Determinants of Healthcare Provider Networks: Risk Selection vs. Administrative Costs

Natalia Serna*

Abstract

Health insurers typically negotiate with providers over network inclusions. This paper shows that insurers' decision to include providers in the network depends on two factors: risk selection and network administrative costs. To decompose the relative importance of these factors, I estimate a structural model of insurer competition in network breadth applied to data from Colombia. I find that insurers risk-select by providing narrow networks in services that unprofitable patients require. Despite selection incentives, some insurers choose to offer broad networks because of heterogeneity in their cost structure. Results have implications for the design of risk adjustment and network adequacy rules.

Keywords: Health Insurance; Provider networks; Risk selection; Cost Structure.

JEL codes: I11, I13, I18, L13.

*Department of Health Policy, Stanford University. e-mail: nserna@stanford.edu. I am deeply grateful to Alan Sorensen, Ken Hendricks, Corina Mommaerts, JF Houde, and Dan Quint, as well as Naoki Aizawa, Lorenzo Magnolfi, Christopher Sullivan, and Ashley Swanson for their mentorship and thoughtful advice. This research has benefited substantially from discussions with Ignacio Cuesta, Kate Ho, Maria Polyakova, Mark Shepard, and Pietro Tebaldi; as well as from comments from participants at BFI's 2024 Women in Empirical Microeconomics, NBER IO 2023 Summer Institute, 2022 Stanford's SITE, IO Workshop at UW-Madison, ASHEcon 2022, IIOC 2023, MEA 2022, Central Bank of Colombia Workshop 2023, and the seminars at Boston University, California Institute of Technology, Kelley School of Business, London School of Economics and Political Science, Ohio State University, Olin School of Business, Quantil Colombia, Ross School of Business, Stanford Health Policy, University of Toronto, University College London, and University of Pennsylvania. I want to thank the Colombian Ministry of Health for providing the data for this research. Findings do not represent the views of any institution involved. All errors are my own.

1 Introduction

In health systems with managed care competition—such as those of the U.S., Israel, Netherlands, Germany, Switzerland, and Colombia—health insurers typically negotiate with healthcare providers over network inclusions. These negotiations result in substantial variation in network breadth across insurers. For example, Dafny et al. (2017) report that physician network breadth ranges from 23% to 34% and hospital network breadth from 75% to 86% across insurance plans in the U.S. Marketplaces.[1] Why does network breadth vary across insurers despite the strong regulation of other plan characteristics (such as premiums, deductibles, and cost-sharing)? And does this variation matter for policy-making?

In answering these questions recent literature has focused on the role of risk selection as the main driver of insurers' incentives to offer narrow networks (e.g., Shepard, 2022).[2] In this paper, I discuss and document a novel dimension that co-determines the formation of provider networks in a health insurance market: insurers' heterogeneous cost structures. If insurers enjoy economies of scale or scope in the number of covered providers, they may have incentives to offer broad networks, counterbalancing the effect of risk selection.

My setting to study the determinants of provider network breadth is the Colombian healthcare system. In this system, private insurers provide one contract that is closely regulated along most dimensions (such as premiums, cost-sharing, and benefits). However, provider networks and provider prices are unregulated. Insurers in Colombia can choose their provider networks separately for every health service covered in the contract. This makes for a convenient empirical setting to isolate and highlight the main determinants of networks and prices.

In terms of service-level risk selection, I start by documenting that the coarseness of the government's risk adjustment formula leaves significant variation in expected patient profitability depending on the types of services the patient is likely to need. Then, I give evidence that provider networks tend to be narrower for less profitable services. Finally, I

[1] The authors measure physician network breadth as the fraction of physicians in a rating area who are included in a plan's network, and hospital network breadth as the fraction of hospital discharges that are represented by a plan's network.

[2] Under risk selection, insurers may attempt to disproportionately enroll healthy (profitable) patients rather than sick (unprofitable) patients by providing networks that appeal the most to the former, which are narrow networks.

show that patients tend to select insurers that have broad networks in services they are likely to need. For example, patients with cardiovascular disease are more likely to choose insurers with broad networks for cardiac care services.

In terms of costs, I document that insurers' total administrative costs (denoted "network administrative costs"), which include expenses related to establishing and maintaining a provider network (such as billing and auditing expenses), are increasing in the breadth of the network. However, per capita administrative costs are decreasing in network breadth, suggesting insurers enjoy economies of scale in the number of covered providers. This contrasts with the positive relation between negotiated prices and network breadth that incentivizes insurers to offer narrow networks. Hence, economies of scale can help explain why some insurers choose to offer broad provider networks despite other elements of the contract being tightly regulated.

Motivated by these descriptive facts, I develop and estimate a model of insurer competition in service-level provider network breadth. This model builds on prior empirical work that uses Horn and Wolinsky (1988)'s Nash-in-Nash bargaining solution to endogenize negotiated prices between buyers and suppliers holding networks fixed (e.g., Grennan, 2013; Gowrisankaran et al., 2015; Ho and Lee, 2017). In my model, however, networks are also determined endogenously. Jointly modelling network and price setting can be excessively complex, if not infeasible.[3] In my analysis, I circumvent this obstacle by redefining the problem as one in which insurers choose, rather than a network, the fraction of providers in a market that deliver a specific service. This ensures tractability while allowing for endogenous networks.

One shortcoming of my approach is that it does not capture changes in insurers' costs that are generated by changes in insurers' or providers' disagreement payoffs during bilateral negotiations. Therefore, while my model captures the key implications of a bargaining environment on insurers' costs in a reduced-form way, it cannot be used to evaluate counterfactual policies that impact these outside options.[4]

[3]Lee and Fong (2013) is one exception. Their empirical approach, nonetheless, is only showcased with simulations, and the dimensions of the problem become quickly intractable in full scale empirical applications as the one I consider here.

[4]Another limitation is that I cannot model provider heterogeneity beyond what is captured by the composition of services that they offer. In the literature, Dafny et al. (2015); Ericson and Starc (2015); Polsky et al.

I assume insurers make a one-time choice of service network breadth, recognizing that this choice will affect both current and future profits as patients age and transition between diagnoses. Expected profits combine a random utility model of demand in which individuals select insurers based on service-level provider network breadth and out-of-pocket costs, and cost functions that allow for heterogeneity in expected costs not only across individuals (risk selection), but also across insurers with varying provider network breadths (network administrative costs).

To estimate the model, I use an administrative dataset that encompasses all enrollees to the contributory healthcare system in Colombia during 2010 and 2011, which represents nearly half of the population in the country (25 million individuals) and their medical claims (650 million). My data also contain the set of providers that insurers cover for every service. Demand estimates show that, conditional on sex and age, willingness-to-pay for service network breadth varies substantially across diagnoses and services, consistent with adverse selection. Moreover, insurers' average cost function exhibits economies of scope across services.

Using the demand and average cost estimates, I recover insurers' network administrative costs. I find that these costs are substantially heterogeneous across insurers and reveal economies of scale in network breadth in line with the descriptive evidence. When comparing my model's prediction of the ratio of total costs (average costs plus network administrative costs) to total revenues against insurers' public income statements, findings show that my model has an accurate fit on untargeted moments.

To quantify the relative importance of risk selection and insurers' cost structure for equilibrium service network breadth, I conduct two sets of counterfactual exercises. To get at risk selection, I examine whether service coverage decisions respond to changes in the government's risk adjustment formula. Without any risk adjustment, I find that average service network breadth would fall 6.6%, corresponding to the exclusion of at most 5 providers from the service network. In contrast, if risk adjustment were made more granular by compensating for diagnoses, average service network breadth would increase 8.0% with effects

(2016); Dafny et al. (2017); Starc and Swanson (2021) use similar approximations to describe the relation between network breadth and premiums.

being larger among services that individuals with chronic health conditions tend to claim.

To get at the importance of insurers' cost structure, I examine whether the heterogeneity in costs can explain why some insurers choose to offer broad service networks despite risk selection. My main finding is that with homogeneous network administrative costs, average service network breadth decreases 7.3% and the decline is larger in services that individuals with chronic diseases tend to claim. However, imposing homogeneity in average costs has no impact on equilibrium network breadth.

Taken together, results indicate that risk selection and network administrative costs weigh equally in the decision to offer network breadth. Risk selection drives the choice of narrow networks while heterogeneous administrative costs drive the choice of broad networks. In my setting, risk adjustment is an effective policy to increase network coverage beyond the observed equilibrium because selection incentives are pervasive. My findings also suggest that network adequacy rules forcing insurers to cover specific providers will improve network breadth because it allows insurers to exploit their economies of scale in administrative costs.[5] More generally, the efficacy of risk adjustment and network adequacy rules will depend on the relative importance of risk selection and cost incentives.

Contributions and literature. This paper makes three contributions to the literature: first, it shows that insurers' cost structure may have an opposite effect relative to risk selection on the decision to offer service-level provider network breadth. To do this, I build on insights from Shepard (2022) who first explicitly modelled selection through star hospital coverage, and from Decarolis and Guglielmo (2017); Carey (2017); Geruso et al. (2019); Aizawa and Kim (2018) who study alternative selection mechanisms such as insurance generosity, drug formulary design, and advertising.

Second, it develops a new model of insurer competition that endogenizes network breadth across several insurers and services in a tractable way, while maintaining a relation with the underlying bargaining game. Prior work has focused on endogenizing prices holding networks fixed (Gowrisankaran et al., 2015; Ho and Lee, 2017). More recently, Ho and Lee (2019) endogenize prices and networks in the context of a monopolist insurer, and Fleitas et al.

[5] These types of network adequacy rules are widely used in Medicare Advantage, the Federal Marketplaces, and Medicaid Managed Care (see Pollitz, 2022).

(2024) extend their theoretical framework to an empirical application with several insurers and providers. Ghili (2022) and Liebman (2022) also provide alternative approximations to endogenous networks.

The third contribution is in quantifying the relative importance of risk selection and costs for equilibrium network breadth and showing how this trade-off impacts policy design. A seminal paper in this area is Ho and Lee (2016) who conclude that certain network adequacy regulations may generate costs that far outweigh the benefit to consumers. Existing literature has also studied the relation between network breadth and premiums (Dafny et al., 2017; Polsky et al., 2016; Dafny et al., 2015), while others have focused on the impact of risk adjustment on selection efforts (Brown et al., 2014; McWilliams et al., 2012; Nicholson et al., 2004) and on other elements of the insurance contract (Cabral et al., 2018; McGuire et al., 2013; Pauly and Herring, 2007).

The remainder of this paper is structured as follows: section 2 describes the institutional background and data, section 3 provides descriptive evidence of risk selection and cost variation, section 4 presents the structural model, section 5 shows estimation results, section 6 studies the impact of risk adjustment and insurers' cost structure, and section 7 concludes.

2 Institutional Background and Data

Background. The Colombian healthcare system was established in 1993 and is divided into a "contributory" and a "subsidized" regime. The first covers formal employees and independent workers who pay monthly payroll taxes (nearly 51% of the population). The second covers individuals who are poor enough to qualify and are unable to contribute (the remaining 49%). The national healthcare system has near-universal coverage, which implies that insurer competition for enrollees is zero-sum.

In this system, private insurers provide one national insurance plan. The plan covers a comprehensive list of more than 7,000 services or procedures and 673 medications as of 2010. Several aspects of this plan are strictly regulated, for example, insurance premiums are zero and cost-sharing rules are a function of the enrollee's monthly income level, but are standardized across insurers and providers.[6]

Instead, the government reimburses insurers at the beginning of every year (ex-ante) with capitated risk-adjusted transfers, and at the end of every year (ex-post) with the High-Cost Account. The ex-ante risk adjustment formula controls for sex, age group, and municipality of residence, but it does not include diagnoses. Appendix 1 describes how this risk-adjusted transfer is calculated. Because of the coarsely defined risk pools, the ex-ante formula poorly fits realized healthcare costs. For example, Riascos et al. (2014, 2017) find that the R^2 of the government's formula is only 0.017. The High-Cost Account compensates insurers that enroll an above-average share of people with certain diagnoses, and reimbursements come from insurers that enroll a below-average share.[7] These ex-post transfers typically represent only 0.4% of total ex-ante transfers per insurer, suggesting these ex-post transfers do not provide much risk adjustment.

Provider networks and negotiated prices are the only dimensions in which insurers can differentiate. Insurers form provider networks separately for each health service offered in the national plan and negotiate prices with in-network providers. Service level negotiations imply, for example, that insurers can choose to offer a broad network for orthopedic care but a narrow network for cardiology.[8]

Enrollees can access their insurer's network in any state of the country (states are similar in size to Metropolitan Statistical Areas in the U.S.). Those who reside in peripheral municipalities (similar to counties in the U.S.) typically travel to the capital city of their state to

[6] Cost-sharing in the national insurance plan follows a three-tiered system. As of 2010, for individuals earning less than 2 times the monthly minimum wage (MMW) the coinsurance rate equals 11.5%, the copay equals 2,100 pesos, and the maximum out-of-pocket amount in a year equals 57.5% times the MMW. This corresponds to an actuarial value of 92%. For those with incomes between 2 and 5 times the MMW, the coinsurance rate is 17.3%, the copay is 8,000 pesos, and the maximum out-of-pocket amount is 230% times the MMW. The associated actuarial value is 84%. Finally, for people with incomes above 5 times the MMW, the coinsurance rate equals 23%, the copay 20,900 pesos, and the maximum out-of-pocket amount is 460% times the MMW, corresponding to an actuarial value of 78%. The average exchange rate during 2011 was $1,847 COP/USD.

[7] Diseases compensated by the High-Cost Account include: cervical cancer, breast cancer, stomach cancer, colon cancer, prostate cancer, lymphoid leukemia, myeloid leukemia, hodgkin lymphoma, non-hodgkin lymphoma, epilepsy, rheumatoid arthritis, and HIV-AIDS. See Resolution 0248 of 2014 from the Ministry of Health.

[8] Although the government does stipulate a set of network adequacy rules to guarantee appropriate access to health services, these rules are very coarse and apply only to the provision of primary care, urgent care, and oncology (MinSalud, 2016).

receive care. While consumers can visit out-of-network providers, the insurer does not cover any of the healthcare costs associated with those visits. Hence, individuals' enrollment and healthcare consumption decisions depend on the provider network that their insurer offers in their state of residence.

During 2011, there were 23 private insurers participating in the contributory healthcare system across the 33 states of the country; 10 of these insurers covered 95% of enrollees. Moreover, participation in the subsidized system is not required for insurers in the contributory system.

Data. At the end of every year, insurers report to the government all the health claims made through the national insurance plan that they reimbursed providers in their network for. The data for this paper are: individual-level insurer choices for all those enrolled in the contributory system during 2010 and 2011 (approximately 25 million people), individuals' medical claims (approximately 650 million claims), and insurers' provider listings per healthcare specialty.

The enrollment files have basic demographic characteristics like sex, age, municipality of residence, and enrollment spell length in the year. Although I do not observe individual income per month, using publicly available, aggregate income data from enrollees to the contributory system, I assign to each individual the average income of their municipality, sex, and age cell. The health claims data report date of provision, procedure code, negotiated procedure price, provider, insurer, and diagnosis codes following the International Classification of Diseases (ICD).

Every claim is associated to a 6-digit procedure code from the national health insurance plan. These codes can be mapped to the healthcare specialties that insurers and providers bargain over, contained in the provider listings. These data report 150 unique specialties, which I aggregate up to 20 "services," corroborating with network inclusions inferred from claims.[9] Some examples of specialties in the data are cardiology, pediatric cardiology, and

[9] Across 2010 and 2011, 90% of provider-services in the provider listing files have at least one claim, while 12% of provider-services with claims appear in the provider listings. This difference is explained by the fact that provider listings are submitted at the beginning of the year, but providers can be included to the network mid-year. The final network data includes all insurer-provider-services from the provider listing files (corresponding to 70% of observations) plus provider-services in the claims data that have at least 10 claims with a given insurer (representing the remaining 30%). My model estimates are robust to more stringent conditions on the minimum number of claims to be considered in-network.

cardiovascular surgery, which I aggregate to cardiac care services. Other specialties are intensive care unit, intermediate care unit, neonatal intensive care unit, and hospitalization, which I aggregate to hospital admission services. Appendix 2 provides the final list of services and an excerpt from the provider listings data that shows that service coverage varies substantially across insurers within a given provider.[10]

Sample restrictions. I focus on the sample of individuals aged 19 or older, enrolled with one of the 10 largest insurers. Of these individuals, 2/3 have continuous enrollment spells or no gaps in enrollment within a year. Of the continuously enrolled, 2/3 are *current enrollees*, that is, individuals who are enrolled throughout 2010 and 2011. The remaining 1/3 are *new enrollees* or individuals who enroll for the first time in 2011. Because there is near universal coverage, new enrollees to the contributory system can be individuals who move from the subsidized system after they find a job, turn 18 and choose a different insurer from their parents', or for some reason were uninsured for 12 continuous months.[11]

Switching rates. Insurers compete mainly on attracting these new enrollees because consumer inertia in this system is substantial. Conditional on staying in the contributory system, less than 6% of all enrollees switch their insurer from 2010 to 2011. Moreover, of those who have continuous enrollment spells, less than 2% switch their insurer.[12] Although new enrollees are relatively healthier than current enrollees (see Appendix Table 6), their patterns of healthcare utilization and spending and their ranking of diseases follow the same trends as for current enrollees (see Appendix Figure 6). Thus, when making network coverage decisions, insurers take into account the fact that individuals who enroll today, will stay with the same insurer tomorrow and their costs and diagnoses will evolve in parallel to their stock

[10] Service level negotiations could generate clinical chaos if, for example, a patient with diabetes is admitted to a hospital for a cardiac episode but the insurer does not cover diabetes care at that hospital. This kind of care fragmentation has received substantial media attention for an issue that came to be called the "rounds of death" (*paseo de la muerte* in Spanish), where patients would have to go from clinic to clinic to be treated and died in the process. To avoid these issues and based on conversations with experts, insurers tend to cover all the services at large clinics and hospitals where inpatient admissions occur. But service coverage varies at smaller providers where outpatient care is delivered. Appendix Figure 2 corroborates this fact.

[11] Even if new enrollees in 2011 had enrollment before the start of my sample period in 2010, Decree 806 of 1998 and Decree 1703 of 2002 established that after three continuous months of non-payment of tax contributions, a person would be disenrolled and lose any information so far reported to the system. Enrollment after non-payment is therefore a "fresh-start" in the contributory system. Moreover, in 2011 only around 500 thousand enrollees (out of 25 million) switched from an insurer in the subsidized system that also had presence in the contributory system.

[12] Appendix Figure 5 presents the switch-in rates for each insurer in the different samples.

of enrollees.

3 Descriptive Evidence

Using the demographic information contained in the enrollment files, I recover the ex-ante risk-adjusted transfers that each insurer received for each of its enrollees. Ex-ante reimbursements range from 162.2 thousand pesos ($87) (for males aged 15-18) to 2.2 million pesos ($1,191) (for females aged 75 or older), while realized costs range from 0 to over 300 million pesos ($162K). The enormous variation in healthcare costs across patients suggests that there may be risk selection incentives and possibly heterogeneous costs across insurers.

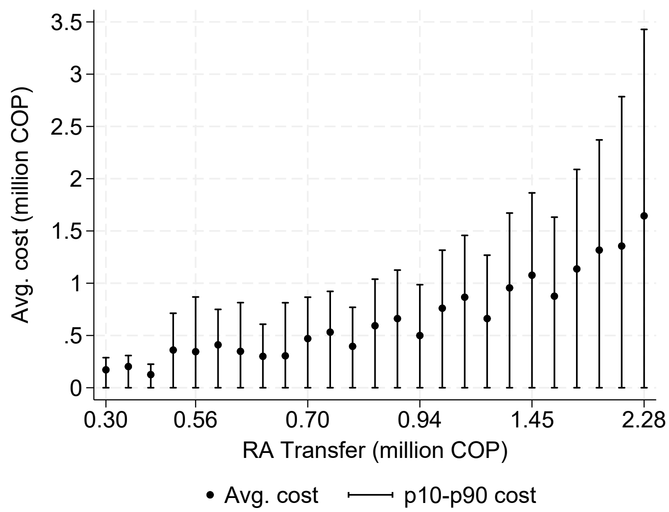


Figure 1: Healthcare cost by risk-adjusted transfer

Note: Figure presents mean and 10th and 90th percentiles of annual healthcare cost across all individuals enrolled in the contributory system, conditional on the government's ex-ante risk-adjusted transfer. Healthcare costs and risk-adjusted transfers are measured in millions of COP. The average exchange rate in 2011 was 1,847 COP/USD.

Figure 1 shows that this variation in healthcare costs holds even after controlling for the risk-adjusted transfers. The figure shows that the mean and the variance (as reflected in the difference between 10th and 90th percentiles) of healthcare costs increase with the government's reimbursement. The rising trend in average costs indicates that insurers can have incentives to enroll costly individuals because they can receive higher reimbursements from the government. The rising trend in variance indicates that there is scope to select consumers in the upper tail of the distribution who are more likely to be overcompensated by the risk adjustment formula (as in Brown et al., 2014).

3.1 Measuring Network Breadth

If insurers respond to selection incentives using their provider networks, then differences in healthcare costs should appear as differences in service network breadth. I define service network breadth as the fraction of providers in a market offering a particular service that are covered by the insurer.[13] A provider is either a hospital, clinic, or physician practice. Because the denominator is the number of providers that are able to deliver a service, this measure of network breadth will capture differences across insurers in service-level negotiations with providers rather than provider-level decisions on whether to offer a specific service line.

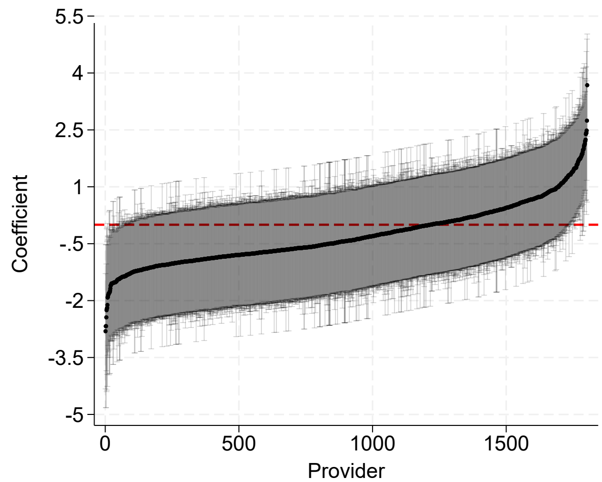


Figure 2: Differences in unobserved provider quality

Note: Figure shows the coefficients and 95% confidence intervals of provider fixed effects in a regression of average provider prices per service on provider and service fixed effects. An observation in this regression is a provider-service combination. Average prices are calculated over the insurers that cover each provider.

