardiology practices and orthopedic practices are defined as any practice that employs at least one physician of the relevant specialty. I consider these three specialties in order to capture three different components 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, and surgical care.

For each service line, I merge in Medicare Part B physician fee schedules from the Center for Medicare and Medicaid Services (CMS) (Center for Medicare and Medicaid Services, 2009). These data contain annual federal updates to each procedure (CPT) code's "Relative-Value-Unit" (RVU) weight, which are constructed to assign each service a measure capturing its relative resource intensity to other procedures. These weights are used by CMS to determine Medicare payment rates for physicians. As such, I use them both a proxy for procedure intensity and in the construction of insurer-physician negotiated rates, described further in subsection C.C6. 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.

Table A.2 shows summary statistics for the physician samples. On average patients going to see 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 that are more resource-intensive and thus assigned higher RVUs. About 85% of primary care patients saw a doctor between 2009 and 2013 that they also have seen previously, while this number was about 64% for cardiologists and about 61% for orthopedists. Distance traveled to PCPs was about 6 miles, on average, and about 10 miles for cardiologists or orthopedists. When seeing a PCP, patients on average visit practices with 41 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 owned by hospitals or owned by health systems.

Table A.2—: Physician Sample Summary Statistics

	PCPs	Cardiologists	Orthopedists
Age	47.92	54.12	44.36
	(15.59)	(13.87)	(18.52)
Female	0.57	0.43	0.52
	(0.50)	(0.49)	(0.50)
RVU	2.61	2.96	5.55
	(1.64)	(4.90)	(12.56)
Used Doc Previously	0.85	0.64	0.61
	(0.36)	(0.48)	(0.49)
Used Med Grp Previously	0.86	0.70	0.65
	(0.35)	(0.46)	(0.48)
Used System Previously	0.86	0.74	0.67
	(0.34)	(0.44)	(0.47)
Distance	5.57	9.54	9.69
	(5.55)	(10.99)	(10.42)
Doctors on Site	41.48	116.86	65.25
	(105.00)	(180.55)	(143.49)
Number of Locations	8.89	9.96	5.51
	(8.63)	(9.29)	(8.31)
Part of Medical Group	0.72	0.72	0.63
	(0.45)	(0.45)	(0.48)
Owned by Hospital	0.26	0.43	0.20
	(0.44)	(0.49)	(0.40)
Owned by System	0.52	0.59	0.32
	(0.50)	(0.49)	(0.47)

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.

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 five-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. Finally, I merge this list of GIC members to external, publicly available data on GIC annual plan premiums and hospital networks. 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.

A2. 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 ensured 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 merge the data by provider name. 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 a 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 provider 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 (e.g. they may 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 I code this "ownership" variable as that group. If the highest level of ownership is a particular hospital (i.e. a hospital-owned physician practice), then this "ownership" variable is coded as that hospital. Finally, if the highest level of ownership is reported as a health system (e.g. Partners Health Care, Steward Health System), then this "ownership" variable is coded as that system. This variable is used primarily in constructing networks (see below).

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 (e.g. Fallon), others only report the list of hospitals. I therefore use the "ownership" variable defined above. I 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), then any physicians or groups that are owned by that system are also excluded.[2] 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 in a particular year is considered in that plan's network. For robustness, I also construct networks that default each of these small practices to being in the plan's network *unless* a majority of claims from a particular plan in a particular year are explicitly flagged as being "out of network."

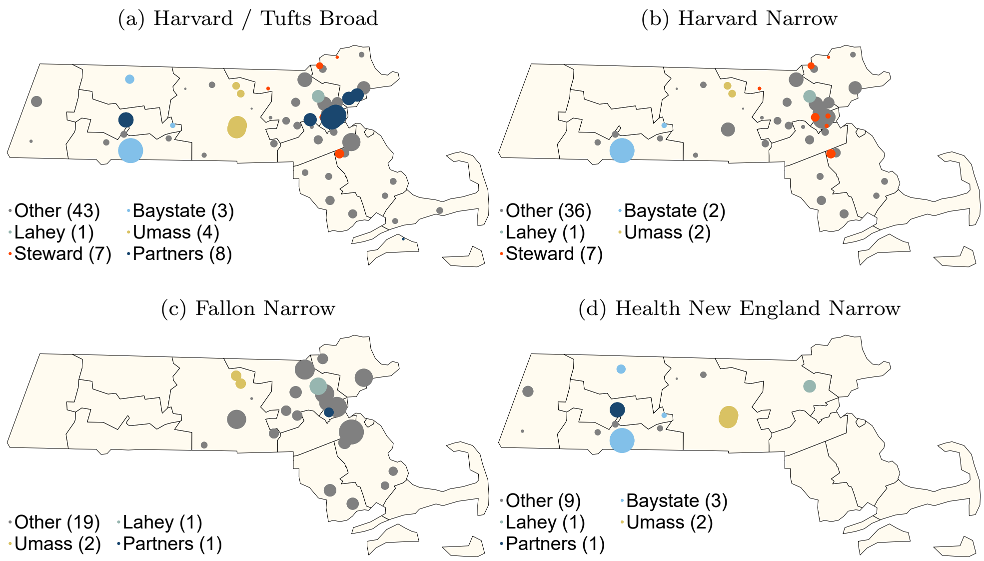
[2]In practice, this is a close approximation of contracts observed on the GIC. Harvard Primary Choice and Tufts Spirit, for instance, cease contracting with all Partners-owned medical groups as well as Partners hospitals.

Additional Descriptives

B1. Additional Network Figures

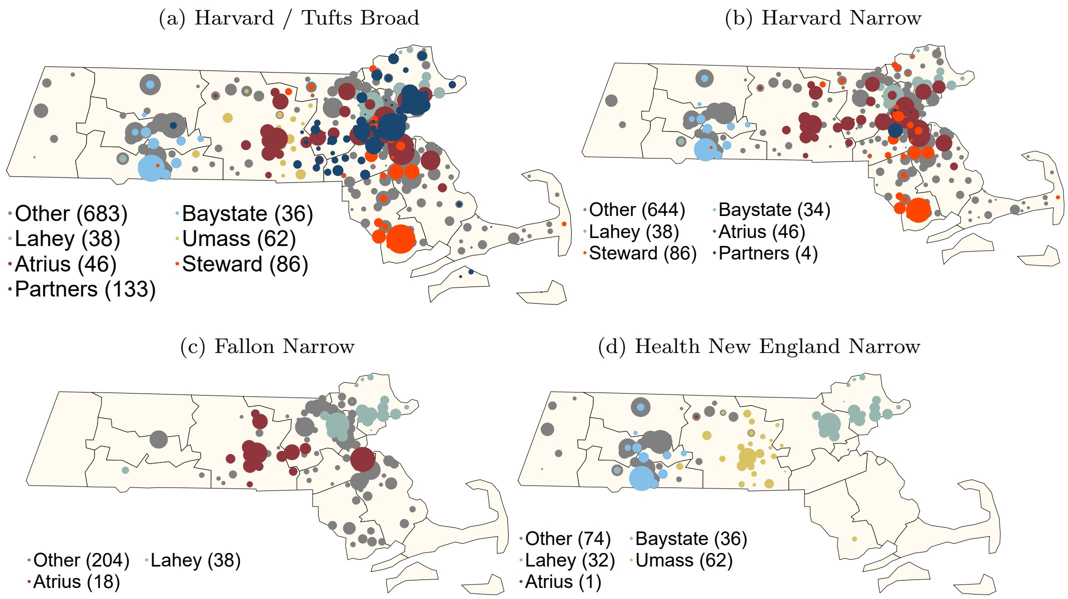
In Figure B.1, Figure B.2, Figure B.3, and Figure B.4, I present additional maps depicting the hospital, PCP, cardiology, and orthpedic practice network coverage across Massachusetts of Harvard and Tufts Broad, Harvard Narrow, Fallon Narrow, and HNE Narrow.

Figure B.1. : Hospital Networks by Plan, 2013



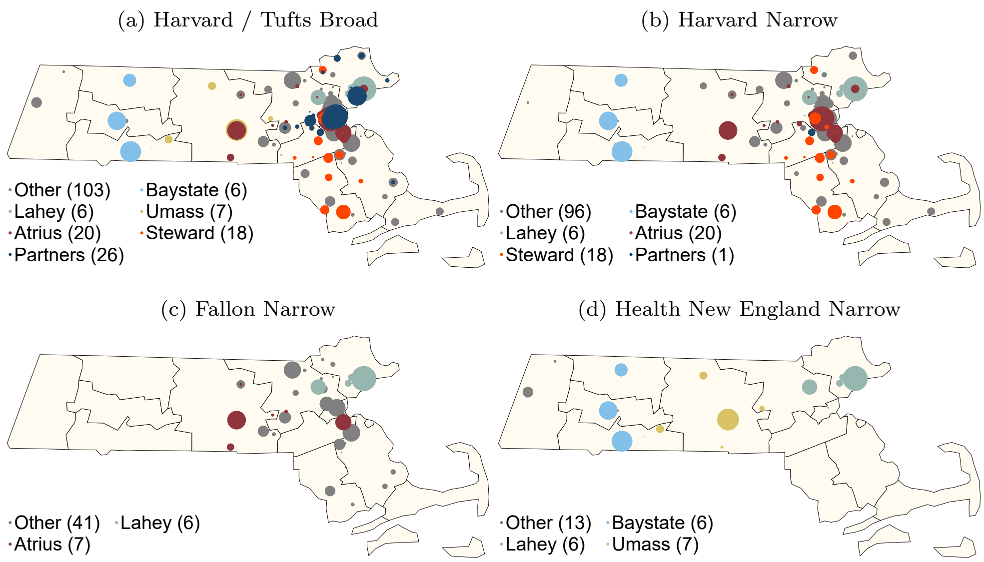
Notes: This figure plots the hospital networks of specified plans on the GIC in 2013. Sizes of the data points reflect relative market shares of the hospitals. Colors reflect ownership status (which health systems owns which hospital).

Figure B.2. : Primary Care Practice Networks by Plan, 2013



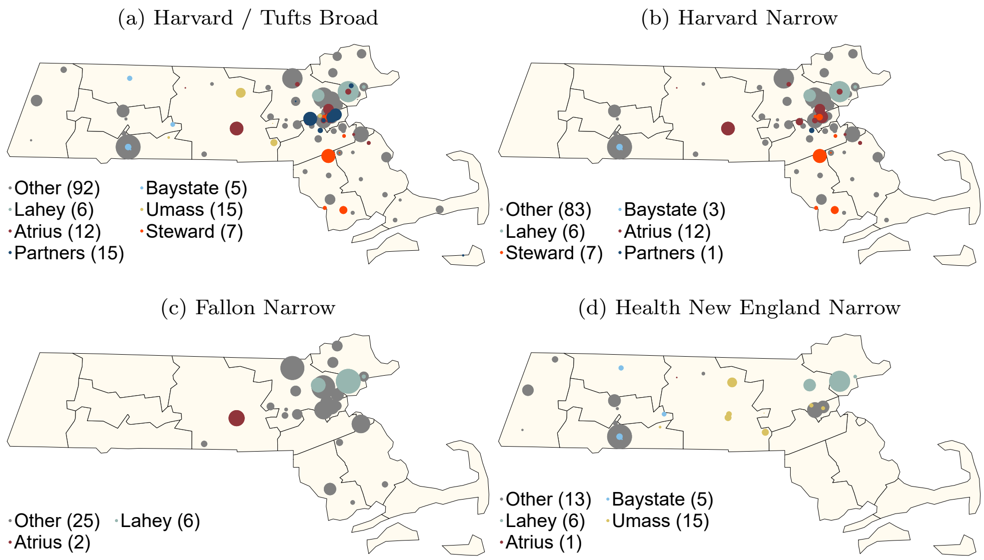
Notes: This figure plots the PCP practice networks of specified plans on the GIC in 2013. Sizes of the data points reflect relative market shares of the practices. Colors reflect ownership status (which health systems owns which practice).

