Market Size and Trade in Medical Services*

Jonathan I. Dingel
Chicago Booth,
CEPR, and NBER

Joshua D. Gottlieb
University of Chicago
and NBER

Maya Lozinski
University of Chicago

Pauline Mourot
Chicago Booth

April 2024

Abstract

We uncover substantial interregional trade in medical services and investigate whether regional increasing returns explain it. In Medicare data, one-fifth of production involves a doctor treating a patient from another region. Larger regions produce greater quantity, quality, and variety of medical services, which they "export" to patients from elsewhere, especially smaller regions. We show that these patterns reflect scale economies: greater demand enables larger regions to improve quality, so they attract patients from elsewhere. Despite concerns about rural access, larger regions have higher marginal returns to spending. We study counterfactual policies that would lower travel costs rather than relocating production.

Keywords: healthcare access, market-size effects, Medicare claims data, trade in services

JEL codes: F12, F14, I11, R12

*We thank Scott Blatte, Steve Buschbach, Cedric Elkouh, Aaron Haefner, Feng Lin, Luke Motley, Nick Niers, Noah Sobel-Lewin, Vaidehi Parameswaran, and Xiaoyang Zhang for excellent research assistance. We are grateful to Rodrigo Adão, Zarek Brot-Goldberg, Don Davis, Rebecca Diamond, Liran Einav, Ed Glaeser, Tali Han, Tom Holmes, Loukas Karabarbounis, Wojciech Kopczuk, Doug Miller, Sean Nicholson, Bruce Schakman, Martin Schneider, Bradley Setzler, Jon Skinner, Felix Tintelnot, discussants Jan David Bakker, Barthelemy Bonadio, Jessie Handbury, Naomi Hausman, Tim Layton, Elena Patel, and many seminar participants for helpful feedback. We thank the Becker Friedman Institute at the University of Chicago for funding support and enabling us to access the Medicare claims data. We thank Antoine Levy and Jacob Moscona for sharing CBSA-level bedrock-depth data and Amitabh Chandra, Maurice Dalton, and Doug Staiger for hospital quality data and code. Dingel thanks the Cohen and Keenoy Faculty Research Fund at the University of Chicago Booth School of Business for support. Lozinski thanks the NIH (T32GM007281 and T32AG051146). jdingel@chicagobooth.edu, jgottlieb@uchicago.edu, mayalozinski@uchicago.edu, pmourot@chicagobooth.edu.

Rural Americans have worse health outcomes (Finkelstein, Gentzkow, and Williams, 2021), but America's doctors are disproportionately located in big cities (Rosenblatt and Hart, 2000). This contrast might suggest a spatial mismatch between consumers and producers of medical services, and discussions of whether physicians are geographically "maldistributed" go back decades (Newhouse et al., 1982*a*; Skinner et al., 2019). We consider an alternative explanation: rather than mismatch, do these patterns reflect the benefits of specialization and trade in an industry featuring scale economies?

We find substantial interregional trade in medical services: one-fifth of Medicare production involves a doctor treating a patient from another region. Applying a trade model to these patient flows, we find that scale economies explain larger regions' production and export of higher-quality medical care. After quantifying regional economies of scale and patients' willingness to travel, we use the estimated model to explore counterfactual policies aimed at improving access for patients in smaller regions.

If medical services have increasing returns to scale, there are benefits to geographically concentrating production. Indeed, medicine has long been suggested as an industry in which the division of labor is limited by the extent of the market (Arrow, 1963; Baumgardner, 1988). Absent trade, however, the only way to serve patients in smaller regions would be to disperse production across space, foregoing the benefits of scale.[1] This is likely necessary for emergency care. But the vast majority of medical spending is not for emergencies. For example, if patients with cancer can travel across regions in search of the ideal oncologist—one specialized in their particular type of cancer, one with a better reputation, or simply a better personal match—the economic geography of medical care may resemble other tradable industries. In that case, society would face a *proximity-concentration tradeoff*: patients who

[1]Many economists assume trade costs for medical services are prohibitively high. Hsieh and Rossi-Hansberg (2021): "Producing many cups of coffee, retail services, or health services in the same location is of no value, since it is impractical to bring them to their final consumers." Jensen and Kletzer (2005): "Outside of education and healthcare occupations, the typical 'white-collar' occupation involves a potentially tradable activity." Bartik and Erickcek (2007): "An industry can bring in new dollars by selling its goods or services to persons or businesses from outside the local economy ('export-base production')...For health care institutions, demand for services tends to be more local."

consume medical services produced far from home incur trade costs yet benefit from the higher quality generated by geographically concentrating production.

We investigate these economic forces using Medicare claims data that we introduce in Section 1. Medicare is the federal government's insurance program for the elderly and disabled and spends four percent of US GDP. Medical service providers submit claims that report the specific medical procedure provided, the treatment location, and where the patient lives. Using hundreds of millions of these claims, we find that "imported" medical procedures—defined as a patient receiving a service provided by a physician in a different region—constitute about one-fifth of US healthcare consumption. Production is geographically concentrated in larger regions, while consumption is much less so. This contrast means exported medical services are disproportionately produced in larger regions and imports are a larger share of consumption in smaller regions. Although some patients travel thousands of kilometers for care, patient flows decline rapidly with distance between regions. While these patterns establish an important role for regional specialization and trade, they do not reveal the strength of regional increasing returns or the consequences of healthcare agglomeration.

To guide our investigation of these mechanisms, Section 2 develops a model of trade in medical services. We adapt standard models to a setting in which the government sets prices, so quality variation clears markets. The model delivers equations that we use to estimate regional service quality, the strength of increasing returns, and market-size effects. If there are regional scale economies, larger markets produce higher-quality care and export it. When economies of scale are sufficiently strong, the model predicts a *strong home-market effect*: greater demand makes larger regions *net* exporters of medical services. Because market size matters more at smaller scales, the model predicts less common medical procedures will respond more to differences in market size. The gravity model of trade flows provides a framework to test this hypothesis, as in Costinot et al. (2019).

In Section 3, we recover revealed-preference estimates of regional service quality by estimating patients' willingness to travel to each exporting region for medical services. These estimates based on trade flows align with conventional measures of quality: they are negatively related to external measures of mortality rates and positively related to other hospital rankings and safety scores.

We use these revealed-preference estimates to estimate the scale elasticity: how does serving more patients raise a region's quality of medical care? This parameter plays a key role in examining health policy when there are regional increasing returns. We find substantial regional increasing returns. We estimate the scale elasticity to be about 0.8: a region producing 10% more because of greater demand delivers about 8% higher quality.

The same trade flows show that regional increasing returns are large enough to explain the geographic patterns of production, consumption, and trade documented in Section 1. We find a strong home-market effect in medical services: greater demand induces a greater increase in exports than imports, making larger markets net exporters of medical care. These scale effects cannot be attributed to larger markets having lower input costs or medical production raising population size. This result remains when estimated in the cross-section, a panel, or instrumented with historical characteristics that predict current production levels.

Section 4 investigates potential sources of regional increasing returns in medical services. We examine how market-size effects and returns to scale vary with procedure characteristics. Rare procedures are traded more and over longer distances. For example, half of the patients having left ventricular assist devices (LVADs) implanted to restore their heart function come from outside the surgeon's region, while only 10% of screening colonoscopies are traded. Consistent with the model, the home-market effect is substantially stronger for less common procedures: a larger residential population drives a greater increase in net exports for rare procedures. Moreover, we estimate that rarer procedures exhibit stronger increasing returns.

A variety of mechanisms could generate these increasing returns to scale: finer specialization among physicians, sharing of lumpy capital equipment, knowledge diffusion, or learning by doing (Marshall, 1890). We show that larger markets have more specialized capital equipment and physicians. Critically, trade enables patients across regions to share in these benefits of scale: imports are more likely to be provided by a specialist—an appropriate and more experienced specialist—and more likely to use rare equipment than locally produced services.

We use our estimates of scale economies and trade costs to explore the proximity-concentration tradeoff quantitatively in Section 5. Counterfactual policies affect regions differently depending on their size and trade patterns. We examine the implications of increasing access to care in one region by either increasing reimbursements or reducing travel costs. Increasing reimbursements has a higher return in more populous regions. When raising reimbursements in the largest regions, the aggregate gain in patient market access per dollar of spending is about 10% higher than when increasing reimbursements in the smallest regions. Increasing reimbursements in one region reduces output quality in neighboring regions, while improving patients' market access to the extent they import from the treated region. Reducing travel costs for one region increases its import demand, which improves both output quality and market access in neighboring regions. The rich pattern of consequences when subsidizing patients in low-output regions highlights the importance of trade and agglomeration for understanding the incidence of these policies on patients and producers.

The higher-quality care available in larger markets may not benefit all patients equally. We find that patients with lower socioeconomic status (SES) are less likely to travel farther for better care, even when we examine travel patterns within the same billing code. Thus, the proximity-concentration tradeoff varies by SES; the pattern of aggregate returns masks important heterogeneity. We also find that the large geographic scope of the United States explains a meaningful part of the correlation between regional income and health.

This paper builds on research in urban, trade, and health economics. Urban economists have documented skill-biased agglomeration in production as knowledge workers have become more numerous and concentrated in skilled cities (Berry and Glaeser, 2005; Moretti, 2011; Diamond, 2016; Davis and Dingel, 2020; Eckert, Ganapati, and Walsh, 2020). Connecting this to the production and trade of services has been more difficult. Most studies of the geography of services analyze restaurants and retailers (Davis et al., 2019; Agarwal, Jensen, and Monte, 2020; Allen et al., 2021; Miyauchi, Nakajima, and Redding, 2021; Burstein, Lein, and Vogel, 2022). We show that—even in a service-based economy—the sizes of both local and potential export markets influence production and quality. This suggests that healthcare can serve as an export base for large markets (Bartik and Erickcek, 2007).

The trade literature has examined market-size effects in manufacturing but investigated services much less. Davis and Weinstein (2003), Hanson and Xiang (2004), and Bartelme et al. (2019) link manufactures' market size to export patterns, in line with the home-market effect of Krugman (1980) and Helpman and Krugman (1985). Dingel (2017) shows that market-size effects drive quality specialization across US cities. Market-size effects for pharmaceuticals have been estimated using demographic variation over time (Acemoglu and Linn, 2004) and across countries (Costinot et al., 2019). Services are much less studied, in part because of the paucity of reliable trade data (Lipsey, 2009; Muñoz, 2022). We advance this literature using the detailed procedure and location information in medical claims data.

The importance of medical care for health and welfare generates substantial public-policy interest. Rural locations have worse health outcomes but fewer doctors per capita. Newhouse et al. (1982a,b,c), Newhouse (1990), and Rosenthal, Zaslavsky, and Newhouse (2005) considered this issue and argued against targeting a uniform geographic distribution of physicians. Building on this, we measure interregional trade in medical services, estimate the impact of geography on patient access, and connect this trade to economies of scale. Modern trade theory guides our modeling, estimation strategy, and counterfactual policy analysis.

1 Empirical setting and geographic patterns

This section describes the data, documents how production and consumption of medical services vary with market size, and shows that trade between regions declines with distance.

1.1 Empirical setting

Our primary dataset is claims data from Medicare, the largest source of medical spending in the United States.[2] Medicare is the federal government's insurance program for the elderly and disabled. It does not directly employ physicians or run its own hospitals. Instead, it pays bills submitted by independent physicians, physician groups, hospitals, and other medical service providers. These bills—called "claims" in industry terminology—report the specific services provided using 5-digit codes from the Healthcare Common Procedure Coding System (HCPCS). There are over 12,000 distinct HCPCS codes, which identify individual procedures at a granular level.[3] In alternative analyses, we use groupings of patient *diagnoses* to account for potential substitution between treatments.

Federal regulation, not pricing decisions by physicians or hospitals, determines the payment for each claim. For physician payments, Medicare sets procedure-specific "reimbursement rates" largely independent of quality, quantity, or region.[4] We remove regional price differences from our spending measure as described in Appendix A.2. This produces an expenditure measure purged of spatial variation in reimbursement rates. Patients pay a share of these reimbursements through copayments and deductibles, but these cost-sharing rules are constant across regions, and most Medicare patients have a Medigap supplemental or Medicaid insurance that covers most or all of this cost sharing (Cabral and Mahoney, 2019).

Medicare claims report the ZIP codes of both the place of service and the patient's residence. Using these, we construct a trade matrix for medical services. We study all medical

[2]Appendix A.1 contains more details about our data and processing.

[3]Codes distinguish between, *e.g.*, flu vaccines that protect against three or four strains of flu, whether administration is intramuscular or intranasal, and patient ages. There are distinct codes for chest X-rays based on whether the images are of ribs, the breastbone, or the full chest, both sides or one side of the body, and the number of images taken (1, 2, 3, or 4+). By contrast, all of health services are aggregated into one of the nine service sectors available in Canadian data on interprovincial trade (Anderson, Milot, and Yotov, 2014) and the seventeen service sectors in international trade data (Borchert et al., 2021).

[4]While Medicare does have some quality incentive programs, the money at stake is a small share of Medicare's overall spending (Gupta, 2021). Medicare has some spatial variation in physician reimbursements, but it is not very large, and has diminished over time; details on payment rules are in Clemens and Gottlieb (2014). Hospital payments also reflect local wage indices, the hospital's occupational composition, and other factors. Gottlieb et al. (2010) show that price variation does not explain spatial variation in expenditure.

care provided by physicians, including both professional and facility charges, outside an emergency room or skilled nursing facility.[5] Because Medicare rarely reimbursed telehealth before 2020, this trade involves traveling to receive care delivered in-person.[6] We aggregate the ZIP-code-level information up to 306 hospital referral regions (HRRs), which are geographic units defined by the *Dartmouth Atlas of Health Care* to represent regional markets for tertiary medical care. HRRs are constructed by aggregating residential areas based on where patients were referred for major cardiovascular surgical procedures and for neurosurgery. Each HRR has at least one city where both categories of surgery were performed. Thus, the construction of these geographic units should tend to minimize trade between HRRs.[7] We construct HRR-to-HRR trade flows by interpreting the patient's residential HRR as the importing region and the service location's HRR as the exporting region.[8]

Physicians, hospitals, pharmacies, and other healthcare providers submit different types of claims. We use Traditional Medicare claims data, primarily from 2017, from hospitals, physicians, and outpatient care providers. One year of data includes around 360 million services, representing $230 billion in spending.[9] These claims are not perfectly representative of all US healthcare, since Medicare beneficiaries are elderly or disabled.[10] But the geographic distribution of Medicare beneficiaries is quite similar to the overall population, and Medicare alone finances one-fifth of medical spending. These data likely capture the key features

[5]Our results are similar when using only physicians' professional fees (Appendix Figure D.2 and Appendix Tables D.8 and D.16).

[6]In 2012, Medicare spent only $5 million—less than 0.001% of its expenditures—on telehealth services (Neufeld and Doarn, 2015), lagging other insurers (Dorsey and Topol, 2016).

[7]We have also used alternative geographies, including core-based statistical areas (CBSAs), metropolitan statistical areas (a subset of CBSAs that excludes the smaller micropolitan areas), and commuting zones (CZs). Because these yield consistent findings, we do not report all such estimates.

[8]The Medicare claims are US patients receiving care at US service facilities. These data do not report any international transactions. Throughout this paper, "imports" and "exports" refer to domestic transactions between regions of the United States.

[9]For one portion of spending (physician services), we observe a random 20% sample of these claims; our spending estimate is scaled up to represent the full Traditional Medicare population.

[10]We validate procedure frequencies in the 20% sample of physician services with two alternative data sources: more comprehensive Medicare data (based on all Traditional Medicare patients, but without information on patients' locations) and Health Care Cost Institute data for privately insured patients (Appendix A.3). We also study the geographic patterns of production by specialty using the physician registry data containing the ZIP code and specialty of all physicians registered to practice in the United States.

of overall healthcare production and consumption. Appendix Table D.1 reports summary statistics of HRR production, consumption, and trade volumes.

1.2 Spatial variation in production and consumption

Figure 1 shows maps of healthcare production and consumption across regions, based on the place of service and patient's residential address, respectively. The consumption map shows substantial variation that has been well documented by the *Dartmouth Atlas* and related literature on geographic variation in healthcare (Fisher et al., 2003*a*,*b*; Finkelstein, Gentzkow, and Williams, 2016). The production map shows even more pronounced variation: more production in large urban agglomerations and less in rural areas. There is substantial variation in production even between immediate neighbors, while consumption varies more smoothly.

The subsequent panels show patterns of trade. Nationally, 19% of production is exported to a patient in another HRR, and 21% is traded between CBSAs.[11] Panel 1(c) shows the ratio of production to consumption; a value larger than one means an HRR is a net exporter. Net-exporting regions tend to be major urban agglomerations, plus places such as Rochester, Minn. and Durham, N.C. that specialize in healthcare. Panel 1(d) shows gross exports as a share of local production for each HRR. Three-quarters of services produced in the Rochester metropolitan area, home to the top-rated Mayo Clinic, are provided to patients from other regions, who travel an average of 545 km to Rochester.[12] As a major healthcare exporter with a population of merely 220,000, Rochester is an outlier: larger regions are responsible for a disproportionate share of medical services production.

Figure 2 plots the average production and consumption per beneficiary across HRRs of different sizes. Production per beneficiary increases with population, with an elasticity of 0.05. Consumption per beneficiary is virtually uncorrelated with population; in fact, the

[11] 42% of physician production is exported outside the provider's Hospital Service Area (HSA), a smaller *Dartmouth Atlas* geographic unit. For comparison, 6% of employees work outside their commuting zone of residence (Monte, Redding, and Rossi-Hansberg, 2018, Appendix Table B.1). For manufactured goods, we obtain a 68% export share for metropolitan areas covered by the Commodity Flow Survey.

[12] These top hospitals tend to have top physicians; Mourot (2024) documents positive assortative matching.

estimated elasticity is slightly negative. The difference between production and consumption is net trade: larger markets are net exporters and smaller markets are net importers. Gross trade flows exceed net trade flows, with imports comprising about one-third of consumption in the smallest regions. Imports per beneficiary decline with an elasticity of -0.35 with respect to population. Exports per beneficiary are approximately flat, which means total exports are increasing with local population.

1.3 Bilateral trade and bilateral distance

Despite the clear patterns in Figure 2, geographic variation in trade is far from entirely explained by market size. The regions with the lowest export shares are Anchorage, Honolulu, and Yakima, Wash., likely reflecting their remoteness. To account for these geographic patterns, we examine bilateral trade flows.

Figure 3 depicts how trade varies with the distance between the patient and place of service. Panel (a) shows the distribution of distances patients travel for care, distinguishing between care provided in the patient's home region and in other regions. Within HRRs, there is a narrow distribution of distances that peaks around 10 km. When visiting providers in a different HRR, patients travel a great variety of distances. There is a local plateau between approximately 30–100 km, suggesting a fair amount of travel to nearby HRRs, perhaps indicating regional medical centers. There is substantial long-distance travel for care: many patients travel hundreds or thousands of kilometers.[13]

Panel 3(b) shows that trade declines with distance. The blue curve depicts trade volume against distance (for pairs of HRRs with positive trade flows) after removing fixed effects for each exporter and each importer.[14] This intensive-margin relationship is roughly log-linear. The red curve shows the extensive margin: the share of pairs with positive trade as a function

[13]Patients' choices to travel these distance underpin our revealed-preference estimates of regional service quality in Section 3.1. The average patient travels 640 km to Chicago and 740 km to New York City, compared with less than 146 km to Urbana-Champaign, Ill. or Charlottesville, Va. An older literature cited in Dranove and Satterthwaite (2000) finds that patients who travel farther incur higher hospital costs.

[14]This application of the Frisch-Waugh-Lovell theorem is only feasible for positive trade volumes.

of distance. This is 100% for nearby pairs and under 80% for the most distant pairs. These patterns motivate the inclusion of distance covariates in our gravity-based analysis.

Patients may vary in their ability or willingness to travel, especially by socioeconomic status. We quantify it here, to the extent feasible in our data, to enrich counterfactual analysis and interpret welfare implications. Panel 3(c) depicts distance elasticities—how fast bilateral trade declines with the distance between regions—estimated separately by neighborhood income decile.[15] We find a strong, nearly monotonic relationship between socioeconomic status and the distance elasticity: patients from the highest neighborhood-income decile exhibit a distance elasticity 25% smaller than those in the lowest decile.[16] This means patients from higher-income neighborhoods are more amenable to traveling farther for medical care. If society faces a proximity-concentration tradeoff, this tradeoff varies by socioeconomic status. This is especially notable given the empirical setting: Medicare insures the near-universe of elderly and disabled Americans.

2 Theoretical framework

This section develops a model of regional trade in medical services tailored to our analysis of US healthcare. We develop a competitive model of a market for one medical procedure that has a fixed price.[17] With fixed prices, quality variation plays a key role in clearing markets. Patients select quality-differentiated services and face trade costs, so patient flows between regions follow a gravity equation. Regional increasing returns cause the quality-adjusted cost of producing a service to decline with scale. The model delivers equations that allow us to estimate the strength of regional increasing returns and the home-market effect. We use

[15] We estimate these distance elasticities using equation (8) as described in Section 2.6.

[16] These estimates are consistent with the interaction that Silver and Zhang (2022) estimate between income and distance to care. These differences in distance elasticities are not driven by differences in the composition of procedures. When we estimate elasticities separately for rare and common services—or even for individual procedures (see Appendix Table D.5)—the income gradient of distance elasticities persists.

[17] For brevity, we present a competitive model, but the consequences of regional increasing returns for trade flows in a fixed-price environment do not hinge on this assumption. Appendix B.1 shows that a monopolistic-competition model with one medical provider in each region delivers the same predictions. As in flexible-price models, many market structures can give rise to a home-market effect (Costinot et al., 2019).

the estimated model to quantify outcomes in counterfactual policy scenarios.