Summarizing insurers' networks with this measure of network breadth is appropriate in my setting because, *conditional on the service*, differences across providers in terms of unobserved quality tend to be negligible. Hence, horizontal differentiation in networks within a service is not central for the analysis. For instance, using a regression of average provider prices on provider and service fixed effects, Figure 2 shows that 99% of provider fixed effects

[13] Although it is mandatory that insurers cover at least one provider for every service in the national insurance plan, network breadth can be determined by the type of consumers that insurers want to risk select upon.

are statistically equal to zero. To the extent that prices reflect quality, this regression evidences that quality differences across healthcare providers are indeed small after controlling for the services they provide.

Table 1 shows that there is substantial variation in my measure of service network breadth across insurers. Even with poor risk adjustment, insurers N and I choose to offer broad service networks, while insurers H and E choose narrow networks.[14] Insurers that have relatively broad networks also tend to have larger market shares, consistent with consumers having a preference for broad provider coverage.

Table 1: Distribution of service network breadth and national market shares in 2011

Insurer	Service network breadth			National market share
	mean (1)	p25 (2)	p75 (3)	(4)
Insurer N	0.572	0.380	0.750	0.128
Insurer I	0.536	0.333	0.714	0.218
Insurer B	0.393	0.214	0.500	0.093
Insurer D	0.380	0.234	0.465	0.051
Insurer J	0.373	0.211	0.500	0.154
Insurer L	0.254	0.064	0.373	0.045
Insurer C	0.238	0.111	0.333	0.040
Insurer K	0.208	0.091	0.267	0.071
Insurer F	0.202	0.038	0.364	0.019
Insurer A	0.176	0.096	0.231	0.017
Insurer G	0.160	0.085	0.192	0.080
Insurer M	0.146	0.049	0.211	0.030
Insurer H	0.127	0.032	0.133	0.016
Insurer E	0.098	0.025	0.111	0.038

Note: Table presents the mean and 25th and 75th percentiles of service network breadth as well as the national market shares in 2011 for each insurer. An observation to construct summary statistics of service network breadth is a combination of insurer, service, and market.

3.2 Network Breadth as a Means of Risk Selection

Variation in service network breadth exemplifies differences in selection efforts and costs *across insurers*. In this subsection, I characterize selection incentives *across services* by

[14] Although by U.S. standards some of these insurers would be considered as having ultra-narrow networks, these standards are based on the coverage of large hospitals (see Bauman et al., 2014). My measure of service network breadth is instead defined over all types of provider organizations. As long as a provider is certified by the Ministry of Health, this provider will be included in my measure. Dafny et al. (2017) report that provider networks in the U.S. based on a similar definition also tend to be much narrower than hospital networks.

replicating figures in Geruso et al. (2019) with data from *all* enrollees in the contributory health system.

Figure 3 shows whether the current risk adjustment systems are effective at neutralizing service-level risk selection. The figure plots the average cost per enrollee against the average revenue per enrollee conditional on patients who make claims for each service. Every circle represents a service weighted by the number of patients who make claims for it. Patients who make claims for several services will be represented in several circles, while patients who make zero claims (and are the most profitable) are not represented in this figure. The red line is the 45 degree line, which splits the space into services that are overcompensated by the risk-adjusted transfers (above the line) and those that are undercompensated (below the line). The main takeaway is that patients who make any claim are likely to be unprofitable; but this is especially true for patients who have claims in certain services such as cardiac care, renal care, and hospital admissions, which are located toward the right of this figure.

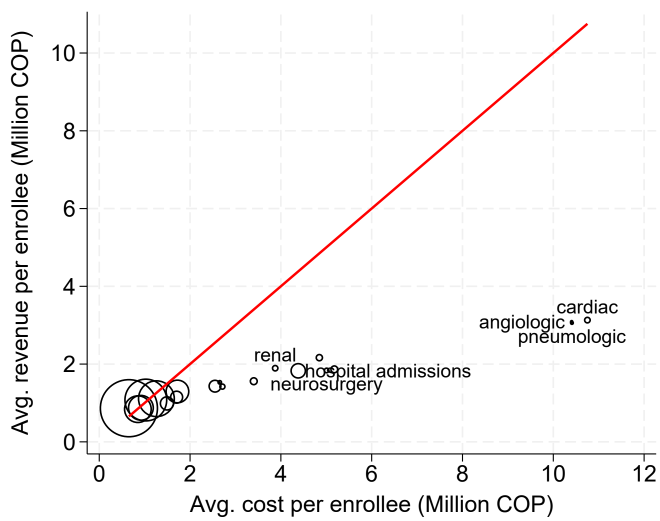


Figure 3: Service-level selection incentives after risk adjustment

Note: Figure presents a scatter plot of average annual revenue and average annual cost per enrollee across all individuals in the contributory system. Each dot is a service weighted by the number of individuals who make claims for the service. Revenues are calculated as the government ex-ante plus ex-post risk-adjusted transfers, plus revenues from copayments and coinsurance rates. The red line is a 45 degree line. One enrollee can be represented in several dots if they make claims for different services. Enrollees who make zero claims are not represented in this figure.

The existence of services that are outliers in terms of profits per enrollee suggests a scope for insurers to engage in service-level risk selection through their choice of provider networks. One way to test whether the data are consistent with selection at the service level is to show whether service network breadth covaries with the profitability of a service, a version of the positive correlation test in Chiappori and Salanie (2000).

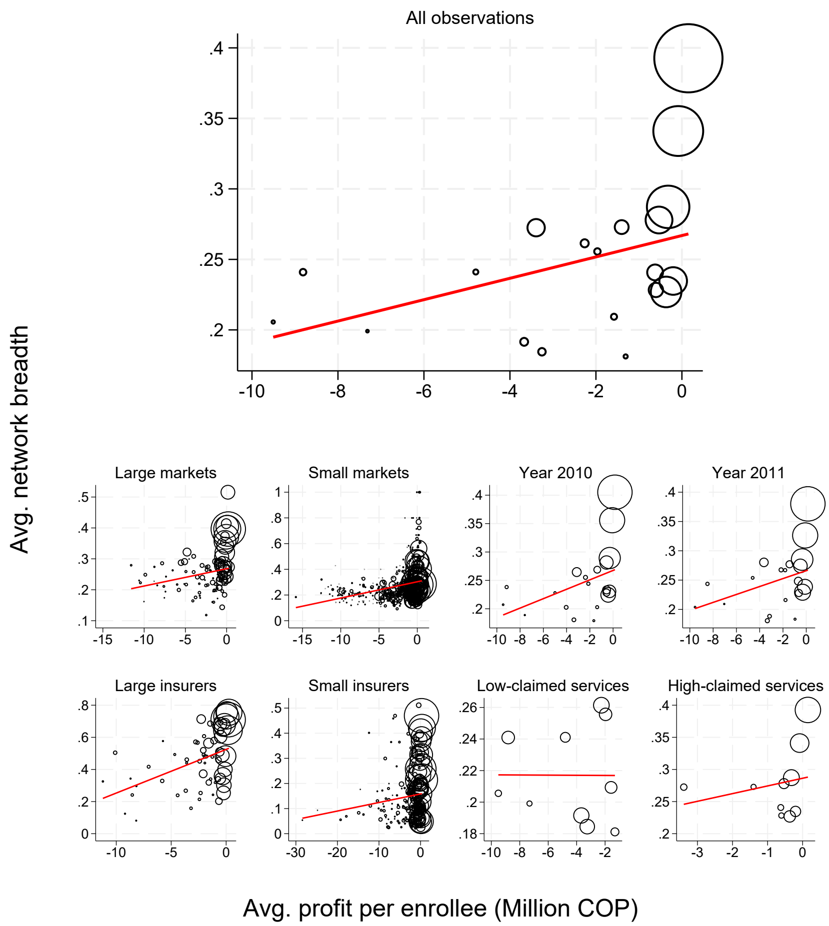


Figure 4: Correlation between network breadth and service profitability

Note: The top panel of the figure presents a scatter plot of average service network breadth and average profit per enrollee across all individuals in the contributory system. Each dot is a service weighted by the number of individuals who make claims for the service. Profits per enrollee are calculated as government's ex-ante plus ex-post risk-adjusted transfers, plus revenues from copayments and coinsurance rates, minus total healthcare costs. The red line corresponds to a linear fit. One enrollee can be represented in several dots if they make claims for different services. Enrollees who make zero claims are not represented in this figure. The bottom panel of the figure presents a scatter plot of average service network breadth and average profit per enrollee conditional on the set of 5 largest markets (states 25, 11, 76, 05, and 13) and the rest of markets, conditional on 2010 and 2011, conditional on the three largest insurers (Insurers I, J, N) and the rest of insurers, and conditional on services with below- and above-median number of enrollees who make claims for it (the median equals 550,000 enrollees).

Figure 4 plots the average accounting profit per enrollee against average service network breadth across insurers and markets. Average profits are calculated conditional on patients who make claims for each service. The red line corresponds to a linear fit and shows that relatively profitable services, such as general medicine and laboratory, tend to have broader networks than relatively unprofitable services, such as cardiac care and renal care. This correlation holds along several dimensions considered in the bottom panels of the figure and is not necessarily driven by services with few claims. Appendix Figures 3 and 4 unpack some of the variation in network breadth per service to give an example of those that are likely to be under-covered across all markets, such as cardiac care and renal care.

Another explanation for why complex services tend to be under-covered is that total demand for those services is relatively low, which would rule out risk selection as driver of network breadth choices. To separate the importance of risk selection, I estimate the correlation between insurer market shares in the number of new enrollees with chronic diseases and network breadth for the services those patients are most likely to need. Table 2 shows that market shares in initial choices are positively correlated with service network breadth, suggestive of adverse selection. For example, insurers with relatively broad networks for renal care tend to have a higher share of new enrollees with renal disease.

Switching decisions. Prior papers on selection in health insurance markets leverage enrollees' switching decisions to test for adverse selection (e.g., Shepard, 2022; Gruber and McKnight, 2016; Newhouse et al., 2015; Brown et al., 2014; Einav et al., 2013). Although this type of analysis would be under-powered in my setting (since, conditional on staying in the contributory system, only 6% of all enrollees switch their insurer between 2010 and 2011), I find evidence consistent with selection on switching decisions in Appendix Table 5. For instance, healthy enrollees are more likely to switch out of insurers with broad networks for primary care, while patients with cardiovascular disease are more likely to switch out of insurers with narrow networks for cardiac care.

Table 2: Market share in initial choices and network breadth

	(1) Healthy	(2) Cancer	(3) Diabetes	(4) Cardio	(5) Renal
Network breadth	0.411	0.332	0.409	0.297	0.327
	(0.022)	(0.042)	(0.022)	(0.046)	(0.064)
Observations	312	312	312	312	312

Note: Table presents OLS regressions using as outcome insurer market shares and as covariate service network breadth. Column (1) uses the sub-sample of individuals without diagnoses and network breadth for general medicine. Column (2) uses the sub-sample of individuals with cancer and network breadth for chemotherapy. Column (3) uses the sub-sample of individuals with diabetes and network breadth for laboratory. Column (4) uses the sub-sample of individuals with cardiovascular disease and network breadth for cardiac care services. Column (4) uses the sub-sample of individuals with renal disease and network breadth for renal care services. All specifications include market fixed effects. An observation in this table is an insurer-market for 2011. Standard errors in parenthesis are clustered at the market level.

3.3 Network Administrative Costs

In addition to service network breadth choices being characterized by risk selection, network breadth is also related to heterogeneous administrative costs, which I refer to as *network administrative costs*. These are costs associated not only with adding providers to the network and the bargaining costs this may entail (such as managers' and lawyers' wages), but more importantly with maintaining the network, which involves billing and auditing expenses.

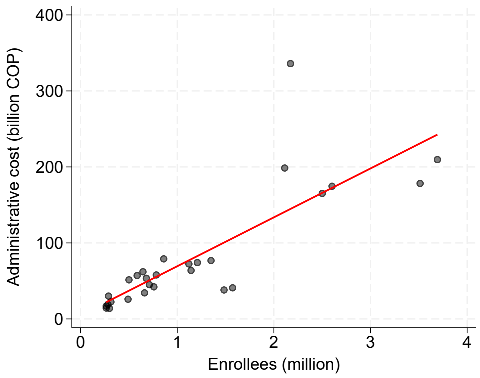
The Colombian Ministry of Health mandates that insurers conduct systematic and periodic audits of in-network healthcare providers in terms of their billing practices and quality of care (Law 1438 of 2011). As a result, insurers maintain substantial billing and auditing departments dedicated to contesting claims submitted by their providers. Such disputes—referred to as *glosas* in Spanish—can result in significant delays in payments and have become a notable source of contention between insurers and providers.

Billing and auditing costs are a major component of healthcare spending in systems with managed care competition. For instance, recent studies in the U.S. indicate that these costs are primarily linked to managing claims from providers (Cutler, 2020; Chernew and Mintz, 2021). Therefore, administrative costs vary with provider network breadth: as the number of in-network providers increases, the insurer faces a greater volume of auditing procedures and billing activities. However, these costs do not vary with enrollee characteristics that determine selection into insurers such as their health status.

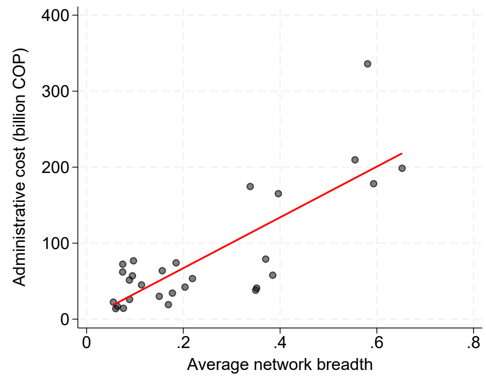
To describe the importance and magnitude of network administrative costs, I use data from insurers' public income statements submitted to the National Health Superintendency in 2010 and 2011.[15] Although I do not use this information for estimating my model in

[15] In 2007, the National General Accounting Agency (*Contaduría General de la Nación, CGN*) adopted the General Plan of Public Accounting (*Plan General de Contabilidad Pública, PGCP*) which standardized balance sheets and income statements for companies and agencies that deliver government services in Colombia, including health insurers (see Resolution 354 of 2007, CGN). Financial reports were regulated following the International Financial Reporting Standards (IFRS), which are used in over 140 countries, including the U.S. The PGCP describes each account that must be reported to the government and the file structure (PGCP). Health insurers are required to submit their income statements by the end of every year to the National Health Superintendency. The government uses insurers' income statements to determine compliance with capital requirements and solvency margins. Non-compliance with these financial measures can lead to an insurer termination. The 2010 and 2011 files I use to produce Figure 5 are available at Supersalud, 2011 and Supersalud, 2010. I use information from account 51 in these files reporting *Gastos de Administración*. According the PGCP, these are operational administrative expenses that include salaries and wages (account 510503, 510506), fees for statutory audit and external auditing (account 511015), fees for legal and financial

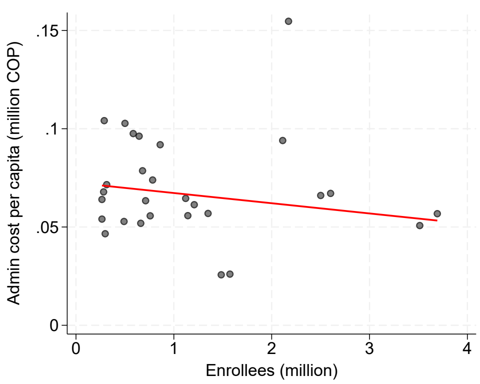
section 4, the data are relevant to describe the factors that affect network breadth choices. Each observation in this dataset corresponds to an insurer-year. The income statements reveal that total administrative costs vary significantly, ranging from 17 to 336 billion pesos across insurers.



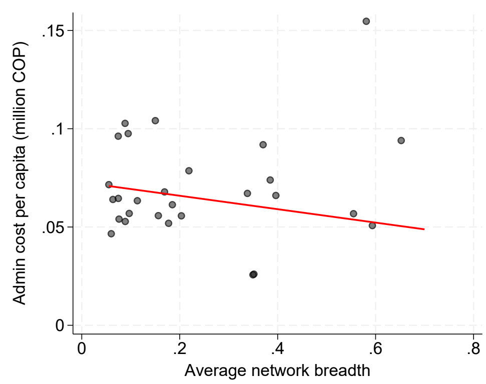
(a) Total administrative cost



(b) Total administrative cost and networks



(c) Administrative cost per capita



(d) Administrative cost per capita and networks

Figure 5: Correlates of insurers' administrative costs

Note: Panel A presents a scatter plot of total administrative costs against the number of enrollees and Panel B against average network breadth (averaged across services and markets). Panel C shows a scatter plot of total administrative cost per capita against the number of enrollees and Panel D against average network breadth (averaged across services and markets). In each panel, a dot is an insurer-year. Red lines correspond to linear fits. In panels C and D the linear fits exclude the outlier with administrative cost per capita over 0.15 million COP. Total administrative costs are obtained from insurers' public income statements submitted to the National Health Superintendency in 2010 and 2011.

Figure 5, Panels A and B show that insurers' total administrative costs are positively correlated with the number of enrollees and with average network breadth across services and

consulting (account 511010), lease of medical or scientific goods and equipment (account 512015), and lease of building and office space (account 512010). The first three accounts describe billing and auditing expenses while the latter two describe expenses in physical inputs that insurers would incur when vertically integrated with providers.

markets. These correlations support the simpler hypothesis that larger insurers have higher administrative costs, which may have nothing to do with the breadth of their networks. However, Panels C and D illustrate that per capita administrative costs are *decreasing* in the number of enrollees and in average network breadth (except for the outlier to the right), suggesting insurers enjoy economies of scale in the number of covered providers.

These economies of scale can potentially overcompensate the healthcare costs associated with providing a broad network as seen in Figure 6. The figure shows that average healthcare costs per service (in black) are increasing, while per-capita administrative costs (in red) are decreasing in network breadth.

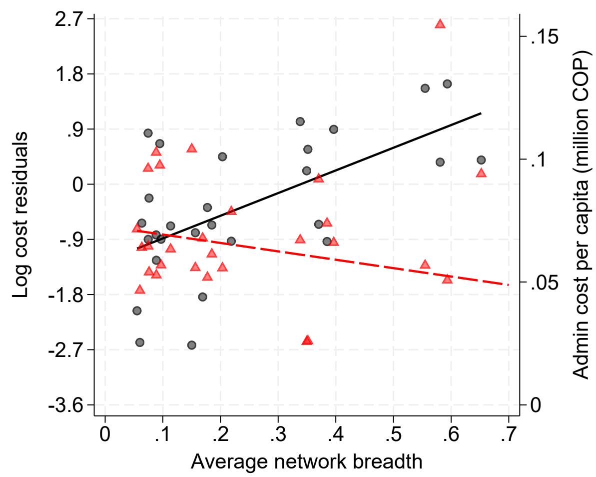


Figure 6: Relation between average costs, administrative costs, and network breadth

Note: Figure shows a scatter plot of the log of total healthcare cost demeaned at the market-service-year against average network breadth (averaged across services and markets) in black. Each black marker is an insurer-year from 2010 to 2011 and is measured in the left vertical axis. The solid black line is a linear fit. On the right vertical axis and in the red markers, the figure shows a scatter plot of the log of administrative cost per capita against average network breadth (averaged across services and markets). Each red marker corresponds to an insurer-year from 2010 to 2011. The dashed red line corresponds to a linear fit and excludes the outlier with administrative cost per capita over 0.15 million COP.

My formulation of insurers' network administrative costs is similar to Tilipman (2022)'s specification of employers' fixed cost of offering a particular health plan and to Prager and Tilipman (2020)'s fixed contracting cost between insurers and hospitals that determines whether they enter a formal contract. The difference in my setting is that the main components of network administrative costs stem from billing and auditing expenses rather than from fixed bargaining costs.

In Colombia, where most aspects of the health insurance plan are closely regulated, differences in these network administrative costs across insurers could explain why equilibrium service network breadth is asymmetric. For example, if consumers' preferences for network breadth in cardiology and the average cost of offering this service are homogeneous across insurers, then we would predict that all insurers choose the same network breadth for cardiology, which is not what we see in the data. Thus, network administrative costs that vary across insurers and services may rationalize the heterogeneity in service network breadth.

4 Model

Motivated by the descriptive evidence, in this section I develop a model of the Colombian insurance market to decompose risk selection and insurers' cost structure as potential mechanisms for service network breadth. I limit my analysis sample moving forward to individuals who have continuous enrollment spells and reside in the main states of the country.[16] This sample distinguishes consumers whose choices are not conflated by variation in income, job loss, or informality, and distinguishes markets with enough variation in service network breadth. Appendix 4 shows some summary statistics and replicates all the descriptive evidence presented earlier for this sample. The appendix also demonstrates stability of an insurer's patient mix, since the trends in healthcare utilization and spending among new enrollees are parallel to those of current enrollees. This stability is relevant for my model since I assume consumers are fully inertial.

4.1 Foundations and Relation to Prior Work

My model is specified over insurers' decision to offer service network breadth. Although substantial research in the U.S. suggests that patients care strongly about whether their preferred provider is included in the network (e.g., Ho, 2006; Shepard, 2022), my model characterizes the identities of specific providers with the composition of services that they offer in an effort to circumvent tractability issues that arise when jointly modelling network

[16] These are the states where the 13 metropolitan areas are located: Bogotá (11), Antioquia (05), Valle del Cauca (76), Atlántico (08), Santander (68), Caldas (17), Risaralda (66), Norte de Santander (54), Nariño (52), Tolima (73), Córdoba (23), Bolívar (13), Meta (50). Numbers in parenthesis correspond to state codes from the National Administrative Department of Statistics (DANE).

and price setting.

Following McFadden (1996), most of the literature to date uses random utility models that allow for preference heterogeneity across providers to derive patients' expected utility for their insurer's network. This expected utility is fed into models of insurer and provider Nash bargaining with the goal of endogenizing negotiated prices (Gowrisankaran et al., 2015; Ho and Lee, 2017). The structural unobservable of the Nash-in-Nash surplus function is typically the provider's marginal cost, but off-equilibrium prices in the event that the insurer and the provider disagree are assumed to be fixed at their equilibrium values (Horn and Wolinsky, 1988).

When trying to endogenize insurers' decision of which providers to include in its network in addition to negotiated prices, the assumption that disagreement payoff prices are fixed can be too strong. Nash-in-Nash proves to be infeasible in this case because both the provider's marginal cost and the off-equilibrium price are unobserved, resulting in a system with more unknowns than equations.[17] Ho and Lee (2019) and Lee and Fong (2013) provide some solutions to this problem. The first group of authors apply their model to a setting with a monopolist insurer and the second group of authors showcase their model through simulations. Here I derive an alternative solution by redefining the problem of *which* providers to include in the network and at what *price* as a problem of *how* many providers to include and at what *cost*.