Figure B.3. : Cardiology Networks by Plan, 2013



Notes: This figure plots the cardiology practice networks of specified plans on the GIC in 2013. 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. : Orthopedic Networks by Plan, 2013



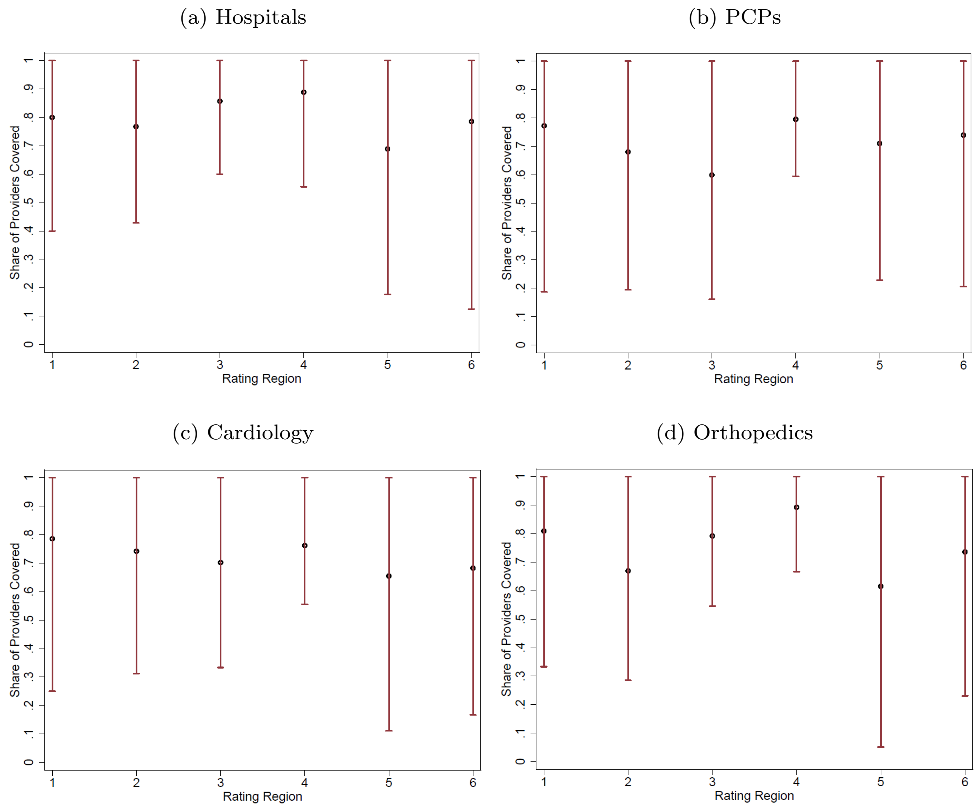
Notes: This figure plots the orthopedic practice networks of specified plans on the GIC in 2013. Sizes of the data points reflect relative market shares of the practices. Colors reflect ownership status (which health systems owns which practice).

Figure B.5 plots the variation in hospital and physician networks across plan and rating region in Massachusetts in 2011.[3] The y-axis represents the share of providers operating in the rating region that each plan covers as an in-network provider (hereafter referred to as "network breadth"). The dots represent the unweighted average network breadth across the plans on the GIC that operate in the respective rating regions, and the bars represent the range of network breadth in that region. For physicians, providers were limited to just the top 50 practices (by number of claims) in each rating region, to avoid measurement error. For each specialty, there is considerable variation in network breadth, both across and within rating region. Across rating regions, *average* unweighted network breadth for PCPs, for instance, ranges from about 60% to about 80%. Within rating region, the broadest plans cover virtually all the top 50 practices and hospitals, while the narrowest plans cover only about 20% of the providers. In Rating Region 5, which includes Boston, average network breadth for hospitals, cardiologists, and orthopedists is quite low, reflecting the fact that many of the narrow-network plans exclude providers in the Boston region. Noticeably, the narrowest plan operating in the region only covers about 10% of the top 50 orthopedic practices

[3]Rating regions are defined according to CMS definitions: https://www.cms.gov/CCIIO/Programs-and-Initiatives/Health-Insurance-Market-Reforms/ma-gra. Rating Region 7 (Cape Cod) is omitted from analysis due to the low number of households on the GIC residing in this region.

in the region.

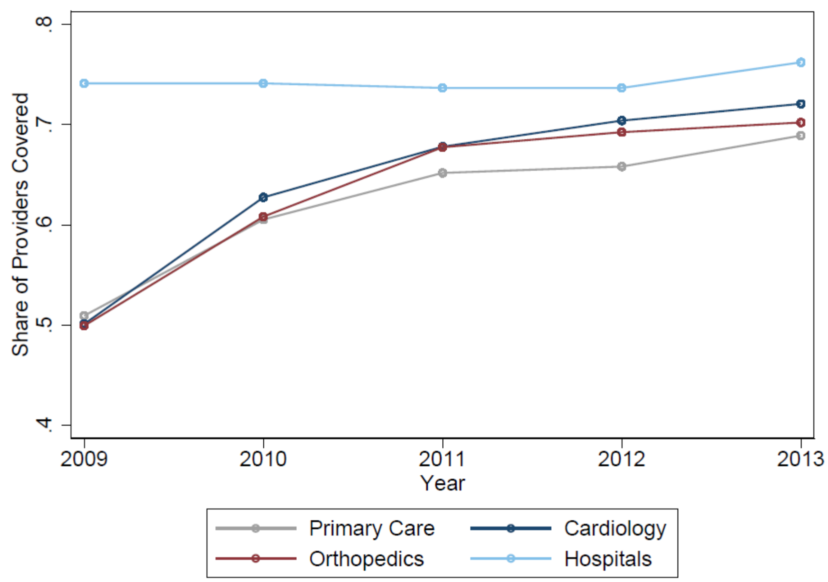
Figure B.5. : Share of Providers Covered by Rating Region and Specialty, 2011



Notes: This figure plots the share of all hospital and physician practices covered by each plan on the GIC by rating region. Each dot represents the unweighted average share of providers in the respective rating covered across all plans operating in those regions. Red bars represent the range of coverage across plans in that region. For PCP, cardiology, and orthopedic networks, data is limited to the top 50 practices (by number of claims) in each rating region.

Figure B.6 displays the unweighted average network breadth over time. The y-axis here represents the share of the state's hospitals and physician practices covered, averaged across all plans operating statewide. This again limits the data to only the top 50 physician practices for each specialty in each rating region. While hospital networks remain fairly stable over time, with the exception of a small uptick in 2013, the network breadth for the three physician specialty groups seem to be increasing over time, ranging from about 50% of physicians covered in 2009 to between 65% and 70% coverage in 2013. This change is driven primarily by three factors. First, during this time period there were some physician exits, as well as mergers between physician groups that resulted in a change in network status.[4] Second, during this period there were significant hospital acquisitions of physician practices. Third, certain narrow plans grew more generous in coverage over time.[5]

Figure B.6. : Share of Providers Covered by Year and Specialty



Notes: This figure plots, by year, the unweighted average share of all hospital and physician practices covered across plans operating statewide on the GIC. For PCP, cardiology, and orthopedic networks, data is limited to the top 50 practices (by number of claims) in each rating region.

B2. Additional Evidence of Inertia

In Table B.1 I present a regression of the probability of enrollment in a narrow-network plan against 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, 8% less likely to be enrolled in a narrow-network plan than new members, suggesting that plan choice inertia may play a large role in explaining broad-network enrollment.

[4]As an example, the Atrius Health system gradually purchased several prominent medical groups,

Table B.1—: Probability of Enrolling in a Narrow Plan

Variable	Coefficient	Standard Error
Existing GIC Member	-0.0807^{***}	0.0015
Age	-0.0004^{***}	0.0000
Female	-0.0267^{***}	0.0015
Chronic Condition	-0.0318^{***}	0.0018
Members in HH	-0.0103^{***}	0.0004
Constant	0.2773^{***}	0.0024
Year FE	Yes	
County FE	Yes	
Obs.	426,288	
Adj R^2	0.27	

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

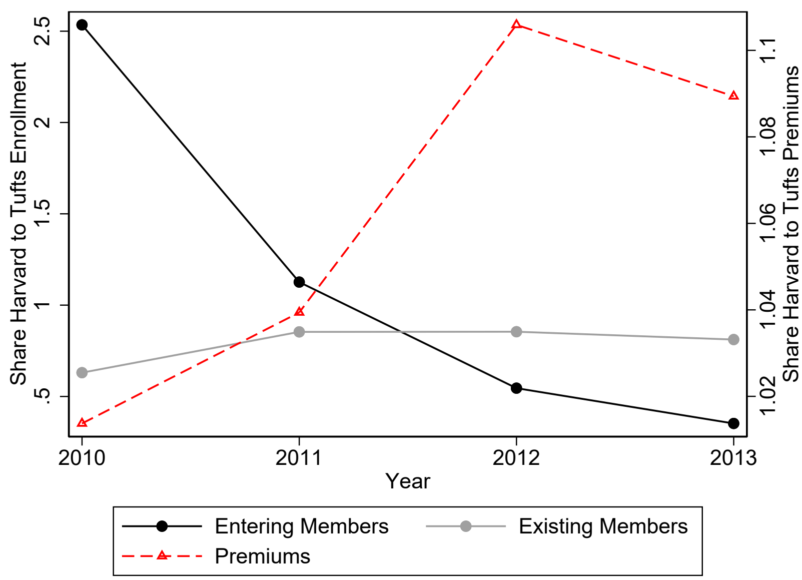
To see more evidence that new cohorts behave differently than existing cohorts, I also report the stickiness of enrollment in broad-network plans as the characteristics of those plans change. To that end, I note that in 2010 the premiums for Harvard and Tufts were fairly similar. However, beginning in 2011, the premium difference between the two plans began to rise, with Harvard Broad growing significantly more expensive than Tufts Broad.

Figure B.7 shows the ratio of enrollment in Harvard Broad versus Tufts Broad over time, along with the ratio of Harvard Broad premiums to Tufts 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 Harvard's premiums rise relative to Tufts', enrollment in Harvard declines dramatically relative to Tufts among new members. By 2012, Tufts' premiums were about 10% less than Harvard's (amounting to about $30 per month for families). Enrollment in Harvard among new members, meanwhile, declined from almost three times that of Tufts in 2009 to about 50% 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.

including Harvard Vanguard and the Fallon Clinic, which were previously separate entities.

[5]For example, Fallon Health Plan did not cover the Partners system until 2013.

Figure B.7. : Share of Members Enrolling in Harvard Broad vs. Tufts Broad by Whether New to GIC



Notes: This figure plots the ratio of members selecting Harvard's broad-network plan over Tufts' broad-network plan as well as the ratio of Harvard to Tufts premiums. 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.

Model Details

C1. Provider Demand Estimation

Market Shares: The probability that patient i with diagnosis l will choose hospital h in time t is given by:

$$\sigma_{ilht} = \frac{exp(\phi_{ilht})}{\sum_{k=1}^{N_{jt}^{H}} exp(\phi_{ilkt})} \tag{C1}$$

where N_{jt}^{H} refers to the hospitals in plan j'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:

$$\sigma_{irdt}^{s} = \frac{exp(\phi_{irdt}^{s})}{\sum_{k=1}^{N_{jt}^{S}} exp(\phi_{irkt}^{s})} \tag{C2}$$

where N_{jt}^{S} is the network of practices of type s in plan j's network.