2.1 Demand

We use a logit model of individuals choosing providers for a given service. Providers and patients are in regions indexed by i or j, with $\mathcal{I}$ denoting the set of regions. All providers in a region are identical. Let N_j denote the number of patients residing in region j who make a choice.[18] Patient k in region j choosing a provider in region i obtains utility

$$U_{ik} = \ln \delta_i + \ln \rho_{ij(k)} + \epsilon_{ik}.$$

The provider-region component δ_i would usually include a product's characteristics and price. With fixed prices, δ_i is simply the quality of care available in region i. The region-pair component ρ_{ij} represents bilateral inverse trade costs (proximity). The idiosyncratic component ϵ_{ik} is independently and identically drawn from a standard Gumbel distribution, so the probability that patient k selects a provider in region i is

$$\Pr(U_{ik} > U_{i'k} \ \forall i' \neq i) = \frac{\exp\left(\ln \delta_i + \ln \rho_{ij(k)}\right)}{\sum_{i' \in 0 \cup \mathcal{I}} \exp\left(\ln \delta_{i'} + \ln \rho_{i'j(k)}\right)}.$$

There is an outside option denoted by $i = 0$, which represents individuals choosing to forgo care, and we normalize its common component to zero, $\ln \delta_0 = \ln \rho_{0j(k)} = 0 \ \forall k$.[19]

This choice probability implies a gravity equation for the quantity of trade between any two regions when we aggregate patients' decisions. Let Q_{ij} denote the quantity of procedures supplied by providers in i to patients residing in j, and let Q_{0j} denote the number of patients in j selecting the outside option. Because each patient selects at most one

[18] Appendix B.2 extends the model to have multiple patient types.

[19] This formulation of demand is familiar from the hospital competition literature, which has been surveyed by, e.g., Gaynor and Town (2011) and Gaynor, Ho, and Town (2015). Our competitive model does not distinguish between hospitals or physicians within a region.

provider, $N_j = \sum_{i \in \mathcal{I} \cup \{0\}} Q_{ij}$. The demand by patients in j for procedures performed in i is

$$Q_{ij} = \delta_i \frac{N_j}{\Phi_j} \rho_{ij}, \tag{1}$$

where $\Phi_j \equiv \sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j}$ is the expected value of the choice set for patients in region j. We call this Φ_j "patient market access," as it captures the quality of care available to patients and their costs of accessing it. Equation (1) is a gravity equation with an origin i component, a destination j component, and an ij pair component. Total demand for care produced in i is

$$Q_i = \delta_i \sum_{j \in \mathcal{I}} \frac{N_j}{\Phi_j} \rho_{ij}. \tag{2}$$

2.2 Production

We assume competitive production of services with free entry and regional increasing returns that are external to individual providers. That is, each price-taking provider chooses its output quality and quantity given total regional production, an exogenous factor price, and an exogenous productivity shifter. A provider in region i that employs L units of the composite input to produce service of quality δ produces the following output quantity:

$$A_i \frac{H(Q_i)}{K(\delta)} L.$$

Improving quality is costly so $K(\delta)$ is increasing. Regional increasing returns to scale are a weakly increasing, concave function $H(Q_i)$ of total regional production, Q_i, which competitive producers take as given (Chipman, 1970). The regional productivity shifter A_i captures any other influences, such as past investments. Provider size L is indeterminate (and unimportant) given the linear production function, external economies of scale, and price-taking behavior. The composite input is supplied to region i at factor price w_i.[20] Thus, the unit

[20]If the regional factor supply were less than perfectly elastic, we would estimate increasing returns net of the cost of hiring additional inputs. That is, if the factor supply elasticity were β, our estimate of the scale elasticity α in equation (4) would instead be an estimate of the effective scale elasticity $\tilde{\alpha} \equiv \alpha - \frac{\beta}{1+\beta}$.

cost of producing quality δ in region i is

$$C(Q_i, \delta_i; w_i, A_i) \equiv \frac{w_i K(\delta_i)}{A_i H(Q_i)}.$$

In our setting, output prices are fixed: the Medicare-determined "reimbursement rate" $\overline{R}$ is independent of quality, quantity, and region. Each provider that produces output of the highest quality produced in region i earns revenue $\overline{R}$ per unit.

Provider optimization and free entry require the unit cost to equal the reimbursement rate in each region. Given factor price w_i and productivity shifter A_i, the free-entry condition

$$C(Q_i, \delta_i; w_i, A_i) = \overline{R} \tag{3}$$

defines a regional isocost curve: the set of quantity-quality combinations for which the average cost of production equals the reimbursement rate. Regional increasing returns make the isocost curve upward-sloping in (Q, δ) space. With free entry and fixed prices, the benefits of scale are realized as higher-quality services in higher-output regions.

While our assumptions thus far suffice for qualitative results, we later specify functional forms for additional predictions and empirical quantification; specifically, $K(\delta_i) = \delta_i$ and $H(Q_i) = Q_i^\alpha$, with a scale elasticity of $\alpha \geq 0$.[21] In this case, the free-entry condition (3) is

$$\overline{R} = \frac{w_i \delta_i}{A_i Q_i^\alpha}. \tag{4}$$

2.3 Equilibrium

Equilibrium equates supply and demand in each region, $Q_i = \sum_j Q_{ij}$. Given exogenous parameters $\overline{R}$, $\{w_i, A_i, N_i\}_{i \in \mathcal{I}}$, and $\{\rho_{ij}\}_{(i,j) \in (\mathcal{I}, \mathcal{I})}$, an equilibrium is a set of quantities and qualities $\{Q_i, \delta_i\}_{i \in \mathcal{I}}$ that simultaneously satisfy equations (2) and (3).

[21] There are increasing returns to scale if $\alpha > 0$. Note that this formulation is compatible with an equilibrium in which region i does not produce the procedure at all: $Q_i = 0 \implies \delta_i = 0$.

2.4 Scale effects in autarky

In autarky, patients choose whether to receive care, but they cannot travel between regions ($\rho_{ij} = 0$ for $i \notin \{0, j\}$). In this case, all demand is local and equation (2) simplifies to

$$Q_{jj} = \frac{\delta_j \rho_{jj}}{1 + \delta_j \rho_{jj}} N_j. \tag{5}$$

The autarkic equilibrium is at the intersection of the demand curve given by equation (5) and the free-entry isocost curve given by equation (3).[22] An increase in population size, $\Delta N_j > 0$, affects equilibrium outcomes by shifting the demand curve.

Figure 4 illustrates how greater demand affects quality in autarky. Panel (a) shows the role of returns to scale. The vertical axis shows quality δ_i and the horizontal axis shows quantity Q_i (on logarithmic scales). Higher quality attracts more patients, so demand is upward-sloping.[23] We show two cases of the free-entry isocost curve defined by equation (4): the horizontal line depicts constant returns ($\alpha = 0$) and the upward-sloping line depicts increasing returns ($\alpha > 0$). With constant returns, a rightward shift in demand ($\Delta N_j > 0$) causes a proportional increase in quantity produced and no change in output quality. With increasing returns, higher demand moves producers up the isocost curve. This quality improvement causes a more-than-proportional increase in quantity produced.

Panel 4(b) shows that an increase in demand raises quality more as the demand curve is increasingly elastic. The panel depicts two demand curves: the one on the left is more elastic, as we would expect for a less common procedure.[24] Shifting each demand curve to the right raises the equilibrium quality of each procedure because of increasing returns to scale. This market-size effect is larger for the less common procedure with more elastic

[22] For the equilibrium to be Marshallian stable, the demand curve must be steeper than the isocost curve at the intersection. There is a stable equilibrium because equation (5) means $Q_{jj} \to N_j$ as $\delta_j \to \infty$.

[23] For visual clarity, we draw a log-linear demand curve. The logit demand function (5) is in fact log-convex, which is consistent with all the comparative statics illustrated in Figure 4.

[24] The demand function (5) is log-convex, so demand is indeed more elastic at lower quality. This is a fixed-price counterpart of Marshall's second law that demand is more elastic at higher prices.

demand because the demand shift is amplified by a larger increase in quantity demanded.[25]

2.5 Market-size effects on trade flows

We now consider trade. With multiple regions and finite trade costs ($\rho_{ij} > 0$), some patients will import—*i.e.*, select a provider located in another region. This trade stems from two sources. First, in the logit demand system with finite trade costs, patients have idiosyncratic preferences that yield a strictly positive probability of choosing every region. Second, when quality varies, regions producing higher-quality services attract more patients.

Fixing the qualities produced in other regions, an increase in one region's demand affects its trade flows through three mechanisms. First, greater total demand for services proportionally increases a region's demand for imports through the N_j term in equation (1). Second, with increasing returns, an increase in N_i elicits an increase in quality δ_i, which raises region i's *gross* exports to each region. Costinot et al. (2019) call this the "weak home-market effect." Third, if increasing returns are sufficiently strong, quality δ_i improves so much that region i's patient market access Φ_i rises such that $\ln \delta_i$ rises more than $\ln (N_i/\Phi_i)$ does. That is, the increase in region i's gross exports exceeds any increase in its gross imports. This is the "strong home-market effect": an increase in local demand raises a region's *net* exports.

Figures 4(c) and 4(d) introduce trade and illustrate the distinction between weak and strong home-market effects.[26] Panel (c) depicts the quality and quantity produced in one region under two scale elasticities. Comparing points B and C, a given increase in demand elicits a larger quality improvement when increasing returns are stronger. Panel (d) depicts equilibrium exports and imports as a function of the region's demand shifter N_j. The import

[25] Alternatively, one could obtain this prediction by assuming that demand is log-linear and the isocost curve is log-concave, so that rare procedures have stronger scale economies. For example, introducing a fixed cost F would yield the isocost curve $\frac{w_i \delta_i}{A_i Q_i^{\alpha}} + F Q_i^{-1} = \bar{R}$. In this case, the scale elasticity $\frac{\partial \delta_i}{\partial Q_i}\frac{Q_i}{\delta_i} = \frac{\alpha \bar{R} + (1-\alpha) F Q_i^{-1}}{\bar{R} - F Q_i^{-1}}$ is decreasing in quantity produced Q_i for $\alpha \in (0,1)$. A rightward shift in demand would cause a larger (log) difference in quality for the low-volume procedure on the steeper part of the isocost curve.

[26] These diagrams are fixed-price analogues of Figures II and III in Costinot et al. (2019). See their discussion of the assumption that one region is large enough to affect its own quality but too small to affect the quality produced in other regions. This assumption is only made for this figure.

curves slope up because an increase in local demand raises demand for imports. The export curves slope up because of increasing returns: higher local demand increases quality, which increases gross exports. This is the weak home-market effect. When the scale elasticity α is larger—the free-entry isocost curve in Panel (c) is steeper—greater demand elicits a larger increase in output quality. This steepens the export curve and flattens the import curve in Panel (d). When the export curve is steeper than the import curve, there is a strong home-market effect: the increase in demand raises exports more than imports.

As in the autarkic case, we predict larger effects of market size for less common procedures. An increase in demand raises quality more when demand is more elastic, leading to a stronger home-market effect for rare procedures. If rare procedures also have greater economies of scale (higher α) that would amplify this contrast. This result motivates a research design comparing home-market effects for common and rare procedures.

These results also hold when an increase in demand in one region affects equilibrium outcomes in other regions. To demonstrate this, we examine the home-market effect in the neighborhood of a symmetric equilibrium. If all regions are the same size, $N_i = \bar{N}\ \forall i$, and trade costs are symmetric, $\rho_{ii} = 1$ and $\rho_{ij} = \rho \in (0,1)\ \forall i \notin \{0, j\}$, there is a symmetric equilibrium with quality $\bar{\delta}$ and patient market access $\bar{\Phi}$ in each region. As detailed in Appendix B.3, we totally differentiate the system of equations in terms of $\{\mathrm{d}\delta_i, \mathrm{d}N_i\}_{i=1}^{\mathcal{I}}$ and evaluate a change in one region's demand N_1 at the symmetric equilibrium.

With increasing returns of any magnitude, there is a weak home-market effect; with sufficiently strong increasing returns, there is a strong home-market effect. When $\alpha \in (0,1)$, an increase in region 1's demand elicits an increase its relative service quality:

$$\mathrm{d}\ln\delta_1 - \mathrm{d}\ln\delta_{j\neq 1} = \left[\frac{1-\alpha}{\alpha}\frac{(\bar{\Phi}-1)}{(1-\rho)\bar{\delta}} + \frac{(1-\rho)\bar{\delta}}{\bar{\Phi}}\right]^{-1} \mathrm{d}\ln N_1 > 0.$$

This higher quality causes region 1 to export more to every other region: $\frac{\mathrm{d}\ln Q_{1j}}{\mathrm{d}\ln N_1} > 0$. The effect on the region's net exports is

$$\mathrm{d}\ln Q_{1,j\neq 1} - \mathrm{d}\ln Q_{j\neq 1,1} = \left[\frac{1-\frac{1-\alpha}{\alpha}\frac{1+(\mathcal{I}-1)\rho}{1-\rho}}{\frac{1-\alpha}{\alpha}\frac{(1+(\mathcal{I}-1)\rho)}{(1-\rho)}+\frac{(1-\rho)\bar{\delta}}{1+(1+(\mathcal{I}-1)\rho)\bar{\delta}}}\right]\mathrm{d}\ln N_1. \tag{6}$$

Net exports increase with population size if and only if $\frac{\alpha}{1-\alpha} > \frac{1+(|\mathcal{I}|-1)\rho}{1-\rho}$. This occurs if increasing returns are sufficiently strong (α is large enough) and trade costs are sufficiently large (ρ is small enough). Otherwise, there is a weak home-market effect but not a strong one. Given a strong home-market effect, the effect in equation (6) is diminishing in the population of patients $\bar{N}$, so we predict a stronger home-market effect for rare procedures.

2.6 Estimating quality, scale economies, and home-market effects

The model yields equations for estimating quality δ_i, the scale elasticity α, and detecting home-market effects. Equation (1) provides a revealed-preference measure of each region's quality, δ_i. We assume the region-pair component is $\ln\rho_{ij} = \gamma X_{ij} + \upsilon_{ij}$, where X_{ij} is a vector of observed trade-cost shifters and υ_{ij} is an orthogonal unobserved component. Taking expectations and then logs of equation (1) yields gross bilateral trade flows:

$$\ln\mathbb{E}\left[S_{ij}\right] = \ln\delta_i + \ln\left(\frac{N_j}{\Phi_j}\right) + \gamma X_{ij}. \tag{7}$$

S_{ij} is the value of procedures exported from region i to patients residing in j.[27] The right side contains three terms: exporter quality $\ln\delta_i$, importer demand shifter $\ln\left(\frac{N_j}{\Phi_j}\right)$, and observed trade costs γX_{ij}. These trade costs may reflect the opportunity cost of patients' time, fiscal costs of travel, and any adverse health consequences of travel. Since our empirical work excludes emergency care, direct health costs are not likely to be critical.

To estimate each region's service quality, we replace the first two terms with exporter

[27]For most of the analysis, we define this as total value across all care, computed as described in Section 1.1. In some cases, we consider subsets of care or even individual procedures. In the procedure-specific case, the dependent variable is the procedure count Q_{ij} rather than spending S_{ij} because no aggregation is required.

fixed effects and importer fixed effects, respectively:

$$\ln \mathbb{E}\left[S_{ij}\right] = \underbrace{\ln \delta_i}_{\text{exporter FE}} + \underbrace{\ln \theta_j}_{\text{importer FE}} + \gamma X_{ij}. \tag{8}$$

The exporter fixed effects are a revealed-preference measure of regional quality, which we validate using external quality metrics. The importer fixed effects, combined with an assumption about the number of potential patients (see Appendix C.2), enable us to compute $\Phi_j = N_j/\theta_j$, a measure of patient market access for those who reside in location j.

The relationship between quality and quantity reveals the scale elasticity of the regional production function, α. Per the free-entry condition (4), quality is an isoelastic function of the quantity produced, conditional on price, cost, and productivity shifters. We take the log of (4), replace $\ln \delta_i$ with its estimate $\widehat{\ln \delta_i}$ from (8), and rearrange terms to obtain the estimating equation:[28]

$$\ln \delta_i = \alpha \ln Q_i + \ln \overline{R} - \ln w_i + \ln A_i. \tag{9}$$

To learn whether these regional increasing returns explain the pattern of trade described in Section 1, we estimate a more parsimonious gravity regression that uses each region's population as a demand shifter. Following Costinot et al. (2019), we differentiate the system of equations (2) and (3) around the symmetric equilibrium and replace $\ln \delta_i$ and $\ln\left(\frac{N_j}{\Phi_j}\right)$ in (7) by log population in the producing and consuming regions, respectively, yielding:

$$\ln \mathbb{E}\left[S_{ij}\right] = \lambda_X \ln \text{population}_i + \lambda_M \ln \text{population}_j + \gamma X_{ij}. \tag{10}$$

A positive coefficient $\lambda_X > 0$ implies a weak home-market effect as defined in Costinot et al. (2019): *gross* exports increase with market size. If $\lambda_X > \lambda_M$, the home-market effect is strong: *net* exports increase with market size.

[28] Appendix A.5 quantifies the potential bias resulting from our observing only the quantity produced for Traditional Medicare beneficiaries, rather than the total quantity produced for all patients. It shows that the bias is small: the estimates in Table 1 should be deflated by about 5%.

To estimate these specifications, we must parameterize observed trade costs γX_{ij}. We first use log distance and a same-region dummy, *i.e.* $\gamma X_{ij} = \gamma_1 \ln \text{distance}_{ij} + \gamma_0 \mathbf{1}(i = j)$. Alternative specifications add $(\ln \text{distance}_{ij})^2$ or replace these continuous distance covariates with indicators for distance deciles. Since observed bilateral trade is zero for many pairs of regions, especially when looking at trade in individual procedures, we use the Poisson pseudo-maximum-likelihood (PPML) estimator (Santos Silva and Tenreyro, 2006).

In other analyses, we estimate slight variants of these specifications. First, we leverage population growth over time using a panel regression with a region-pair fixed effect:

$$\ln \mathbb{E}\left[S_{ijt}\right] = \lambda_X \ln \text{population}_{it} + \lambda_M \ln \text{population}_{jt} + \phi_{ij} + \gamma_t X_{ij}, \tag{11}$$

where S_{ijt} denotes the gross bilateral trade flow in year t. In this specification, changes in population across time periods t identify λ_X and λ_M. Second, we contrast common and rare services by dividing procedures into two groups based on whether their national volume is above or below that of the median procedure.[29] We estimate equation (8) separately for these two groups to obtain the qualities of rare and common procedures. Let S_{ijc} denote exports from i to j in category of care $c \in \{\text{common}, \text{rare}\}$. The model predicts a stronger home-market effect for rare procedures, so we estimate the following specification:

$$\begin{aligned} \ln \mathbb{E}\left[S_{ijc}\right] = {} & \lambda_X \ln \text{population}_i + \lambda_M \ln \text{population}_j + \gamma X_{ij} \\ & + \left(\mu_X \ln \text{population}_i + \mu_M \ln \text{population}_j + \psi X_{ij}\right) \cdot \mathbf{1}(c = \text{rare}). \end{aligned} \tag{12}$$

In the presence of an overall strong home-market effect, theory predicts $\mu_X > \mu_M$. An alternate specification introduces ij-pair fixed effects, which absorb all the covariates not interacted with $\mathbf{1}(c = \text{rare})$. We also estimate similar models with heterogeneity in other dimensions, such as procedure intensity and frequency of patient engagement, and for specific procedures.

[29] Appendix Figure D.1(b) plots the distribution of import shares across regions for the two groups.

3 Regional increasing returns in medical services

This section tests for increasing returns in healthcare and the implications for trade flows. Section 3.1 estimates region-level quality of medical care and shows that external quality indicators line up with revealed-preference measures based on trade flows. Using these measures, Section 3.2 finds substantial regional increasing returns in healthcare production. Section 3.3 demonstrates a strong home-market effect, implying that scale economies indeed generate the observed patterns of production and trade.

3.1 Quality estimates

Our revealed-preference measures of regional service quality are the exporter fixed effects $\widehat{\ln \delta_i}$ from estimating equation (8). Figure 5 relates these fixed effects to external measures of regional hospital quality. We aggregate estimated hospital mortality from Chandra, Dalton, and Staiger (2023) and the Centers for Medicare and Medicaid Services (CMS) by HRR. We also count the number of times each region's hospitals appear on *U.S. News* Best Hospitals.[30] Panels 5(a)–(c) establish that patients travel farther to obtain care from regions with lower hospital mortality rates or better *U.S. News* rankings. Appendix Table D.2 reports the corresponding regressions. It also shows a regression relating the exporter fixed effects simultaneously to a variety of other hospital safety, mortality, accreditation, and other quality measures. This regression has an adjusted R^2 of 0.60. While our revealed-preference estimates capture all attributes that influence patient choices, their relationship to these other quality measures suggest that our estimates capture many characteristics of clinical interest.

[30] Compared with the CMS measures, Chandra, Dalton, and Staiger (2023) use empirical Bayes estimation to account for differences in hospital volume and quality drift over time. Appendix A.1 explains how we use the *U.S. News* rankings. Appendix Figure D.3 also presents results using Hospital Safety Grades from the Leapfrog Group.

3.2 Estimating the scale elasticity

3.2.1 Empirical implementation

We use equation (9) to estimate the strength of regional increasing returns. Regional output quality depends on the volume of production and exogenous shifters of the isocost curve. One potential concern with estimating equation (9) by ordinary least squares is reverse causality. Shifts of the isocost curve would cause movements along the upward-sloping demand curve, biasing the estimated scale elasticity upwards. We address this by using three instruments for production and differences over time.

Our first instrument is current population. Population is relevant for healthcare output, since larger populations clearly require more healthcare.[31] The instrument is valid if it shifts only demand and not supply. We show that urban amenities do not make healthcare labor inputs cheaper in larger markets.

Nevertheless, one potential concern with this instrument is reverse causality. Suppose that success in exporting medical services serves as an employment base that raises current population size, as epitomized by "anchor institutions."

We use two further instruments to address this concern. Our second instrument is historical population. In 1940, medicine was a far smaller industry and did not drive local population in the way it might today. Because local populations persist, population in 1940 predicts current population, so we use the former as an instrument for the latter.