Appendix 5 shows that this model is a reduced-form representation of one with provider choice and Nash-in-Nash bargaining after imposing two restrictions: first, that providers are price-takers and, second, that insurer costs do not depend on the disagreement payoffs. That is, my model assumes that changes in insurer $g \neq j$'s costs do not affect insurer j's costs when j excludes a provider from its network. Although these restrictions may seem prohibitive, the model captures the key features of a bargaining environment while maintaining computational tractability: that including a provider in insurer j's network will increase j's costs if enrollment rises, and that removing the provider will lower insurer j's costs if providers are highly substitutable.

[17] Crawford and Yurukoglu (2012) discuss some of the limitations of allowing for endogenous networks in Nash-in-Nash in the context of television markets.

My model is an exact representation of markets where providers are homogeneous *conditional on the service*, an assumption that appropriately characterizes the Colombian healthcare system as seen in Figure 2. When applied to settings with heterogeneous providers, my model would derive a lower bound on consumer surplus. Moreover, the model allows me to quantify by how much would service network breadth and insurer costs change under certain counterfactual policies, but it does not allow me to say which providers are being included or excluded from the networks or to evaluate policies that impact disagreement payoffs.[18]

4.2 Insurer Demand

I model insurer demand in the sample of new enrollees in 2011, who do not experience inertia when making their first enrollment choice. Assume a new enrollee i living in market m is of type θ. The individual assigns a probability $q_{\theta k}$ to claiming each of the $k = \{1, ..., K\}$ services. An individual's type is given by a combination of sex, age group (19-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, ≥75), and diagnosis $d \in D =$ {cancer, diabetes, cardiovascular disease, pulmonary disease, renal disease, other chronic disease, no diseases}.[19]

I assume new enrollees know their diagnoses before making their *first* enrollment choice in the contributory system. This is possible either because they are aware of medical family history or because they went to the doctor and received a diagnosis while uninsured. Once they choose this *first* insurer, they remain with it for subsequent years. Individuals' private information regarding health status, which I observe in the data, implies that selection in my model will occur on observable, un-reimbursed (or poorly reimbursed) characteristics such as diagnoses.[20]

[18]Estimating a richer model of provider choice and Nash-in-Nash bargaining is also relatively infeasible in my setting because it involves estimating 20 demand functions (one for each service) and because the size of consumer choice sets for relatively common services such as general medicine equals 179 in the largest market. This choice set would imply solving $179 \times |J_m|$ bilateral negotiations in this market and for general medicine alone.

[19]Diagnoses in the list are groupings of ICD-10 codes following Riascos et al. (2014). These diagnoses were chosen for being the most expensive in Colombia and thus the most likely to be undercompensated by the current risk adjustment formula. For example, the most expensive patients with renal disease had annual healthcare cost of over 55 million pesos in 2011, more than 100 times the monthly minimum wage. For individuals with several comorbidities, I assign the most expensive disease.

[20]Results are also robust to a version of the model where individuals are uncertain about their diagnoses

Denote by u_{ijm} the indirect utility of a new enrollee i in market m for insurer j, which takes the following form:

$$u_{ijm} = \beta_{im} \sum_k q_{\theta k} H_{jkm} - \alpha_i c_{\theta jm}(H_{jm}) + \phi_j + \varepsilon_{ijm} \tag{1}$$

where $\beta_{im} = (x_i \;\; x_m)'\beta$ and $\alpha_i = x_i'\alpha$. The vector x_i includes dummies for sex, age group, and diagnosis, and x_m are market dummies. The average out-of-pocket cost of consumer type θ at insurer j is given by $c_{\theta jm}$ and depends on the insurer's vector of service network breadth $H_{jm} = \{H_{jkm}\}_{k=1}^{K_m}$. The coefficient ϕ_j is an insurer fixed effect that captures unobserved insurer quality that is common across individuals and markets. Finally, ε_{ijm} is an *iid* unobserved shock to preferences assumed to follow a type I extreme value distribution.

Average out-of-pocket costs are the sum of coinsurance payments, copays, and tax contributions to the system:

$$c_{\theta jm} \equiv \text{Coins}_{\theta jm} + \text{Copay}_{\theta jm} + \text{Tax}_{\theta} = r_\theta AC_{\theta jm}(H_{jm})$$

Coinsurance payments and copays are indexed by j because they are a function of the insurer's negotiated service prices with in-network providers. These prices are correlated with service network breadth because an insurer's bargaining position depends on how many providers it has included in the network.[21] I capture this correlation by noting the pass-through of insurers' costs to consumers' out-of-pocket costs via cost-sharing. Out-of-pocket costs equal the individual's coinsurance rate times the insurers' average cost per enrollee, which in turn depends on service network breadth. This dependence is needed to rationalize the existence of narrow network insurers in equilibrium since myopic, healthy new enrollees will disproportionately choose narrow-network insurers with lower implied out-of-pocket costs.

The probability that a consumer type θ makes a claim in service k is calculated outside of the model as the average prediction of a logistic regression, which implies that these

(see Appendix Table 17).

[21] Appendix Table 8 shows that negotiated prices and healthcare costs per service are positively correlated with service network breadth.

probabilities are non-zero for every θk. Appendix 6.1 explains this procedure in more detail. Given the distribution of the preference shock, the probability that consumer i in market m enrolls with insurer j is:

$$s_{ijm}(H_m) = \frac{\exp\left(\beta_{im}\sum_k q_{\theta k}H_{jkm} - \alpha_i c_{\theta jm}(H_{jm}) + \phi_j\right)}{\sum_{j'\in\mathcal{J}_m}\exp\left(\beta_{im}\sum_k q_{\theta k}H_{j'km} - \alpha_i c_{\theta j'm}(H_{j'm}) + \phi_{j'}\right)}$$

Identification. To identify the parameters associated with service network breadth, I rely on variation in market demographics across markets, which generates exogenous variation in the claim probabilities. I also rely on within-market variation in insurer market shares. For example, if an insurer offers the same network breadth for cardiac care in two different markets, but one of these markets has a higher prevalence of cardiovascular conditions, then we should observe higher insurer demand in the market where people are relatively sicker. Appendix Table 9 shows that service network breadth is uncorrelated with the claim probabilities conditional on the consumer type, providing suggestive evidence of the exogeneity of market demographics and of the relatively small impact of moral hazard (i.e., consumers are not more likely to make claims the broader is the network).

Identification is threatened if service network breadth is correlated with unobserved insurer quality or unobserved consumer characteristics, such as their valuation for specific providers. Service network breadth could also be correlated with how good the insurer is in processing health claims. These types of unobserved insurer characteristics potentially do not vary across markets conditional on the consumer type. Therefore, the inclusion of insurer fixed effects allows me to identify preferences for network breadth from the exogenous variation in market demographics.

To identify the parameters associated with the out-of-pocket cost, I use variation in income across patients within a market, which generates exogenous variation in the coinsurance rates. This variation may not be sufficient for identification if negotiated service prices are correlated with unobserved provider quality. For example, if an insurer covers a star hospital, demand and negotiated prices for that insurer will be relatively high across all income groups, and my model would interpret consumers as having low sensitivity to out-of-pocket costs. This endogenous variation is specific to an insurer, hence the inclusion of insurer fixed effects helps me isolate the variation in out-of-pocket costs that is due to the coinsurance rates. I also conduct robustness checks in section 5 to verify that potential differences in provider quality conditional on the service are irrelevant for my results.

4.3 Insurer Average Costs per Enrollee

I approximate the expected cost of type-θ individuals as the average cost across all consumers i that are of type θ. Then, I model the logarithm of average cost per consumer type as a quadratic function of service network breadth as follows:

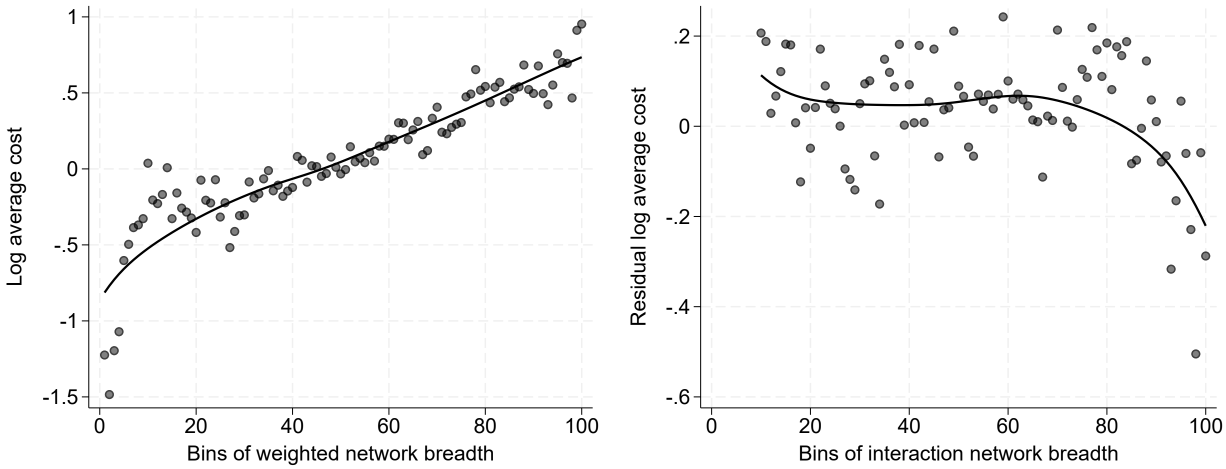
$$\log(AC_{\theta jm}(H_{jm})) = \tau_0\Big(\sum_{k}^{K_m} q_{\theta k} A_k\Big) + \tau_1\Big(\sum_{k}^{K_m} q_{\theta k} H_{jkm}\Big) + \frac{1}{2K_m}\tau_2 \sum_{k}^{K_m}\sum_{l\neq k}^{K_m} q_{\theta k} q_{\theta l} H_{jkm} H_{jlm} + \lambda_\theta + \eta_m + \delta_j + \varepsilon_{\theta jm} \quad (2)$$

where K_m is the number of services available in market m; A_k is the government's reference price for service k, guiding negotiations between insurers and providers (explained in more detail in Appendix 7); λ_θ, η_m, and δ_j are consumer type, market, and insurer fixed effects, respectively; and $\varepsilon_{\theta jm}$ is white noise.[22]

Equation (2) captures how consumers of different types will be associated with different average costs to the insurer conditional on service network breadth. This reduced-form approach is very much in the spirit of Tebaldi (2024) who also models insurers' expected costs as an exponential function of consumer characteristics and generosity of coverage. My choice of functional form is supported by the underlying bargaining game as seen in Appendix 5 and by the empirical relationship between log average costs with service network breadth and with the interaction of network breadth across pairs of services as seen in Figure 7, Panels A and B, respectively.[23]

[22] My specification for insurers' log average cost per enrollee is greater than the log average cost obtained from aggregating service-level costs per enrollee with weights given by the claim probabilities.

[23] The micro-foundation in Appendix 5 shows that the average cost has insurer-specific slopes with respect to service network breadth. Equation (2) has one slope for all insurers, which is less demanding of the data.



(a) Service network breadth

(b) Service network breadth interactions

Figure 7: Empirical relation between average costs and networks

Note: Panel A presents a scatter plot of the log of average costs per enrollee by percentiles of weighted network breadth $\sum_k q_{\theta k} H_{jkm}$. Panel B presents a scatter plot of the residual log of average costs per enrollee after controlling for weighted network breadth by percentiles of the measure of scope economies $\sum_k^{K_m} \sum_{l \neq k}^{K_m} q_{\theta k} q_{\theta l} H_{jkm} H_{jlm}$. Solid lines correspond to non-parametric fits.

The coefficient τ_0 captures whether insurers bargain higher or lower prices than the reference price with the average provider in their network. τ_1 represents the direct effect of network breadth for service k on average costs. τ_2 captures the average degree of complementarity between pairs of services. If $\tau_2 < 0$, insurers have economies of scope across services, thus greater coverage for service $l \neq k$ makes it more attractive to provide higher coverage for service k. This measure of scope economies helps rationalize that insurers with broad networks in one service, tend to offer broad networks in other services as well (see Appendix Figure 13).

Identification. The parameters of equation (2) are identified from variation in average costs within consumer types and across insurers that are identical except for their service network breadth. My source of identification does not rely on different consumers implying different costs for similar insurers as in Tebaldi (2024) but, *conditional* on the composition of enrollee pools, for different service coverage levels to imply different costs to the insurer. In this case, variation in service network breadth across insurers is exogenous conditional on the rich set of fixed effects. However, one worry is that consumers may select into insurers based on their unobservables. One way to check this is to test whether estimates are robust to more granular definitions of consumer types. I conduct a robustness check of this style in Appendix 8.

4.4 Competition in Network Breadth

Insurers compete separately in every market choosing their service network breadths after taking expectations of demand and costs. Let $\pi_{ijm}(H_m, \theta)$ be insurer j's annual per-enrollee profit in market m, which depends on j's service network breadth and its rivals' $-j$, all collected in the vector $H_m = \{H_{jm}, H_{-jm}\}$. The annual per-enrollee profit is given by:

$$\pi_{ijm}(H_m, \theta) = (R_{\theta m} - (1 - r_\theta)AC_{\theta jm}(H_{jm}))s_{ijm}(H_m)$$

where $R_{\theta m}$ is the per-capita revenue (including ex-ante and ex-post risk-adjusted transfers plus average copayments), $AC_{\theta jm}$ is the average cost of a type-θ consumer net of patients' coinsurance payments with r_θ denoting the coinsurance rate, and s_{ijm} is consumer i's choice probability for insurer j in market m from the demand model.

I focus on a static Nash equilibrium in which insurers choose service network breadth simultaneously to maximize the net present discounted value of their profits:

$$\Pi_{jm}(H_m) = \sum_\theta \Big(\underbrace{\pi_{ijm}(H_m, \theta)N_{\theta m}}_{\text{current profit}} + \underbrace{\sum_{s=t+1}^{T} \zeta^s \sum_{\theta'} (1 - \rho_{\theta' m})\mathcal{P}(\theta'|\theta)\pi_{ijm}(H_m, \theta')N_{\theta' m}}_{\text{future profit}} \Big)$$
$$- \underbrace{\sum_k (\omega H_{jkm} + \xi_{jkm})H_{jkm}}_{\text{network administrative cost}}$$

Insurers take into account the future profits associated with each enrollee since, after making their first enrollment choice, individuals do not switch (see section 2). $N_{\theta m}$ is the fixed market size of consumers type θ, which I compute counting both new and current enrollees. In the expression for future profits, $\rho_{\theta m}$ represents the probability that a type-θ consumer drops out of the contributory system. This probability is exogenous to the choice of network breadth as it is mostly governed by the event of being unemployed. $\mathcal{P}(\theta'|\theta)$ is the transition probability from type θ in period t to type θ' in period $t + 1$. Appendix Table 10 provides evidence of the exogeneity of transition probabilities by showing that, conditional on not having a diagnosis in 2010, the likelihood of being diagnosed in 2011 is uncorrelated with service network breadth. Future profits in year t are discounted by a factor of ζ^t, which I set to 0.95 and forward simulate this profit function for $T = 100$ periods.[24]

In addition to its indirect effect on insurer profits through expected costs and demand, I assume service network breadth involves a direct network administrative cost to the insurer, that is, an administrative cost that does not vary with the type of consumers that enroll but varies with the fraction of covered providers. The network administrative cost is non-linear in service network breadth and heterogeneous across insurers and markets (as seen in Figure 5), with $\xi_{jkm} = \xi_j + \xi_m + \vartheta_{jkm}$. In this specification, ξ_j represents the insurer-specific cost component, ξ_m the market-specific cost component, and ϑ_{jkm} the idiosyncratic cost shock that is observed by insurers but unobserved to the econometrician. Appendix Figure 10 shows that insurers that offer broad networks in one market tend to do so in other markets as well, supporting my formulation of insurer-market specific network administrative costs.

Profit maximization involves a set of $|J| \times |K|$ first-order conditions (FOCs) in each market that, assuming an interior solution in service network breadth, is given by:

$$\underbrace{\sum_i \left(\frac{\partial \pi_{ijm}}{\partial H_{jkm}} N_{\theta m} + \sum_{s=t+1}^{T} \zeta^s \sum_{\theta'} (1 - \rho_{\theta' m}) \mathcal{P}(\theta' | \theta) \frac{\partial \pi'_{ijm}}{\partial H_{jkm}} N_{\theta' m} \right)}_{\text{MVP}_{jkm}} = 2\omega H_{jkm} + \xi_{jkm} \qquad (3)$$

The left-hand side of equation (3) represents the marginal variable profit MVP_{jkm} and the right-hand side is the marginal network administrative cost. The derivative of the short-run per enrollee profit, which enters the MVP_{jkm}, is:

[24] In the formulation of insurer profits, I use θ to denote sex-age-diagnosis combinations instead of sex-age group-diagnosis, for simplicity in notation.

$$\frac{\partial \pi_{ijm}}{\partial H_{jkm}} = \underbrace{\overbrace{R_{\theta m}\frac{\partial s_{ijm}}{\partial H_{jkm}}}^{\substack{\text{Selection}\\ \text{incentives}}} + \overbrace{R_{\theta m}\frac{\partial s_{ijm}}{\partial AC_{\theta jm}}\frac{\partial AC_{\theta jm}}{\partial H_{jkm}}}^{\text{Cost incentives}}}_{\text{Marginal revenue}} \tag{4}$$

$$\underbrace{- (1-r_\theta)\Big(\overbrace{AC_{\theta jm}\frac{\partial s_{ijm}}{\partial H_{jkm}} + s_{ijm}\frac{\partial AC_{\theta jm}}{\partial H_{jkm}}}^{\text{Selection incentives}} + \overbrace{AC_{\theta jm}\frac{\partial s_{ijm}}{\partial AC_{\theta jm}}\frac{\partial AC_{\theta jm}}{\partial H_{jkm}}}^{\text{Cost incentives}}\Big)}_{\text{Marginal cost}}$$

Equation (4) shows how selection and average cost incentives affect insurers' service network breadth choices. If an insurer unilaterally increases its network breadth for a particular service, marginal revenues will increase because demand from individuals with high willingness-to-pay for that service is higher (selection effect). Marginal costs also increase because patients with high willingness-to-pay for the service are the most expensive in that service, and because changes in service network breadth increase the cost of the marginal consumer (selection effect). Average cost incentives have opposite effects on marginal revenues and costs. Increasing network breadth for a particular service raises consumers' out-of-pocket costs and puts a downward pressure on demand and marginal revenues. An increase in service network breadth also reduces marginal costs because if relatively sicker consumers disenroll due to higher out-of-pocket payments, then the marginal consumer becomes cheaper.

Identification. Rewriting the FOC as

$$\text{MVP}_{jkm}(H_{jkm}) = 2\omega H_{jmk} + \xi_j + \xi_m + \vartheta_{jkm}, \quad \forall\ H_{jkm} \in (0,1) \tag{5}$$

makes explicit the endogeneity problem between H_{jkm} and the network administrative cost shocks, ϑ_{jkm}. Insurers observe ϑ_{jkm} before or at the same time as they decide on their service network breadths. For instance, if an insurer hires a highly trained manager to bargain with providers or if an insurance company is vertically integrated with its network, then $E[\vartheta_{jkm}|H_{jkm}] < 0$.[25] Thus, to identify ω I require an instrument. Specifically, I use the average service claim probabilities among healthy consumers. Intuitively, the claim

[25] Vertical integration is restricted by the Colombian government to up to 30% of an insurer's assets. Thus, endogeneity stemming from vertical integration is unlikely.

probabilities only matter for insurers' marginal variable profits through their interaction with service network breadth as specified in the model. Exogenous changes in these claim probabilities will shift the variable profits for each service, helping trace out the network administrative cost curve.

Equation (5) also suggests that if the rich demand and average cost functions (in terms of consumer and insurer heterogeneity) predict that an insurer has a positive MVP, then the only way to rationalize that this insurer chooses relatively low network breadth is that it also faces high marginal network administrative costs. This intuition relies on the demand and average cost functions being well-specified, hence I will provide evidence of the fit of these models in the next section. I estimate the parameters of the network administrative cost from the FOC via 2SLS to incorporate the instrument. This is possible because in my estimation sample, only 1.8% of observations correspond to corner solutions in H_{jkm}.

5 Estimation Results

5.1 Insurer Demand

The insurer demand model is a conditional logit estimated by maximum likelihood. To reduce the computational burden, I estimate equation (1) on a random sample of 500,000 new enrollees. Results in Table 3 show that insurer demand is decreasing in out-of-pocket costs and increasing in service network breadth. A 10 thousand pesos increase in out-of-pocket costs ($6) reduces the choice probability by 16%, corresponding to an average elasticity of -0.16.[26] Moreover, a 10 percentage point increase in network breadth across all services, increases the choice probability by 34%.[27] These results suggest not only that there is sorting based on service network breadth but also that consumers prefer broad service networks overall.

Interactions between consumer and insurer characteristics matter for enrollment decisions. Patients aged 65 or older have stronger preferences for network breadth and are more

[26] The elasticity with respect to out-of-pocket costs is $\frac{\partial s_{ijm}}{\partial c_{\theta jm}} \frac{c_{\theta jm}}{s_{ijm}}$, which I average across consumers and insurers.

[27] Calculated as $\beta_{ij} \sum_k q_{\theta k}$ and averaged across consumers and insurers.

sensitive to out-of-pocket costs compared to younger patients. One explanation for this latter result is that old individuals need more expensive care. Individuals with renal disease have stronger preferences for broader networks than those without diagnoses. Consumers with chronic conditions are also significantly less responsive to out-of-pocket costs. Appendix 6.3 presents some measures of in-sample model fit.

Table 3: Insurer demand

Variable	Network breadth		OOP spending (million)	
	coef	se	coef	se
Mean	3.429	(0.021)	-1.602	(0.117)
<u>Interactions</u>				
Male	0.543	(0.010)	0.121	(0.066)
Cancer	-0.601	(0.013)	0.003	(0.092)
Cardiovascular	-0.901	(0.011)	-0.205	(0.075)
Diabetes	-0.464	(0.023)	0.008	(0.105)
Other disease	-0.783	(0.015)	0.465	(0.068)
Pulmonary	-0.610	(0.031)	0.841	(0.095)
Renal	0.039	(0.037)	0.873	(0.069)
Age 19-24	0.055	(0.020)	0.566	(0.158)
Age 25-29	-0.575	(0.019)	0.326	(0.120)
Age 30-34	-0.560	(0.019)	0.338	(0.130)
Age 35-39	-0.456	(0.020)	-0.312	(0.215)
Age 40-44	-0.356	(0.020)	0.442	(0.161)
Age 45-49	-0.384	(0.019)	0.237	(0.141)
Age 50-54	-0.324	(0.020)	0.314	(0.133)
Age 55-59	-0.246	(0.021)	0.477	(0.122)
Age 60-64	-0.147	(0.023)	0.160	(0.121)
N	5200890			
Pseudo-R^2	0.112			

Note: Table presents a conditional logit model of insurer choice estimated by maximum likelihood on a random sample of 500,000 new enrollees. An observation is a combination of individual and insurer. Specification includes insurer fixed effects. Robust standard errors in parenthesis.