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, described below. Further, I estimate the models separately by the seven Massachusetts rating regions and by specialty group (PCP, cardiology, and orthopedics).

All models include patient characteristics interacted with provider characteristics, travel time interacted with both patient and provider characteristics, and a full set of provider fixed effects 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 Classifications 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.[6] 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 interact hospital fixed effects with the CCS categories.

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 PCPs), 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.

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 interact each inertia variable with patient RVU as well as with a proxy measure for the length of a particular patient-provider relationship. To construct this, I infer from the claims the earliest visit a particular patient had with a particular provider, and calculate the number of years to the present day.

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.

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). I address these potential selection concerns in subsection C.C4.

Dimensionality Reduction Perhaps the most salient issue in estimation of the physician models is the presence of tens 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

[6]The construction of these weights follow closely to work by Shepard (2016); a discussion of their construction can be found in subsection C.C6.

the dimensionality problem. The first is that 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 individual practices are location-specific, this allows me to include a larger span of the full physician practice space in my estimation. In addition, it allows for estimation of flexible parameters that vary by region.

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. $\{T_{id} \in b\} = T_{ib} = b^{mid}$. As an example, $b^{mid} = 2.5$ for distance band $b =0$ to 5 miles. Given these assumptions, and dropping the region and time subscripts for convenience, the model in Equation 1 becomes:

$$u^s_{ird} = \underbrace{\phi^s_{ird} + \varepsilon^s_{ird}}_{\text{Utility for Top 50 Practices}} \tag{C3}$$

$$u^s_{ird} = \underbrace{\sum_b \mathbb{1}\{T^s_{id} \in b\}(T^s_{ib}\lambda^s_1 + T^s_{ib}v_{ir}\lambda^s_2 + N^s_{ib}\gamma^s_b) + \varepsilon^s_{irb}}_{\text{Utility for Practices Outside Top 50}} \tag{C4}$$

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 valuations of these options to vary by the number of doctors in those groups. As an example, if patient i's physician network removed a physician practice in distance band b, patient i's utility would decrease by γ^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, two empirical facts support this claim. First, the top 50 practices by market share in a given region account for most of patient claims.[7] Second, most practices outside

[7]In Boston, for instance, where there is the highest density of physicians, the top 50 PCP practices

the top 50 are included in all plans' networks. As a result, most of the variation in networks across plans comes from variation in coverage of 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 tractable, but also is likely a reasonable assumption given this context.

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 the outside option to 0 in all models.

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 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. I abstract away from decomposing the extent to which λ_5^s is driven by true switching costs as opposed to unobserved preference heterogeneity. In particular, persistence in physician choice may be driven by three factors: physician-patient capital accumulated through repeated interactions (i.e. the patient *develops* utility for a particular physician ex-post); unobserved physician quality (i.e. the patient stays with the physician for factors unobserved to the econometrician); and true switching frictions or hassle costs irrespective of physician quality. In my setting, I choose to focus on the most conservative interpretation of physician inertia: that λ_5^s entirely reflects physician-patient capital. In counterfactual exercises, when patients lose access to their previously used physicians or practices, I treat this as a "welfare-relevant" utility loss.[8] However, to test the robustness of this assumption, in Appendix E I also present estimates of the key parameters of my employer objective function that treat this inertia term as being driven by the two other aforementioned sources.

C2. Hospital Demand Estimates

Table C.1 reports the results for the hospital demand model. Column 1 displays the main results, which are run on the full sample of hospital admissions in Massachusetts between 2009 and 2013. These models incorporate flexible distance

account for approximately 70% of all claims, while the top 50 cardiology and orthopedic practices account for nearly 90% of all claims.

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

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 reported are for Barnstable County (the base county), Worcester (Central Massachusetts), Hampden (Western Massachusetts), and Suffolk (Eastern Massachusetts). Consistent with prior literature, the distance coefficients are negative and significant in all reported counties, implying that patients prefer to go to hospitals that are close to where they live. Notably, patients are far less reactive to distance in Barnstable, Hampden, and Worcester (where they are more likely to drive by car to find a hospital) than they are in Suffolk (which contains metropolitan Boston). While these coefficients are difficult to interpret (the measure is in utils instead of a dollarized amount), comparing them with other parameter estimates shed some light on their practical magnitude. For instance, the estimates imply that hospital patients in Suffolk are, on average willing, to travel approximately 25 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 previously. This "willingness-to-travel" to a hospital the patient has previously used varies by county, conditional on age, disease, and hospital characteristics. The estimates imply that consumers in Barnstable, for instance, are willing to travel an additional 13 miles on average 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.

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 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 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 MRIs.

It is worth mentioning that this model so far omits copayments that plans charge

Table C.1—: Results of Hospital Demand Model

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 R^2	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.

to visit different hospitals. On the GIC, plans are differentiated in their premiums, their networks, and the copays that patients pay for a hospital admission across *plans*, across *hospitals*, and over time (Prager, 2016). In 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.

C3. Physician Demand Estimates

Table C.2 reports the results of the physician demand models for PCP practices, cardiology practices, and orthopedic practices. 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. I omit year 2009, the earliest year of data in the claims, as I cannot observe patients' prior use of physicians in that year. 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. Table C.3 shows analogous parameter estimates for the Worcester region, for comparison.

Consistent with the results of the hospital demand model, distance plays an extremely important role in patient choice of physician. Across the three specialist groups, distance has a negative and significant effect on utility.[9] Patients, on average, prefer visiting practices owned by hospitals or health systems, though the effect is considerably stronger for cardiology practices.[10]

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 latter is consistent with the result from hospital demand, namely that conditional on risk, older individuals prefer to travel smaller distances to seek care, particularly for routine primary care treatment. For cardiologists and orthopedic

[9] 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 scaling of the logit error.

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

Table C.2—: Results of Physician Demand Models (Boston)

Variable	PCP Practices	Cardiology Practices	Orthopedic Practices
Distance	−0.4168***	−0.2994***	−0.2575***
	(0.0186)	(0.0147)	(0.0159)
Owned by Hosp. or System	0.2813***	1.4316***	0.6287***
	(0.0977)	(0.0895)	(0.0867)
Used Prac Previously	3.6494***	1.0508***	2.4033***
	(0.0401)	(0.0354)	(0.0406)
x Length of Relationship	0.3751***	0.1947***	−0.1962***
	(0.0097)	(0.0104)	(0.0118)
x RVU	−0.0643***	0.0646***	0.0090***
	(0.0094)	(0.0053)	(0.0019)
Used Med Grp Previously	1.4447***	1.6987***	1.5252***
	(0.0401)	(0.0345)	(0.0431)
Used System Previously	0.5677***	0.8538***	1.0928***
	(0.0362)	(0.0303)	(0.0356)
		Interactions with Patient Characteristics	
DistxFemale	0.0004	−0.0046	−0.0064*
	(0.0015)	(0.0029)	(0.0034)
DistxAge	−0.0008***	0.0000	0.0001
	(0.0000)	(0.0001)	(0.0001)
DistxChronic	−0.0015	0.0297***	0.0261***
	(0.0019)	(0.0079)	(0.0055)
		Interactions with Provider Characteristics	
DistxNumDocs	0.0004***	−0.0004***	−0.0002***
	(0.0000)	(0.0000)	(0.0000)
DistxNumLocs	0.0036***	−0.0036***	−0.0054***
	(0.0009)	(0.0006)	(0.0007)
DistxMedGrp	0.0260***	−0.0381***	−0.0507***
	(0.0075)	(0.0080)	(0.0074)
AgexNumDocs (00s)	−0.0020***	0.0045***	0.0000
	(0.0000)	(0.0000)	(0.0000)
AgexNumLocs (00s)	0.0192	0.1423***	0.0791***
	(0.0165)	(0.0133)	(0.0149)
AgexMedGrp	0.0027*	0.0102***	0.0243***
	(0.0014)	(0.0015)	(0.0014)
Practice FE	Yes	Yes	Yes
Obs.	3,289,932	1,853,631	1,634,164
Pseudo R^2	0.64	0.59	0.57

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 at the practice, share of physicians at the practice who are specialists, number of doctors across the entire system, and number of practices owned by the system, Model contains a full set of practice fixed effects. Note that AgexNumDocs and AgexNumLocs are reported in hundreds (00s). Length of relationship is measured in years.

practices, 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.

Patients seeking primary care are willing to travel farther to access practices with more physicians on site. In addition, they are willing to travel farther 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.[11] This is particularly pronounced for cardiologists, where the age effect on visiting larger practices is considerably larger than the other specialty groups.

All three of the physician inertia indicators are highly predictive of 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 an individual, on average, would be willing to travel an additional 8.5 miles to access the same PCP practice, 5.8 miles to access the same cardiology practices, and 9.5 miles to access the same orthopedic practices. The magnitudes are quite similar to the magnitudes in the hospital demand model. The stickiness to previously used providers also varies significantly with patient health and the length of the patient-provider relationship. For PCPs and cardiologists, the longer a patient has been seeing a physician, the more likely they are to use the physician again next time. For orthopedic practices, this is reversed: the longer time has elapsed since the first time seeing the provider, the *less* likely a patient is to see that orthopedist again. This may be driven by the short-term nature of orthopedic care, which tends to more often than PCPs or cardiologists treat specific injuries on a one-off basis. For cardiologists and orthopedists, patients needing more intensive procedures (i.e. those who have higher RVU weights) are more likely to use physicians they have used in the past. However, this is not the case for PCPs, where those who have more intensive needs are likely to see a new PCP. Altogether, these results imply that inertia to previously used physicians play a significant role in provider choice.

For comparison, Table C.3 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 are on

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

average willing to travel an additional 8.5 miles to access the same PCP practice, this figure is approximately 30 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.[12] 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 reflective of the fact that, unlike Boston, Worcester has fewer prestigious academic medical centers.[13] Much like in Boston, older patients seeking specialist care significantly prefer doctors that are part of medical groups and that work for practices which have multiple locations.

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

[13]Worcester does, however, contain a prominent medical group: the Fallon Clinic (later renamed Reliant Medical Group).

Table C.3—: Results of Physician Demand Models (Worcester)

Variable	PCP Practices	Cardiology Practices	Orthopedic Practices
Distance	−0.1546***	−0.1595***	−0.2210***
	(0.0118)	(0.0104)	(0.0107)
Owned by Hosp. or System	−0.1576**	0.1139	−0.2913***
	(0.0690)	(0.0970)	(0.0902)
Used Prac Previously	4.6802***	1.3632***	3.1657***
	(0.0446)	(0.0525)	(0.0657)
x Length of Relationship	0.2562***	0.0940***	−0.3311***
	(0.0118)	(0.0185)	(0.0248)
x RVU	−0.1352***	0.0192***	0.0106***
	(0.0114)	(0.0053)	(0.0031)
Used Med Grp Previously	0.6712***	1.2439***	1.0773***
	(0.0416)	(0.0532)	(0.0629)
Used System previously	0.7626***	0.8348***	1.0247***
	(0.0411)	(0.0512)	(0.0594)
	Interactions with Patient Characteristics		
DistxFemale	0.0018	−0.0004	−0.0029
	(0.0011)	(0.0025)	(0.0028)
DistxAge	−0.0003***	−0.0002*	0.0002**
	(0.0000)	(0.0001)	(0.0001)
DistxChronic	0.0071***	0.0182***	0.0519***
	(0.0017)	(0.0056)	(0.0046)
	Interactions with Provider Characteristics		
DistxNumDocs	−0.0001*	0.0000	−0.0002***
	(0.0001)	(0.0000)	(0.0000)
DistxNumLocs	−0.0099***	−0.0018***	0.0005
	(0.0013)	(0.0006)	(0.0007)
DistxMedGrp	0.0340***	0.0033	0.0021
	(0.0069)	(0.0059)	(0.0056)
AgexNumDocs (00s)	−0.0036**	−0.0026	0.0149***
	(0.0018)	(0.0024)	(0.0026)
AgexNumLocs (00s)	0.3956***	0.1837***	0.0188
	(0.0373)	(0.0405)	(0.0421)
AgexMedGrp	−0.0024	0.0109***	0.0098***
	(0.0023)	(0.0032)	(0.0022)
Practice FE	Yes	Yes	Yes
Obs.	2,662,897	686,687	560,253
Pseudo R^2	0.60	0.62	0.62

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 at the practice, share of physicians at the practice who are specialists, number of doctors across the entire system, and number of practices owned by the system, Model contains a full set of practice fixed effects. Note that AgexNumDocs and AgexNumLocs are reported in hundreds ($00s$). Length of relationship is measured in years.