The next instrument goes farther back than 1940 and uses local geology to predict healthcare production. Rosenthal and Strange (2008) and Levy and Moscona (2020) show that shallower subterranean bedrock makes construction easier, increasing population density. Bedrock depth also predicts population size, so it is our third instrument for local demand.[32]

[31]The variation in demand predicted by population is overwhelmingly driven by headcount, not income per capita: the population elasticity of income per capita is below 0.1, and the income elasticity of health spending is likely below 1.0 (Acemoglu, Finkelstein, and Notowidigdo, 2013).

[32]This instrument is currently only available for CBSAs and CZs, not HRRs. Our main results hold when defining regions as CBSAs and CZs. Levy and Moscona (2020) show that the bedrock instrument has ample first-stage power for predicting CBSA population density; the same is true for our endogenous variables (population levels).

Our final research design uses variation over time. We estimate equation (9) in first differences using changes from 2013 to 2017. This uses the relationship between quality growth and production growth across regions to estimate the scale elasticity. To similarly confound both the cross-sectional and first-difference regressions, omitted supply shifters would have to exhibit similar cross-sectional and temporal variation.

3.2.2 Scale improves quality

Estimated service quality $\widehat{\ln \delta_i}$ rises substantially with the regional production volume $\ln Q_i$. Figure 5(d) depicts this relationship, which appears consistent with our isoelastic specification. Table 1 reports regression estimates. The estimated scale elasticity is around 0.8 and stable under various estimation approaches. The first row shows our baseline estimate for the cross-section, and the second uses variation over time. The third and fourth rows instrument for output using contemporaneous or historical population, which exhibit strong first-stage regressions. The second column omits the diagonal S_{ii} observations when estimating the gravity equation (8), to avoid any bias from having a region's own local consumption influence both the quality measures and output. The third column controls for (small) spatial variation in reimbursements, mean two-bedroom property value, and mean annual earnings for non-healthcare workers. Instrumenting for output tends to reduce the estimated scale elasticity. Excluding the diagonal of the trade matrix when estimating quality tends to raise it. Across all 12 estimates, the lowest elasticity is 0.63 and the highest is 1.04.[33] The results for CBSAs and CZs, reported in Appendix Tables D.3 and D.4, are also stable across specifications and when using the bedrock instrument.

The rest of this paper demonstrates the economic implications of these scale economies in three ways. First, Section 3.3 shows that they are large enough to generate the observed patterns of production and trade in medical services. Second, Section 4 demonstrates mech-

[33] These estimates lie in the middle of other estimated agglomeration elasticities. Kline and Moretti (2013) estimate an elasticity of 0.4–0.47 from the Tennessee Valley Authority's investments. In manufacturing, Greenstone, Hornbeck, and Moretti (2010) report an analogous elasticity above 1 (a 12% increase in total factor productivity caused by adding a plant representing 8.6% of the county's prior output).

anisms that contribute to regional increasing returns. Third, Section 5 uses the parameter estimates to analyze counterfactual scenarios that illustrate how these forces work and provide guidance for healthcare policy in a world with increasing returns.

3.3 A strong home-market effect in medical services

While we have established there are regional increasing returns, we have yet to see if they are sufficiently strong to explain the fact that larger markets are net exporters. This question determines whether market size enables larger regions to specialize in healthcare production and to share the benefits of scale with patients in other regions. We now test for a strong home-market effect in medical services by estimating equation (10).

Table 2 presents the results. The first column shows significant, positive coefficients on both provider- and patient-market population.[34] The coefficient on provider-market population is substantially greater than that on patient-market population. This demonstrates a *strong* home-market effect (Costinot et al., 2019). Not only does a larger population increase gross exports, but it does so more than it increases gross imports by local patients.

The distance elasticity of medical services trade between hospital referral regions is -1.6. This is substantially larger than the distance elasticity of -0.95 estimated for trade in manufactures between CBSAs (Dingel, 2017).[35] This suggests that trade in personal services incurs greater distance-related costs, relative to the degree of product differentiation across regions, than trade in manufactured goods. The most obvious difference is that patients themselves must travel to the provider.

The next two columns of Table 2 demonstrate that more flexible distance-covariate specifications do not meaningfully alter the estimated home-market effect. Column 2 introduces the square of log distance as an additional covariate. Column 3 replaces the parametric distance controls with dummies for deciles of distance. The result is stable across the

[34] In all estimates using trade matrices, we two-way cluster standard errors by patient and provider market.

[35] We find a distance elasticity of medical services trade between CBSAs of -2.4. The analogous elasticity of health care and social assistance services trade between Canadian provinces is -1.42 (Anderson, Milot, and Yotov, 2014). The distance elasticity of international trade is typically near -0.9 (Disdier and Head, 2008).

columns: both gross and net exports increase with market size.

The fourth column of Table 2 uses the historical population instrument to address concerns about reverse causality. We obtain similar home-market-effect estimates to our baseline results. Appendix Tables D.6 and D.7 report similar results when using CBSAs or CZs rather than HRRs as our geographic unit. Appendix Tables D.6 and D.7 also show the results are robust to instrumenting with historical population and with bedrock depth.

The final column of Table 2 uses changes in population over time to proxy for changes in demand, per equation (11). The market-pair fixed effects in this specification absorb all cross-sectional variation. We find a strong home-market effect that is similar to the cross-sectional estimate.

Competing explanations. Our instruments address reverse causality, but all of them operate through population size. So other channels could generate relationships similar to the market-size effect we estimate. Most significantly, if doctors prefer to live in big cities (Lee, 2010), as college graduates generally do (Diamond, 2016), they could accept lower nominal wages and thus reduce healthcare production costs w_i in such cities. This would raise quality in large markets, but through a different mechanism than increasing returns.

Before we address this problem, first note what is *not* a problem: physicians preferring to work in larger regions for job-related reasons. If a region's scale enables it to support large academic medical centers, which attract workers, this is an agglomeration benefit that the scale elasticity α ought to capture.[36] More patients enable physicians to specialize, conduct research, and train medical students. These forces operate through the scale of healthcare production in the region, and academic medical centers are part of this production function.

The challenge to our interpretation arises if physicians prefer larger markets for non-

[36]The most salient example is Cornell University: after an abortive attempt to have medical training in both Ithaca and New York City, the Cornell Trustees quickly closed down the Ithaca location and centered the medical school in New York—where the patients and doctors were more abundant—in the early 20th century (Flexner, 1910; Gotto and Moon, 2016). As this history illustrates, the potential local demand for care can drive the location of medical training. In general education, in contrast, university placement induces economic growth (Moretti, 2004).

medical reasons, and this labor supply shift increases quality or reduces costs in larger markets. We investigate whether this mechanism is strong enough to reduce net costs in larger markets. Doctors are cheaper in larger markets (Gottlieb et al., 2023), but other costs rise with population size. Appendix Figure D.4 shows that the population elasticity of doctors' earnings is -0.01, but that for non-physicians is 0.043.[37] To compute the population elasticity of overall labor costs, note that non-physician labor's share of healthcare production is twice the physician labor's share (Appendix Table D.9). The population elasticity of labor costs is thus positive. The higher cost of real estate in larger markets reinforces these higher labor costs. This spatial variation undermines the idea that amenities make production cheaper in larger markets. Section 4 further shows that specialty-specific income elasticities are close to zero on average and unrelated to the number of specialists.

One final concern is measurement error in Medicare's records of patients' residences. To address this, Appendix A.4 first demonstrates our results are robust to excluding states with large seasonal populations. Second, we examine how far dialysis patients appear to travel. We find that residential measurement error is limited and does not drive our results.

4 Mechanisms

This section explores potential sources of scale economies and asks if trade expands the geographic scope of these mechanisms. We focus on professional fees for which payments are made at the service level.[38] Section 4.1 examines how market-size effects vary with procedure characteristics, especially the rare-vs-common comparisons suggested by the model. Consistent with our theoretical predictions, rare services exhibit stronger market-size effects. They also exhibit stronger increasing returns to scale. Section 4.2 shows that division of labor and lumpy capital may be substantial sources of regional increasing returns for medical services. Section 4.3 directly connects trade to these mechanisms: imported care is more likely to be

[37]Appendix A.1 discusses subtleties of the income data.

[38]We observe 210 million claim lines representing $89 billion in spending on professional services.

performed by specialists, by appropriate specialists, by more experienced physicians, and using scarce capital equipment, especially for smaller regions' imports. Trade thus expands the population benefiting from specialization.

4.1 How market-size effects vary with procedure characteristics

4.1.1 Spatial variation in production and consumption by frequency

Larger regions produce a wider variety of procedures. In Figure 6(a), we count the number of procedures produced in each region and graph these counts within seven procedure categories against regional population. Larger regions produce a greater variety of care in all seven categories. Panel (b) shows that this is true for consumption, but to a lesser extent: Aggregating across all seven categories, HRRs in the top decile of population produce 3 times, but consume only 2 times, as many unique procedures as those in the bottom decile.

To better understand the characteristics of the procedures performed in larger markets, we estimate each procedure's population elasticity of production per Medicare beneficiary.[39] Let Q_{pi} denote the count of procedure p produced in region i. Let M_i denote the number of Medicare beneficiaries residing in i. For each procedure p, we estimate the following relationship across regions:

$$\ln \mathbb{E}\left[\frac{Q_{pi}}{M_i}\right] = \zeta_p + \beta_p \ln \text{population}_i. \tag{13}$$

The estimated elasticity, $\hat{\beta}_p$, describes how production varies with market size, and we estimate it using Poisson pseudo-maximum-likelihood.[40] If the quantity produced were simply proportional to population, β_p would be zero. Our model suggests that scale effects play a larger role for rare procedures. It predicts less common services will have higher population

[39]Davis and Dingel (2020) relate population elasticities to other measures of geographic concentration, such as location quotients, and estimate population elasticities of employment for various skills and sectors.

[40]In a robustness check, we have also estimated a zero-inflated Poisson model, to account for the possibility that fixed costs are especially important for the decision of whether to provide the first instance of a service in a region. These results (not reported here) are quite similar.

elasticities of production.

Production per beneficiary indeed rises with market size, especially for less common procedures. The blue dots in Figure 6(c) relates the population elasticity of production per beneficiary $\hat{\beta}_p$ to the procedure's national volume, $Q_p = \sum_i Q_{pi}$. Across values of Q_p, procedure output per beneficiary increases with market size. Less common procedures have higher elasticities, consistent with stronger economies of scale in rare procedures.

This finding raises questions about patients' access to care. What happens to patients who live in smaller markets but need rare services? To investigate this, we separately estimate the population elasticity of *consumption* per beneficiary for each procedure. Let G_{pi} denote the count of procedure p consumed by patients *residing* in region i. We then estimate an analogue of equation (13) where we replace the dependent variable with $\ln \mathbb{E}\left[G_{pi}/M_i\right]$, and denote by β_p^C the resulting coefficient. If $\beta_p^C \neq \beta_p$, there is size-predicted net trade in procedure p.

The population elasticity of consumption per beneficiary is smaller for the vast majority of procedures and less steeply related to a procedure's national frequency. The red squares in Figure 6(c) plot the population elasticity of consumption per beneficiary $\hat{\beta}_p^C$ of a procedure against its national volume $\ln Q_p$. While the relationship is negative, the slope for consumption is only one-third that for production. Appendix Table D.10 reports the production, consumption, and trade patterns for two exemplar procedures: screening colonoscopy and LVAD implantation. Colonoscopies are common and geographically dispersed, while LVAD procedures are rare, geographically concentrated, and traded over longer distances.

We have thus far modeled patients as demanding (and providers as producing) specific service codes. An alternative view is that patients have a particular medical condition that requires treatment, but the patients may not know what particular care they need; they simply know they require care. As physicians might use different treatments across regions for the same condition, our procedure-level results could reflect substitution across procedures. We address this by conducting a similar analysis across diagnoses.

Figure 6(d) shows production and consumption elasticities by diagnosis, rather than by procedure. The key patterns remain similar: production elasticities are higher, and decline more rapidly with national patient volume, than consumption elasticities. Both sets of elasticities have less steep relationships with national volume than for procedures. This could reflect measurement error within each category: the 482 diagnosis categories we use are far coarser than the 8,210 procedures in Figure 6(c). Alternatively, it could indicate true substitution among procedures within a condition that varies with location.

The contrasting population elasticities of production and consumption summarized in Figure 6 imply trade in medical services between markets of different sizes. Just as theories of trade with scale effects would predict, larger markets export rare procedures and smaller markets import them. For almost all procedures, production increases more than proportionately with market size. Consumption also increases more than proportionately with market size, but much less so than production. The differences between these elasticities mean net exports vary with market size. The implied net trade between markets of different sizes is particularly large for procedures that have small national volumes.

4.1.2 Market-size effects are stronger for rare and resource-intensive procedures

We now use trade flows to explicitly understand where rare medical services are provided and why. Since Section 3.3 found a strong home-market effect overall, the model predicts that this should be especially true for rare services. Trade is indeed more prevalent in rare services (Appendix Figure D.1). We now use the specific pattern of these trade flows to test the model's prediction that scale economies drive the production and trade of rare services.

Table 3 reports gravity regressions in which each pair of locations has two observations: one for rare services and one for common. Column 1 repeats our baseline regression from Table 2 for professional fees and reports similar results. Column 2 limits the sample to pairs of location that have positive trade in at least one of the two procedure groups, which is the estimation sample used in the remainder of the table. We then estimate equation (12), interacting both provider-market and patient-market population with an indicator for rare services.

The home-market effect is stronger for rare services. In column 3 of Table 3, the coefficient on provider-market population increases by about 60% relative to common services. The coefficient on patient-market population shrinks by more than half. Column 4 introduces location-pair fixed effects. Columns 5 and 6 are analogues of the previous two, but add a quadratic log-distance term. These results are statistically indistinguishable from the previous columns. Columns 7 and 8 demonstrate that our result holds when we look across diagnoses rather than procedures. As with the production and consumption elasticities in Figure 6(d), the magnitude of the difference between rare and common care shrinks. This could reflect substitution across care within a diagnosis or a less precise classification of diagnoses than of procedures. But the qualitative pattern holds and remains significant, consistent with the model's prediction.

Appendix Table D.11 shows that these results are robust to using our instruments for market size. Columns 1 and 2 show estimates for common and rare services, respectively, when instrumenting for population in each region by its 1940 population. Columns 3 and 4 repeat the exercise using CBSAs rather than HRRs, and columns 5 and 6 switch to the bedrock-depth instrument. The results are consistent regardless of geographic unit or instrument.[41] This stability suggests that neither the aggregate result nor the variation with procedure frequency is driven by anchor institutions or similar omitted variables.

The finding that less common procedures exhibit stronger home-market effects is robust to different ways of defining rare and common care. Appendix Figure D.5 shows that the pattern holds when splitting by deciles of national frequency. Appendix Table D.13 shows the same pattern among illustrative procedures.

The home-market effect is also stronger for rare procedures when controlling for how often an individual patient receives the same procedure, which we call a procedure's "engagement".

[41] Appendix Table D.12 reports analogous results based on commuting zones.

If patients are less willing to travel for high-engagement services and these services are more common, higher engagement could drive the stronger home-market effect we observe for rare procedures.[42] Appendix Table D.14 shows that patients are indeed more sensitive to distance for high-engagement procedures, but controlling for this heterogeneity does not meaningfully alter the estimated differential impacts of population size for rare procedures.

The model predicts larger home-market effects for rare services when they have more elastic demand or larger scale economies. Fixed costs are one reason the scale elasticity may be larger for rarer services (see footnote 25). When estimating equation (9) separately by procedure frequency, we find that scale elasticity is substantially larger for rare services. It is near 1.0, as shown in the second panel of Appendix Table D.16.

Some procedures require more capital and specialized knowledge, and the consequences of market size for these resource-intensive procedures may illuminate sources of increasing returns. Relative value units (RVUs) capture the extent of physician work, expertise, and additional resources required to perform the procedure and determine the procedure's Medicare payment. We divide procedures into terciles by intensity as measured by the total RVUs CMS assigns to a procedure. We find that the home-market effect is increasing in procedure intensity (Appendix Table D.15). This motivates our investigation of physician specialization and equipment usage in Section 4.2.

The potential concern about omitted cost shifters from Section 3.3 has an analogue here: Do the doctors who provide rare services benefit more from urban amenities than those providing common ones, lowering the cost of producing rare services in larger markets? This has facial plausibility if rare services are produced by elite specialists who earn more and might be more willing to pay for urban amenities through lower compensation.

Examining the population elasticities of physician earnings for each specialty alleviates this concern. If urban amenities drive specialists' locations, earnings elasticities should be

[42] In fact, the national frequency of a service has a very low correlation with various measures of engagement for that service, so it does not confound this result. The correlation between the share of patients who had more than one claim for the procedure in a given year and the procedure's frequency is 0.14.

negative, especially for rare specialties. But Figure 7(a) shows that the income elasticities are close to zero on average and uncorrelated with the specialty's national abundance. However urban amenities affect physicians' choices, these choices do not exhibit the compensating differentials necessary to explain the relationship between market size and specialization.

4.2 Mechanisms for strong regional increasing returns

This section shows that division of labor and lumpy capital are mechanisms behind regional increasing returns to scale in medical services. This evidence does not preclude other agglomeration mechanisms from also playing a role. Knowledge diffusion (Baicker and Chandra, 2010) and thicker input markets could also be important productivity benefits of scale. We focus on specialization, physician experience, and equipment use because they can be measured using claims data and are closely related to procedure-level returns to scale.

4.2.1 Scale facilitates the division of labor

One source of increasing returns could be division of labor among physicians. In particular, the specialized labor required to produce rare services could drive the patterns we found across treatments and diagnoses. Specialized services may require physicians with specific training, whose presence may require high demand (Dranove, Shanley, and Simon, 1992).

To study this mechanism, we estimate the population elasticity of physicians per capita for each specialty and relate it to the number of physicians in the specialty. Let Y_{si} be the number of doctors of specialty s in location i.[43] We estimate a Poisson model,

$$\ln \mathbb{E}\left[\frac{Y_{si}}{\text{population}_i}\right] = \zeta_s^S + \beta_s^S \ln \text{population}_i, \tag{14}$$

for each specialty s by maximum likelihood.

Figure 7(b) shows a clear negative relationship between a specialty's per capita population

[43] Data are from the National Plan and Provider Enumeration System (NPPES); see Appendix A.1.

elasticity $\hat{\beta}_s^S$ and the national number of physicians in that specialty.[44] A natural explanation for rare procedures and rare specialties both being geographically concentrated in larger regions is that the size of the market limits the division of labor. To the extent that producing rare procedures requires specialized physicians, a larger volume of patients makes production economically viable.

4.2.2 Lumpy capital as a source of regional increasing returns

Another potential source of increasing returns is lumpy capital, namely expensive medical equipment. For instance, if a piece of medical equipment is rarely used in procedures, small markets may not use it enough to justify the investment.

We provide evidence consistent with this hypothesis using a CMS dataset that enumerates the types of equipment used by each medical procedure (HCPCS code).[45] We measure equipment use by linking this dataset to procedure-level production data. The frequency of equipment use H_{di} for a piece of equipment d in region i is defined as the sum of the procedure volume in region i across all procedures that use equipment d. We use a Poisson model to estimate the population elasticity of per-capita equipment use:

$$\ln \mathbb{E}\left[\frac{H_{di}}{\text{population}_i}\right] = \zeta_d^K + \beta_d^K \ln \text{population}_i. \tag{15}$$

Figure 7(c) plots the population elasticity of per-capita equipment use β_d^K against log national use, $\ln H_d$. The strong negative correlation indicates that equipment used less often tends to concentrate in larger markets. Similar to the patterns for physician specializations, rare capital equipment is disproportionately employed in larger regions.

[44] This pattern is not attributable to spatial sorting driven by rare specialties commanding higher earnings. In fact, a specialty's number of physicians and mean earnings are uncorrelated. Appendix Table D.17 shows that controlling for a specialty's earnings has no effect on the negative relationship between population elasticity and number of physicians across specialties.

[45] See Appendix A.1 for details.

4.3 Travel to access specialized services

We next ask whether the distribution of specialist physicians helps explain trade. Figure 8(a) shows the share of imports and of locally produced consumption that are provided by specialists as a function of regional population.[46] Imports are significantly more specialist-intensive than local production. This difference is especially pronounced in the smallest regions, and it remains true throughout the population distribution.

Does trade match patients with the appropriate specialist? Among all specialty care, we determine the two most common specialties to provide each unique service and label these the "standard" specialties for that care. We then determine whether each instance of the treatment was provided by a standard or non-standard specialist.

Figure 8(b) shows the share of imports and of locally produced care provided by non-standard specialties. Imports are less likely to come from non-standard specialists than local care, and the distinction is especially pronounced in the smallest regions. The difference is substantial: Local care in the smallest regions is 40% more likely to be provided by a non-standard specialist than in the largest regions (7% vs. 5%). When importing medical services, this share falls to 5%—indistinguishable from the largest regions' locally produced care.

We conduct a similar analysis based on provider experience. Using the public Medicare provider data (based on all Traditional Medicare patients), we count the number of times the physician billed for the specific service in the previous year. We divide this experience measure by the procedure's national mean and average it across all procedures provided to patients in an HRR. We then rescale HRR-level experience by the mean across HRRs. Figure 8(c) shows that, at all population sizes, care imported from other regions is produced by more experienced providers than locally produced care.[47] Patients in larger regions see

[46] We define "specialist" to mean all physicians except those whose primary specialty is internal medicine, general practice, or family practice.

[47] This comparison restricts attention to the 143 procedures performed in all HRRs. Thus, regional variation does not reflect the fact that larger markets produce a greater number of distinct codes (Figure 6(a)).

more experienced providers for both imported and locally produced care.

Specialists are disproportionately located in larger markets, as are physicians with more experience in their procedures. Since imported care is predominantly specialized and provides patients access to higher experience, we conclude that visiting the appropriate specialist based on training or experience is part of the value proposition for trade in medical care.

Figure 8(d) shows analogous patterns for rare capital equipment. We measure the share of care that requires a piece of rare capital equipment and plot these shares separately for local and imported care against regional population. While locally produced care in larger markets is more likely to use rare equipment, imported care has a higher rare-equipment share throughout the population distribution.