With my estimates of the preference for service network breadth and out-of-pocket costs, I calculate patient willingness-to-pay (wtp) for an additional percentage point of network breadth in each service as $\frac{1}{-\alpha_i}\frac{\partial s_{ijm}}{\partial H_{jkm}}$. Differences in wtp across consumer types will be suggestive of patient sorting based on service network breadth. Table 4 presents the average wtp for some services among patients with chronic diseases, normalizing individuals without diagnoses (labelled "healthy") to 1. Patients with chronic conditions have a significantly higher wtp for network breadth across all services compared to healthy individuals. For example, patients with renal disease are willing to pay 14 times more than a healthy individual for an additional provider in the network for renal care services.[28] In Appendix Table 13 I also report *wtp* for service network breadth by age groups, finding that older adults have higher willingness-to-pay for complex care. This variation in *wtp* implies that, in principle, insurers can avoid unprofitable patients by offering narrow networks in the services they require.

Table 4: Average willingness-to-pay per service and diagnosis

Diagnosis	Cardiac care	Renal care	Imaging	General medicine	Laboratory	Hospital admissions
Cancer	2.501	2.501	1.458	0.774	1.197	2.317
Diabetes	2.254	2.254	1.421	0.800	1.192	2.115
Cardio	1.718	1.718	1.093	0.621	0.920	1.614
Pulmonary	7.405	7.406	4.226	2.222	3.457	6.828
Renal	13.848	13.850	6.972	3.412	5.545	12.463
Other disease	3.698	3.698	2.199	1.189	1.818	3.435
Healthy	1.000	1.000	1.000	1.000	1.000	1.000

Note: Table presents the average willingness-to-pay for a percentage point increase in network breadth for the service in the column. Willingness-to-pay is calculated as $\frac{1}{-\alpha_i}\frac{\partial s_{ijm}}{\partial H_{jkm}}$, averaged across all consumers with the diagnosis in the row, and normalized to 1 for healthy individuals.

Alternative specifications. I estimate several alternative demand specifications to provide encouraging evidence of my identification arguments and modelling choices. Appendix Table 14 presents a demand function that includes an indicator of star hospital coverage, showing that the coefficient on this indicator is economically small and that including this variable has no impact on the coefficients on service network breadth and out-of-pocket costs.[29] Thus, variation in provider quality does not bias my main estimates.

Because recovering new enrollees' diagnoses from claims submitted throughout the year can create mechanical bias (since broader networks tend to cover individuals in worse health and the likelihood of receiving a diagnosis might be higher the broader is the network), in Appendix Table 15 I identify diagnoses using only the information from claims filed in January 2011. My main estimates are robust to this alternative definition, with a slight increase in the magnitude of the coefficient associated with the out-of-pocket cost.

[28]The measure of willingness-to-pay can also be interpreted in terms of travel times to the nearest provider as seen in Appendix Figure 11. For example, the estimates imply that patients with renal disease are willing to pay 14 times more than a healthy individual for a reduction of approximately 10 minutes in travel time per visit to the nearest provider that offers renal care services.

[29]Out of the roughly 11,000 providers in the country (hospitals, clinics, physician practices), around 1,800 are hospitals. Of these hospitals, 47 can be considered top-tier medical centers (see Caicedo, 2023). I use this list of hospitals to construct my measure of star hospital coverage.

Appendix Table 16 presents a demand model where service network breadth is weighted by each provider's number of beds. Although this specification is not supported by the micro-foundation, results shows that my estimates are robust to accounting for provider size. Appendix Table 17 shows a version of the demand model in which consumers are uncertain about their diagnoses. The coefficients in this table are larger in magnitude than in my main specification (but produce the same qualitative results) for two reasons: first, the specification does not include interactions of service network breadth and out-of-pocket costs with diagnoses; second, out-of-pocket costs are weighted by the diagnosis probabilities and summed across diagnoses for every individual, resulting in smaller magnitudes of the out-of-pocket cost relative to the main specification.

5.2 Insurer Average Costs Per Enrollee

I estimate equation (2) in the sample of new and current enrollees, conditional on observed choices in 2010 and 2011. Table 5 shows the results; each column explores the impact of incorporating different sets of fixed effects. Column (4) is my preferred specification and Appendix Figure 14 presents the associated consumer type fixed effects with their corresponding 95% confidence intervals.

I find that average costs are increasing in service network breadth and decreasing in the interaction between network breadth for different pairs of services. Hence, insurer coverage decisions are characterized by economies of scope: a 1% increase in network breadth for service k reduces the average cost of providing service $l \neq k$ by 0.5% per enrollee.[30] Moreover, the estimate for τ_1 indicates that a 1% increase in service network breadth raises average costs by 4.4% per enrollee.[31]

One reason why insurers enjoy economies of scope is that they may receive price discounts at provider h when they cover it for several services. Put differently, when providers can deliver several services, they increase competition for those services, lowering the providers' disagreement payoffs. For example, if provider $-h$ is dropped from the network of laboratory testing, then demand for other diagnostic services like imaging will increase at $-h$ but will

[30] Calculated as the average of $100 \times \frac{1}{2K_m} \hat{\tau}_2 \sum_{l \neq k} q_{\theta k} q_{\theta l} H_{jlm}$

[31] Calculated as the average of $100 \times \hat{\tau}_1 q_{\theta k}$

decrease at provider h. This implies that the equilibrium price that h can charge to the insurer for imaging is lower than it would be without the interaction with laboratory testing. Appendix Table 18 provides evidence of this mechanism by showing that negotiated prices for service k between insurer j and provider h are negatively correlated both with the fraction of other services $l \neq k$ for which the provider is in network and with average network breadth across all other services $l \neq k$.

Table 5: Insurer average costs per enrollee

	Log average cost per enrollee							
	(1)		(2)		(3)		(4)	
Variable	coef	se	coef	se	coef	se	coef	se
Service network breadth	0.357	(0.038)	0.158	(0.034)	0.132	(0.033)	0.274	(0.047)
Scope economies	-2.827	(0.545)	0.001	(0.472)	0.621	(0.470)	-1.100	(0.580)
Reference price	9.244	(0.119)	3.684	(0.509)	3.656	(0.519)	3.372	(0.544)
<u>Insurer FE</u>								
Insurer A							0.156	(0.038)
Insurer B							-0.084	(0.022)
Insurer C							0.020	(0.025)
Insurer D							-0.137	(0.027)
Insurer E							0.279	(0.047)
Insurer F							-0.012	(0.062)
Insurer G							0.062	(0.033)
Insurer H							0.043	(0.040)
Insurer I							-0.002	(0.017)
Insurer J							0.099	(0.020)
Insurer K							-0.135	(0.033)
Insurer L							0.116	(0.035)
Insurer M							-0.059	(0.032)
Constant	-2.108	(0.028)	-1.490	(0.086)	-1.306	(0.088)	-1.368	(0.097)
Consumer type FE	No		Yes		Yes		Yes	
Market FE	No		No		Yes		Yes	
N	18369		18369		18369		18369	
R^2	0.339		0.592		0.605		0.611	

Note: Table presents OLS regressions of the log of average costs per consumer type on service network breadth, the measure economies of scope, and the service reference price. An observation is a combination of insurer, consumer type, market and year. Estimation uses data from all continuously enrolled individuals in 2010 and 2011. Column (1) does not include fixed effects, column (2) includes consumer type fixed effects, column (3) includes consumer type and market fixed effects, and column (4) includes consumer type, market, and insurer fixed effects. Robust standard errors in parenthesis.

Comparing the R^2 across the different specifications, Table 5 also indicates that service network breadth and reference prices alone explain nearly 34% of the variation in insurers' average costs per enrollee (column 1). The heterogeneity across insurers captured by the insurer fixed effects explains only an additional 6 percentage points of the variation after controlling for consumer types and markets (column 4), suggesting this heterogeneity is unlikely to generate asymmetric network breadth choices among insurers. My average cost model accurately fits the data moments that identify the coefficients on service network breadth and scope economies as seen in Appendix Figure 15.

Alternative specifications. I conduct several robustness checks on the average cost function to provide evidence of my identification approach and choice of functional form. Appendix Table 19 shows that my model is robust to more granular definitions of consumer type, providing support for the limited relevance of unobserved cost heterogeneity within consumer types. Appendix Table 20 presents results including a star hospital indicator, constructed in the same way as for the robustness check on the demand model. I find that the coefficient on this indicator is relatively small and that my main estimates remain unchanged. Together with results on demand, this exercise supports my argument that variation in provider quality conditional on the service is negligible and does not bias my main estimates. Finally, Appendix Table 21 shows a version of the average cost model in which network breadth is weighted by each provider's number of beds, showing that my estimates are robust to accounting for provider size.

5.3 Competition in Network Breadth

The third piece of the insurers' profit function to estimate is the network administrative cost. To do this, I use insurers' FOCs. Demand and average cost estimates allow me to compute MVPs in the left-hand side of equation (3). Dropout and transition probabilities are calculated outside of the model non-parametrically. Appendix 9 presents summary statistics of these probabilities and MVPs. The fact that MVPs are positive for all insurer-market-services and that demand and average costs have a strong fit, suggests that network administrative costs play an important role in the profit maximizing choices of service network breadth.

Table 6 presents the results of equation (5) for the log of marginal variable profits. Each column includes a different set of fixed effects and uses as instrument for service network breadth the average claim probability among healthy consumers. Column (3) presents my preferred specification and Appendix Table 25 presents the associated first-stage results. I find that while $\widehat{\omega}$ is positive and significant, network administrative costs are decreasing in service network breadth because of insurers' cost shocks $\widehat{\xi}_{jkm}$, that is, insurers that offer more generous coverage are also more efficient in their billing and auditing activities. The first-stage F-statistic suggests that my instrument is relevant; without the instrument, the coefficient on service network breadth would be one order of magnitude smaller, consistent with the direction of the bias discussed in section 4.4.

Table 6: Model of insurer network administrative costs

	Log marginal variable profit					
	(1)		(2)		(3)	
Variable	coef	se	coef	se	coef	se
Service network breadth	29.17	(2.379)	29.17	(2.253)	29.19	(1.903)
Insurer FE						
Insurer A					7.762	(0.744)
Insurer B					3.067	(0.464)
Insurer C					6.419	(0.643)
Insurer D					5.391	(0.551)
Insurer E					6.955	(0.932)
Insurer F					8.307	(0.989)
Insurer G					9.376	(0.772)
Insurer H					1.441	(1.105)
Insurer I					2.107	(0.423)
Insurer J					4.741	(0.523)
Insurer K					5.479	(0.616)
Insurer L					-1.694	(0.604)
Insurer M					8.122	(0.833)
Constant	-2.029	(0.791)	-3.264	(0.861)	-8.799	(1.130)
Market FE	No		Yes		Yes	
First-stage F-stat	160.69		167.93		249.47	
N	2280		2280		2280	
Unadjusted R^2	0.484		0.537		0.670	

Note: Table presents 2SLS regression of the log of marginal variable profit on service network breadth. An observation is a combination of insurer, service, and market. The instrument for service network breadth is the average claim probability for each service among healthy consumers. Table reports the F-statistic for the first stage regression. Column (1) does not include fixed effects, column (2) includes market fixed effects, and column (3) includes market and insurer fixed effects. Robust standard errors in parenthesis.

A comparison of the unadjusted R^2 across the different specifications, indicates that service network breadth explains 48% of the variation in insurers' marginal variable profits (column 1). The heterogeneity across insurers in their network administrative costs is also substantial, as including insurer fixed effects (column 3) increases the R^2 by 13 percentage points relative to including only market fixed effects (column 2).

Figure 8 illustrates the implications of insurer heterogeneity in average and network administrative costs for equilibrium service network breadth. The figure presents the model-predicted log average cost per enrollee (in black) and the model-predicted log administrative cost per enrollee (in red) against percentiles of average service network breadth. Without network administrative costs, risk selection incentives would unambiguously generate an equilibrium in which all insurers choose narrow service networks, because average costs per enrollee increase monotonically in network breadth. Instead, the presence of network administrative costs that decrease in the fraction of covered providers raises the possibility that equilibrium service networks are asymmetric within a service and market.

The relation depicted in Figure 8 is consistent with the data patterns reported in Figure 6, even though information from insurers' public income statements is not used in estimation. Appendix Figure 16 shows that my model of network administrative costs fits *untargeted* moments coming from these public income statements, such as the ratio of total costs (average costs plus network administrative costs) to total revenues.

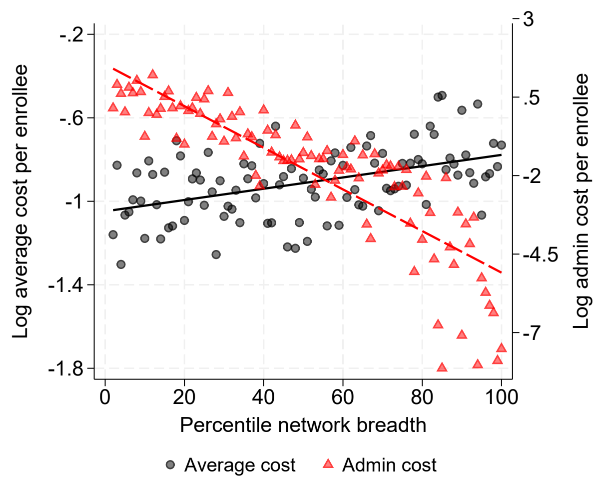


Figure 8: Model-predicted relation between average costs, administrative costs, and networks

Note: Figure presents a scatter plot of the log of model-predicted average cost per enrollee (for every insurer and market) within each percentile of average network breadth (averaged across services) in black. Log average costs are depicted on the left vertical axis. The solid black line corresponds to a linear fit. On the right vertical axis and in the red markers, the figure presents a scatter plot of the log of model-predicted network administrative cost per enrollee within each percentile of average network breadth (averaged across services). The dashed red line corresponds to a linear fit.

Alternative specifications. In Appendix Table 26 I test the robustness of my network administrative cost model to accounting for provider size (measured by the number of beds) in service network breadth. Results in the appendix show that my coefficients are robust to this alternative specification.

6 Risk Selection vs. Cost Structure

In this section, I use my model estimates to assess the magnitude of risk selection and insurers' cost structure for determining service network breadth. This exercise is important to inform the types of policies that can increase network coverage for patients who need it the most. For example, if risk selection is the main driver of service network breadth decisions, then implementing network adequacy rules that mandate the inclusion of specific providers could be counterproductive, as insurers might simply avoid enrolling patients who require those providers. In this case, a policy like risk adjustment would be more effective at encouraging insurers to create broader networks. On the other hand, if insurers' cost incentives, such as economies of scale or scope, predominantly drive their network breadth decisions, then requiring insurers to include certain providers—as proposed in U.S. network adequacy regulations—could effectively enhance provider coverage.

To evaluate the importance of risk selection, I conduct two counterfactual simulations: first, I eliminate risk adjustment by imposing the same transfer across all consumer types. Second, I improve the government's formula by compensating for a list of 14 health conditions listed in Appendix Table 27. In these analyses, I hold long-run government spending fixed across all markets, so that changes in service network breadth are determined only by changes in how resources are redistributed across insurers, but not by the level of the transfer itself.

Then, to quantify the relevance of cost incentives, I conduct two additional counterfactuals: I eliminate average cost and network administrative cost heterogeneity by imposing the median fixed effect, $\delta_j = \overline{\delta}$ and $\xi_j = \overline{\xi}\ \forall j$, respectively. While these exercises are not achievable through policy mechanisms since they require that the regulator knows insurers' cost structures, they provide evidence of the trade-off between risk selection and cost incentives, which informs whether policies like network adequacy standards are effective. Note also that these counterfactuals will speak to the role of heterogeneity in insurers' network administrative costs (differences in productivity) rather than to the existence of these costs at all.

One concern in the counterfactual analyses is that the model may admit multiple equilibria. For instance, my measure of scope economies can make it such that every firm choosing complete networks or no coverage at all are both feasible equilibria. While a direct proof of uniqueness is challenging, in Appendix 10 I provide suggestive evidence of uniqueness by computing the second partial derivative of the insurers' profit function with respect to service network breadth, all else equal. The rich preference and cost heterogeneity prevent multiple equilibria from arising. In computing the counterfactual analyses, I also use several different starting values for the vector of service network breadth to confirm that they all converge to the same equilibrium. For tractability, I conduct all my counterfactuals with data from the largest market, Bogotá, where 29% of all continuously enrolled individuals reside and where all private insurers compete.

Table 7 presents the results. Panel A shows the percentage change in average network breadth, average costs per enrollee, total average costs, and long-run consumer surplus for individuals with and without diagnoses, relative to the observed scenario. Panel B shows the percentage change in network breadth for specific services. Columns (1) and (2) show results under no risk adjustment and with improved risk adjustment, respectively. And columns (3) and (4) show results imposing homogeneity in average costs and network administrative costs, respectively.

I find that without risk adjustment, average network breadth falls 6.6%, which translates into excluding at most 5 providers from the service network. The reduction in service network breadth generates relevant welfare effects: long-run consumer surplus falls 1.3% for individuals with diagnoses and 1.0% for those without diagnoses. Consumer surplus for individuals with chronic conditions falls by a greater magnitude because networks for the relatively expensive services that they tend to claim become much narrower as seen in Panel B. For instance, average network breadth for hospital admissions decreases approximately 6.0% relative to the observed scenario, while average network breadth for general medicine is essentially unchanged.

Column (2) presents qualitatively opposite results. With improved risk adjustment, average network breadth increases 8.0%, which corresponds roughly to adding at most 7 providers to the network. Effects are larger for services that mostly sick patients claim, consistent with weakened selection incentives and with risk selection being a determinant of narrow networks. For example, average network breadth for cardiac care increases 8.6%, while average network breadth for general medicine increases only 2.5%. In this case, I find that long-run consumer surplus increases for both types of consumers relative to the observed scenario and is slightly higher for those with diagnoses.

Table 7: Networks, costs, and welfare after changes in risk adjustment and cost structures

Variable	Risk adjustment: No RA (1)	Risk adjustment: RA (2)	Cost structure: Avg cost (3)	Cost structure: Admin cost (4)
Panel A. Overall				
Average network breadth	-6.593	8.012	0.124	-7.257
Average cost per enrollee	-0.458	0.276	-2.590	-1.308
Total average cost	-0.527	0.284	-0.024	2.509
Consumer surplus (with diagnoses)	-1.267	0.826	-0.083	0.557
Consumer surplus (without diagnoses)	-1.002	0.667	-0.004	0.569
Panel B. Service network breadth				
Otorhinolaryngologic care	-5.503	8.236	0.129	-8.692
Cardiac care	-8.185	8.579	0.119	-8.133
Gastroenterologic care	-7.317	8.488	0.120	-8.280
Renal care	-8.666	9.081	0.126	-8.608
Gynecologic care	-7.506	8.747	0.124	-7.675
Orthopedic care	-8.109	8.643	0.120	-8.230
Imaging	-1.909	4.929	0.100	-4.986
General medicine	0.242	2.473	0.105	-5.676
Laboratory	-1.033	4.138	0.102	-5.439
Hospital admission	-6.021	7.335	0.105	-6.361

Note: Panel A presents the percentage change relative to the observed scenario in average network breadth, average cost per enrollee, total average cost, and long-run consumer surplus for sick and healthy individuals, in the scenario without risk adjustment in column (1), the scenario with improved risk adjustment in column (2), the scenario with homogeneous average costs in column (3), and the scenario with homogeneous network formation costs in column (4). Panel B presents the percentage change relative to the observed scenario in average network breadth for a few service categories. Simulations use data from Bogotá.

Although the results in column (2) are largely symmetric to those reported in column (1), this symmetry is not necessarily expected. My counterfactual results indicate that service network breadth is increasing in the predictive power of the risk adjustment formula. In column (2) I have designed a risk adjustment system where compensating for the 14 conditions listed in Appendix Table 27 results in networks that are twice as broad as those in column (1). However, reducing the predictive power of the risk adjustment formula would result in smaller impacts on service network breadth.

In terms of insurers' cost structures, my analysis reveals that, absent heterogeneity in network administrative costs, service network breadth would significantly decrease, as seen in column (4), while average cost heterogeneity has minimal influence on network breadth, as shown in column (3). These results go in line with my model estimates indicating that insurer fixed effects help explain a small variation in average costs but a large variation in network administrative costs.

Focusing on column (4), Panel A shows that if insurers had homogeneous network administrative costs, average service network breadth would decrease 7.3% relative to the observed scenario. Insurers' total average cost would increase 2.5% because they can no longer take advantage of scope economies. Long-run consumer surplus for individuals with and without diagnoses would increase by a moderate amount, suggesting welfare gains from lower out-of-pocket costs slightly overcompensate welfare losses from reduced network coverage.

Panel B shows that the reduction in network breadth is larger for services that mostly sick individuals tend to claim, hence economies of scale push service network breadth in the opposite direction relative to risk selection. For instance, network breadth for general medicine decreases approximately 5.7%, while network breadth for renal care falls around 8.6% relative to the observed scenario. In Appendix Table 28 I also show that results are qualitatively similar when I impose homogeneity across insurers by assigning the mean rather than the median insurer fixed effect.

As mentioned in section 4.1, my model permits simulating counterfactual policies that do not impact insurers' or providers' disagreement payoffs. Modifying the risk adjustment formula and imposing homogeneity in insurers' cost structures likely does not affect disagreement payoffs differentially across insurer-provider pairs relative to the observed equilibrium. Appendix Figure 18 presents some evidence of this by showing that the functional form of my average and network formation cost models remain stable in these counterfactuals.

Taken together, results indicate that risk selection and network administrative costs weigh equally in the decision to offer service network breadth. Risk selection drives the choice of narrow networks, while network administrative costs drive the choice of broad networks. Thus, policies like risk adjustment can improve coverage for individuals with and without chronic conditions beyond what the market can produce given the distribution of insurer productivity. Network adequacy rules either forcing insurers to cover specific providers or establishing minimum provider-to-enrollee ratios are also desirable in this setting due to the substantial economies of scale that insurers enjoy in their network administrative costs.

6.1 Discussion

The Colombian health care system is characterized by imperfect competition between insurers that design their provider networks to attract new consumers. In all my counterfactual simulations, I hold this competitive landscape fixed. However, understanding the implications of imperfect competition for equilibrium service network breadth is also crucial for external validity. Veiga and Weyl (2016) and Azevedo and Gottlieb (2017) show that strong competition in markets with adverse selection can lead to lower quality because a firm can always enter the market with a cheap, low-quality product, stealing all the profitable consumer types who are relatively more price-sensitive.[32]

In my setting, this pricing mechanism does not exist because premiums are regulated and equal to zero. A monopolist insurer does not internalize any of the social value of providing a broad network because it cannot set premiums. Thus, the monopolist and the social planner would opt for different network breadths. Imperfect competition in my setting reinforces risk selection incentives, pushing equilibrium outcomes towards narrow networks. Hence, the fact that I observe heterogeneity in service network breadth among insurers despite risk selection and imperfect competition suggests there are other factors that determine the decision to offer network breadth. This paper argues that insurers' network administrative costs can help explain both the asymmetric equilibrium in network breadth and why some insurers choose broad networks in imperfectly competitive markets without premiums.