C4. Selection on Unobservables in Provider Demand

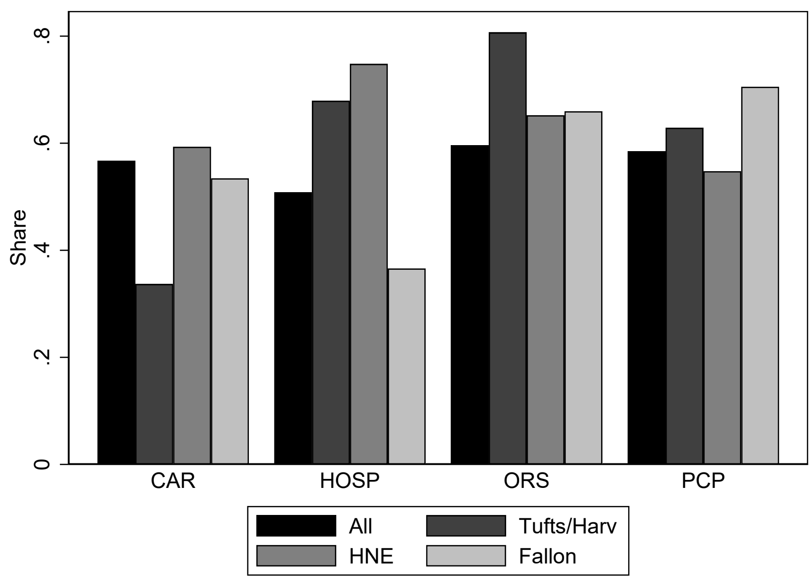
A concern with two-part multinomial logit demand models of the type presented in section II is that they may suffer from a problem with selection on unobservables as a consequence of being estimated separately. Due to the fact that the models condition on the hospital and physician networks of each patient i at time t, N^H_{jt} and N^S_{jt}, the expected utility of a particular hospital and physician network, EU^H_{Ijt} and EU^S_{ijt} (defined below), 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, Rosenbaum and Wilson, 2019).

I present here some reduced-form evidence suggesting that such selection is not a major concern in my setting. Figure C.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. The model also predicts hospital choices quite well, with a particularly good fit for patients in Health New England and the Tufts/Harvard narrow networks. The model does extremely well for orthopedic surgeons, predicting nearly 80% of choices accurately for Tufts/Harvard. However, the model performs slightly worse for cardiologists for Tufts/Harvard, and somewhat worse for hospitals for patients enrolled in Fallon.

In addition, Figure C.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 predicting accurate market shares in 2009, but then converges for every year after 2010.[14] 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

[14]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).

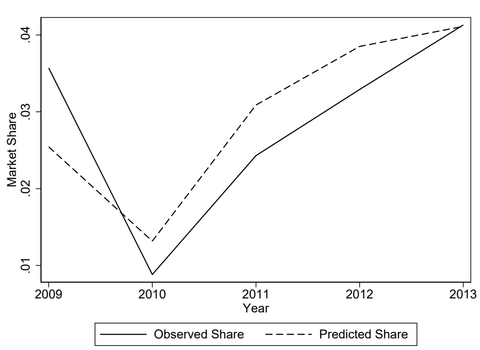
Figure C.1. : Share of Actual Choices Accurately Predicted, by Specialty



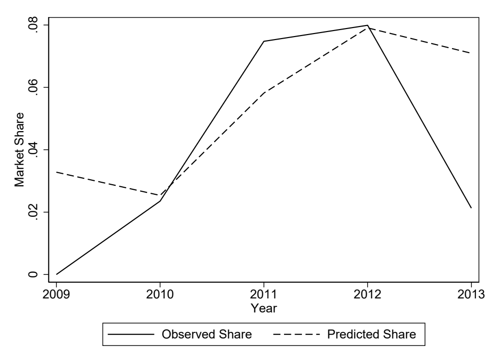
Notes: This figure plots the share of choices of providers made by individuals in narrow-network plans that are accurately predicted. Parameters used for prediction were estimated from a demand model among only individuals in broad-network plans. "Tufts/Harv" refers to Tufts Narrow and Harvard Narrow.

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

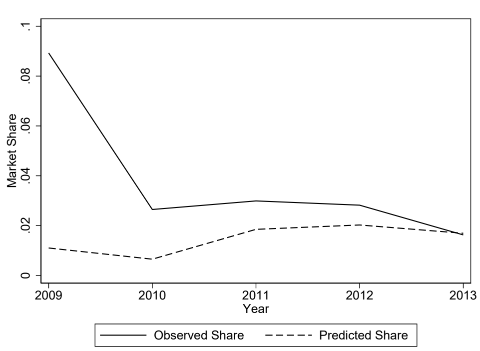
(a) Tufts Medical Center Share



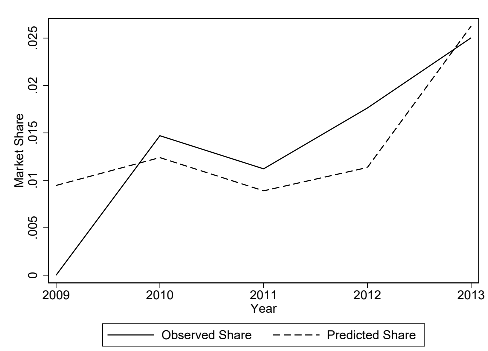
(b) Beth Israel Hospital Share



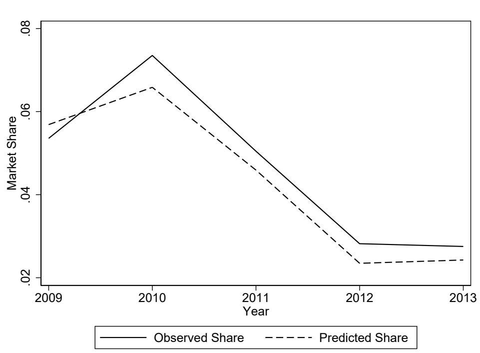
(c) Boston Medical Center Share



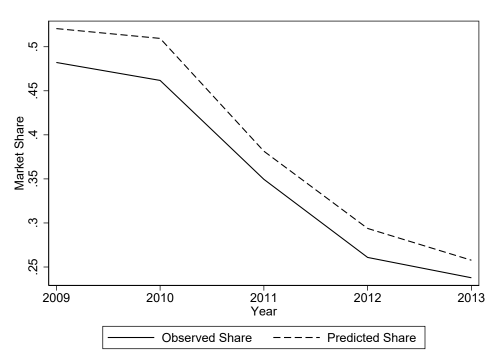
(d) Lahey Hospital Share



(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.

Berkshire and Baystate medical centers, both of which are located in Eastern Massachusetts.

C5. Plan Demand

Construction of EU_{Ijt}: I define the expected utility for hospitals and physicians, respectively, as:

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

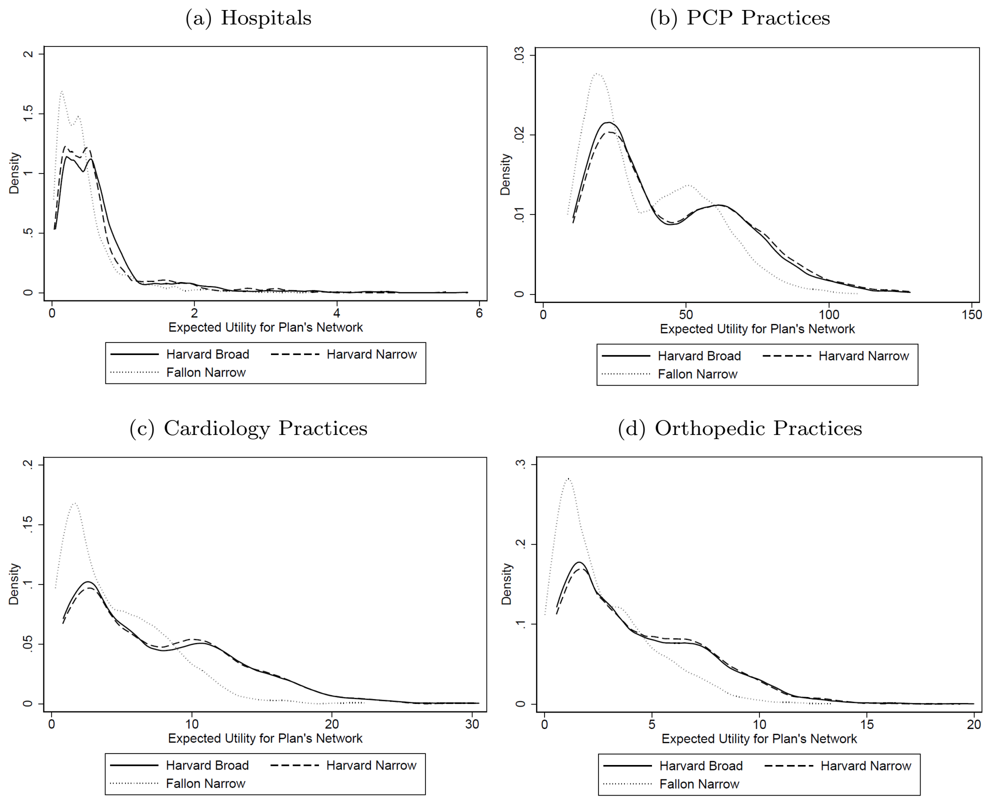
$$EU^s_{Ijt} = \sum_{i \in I} \sum_{r} f_{ir} log \left(\sum_{d \in N^S_{jt}} exp(\phi^s_{irdt}) \right)$$

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 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.

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 into the probability of falling into discrete RVU bins within each specialty: 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 complexity.

Figure C.3 plots the density of each household's expected utility for hospitals and physician specialties for three plans' 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 plan yields lower utility than its broad plan, 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 (see Appendix B)—almost all excluding those owned by Partners—whereas Fallon covers significantly fewer providers in Boston.

Figure C.3. : Expected Utility for Various Networks, Boston Rating Region



Notes: This figure plots the distribution of EU^H_{Ijt} and EU^s_{Ijt} for hospitals and 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 Harvard Broad network does skew slightly to the right to that of both narrow networks, the three network utilities virtually overlap one another for a significant portion of the density plot. Looking at Panel (b), which shows the utility distribution for PCPs, EU^{PCP}_{Ijt}, 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.