This analysis also provides a second validation of our interpretation that trade reflects quality variation. Patients travel to regions with highly-ranked hospitals, which larger markets tend to have—along with rare equipment and the ability to provide rare services. This market-size effect strongly predicts gross and net exports. Together, this suggests that economies of scale play an important role in increasing the quality of care, and trade between regions enables patients from many regions to share the benefits of this agglomeration.

5 Health policy with trade and increasing returns

Given the estimated strength of regional increasing returns, geographically concentrating healthcare production has substantial benefits. Larger regions support more specialists, physician experience, specialized equipment, and procedures. But geographic concentration implies that patients in smaller regions may have limited access to care. We use our estimates of the scale elasticity α, region-specific qualities δ_i, and observed trade flows to quantify how various counterfactual policy scenarios would change each region's patient market access for non-emergency care.[48] Our results underline the importance of distinguishing between the

[48]Because our estimation sample excludes emergency services, our estimates and counterfactual scenarios omit any complementarities between emergency and non-emergency care. Our partial-equilibrium model

quality of locally produced services and the quality of services to which local residents have access. They also show how geography may contribute to striking patterns of inequality in Americans' health.

5.1 Method to compute counterfactual outcomes

We compute counterfactual equilibrium outcomes relative to the baseline equilibrium. For the baseline equilibrium, define export shares $x_{ij} \equiv \frac{Q_{ij}}{\sum_{j'} Q_{ij'}}$ and import shares $m_{ij} \equiv \frac{Q_{ij}}{N_j}$. For every variable or parameter y, denote the ratio of its counterfactual value y' to its baseline value y by $\hat{y} \equiv \frac{y'}{y}$. Appendix C.1 shows how we solve for the relative counterfactual endogenous qualities $(\hat{\delta})$ using baseline equilibrium shares (x_{ij}, m_{ij}), the scale elasticity (α), and relative counterfactual exogenous parameters $(\hat{A}, \hat{R}, \hat{w}, \hat{\rho}, \hat{N})$. In particular, counterfactual qualities are given by a system of $\mathcal{I}$ equations with unknowns $\{\hat{\delta}_i\}_{=1}^{\mathcal{I}}$:

$$\hat{\delta}_i = \left(\hat{R}_i \hat{A}_i / \hat{w}_i\right)^{\frac{1}{1-\alpha}} \left(\sum_{j \in \mathcal{I}} \frac{x_{ij} \hat{\rho}_{ij} \hat{N}_j}{m_{0j} + \sum_{i' \in \mathcal{I}} m_{i'j} \hat{\delta}_{i'} \hat{\rho}_{i'j}}\right)^{\frac{\alpha}{1-\alpha}}.$$

The first term of this expression, $\left(\hat{R}_i \hat{A}_i / \hat{w}_i\right)^{\frac{1}{1-\alpha}}$, shows that the scale elasticity α governs the effect of exogenous supplier shifters, including reimbursements $\hat{R}_i$, on quality produced in a region. Reimbursement rates shift the scale of production, and stronger scale economies (higher α) amplify these shifts. The second term shows how changes in other regions influence local outcomes through trade, combined with scale. Thus, our counterfactual scenarios rely on both our estimates of the scale elasticity α and observed trade patterns.[49,50]

assumes elastic factor supplies. Our model and estimates represent long-run elasticities, omitting any adjustment costs or short-run diseconomies of scale due to crowding or queuing.

[49]The qualitative results are not sensitive to our exact estimate of α. For instance, cutting α by half changes the magnitudes but not the geographic patterns.

[50]To compute import shares, we assume that the number of potential patients is proportional to the number of Traditional Medicare beneficiaries and infer the share choosing the outside option in each region. See Appendix C.2. The qualitative and spatial patterns of counterfactual outcomes do not depend on what share we assume choose the outside option. Appendices C.3 and C.4 generalize this method of computing counterfactual outcomes to the model with multiple types of patients introduced in Appendix B.2.

5.2 Counterfactual reimbursement rates and travel costs

Local reimbursement increase. Policymakers concerned about a region's healthcare access might consider local production subsidies. Figure 9 contrasts the consequences of raising reimbursements by 30% in Boston and in Paducah, Ky. For the Boston scenario, Panel (a) depicts the impact on quality of care in each region relative to its baseline value. Free entry means that higher reimbursements translate to higher-quality care produced in Boston. Quality declines in the rest of New England as patients substitute away and scale economies translate lower volumes into lower quality (an "agglomeration shadow", as in Fujita and Krugman, 1995). These effects diminish with distance to Boston.

Yet regions with larger declines in output quality due to Boston's expansion experience larger improvements in patient market access, $\hat{\Phi}_i$ (Panel 9(b)). This measure accounts for changes in the quality of care patients receive and their costs of traveling when they choose to do so. Patients in Boston benefit the most from the higher reimbursement of their local production. Outside Boston, regional changes in patient market access are nearly opposite the changes in local output quality. Nearby regions import enough that the benefits of improved quality in Boston exceed the declines in the quality of local production, so their patient market access improves. Regions closer to Boston experience larger declines in the quality of local production precisely because their residents' choice sets improve more, spurring more substitution. In more distant regions, the welfare impacts are virtually zero.

Patients who live in higher-income neighborhoods benefit more. The value of market access increases by 120% more for third- than for first-tercile patients nationwide (Appendix Table D.18). This reflects greater consumption of Boston production (column 3).

The consequences of higher reimbursement rates in Paducah, Ky. exhibit very different spatial patterns than in Boston. Figures 9(c) and 9(d) depict the regional changes in output quality and patient market access, respectively, caused by a 30% reimbursement increase in Paducah. Unlike Boston, Paducah is a net importer: its consumption of medical services exceeds local production by more than one-third. Higher reimbursements that improve output quality in Paducah cause Paducahans to reduce their imports from neighboring regions. This reduces the quantity produced in neighboring regions, lowering their output quality. But Panel (d) shows that—unlike the pattern of outcomes in the Boston scenario—those regions where output quality declines more are the regions where patient market access declines more.

The contrasting outcomes reflect trade flows in the baseline equilibrium: Boston is a net exporter of medical services and Paducah is a net importer. Higher reimbursements in Boston cause output quality declines in nearby regions—largely because residents of those regions import more when Boston's quality improves. In contrast, higher reimbursements in Paducah reduce neighboring regions' output quality largely because Paducah residents demand fewer exports from these regions when Paducah's quality improves. Nearby regions import little from Paducah, so they benefit little from its improved quality. Appendix Figure D.6 shows that the lessons from Boston and Paducah generalize: the pattern of spillovers from increasing reimbursements in one region is driven by that region's net trade in medical care. To summarize, the spillover consequences of subsidizing production in one region depend on trade patterns; changes in regional output quality need not align with changes in regional patient market access.

The distributional consequences of region-specific subsidies depend on which region is subsidized. We compute the nationwide gains in market access from subsidizing production in each region, one at a time. Figure 10 shows this gain, scaled by the increase in total spending, as a function of region size. The aggregate gain in market access per dollar spent is higher in larger markets: further concentration of production has larger benefits. The graph also shows the gains per dollar separately by income tercile. Subsidizing production in less populous regions benefits lower-income ZIP codes more. These contrasts reflect geographic divides in incomes: lower-income patients are more likely to live in and near smaller regions.

Reducing travel costs. Rather than subsidizing local production, policies might improve patient market access in a particular region by facilitating trade. We examine the consequences of a policy that reduces travel costs for Paducahans obtaining care elsewhere.[51] Figure 11 shows that, unlike an increase in Paducah reimbursements, this policy has positive spillovers on neighboring regions. These regions increase their exports to Paducah, and thus their own scale and quality. This improves their residents' market access.

Reducing travel costs for Paducahans benefits both Paducah and its neighbors—though we do not estimate the fiscal costs of subsidizing travel. Paducahans benefit even though facilitating travel reduces the quantity—and thus the quality—produced in Paducah. Analysts looking at the impact of travel subsidies on the quantity or quality of care provided in Paducah itself would reach very different conclusions than those looking at the impact on patient market access.

These counterfactual scenarios are subject to significant caveats, and we have not attempted to identify the optimal policy. Even so, this simple model rationalizes important aspects of the economic geography of US healthcare policy. The counterfactual scenarios highlight our main findings: Healthcare production has substantial local increasing returns, and patient travel plays a meaningful role in enabling access to higher-quality care. Given these economic mechanisms, regional spillovers are larger when economies of scale are stronger, depend on the pattern of trade flows, and differ depending on whether policies subsidize production or travel. This shows the importance of distinguishing between regional output quality and regional patient access when evaluating healthcare policies.

5.3 Geography and healthcare inequality across countries

Looking at healthcare geography more broadly, a striking fact is that the United States has a steeper relationship between regional income and mortality than other rich economies.

[51] Specifically, $\hat{\rho}_{i,\text{Paducah}} = 1.3$ when $i \neq$ Paducah. The impact of this change on Paducah residents' market access $\hat{\Phi}_{\text{Paducah}}$ is similar to a 3% increase in reimbursements in Paducah.

Appendix Figure D.8(a) compares the US health-income gradient with Germany's. We use our model to explore how much geography can explain this difference in gradients.

To do so, we compute the changes in trade costs that would make the distribution of market potential across US regions comparable to that across German regions.[52] This requires broad reductions in travel costs, effectively condensing the United States. It requires particularly large declines in travel costs for remote US regions.

This transformation reduces the Φ_i-income elasticity by 21% (Appendix Figure D.8(b)). This suggests that a substantial part of the inequality in market access can be explained by the US's greater geographic area. If the US were Germany's size, patients from lower-income areas would see particular benefit: these areas are disproportionately rural, and thus experience the largest gains in this counterfactual. Viewed from the opposite perspective, 27% of the larger Φ_i-income elasticity in the United States is because of its geographic scope.

6 Conclusion

Smaller markets have fewer specialized physicians, produce less medical care per capita, and have worse health outcomes than larger markets. Thanks to trade in medical services, less production does not translate one for one into less consumption of medical services. Instead, trade affords patients who live in smaller markets access to higher-quality care. This quality comes in part from consuming services that would otherwise be unavailable, visiting appropriate specialists, and accessing experienced physicians.

This trade amplifies the scale advantages of large markets and hence the quality of care they produce. This means the healthcare industry can serve as an export base for large cities. Substantial scale economies also imply that policies to reallocate care across regions may impact the quality of care available. The rich and varied patterns of consequences when subsidizing production or travel in "under-served" markets highlight the importance of trade and agglomeration for the incidence of these policies on patients and producers.

[52]See Appendix C.5 for details. The analysis is similar when we compare the United States to France.

References

AAPC. 2021. "What Is CPT?" Available online at https://www.aapc.com/resources/medical-coding/cpt.aspx.

Acemoglu, Daron, Amy Finkelstein, and Matthew J. Notowidigdo. 2013. "Income and Health Spending: Evidence from Oil Price Shocks." *The Review of Economics and Statistics*, 95(4): 1079–1095.

Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *The Quarterly Journal of Economics*, 119(3): 1049–1090.

Agarwal, Sumit, J. Bradford Jensen, and Ferdinando Monte. 2020. "Consumer Mobility and the Local Structure of Consumption Industries." NBER Working Paper 23616.

Allen, Treb, Simon Fuchs, Sharat Ganapati, Alberto Graziano, Rocio Madera, and Judit Montoriol-Garriga. 2021. "Urban Welfare: Tourism in Barcelona."

Anderson, James E., Catherine A. Milot, and Yoto V. Yotov. 2014. "How much does geography deflect services trade? Canadian answers." *Int'l Econ. Rev.*, 55(3): 791–818.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review*, 53(5): 941–973.

Baicker, Katherine, and Amitabh Chandra. 2010. "Understanding Agglomerations in Health Care." *Agglomeration Economics*, 211–236. University of Chicago Press.

Bartelme, Dominick G., Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2019. "The Textbook Case for Industrial Policy: Theory Meets Data." NBER NBER Working Paper 26193.

Bartik, Timothy, and George Erickcek. 2007. "Higher Education, the Health Care Industry, and Metropolitan Regional Economic Development: What Can 'Eds & Meds' Do for the Economic Fortunes of a Metro Area's Residents?" Upjohn Inst. Working Paper 08-140.

Baumgardner, James R. 1988. "Physicians' Services and the Division of Labor across Local Markets." *Journal of Political Economy*, 96(5): 948–982.

Berenson, Robert A., Jonathan H. Sunshine, David Helms, and Emily Lawton. 2015. "Why Medicare Advantage plans pay hospitals traditional Medicare prices." *Health Affairs*, 34(8): 1289–1295.

Berry, Christopher R., and Edward L. Glaeser. 2005. "The divergence of human capital levels across cities." *Papers in Regional Science*, 84(3): 407–444.

Borchert, Ingo, Mario Larch, Serge Shikher, and Yoto V. Yotov. 2021. "The International Trade and Production Database for Estimation (ITPD-E)." *International Economics*, 166(C): 140–166.

Burstein, Ariel, Sarah Lein, and Jonathan Vogel. 2022. "Cross-border shopping: evidence and welfare implications for Switzerland."

Cabral, Marika, and Neale Mahoney. 2019. "Externalities and taxation of supplemental insurance: A study of Medicare and Medigap." *American Economic Journal: Applied Economics*, 11(2): 37–73.

Centers for Medicare and Medicaid Services. 2022. "National Health Expenditure Data."

Chandra, Amitabh, Maurice Dalton, and Douglas O Staiger. 2023. "Are Hospital Quality Indicators Causal?" NBER Working Paper 31789.

Chipman, John S. 1970. "External Economies of Scale and Competitive Equilibrium." *The Quarterly Journal of Economics*, 84(3): 347–385.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Treatment Patterns and Patient Health?" *American Economic Review*, 104(4): 1320–1349.

Costinot, Arnaud, Dave Donaldson, Margaret Kyle, and Heidi Williams. 2019. "The More We Die, The More We Sell? A Simple Test of the Home-Market Effect." *The Quarterly Journal of Economics*, 134(2): 843–894.

Davis, Donald R., and David E. Weinstein. 2003. "Market access, economic geography and comparative advantage: an empirical test." *J. of International Economics*, 59(1): 1–23.

Davis, Donald R., and Jonathan I. Dingel. 2020. "The comparative advantage of cities." *Journal of International Economics*, 123(C).

Davis, Donald R., Jonathan I. Dingel, Joan Monras, and Eduardo Morales. 2019. "How Segregated Is Urban Consumption?" *Journal of Political Economy*, 127(4): 1684–1738.

Diamond, Rebecca. 2016. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." *American Economic Review*, 106(3): 479–524.

Dingel, Jonathan I. 2017. "The Determinants of Quality Specialization." *Review of Economic Studies*, 84(4): 1551–1582.

Dingel, Jonathan I., and Felix Tintelnot. 2021. "Spatial Economics for Granular Settings." NBER Working Paper 27287.

Disdier, Anne-Célia, and Keith Head. 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics*, 90(1): 37–48.

Dorsey, E. Ray, and Eric J. Topol. 2016. "State of Telehealth." *New England Journal of Medicine*, 375(2): 154–161. PMID: 27410924.

Dranove, David, and Mark A. Satterthwaite. 2000. "The Industrial Organization of Health Care Markets." *Handbook of Health Economics*, 1: 1093–1139.

Dranove, David, Mark Shanley, and Carol Simon. 1992. “Is Hospital Competition Wasteful?” *The RAND Journal of Economics*, 23(2): 247–262.

Eckert, Fabian, Sharat Ganapati, and Conor Walsh. 2020. “Skilled Scalable Services: The New Urban Bias in Economic Growth.” CESifo Working Paper Series 8705.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. “Sources of geographic variation in health care: Evidence from patient migration.” *QJE*, 131(4): 1681–1726.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. “Place-Based Drivers of Mortality: Evidence from Migration.” *American Economic Review*, 111(8): 2697–2735.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. Lee Lucas, and Etoile L. Pinder. 2003*a*. “The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care.” *Ann. Intern. Med.*, 138(4): 273–287.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. Lee Lucas, and Etoile L. Pinder. 2003*b*. “The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care.” *Ann. Intern. Med.*, 138(4): 288–298.

Flexner, Abraham. 1910. “Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching.” Carnegie Fnd. Bulletin 4.

Fowler, Christopher S, and Leif Jensen. 2020. “Bridging the gap between geographic concept and the data we have: The case of labor markets in the USA.” *Environment and Planning A: Economy and Space*, 52(7): 1395–1414.

Fujita, Masahisa, and Paul Krugman. 1995. “When is the economy monocentric?: von Thünen and Chamberlin unified.” *Regional Science and Urban Econ.*, 25(4): 505–528.

Gaynor, Martin, and Robert J Town. 2011. “Competition in health care markets.” *Handbook of Health Economics* Vol. 2, 499–637. Elsevier.

Gaynor, Martin, Kate Ho, and Robert J Town. 2015. “The industrial organization of health-care markets.” *Journal of Economic Literature*, 53(2): 235–284.

Gottlieb, Daniel J, Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan S Skinner, and Jason M Sutherland. 2010. “Prices don’t drive regional Medicare spending variations.” *Health Affairs*, 29(3): 537–543.

Gottlieb, Joshua D., Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova. 2023. “Who Values Human Capitalists’ Human Capital? The Earnings and Labor Supply of U.S. Physicians.” National Bureau of Economic Research Working Paper 31469.

Gotto, Antonio M., and Jennifer Moon. 2016. *Weill Cornell Medicine: A History of Cornell’s Medical School.* Ithaca: Cornell University Press.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. “Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings.” *Journal of Political Economy*, 118(3): 536–598.

Gupta, Atul. 2021. "Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program." *American Economic Review*, 111(4): 1241–83.

Hanson, Gordon H., and Chong Xiang. 2004. "The Home-Market Effect and Bilateral Trade Patterns." *American Economic Review*, 94(4): 1108–1129.

Helpman, Elhanan, and Paul R. Krugman. 1985. *Market Structure and Foreign Trade.* MIT Press.

Hsieh, Chang-Tai, and Esteban Rossi-Hansberg. 2021. "The Industrial Revolution in Services." Center for Economic Studies, U.S. Census Bureau Working Papers 21-34.

Jensen, J. Bradford, and Lori G. Kletzer. 2005. "Tradable Services: Understanding the Scope and Impact of Services Offshoring." *Brookings Trade Forum*, 75–116.

Kline, Patrick, and Enrico Moretti. 2013. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics*, 129(1): 275–331.

Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review*, 70(5): 950–59.

Lee, Sanghoon. 2010. "Ability sorting and consumer city." *Journal of Urban Economics*, 68(1): 20–33.

Levy, Antoine, and Jacob Moscona. 2020. "Specializing in Density: Spatial Sorting and the Pattern of Trade." Mimeo, MIT.

Lipsey, Robert E. 2009. "Measuring International Trade in Services." In *International Trade in Services and Intangibles in the Era of Globalization.* , ed. Marshall B. Reinsdorf and Matthew J. Slaughter, 27–74. University of Chicago Press.

Lopez, Eric, and Gretchen Jacobson. 2020. "How Much More Than Medicare Do Private Insurers Pay? A Review of the Literature."

Marshall, Alfred. 1890. *Principles of Economics.* London: Macmillan and Co.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J. Redding. 2021. "The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data." NBER Working Paper 28497.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, migration, and local employment elasticities." *American Economic Review*, 108(12): 3855–3890.

Montiel Olea, José Luis, and Carolin Pflueger. 2013. "A robust test for weak instruments." *Journal of Business & Economic Statistics*, 31(3): 358–369.

Moretti, Enrico. 2004. "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data." *J. of Econometrics*, 121(1-2): 175–212.

Moretti, Enrico. 2011. "Local labor markets." In *Handbook of Labor Economics*. Vol. 4, 1237–1313. Elsevier.

Mourot, Pauline. 2024. "Should Top Surgeons Practice at Top Hospitals? Sorting and Complementarities in Healthcare." University of Chicago, mimeo.

Muñoz, Mathilde. 2022. "Trading Non-Tradables: The Implications of Europe's Job Posting Policy."

Neufeld, Jonathan D., and Charles R. Doarn. 2015. "Telemedicine Spending by Medicare: A Snapshot from 2012." *Telemed J E Health*, 21(8): 686–693.

Newhouse, Joseph P. 1990. "Geographic access to physician services." *Annual Review of Public Health*, 11(1): 207–230.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz. 1982*a*. "Does the Geographical Distribution of Physicians Reflect Market Failure?" *The Bell Journal of Economics*, 13(2): 493–505.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz. 1982*b*. "The geographic distribution of physicians: Is the conventional wisdom correct?" Santa Monica: RAND Corp. Publ. No. R-2734.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz. 1982*c*. "Where have all the doctors gone?" *JAMA*, 247(17): 2392–2396.

Rosenblatt, R. A., and L. G. Hart. 2000. "Physicians and rural America." *The Western Journal of Medicine*, 173(5): 348–351.

Rosenthal, Meredith B., Alan Zaslavsky, and Joseph P. Newhouse. 2005. "The Geographic Distribution of Physicians Revisited." *Health Services Research*, 40(6p1): 1931–1952.

Rosenthal, Stuart S., and William C. Strange. 2008. "The attenuation of human capital spillovers." *Journal of Urban Economics*, 64(2): 373–389.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek. 2022. "IPUMS USA: Version 12.0." Minneapolis, MN: IPUMS.