7 Conclusions

Private health insurers respond to different incentives when crafting the various elements of their insurance contracts. This paper shows that risk selection and insurers' cost structure are the main drivers of the decision to offer provider network breadth. Risk selection induces

[32] Mahoney and Weyl (2017) and Cuesta and Sepúlveda (2021) provide some empirical applications where competition can be harmful for consumers in the subprime auto lending market and the market for consumer credit, respectively.

insurers to offer narrow networks, while network administrative cost heterogeneity induces insurers to offer broad networks despite selection incentives. I use a structural model of insurer competition in service network breadth to decompose the relative importance of these factors in counterfactuals. The empirical setting is Colombia, where the government regulates premiums and cost-sharing (similar to Medicaid Managed Care and the Exchanges in the U.S.), and allows insurers to choose only which and how many providers to cover for each health service.

To quantify the equilibrium impact of risk selection on service network breadth, I modify the risk adjustment formula. Without risk adjustment, average service network breadth would decrease 6.6%, consistent with increased selection incentives. Instead, improving the risk adjustment formula by compensating for a granular list of diagnoses would increase average service network breadth by 8.0%. To quantify the equilibrium impact of insurers' cost structure, I force network administrative costs to be homogeneous across insurers. Results show that average service network breadth would decrease 7.3%, with reductions being larger in services that sick individuals require the most.

The findings of this paper provide new evidence of selection on provider networks and speak to the increasing use of network adequacy rules in markets where narrow-network plans have proliferated. A direct implication of my results is that when risk selection weighs more than insurers' cost structure in the decision to offer provider networks, there is little scope for the use of network adequacy rules.

References

Aizawa, N. and Kim, Y. (2018). Advertising and Risk Selection in Health Insurance Markets. *American Economic Review*, 108(3):828–867.

Azevedo, E. M. and Gottlieb, D. (2017). Perfect Competition in Markets with Adverse Selection. *Econometrica*, 85(1):67–105.

Bauman, N., Coe, E., Ogden, J., and Parikh, A. (2014). Hospital Networks: Updated National View of Configurations on the Exchanges. *McKinsey Report*. Accessed through https://www.mckinsey.com/industries/healthcare/our-insights/hospital-networks-updated-national-view-of-configurations-on-the-exchanges/.

Brown, J., Duggan, M., Kuziemko, I., and Woolston, W. (2014). How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program. *American Economic Review*, 104(10):3335–3364.

Cabral, M., Geruso, M., and Mahoney, N. (2018). Do Larger Health Insurance Subsidies Benefit Patients or Producers? Evidence from Medicare Advantage. *American Economic Review*, 108(8):2048–2087.

Caicedo, E. (2023). En Colombia están 47 de los 2300 mejores hospitales y clínicas del mundo. *El Tiempo*. Accessed through https://www.eltiempo.com/salud/en-colombia-estan-47-de-los-2-300-mejores-hospitales-y-clinicas-del-mundo-784388.

Carey, C. (2017). Technological Change and Risk Adjustment: Benefit Design Incentives in Medicare Part D. *American Economic Journal: Economic Policy*, 9(1).

Chernew, M. and Mintz, H. (2021). Administrative Expenses in the US Health Care System: Why So High? *JAMA*, 326(17):1679–1680.

Chiappori, P. and Salanie, B. (2000). Testing for Asymmetric Information in Insurance Markets. *Journal of Political Economy*, 108(1):56–78.

Crawford, G. S. and Yurukoglu, A. (2012). The Welfare Effects of Bundling in Multichannel Television Markets. *American Economic Review*, 102(2):643–685.

Cuesta, J. I. and Sepúlveda, A. (2021). Price Regulation in Credit Markets: A Trade-off Between Consumer Protection and Credit Access. *Available at SSRN 3282910*.

Cutler, D. (2020). Reducing administrative costs in US health care. *The Hamilton Project Policy Proposal*, 9:3–25.

Dafny, L., Hendel, I., and Wilson, N. (2015). Narrow Networks on the Health Insurance Exchanges: What do they Look Like and How do they Affect Pricing? A Case Study of Texas. *American Economic Review: Papers & Proceedings*, 105(5):110–14.

Dafny, L. S., Hendel, I., Marone, V., and Ody, C. (2017). Narrow networks on the health insurance marketplaces: Prevalence, pricing, and the cost of network breadth. *Health Affairs*, 36(9):1606–1614.

Decarolis, F. and Guglielmo, A. (2017). Insurers' response to selection risk: Evidence from Medicare enrollment reforms. *Journal of Health Economics*, 56(1):383–396.

Einav, L., Finkelstein, A., Ryan, P., Schrimpf, D., and Cullen, M. (2013). Selection on Moral Hazard in Health Insurance. *American Economic Review*, 103(1):178–219.

Ericson, K. M. and Starc, A. (2015). Measuring Consumer Valuation of Limited Provider Networks. *American Economic Review*, 105(5):115–19.

Fleitas, S., Gowrisankaran, G., Starc, A., and Swanson, A. (2024). The Power of Exclusion: Pharmacy Networks and Bargaining in Medicare Part D. Working paper.

Geruso, M., Layton, T., and Prinz, D. (2019). Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges. *American Economic Journal: Applied Economics*, 11(2):64–107.

Ghili, S. (2022). Network Formation and Bargaining in Vertical Markets: The Case of Narrow Networks in Health Insurance. *Marketing Science*, 41(3):501–527.

Gowrisankaran, G., Nevo, A., and Town, R. (2015). Mergers When Prices Are Negotiated: Evidence from the Hospital Industry. *American Economic Review*, 105(1):172–203.

Grennan, M. (2013). Price Discrimination and Bargaining: Empirical Evidence from Medical Devices. *American Economic Review*, 103(1):145–77.

Gruber, J. and McKnight, R. (2016). Controlling Health Care Costs Through Limited Network Insurance Plans: Evidence from Massachusetts State Employees. *American Economic Journal: Economic Policy*, 8(2):219–250.

Ho, K. (2006). The Welfare Effects of Restricted Hospital Choice in the US Medical Care Market. *Journal of Applied Econometrics*, 21(7):1039–1079.

Ho, K. and Lee, R. (2017). Insurer Competition in Health Care Markets. *Econometrica*, 85(2):379–417.

Ho, K. and Lee, R. (2019). Equilibrium Provider Networks: Bargaining and Exclusion in Health Care Markets. *American Economic Review*, 109(2):473–522.

Ho, K. and Lee, R. S. (2016). Narrow Medical Provider Networks: Welfare Implications and Approaches to Market Design. *More Fair by Design: Economic Design Responses to Inequality*.

Horn, H. and Wolinsky, A. (1988). Bilateral Monopolies and Incentives for Merger. *The RAND Journal of Economics*, pages 408–419.

Lee, R. and Fong, K. (2013). Markov-Perfect Network Formation: An Applied Framework for Bilateral Oligopoly and Bargaining in Buyer-Seller Networks. Working paper.

Liebman, E. (2022). Bargaining in Markets with Exclusion: An Analysis of Health Insurance Networks. Working paper.

Mahoney, N. and Weyl, G. (2017). Imperfect Competition in Selection Markets. *The Review of Economics and Statistics*, 99(4):637–651.

McFadden, D. (1996). Computing Willingness-to-Pay in Random Utility Models. *University of California at Berkeley, Econometrics Laboratory Software Archive, Working Papers*.

McGuire, T., Glazer, J., Newhouse, J., Normand, S., Shi, J., Sinaiko, A., and Zuvekas, S. (2013). Integrating risk adjustment and enrollee premiums in health plan payment. *Journal of Health Economics*, 32(6):1263–1277.

McWilliams, J., Hsu, J., and Newhouse, J. (2012). New Risk-Adjustment System Was Associated With Reduced Favorable Selection In Medicare Advantage. *Health Affairs*, 31(12):2630–2640.

MinSalud (2016). Redes integrales de prestadores de servicios de salud: Lineamientos para el Proceso de Conformación, Organización, Gestión, Seguimiento y Evaluación. *Ministerio de Salud y Protección Social*.

Newhouse, J. P., Price, M., McWilliams, J. M., Hsu, J., and McGuire, T. G. (2015). How Much Favorable Selection is Left in Medicare Advantage? *American journal of health economics*, 1(1):1–26.

Nicholson, S., Bundorf, K., Stein, R., and Polsky, D. (2004). The Magnitude and Nature of Risk Selection in Employer-Sponsored Health Plans. *Health Services Research*, 39(6):1817–1838.

Pauly, M. and Herring, B. (2007). Risk Pooling and Regulation: Policy and Reality in Today's Individual Health Insurance Market. *Health Affairs*, 26(3):770–779.

Pollitz, K. (2022). Network Adequacy Standards and Enforcement. *Kaiser Family Foundation Report*. Accessed through https://www.kff.org/affordable-care-act/issue-brief/network-adequacy-standards-and-enforcement/.

Polsky, D., Cidav, Z., and Swanson, A. (2016). Marketplace plans with narrow physician networks feature lower monthly premiums than plans with larger networks. *Health Affairs*, 35(10):1842–1848.

Prager, E. and Tilipman, N. (2020). Regulating Out-of-Network Hospital Payments: Disagreement Payoffs, Negotiated Prices, and Access. Working Paper.

Riascos, A., Alfonso, E., and Romero, M. (2014). The performance of risk adjustment models in colombian competitive health insurance market. *Documentos CEDE*, (33).

Riascos, A., Romero, M., and Serna, N. (2017). Risk Adjustment Revisited using Machine Learning Techniques. *Documentos CEDE*, (27).

Ruiz, F., Amaya, L., Garavito, L., and Ramírez, J. (2008). *Precios y Contratos en Salud: Estudio Indicativo de Precios y Análisis Cualitativo de Contratos*. Ministerio de la Protección Social.

Shepard, M. (2022). Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange. *American Economic Review*, 112(2):578–615.

Starc, A. and Swanson, A. (2021). Preferred pharmacy networks and drug costs. *American Economic Journal: Economic Policy*, 13(3):406–46.

Tebaldi, P. (2024). Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA. *Forthcoming, Review of Economic Studies.*

Tilipman, N. (2022). Employer incentives and distortions in health insurance design: Implications for welfare and costs. *American Economic Review*, 112(3):998–1037.

Veiga, A. and Weyl, E. G. (2016). Product Design in Selection Markets. *The Quarterly Journal of Economics*, 131(2):1007–1056.

Online Appendix

Appendix 1 Current risk adjustment system

For year t, the base un-adjusted capitated transfer is calculated using the claims data from year $t-2$. The per-enrollee transfer is equal to the average annual healthcare cost in the population multiplied by a risk adjustment factor that is specific to a combination of sex, age group, and municipality. Appendix Table 1 shows the national un-adjusted transfer and Appendix Table 2 shows the risk adjustment multipliers.

Appendix Table 1: Base capitated transfer for the Contributory System during 2011

Department/city		Transfer
National (pesos)		525,492
Market multiplier a_m	Amazonas	$\times$ 1.10
	Arauca, Arauca	$\times$ 1.10
	Yopal, Casanare	$\times$ 1.10
	Florencia, Caquetá	$\times$ 1.10
	Chocó	$\times$ 1.10
	Riohacha, Guajira	$\times$ 1.10
	Guainía	$\times$ 1.10
	Guaviare	$\times$ 1.10
	Villavicencio, Meta	$\times$ 1.10
	Putumayo	$\times$ 1.10
	San Andrés y Providencia	$\times$ 1.10
	Sucre, Sincelejo	$\times$ 1.10
	Vaupés	$\times$ 1.10
	Vichada	$\times$ 1.10
	Soacha, Cundinamarca	$\times$ 1.06
	Bello, Antioquia	$\times$ 1.06
	Itaguí, Antioquia	$\times$ 1.06
	Envigado, Antioquia	$\times$ 1.06
	Sabaneta, Antioquia	$\times$ 1.06
	Soledad, Antioquia	$\times$ 1.06
	Bogotá	$\times$ 1.06
	Medellín, Antioquia	$\times$ 1.06
	Barranquilla, Atlántico	$\times$ 1.06

Note: Table reports national base risk-adjusted transfer which includes payments for promotion and prevention programs. Table also reports risk-adjustment multipliers for each market.

Appendix Table 2: Risk Adjustment Factors in the Contributory System during 2011

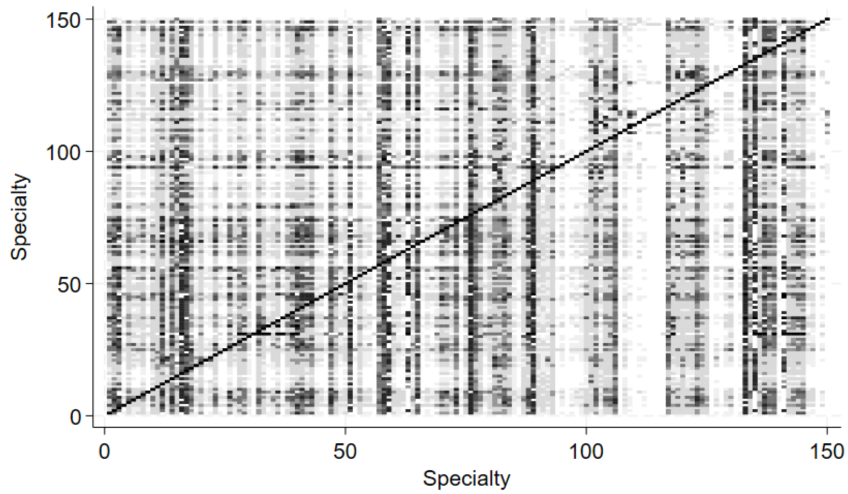
Age group	Sex	Multiplier
Less than 1	—	3.0000
1-4	—	0.9633
5-14	—	0.3365
15-18	M	0.3207
15-18	F	0.5068
19-44	M	0.5707
19-44	F	1.0588
45-49	—	1.0473
50-54	—	1.3358
55-59	—	1.6329
60-64	—	2.1015
65-69	—	2.6141
70-74	—	3.1369
More than 74	—	3.9419

Note: Table reports government risk-adjustment multipliers by sex and age group.

Appendix 2 Service categories

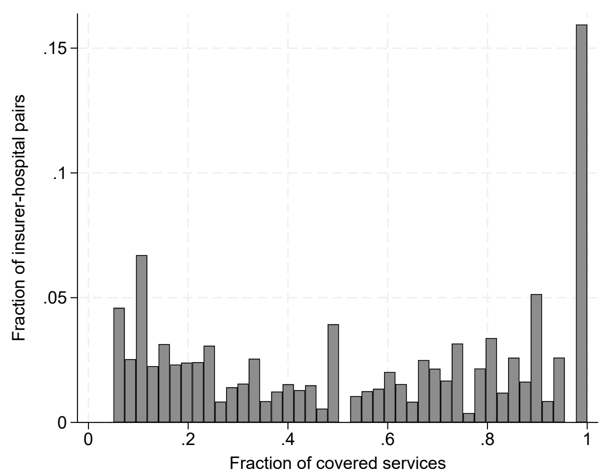
The service-level provider network data reports 150 unique specialities over which insurers and providers bargain. Some of these specialties are highly correlated in the sense that insurers tend to include them together at a particular provider. Appendix Figure 1 presents a heatmap of the fraction of insurer-provider pairs that include the specialty in the horizontal axis, and also include the specialty in the vertical axis. Dark colors represent higher fractions. The heatmap shows that (i) there are very common specialties such as general medicine and internal medicine seen in the vertical dark lines, and (ii) some specialties are correlated along the diagonal.

Appendix Figure 2 shows that most insurers cover all the services at a particular provider, but there is still substantial variation in service coverage within insurer-provider pair. I group the different specialties of the network data into a final list 20 service categories, which can be mapped to the claims data based on the 6-digit service code reported for each claim. Appendix Table 3 provides the final list of services and Appendix Table 4 provides a data excerpt of insurer's network inclusions for three hospitals and three services.



Appendix Figure 1: Heatmap of specialty pairs network inclusions

Note: Figure presents a heatmap of the fraction of insurer-hospital pairs in the network data that include the specialty in the horizontal axis and the specialty in the vertical axis. Darker colors represent higher fractions.



Appendix Figure 2: Service inclusions within hospital

Note: Figure presents the distribution of the fraction of services that the provider can deliver which are covered by the insurer.

Appendix Table 3: List of services

Service code	Description
01	**Neurosurgery:** Procedures in skull, brain, and spine
02	**Other neurologic care:** Procedures in nerves and glands
03	**Otorhinolaryngologic care:** Procedures in face and trachea
04	**Pneumologic care:** Procedures in lungs and thorax
05	**Cardiac care:** Procedures in cardiac system
06	**Angiologic care:** Procedures in lymphatic system and bone marrow
07	**Gastroenterologic care:** Procedures in digestive system
08	**Hepatologic care:** Procedures in liver, pancreas, and abdominal wall
09	**Renal care:** Procedures in urinary system
10	**Gynecologic care:** Procedures in reproductive system
11	**Orthopedic care:** Procedures in bones and joints
12	**Other orthopedic care:** Procedures in tendons, muscles, and breast
13	**Diagnostic aid:** Diagnostic procedures in skin and subcutaneous cellular tissue
14	**Imaging:** Radiology and non-radiology imaging
15	**Internal and general medicine:** Consultations
16	**Laboratory:** Laboratory and blood bank
17	**Nuclear medicine:** Nuclear medicine and radiotherapy
18	**Rehab and mental health:** Rehabilitation, mental health care, therapy
19	**Therapy (chemo and dialysis):** Prophylactic and therapeutic procedures
20	**Hospital admissions:** Inpatient services

Note: Table presents the final list of 20 services and their description.

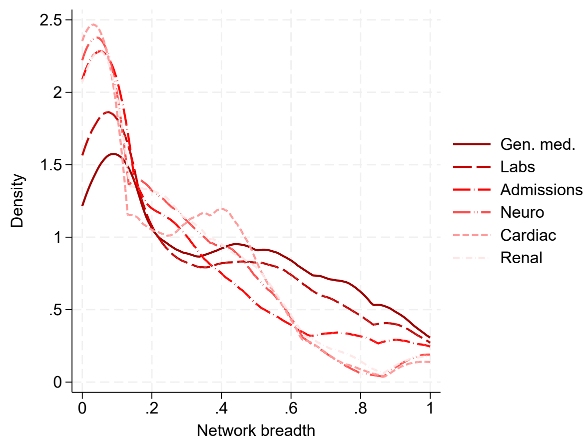
Appendix Table 4: Service coverage at hospitals

	Cardiac care			Renal care			Hospital admissions		
Insurer	Valle del Lili	Santa Fe	Pablo Tobón	Valle del Lili	Santa Fe	Pablo Tobón	Valle del Lili	Santa Fe	Pablo Tobón
Insurer A	1	0	0	1	0	0	1	1	1
Insurer B	1	0	1	1	0	1	1	1	1
Insurer C	1	0	0	0	0	0	1	1	1
Insurer D	1	1	1	1	1	1	1	1	1
Insurer E	1	1	0	1	1	0	1	1	0
Insurer F	0	0	1	0	0	1	0	0	1
Insurer G	1	1	1	1	1	1	1	1	1
Insurer H	1	1	0	1	1	0	1	1	0
Insurer I	1	0	1	1	0	1	1	1	1
Insurer J	1	1	1	1	1	1	1	1	1
Insurer K	0	1	0	0	1	0	1	1	1
Insurer L	1	1	1	1	1	1	1	1	1
Insurer M	0	0	0	0	0	0	1	1	1
Insurer N	1	1	1	1	1	1	1	1	1

Note: Table presents service coverage per insurer at three hospitals in the country and for three services. Data comes from the National Health Superintendency.

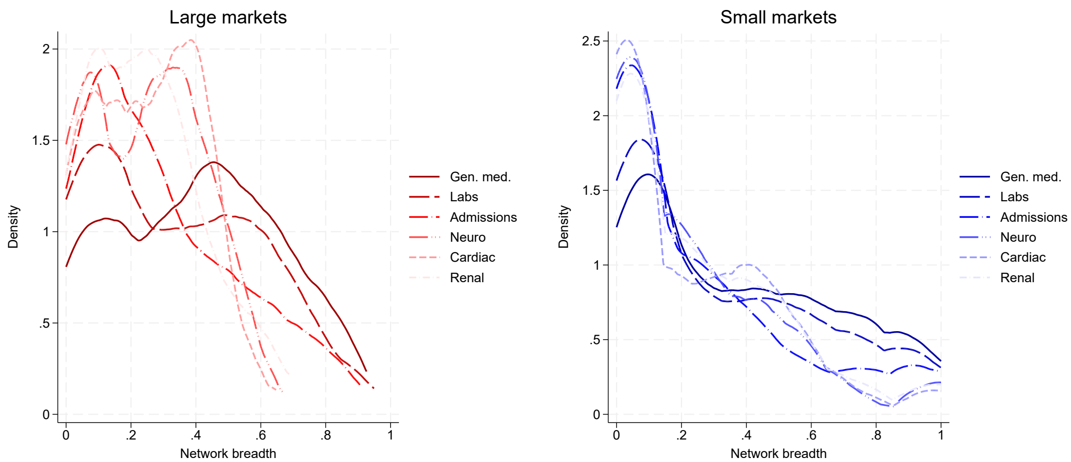
Appendix 3 Additional descriptives in full sample

This appendix presents additional descriptive evidence in the full sample of enrollees. Appendix Figure 3 shows the distribution of network breadth during 2011 for general medicine, laboratory testing, hospital admissions, neurological care, cardiac care, and renal care. Appendix Figure 4 presents these distributions conditional on the 5 largest markets and the rest of markets. These two figures illustrate the types of services that are under-covered across all markets. Appendix Table 5 shows OLS regressions of an indicator for switching out of an insurer on that insurer's service network breadth conditional on patients with different health conditions. The estimation uses the full sample of individuals and findings provide evidence of adverse selection in service network breadth.



Appendix Figure 3: Distribution of network breadth per service

Note: Figure presents kernel density estimates for the distribution of network breadth conditional on six services: general medicine, laboratory testing, hospital admissions, neurological care, cardiac care, and renal care. An observation to construct these distribution is a combination of insurer, market, and year.



Appendix Figure 4: Distribution of service network breadth per market

Note: Figure presents kernel density estimates of the distribution of network breadth for general medicine, laboratory testing, hospital admissions, neurological care, cardiac care, and renal care in the 5 largest markets in the left panel and in the rest of markets in the right panel.