Estimation Details: I leverage the panel structure of my data—the fact that I observe a sequence of household I making plan choices of plans J over time periods T—to estimate the plan demand model using maximum simulated likelihood, following the procedure outlined by Train (2009). Specifically, the probability that I observe household I making any particular sequence of choices over time is given by:

$$s_I = \int \sum_{t=1}^{T} \sum_{j=1}^{J} \left[\frac{exp(\delta_{Ijt}(\beta))}{\sum_{k=1}^{J} exp(\delta_{Ikt}(\beta))} \right]^{y_{Ijt}} F(\beta) d\beta \tag{C5}$$

where y_{Ijt} is equal to 1 if household I chose plan j at time t and 0 otherwise. To construct a simulated likelihood function, I take r draws for household I from the distribution of β as outlined in Equation 4. For each draw, the likelihood function becomes:

$$\mathcal{L} = \sum_{I} ln \left\{ \frac{1}{R} \sum_{r=1}^{R} \sum_{t=1}^{T} \sum_{j=1}^{J} \left[\frac{exp(\delta_{Ijt}(\beta^r))}{\sum_{k=1}^{J} exp(\delta_{Ikt}(\beta^r))} \right]^{y_{Ijt}} \right\} \tag{C6}$$

where β^r is draw r from the distribution of β. I search over 500 independent draws.

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.

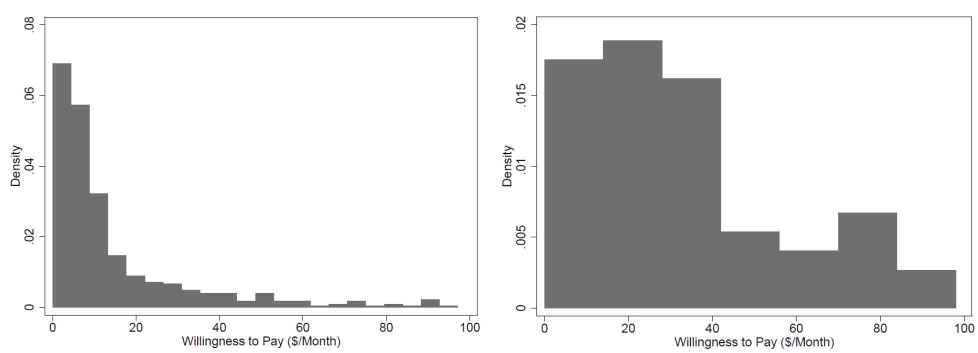
A full set of plan fixed effects are included. As with the provider demand model, I include an indicator variable for whether a particular plan matches an enrollee's plan choice from the previous year. This follows prior literature on plan inertia (Handel, 2013; Polyakova, 2016; Shepard, 2016) and is 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).

For the year 2012 (the year in which the GIC implemented 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 from the corresponding municipalities who joined during the corresponding years.

Distribution of WTP: Figure C.4 shows the distribution of estimated WTP from the plan demand model for Harvard Broad vs. Harvard Narrow, and Fallon Broad vs. Fallon Narrow. Two conclusions emerge from this figure. First, although the mean reported values for Harvard reported in Table 2 are around \$19 per month, there is clearly significant heterogeneity, with certain households willing to pay nearly \$100 per month for access to the broader network. Second, the overall 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.

Figure C.4. : Willingness-to-Pay for Broad Versus Narrow Networks

(a) Harvard Broad vs. Harvard Narrow (b) Fallon Broad vs. Fallon Narrow



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.

C6. Premium Setting Stage

Construction of p_{jht} and p^s_{jdt}: In order to complete Equation 8 and construct the employer objective function, I construct a measure for the base reimbursement 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 scale this price by a series of resource weights to arrive at a payment for each diagnosis and procedure. I use observed "allowed amounts" to specify a base rate for each insurer-provider combination.[15]

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, White and Lyons (2015). I assume that price takes the following form:

$$A^s_{irjdt} = p^s_{jdt} * RVU_{rt} \tag{C7}$$

A^s_{irjdt} 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^s_{jdt} between plan j and practice d, multiplied by the RVU weight for the procedure, RVU_{rt}. The resulting base price can therefore be interpreted as the negotiated rate between plan j and physician practice d for one RVU of care.

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. 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 insurer-hospital base prices, by running the following model:

$$ln(A_{iljht}) = \gamma_{jht} + \psi_{lt} + x_{ilt} + \epsilon_{iljh} \tag{C8}$$

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, I proxy for diagnoses by including ψ_{lt}. These are a set of fixed effects for the 18 CCS diagnosis categories used in the hospital demand model. The model is therefore similar to the physician price construction, except that by including these fixed effects, I estimate weights for each diagnosis rather

[15]Similar approaches have been taken by Gowrisankaran, Nevo and Town (2015), Ho and Lee (2017), and others.

than using observed weights. The model also includes Elixhauser comorbidity indices 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. I scale the predicted price by this factor in order to arrive at the predicted base price for a standardized unit of care, p_{jht}.

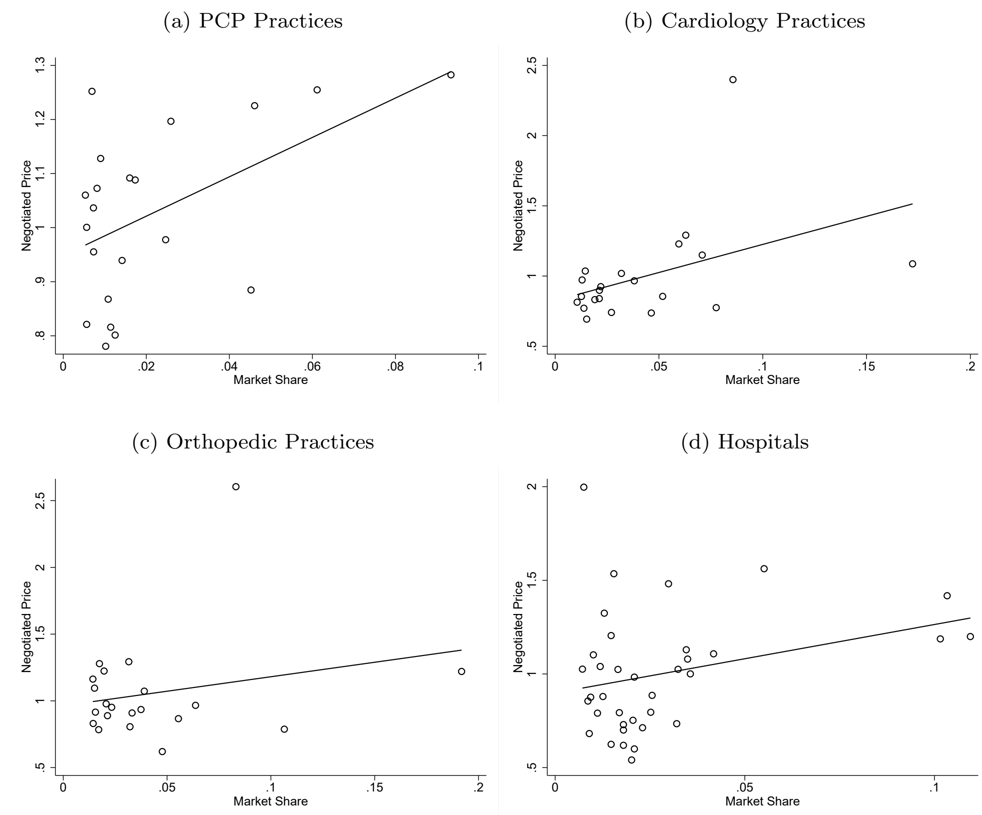
Table C.4 reports the average negotiated base prices for hospitals and physicians and average weights by type of provider and facility type in 2011.[16] The table suggests that negotiated prices do not vary considerably across medical specialties in Massachusetts. Within specialty, however, there is considerable variation. Facility-based cardiology practices, for instance, receive an average price-per-RVU of $57, 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 about $20,000 while the minimum was about $3,000. 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 service, yielding an average of $123 per procedure. Orthopedists, however, perform an average of 4 RVUs per service, implying an average payment of $221 per procedure.

I next examine whether the preference for broad-network plans translates into higher negotiated rates for those providers. Figure C.5 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 market share.

It is clear from the graphs that there is a distinct positive relationship between provider price and consumer valuation of a provider within the insurer's network. The relationship appears strongest for hospitals and cardiologists. However, there is still a positive relationship for PCPs and orthopedists as well. 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 employers in offering plan choice thus becomes clear: to offer a broad-network plan to consumers would yield greater consumer surplus through the inclusion

[16] I define practices that are "office-based" 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 C.5. : Insurer Negotiated Price by Market Share, Boston Rating Region 2011



Notes: This figure plots the negotiated price for hospitals, p_{jht}, and for physician practices, p^s_{jdt}, against market share. Prices are reported for a single insurer and relative to the insurer-specific mean. Data is for year 2011. All plots are for Boston rating region only, except for Panel (d), which is reported for all of Massachusetts.

Table C.4—: Estimated Price and Weight Measures, 2011

	PCPs	Cardiologists	Orthopedists	Hospitals
		Office-Based		
Average Base Price	56.56	56.29	55.33	–
	(12.43)	(14.80)	(16.94)	–
Average Weight	2.19	2.74	3.99	–
	(0.60)	(1.28)	(2.46)	–
		Facility-Based		
Average Base Price	58.52	56.60	52.16	8,145.12
	(15.67)	(19.78)	(14.16)	(3,028.49)
Average Weight	2.52	2.05	6.44	1.02
	(1.44)	(1.69)	(5.17)	(0.12)

Notes: "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 $RVU_{rt} = 1$. In the case of hospitals, this refers to the case where $w_{lt} = 1$. Hospital weights are scaled so that the yearly average is one, meaning that hospital base prices refer to the price for an admission of average weight. "Office-based" settings are defined as practices where more than 70% of claims are flagged as in an office-based setting.

of high-valuation hospitals and doctors, but would also reduce surplus through higher premiums.

Estimating Unobserved Marginal Costs: To estimate $c^u_{Ijt}(N_{jt})$, I rely on standard inversion of the first-order condition specified in Equation 9. 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 claims data, to recover *unobserved* marginal costs, I parameterize them as $c^u_{Ijt}(N_{jt}) = c^u_{jt}(N_{jt})\theta^c_I$, where θ^c_I scales base plan-specific unobserved costs, $c^u_{jt}(N_{jt})$, across household type I. I assume that unobserved marginal costs only vary by whether the household is an individual or family. I infer θ^c_I directly from the data by aggregating all claims from providers that are not hospitals, PCPs, cardiologists, and orthopedists, and regressing the observed allowed amounts for these claims on household type.[17] This reduces the number of unknowns to JT, allowing for full recovery of $c^u_{jt}(N_{jt})$.

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

[17]The critical assumption here is that all marginal costs that vary by more granular risk types are captured through *observed* hospital and physician costs, whereas *unobserved* costs only vary by family type. While strong, this seems reasonable as a first-order approximation. I report robustness on this assumption in subsection E.E3.

insurer subscript m back) such that:

$$c^u_{mjt}(N_{mjt}) = \kappa x_{mjt} + \gamma_m + \gamma_t + \varepsilon_{mjt} \tag{C9}$$

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.

Cost Estimates: Table C.5 reports the results of Equation C9, with the log of unobserved marginal costs as the dependent variable. Year 2012 is omitted due to potential bias from it being the year of the premium holiday.