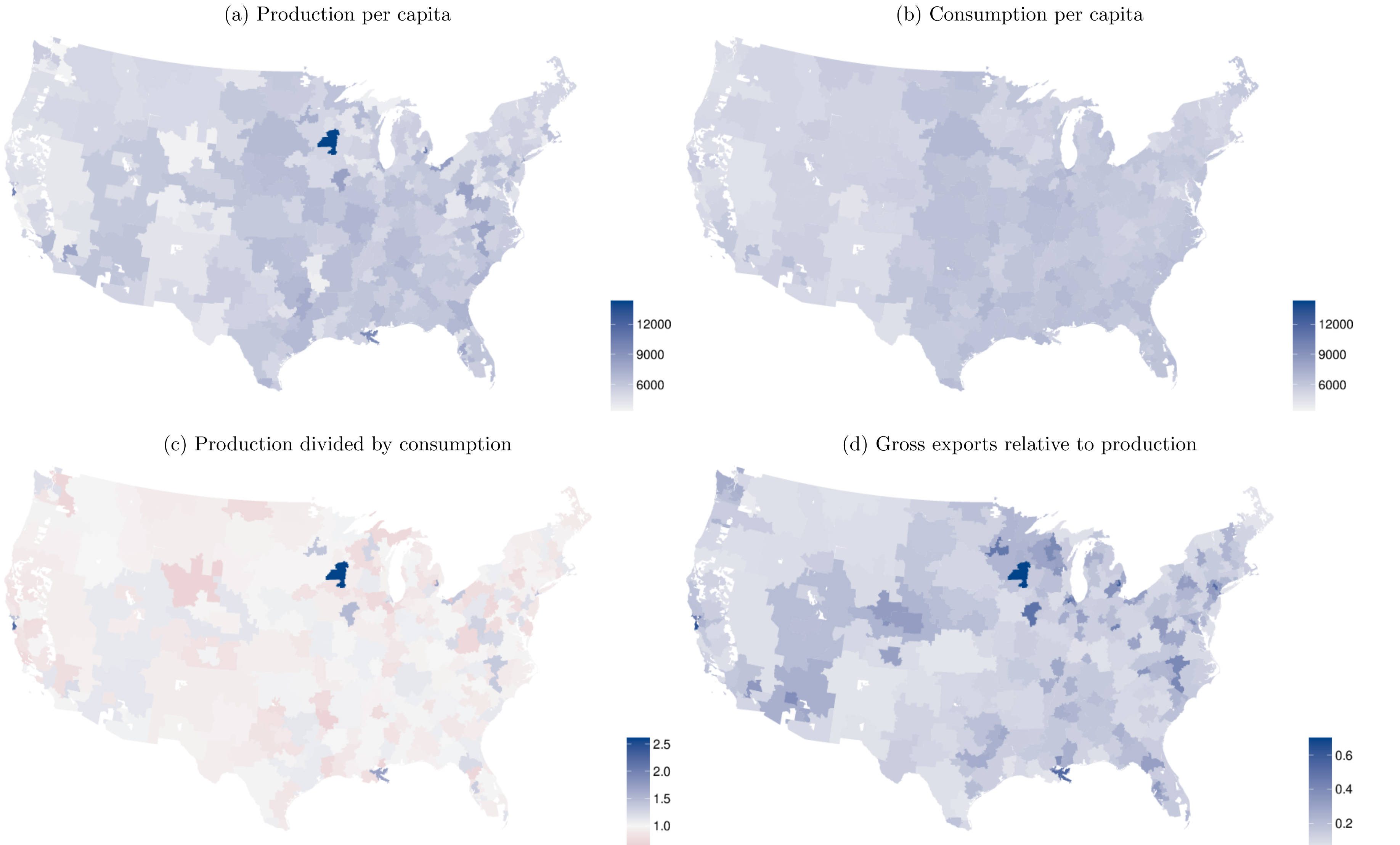
Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

Silver, David, and Jonathan Zhang. 2022. "Impacts of Basic Income on Health and Economic Well-Being: Evidence from the VA's Disability Compensation Program." NBER Working Paper 29877.

Skinner, Lucy, Douglas O. Staiger, David I. Auerbach, and Peter I. Buerhaus. 2019. "Implications of an aging rural physician workforce." *NEJM*, 381(4): 299–301.

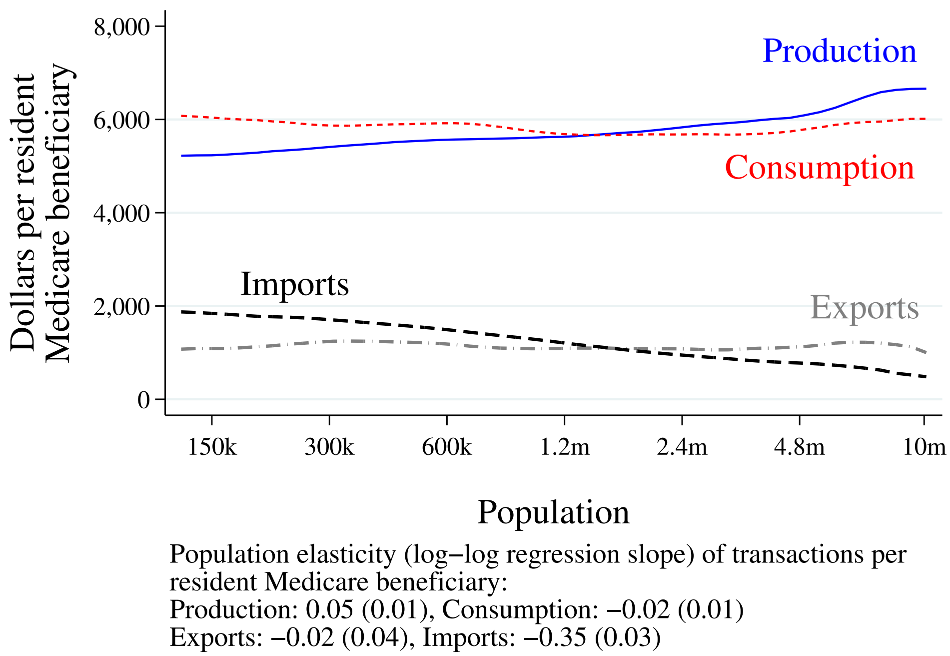
Zuckerman, Stephen, Laura Skopec, and Joshua Aarons. 2021. "Medicaid Physician Fees Remained Substantially Below Fees Paid By Medicare In 2019." *Health Affairs*, 40(2): 343–348.

Figure 1: Production, consumption, and trade across regions



Notes: These maps depict production, consumption, and trade by hospital referral region (HRR). Panel (a) shows dollars of production per capita. The HRR of production is the location where the service is provided. Panel (b) shows dollars of consumption per capita. The HRR of consumption is based on the patient's residential address. Panel (c) shows the ratio of production per capita to consumption per capita. Panel (d) shows gross exports as a share of production. Section 1.1 and Appendices A.1 and A.2 detail how we compute trade flows in physician services (excluding emergency-room care and skilled nursing facilities) at standardized prices from the Medicare 20% carrier, 100% MedPAR, and 100% outpatient claims Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. The Anchorage and Honolulu HRRs are not depicted.

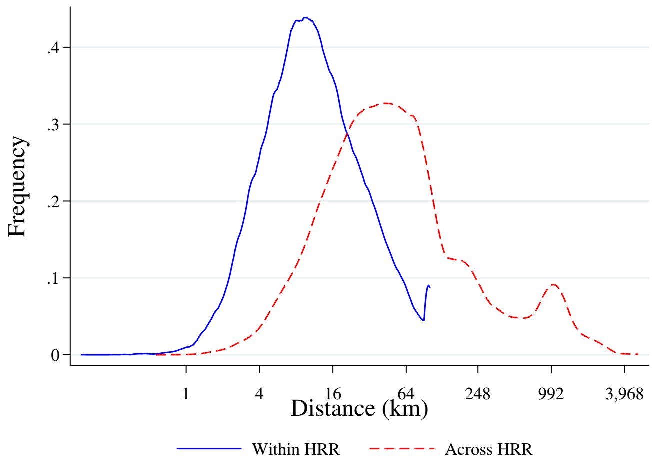
Figure 2: Production and consumption of medical care across regions



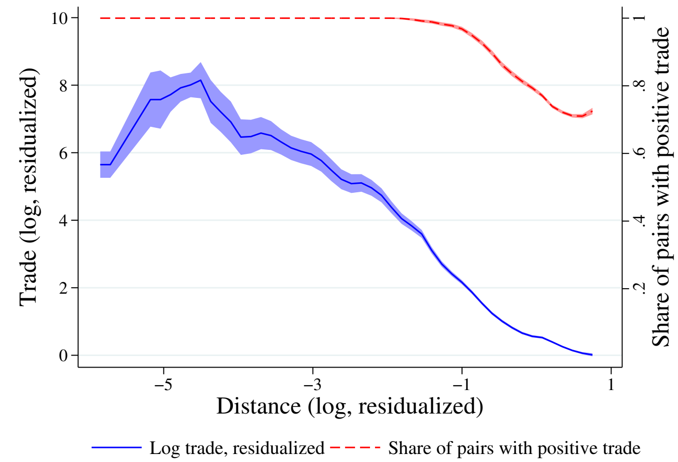
Notes: This figure shows production, consumption, and trade per capita of Medicare services across hospital referral regions (HRRs) of different sizes, all smoothed via local averages. The blue series shows production of medical care per Medicare beneficiary residing in the HRR of production. The red series shows consumption of medical care per Medicare beneficiary residing in the HRR of consumption. The dashed gray series shows interregional "exports" of medical care and the dashed black series shows interregional "imports" of medical care, again per Medicare beneficiary. Section 1.1 and Appendices A.1 and A.2 detail how we compute trade flows in physician services (excluding emergency-room care and skilled nursing facilities) at standardized prices from the Medicare 20% carrier, 100% MedPAR, and 100% outpatient claims Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project.

Figure 3: Patients travel between regions and trade declines with distance, more so for lower-income patients

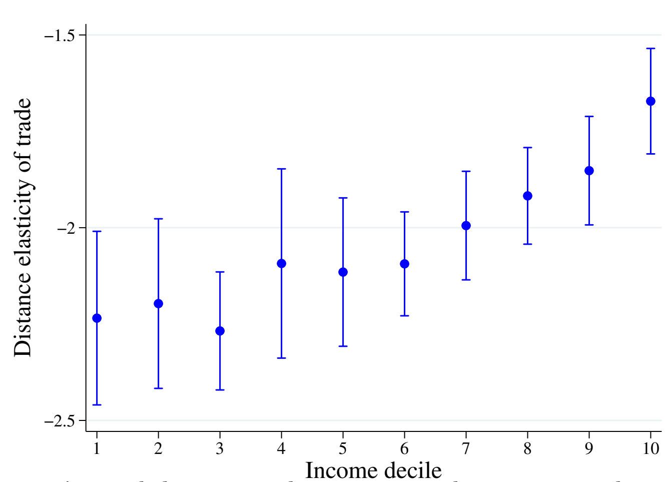
(a) Distribution of travel distances within and across HRRs



(b) Trade volume and extensive margin by distance

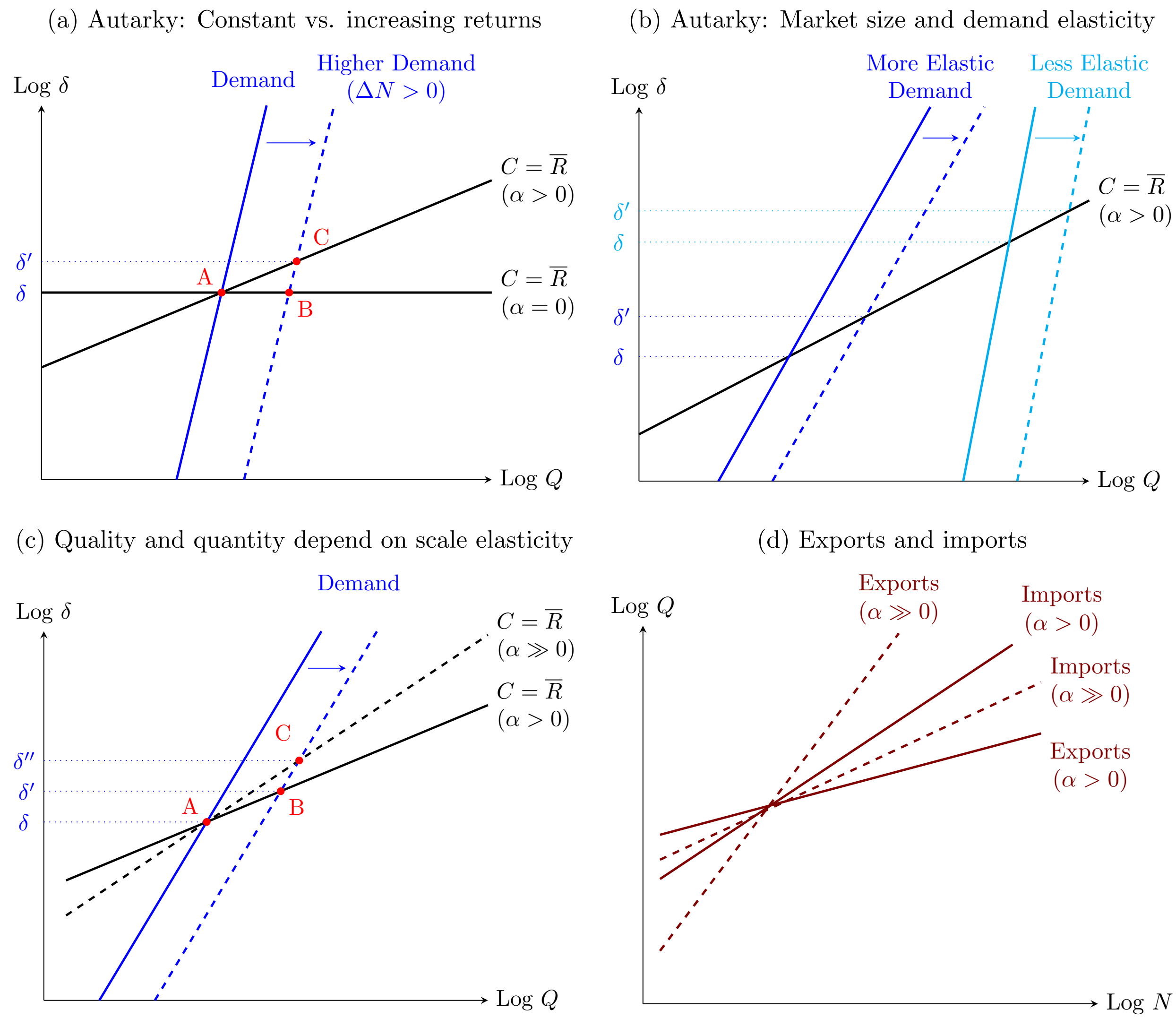


(c) Higher-income patients are less sensitive to distance



Notes: Panel (a) shows the distribution of patients' travel distances when patients obtain care within their home HRR (blue distribution) and when they travel across HRRs (red distribution). Travel distances measure the distance between home and treatment locations. For travel within a hospital referral region, the distance measure reflects the distance between the centroid of the patient's residential ZIP code and the ZIP code of the service location. We use ZCTA-to-ZCTA distances downloaded from the National Bureau of Economic Research; those exceeding 160 kilometers are winsorized at 160 kilometers. For travel across HRRs, we use ZCTA-to-ZCTA distances when they are within 160 kilometers and (for computational ease) use HRR-to-HRR distances beyond 160 kilometers. In Panel (b), the blue series depicts the volume of trade against distance, after conditioning out the fixed effects in equation (8), for positive-trade pairs of locations. The red series shows the share of HRR pairs with positive trade as a function of the distance between them, after conditioning out the importer fixed effects and exporter fixed effects, as in equation (8). Panel (c) depicts the coefficient on log distance obtained by estimating equation (8) separately for each decile of the national ZIP-level median-household-income distribution. The 95% confidence intervals are computed using standard errors two-way clustered by both patient HRR and provider HRR. Patients from higher-income ZIP codes are less sensitive to distance.

Figure 4: Illustrative model diagrams

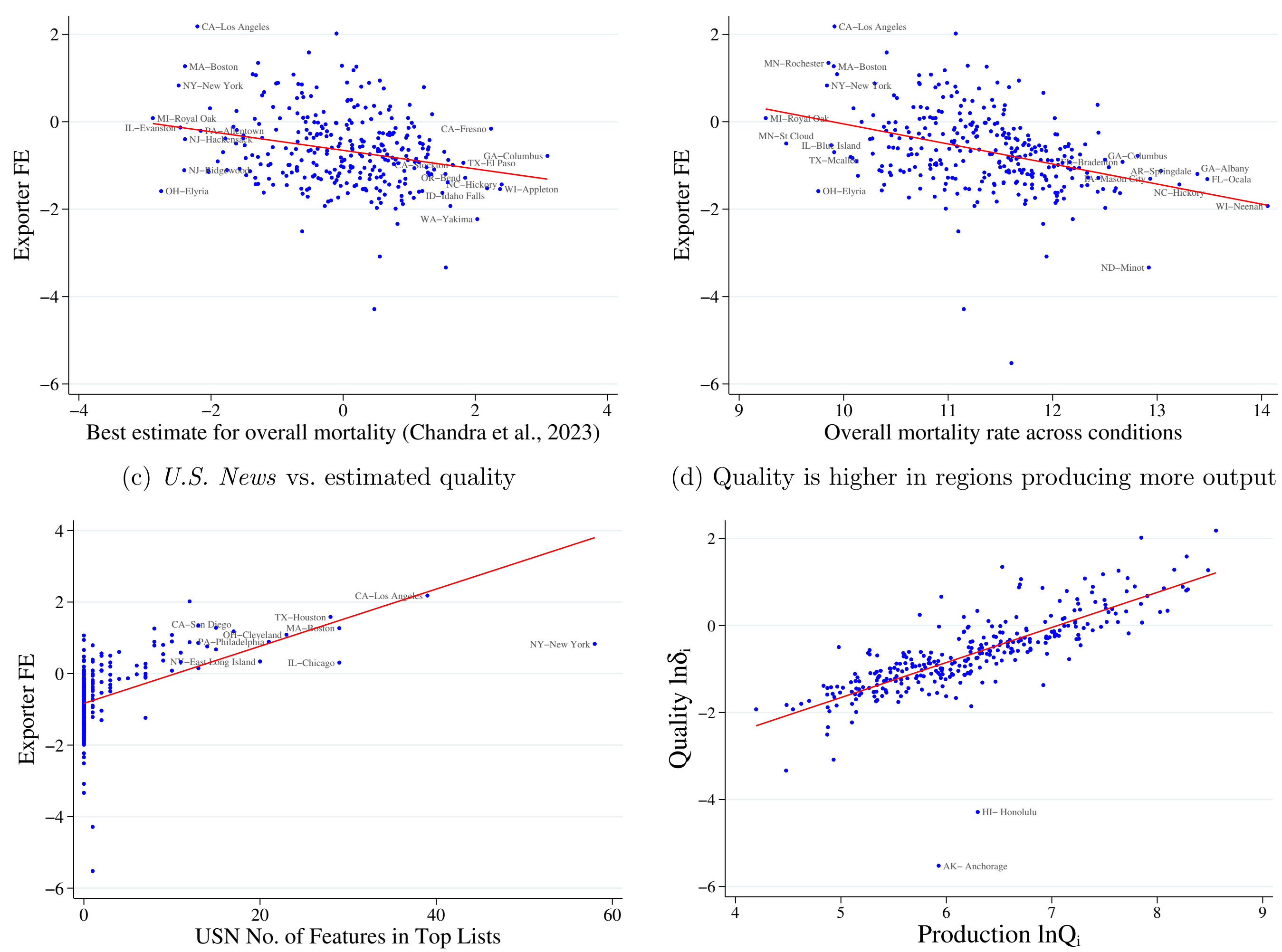


Notes: This figure depicts how increasing demand in one region affects its equilibrium outcomes. In Panels (a)–(c), quantity produced Q is on the horizontal axis and service quality δ is on the vertical axis. The black lines depict the free-entry isocost curve, $C = \overline{R}$, given by equation (3). The blue and cyan lines depict demand for the region's service, which we depict as log-linear for visual clarity. (The logit demand function is actually log-convex, which is consistent with all the depicted comparative statistics.) Equilibrium is the intersection of the demand and isocost curves. An increase in demand is the rightward shift from the solid to the dashed demand curve. This shift increases equilibrium quality from δ to δ'. Panel (a) shows that higher demand elicits higher quality if there are increasing returns to scale. Panel (b) shows that this quality improvement is larger when demand is more elastic. Panels (c) and (d) introduce trade and compare the extent of quality improvement under two different magnitudes of increasing returns ($\alpha > 0$ and $\alpha \gg 0$). These magnitudes govern the patterns of interregional trade, shown in Panel (d) as a function of the number of potential patients N. Imports from other regions rise with N. With increasing returns to scale ($\alpha > 0$), exports to other regions also rise with N (a weak home-market effect). When the scale elasticity α is larger ($\alpha \gg 0$), the import curve is flatter and the export curve is steeper. With sufficiently strong increasing returns, an increase in local demand causes a greater increase in exports than imports (a strong home-market effect).

Figure 5: Estimated quality is positively correlated with total output and external quality metrics

(a) Quality estimates vs. mortality rate (Chandra et al.) (b) Quality estimates vs. mortality rate (CMS)

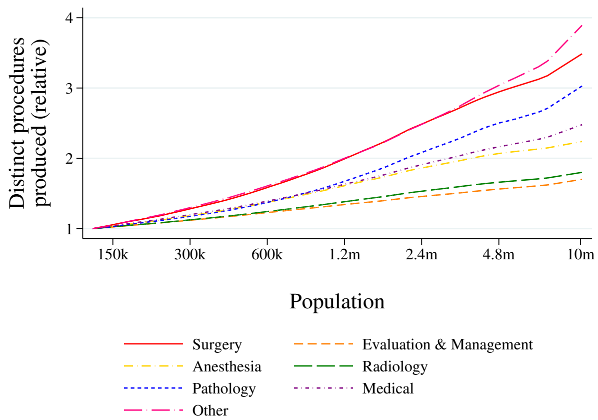
(c) *U.S. News* vs. estimated quality (d) Quality is higher in regions producing more output



Notes: The first three panels show the relationship between the exporter fixed effects (our revealed-preference measure of quality) and external quality measures. The vertical axis shows the exporter fixed effects for each HRR estimated from equation (8). The horizontal axis in Panel (a) is an average of the hospital mortality rates estimated by Chandra, Dalton, and Staiger (2023) in a region, and in Panel (b) an average of those estimated by CMS. The negative correlations indicate patients travel farther to obtain care from regions with lower hospital mortality rates. The horizontal axis in Panel (c) is a count of the number of times each region's hospitals appear on the *U.S. News* list of best hospitals. *U.S. News* produces an overall ranking as well as rankings for 12 particular specialties. We count the number of times each HRR's hospitals appear on any of these 13 lists. The relationship is positive, indicating that patients travel farther to obtain care from regions highly ranked by *U.S. News*. The horizontal axis in Panel (d) is the volume of production. Section 1.1 and Appendices A.1 and A.2 detail how we compute production and trade of physician services (excluding emergency-room care and skilled nursing facilities) at standardized prices from the Medicare 20% carrier, 100% MedPAR, and 100% outpatient claims Research Identifiable Files.