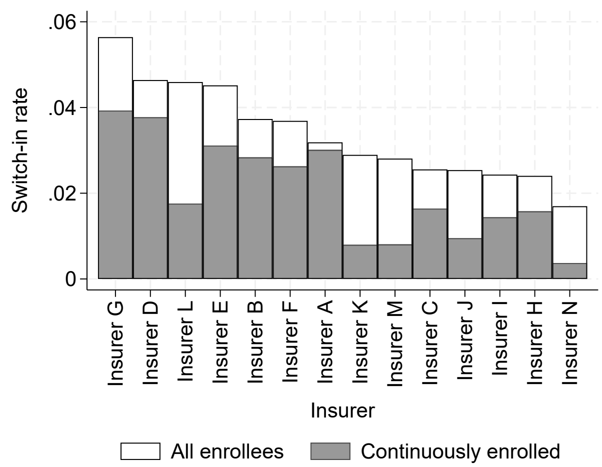
Appendix Table 5: Determinants of switching

Sample:	Healthy	Cancer	Diabetes	Cardio
Service:	General medicine	Therapy	Laboratory	Cardiac care
	(1)	(2)	(3)	(4)
Service network breadth	0.0721	-0.0484	-0.0134	-0.0249
	(0.00106)	(0.00234)	(0.00251)	(0.00188)
Demographics (male, age)	Yes	Yes	Yes	Yes
Enrollment spell length	Yes	Yes	Yes	Yes
Market FE	Yes	Yes	Yes	Yes
N	10703097	770951	346017	1718751
R^2	0.145	0.0740	0.0790	0.0808

Note: Table presents OLS regression of a switching-out indicator on service network breadth for the 2010 insurer. Estimation uses the full sample. Column (1) uses the sub-sample of individuals without diagnoses and network breadth for general medicine. Column (2) uses the sub-sample of individuals with cancer and network breadth for chemotherapy. Column (3) uses the sub-sample of individuals with diabetes and network breadth for laboratory. Column (4) uses the sub-sample of individuals with cardiovascular disease and network breadth for cardiac care services. All specifications control for enrollees' demographic characteristics, enrollment spell length, and market fixed effects. Robust standard errors in parenthesis.

Appendix 4 Descriptives in subsample

This appendix replicates the descriptive evidence in the main text on the sample for model estimation which comprises individuals who have continuous enrollment spells. Appendix Figure 5 presents the fraction of consumers who switch into each insurer in 2011 relative to 2010 among the full sample and the sample of continuously enrolled.



Appendix Figure 5: Insurer switch-in rates in the continuously enrolled

Note: Figure presents the number of enrollees that switch into each insurer in 2011 relative to 2010 divided by the total number of enrollees at each insurer in 2011. The white bars use the full sample of enrollees. The gray bars use the sample of continuously enrolled.

Appendix Table 6 presents summary statistics of the continuously enrolled, new enrollees, and the random sample of new enrollees used for demand model estimation. New enrollees are relatively healthier and have lower healthcare spending than the continuously enrolled. However, Appendix Figure 6, Panels A and B show that trends in monthly healthcare utilization and spending are parallel between the two groups. Panel C further illustrates that several diagnoses have the same ranking among the continuously enrolled as among the new enrollees. Appendix Figures 7 and 8 replicate the evidence on selection incentives at the service level and positive correlation between service profitability and network breadth among the continuously enrolled. Finally, Appendix Table 7 shows results of OLS regressions of an indicator for switching out of an insurer on that insurer's network breadth conditional on the continuously enrolled with certain health conditions.

Appendix Table 6: Summary statistics of new and continuous enrollees in 2011

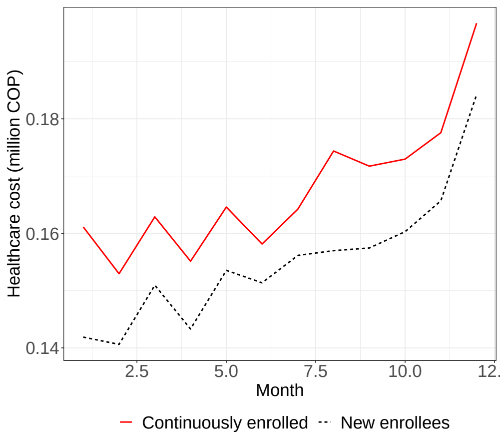
	(1) Continuous		(2) New		(3) Demand sample	
	mean	sd	mean	sd	mean	sd
Male	0.447	0.497	0.508	0.500	0.507	0.500
Age	45.526	15.960	41.791	15.343	41.849	15.354
Cancer	0.088	0.283	0.058	0.233	0.057	0.232
Diabetes	0.042	0.200	0.018	0.135	0.019	0.136
Cardiovascular disease	0.194	0.396	0.099	0.299	0.099	0.299
Pulmonary disease	0.021	0.142	0.010	0.099	0.010	0.100
Renal disease	0.018	0.133	0.007	0.081	0.007	0.082
Other disease	0.071	0.257	0.039	0.193	0.039	0.193
Total healthcare cost†	0.591	3.623	0.358	3.708	0.360	2.219
Risk-adjusted transfer†	0.761	0.477	0.666	0.415	0.668	0.416
Income†	1.280	0.431	1.238	0.402	1.256	0.385
Observations	7675021		3101064		500000	

Note: Table presents mean and standard deviation in parenthesis of demographic and health characteristics of the continuously enrolled in column (1), new enrollees in column (2), and the sample of new enrollees for demand model estimation in column (3) during 2011. (†) measured in millions of COP.

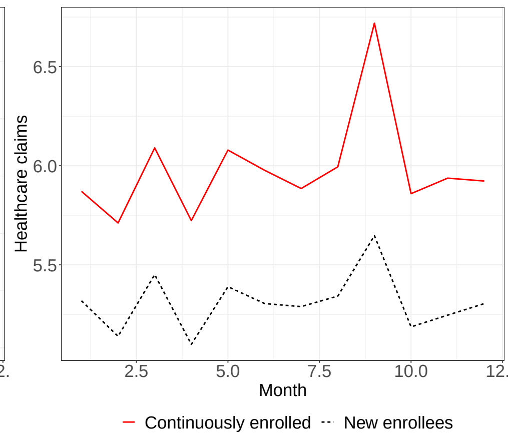
Appendix Table 7: Determinants of switching in the continuously enrolled

Sample:	Healthy	Cancer	Diabetes	Cardio
Service:	General medicine	Therapy	Laboratory	Cardiac care
	(1)	(2)	(3)	(4)
Service network breadth	-0.0010	-0.00048	-0.00024	-0.00093
	(0.00013)	(0.00022)	(0.00016)	(0.00015)
Demographics (male, age)	Yes	Yes	Yes	Yes
Enrollment spell length	Yes	Yes	Yes	Yes
Market FE	Yes	Yes	Yes	Yes
N	2783571	422246	243125	1147976
R^2	0.00027	0.00027	0.00011	0.00025

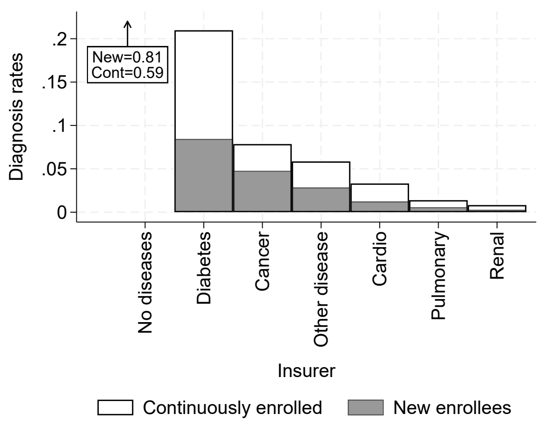
Note: Table presents OLS regression of a switching indicator on service network breadth for the 2010 insurer. Estimation uses the sample of continuously enrolled. Column (1) uses the sub-sample of individuals without diagnoses and network breadth for general medicine. Column (2) uses the sub-sample of individuals with cancer and network breadth for chemotherapy. Column (3) uses the sub-sample of individuals with diabetes and network breadth for laboratory. Column (4) uses the sub-sample of individuals with cardiovascular disease and network breadth for cardiac care services. All specifications control for enrollees' demographic characteristics, days enrolled, and market fixed effects. Robust standard errors in parenthesis.



(a) Monthly healthcare spending



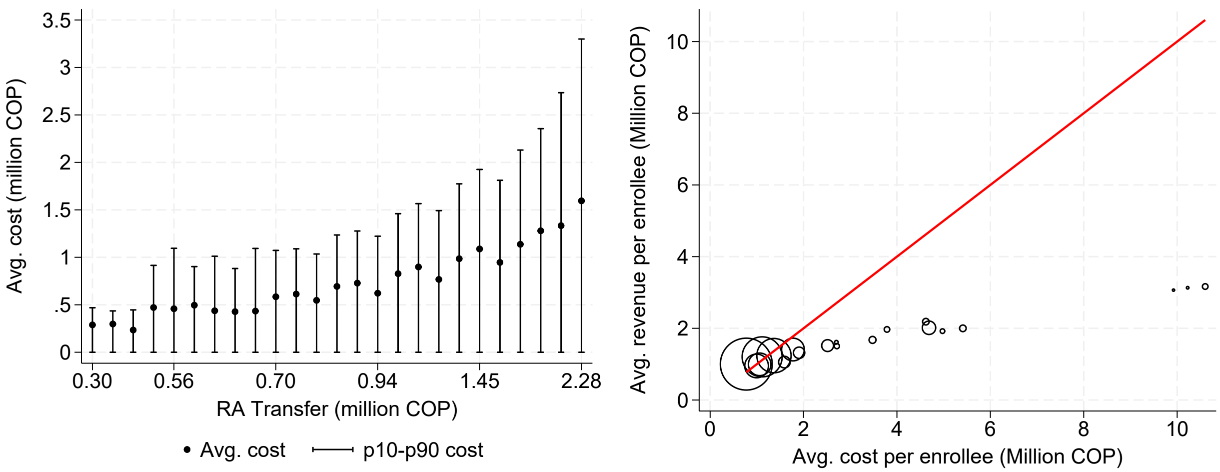
(b) Monthly number of claims



(c) Diagnosis rates

Appendix Figure 6: Utilization trends among new and continuous enrollees

Note: Panel A presents the trend in average monthly healthcare spending per enrollee (in million COP) during 2011 among the continuously enrolled in the solid red line and among new enrollees in the dashed black line. Panel B presents the trend in the average monthly number of claims per enrollee among the continuously enrolled in the solid red line and among new enrollees in the dashed black line. Panel C presents the fraction of continuously enrolled individuals and of new enrollees that have diabetes, cancer, cardiovascular disease, pulmonary disease, renal disease, other diseases, and no diseases during 2011.

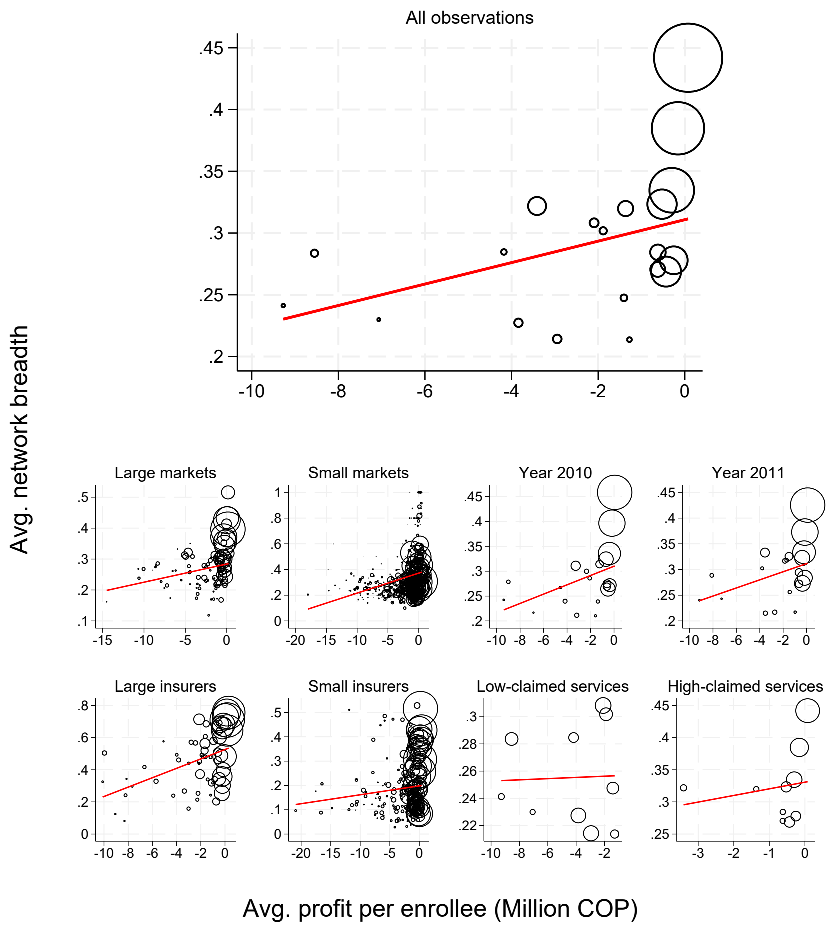


(a) Cost by risk-adjusted transfer

(b) Service-level selection incentives

Appendix Figure 7: Costs and selection incentives in the continuously enrolled

Note: Panel (a) of the figure presents mean, and 10th and 90th percentiles of annual healthcare cost by ex-ante government's risk-adjusted transfer in the sample of continuously enrolled. Panel (b) presents a scatter plot of average cost per enrollee against average revenue per enrollee in the sample of continuously enrolled. Each dot is a service weighted by the number of individuals who make claims for the service. The red line is a 45 degree line. One enrollee can be represented in several dots if she makes claims for different services. Enrollees who make zero claims are not represented in this figure.



Appendix Figure 8: Network breadth and service profitability in the continuously enrolled

Note: Figure presents a scatter plot of average service network breadth against average profit per enrollee in the sample of continuously enrolled. Each dot is a service weighted by the number of individuals who make claims for the service. Profits are calculated as government ex-ante and ex-post transfers, plus revenues from copays and coinsurance rates, minus total healthcare costs. The red line corresponds to a linear fit. One enrollee can be represented in several dots if she makes claims for different services. Enrollees who make zero claims are not represented in this figure.

Appendix 5 Micro-foundation

Insurer demand. For the demand side, take one market and consider a simple model of provider choice where individual i's indirect utility from choosing provider h for service k in the network of insurer j is:

$$u_{ijkh} = \overline{\xi_k} H_{jk} + \nu_{ijkh}$$

This model assumes that providers have identical quality *conditional on the service* which is equal to $\overline{\xi}_k$ weighted by the fraction of covered providers H_{jk}. Moreover, ν_{ijkh} is a preference shock distributed T1EV. Following McFadden (1996), individual i's value for insurer j's network of providers in service k, G_{jk}, is:

$$w_{ijk} = \log\Big(\sum_{h \in G_{jk}} \exp(\overline{\xi}_k H_{jk})\Big)$$

which simplifies to:

$$w_{ijk} = \log\Big(\sum_{h \in G_{jk}} \exp(\overline{\xi}_k H_{jk})\Big) = \log(|G_{jk}| \exp(\overline{\xi}_k H_{jk})) = \log(|G_{jk}|) + \overline{\xi}_k H_{jk} = \phi_{jk} + \overline{\xi}_k H_{jk}$$

where $|G_{jk}|$ is the number of providers in insurer j's network for service k and $\log(|G_{jk}|) = \phi_{jk}$. Summing across services yields:

$$\sum_k w_{ijk} = \phi_j + \sum_k \overline{\xi}_k H_{jk}$$

where $\phi_j = \sum_k \phi_{jk}$. This shows that insurer demand can be modelled as a function of $\sum_k \overline{\xi}_k H_{jk}$ and insurer fixed effects ϕ_j. It also shows that a demand function defined over network breadth is an exact representation of markets where providers are homogeneous *conditional on the service.*

The relation between network valuation and network breadth can be extended to a model where providers differ in quality and where consumers have heterogeneous preferences as follows. Suppose the utility function is:

$$u_{ijkh} = x_{\theta(i)}\xi_{hk} + \varepsilon_{ijkh}$$

where $x_{\theta(i)}$ is a vector of observed consumer characteristics describing a consumer type θ. Let γ_θ be the fraction of consumers type θ in the population and $|G_k|$ the total number of providers that deliver service k. Then:

$$\begin{aligned}\sum_\theta \gamma_\theta w_{\theta(i)jk} &= \sum_\theta \gamma_\theta \log\Big(\sum_{h\in G_{jk}} \exp(x_{\theta(i)}\xi_{hk})\Big) \geq \sum_\theta \gamma_\theta \log\Big(\frac{1}{|G_k|}\sum_{h\in G_{jk}} \exp(x_{\theta(i)}\xi_{hk})\Big)\\ &\geq \sum_\theta \gamma_\theta \frac{1}{|G_k|}\sum_{h\in G_{jk}} \log(\exp(x_{\theta(i)}\xi_{hk})) = \sum_\theta \gamma_\theta \frac{1}{|G_k|}\sum_{h\in G_{jk}} x_{\theta(i)}\xi_{hk}\\ &= \sum_\theta \gamma_\theta \frac{|G_{jk}|}{|G_k|}\sum_{h\in G_{jk}} \frac{1}{|G_{jk}|} x_{\theta(i)}\xi_{hk} = \sum_\theta \gamma_\theta x_{\theta(i)}\bar{\xi}_{jk}H_{jk}\end{aligned}$$

where the second inequality follows from Jensen's inequality and $\bar{\xi}_{jk} = \frac{1}{|G_{jk}|}\sum_{h\in G_{jk}} \xi_{hk}$ is the average quality of providers in insurer j's network. This derivation indicates that when providers differ in quality conditional on the service and when consumers have heterogeneous preferences, a model of insurer demand defined over $\gamma_\theta x_{\theta(i)}\bar{\xi}_{jk}H_{jk}$ will result in a lower bound for consumer surplus relative to a demand function defined over $\gamma_\theta w_{\theta(i)jk}$.

Insurer costs. Moving to the supply side, suppose that insurer j and provider h engage in bilateral negotiations over service prices. Let $D_j(\cdot)$ be insurer j's demand, R the per-capita risk-adjusted transfer, $D_{jhk}(\cdot)$ provider h's demand for service k from j's enrollees, p_{jhk} the negotiated price, m_{hk} provider h's marginal cost of providing service k, H_{jk} the set of providers in insurer j's network for service k, and J_{hk} the set of insurers that cover provider h for service k. Insurer profits can be written as $\pi^j = D_j(\cdot)R - \sum_k\sum_{h\in H_{jk}} D_{jhk}(\cdot)p_{jhk}$ and provider profits as $\pi^h = \sum_k\sum_{j\in J_{hk}} D_{jhk}(\cdot)(p_{jhk} - m_{hk})$. For simplicity, suppose that $D_{jhk}(\cdot)$ does not depend on prices as in Ho and Lee (2017).

The log Nash surplus function is:

$$S_{jhk} = \beta\log(\pi_j - t_h^j) + (1-\beta)\log(\pi_h - t_j^h)$$

where β is the bargaining power of the insurer, and t_h^j and t_j^h are the insurer and provider disagreement payoffs, respectively. The insurer disagreement payoff is defined as the profit it would enjoy if it excludes provider h from the network, while reimbursing the rest of providers at their equilibrium prices. Provider disagreement payoffs are defined analogously. The FOC of the log Nash surplus function with respect to the negotiated price is:

$$\sum_k D_{jhk} p_{jhk} = \beta\Big(\sum_k D_{jhk} m_{hk} - \sum_k \sum_{n \in J_h \backslash j} \Delta D_{nhk}(\cdot)(p_{nhk} - m_{hk})\Big)$$
$$+ (1-\beta)\Big(\Delta D_j(\cdot) R - \sum_k \sum_{l \in H_j \backslash h} \Delta D_{jlk}(\cdot) p_{jlk}\Big)$$

Adding these FOCs across all providers in the market for service k, imposing symmetry across providers for service k, and dividing on both sides by insurer j's demand, yields the following expression for the insurer's average cost per enrollee:

$$\begin{aligned}
AC_j &= \frac{1}{D_j}\beta\Big(\sum_k \overline{D}_{jk}\overline{m}_k H_{jk} - \sum_k \Delta\overline{D}_{nk}(\overline{p}_{nk} - \overline{m}_k)(|J_k|-1)H_{jk}\Big) \\
&+ \frac{1}{D_j}(1-\beta)\Big(\Delta D_j R H_{jk} - \sum_k \Delta\overline{D}_{jk}\overline{p}_{jk}\ (|G_{jk}|-1)H_{jk}\Big) \\
&= \frac{1}{D_j}\beta\sum_k\Big(\overline{D}_{jk}\overline{m}_k - \Delta\overline{D}_{nk}(\overline{p}_{nk} - \overline{m}_k)(|J_k|-1)\Big)H_{jk} \\
&+ \frac{1}{D_j}(1-\beta)\Big(\Delta D_j R\Big)H_{jk} + \frac{1}{D_j}(1-\beta)\sum_k\Big(\Delta\overline{D}_{jk}\overline{p}_{jk}\Big)H_{jk} \\
&- \frac{1}{D_j}(1-\beta)\sum_k\Big(\Delta\overline{D}_{jk}\overline{p}_{jk}\ |G_k|\Big)H_{jk}^2 \\
&= f(H_{jk}, H_{jk}^2)
\end{aligned}$$

where variables with over-lines denote the value for the average provider in service k. This derivation shows that an average cost function that is quadratic in network breadth is a correct simplification when providers are homogeneous conditional on the service. Together with results on demand, this shows that my proposed model is internally consistent.

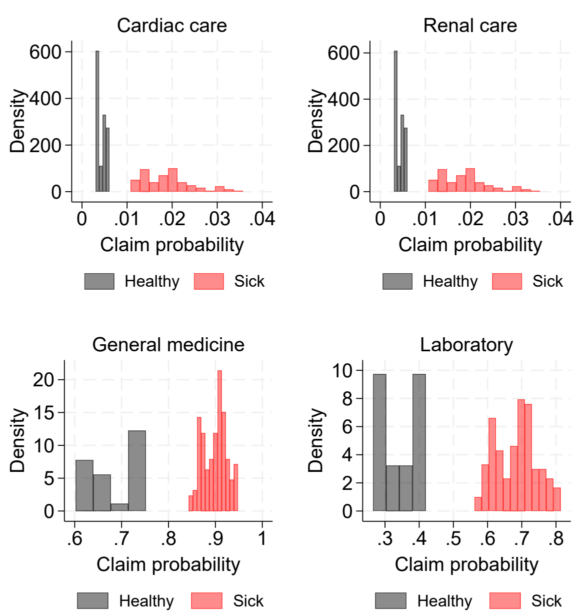
Appendix 6 Model inputs

6.1 Service claim probabilities

I estimate the claim probabilities using the following logistic regression:

$$\text{logit}(\text{any claims})_{ik} = \psi_k + \psi_\theta + \epsilon_{ik} \tag{6}$$

The dependent variable is an indicator for whether patient i makes a claim for service k. On the right side, ψ_k and ψ_θ are service and consumer type fixed effects, respectively. ϵ_{ikm} is a mean zero shock to the claim probability that is independent of network breadth conditional on consumer observables. I assume that new enrollees' expectations over the services they will need are correct on average and that these expectations do not depend on the insurer they enroll with. I estimate equation (6) on data from both current and new enrollees in 2010 and 2011. Appendix Figure 9 presents the resulting distribution of $q_{\theta k}$ for a few services such as cardiac care, renal care, general medicine, and laboratory testing.