Table C.5—: Unobserved Marginal Cost Estimates

	Coefficient	Standard Error
Narrow Network	-0.164^{***}	0.019
Harvard Pilgrim	0.059^{**}	0.022
Health New England	-0.064^{**}	0.026
Neighborhood Health Plan	-0.039	0.026
Tufts Health Plan	0.057^{**}	0.022
2010	0.002	0.023
2011	0.035	0.022
2013	0.080^{***}	0.022
Constant	5.911^{***}	0.021
Observations	28	
R^2	0.93	

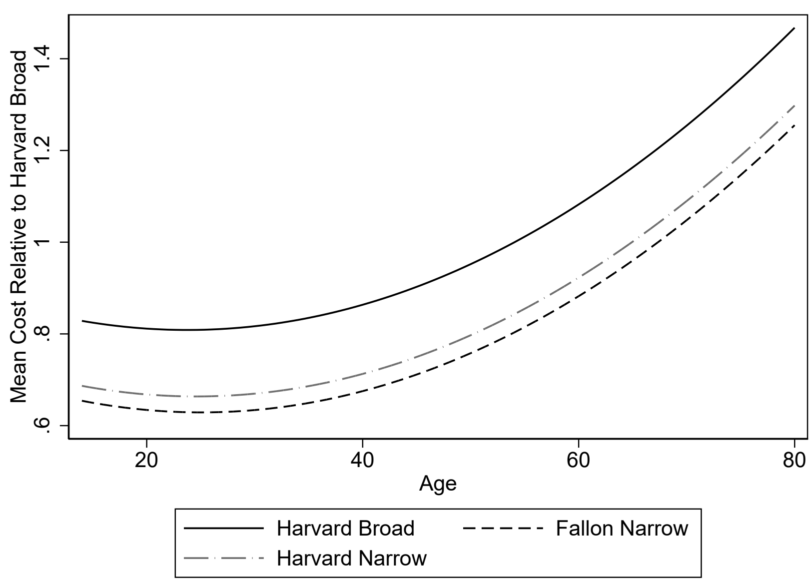
Notes: Results from marginal cost estimation. Dependent variable is the log of unobserved marginal costs. Omitted insurer is Fallon Health Plan. Omitted year is 2009. Year 2012 is also omitted from the analysis due to concern of bias from the enactment of the premium holiday.

The results indicate that being a narrow-network plan reduces unobserved marginal costs of health care by approximately 16%. 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 a larger set of providers unaccounted for by the chosen specialties.[18] 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.

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

Figure C.6 plots the total estimated marginal costs of health care (hospital + PCP + cardiology + orthopedics + unobserved) against age for single-member households. I report estimated cost-curves for Harvard Broad, Harvard Narrow, and Fallon Narrow and, for confidentiality reasons, only report estimated costs relative to the Harvard Broad mean. 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 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.

Figure C.6. : 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.

C7. Additional Details on Estimating the Employer Objective Function

I make several assumptions to proceed with the estimation of ρ and FC_j in Equation 11. 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.

The latter refers to all other uncertainty in demand, including measurement error. Both disturbances are unknown to the employer and the econometrician. I assume that $E[v^b_{1,\delta_{Jt}}] = 0$.

Rather than relying on instruments within the employer's information set, I instead use observed data on municipal entrants by year to specify a distribution of household entrants over which the employer 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.[19] 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 sample means, this implies:

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

In the estimation of Equation 17, I take the average of 10 disturbances of ω_k. That is, I estimate the moment inequalities assuming 10 different potential random samples of entrants in each year given the number of municipalities who entered the previous year. For each set of potential entrants, I also simulate 10 different distributions of demand coefficients, β^s_{2I}, to account for the presence of unobserved heterogeneity. This makes a total of 100 evaluations of each moment in each year.[20]

The second assumption is that there is no presence of a structural error component that the employer knows when making decisions, but the econometrician does not. Prior work has treated such structural errors as disturbances in the fixed cost term. For instance, it could be that: $FC_j = FC + v_{2,j}$, 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 formulation if the error term varied by the type of product offered. In my setting, the GIC might choose to contract with certain insurers, offer certain

[19]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 entrants 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.

[20]In previous drafts of this paper, I have evaluated the employer objective function using 100 different samples of municipal entrants. However, this poses a major computational burden with the inclusion of random coefficients. The current results are encouragingly similar to higher-order simulations.

products, or offer certain networks for which the fixed costs are lower. Without addition assumptions, 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. $FC_j = FC$. Similar assumptions were made by Nosko (2014). 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. 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, the differences in fixed costs *within* an employer group are likely smaller. This assumption may be violated if, for instance, offering a product that was broader in network size than another product also meant an increase in 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 their maximum network, 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.

Details on the Private Employer Sample

Sample Construction: To construct the sample of large, private employers used in section III, I limit the claims data to members employed by non-government firms with more than 50 employees and those who have at least one commercial insurance product that is "self-insured." Restricting the sample to large, self-insured firms makes the estimation of the moment inequalities considerably simpler than if I also included small employers, as it allows me to construct similar premium pricing rules as for the GIC and abstract from incorporating insurer profit functions. As the APCD does not contain firm identifiers, I instead create a sample of firms using the employer zip code field (herafter referred to as "employerzip"), Standard Industrial Classification (SIC) code, and product (plan) identifiers (IDs), the latter typically being unique within firm. As employers can offer multiple different plans—and therefore have multiple plan IDs—I use employee flows across plan IDs to determine the likelihood that any two IDs belong to the same firm. Specifically, if I observe two different plan IDs within an employerzip-SIC grouping and also observe that a non-trivial share of employees switch from one ID to the another (and vice versa), I assume that both IDs are part of the same firm.

I then simulate a distribution of firms offering narrow-network insurance plans using external micro data from the Kaiser Family Foundation and Health Research & Education Trust (HRET). The Kaiser/HRET annual survey of employer-sponsored health benefits contains questions about employers' general characteristics, plan offerings, enrollment, health risk appraisals, and other topics. Beginning in 2014, the survey asked whether firms offered narrow-network plans. Since my APCD sample ranges from 2009 - 2013, I limit the APCD to 2013 and simulate firm offer distributions from the 2014 Kaiser/HRET survey. Unfortunately, the survey only contains geographic information up to broad Census region categories. I am therefore not able to match the distribution of firms and plan offers to Massachusetts firms directly. Instead, I limit the Kaiser/HRET sample to only firms in the Northeast United States, and match narrow-network offer rates using data on firm size, industry, and number of plans offered.

I make a few additional simplifying assumptions in creating the sample. First, I limit the APCD to only members covered by the same insurers as in the GIC. This primarily has the effect of removing Blue Cross Blue Shield (BCBS) members from the data. While this represents a non-trivial share of commercial enrollment in Massachusetts, it is nonetheless a sensible restriction to make.[21] During my sample period, BCBS was the only carrier that did not offer *any* narrow-network products on the market (Office of the Attorney General Martha Coakley, 2013). Moreover, focusing exclusively on GIC carriers reduces computation burden significantly, as it enables me to use already-estimated demand parameters and

[21] BCBS had about 45% of the commercial payer market share in Massachusetts in 2012 (Center for Health Information and Analytics, 2013).

negotiated prices (see subsection C.C6) rather than re-estimating demand for the set of BCBS members, for which I have no estimated "brand effect."

The second simplifying assumption I make is in the network offerings of employers. For products outside the GIC, I do not observe the network breadth of each plan. One option would be to infer networks based on observed claims.[22] However, this approach is prone to significant measurement error, particularly for firms with fewer employees. Instead, I leverage institutional features of the Massachusetts insurance market. In particular, outside of the GIC, Harvard Pilgrim and Tufts Health Plan each only marketed one narrow-network insurance product to employers as of 2013, known as "Harvard Focus" and "Tufts Select" (Office of the Attorney General Martha Coakley, 2013). I therefore assume that any firm in the private sample that I simulate offering a Harvard Pilgrim narrow-network product covers the same hospitals and physicians as Harvard Focus. Similarly, for firms that I simulate offering Tufts narrow products, I assign the network breadth of Tufts Select. I impute the networks of these products using publicly-available network brochures for each of these plans (in a similar way to the construction of GIC networks, detailed in Appendix A).

Table D.1 reports summary statistics for the simulated private employer sample and compares them to the GIC. Overall, the sample contains 123 simulated large private employers in the state. Though many of the characteristics of the simulated sample look similar to the GIC, there are some notable differences. Approximately 8% of those employers offer narrow-network plans (Column 1), consistent with the share seen in the Kaiser/HRET survey. However, only 2% of *employees* across the state actually enrolled in narrow-network plans in 2013, compared with about 12% in the GIC (Columns 2 and 3).[23] The GIC sample is slightly older, with about 20% of employees being over age 55, compared to about 16% in the private employer sample. Together, the health care and service industry comprised 70% of the private firms. In terms of geographic distribution, most large private employers (55%) are headquartered in Boston (Rating Region 5). At the employee level, this translates to about 37% of all private employees working for firms in Boston, with the next largest share (17%) working for firms in Rating Region 4 (the North Shore). On the GIC, conversely, employees were more evenly distributed across regions. For instance, 24% of employees lived in Rating Region 5 and 22% of employees lived in Rating Region 4. Overall, then, private employers skew more heavily towards dense, urban areas than employees on the GIC.

Estimation of Employer Objective Function: Estimation of the employer objective function for private employers follows a very similar procedure outlined in subsection II.D. However, I make several assumptions to accommodate features

[22]See Gruber and McKnight (2016) for such an approach.
[23]Recall that the recent premium holiday implemented in 2012 was somewhat responsible for this high share of enrollment.

Table D.1—: Summary Statistics for Simulated Private Employer Sample

Variable	Private Emp.	Private Emp.	GIC
	Firm Level	Employee Level	
Offered a Narrow Network	0.079	–	–
	(0.271)	–	–
Enrolled in a Narrow Network	–	0.018	0.118
	–	(0.134)	(0.323)
Age 55+	–	0.159	0.201
	–	(0.366)	(0.401)
Female	–	0.536	0.518
	–	(0.499)	(0.500)
Rating Area			
1	0.131	0.064	0.158
	(0.339)	(0.245)	(0.365)
2	0.071	0.087	0.127
	(0.259)	(0.282)	(0.333)
3	0.071	0.172	0.100
	(0.259)	(0.377)	(0.300)
4	0.143	0.171	0.218
	(0.352)	(0.376)	(0.413)
5	0.548	0.372	0.240
	(0.501)	(0.483)	(0.427)
6	0.036	0.125	0.135
	(0.187)	(0.331)	(0.342)
Industry			
Health Care	0.425	0.391	–
	(0.497)	(0.488)	–
Service	0.310	0.397	–
	(0.465)	(0.489)	–
Wholesale	0.023	0.023	–
	(0.151)	(0.148)	–
Transportation, Communications, Utilities	0.115	0.072	–
	(0.321)	(0.259)	–
Manufacturing	0.126	0.117	–
	(0.334)	(0.322)	–
Number of Employers	123		

Notes: Summary statistics for simulated sample of private employers in Massachusetts in 2013 (Columns 1 and 2) and employees of the Group Insurance Commission (Column 3). First column reports characteristics at the firm level, while last two columns report characteristics at the employee level.

of the simulated sample. First, I use the same demand parameters as estimated in Table 2, essentially assuming that employees of large, self-insured, private firms, conditional on observables, have similar demand for health insurance as employees of the GIC. Second, in order to circumvent selection issues with estimating fixed costs across different employers (noted in subsection C.C7), I restrict the moments for each employer to have the same number of plans they currently offer. For example, if an employer currently offers two plans, then for that employer, I only consider alternate plan menus/networks in which that employer offers two plans. This allows me to isolate the effect on the employer-employee mismatch term, ρ. Finally, for each alternate plan menu, I now construct moments by taking sample averages across employers. In other words, the moment equation from Equation 17 becomes:

$$m(\delta_J, \delta_J^a, \theta, z) = \sum_{s=1}^{10} \left(\frac{1}{F} \sum_f \left[(W(\delta_J, \theta_s) - W(\delta_J^a, \theta_s)) \otimes g(z) \right] \right) \geq 0 \tag{D1}$$

where f is the subscript for employer f and F is the total number of private firms sampled.