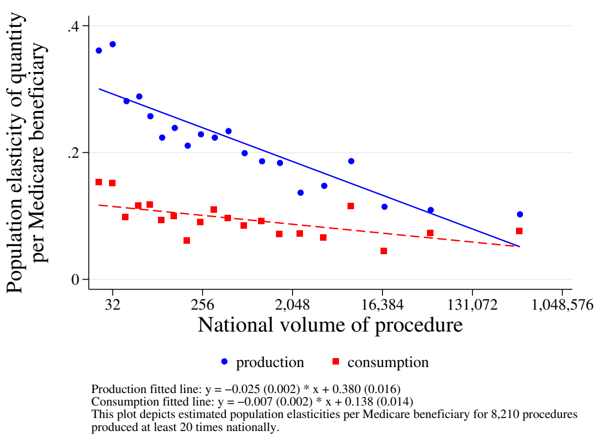
Figure 6: Production and consumption of medical care by market size and service

(a) Variety of procedures produced by market size

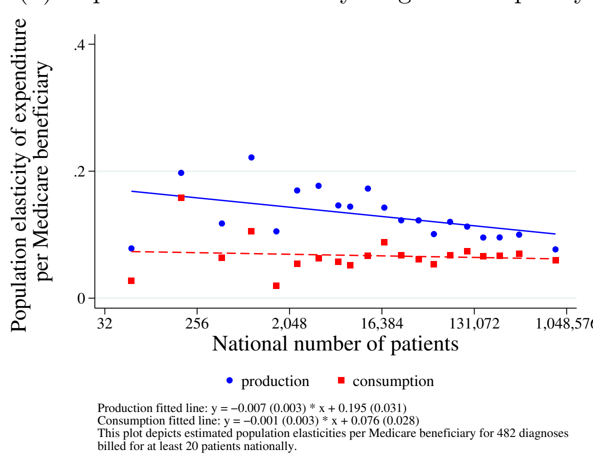


(b) Variety of procedures consumed by market size

(c) Population elasticities by procedure frequency



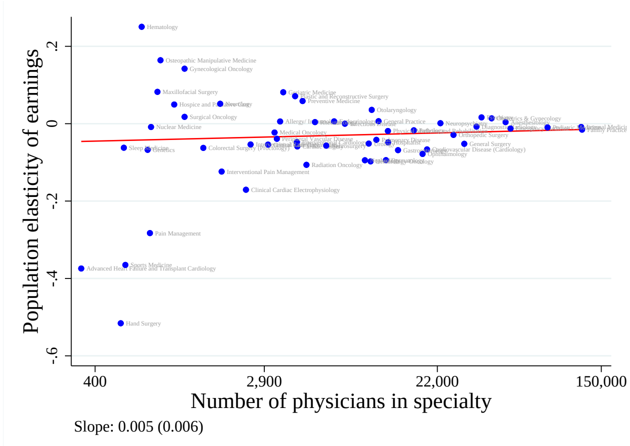
(d) Population elasticities by diagnosis frequency



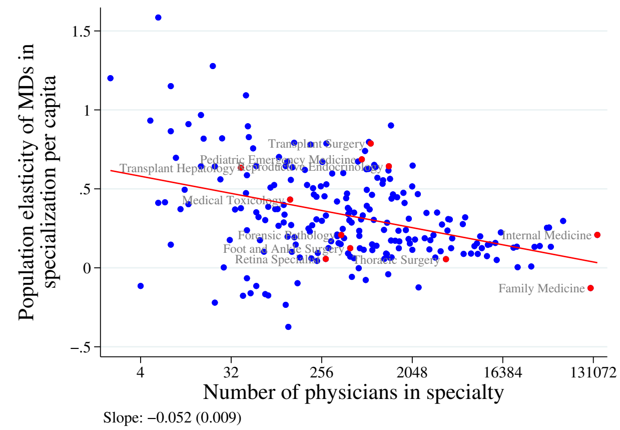
Notes: Panel (a) depicts the number of distinct services produced in an HRR as a function of population. We use procedure classifications from the American Academy of Professional Coders, which groups codes into surgeries, anesthesia, radiology, pathology, medical, and evaluation & management services (AAPC, 2021). We combine Category II codes, Category III codes and Multianalyte Assays into "other." Within each group, the procedure count is relative to that of the smallest regions. More populous HRRs produce a greater number of services. Panel (b) depicts the number of distinct services consumed in an HRR as a function of population. In Panels (c) and (d), the vertical axes are the population elasticities of quantity of medical care produced (blue dots) and consumed (red squares) per local Medicare beneficiary. The elasticities are computed using the Poisson model in equation (13) based on place of service and patients' residential location, respectively. Panel (c) estimates these elasticities for each of the procedures provided at least 20 times nationally in the Medicare data. Panel (d) estimates the elasticities for care provided to treat each of the Clinical Classifications Software Refined (CCSR) diagnoses billed for at least 20 patients nationally in the Medicare data. Each panel depicts these population elasticities as a function of the national volume (of procedures and diagnoses, respectively). Section 1.1 and Appendices A.1 and A.2 detail how we compute production and consumption of physician services (excluding emergency-room care and skilled nursing facilities) at standardized prices from the Medicare 20% carrier, 100% MedPAR, and 100% outpatient claims Research Identifiable Files. The contrasting population elasticities of production and consumption imply trade in medical services between markets of different sizes, with more net trade for rare procedures and rare diseases.

Figure 7: Larger regions produce more specialized care, with more specialized equipment

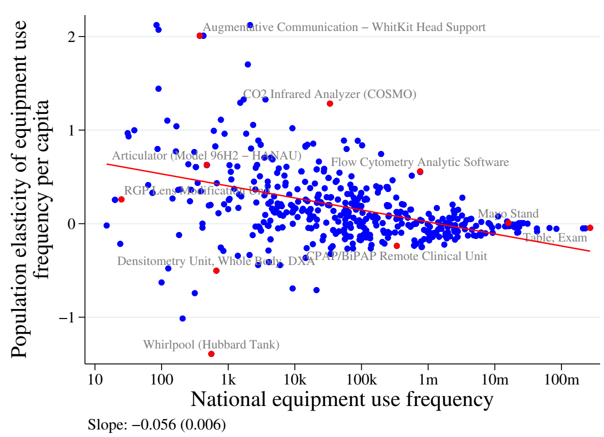
(a) Population elasticities of specialties' earnings



(b) Population elasticities of physician specializations

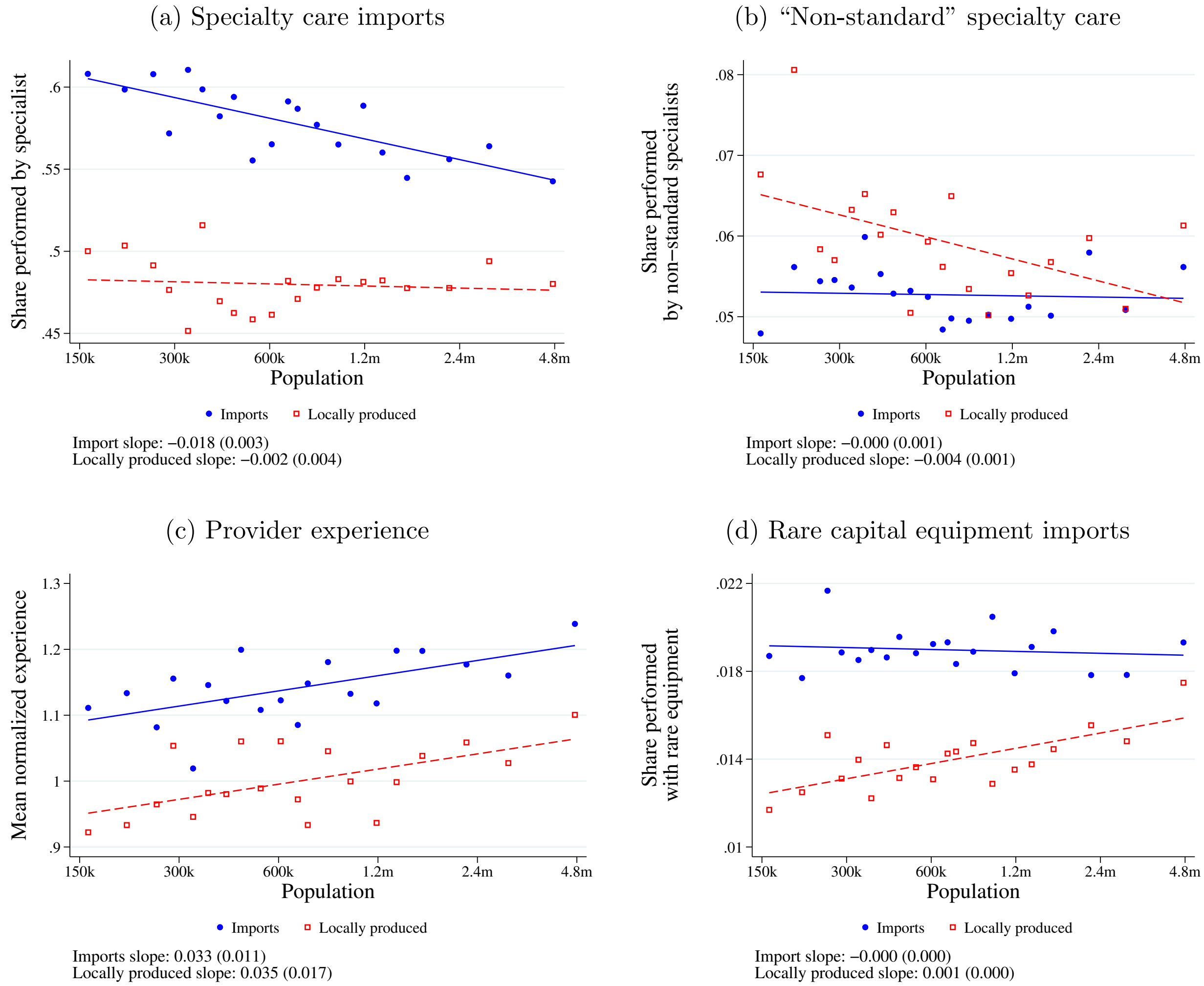


(c) Population elasticities of equipment use



Notes: Panel (a) shows the population elasticity of income for different medical specialties against the total number of physicians in those specialties. For each specialty, we estimate the elasticity of income with respect to population across commuting zones, using data from Gottlieb et al. (2023). The regression line weights each specialty with the number of physicians in that specialty. The graph shows that these elasticities are unrelated to the total national count of physicians in those specialties. The vertical axis of Panel (b) depicts the population elasticities of quantity of physicians in an HRR. The population elasticities are computed for each specialty using the Poisson model in equation (14). The horizontal axis shows the nationwide number of physicians in each specialty. The negative relationship indicates that rare specialties are disproportionately concentrated in high-population regions. Panel (c) plots the population elasticity of equipment use frequency per capita against the national frequency of equipment use, for each piece of equipment. Pieces of equipment with elasticities above the 98th percentile are excluded. The graph shows that rarely-used equipment is used more intensively in large markets.

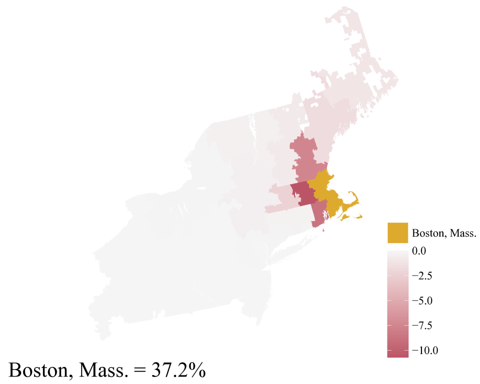
Figure 8: Imports are specialist-intensive, especially in smaller regions



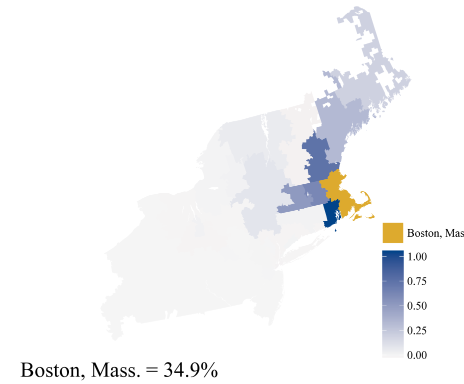
Notes: Panel (a) shows the share of procedures that are performed by a specialist, for imports and locally produced procedures, by market size. We define generalists as internal-medicine, general-practice, and family-practice physicians and define specialists as all other physicians. Imports are more likely to be performed by a specialist, and smaller markets' imports especially so. Panel (b) examines procedures that are typically performed by specialists, and classifies the "standard" specialists as the top two specialties performing the procedure nationally. It shows the shares of procedures performed by the "non-standard" specialties in imported specialty care and locally produced specialty care as a function of local population size. Imports are less likely to be performed by "non-standard" specialties, especially for smaller regions. Panel (c) shows the mean relative experience of providers for care produced locally and imported by population size of the patient's region. This panel describes only procedures that are performed in all hospital referral regions (143 procedures). In public-use Medicare data, we define a provider's experience for a given procedure as the number of times they performed the procedure for Traditional Medicare patients in the prior calendar year. Before aggregating to the regional level, we rescale experience in each procedure so that its mean is one. We further normalize at the regional level so that the experience of the average HRR is one. On average, patients in larger markets obtain treatment from more experienced providers. At all population levels, imported care is produced by more experienced providers than local care. Panel (d) shows the share of procedures that use rare capital equipment in imported and locally produced care by market size. Rare equipment is defined as those pieces of equipment with below-median use in the Medicare data, defined as in Appendix A.1. We see that locally produced care in larger markets is more likely to use rare equipment, and imported care has a higher rare-equipment share throughout the population distribution

Figure 9: Counterfactual outcomes for higher reimbursements in one region

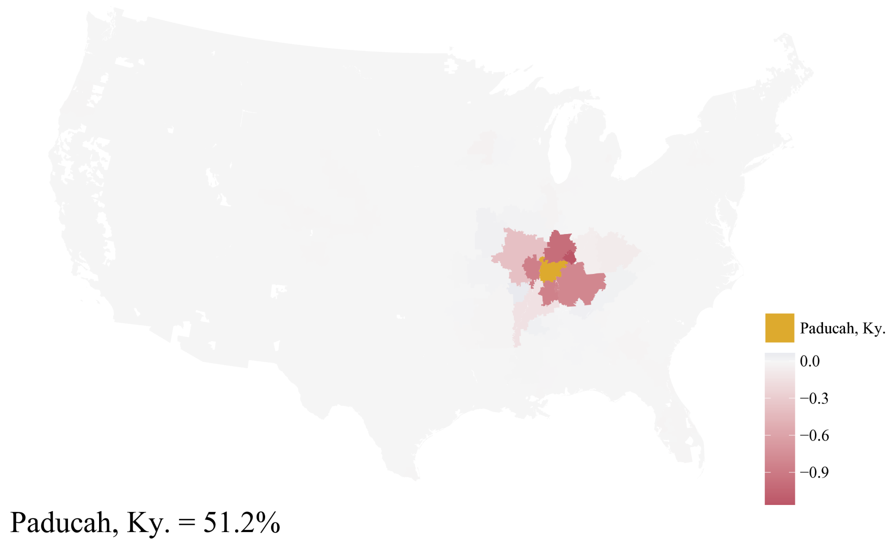
(a) Change (%) in output quality δ_i: higher reimbursement in Boston, Mass.



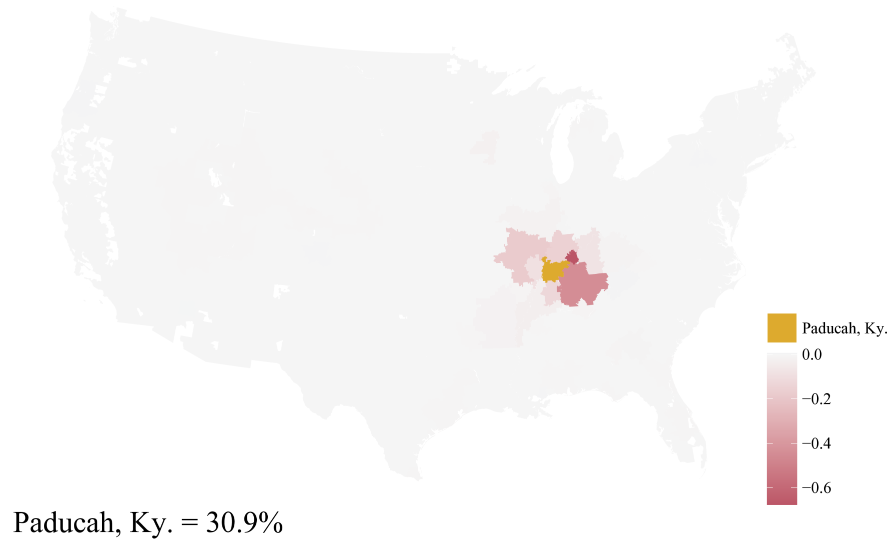
(b) Change (%) in market access Φ_i: higher reimbursement in Boston, Mass.



(c) Change (%) in output quality δ_i: higher reimbursement in Paducah, Ky.

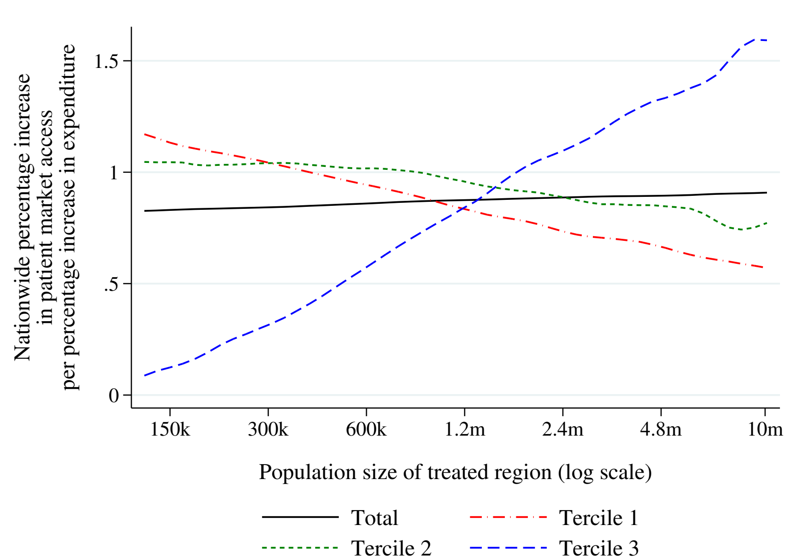


(d) Change (%) in market access Φ_i: higher reimbursement in Paducah, Ky.



Notes: Panels (a) and (b) show the impacts of increasing reimbursements by 30% in the Boston, Mass. HRR ($\hat{R}_i = 1.3$) based on our estimated model. Panel (a) illustrates the percentage change in quality of care δ_i provided in each region. Panel (b) illustrates the percentage change in the value of market access Φ_i for patients who live in an region. Panels (c) and (d) are analogous, but for a 30% increase in reimbursements in Paducah, Ky., a net importer. In all panels, the predicted change for the region whose reimbursement changes ("treated region") is listed on the map itself. In both cases, the quality produced in neighboring regions declines (Panels (a) and (c)). Patients in regions near Boston benefit from increased access to the treated region (Panel (b)), so there is a negative relationship between the percentage changes in δ and Φ across regions. In contrast, patients in regions near Paducah suffer a decrease in access (Panel (d)). The contrasting outcomes stem from Boston being a net exporter and Paducah being a net importer in the baseline equilibrium. The exercise is described in detail in Section 5.

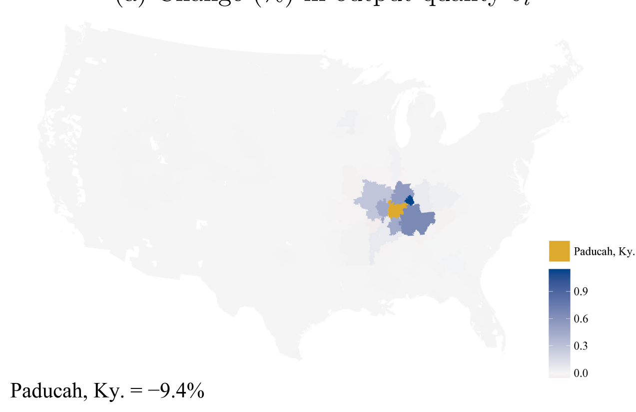
Figure 10: Returns to higher reimbursements in one region by region size



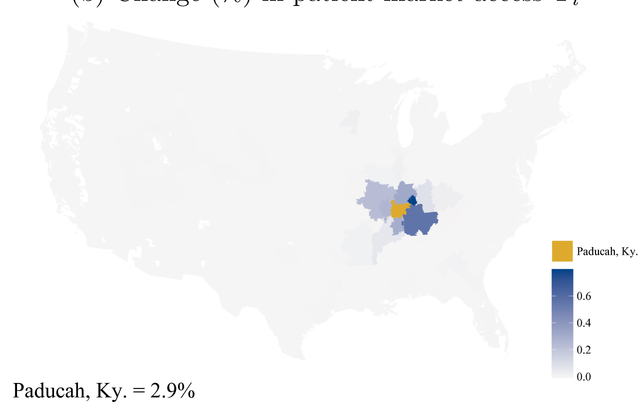
Notes: This figure summarizes the counterfactual outcomes of 30% higher reimbursements in one HRR as a function that HRR's population size. The nationwide return is the percentage increase in patient market access $\sum_\kappa \sum_j N_{j\kappa}\Phi_{j\kappa}$ per percentage increase in nationwide expenditures $\sum_i Q_i R_i$. The tercile-specific return is the increase in tercile-specific patient market access $\sum_j N_{j\kappa}\Phi_{j\kappa}$. Increasing reimbursements in more populous HRRs has the highest return when measured as impact on aggregate market access. Subsidies in less populous regions favor lower-income patients, primarily because there are more low-income patients living in and close to smaller regions.