Appendix Figure 9: Distribution of service claim probability

Note: Figure presents the distribution of the probability of making a claim in a few specific services, separately for individuals with (sick) and without (healthy) diagnoses. Services reported in the figure include cardiac care, renal care, general medicine, and laboratory testing.

6.2 Support for model assumptions

This subsection presents supporting evidence for the assumptions of my model. Appendix Table 8 shows that negotiated service prices and total service healthcare costs are higher the broader is the network. Hence, consumers who enroll broad-network insurers face a trade-off between higher out-of-pocket payments and greater coverage. Appendix Table 9 shows that service claim probabilities are uncorrelated with service network breadth conditional on the consumer type, providing suggestive evidence for using demographic variation in claim probabilities to identify the parameter on network breadth in the demand model. Appendix Table 10 illustrates that the probability of receiving a chronic disease diagnosis in 2011 conditional on not having one in 2010 is uncorrelated with network breadth. Thus, transition probabilities in the insurer profit function are orthogonal to network breadth. Appendix Figure 10 displays each insurer's average network breadth in each market. This figure evidences that insurers that offer broad networks in one market, tend to do so in other markets as well, supporting my specification of network formation costs being insurer-market specific.

Appendix Table 8: Correlation between negotiated prices and network breadth

	Log negotiated price (1)	Log total cost (2)
Service network breadth	0.423 (0.154)	3.793 (0.195)
Fraction male	Yes	Yes
Fraction of enrollees with diseases	Yes	Yes
Log average income	Yes	Yes
Number of enrollees	Yes	Yes
Market-service-year FE	Yes	Yes
N	7014	7014
R^2	0.424	0.597

Note: Table presents an OLS regression using as outcomes the log of negotiated service prices in column (1) and the log of total healthcare costs per service in column (2). An observation is an insurer-service-market-year. Negotiated prices are calculated as the average price across providers weighted by number of claims. Specifications control for the fraction of males; fraction of enrollees with cancer, diabetes, cardiovascular disease, pulmonary disease, renal disease, and other other diseases; log of average income; and number of enrollees. Specifications include market-service-year fixed effects. Robust standard errors in parenthesis.

Appendix Table 9: Correlation between claim probabilities and network breadth

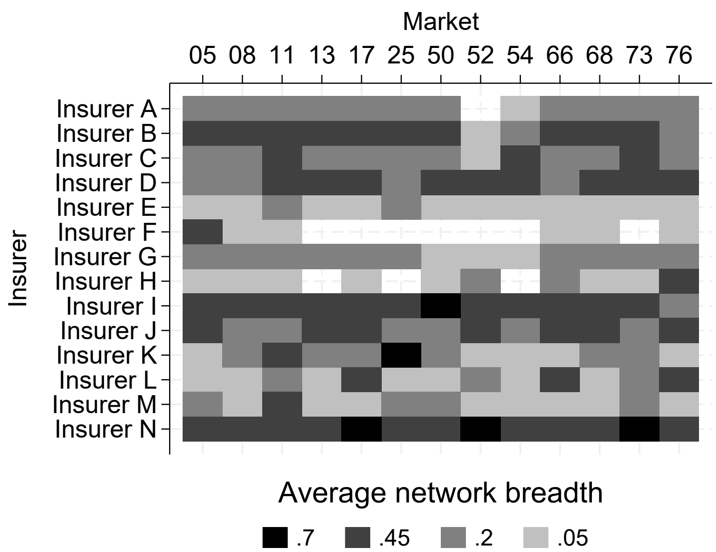
	Claim probability
Service network breadth	0.00033
	(0.0006)
Constant	0.0957
	(0.0002)
Consumer type FE	Yes
Service FE	Yes
Market FE	Yes
N	10000000
R^2	0.945

Note: Table presents an OLS regression of the claim probability on service network breadth. An observation is an individual-service. Estimation uses the subsample of new enrollees for model estimation. Specification includes consumer type, service, and market fixed effects. Standard errors are clustered at the consumer type level.

Appendix Table 10: Correlation between diagnosis probability and network breadth

	Any diagnosis in 2011
Claim-weighted service network breadth	0.00022
	(0.00072)
Claim probabilities	Yes
Market FE	Yes
N	2783735
R^2	0.997

Note: Table presents an OLS regression of an indicator of being diagnosed with any chronic condition in 2011 on service network breadth weighted by the claim probabilities. Estimation uses the subsample of continuously enrolled individuals who did not have a diagnosis in 2010. Specification controls for the service claim probabilities and includes market fixed effects. Standard errors are clustered at the insurer and market level.



Appendix Figure 10: Average network breadth across markets

Note: Figure presents the average network breadth (averaged across services) for each insurer in the row and each market in the column. Darker colors indicate higher network breadth.

6.3 In-sample demand model fit and additional results

This appendix shows the observed and model-predicted national market shares and market shares in the three largest markets for every insurer.

Appendix Table 11: National market shares

Insurer	Observed	Predicted
EPS001	2.219	2.228
EPS002	8.654	8.583
EPS003	4.441	4.444
EPS005	5.655	5.667
EPS008	5.039	5.019
EPS009	1.995	2.010
EPS010	7.973	7.992
EPS012	1.620	1.640
EPS013	14.878	14.887
EPS016	16.702	16.687
EPS017	7.572	7.548
EPS018	4.280	4.323
EPS023	2.961	2.964
EPS037	16.011	16.008

Note: Table presents observed and model-predicted national market shares using the demand model.

Appendix Table 12: Market shares in three largest markets

	Market 05		Market 11		Market 76	
Insurer	Obs	Pred	Obs	Pred	Obs	Pred
EPS001	0.778	3.295	4.298	2.194	1.183	1.117
EPS002	5.155	6.209	9.547	11.051	2.952	3.463
EPS003	3.070	3.572	8.152	7.541	.875	1.670
EPS005	1.420	3.797	11.346	7.427	2.634	4.505
EPS010	26.999	16.899	3.155	5.966	4.451	4.331
EPS013	11.335	17.231	9.206	12.304	7.424	7.491
EPS016	25.154	21.606	3.691	6.924	27.331	28.424
EPS023	2.300	2.637	6.541	6.731	1.885	1.084
EPS037	14.210	14.851	12.273	7.534	16.888	16.114

Note: Table presents observed and model-predicted market shares in the three largest markets using the demand model.

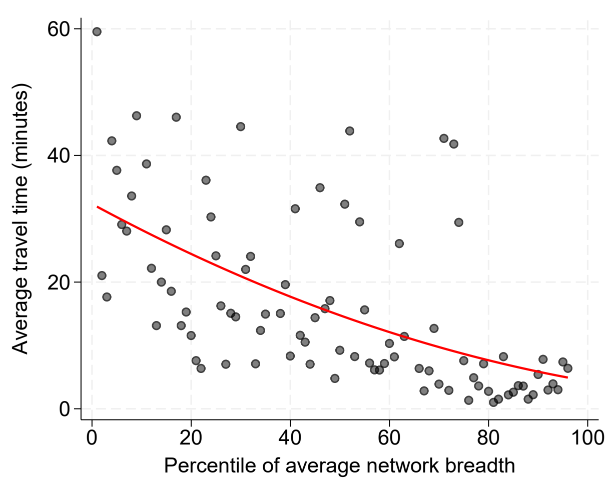
6.4 Alternative specifications

This appendix presents alternative demand specifications to provide encouraging evidence of my identification arguments and modelling choices.

Appendix Table 13: Average willingness-to-pay per service and age group

Age group	Cardiac care	Renal care	Imaging	General medicine	Laboratory	Hospital admissions
Age 19-24	1.111	1.111	1.395	1.790	1.507	1.150
Age 25-29	0.885	0.885	1.051	1.217	1.108	0.909
Age 30-34	0.841	0.841	1.022	1.223	1.087	0.867
Age 35-39	0.519	0.519	0.643	0.794	0.689	0.537
Age 40-44	0.952	0.952	1.151	1.396	1.224	0.980
Age 45-49	0.810	0.809	0.958	1.139	1.012	0.831
Age 50-54	0.845	0.845	1.005	1.185	1.061	0.869
Age 55-59	1.101	1.101	1.252	1.424	1.305	1.124
Age 60-64	1.002	1.002	1.091	1.177	1.118	1.017
Age 65-69	0.912	0.912	0.983	1.054	1.005	0.924
Age 70-74	0.920	0.920	0.974	1.026	0.991	0.929
Age 75+	1.000	1.000	1.000	1.000	1.000	1.000

Note: Table presents average willingness-to-pay for a percentage point increase in network breadth for the service in the column. Willingness-to-pay is calculated as $\frac{1}{-\alpha_i}\frac{\partial s_{ijm}}{\partial H_{jkm}}$, averaged across all consumers with the age group in the row, and normalized to 1 for individuals aged 75 or older.



Appendix Figure 11: Correlation between network breadth and travel times

Note: Figure presents the a scatter plot of average travel time (in minutes) for every percentile of average network breadth (averaged across services). Travel times are calculated between the municipality centroid and the nearest in-network provider for every insurer. The red line corresponds to a quadratic fit.

Appendix Table 14: Insurer demand with star hospital indicator

	Network breadth		OOP spend		Star hospital	
	coef	se	coef	se	coef	se
Mean	3.441	(0.021)	-1.586	(0.117)	0.037	(0.005)
<u>Interactions</u> Male	0.543	(0.010)	0.121	(0.066)		
Cancer	-0.601	(0.013)	-0.012	(0.093)		
Cardiovascular	-0.902	(0.011)	-0.209	(0.076)		
Diabetes	-0.465	(0.023)	0.007	(0.105)		
Other disease	-0.785	(0.015)	0.457	(0.068)		
Pulmonary	-0.611	(0.031)	0.836	(0.095)		
Renal	0.039	(0.037)	0.860	(0.069)		
Age 19-24	0.055	(0.020)	0.577	(0.158)		
Age 25-29	-0.576	(0.019)	0.328	(0.120)		
Age 30-34	-0.560	(0.019)	0.346	(0.130)		
Age 35-39	-0.456	(0.020)	-0.306	(0.214)		
Age 40-44	-0.356	(0.020)	0.456	(0.160)		
Age 45-49	-0.384	(0.019)	0.236	(0.141)		
Age 50-54	-0.324	(0.020)	0.319	(0.134)		
Age 55-59	-0.246	(0.021)	0.476	(0.122)		
Age 60-64	-0.147	(0.023)	0.162	(0.121)		
N			5200890			
Pseudo-R^2			0.112			

Note: Table presents the insurer demand model including a measure of star hospital coverage equal to $\sum_k q_{\theta k} Star_{jkm}$, where $Star_{jkm}$ is an indicator for whether insurer j covers a star hospital in market m for service k. Star hospitals are the top 50 academic medical centers in the country. Specification includes insurer fixed effects. Robust standard errors in parenthesis.

Appendix Table 15: Insurer demand with diagnosis received in January

Variable	Network Breadth		OOP spend	
	coef	se	coef	se
Mean	2.725	(0.021)	-1.048	(0.096)
<u>Interactions</u>				
Male	0.638	(0.010)	-0.013	(0.068)
Cancer	-0.512	(0.031)	-0.164	(0.234)
Cardiovascular	-0.542	(0.019)	-0.605	(0.168)
Diabetes	-0.645	(0.042)	0.338	(0.227)
Other disease	-0.674	(0.033)	0.233	(0.155)
Pulmonary	-0.933	(0.065)	0.830	(0.203)
Renal	-0.519	(0.083)	0.810	(0.087)
Age 19-24	0.411	(0.020)	0.489	(0.148)
Age 25-29	-0.234	(0.019)	0.232	(0.117)
Age 30-34	-0.237	(0.019)	0.375	(0.126)
Age 35-39	-0.163	(0.020)	-0.394	(0.236)
Age 40-44	-0.094	(0.020)	0.547	(0.152)
Age 45-49	-0.176	(0.020)	0.430	(0.140)
Age 50-54	-0.195	(0.020)	0.420	(0.134)
Age 55-59	-0.158	(0.021)	0.337	(0.118)
Age 60-64	-0.090	(0.023)	0.191	(0.130)
N		5200890		
Pseudo-R^2		0.107		

Note: Table presents the insurer demand model obtaining individuals' diagnoses from claims made in January 2011. Specification includes insurer fixed effects. Robust standard errors in parenthesis.

Appendix Table 16: Insurer demand with network breadth weighted by number of beds

Variable	Network Breadth		OOP spend	
	coef	se	coef	se
Mean	4.325	(0.026)	-1.263	(0.111)
Interactions				
Male	0.846	(0.014)	-0.066	(0.061)
Cancer	-0.636	(0.019)	0.018	(0.080)
Cardiovascular	-1.243	(0.016)	-0.117	(0.068)
Diabetes	-0.609	(0.032)	-0.013	(0.102)
Other disease	-1.073	(0.021)	0.247	(0.065)
Pulmonary	-0.905	(0.041)	0.669	(0.097)
Renal	-0.069	(0.057)	0.481	(0.065)
Age 19-24	0.126	(0.029)	1.034	(0.161)
Age 25-29	-0.828	(0.026)	0.612	(0.113)
Age 30-34	-0.807	(0.027)	0.767	(0.116)
Age 35-39	-0.669	(0.028)	0.154	(0.185)
Age 40-44	-0.528	(0.028)	0.878	(0.140)
Age 45-49	-0.551	(0.027)	0.641	(0.130)
Age 50-54	-0.472	(0.028)	0.597	(0.126)
Age 55-59	-0.351	(0.030)	0.641	(0.120)
Age 60-64	-0.215	(0.032)	0.272	(0.120)
N	5200890			
Pseudo-R^2	0.116			

Note: Table presents the insurer demand model with service network breadth weighted by each provider's number of beds. The denominator in network breadth is the total number of beds in the market and the numerator is the number of beds included in the insurer's network. Specification includes insurer fixed effects. Robust standard errors in parenthesis.

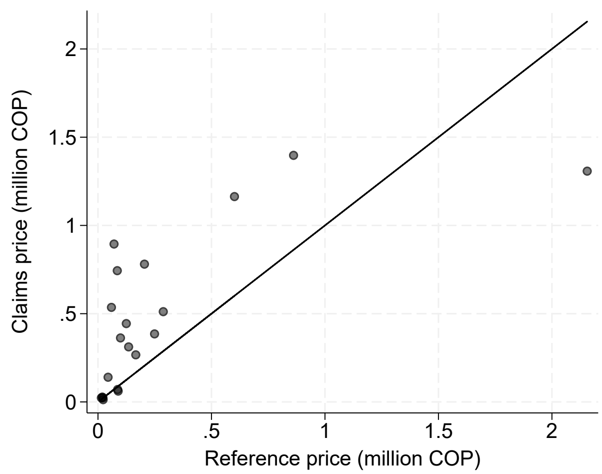
Appendix Table 17: Insurer demand with expectations over diagnoses

Variable	Network Breadth		OOP spend	
	coef	se	coef	se
Mean	2.381	(0.018)	-3.865	(0.134)
Interactions				
Male	0.688	(0.009)	-1.114	(0.135)
Age 19-24	1.038	(0.018)	8.139	(0.310)
Age 25-29	0.267	(0.016)	-0.175	(0.213)
Age 30-34	0.232	(0.017)	-1.721	(0.291)
Age 35-39	0.281	(0.018)	-2.919	(0.297)
Age 40-44	0.325	(0.017)	1.076	(0.351)
Age 45-49	0.160	(0.018)	0.290	(0.262)
Age 50-54	0.058	(0.018)	-0.370	(0.247)
Age 55-59	0.085	(0.018)	1.344	(0.235)
Age 60-64	0.031	(0.020)	0.197	(0.229)
N	5200890			
Pseudo-R^2	0.116			

Note: Table presents the insurer demand model assuming consumers form expectations over diagnoses and services. Let $\tilde{\theta}$ be a combination of sex and age group. Network breadth is calculated as $\sum_d \sum_k \gamma_{d\tilde{\theta}} q_{\theta k} H_{jkm}$ and out-of-pocket costs are calculated as $\sum_d \gamma_{d\tilde{\theta}} c_{\theta jm}$, where $\gamma_{d\tilde{\theta}}$ is the probability of having diagnosis d conditional on $\tilde{\theta}$. Specification includes insurer fixed effects. Robust standard errors in parenthesis.

Appendix 7 Service reference prices

In 2005, the Colombian government published a list of reference prices for all the services included in the national health insurance plan. These prices were created by a group of government officials and medical experts with the purpose of reimbursing healthcare providers in the event of terrorist attacks, natural disasters, and car accidents (Decree 2423 of 1996). Although they were not meant to guide price negotiations between insurers and providers, there is evidence that insurers use these reference prices as starting points in their negotiations with providers (Ruiz et al., 2008). Appendix Figure 12 shows that references prices are highly correlated with negotiated prices from the claims data and that negotiated prices tend to be higher than the reference price.

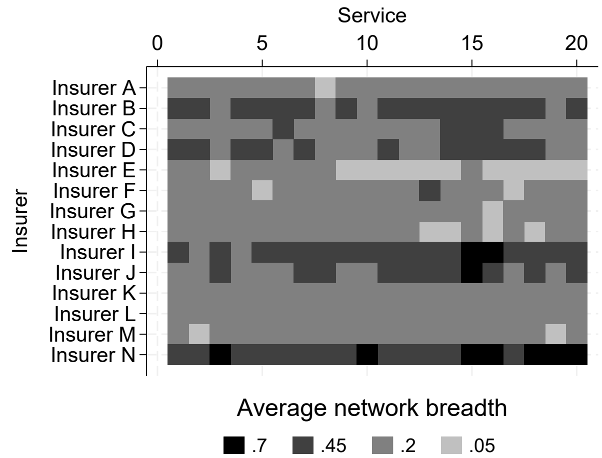


Appendix Figure 12: Negotiated prices and reference prices

Note: Figure presents a scatter plot of average negotiated price obtained from the claims data and average reference price per service. The black line is the 45 degree line.

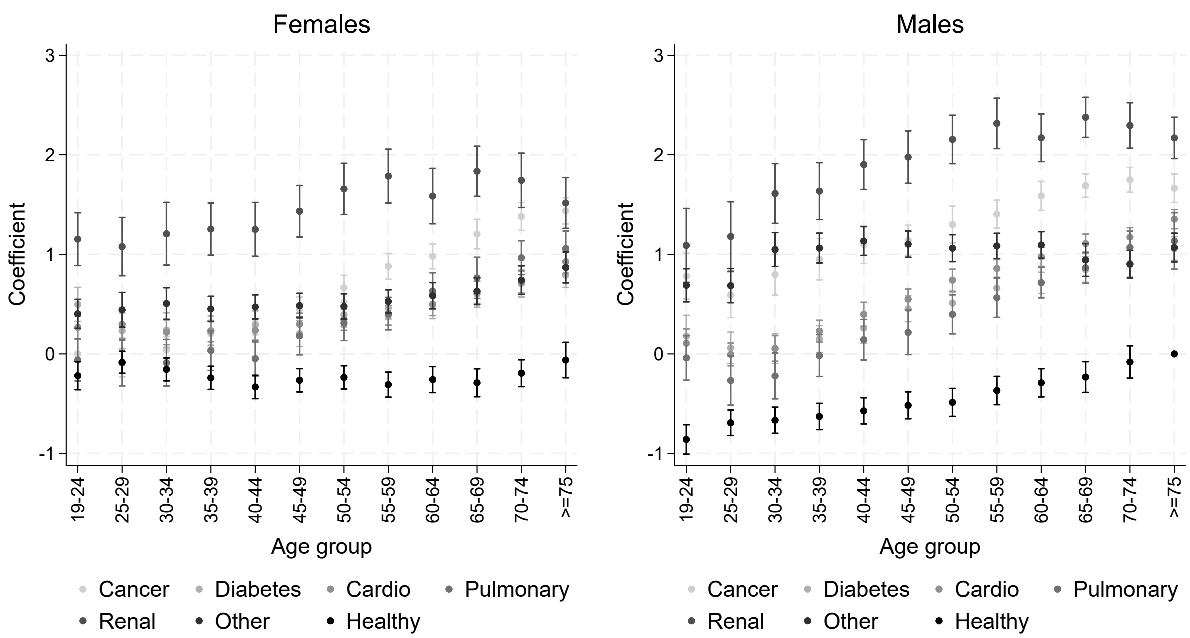
Appendix 8 Additional average cost results

This appendix presents additional descriptive evidence and robustness checks to support my identification and functional form assumption for the average cost model. Appendix Figure 13 shows that insurers' network breadth decisions are correlated across services. Appendix Figure 14 shows the consumer type fixed effects in my main specification. Appendix Figure 15 presents the average cost model fit. Appendix Table 18 shows evidence that insurers that cover several services within a provider receive price discounts, suggestive of economies of scope.



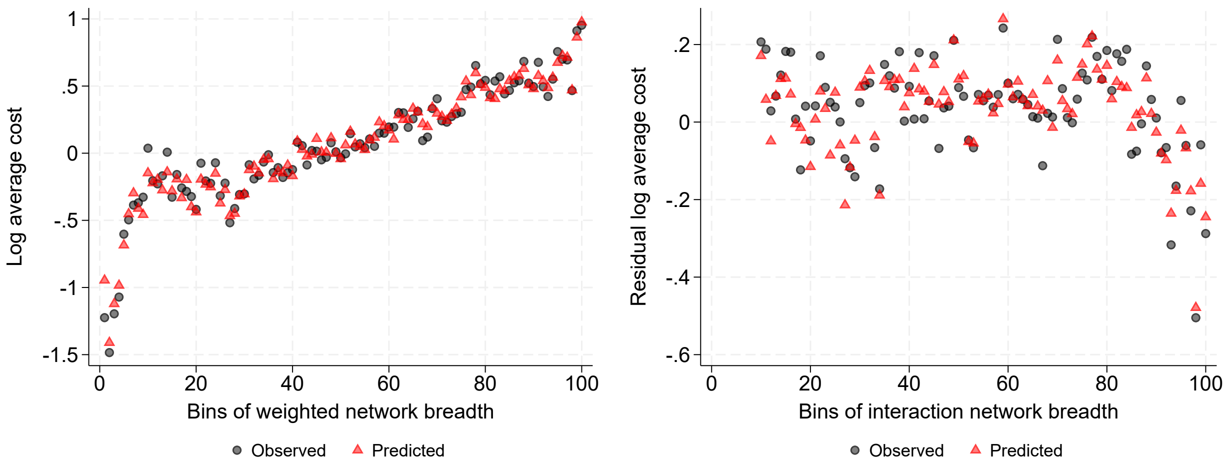
Appendix Figure 13: Average network breadth per insurer and service

Note: Figure presents the average service network breadth (averaged across markets) for each insurer in the row and each service in the column. Darker colors represent higher network breadth.