Robustness on Employer Objective Function

E1. Households with Prior Provider Relationships

In Table E.1, I report estimates of ρ and FC_j where I reweight the population of the employee pool such that 90% of households have had a prior relationship with a provider. Here, I define a "prior relationship" as having previously (as of year t) visited a provider. In particular, I consider four different populations: employees with a prior relationship with a Partners provider (either a hospital or physician), a Umass provider, an Atrius physician, or any provider that is *only* covered by a broad-network plan. The results do show that the mismatch parameter declines for households with prior provider relationships, implying that employers may overweight these populations in their network design. In particular, ρ declines from 3.70 at baseline to 2.19 for the population with any prior provider relationship. The results are similar for households with a prior relationship with a Partners provider or Umass provider.

Notably this is not the case for households with a prior relationship with an Atrius provider. In addition, the estimates of ρ from the regional analyses in Table 4 are comparable, and in the case of the North Shore (Region 4) they are substantially lower. The combination of these facts suggests that, while employers do appear to be motivated by the preferences of households with prior provider relationships, these effects are likely strongest in regions where dominant providers (e.g. Partners) face less competition. This may explain why the mismatch parameter increases for households with prior Atrius relationships: the fact that Atrius primarily operates in dense regions (e.g. Boston) suggest that its removal from a network may not affect utility by as large a magnitude as the decrease in premium spending.

Table E.1—: Employer Objective Function Parameters For Populations with Prior Provider Relationships

	(1) Baseline	(2) Any Provider	(3) Partners Provider	(4) Umass Provider	(5) Atrius Provider
ρ	3.70	2.19	2.23	2.08	4.04
FC_j ($Millions)	3.98	2.20	2.64	1.99	4.53

Notes: Results from ρ and FC_j estimation for 2009-2013. Column 1 presents estimates for the baseline population of GIC enrollees. Columns 2-5 present estimates that reweight the population such that 90% of the population have a prior relationship with a provider of a certain type. FC_j reported in millions of dollars.

E2. Physician Inertia and Active Choice Frictions

Alternate Assumptions on Physician Inertia: In my model of provider demand, my baseline estimates treat persistence in provider choice as a welfare-relevant utility component driven by the formation of patient-physician capital. Under this assumption, if a physician were removed from the network and a patient had seen that physician previously, that patient would incur a substantial utility loss. However, the characteristics of that physician would not necessarily be informative as to which physician the patient would choose in his or her absence.

There are two alternate interpretations of physician inertia. The first is that persistence in choice of providers is driven by unobserved physician quality and not necessarily the patient-provider match. Here, the loss of a physician from the network would also imply a welfare-relevant loss. However, the main distinction from the baseline assumption is that the utility change from the loss of a provider will vary by (a) the patient's characteristics and preferences; (b) the characteristics of the provider and; (c) the characteristics of the remaining providers in the choice set. For example, if a high-quality physician were removed from the network with no close substitute in the resulting smaller network, the patient would incur a substantially higher utility loss than the baseline estimate. Conversely, if a physician were removed and the resulting network had many physicians remaining of similar quality, the utility loss—and hence welfare implications —would be smaller than the baseline.

Finally, the inertia term may reflect switching or hassle costs irrespective of physician quality or match. Here, if a physician were removed from a network, the model ought to predict a similar second choice as with the baseline assumption. However, if persistence were driven by hassle costs, then it is possible the employer would not view such costs as welfare-relevant in its decision-making about network offers.

Each of these interpretations, through their impact on consumer utility of a network change, can have significant impacts on the estimates of the employer objective function. In Table E.2, I report results on ρ and FC_j assuming that the entirety of the inertia term were driven by these various forces.[24] To test the impact of treating physician inertia as a switching/hassle cost, I re-estimate the employer objective function assuming that the utility change from losing a provider were "welfare irrelevant" from the eyes of the employer. In doing so, the estimate of the employer-employee mismatch increases significantly, from a baseline of 3.70 to 6.39 (Column 2). This result makes sense: in this scenario, any potential narrowing of a network would result in a smaller utility loss, but a similar decline in health spending. As such, the fact that the employer does *not* narrow the network implies a much larger mismatch between employer incentives and employee preferences.

To test the impact of treating physician inertia as unobserved provider quality, I re-estimate the provider demand model only on patients who had never seen *any* provider prior to their current visit. I then use these demand estimates in

[24]Indeed, the inertia term might be driven by a combination of these forces. Treating the entirety of the term as being driven by one force or another is meant to show bounds on the relevant parameters for the employer.

Table E.2—: Employer Objective Function Parameters Under Alternate Provider Inertia Assumptions

	(1) Inertia = Pat./Prov. Match	(2) Inertia = Switching Costs	(3) Inertia = Unobserved Quality
ρ	3.70	6.39	4.00
FC_j ($Millions)	3.98	7.90	4.55

Notes: Results from ρ and FC_j estimation for 2009-2013. Columns 1-3 presents estimates under different assumptions of the interpretation of physician inertia. FC_j reported in millions of dollars.

estimation of the employer objective function. The assumption here is that if persistence in physician choice were driven mainly by unobserved physician quality (irrespective of physician-provider-specific match), this ought to be reflected in the first-time choices made by brand new patients. Under this interpretation, the employer-employee mismatch again rises, though very slightly, from baseline. (Column 3). This suggests two things. First, removing any physician yields a smaller utility loss for patients than the baseline assumption, implying that patients are typically able to find close substitutes. Second, the baseline model does reasonably well at estimating unobserved provider quality.[25] Taken together, the fact that the baseline model yields the smallest mismatch parameter implies that it is most conservative interpretation of physician inertia. The "true" mismatch parameter, then, likely lies somewhere between 3.70 and 6.39, but it always considerably greater than 1.

Active Choice Frictions: While the demand model estimated in subsection II.B incorporates plan inertia and switching costs, it does not explicitly model active choice frictions that might apply to both new and existing enrollees. Recent work has shown that choice complexity, information asymmetry, and choice overload drive enrollees to, for example, opt into dominated health plans (Bhargava, Loewenstein and Sydnor, 2017; Handel and Kolstad, 2015; Abaluck and Gruber, 2020). For instance, it may be the case that employees select broad networks not out of a "true" valuation of the network, but rather out of a lack of full information about plan features and an aversion to potential out-of-network bills they might incur. My model treats these frictions as "welfare-relevant" in the sense that employers observe household plan selections ex-post and make their plan offer decisions assuming those choices reflect full information. However, to the extent that observed household choices reflect these frictions—and are therefore not welfare-relevant—estimates of $CS(\delta_{Jt}, \theta)$ (and therefore ρ) may be biased. Though I do not test robustness to this formally, the exercises with physician inertia shown affect the employer objective function through a similar mechanism as

[25]If the estimate revealed that the mismatch parameter substantially *declined*, this would imply the baseline model was not accurately capturing the utility loss from the removal of a flagship or high-quality provider from the network.

would active choice frictions. Part of the valuation of a broad-network plan may reflect frictions that may alter computations of consumer surplus if the employer treated them as welfare-irrelevant. In those simulations, I show that estimates of ρ *increase* when employers treat physician inertia as welfare-irrelevant. Intuitively, households lose less utility from a shift to narrow-network plans and, as such, the employer-employee mismatch rises. I therefore take the estimates of ρ in Table 3 as conservative estimates.

E3. Employer Mistakes and Additional Robustness

The Role of Switching Costs: A possible explanation for employer persistence in offering broad-network plans is that employers misperceive the true loss in employee utility from a loss of providers. This would most commonly be the case if they mistook enrollee inertia for "true" network utility. This is a fairly difficult phenomenon to test for. Indeed, if the entirety of the switching cost parameter were shifted to network utility, then the mismatch parameter would mechanically shift downward as the utility gap between broad and narrow networks would widen. To get a sense of the precise magnitude, one possibility would be to re-estimate the plan demand model but simply omit the plan switching cost term. However, as seen in Table 2, this results in implausibly low premium sensitivity estimates.[26] Another approach would be to shift some portion of the switching cost estimate towards the network utility. However, this approach is difficult to implement empirically as it requires making assumptions as to how switching costs–a flat per-plan cost–maps to network utility, which scales by plan.

I instead take an alternate approach: I re-estimate the plan demand model for a specific sub-segment of the population for whom it is likely the employer believes have strong preferences for broad networks. I then apply these estimates to the entire population. Specifically, I focus on new entrants to the GIC coming specifically from municipalities entering for the first time. This solves the premium elasticity issue mentioned above, as these employees, by definition, have no inertia. However, unlike other new employees, these municipal entrants have previously lived and worked in the state, and have also been previously enrolled in private health insurance. Moreover, prior plans were uniformly generous, broad-network plans.[27]

Panel A of Table E.3 reports these estimates, focusing only on moments that fix the number of products to isolate the effects on ρ.[28] The employer-employee mismatch term drops to 1.89, about a 49% decline from the baseline estimate (Column 1). Indeed, this does suggest a role for employer misperceptions. However,

[26] Some older households under this specification are predicted to have *positive* utility from higher premiums.

[27] About 50% of municipal entrants were previously insured by Blue Cross Blue Shield, which at the time had no narrow-network products. About 90% of entrants were enrolled in a plan with zero deductible.

[28] Because much of the switching cost is loaded onto network utility in this specification, the estimated upper bound on fixed costs would be exceedingly large.

even in this case, the employer continues to overweight the average household's preferences by about 2-to-1. Moreover, the same patterns across populations observed in Table 4 persist. In fact, assuming that the *entire* discrepancy between the baseline estimates and these estimates is driven by misperceptions, this still implies that up to a quarter of the mismatch can potentially be attributed to unequal weighting in household preferences.

Table E.3—: Additional Specifications for the Employer Objective Function

	(1) Baseline	(2) Older	(3) Older, Regions 1,4,6	(4) Older, Region 4
Panel A: Estimation On New, Municipal Entrants				
ρ	1.89	1.49	1.52	1.25
FC_j ($Millions)	–	–	–	–
Panel B: Estimation With Alternate Marginal Costs				
ρ	2.94	2.51	2.05	1.65
FC_j ($Millions)	3.65	3.60	2.24	2.65
Panel C: Estimation Restricting Harvard Broad to Narrow				
ρ	3.67	3.07	2.64	1.93
FC_j ($Millions)	3.88	3.35	2.47	2.82
Panel D: Estimation Restricting Tufts Broad to Medium				
ρ	3.70	3.08	2.65	1.93
FC_j ($Millions)	3.98	3.39	2.50	2.83
Panel E: Estimation Restricting Tufts Narrow to Harvard Medium				
ρ	4.38	3.26	2.67	1.86
FC_j ($Millions)	4.68	3.44	2.61	2.73
Panel F: Estimation Restricting Harvard Broad to Medium				
ρ	1.97	1.67	1.30	0.98
FC_j ($Millions)	2.57	2.27	1.57	1.95

Notes: Results from ρ and FC_j estimation for 2009-2013. Column 1 presents estimates for the baseline population of GIC enrollees. Columns 2-4 present estimates that reweight the population such that 90% of households are older (age 55 and older) and/or reside in certain rating regions. Panel A reports estimates on only new, municipal entrants. Panel B reports estimates from alternate marginal cost assumptions. Panels C-E present estimates restricting the GIC from altering certain plans.