Figure 11: Counterfactual outcomes when changing travel costs for Paducah, Ky. residents

(a) Change (%) in output quality δ_i



(b) Change (%) in patient market access Φ_i



Notes: This figure shows the impacts of a counterfactual 30% decrease in travel costs for Paducah residents ($\hat{\rho}_{ij} = 1.3 \ \forall i \neq$ Paducah). Panel (a) depicts the percentage change in quality of care δ_i produced in each region. Panel (b) depicts the percentage change in the value of market access Φ_i for patients who live in a region. The notes report the changes for Paducah.

Table 1: Scale elasticity estimates

All services	Baseline	No Diagonal	Controls
OLS: 2017	0.806	0.961	0.786
	(0.031)	(0.047)	(0.041)
OLS: 2013–2017 difference	0.999	1.045	1.018
	(0.079)	(0.083)	(0.082)
2SLS: population (log)	0.800	0.905	0.777
	(0.037)	(0.057)	(0.050)
	[2141]	[2141]	[1621]
2SLS: population (1940, log)	0.697	0.924	0.633
	(0.063)	(0.093)	(0.070)
	[163]	[163]	[206]

Notes: This table reports estimates of α from ordinary least squares (OLS) or two-stage least squares (2SLS) regressions of the form $\widehat{\ln \delta}_i = \alpha \ln Q_i + \ln R_i + \ln w_i + u_i$ using Hospital Referral Regions (HRRs) as the geographic units. The dependent variable $\widehat{\ln \delta}_i$ is estimated in equation (8) using a same-region dummy and a quadratic function of log distance. Q_i is region i's total production for Medicare beneficiaries, R_i is Medicare's Geographic Adjustment Factor, the w_i covariate includes mean two-bedroom property value and mean annual earnings for non-healthcare workers, and u_i is an error term. In the "no diagonal" column, S_{ii} observations were omitted when estimating $\widehat{\ln \delta}_i$ in equation (8). In the third column, the $\ln R_i$ and $\ln w_i$ controls are included in the regressions (coefficients not reported). The first-difference specification estimates $\Delta \widehat{\ln \delta}_i = \alpha \Delta \ln Q_i + \Delta u_i$. In the rows labeled "2SLS," we instrument for $\ln Q_i$ using the specified instruments. The standard errors in parentheses are robust to heteroskedasticity. For 2SLS estimates, first-stage effective F-statistics (Montiel Olea and Pflueger, 2013) are reported in square brackets. All estimates reveal substantial scale economies.

Table 2: Aggregate medical services exhibit a strong home-market effect

	(1)	(2)	(3)	(4)	(5)
				IV: 1940	2013–2017
	Cross-sectional PPML			population	panel
λ_X Provider-market population (log)	0.671	0.681	0.671	0.757	0.939
	(0.0543)	(0.0505)	(0.0366)	(0.0547)	(0.151)
λ_M Patient-market population (log)	0.260	0.252	0.286	0.284	-0.205
	(0.0547)	(0.0501)	(0.0346)	(0.0467)	(0.148)
Distance (log)	-1.627	0.344		0.377	
	(0.0489)	(0.304)		(0.250)	
Distance (log, squared)		-0.199		-0.201	
		(0.0305)		(0.0247)	
Distance (log) × 2017					-0.00117
					(0.00667)
p-value for H_0: $\lambda_X \leq \lambda_M$	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	93,636	93,636	93,636	93,636	162,678
Fixed effects					ij
Distance elasticity at mean		-1.59		-1.57	
Distance deciles			Yes		

Notes: This table reports estimates of equations (10) and (11), which evaluate the presence of weak or strong home-market effects. The sample is all HRR pairs ($N = 306^2$), multiplied by two for the last column. The dependent variable in all regressions is the value of trade when including professional and facility (inpatient and outpatient) fees at national average prices. Columns 1 through 4 use 2017 data. In column 1, the independent variables are patient- and provider-market log population, log distance between HRRs, and an indicator for same-HRR observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments. Column 5 presents estimates of equation (11) including ij fixed effects using data from 2013 and 2017. Section 1.1 and Appendices A.1 and A.2 detail how we compute trade flows in physician services (excluding emergency-room care and skilled nursing facilities) at standardized prices from the Medicare 20% carrier, 100% MedPAR, and 100% outpatient claims Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table 3: The home-market effect is stronger for rare procedures and diagnoses

	Procedure		Procedure		Procedure		Diagnosis	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
λ_X Provider-market population (log)	0.618	0.605	0.603		0.606		0.601	
	(0.0516)	(0.0493)	(0.0493)		(0.0488)		(0.0489)	
λ_M Patient-market population (log)	0.360	0.364	0.366		0.364		0.371	
	(0.0519)	(0.0492)	(0.0492)		(0.0486)		(0.0486)	
μ_X Provider-market population (log) $\times$ rare			0.344	0.329	0.362	0.317	0.120	0.110
			(0.0447)	(0.0405)	(0.0452)	(0.0392)	(0.0232)	(0.0206)
μ_M Patient-market population (log) $\times$ rare			-0.241	-0.239	-0.250	-0.220	-0.0986	-0.0915
			(0.0606)	(0.0587)	(0.0612)	(0.0564)	(0.0186)	(0.0172)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.005	0.006	0.007		0.005		0.008	
p-value for H_0: $\mu_X \leq \mu_M$			<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	187,272	110,402	110,402	110,402	110,402	110,402	109,658	109,658
Distance [linear] controls	Yes	Yes	Yes	Yes				
Distance [quadratic] controls					Yes	Yes	Yes	Yes
Patient-provider-market-pair FEs				Yes		Yes		Yes

Notes: This table reports estimates of equation (12), which introduces interactions with an indicator for whether a procedure or diagnosis is "rare" (provided less often than the median, when adding up all procedures provided nationally). The interactions with patient- and provider-market population reveal whether the home-market effect is larger for rare procedures/diagnoses. The unit of observation is {rare indicator, exporting HRR, importing HRR} so the number of observations is 2×306^2 in column 1, and the dependent variable in all regressions is the value of trade. Columns 1–6 reflect estimates by procedure and Columns 7 and 8 reflect estimates by diagnosis. Columns 2 onwards drop HRR pairs with zero trade in both procedure groups, and column 2 shows that this restriction has a negligible impact on the estimated log population coefficients. Columns 3 onwards include the rare indicator interacted with patient- and provider-market populations and distance covariates. Columns 1–4 control for distance using the log of distance between HRRs. Columns 5–8 add a control for the square of log distance. Columns 4, 6, and 8 introduce a fixed effect for each ij pair of patient market and provider market, so these omit all covariates that are not interacted with the rare indicator. The positive coefficient on provider-market population $\times$ rare across all columns indicates that the home-market effect is stronger for rare than for common services. The negative coefficient on patient-market population $\times$ rare across all columns indicates that the *strong* home-market effect has a larger magnitude for rare services. Valid primary diagnoses observed in 1,000 distinct claims or more nationally in the professional fees 20% sample are included. Section 1.1 and Appendices A.1 and A.2 detail how we compute these trade flows for physician services (excluding emergency-room care and skilled nursing facilities) at standardized prices from the Medicare 20% carrier, 100% MedPAR, and 100% outpatient claims Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Appendix – For Online Publication

Market Size and Trade in Medical Services

Jonathan I. Dingel, Joshua D. Gottlieb, Maya Lozinski, and Pauline Mourot

April 2024

A Data appendix

A.1 Data and processing

Medicare data. For patients enrolled in Traditional (fee-for-service) Medicare, our data contain a 100% sample of hospital inpatient and outpatient claims. For physician care (provided in the Carrier files), we have all claims from a random 20% sample of patients. Our main analysis uses data from 2017, but panel analysis uses data from 2013–2017.

One-third of Medicare patients opt out of the traditional version of Medicare, where care is paid directly by the government, in favor of a private insurance scheme ("Medicare Advantage"). In these private schemes, the government pays the insurer a fixed amount per patient and the insurers are responsible for the patient's care. Because Medicare does not pay claim-level bills in these private insurance schemes, the availability and quality of data for the privately insured patients is lower. We exclude these patients from our analysis.

Geography. We assign ZIP codes to hospital referral regions (HRRs) using a 2016 crosswalk from the Dartmouth Atlas of Healthcare. We assign ZIP codes to core-based statistical areas (CBSAs) by using the 2010 ZCTA to County Relationship File, updated county definitions, and the September 2018 definitions of CBSAs. For travel within an HRR, we measure the distance between the centroids of the patient's residential ZIP code and the ZIP code of the service location. We obtain the centroid coordinates from the Census Bureau's corresponding ZIP code tabulation areas (ZCTAs). For travel across HRRs, we use ZCTA-to-ZCTA distances when they are within 160 km, and (for computational ease) use HRR-to-HRR distances beyond 160 km.

Socioeconomic status. Our data do not contain patients' wealth or income. We use median household income in each patient's residential ZIP code to proxy for their socioeconomic status and estimate equation (8) separately for each decile of this proxy.

Physician earnings. We use data from Gottlieb et al. (2023) to measure the population elasticities of doctors' earnings and the American Community Survey (Ruggles et al., 2022) to examine other healthcare workers' earnings and real estate costs.The Gottlieb et al. (2023) earnings data depicted in Appendix Figure D.4 are only available for 111 commuting zones. The American Community Survey (ACS) covers far more CBSAs, but this source top-codes income for a substantial share of doctors. We use these same sources to compute physicians' share of the labor costs.

Physician specialties. Data come from the National Plan and Provider Enumeration System (NPPES) data, which cover all physicians, not just those serving Medicare patients. These data only report the number of doctors/specialists and their location, but contain no further information about procedures performed. We restrict attention to the 223 specializations within Allopathic & Osteopathic Physicians. We restrict attention to national provider identifiers of the "individual" entity type (as opposed to "organization"). We consider each physician's primary specialty, as indicated in the NPPES file. The results shown in Figure 7b look similar (not reported here) when we allow for multiple specialties per physician, a common occurrence in the NPPES data.

Capital equipment. The data on equipment use are obtained from the CMS Physician Fee Schedule Final Rule documentation for the calculation of practice expense Relative Value Units. For each HCPCS code, the data report the types of medical equipment that are used for that procedure. They also report the estimated price, useful life, utilization rate, used minutes per procedure, etc., of each piece of equipment (used minutes per procedure for the same piece of equipment can differ by HCPCS code).

We combine the CMS equipment requirement data with procedure frequency by geographic area, computed from the public use Medicare data, to obtain the "frequency of equipment use" for each type of equipment. Specifically, for each HRR, we locate all procedures that use a specific piece of equipment (equipment X), add up the number of billings for each of those procedures from that HRR, and interpret this total as the number of times equipment X is being used in that HRR. Similarly, we define the national frequency of equipment use by adding up the number of billings nationwide across all procedures that use equipment X. Finally, equipment use frequency per capita is defined as the equipment use frequency in a geographic unit divided by its population. We then estimate the population elasticity of *equipment use frequency per capita* at the HRR level, the same approach as we take for procedures and specialists. A population elasticity is estimated for each piece of equipment, for around 580 pieces of equipment (excluding outliers).

Chandra, Dalton, and Staiger (2023) hospital quality measures. Chandra, Dalton, and Staiger (2023) provide empirical Bayes estimates of hospital mortality across five large diagnostic cohorts: heart attacks, hip fractures, pneumonia, congestive heart failure, and strokes. We use the 2014 estimates as this is the closest year to our main analyses. They also report the number of episodes at the hospital-cohort-year level, which we use as weights when aggregating across hospitals. We map hospitals to HRRs using the Dartmouth Atlas (https://data.dartmouthatlas.org/supplemental/#hospital).

We aggregate the hospital-cohort level data to the HRR level in two steps. First, we compute the overall quality measure at the hospital level by regressing the empirical Bayes estimates on a set of hospital dummies and a set of cohort dummies; hospital-level estimates are simply the hospital fixed effects from this regression. Next, we aggregate across hospitals within an HRR by computing the average quality estimate across hospitals in that HRR, weighted by the number of episodes in each hospital. We standardize the exporter fixed effects before plotting them against our revealed-preference quality measures.

U.S. News and World Report. The publication produces an overall ranking and rankings for 12 particular specialties. We count the number of times each HRR's hospitals appear on any of these 13 lists.[53] Thus, higher ranking on the horizontal axis indicates a region has some combination of more top-ranked hospitals, or each of its hospitals performs well in many specialty areas.

The *U.S. News* rankings are intended to capture the "Best Hospitals," a concept associated with providing highly specialized care. So it is natural that there is a stronger relationship between the *U.S. News* rankings and exporter fixed effects for rare services; the slope in Figure D.3(b) is twice as large as that for common services in Figure D.3(a). In contrast, safety grades are not differentially relevant for rare services: Figures D.3(c) and D.3(d) show virtually identical slopes.

Depth to bedrock. Following Levy and Moscona (2020), we obtain 250m-cell resolution absolute depth to bedrock raster data from the International Soil Reference and Information Centre (ISRIC) SoilGrids 2017 data release. To define boundaries for commuting zones, we use the 2018-vintage cartographic boundary file for ZCTAs from the Census Bureau and match them to commuting zones using definitions from Fowler and Jensen (2020). Median depth to bedrock is computed over commuting zone boundaries which are based on the commonly used geodetic coordinate system EPSG:4326.

Clinical Classifications Software Refined (CCSR) diagnoses. We use the Clinical Classifications Software Refined (CCSR) diagnosis categories produced by the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project. CCSR aggregates over 70,000 ICD-10-CM diagnosis codes into "clinical categories," of which 482 have at least 20 patients each in our data. We split these categories at the median frequency to separate common from rare diagnoses.

[53] Results are similar when we use other methods to aggregate the rankings information, including when we account for the ordered nature of the lists.

A.2 Geographic price adjustments

Professional fees. To adjust for geographic price variation in the professional fees, we compute a national average price per Healthcare Common Procedure Coding System (HCPCS) code as the sum of the line allowed amount, which includes the line item's Medicare-paid and beneficiary-paid amounts (i.e., deductible, copayment, and coinsurance), divided by the sum of the line service count per HCPCS code nationally. We then apply this average price to all billing for the HCPCS code when computing total spending across services.

Hospital inpatient fees. For 2014–2017, we use the field "final standard payment amount" in the MedPAR file, which is computed as described in Finkelstein, Gentzkow, and Williams (2016) and Gottlieb et al. (2010). This represents "a standard Medicare payment amount, without the geographical payment adjustments and some of the other add-on payments that go to the hospitals" according to the data documentation. To ensure that total national spending remains identical, we multiply the standardized amount by a constant factor equal to the ratio of total national spending over the total standardized amount nationally. This field does not exist for 2013, the first year of our panel. Instead, we use the following process to infer the price adjustment from data available in 2013. We first compute the total expenditure as the sum of the Medicare payment amount, the primary payer amount, the pass-through amount, the Part A beneficiary co-insurance amount, the beneficiary deductible amount, and the beneficiary blood deductible amount for 2014 – the last year with the final standard payment amount field – and 2013. We then compute the ratio of the calculated total expenditure and the provided standard payment amount for 2014 and scale our total expenditures in 2013 by that ratio to infer the standardized expenditure amount. To test this process, we implement an analogous exercise using 2015 data to infer the standardized price for 2014, and find an $R^2 \geq 0.995$ between our computed variable and the one provided in the 2014 data.

Hospital outpatient fees. To adjust for geographic price variation in hospital outpatient fees, we compute a national average price per Healthcare Common Procedure Coding System (HCPCS) code, Ambulatory Payment Classifications (APC) code, and revenue center code. HCPCS codes reflect the procedure performed and APC codes reflect a prospective payment system applicable to outpatient analogous to Diagnosis Related Groups (DRGs) for inpatient claims. Revenue center contains information on the place of service, e.g. rehabilitation or acute care, so we consider two procedures performed in different revenue centers as different procedures for price adjustment purposes.

The total amount per claim line is calculated as the sum of the claim (Medicare) payment amount, the primary payer amount, the Part B beneficiary co-insurance amount, the beneficiary Part B deductible amount, and the beneficiary blood deductible amount. These amounts are summed nationally for each {HCPCS code, APC code, revenue center code} triplet, and divided by the frequency of that triplet to obtain a national average price. We then apply this average price to all instances of that {HCPCS code, APC code, revenue center code} combination when computing total spending across services. To ensure that total national spending remains identical, we multiply the price-adjusted measures by a constant factor equal to the ratio of the total national spending over the total price-adjusted spending nationally.

A.3 Procedure frequency in estimation sample matches other sources

Medicare provides two public-use files based on 100 percent claims data. Neither one is suitable for our main analysis because they do not contain patient location, so cannot be used to study trade flows. Instead, we use both of the public files to validate aspects of the confidential (20 percent) sample we must use for our main analysis.

The first public-use file contains the complete count of procedures billed by HCPCS code but does not have information about providers. We use it to confirm that procedure counts based on the confidential data do not suffer substantial sampling bias. In Figure D.7(a), we split procedure codes into deciles based on their national frequencies, separately in the confidential and public datasets. This generates a 100-cell matrix by decile pair. We plot the share of procedures in each cell in this matrix to determine how well the two datasets align. The vast majority of the codes are on the diagonal, with almost all of the remainder adjacent to the diagonal. This suggests that sampling error is not causing us to mischaracterize procedure frequency.

Medicare provides a second public file at the level of physician-by-procedure (HCPCS code). This summary does not contain any patient-level information so cannot be used to study trade flows, but we can use it to replicate analyses based on the location of production and physician experience. This file is censored such that physician-by-procedure pairs with 10 or fewer observations per year are suppressed, which makes for a more complicated bias than simple 20 percent random sampling. Nevertheless, all of the results that can be tested on this sample confirm those found in the 20 percent sample.

Since our procedure frequency measures rely on Medicare data, we would mismeasure frequency if the Medicare population uses a substantially different composition of care from the broader population. For example, childbirth is less common among Medicare beneficiaries. So our frequency measures may not capture the true national frequency of a procedure.

We address this by comparing procedure frequencies between the Medicare public data and private data from the Health Care Cost Institute (HCCI). The HCCI data contain claims for about 55 million privately insured patients (about 35% of individuals with employer-based insurance). We only consider HCPCS codes performed on at least eleven patients in the HCCI data. Note that frequencies are computed for all providers here, not only MDs and DOs. The authors acknowledge the assistance of the Health Care Cost Institute (HCCI) and its data contributors, Aetna, Humana, and Blue Health Intelligence, in providing the claims data analyzed in this section.

We examine whether procedures classified as above median frequency in one dataset are above median frequency in the other dataset. Table D.22 shows that 88% of the services above median frequency in Medicare are also as above median frequency in the HCCI data. Similarly, 82% of the services below median frequency in Medicare are also below median frequency in the HCCI data.

We next compare classifications of procedures' frequency deciles in Figure D.7(b). Analogous to Figure D.7(a), this plot visualizes the share of procedures which fall into each of pair of frequency decile bins in HCCI and Medicare data. The two classifications appear to coincide relatively well, with slightly stronger agreement for very frequent procedures compared to rare procedures in the Medicare public-use data. Overall, the frequency classifications of procedures coincide well between Medicare public-use data and HCCI data.

A.4 Residential measurement error

This appendix uses two methods to investigate potential measurement error in patients' residential locations. The first source of potential error is "snowbird" patients, who have multiple residences and therefore may appear to travel farther than they actually do. They may need medical care while spending months in a warmer HRR that is not the one listed as their main residence (or vice versa). Our results are robust to two methods of removing potential snowbirds: excluding Arizona, California, and Florida, following Finkelstein, Gentzkow, and Williams (2016), and excluding the 10% of HRRs with the highest share of second homes in American Community Survey data. These results are in Tables D.19 and D.20. The results are little changed by these sample restrictions.

We test for more general location measurement error by examining how far patients appear to travel for dialysis. Since Medicare patients requiring dialysis must generally visit a dialysis center thrice weekly, they are unlikely to go substantial distances for this service. Table D.21 compares travel distances for dialysis with other care. Dialysis patients appear to travel less than one-quarter as often as other patients—and even less when excluding snowbird states—suggesting that our residential location assignment is largely accurate.

A.5 Scale elasticity estimation with unobserved market segments

Our data only contain procedure-level production and consumption in Traditional Medicare (TM), not for Medicare Advantage (MA) or non-Medicare (NM) patients. We quantify how this biases our estimate of the scale elasticity, α, based on geographic variation. Suppose the production function is

$$\ln \delta_i = \alpha \ln Q_i + u_i,$$

where $Q_i = Q_i^{\mathrm{TM}} + Q_i^{\mathrm{MA}} + Q_i^{\mathrm{NM}}$ is the total quantity produced in region i, of which we only observe Q_i^{TM}. When we estimate the scale elasticity α using Q_i^{TM} as a proxy for Q_i, our regression coefficient may be biased:

$$\frac{\mathrm{Cov}(\ln \delta_i, \ln Q_i^{\mathrm{TM}})}{\mathrm{Var}(\ln Q_i^{\mathrm{TM}})} = \frac{\mathrm{Cov}(\alpha \ln Q_i, \ln Q_i^{\mathrm{TM}})}{\mathrm{Var}(\ln Q_i^{\mathrm{TM}})} + \frac{\mathrm{Cov}(u_i, \ln Q_i^{\mathrm{TM}})}{\mathrm{Var}(\ln Q_i^{\mathrm{TM}})} = \alpha\zeta,$$

where ζ, which governs the bias, is the regression coefficient from $\ln Q_i = \zeta \ln Q_i^{\mathrm{TM}} + \mathfrak{u}_i$.