Appendix Figure 14: Consumer type fixed effects

Note: Figure presents point estimates and 95 percent confidence intervals of the consumer type fixed effects in the average cost function. The left panel presents fixed effects for females separately by disease category and age group. The right panel presents fixed effects for males separately by disease category and age group.



(a) Service network breadth

(b) Service network breadth interactions

Appendix Figure 15: Average cost model fit

Note: Panel A presents a scatter plot of observed (black circles) and predicted (red triangles) log average cost per enrollee by percentiles of weighted service network breadth ($\sum_k q_{\theta k} H_{jkm}$). Panel B presents a scatter plot of observed (black circles) and predicted (red triangles) residual log average cost per enrollee after controlling for service network breadth by percentiles of scope economies ($\sum_k^{K_m} \sum_{l \neq k}^{K_m} q_{\theta k} q_{\theta l} H_{jkm} H_{jlm}$).

Appendix Table 18: Price mechanism for economies of scope

	Log negotiated price	
	(1)	(2)
Provider in network for other services	-0.091	
	(0.002)	
Avg. network breadth for other services		-1.536
		(0.027)
Demographics (male, diagnoses, income)	Yes	Yes
Number of enrollees	Yes	Yes
Market-Service-Year FE	Yes	Yes
N	77312	80129
R^2	0.655	0.660

Note: Table presents OLS regression of the log of the negotiated price for service k between insurer j and provider h on the fraction of other services $l \neq k$ for which the provider is in network in column (1) and on average network breadth across all other services $l \neq k$ in column (2). An observation in this table is a combination of insurer, provider, service, market, year. Specifications control for the fraction of males; fraction of enrollees with cancer, diabetes, cardiovascular disease, pulmonary disease, renal disease, and other other diseases; log of average income; and number of enrollees for the insurer. Specifications include market-service-year fixed effects. Estimation is conditional on the provider being in-network for service k. The reduction in the number of observations in column (1) is due to providers that are in-network for service k only and for which the fraction is indeterminate because it is divided by zero. Robust standard errors in parenthesis.

Appendix Tables 19, 20, and 21 present robustness on my average cost model using patient-level data, including an indicator for star hospital coverage, and weighting network breadth by each provider's number of beds.

Appendix Table 19: Patient-level estimates of average cost

	Log total healthcare cost+1	
	coef	se
Service network breadth	5.591	(0.024)
Scope economies	-54.07	(0.265)
Reference price	-2.169	(0.062)
Insurer FE		
Insurer A	-0.645	(0.395)
Insurer B	0.762	(0.292)
Insurer C	0.954	(0.459)
Insurer D	1.803	(0.259)
Insurer E	2.588	(0.298)
Insurer F	0.976	(0.442)
Insurer G	1.968	(0.268)
Insurer H	0.248	(0.396)
Insurer I	0.650	(0.250)
Insurer J	0.827	(0.254)
Insurer K	1.214	(0.306)
Insurer L	0.106	(0.298)
Insurer M	1.897	(0.426)
Constant	-7.664	(0.167)
Individual FE	Yes	
Market FE	Yes	
N	8003976	
R^2	0.640	

Note: Table presents OLS regression of log healthcare cost (plus 1) per patient on network breadth, economies of scope, and service reference price. Uses a random sample of 500,000 patients. Includes insurer, market, and consumer type fixed effects. Reference price is omitted due to multicollinearity. Robust standard errors in parenthesis.

Appendix Table 20: Average cost with star hospital indicator

Variable	Log average cost per enrollee coef	se
Service network breadth	0.273	(0.047)
Scope economies	-1.147	(0.579)
Star hospital	0.041	(0.009)
Reference price	—	—
Insurer FE		
Insurer A	0.154	(0.038)
Insurer B	-0.093	(0.022)
Insurer C	0.017	(0.025)
Insurer D	-0.152	(0.027)
Insurer E	0.273	(0.047)
Insurer F	0.010	(0.061)
Insurer G	0.071	(0.034)
Insurer H	0.036	(0.040)
Insurer I	-0.008	(0.018)
Insurer J	0.107	(0.021)
Insurer K	-0.132	(0.033)
Insurer L	0.113	(0.035)
Insurer M	-0.050	(0.032)
Constant	-0.330	(0.053)
Consumer type FE	Yes	
Market FE	Yes	
N	18369	
R^2	0.612	

Note: Table presents OLS regression of log average cost per consumer type excluding the capital city, on network breadth, scope economies, and service reference prices. Includes insurer, market, and consumer type fixed effects. Robust standard errors in parenthesis.

Appendix Table 21: Average cost with with network breadth weighted by number of beds

	Log average cost per enrollee	
Variable	coef	se
Service network breadth	0.307	(0.065)
Scope economies	-1.272	(0.683)
Reference price	2.818	(0.570)
<u>Insurer FE</u>		
Insurer A	0.093	(0.037)
Insurer B	-0.166	(0.021)
Insurer C	-0.050	(0.023)
Insurer D	-0.242	(0.025)
Insurer E	0.195	(0.046)
Insurer F	0.001	(0.057)
Insurer G	0.015	(0.034)
Insurer H	0.012	(0.040)
Insurer I	-0.028	(0.017)
Insurer J	0.091	(0.021)
Insurer K	-0.212	(0.032)
Insurer L	0.133	(0.035)
Insurer M	-0.122	(0.031)
Constant	-1.393	(0.104)
Consumer type FE	Yes	
Market FE	Yes	
N	18369	
R^2	0.610	

Note: Table presents OLS regression of log average cost per consumer type excluding the capital city, on network breadth, scope economies, and service reference prices. Includes insurer, market, and consumer type fixed effects. Robust standard errors in parenthesis.

Appendix 9 Additional network administrative cost results

9.1 Dropout and transition probabilities

To estimate the marginal cost of network formation in the third step of my model, I first compute the probability that consumer type θ drops out of the contributory system and the probability that consumer type θ in period t transitions into θ' in period $t+1$. I use data from *all* enrollees to the contributory system in 2010 and 2011, regardless of their enrollment spell length, to compute dropout probabilities. For each consumer type θ, I calculate the probability that they drop out of the system non-parametrically as the number of individuals of type θ observed only in 2010 but not 2011, divided by the total number of type θ individuals in 2010. Appendix Table 22 presents the mean and standard deviation of the dropout probability conditional on diagnoses, sex, and age.

To compute transition probabilities I use a non-parametric approach as well, using data from all the continuously enrolled in 2010 and 2011. The probability that type θ transitions into θ' equals the number of type θ in 2010 that end up with diagnosis l' in 2011, divided by the number of type θ individuals in 2010. Appendix Table 23 presents the mean and standard deviation in parenthesis of transition probabilities from having cancer, diabetes, cardiovascular disease, pulmonary disease, renal disease, other diseases, and no diseases in period t to having each of these 7 health conditions in period $t+1$.

9.2 Marginal variable profits

This appendix presents additional descriptive evidence for marginal variable profits and measures of out-of-sample model fit.

Appendix Table 22: Dropout probabilities

	mean	sd
Female	0.084	(0.131)
Male	0.106	(0.165)
Age 19-24	0.120	(0.177)
Age 25-29	0.087	(0.134)
Age 30-34	0.081	(0.135)
Age 35-39	0.085	(0.141)
Age 40-44	0.085	(0.146)
Age 45-49	0.085	(0.149)
Age 50-54	0.089	(0.153)
Age 55-59	0.091	(0.158)
Age 60-64	0.092	(0.158)
Age 65-69	0.096	(0.159)
Age 70-74	0.104	(0.160)
Age 75+	0.124	(0.164)
Cancer	0.048	(0.024)
Diabetes	0.027	(0.008)
Cardiovascular	0.028	(0.009)
Pulmonary	0.040	(0.015)
Renal	0.044	(0.018)
Other disease	0.026	(0.011)
Healthy	0.450	(0.073)

Note: Mean and standard deviation in parenthesis of dropout probabilities conditional on diagnosis in the first panel, age group in the second panel, and sex in the third panel.

Appendix Table 23: Transition probabilities

Diagnosis	Stat	Cancer	Cardio	Diabetes	Renal	Pulmonary	Other	Healthy
Cancer	mean	0.316	0.017	0.139	0.014	0.007	0.047	0.460
	sd	0.067	0.014	0.090	0.013	0.006	0.019	0.176
Diabetes	mean	0.030	0.557	0.170	0.009	0.013	0.021	0.200
	sd	0.026	0.078	0.100	0.010	0.011	0.010	0.140
Cardio	mean	0.043	0.028	0.554	0.014	0.011	0.034	0.316
	sd	0.036	0.018	0.205	0.012	0.010	0.009	0.224
Pulmonary	mean	0.055	0.019	0.191	0.234	0.007	0.078	0.416
	sd	0.046	0.014	0.089	0.152	0.006	0.034	0.231
Renal	mean	0.044	0.036	0.214	0.012	0.371	0.058	0.265
	sd	0.035	0.030	0.132	0.013	0.062	0.031	0.154
Other	mean	0.056	0.016	0.156	0.023	0.008	0.343	0.398
	sd	0.040	0.013	0.106	0.020	0.004	0.058	0.095
Healthy	mean	0.055	0.012	0.108	0.014	0.004	0.045	0.762
	sd	0.042	0.008	0.068	0.014	0.003	0.021	0.109

Note: Table presents mean and standard deviation in parenthesis of transition probabilities across diagnoses. Summary statistics are calculated across sex-age combinations in each cell.

Appendix Table 24: Summary statistics of marginal variable profits

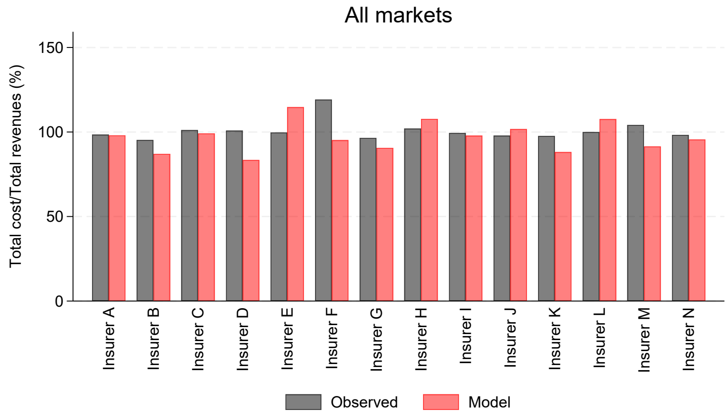
Insurer	mean	sd
Insurer A	6697.998	28615
Insurer B	34998.53	156924.5
Insurer C	16191.7	86058.47
Insurer D	23004.19	111649.9
Insurer E	93551.45	277196.6
Insurer F	11046.99	30985.02
Insurer G	27080.69	97283.26
Insurer H	41008.3	88315.88
Insurer I	31207.44	137559.4
Insurer J	25403.59	95181.3
Insurer K	57247.54	275614.2
Insurer L	14026.04	51929.83
Insurer M	40024.9	150600.8
Insurer N	27492.63	99826.83

Note: Table presents mean and standard deviation of marginal variable profits per insurer. Measured in millions of COP per service.

Appendix Table 25: First-stage regression for network formation cost estimation

	Service network breadth	
	coef	se
Claim probability, healthy	0.310	(0.019)
<u>Insurer FE</u>		
Insurer A	-0.322	(0.013)
Insurer B	-0.103	(0.016)
Insurer C	-0.255	(0.014)
Insurer D	-0.199	(0.014)
Insurer E	-0.249	(0.020)
Insurer F	-0.301	(0.026)
Insurer G	-0.330	(0.014)
Insurer H	-0.031	(0.029)
Insurer I	-0.074	(0.016)
Insurer J	-0.178	(0.015)
Insurer K	-0.198	(0.020)
Insurer L	-0.034	(0.026)
Insurer M	-0.303	(0.018)
Constant	0.524	(0.015)
Market FE	Yes	
F-statistic	249.47	
N	2280	

Note: Table presents first stage regression for my preferred specification of the network administrative cost. The outcome variable is service network breadth and the covariates are insurer fixed effects, market fixed effects, and the instrument which is the average service claim probability among healthy consumers. The table reports the F-statistic associated with the instrument. Robust standard errors in parenthesis.



Appendix Figure 16: Out-of-sample model fit

Note: Figure presents the model-predicted ratio of total costs (average costs plus network administrative costs) to total revenues and the observed ratio from insurers' public income statements submitted to the National Health Superintendency. Because my model is estimated on the sample of new enrollees with continuous enrollment and public income statements correspond to all enrollees, I scale up estimated insurer revenues and costs by multiplying by the total number of enrollees in the country and dividing by the number of new enrollees.

Appendix Table 26: Model of network administrative costs with service network breadth weighted by number of beds

	Log marginal variable profit	
Variable	coef	se
Service network breadth	21.50	(1.171)
Insurer FE		
Insurer A	4.557	(0.473)
Insurer B	1.156	(0.339)
Insurer C	3.559	(0.422)
Insurer D	2.315	(0.352)
Insurer E	4.346	(0.684)
Insurer F	5.067	(0.779)
Insurer G	6.527	(0.520)
Insurer H	-0.434	(0.875)
Insurer I	1.468	(0.325)
Insurer J	4.533	(0.407)
Insurer K	3.064	(0.433)
Insurer L	-1.099	(0.498)
Insurer M	4.914	(0.582)
Constant	-5.741	(0.889)
Market FE	Yes	
First-stage F-stat	325.55	
N	2206	

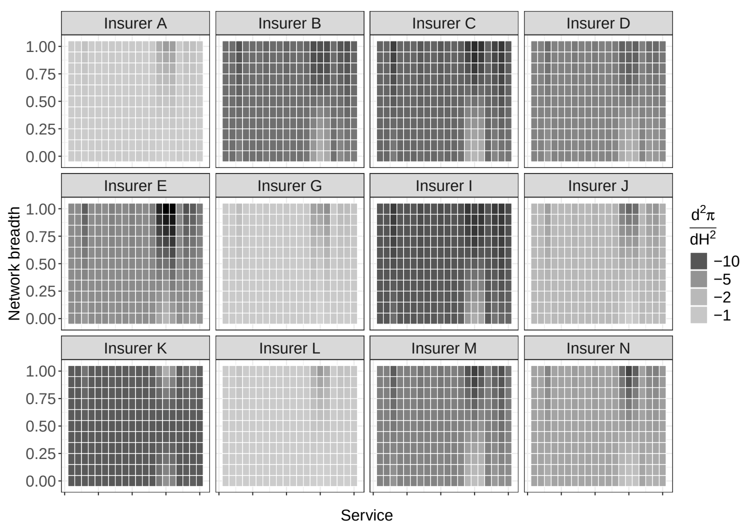
Note: Table presents 2SLS regression of the log of marginal variable profit on service network breadth, insurer fixed effects, and market fixed effects. Service network breadth is calculated using each provider's number of beds as weights. An observation is a combination of insurer, service, and market. The instrument for service network breadth is the average service claim probability among healthy consumers. The table reports the F-statistic for the first stage regression. Robust standard errors in parenthesis.

Appendix 10 Concavity of the profit function

The second partial derivative of the short-run profit function with respect to network breadth for service k is:

$$\frac{\partial^2 \pi_{jm}}{\partial H_{jkm}^2} = \sum_i \left((R_{\theta m} - (1-r_i)AC_{\theta jm}) \frac{\partial^2 s_{ijm}}{\partial H_{jkm}^2} - 2(1-r_i)\frac{\partial s_{ijm}}{\partial H_{jkm}} \frac{\partial AC_{\theta jm}}{\partial H_{jkm}} - (1-r_i)s_{ijm}\frac{\partial^2 AC_{\theta jm}}{\partial H_{jkm}^2} \right)$$

To check whether this derivative is negative at all values of network breadth, I conduct a partial equilibrium exercise where each insurer is allowed to deviate and set $H_{jkm} = \{0, 0.1, 0.2, 0.3, ..., 1\}$ for each service k, holding its rivals' choices fixed at observed levels. I compute this exercise with data from Bogotá as in my counterfactuals. Appendix Figure 17 presents the results. Each panel corresponds to the deviating insurer, and displays the value of the second partial derivative for each service in the horizontal axis and for each value of network breadth in the vertical axis. Results show that the second partial derivative of the short-run profit function is negative for all insurers and services.



Appendix Figure 17: Second partial derivative of short-run profit function

Note: Figure presents the second partial derivative of the insurer short-run profit function for every service. Each panel corresponds to an insurer, the horizontal axis is a service, and the vertical axis is the value of service network breadth.

Appendix 11 Additional counterfactual results

Appendix Table 27: Insurer demand

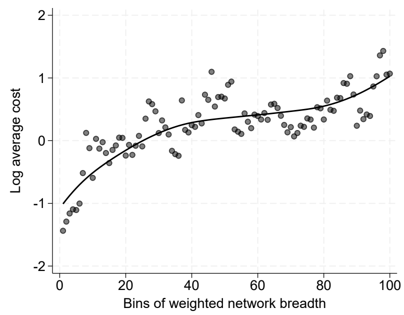
Diagnosis list
Healthy
Cardiovascular disease
Other Disease
Cervical Cancer
Breast Cancer
Other Renal Disease
Other Cancer
Chronic Kidney Disease
Diabetes
Skin Cancer
Lymphoma
Stomach Cancer
HIV-AIDS
Lung Cancer

Note: Table presents list of diagnoses used in the counterfactual with improved risk adjustment.

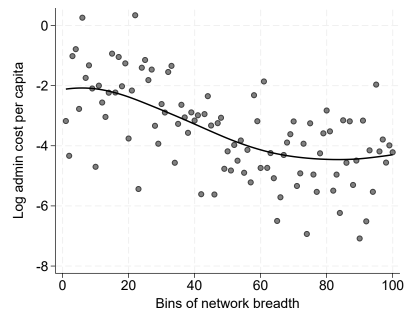
Appendix Table 28: Networks, costs, and welfare under homogeneous costs

Variable	(1) Avg cost	(2) Admin cost
Panel A. Overall		
Average network breadth	-0.218	-1.443
Average cost per enrollee	-0.408	-0.152
Total average cost	2.219	3.649
Consumer surplus (with diagnoses)	-0.246	2.885
Consumer surplus (without diagnoses)	-0.139	2.802
Panel B. Service network breadth		
Otorhinolaryngologic care	-0.217	-2.371
Cardiac care	-0.243	-2.265
Gastroenterologic care	-0.237	-2.285
Renal care	-0.258	-2.397
Gynecologic care	-0.244	-1.941
Orthopedic care	-0.245	-2.289
Imaging	-0.115	-0.402
General medicine	-0.036	-1.449
Laboratory	-0.088	-0.899
Hospital admission	-0.203	-1.475

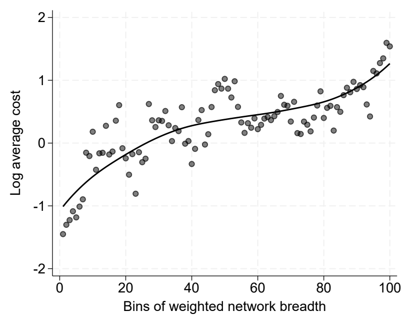
Note: Panel A presents the percentage change relative to the observed scenario in average network breadth, insurer total average costs, short-run average cost per enrollee, and long-run consumer welfare for individuals with and without diagnoses, in the scenario with homogeneous average costs in column (1), and the scenario with homogeneous average and network formation costs in column (2). Insurer fixed effects in average costs and network formation costs are set to the average fixed effect. Panel B presents the percentage change relative to the observed scenario in average network breadth by service category.



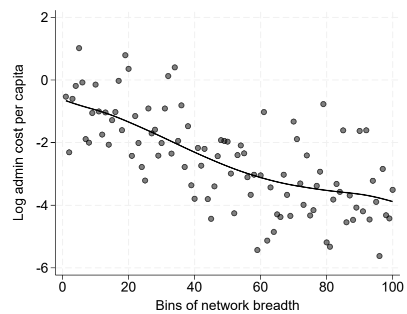
(a) No RA: avg. cost



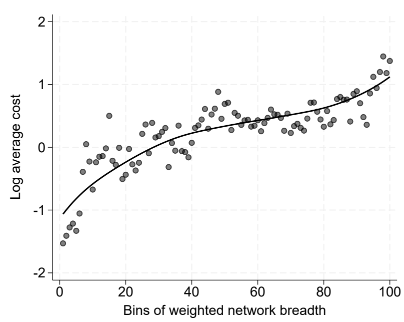
(b) No RA: admin. cost



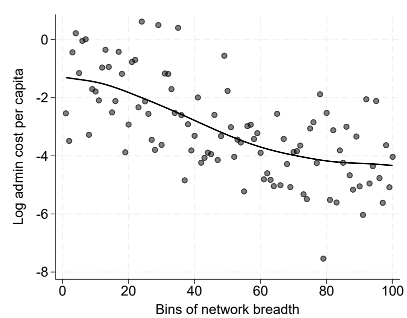
(c) RA: avg. cost



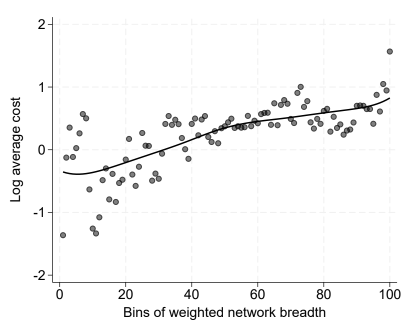
(d) RA: admin. cost



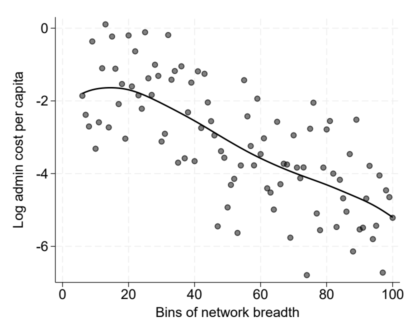
(e) Hom. Avg: avg. cost



(f) Hom. Avg: admin. cost



(g) Hom. admin: avg. cost



(h) Hom. admin: admin. cost

Appendix Figure 18: Functional form stability in counterfactuals

Note: Figure presents the predicted log average cost per enrollee and the predicted log network administrative cost per enrollee by percentile of service network breadth. Each row corresponds to a counterfactual: without risk adjustment, with improved risk adjustment, and imposing homogeneity in average and network administrative costs.