Alternate Assumptions on Unobserved Marginal Costs: I also test my model's sensitivity to the estimation of unobserved marginal costs, $c^u_{jt}(N_{jt})$, as detailed in subsection C.C6. In particular, if the cost differential from switching to a narrow-network plan is lower than the cost differential I estimate, then the estimate of ρ may be inflated. I address this in two ways. First, rather than estimating the cost equation in Equation C9, I instead predict $c^u_{jt}(N_{jt})$ for each counterfactual network non-parametrically. Specifically, I take the ratio of estimated unobserved marginal costs to *observed* marginal costs, $c^o_{jt}(N_{jt})$, for each household and insurer. I then scale the observed marginal costs for each counterfactual network by this ratio. This has the effect of allowing unobserved marginal costs to vary not just by whether a counterfactual network is narrow or not, but by the size of its network relative to a broad-network plan. To the extent that small changes in networks result in smaller changes in unobserved marginal costs than currently estimated in Table C.5, this specification ought to address this. Second, I allow more flexibility in θ_I^c in Equation 9 by allowing unobserved marginal costs to vary not only by household type (individual or family), but also by age and rating region.

Note that these are fairly conservative assumptions. My model, for instance, assumes that provider prices are fixed in equilibrium. Prior literature on narrow networks have shown that more prevalent use of these plans may result in decreases in equilibrium provider prices (Ho and Lee, 2019; Ghili, 2020; Liebman, 2018). In addition, I am assuming a fixed *quantity* of care in equilibrium. To the extent that narrow networks induce reductions in this quantity, I may be underestimating the cost differential (Gruber and McKnight, 2016; LoSasso and Atwood, 2016). Finally, in Table C.5, the estimated differential for $c^u_{jt}(N_{jt})$ between a broad and narrow plan within the same insurer is about 16%, even though the premium differential between these plans is, on average, 20%. The assumptions here reduce these differentials further when the network changes are small.

Panel B of Table E.3 shows these results. Indeed, assuming that unobserved marginal costs decline by smaller magnitudes reduces the estimated ρ, though this is somewhat mechanical. Importantly, even under these conservative assumptions on costs, my estimates still imply employers overweight consumer surplus by about three times relative to the preferences of the average consumer.

Restricting Certain Moments: I run a series of robustness checks in which I restrict the estimation to certain moments to ascertain which plan changes drive the primary estimates of the employer-employee mismatch in Table 3. The results are displayed in Panels C-E of Table E.3. In Panel C, I restrict the GIC's ability to reduce Harvard's broad network to a "narrow" equivalent to network "N1" as described in section IV. This is a fairly network of both hospitals and physicians, with both Partners and Atrius removed, as well as additional hospitals and physician groups. In Panel D, I restrict the GIC from reducing Tufts' broad network to a "medium" sized network ("M."). In this network, the only major change is that Partners is removed. Atrius physicians, however, are preserved. The mismatch parameter is robust to each of these specifications.

In Panel E, I restrict the GIC from *broadening* Tufts' narrow network by limiting all moments in which it switches to a Harvard "M" network. Here, the baseline estimate of ρ increases slightly to 4.38. This is driven by the fact that these moments serve as relevant "upper bounds" on the parameter. Removing them therefore changes the bound on the estimate. Even still, the mismatch parameter is largely consistent with the baseline specification.

The one glaring exception is when I restrict the GIC's ability to reduce the network breadth of Harvard Broad to either its "M" network or a similarly-sized network on the small group market ("N2"). Again, these networks are somewhat larger than the others considered in the choice set, as they primarily remove Partners, but preserve a wide network of physicians in other systems. Panel F reports these estimates. Here, the mismatch parameter changes considerably: the estimate of ρ for the baseline population falls from 3.70 to 1.97.

The implication is that the mismatch appears to be largely driven by employers' unwillingness to make small network changes, yet those with significant implications on premium spending. In this instance, the moments responsible for the result are the ones in which the employer could remove from its network some flagship and costly hospital systems (e.g. Partners) but preserve a wide network of *physicians* (e.g. Atrius). If the employer made this move, the utility differences for most employees would be small, given consumers' strong preferences for physician networks seen in Table 2. The cost implications, however, would be substantial. This result is consistent with the insights from Shepard (2016). Interestingly, even in simulations restricting the GIC's ability to narrow Harvard's broad-network plan, the same heterogeneity implications from Table 4 continue to persist. Reweighting the population to the older sample described above, the estimate of ρ falls further to 1.67, while it falls to 1.30 when rescaling towards the employees residing in regions 1,4, and 6.

Additional Counterfactual Details

F1. Simulation Procedure

I now describe the procedure used to implement the policy simulations in section IV. In order to reduce the dimensionality of the computation, as with the employer objective function estimation, I restrict to the same product space used in subsection II.D and detailed below. This leaves a possible set of 14 products for the employer to offer. I proceed computing the equilibrium networks offered in a series of steps:

1) Construct a matrix of $2^{14} = 16,384$ possible combinations of products offers.
2) For each product combination, 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 marginal costs of health care to the employer, 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" marginal costs of health care, c^u_{Ijt}, using the parameters estimated from Equation C9.
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 9.
6) Compute the estimated consumer surplus, $CS(\delta_{Jt}, \theta)$, and total outlays for the employer under the current product menu offered.
7) Compute the employer's objective function using estimated $CS(\delta_{Jt}, \theta)$, total expenditures, the estimated mismatch parameter, ρ, and fixed costs, FC_j.
8) Repeat this procedure for each vector of possible plan menus, and take the max of all the computed employer objective functions.

F2. Additional Counterfactual Tables and Figures

Premiums and Market Shares Under the Enthoven Counterfactual: Table F.1 shows the counterfactual premiums and market shares for each health plan under the Enthoven approach, assuming the product menu remains fixed. Under this counterfactual, individual co-premiums for broad-network plans increase substantially, while co-premiums for narrow-network plans remains relatively similar to their baseline. Harvard Broad, in particular, increases from its observed value of $152 per month to $311 per month. Not coincidentally, these plans also see substantial shifts in enrollment. The share of enrollees in Harvard Broad declines from 34% to just 11%, with many of those households shifting to Fallon, HNE, and NHP. As a result of these shifts, the overall spending burden for the GIC falls by *more* than the loss in consumer surplus.

Table F.1—: Counterfactuals: Shares and Premiums for Enthoven Approach, 2011

Insurer	Network	Baseline	Counterfactual	Baseline	Counterfactual
		Market Shares		Co-Premiums	
Fallon	Very Narrow	0.02	0.08	$105	$111
Fallon	Broad	0.05	0.03	$126	$197
Harvard	Med	0.04	0.05	$122	$179
Harvard	Broad	0.34	0.11	$152	$311
HNE	Narrow	0.11	0.22	$105	$105
NHP	Narrow	0.03	0.29	$106	$109
Tufts	Narrow	0.01	0.02	$117	$159
Tufts	Broad	0.41	0.19	$147	$272

Notes: Market shares and individual monthly co-premiums for baseline and counterfactual predictions, holding the GIC's product menu fixed. Individual co-premiums for counterfactual plans computed only in regions where Health New England was offered.

Policy Simulations Assuming No Logit Error: I re-estimate the region-rating counterfactuals presented in section IV under the assumption of no logit error. Table F.2 reports these results. The results remain largely consistent with those reported in Table 6. In particular, the employer still offers predominantly broad-network plans in Rating Region 4 and predominantly narrow-network products in Rating Region 5. The most notable change is the the *number* of product offered drops somewhat, with the GIC offering five plans in Rating Region 1 and Rating Region 4, while only offering four plans in Rating Region 5. Despite these changes, the welfare implications remain similar, but for a slight increase in total surplus (Panel C) relative to the estimates in Table 6. This is driven by the fact that the fixed cost estimates are smaller with the logit error removed (as in Table 3) and, as a result, social surplus is somewhat higher relative to the baseline scenario.

Table F.2—: Counterfactuals: Equilibrium Networks Chosen Under Region-Based Pricing, No Logit Error

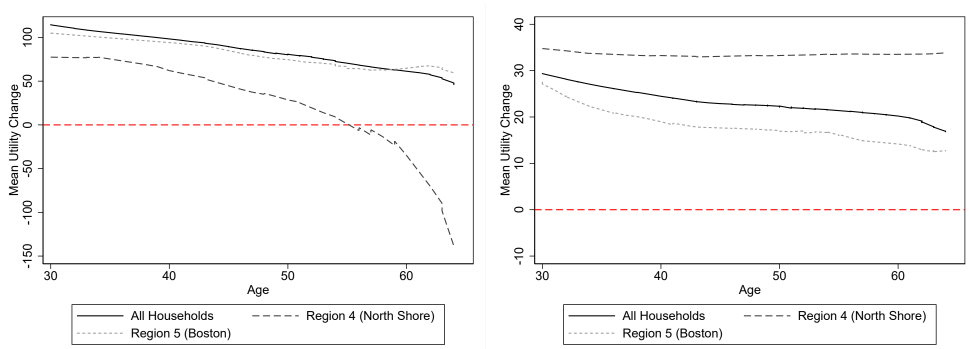
Insurer	Network	(1) Observed	(2) R1	(3) R4	(4) R5
Panel A: Equilibrium Plan Menus/Networks					
Fallon	VN	x		x	
Harvard	VN		x		
Tufts	VN				
Harvard	N1				
Tufts	N1	x			
Harvard	N2				x
Tufts	N2				
HNE	N	x	x		
NHP	N	x		x	x
Harvard	M	x	x		x
Tufts	M				
Fallon	B	x	x	x	
Harvard	B	x		x	
Tufts	B	x	x	x	x
Total Plans		8	5	5	4
Panel B: Welfare/Spending with Fixed Menu					
ΔCS				$-\$0.55$	
$\Delta Costs$				$-\$1.26$	
ΔFC				–	
$\Delta Surplus$				\$0.71	
Panel C: Welfare/Spending with Menu Changes					
ΔCS				$-\$7.93$	
$\Delta Costs$				$-\$37.31$	
ΔFC				\$3.85	
$\Delta Surplus$				\$25.54	

Notes: GIC observed and predicted products offered under region-based rating. "R1" refers to plan networks for Region 1, etc. Panel B reports the welfare and cost changes assuming plan menus remain fixed. Panel C reports these quantities allowing the employer changes to menus. "ΔCS" refers to change in consumer surplus per-household-per-month. "$\Delta Costs$" refer to the change in GIC health costs per-household-per-month. "ΔFC" refer to changes in fixed costs from the new menus. $\Delta Surplus$ refers to the change in total surplus.

Distributional Consequences for Individuals: Figure F.1 plots results from Figure 4, but for individuals rather than families. The results are quite similar, except that individuals see fewer surplus losses from either approach. For the Enthoven approach, individuals living in the North Shore do not see surplus losses until around age 55, while in the region-rating approach, *no* individuals see surplus losses.

Figure F.1. : Total Surplus Changes by Age

(a) Enthoven Approach (b) Region-Rating Approach



Notes: This figure plots the average utility change across households with individual plans by age from implementing an Enthoven pricing approach (Panel A) and a region-rating approach (Panel B). All estimates allow the GIC to alter its plan menus. Curves are plotted for all households, for households in rating region 4 (the North Shore of Massachusetts), and for rating region 5 (which includes the Boston metro area). Surplus is presented in dollarized terms, net of the predicted change in spending incurred by the GIC.