To compute ζ we differentiate the identity $Q_i = Q_i^{\mathrm{TM}} \left(1 + \frac{Q_i^{\mathrm{MA}}}{Q_i^{\mathrm{TM}}} + \frac{Q_i^{\mathrm{NM}}}{Q_i^{\mathrm{TM}}}\right)$ with respect to Q_i^{TM}, which we observe:

$$\frac{\mathrm{d} \ln Q_i}{\mathrm{d} \ln Q_i^{\mathrm{TM}}} = 1 + s_i^{\mathrm{MA}} \varrho_i^{\mathrm{MA}} + s_i^{\mathrm{NM}} \varrho_i^{\mathrm{NM}},$$

where $s_i^{\mathrm{MA}} \equiv \frac{Q_i^{\mathrm{MA}}}{Q_i^{\mathrm{TM}} + Q_i^{\mathrm{MA}} + Q_i^{\mathrm{NM}}}$ is the MA share of production in region i, $\varrho_i^{\mathrm{MA}} \equiv \frac{\mathrm{d} \ln \frac{Q_i^{\mathrm{MA}}}{Q_i^{\mathrm{TM}}}}{\mathrm{d} \ln Q_i^{\mathrm{TM}}}$ is the TM production elasticity of relative production, and s_i^{NM} and ϱ_i^{NM} are similarly defined for non-Medicare (NM) insurance. To make it feasible to estimate these elasticities, we assume that they are constant across regions. If relative quantities produced are uncorrelated with the Traditional Medicare quantity produced ($\varrho^{\mathrm{MA}} = \varrho^{\mathrm{NM}} = 0$), then $\zeta = 1$ and $\alpha\zeta$ is an unbiased estimate of the scale elasticity α.[54] Otherwise, we need estimates of the average production shares $\bar{s}^{\mathrm{MA}}$ and $\bar{s}^{\mathrm{NM}}$ and the regression coefficients ϱ^{MA} and ϱ^{NM} to compute ζ.

[54] A special case would be if the quantity of care produced outside of TM is perfectly correlated with volume inside TM, so the shares s_i^{MA} and s_i^{NM} are constant.

We compute the production shares using data on aggregate expenditures and price deflators from prior research. Medicare, including both TM and MA, paid for $153 billion of the $525 billion spent nationally on physician services in 2017 (Centers for Medicare and Medicaid Services, 2022). Per capita spending and prices are similar between the two parts of Medicare (Berenson et al., 2015). Given this similarity, we apportion Medicare's production between TM and MA based on relative enrollment and obtain $\bar{s}^{\mathrm{MA}} = 0.111$. Next we consider Non-Medicare (NM) production. Private insurance spent $226 billion, which we deflate by a factor of 1.43 to account for the higher prices private insurance pays to make quantities comparable to Medicare (Lopez and Jacobson, 2020). Medicaid spent roughly $41 billion, which we deflate by its relative price of 0.72 (Zuckerman, Skopec, and Aarons, 2021). We incorporate other residual categories of production without price adjustments.[55] Combining these, we obtain an average $\bar{s}^{\mathrm{NM}} = 0.676$.

To estimate ϱ^{MA} and ϱ^{NM}, we assume that relative production is proportionate to relative resident beneficiaries. We obtain the number of TM beneficiaries and number of MA beneficiaries by HRR from Medicare enrollment data and compute the number of NM patients as total population minus Medicare enrollees.[56] Regressing the respective beneficiary ratios on log TM production yields $\hat{\varrho}^{\mathrm{MA}} = 0.089$ and $\hat{\varrho}^{\mathrm{NM}} = 0.075$. Putting these together means $\hat{\zeta} = 1.060$, so our estimated $\alpha\zeta = 0.80$ from Table 1 implies a scale elasticity of $\alpha = \frac{0.80}{1.060} = 0.75$.

[55] These other categories in the National Health Expenditure data are labeled Other Health Insurance Programs and Other Third Party Payers, along with out-of-pocket spending. Our simplifying approach here amounts to assuming Medicare prices for these residual categories.

[56] Ideally we would like to use the quantity of production in NM and MA markets, but we do not have this available at the HRR level. Beneficiaries might seem like a problematic proxy because the composition of NM beneficiaries varies widely across space, with some regions having a high Medicaid share and others a high private share. In aggregate, these two markets turn out to have similar per capita quantities of physician service spending: while private spending is $1,118 per capita and Medicaid spending is $550 per capita, the price adjustments mentioned above the quantities are relatively similar at $782 and $764, respectively, when valued at Medicare prices.

B Theory appendix

B.1 Monopolistic competition with one firm per region

Suppose that there is a single firm in each region that offers fixed-price services to patients under monopolistic competition with the firms in other regions. Assume $K(\delta_i) = \delta_i$ and $H(Q_i) = Q_i^\alpha$. The profit-maximizing choice of quality δ_i by the firm in region i is

$$\max_{\delta_i} \pi_i = \left(\bar{R} - \frac{w_i \delta_i}{A_i Q_i^\alpha}\right) Q_i \quad \text{where } Q_i = \sum_j Q_{ij} = \delta_i \sum_j \frac{N_j}{\Phi_j} \rho_{ij}$$

$$\frac{\partial \pi_i}{\partial \delta_i} = 0 \implies \frac{\bar{R}}{2-\alpha} = \frac{w_i \delta_i}{A_i Q_i^\alpha} = C(Q_i, \delta_i; w_i, A_i)$$

This expression replaces the free-entry condition (4) in the definition of equilibrium. Changing the value of the constant on the left side of this equality does not change any of the subsequent theoretical predictions. In this respect, the monopolistic-competition model with one firm per region is isomorphic to the perfect-competition model with external economies of scale.

B.2 Model with multiple types of patients

This section extends the model to feature multiple types of patients who face different trade costs. There is a finite set of patient types, which are indexed by κ. A patient type is defined by the trade costs $\rho_{ij(k)} = \rho_{ij}^\kappa \forall k \in \kappa$. Qualities δ_i, including the outside option δ_0, are the same for all patient types. The demand by patients of type κ residing in location j for procedures performed by providers in location i is now given by

$$Q_{ij}^\kappa = \frac{\delta_i N_j^\kappa}{\Phi_j^\kappa} \rho_{ij}^\kappa.$$

The aggregate gravity equation is the sum of type-specific gravity equations:

$$Q_{ij} = \sum_{\kappa} Q_{ij}^{\kappa} = \delta_i \sum_{\kappa} \frac{N_j^{\kappa}}{\Phi_j^{\kappa}} \rho_{ij}^{\kappa}. \tag{B.1}$$

The free-entry condition (4) remains unchanged with the introduction of multiple patient types:

$$R_i = \frac{w_i \delta_i}{A_i Q_i^{\alpha}}.$$

In equilibrium, market clearing requires that

$$Q_i = \left(\frac{w_i \delta_i}{A_i R_i}\right)^{1/\alpha} = \delta_i \sum_j \sum_{\kappa} \frac{N_j^{\kappa}}{\Phi_j^{\kappa}} \rho_{ij}^{\kappa} \implies \delta_i = \left(\frac{A_i R_i}{w_i}\right)^{1/(1-\alpha)} \left(\sum_j \sum_{\kappa} \frac{N_j^{\kappa}}{\Phi_j^{\kappa}} \rho_{ij}^{\kappa}\right)^{\alpha/(1-\alpha)}.$$

B.3 Derivations of results in Section 2.5

Abusing notation so that $\mathcal{I}$ is both the set and number of regions, equations (2) and (3) together constitute $2\mathcal{I}$ equations with $2\mathcal{I}$ unknowns. For the special case of $H(Q_i) = Q_i^{\alpha}$ and $K(\delta_i) = \delta_i$, this reduces to the following $\mathcal{I}$ equations with the unknowns $\{\delta_i\}_{i=1}^{\mathcal{I}}$:

$$\delta_i = \left(\frac{\overline{R} A_i}{w_i}\right)^{\frac{1}{1-\alpha}} \left(\sum_{j \in \mathcal{I}} \frac{\rho_{ij}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j}} N_j\right)^{\frac{\alpha}{1-\alpha}}$$

Following Costinot et al. (2019), we examine the home-market effect in the neighborhood of a symmetric equilibrium. For brevity, assume $\frac{\overline{R} A_i}{w_i} = 1 \; \forall i$. Note that at the symmetric equilibrium:

$$\bar{\delta}^{\frac{1-\alpha}{\alpha}} = \frac{1}{1 + \bar{\delta} + \sum_{i' \neq i} \bar{\delta} \rho} \bar{N} + \sum_{j \neq i} \frac{\rho}{1 + \bar{\delta} + \sum_{i' \neq j} \bar{\delta} \rho} \bar{N} = \frac{1 + (\mathcal{I} - 1)\rho}{\bar{\Phi}} \bar{N} = \frac{\bar{\Phi} - 1}{\bar{\Phi}} \frac{\bar{N}}{\bar{\delta}}. \tag{B.2}$$

Given $\alpha > 0$, totally differentiating the above system of equations in terms of $\{\mathrm{d}\delta_i, \mathrm{d}N_i\}_{i=1}^{\mathcal{I}}$

and evaluating it at the symmetric equilibrium yields the following expression:

$$\frac{\bar{\Phi}^2}{\bar{N}}\frac{1-\alpha}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}}\mathrm{d}\delta_i = -\left[\mathrm{d}\delta_i + \rho\sum_{i'\neq i}\mathrm{d}\delta_{i'}\right] + \bar{\Phi}\frac{\mathrm{d}N_i}{\bar{N}} + \sum_{j\neq i} -\rho\left[\mathrm{d}\delta_j + \rho\sum_{i'\neq j}\mathrm{d}\delta_{i'}\right] + \sum_{j\neq i}\rho\bar{\Phi}\frac{\mathrm{d}N_j}{\bar{N}}.$$

Given $\mathrm{d}N_1 > 0$ and $\mathrm{d}N_j = 0\ \forall j \neq 1$, we obtain the following expression for $\mathrm{d}\ln\delta_1$:

$$\mathrm{d}\ln\delta_1 = \frac{\frac{\bar{\Phi}}{\bar{\delta}}\mathrm{d}\ln N_1 - (\mathcal{I}-1)(2\rho + ((\mathcal{I}-2)\rho^2))\mathrm{d}\ln\delta_{j\neq 1}}{\frac{\bar{\Phi}^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} + 1 + (\mathcal{I}-1)\rho^2}. \tag{B.3}$$

Further tedious algebra delivers the following expression for quality changes:

$$\mathrm{d}\ln\delta_1 - \mathrm{d}\ln\delta_{j\neq 1} = \frac{(1-\rho)}{\frac{\bar{\Phi}^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} + (1-\rho)^2}\frac{\bar{\Phi}}{\bar{\delta}}\mathrm{d}\ln N_1. \tag{B.4}$$

Equation (B.2) implies that $\frac{\bar{\Phi}^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} = \left(\frac{1-\alpha}{\alpha}\right)\frac{\bar{\Phi}(\bar{\Phi}-1)}{\bar{\delta}}$ and therefore

$$\mathrm{d}\ln\delta_1 - \mathrm{d}\ln\delta_{j\neq 1} = \frac{(1-\rho)}{\left(\frac{1-\alpha}{\alpha}\right)\frac{\bar{\Phi}(\bar{\Phi}-1)}{\bar{\delta}} + (1-\rho)^2}\frac{\bar{\Phi}}{\bar{\delta}}\mathrm{d}\ln N_1 \quad = \left[\frac{1-\alpha}{\alpha}\frac{(\bar{\Phi}-1)}{(1-\rho)\bar{\delta}} + \frac{(1-\rho)\bar{\delta}}{\bar{\Phi}}\right]^{-1}\mathrm{d}\ln N_1 > 0.$$

The last expression above is reported in Section 2.5.

Prior to deriving the weak and strong home-market effects, we obtain an expression for $\frac{\mathrm{d}\ln\delta_j}{\mathrm{d}\ln N_1}$ for $j \neq 1$ around the symmetric equilibrium. Define $\bar{\mathcal{Q}} \equiv \frac{\bar{\Phi}^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} > 0$. Combining the expressions for $\mathrm{d}\ln\delta_1$ from equation (B.3) and for $\mathrm{d}\ln\delta_1 - \mathrm{d}\ln\delta_{j\neq 1}$ from equation (B.4) yields the following:

$$\frac{\mathrm{d}\ln\delta_{j\neq 1}}{\mathrm{d}\ln N_1} = \frac{\bar{\Phi}}{\bar{\delta}}\frac{\bar{\mathcal{Q}}\rho + \rho^3(\mathcal{I}-1) - \rho^2(\mathcal{I}-2) - \rho}{(\bar{\mathcal{Q}} + (1-\rho)^2)(\bar{\mathcal{Q}} + 1 + \rho^2 + 2\rho(\mathcal{I}-1) + \mathcal{I}\rho^2(\mathcal{I}-2))}$$

The weak home-market effect is derived as follows:

$$
\begin{aligned}
\ln Q_{1,j\neq 1} &= \alpha \ln Q_1 + \ln \rho - \ln \Phi_j + \ln N_j \\
\frac{\mathrm{d} \ln Q_{1,j\neq 1}}{\mathrm{d} \ln N_1} &= \alpha \frac{\mathrm{d} \ln Q_1}{\mathrm{d} \ln N_1} - \frac{\alpha}{\Phi_j}\left(\rho Q_1^{\alpha-1} \frac{\mathrm{d} Q_1}{\mathrm{d} \ln N_1} + Q_j^{\alpha-1} \frac{\mathrm{d} Q_j}{\mathrm{d} \ln N_1} + \rho \sum_{i'\neq 1,j} Q_{i'}^{\alpha-1} \frac{\mathrm{d} Q_{i'}}{\mathrm{d} \ln N_1} \right) \\
&= \frac{\mathrm{d} \ln \delta_1}{\mathrm{d} \ln N_1} - \frac{1}{\Phi_j}\left(\rho \delta_1 \frac{\mathrm{d} \ln \delta_1}{\mathrm{d} \ln N_1} + \delta_j \frac{\mathrm{d} \ln \delta_j}{\mathrm{d} \ln N_1} + \rho \sum_{i'\neq 1,j} \delta_{i'} \frac{\mathrm{d} \ln \delta_{i'}}{\mathrm{d} \ln N_1} \right) \\
&= \left(\frac{\bar{N} - Q_{1j}}{\bar{N}} \right) \frac{\mathrm{d} \ln \delta_1}{\mathrm{d} \ln N_1} - \left(\frac{\bar{N} - Q_{0j} - Q_{1j}}{\bar{N}} \right) \frac{\mathrm{d} \ln \delta_j}{\mathrm{d} \ln N_1} \\
&= \left(\frac{\bar{N} - Q_{1j}}{\bar{N}} \right) \left[\frac{\mathrm{d} \ln \delta_1}{\mathrm{d} \ln N_1} - \frac{\mathrm{d} \ln \delta_j}{\mathrm{d} \ln N_1} \right] + \frac{Q_{0j}}{\bar{N}} \frac{\mathrm{d} \ln \delta_j}{\mathrm{d} \ln N_1} \\
&= \frac{\Phi}{\bar{\delta} \bar{N}} \frac{1}{\bar{\mathcal{Q}} + (1-\rho)^2} \Big[(Q_{jj} + (\mathcal{I}-2) Q_{1j})(1-\rho) + \frac{Q_{0j}}{\bar{\mathcal{Q}} + 1 + \rho^2 + 2\rho(\mathcal{I}-1) + \mathcal{I}\rho^2(\mathcal{I}-2)} \\
&\quad \times \left\{ \bar{\mathcal{Q}} + (\rho-1)^2 + 2(\mathcal{I}-1)(\rho-\rho^2) + (\mathcal{I}-1)(\mathcal{I}-2)[\rho^2 - \rho^3] \right\} \Big] \\
&> 0.
\end{aligned}
$$

The condition for the strong home-market effect is derived as follows:

$$
\begin{aligned}
Q_{1,j\neq 1} - Q_{j\neq 1,1} &= \frac{Q_1^\alpha \rho}{1 + Q_1^\alpha \rho + Q_j^\alpha + \sum_{i\neq 1,j} Q_i^\alpha \rho} N_j - \frac{Q_j^\alpha \rho}{1 + Q_1^\alpha + Q_j^\alpha \rho + \sum_{i\neq 1,j} Q_i^\alpha \rho} N_1 \\
\mathrm{d} \ln Q_{1,j\neq 1} - \mathrm{d} \ln Q_{j\neq 1,1} &= \mathrm{d} \ln N_j - \mathrm{d} \ln N_1 + \alpha \left[1 + (1-\rho) \frac{\bar{Q}^\alpha}{\bar{\Phi}} \right] (\mathrm{d} \ln Q_1 - \mathrm{d} \ln Q_j) \\
&= -\mathrm{d} \ln N_1 + \left[1 + (1-\rho) \frac{\bar{\delta}}{\bar{\Phi}} \right] (\mathrm{d} \ln \delta_1 - \mathrm{d} \ln \delta_j) \\
&= \left[\frac{1 - \frac{1-\alpha}{\alpha} \frac{1+(\mathcal{I}-1)\rho}{1-\rho}}{\frac{1-\alpha}{\alpha} \frac{(1+(\mathcal{I}-1)\rho)}{(1-\rho)} + \frac{(1-\rho)\bar{\delta}}{1+(1+(\mathcal{I}-1)\rho)\bar{\delta}}} \right] \mathrm{d} \ln N_1.
\end{aligned}
$$

There is a strong home-market effect in the neighborhood of the symmetric equilibrium if and only if $\mathrm{d} \ln Q_{1,j\neq 1} - \mathrm{d} \ln Q_{j\neq 1,1} > 0$.

$$
\left[\frac{1 - \frac{1-\alpha}{\alpha} \frac{1+(\mathcal{I}-1)\rho}{1-\rho}}{\frac{1-\alpha}{\alpha} \frac{(1+(\mathcal{I}-1)\rho)}{(1-\rho)} + \frac{(1-\rho)\bar{\delta}}{1+(1+(\mathcal{I}-1)\rho)\bar{\delta}}} \right] \mathrm{d} \ln N_1 > 0 \iff \frac{\alpha}{1-\alpha} > \frac{1+(\mathcal{I}-1)\rho}{1-\rho}
$$

This is true if α is large enough and ρ is small enough.

Our rare-versus-common prediction concerns how the effect of market size on net exports varies with the number of potential patients $\bar{N}$. Given the scale elasticity α and (inverse) trade costs ρ, the denominator of the right side of equation (6) is increasing in the symmetric-equilibrium quality $\bar{\delta}$. For two procedures that both exhibit a strong home-market effect because they have the same scale elasticity and trade costs, the effect of population size on net exports will be larger for the procedure with lower service quality. The symmetric-equilibrium service quality is increasing in the number of potential patients $\bar{N}$ because there are increasing returns (see equation (B.2)). Thus, in the neighborhood of the symmetric equilibrium, the strength of a strong home-market effect is decreasing in the number of potential patients.

C Details of counterfactual calculations

Section C.1 describes how we compute counterfactual equilibrium outcomes relative to baseline equilibrium outcomes in the model. Section C.2 describes the assumptions we make to infer the number of potential patients N_j and hence import shares m_{ij}, which are inputs into these calculations. Section C.3 describes how to compute counterfactual outcomes in the model when there are multiple (observed) types of patients who differ in their trade costs. Section C.4 describes how we infer the number of potential patients of each type.

C.1 Computing equilibrium outcomes in counterfactual scenarios

We compute counterfactual equilibrium outcomes relative to baseline equilibrium outcomes by rewriting the equilibrium system of equations in terms of the initial allocation, constant elasticities, relative exogenous parameters, and relative endogenous equilibrium outcomes, a technique known as "exact hat algebra" in the trade literature.

If $K(\delta) = \delta$ and $H(Q) = Q^{\alpha}$, an equilibrium is a set of quantities and qualities $\{Q_i, \delta_i\}_{i \in \mathcal{I}}$

that simultaneously satisfy equations (4) and (1) and $Q_i = \sum_j Q_{ij}$. Consider two equilibria: the baseline equilibrium and the counterfactual equilibrium. Define export shares $x_{ij} \equiv \frac{Q_{ij}}{\sum_{j'} Q_{ij'}}$ and import shares $m_{ij} \equiv \frac{Q_{ij}}{N_j}$ in the baseline equilibrium. Denote the counterfactual parameters and equilibrium outcomes by primes. Plugging $Q_i = \sum_j Q_{ij}$ into equation (4), we can write the system of equations for each equilibrium as

$$\begin{array}{ll} \delta'_i = \left(\frac{R'_i A'_i}{w'_i}\right)\left(\sum_j Q'_{ij}\right)^{\alpha} & \qquad Q'_{ij} = \delta'_i \frac{\rho'_{ij}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta'_{i'} \rho'_{i'j}} N'_j \\ \delta_i = \left(\frac{R_i A_i}{w_i}\right)\left(\sum_j Q_{ij}\right)^{\alpha} & \qquad Q_{ij} = \delta_i \frac{\rho_{ij}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j}} N_j \end{array} .$$

Define $\hat{y} \equiv \frac{y'}{y}$ for every variable y. For example, $\hat{\delta}_i \equiv \frac{\delta'_i}{\delta_i}$.

We now rewrite the counterfactual equilibrium equations in terms of baseline equilibrium shares x_{ij}, m_{ij}, the scale elasticity α, (relative) counterfactual exogenous parameters $\hat{A}, \hat{R}, \hat{w}, \hat{\rho}, \hat{N}$, and (relative) counterfactual endogenous qualities $\hat{\delta}$.

First, divide the counterfactual free-entry condition by the baseline free-entry condition to obtain an expression for relative quality:

$$\frac{\delta'_i}{\delta_i} = \frac{\hat{R}_i \hat{A}_i}{\hat{w}_i}\left(\frac{\sum_{j \in \mathcal{I}} Q'_{ij}}{\sum_{j \in \mathcal{I}} Q_{ij}}\right)^{\alpha} = \frac{\hat{R}_i \hat{A}_i}{\hat{w}_i}\left(\sum_{j \in \mathcal{I}} \frac{Q_{ij}}{\sum_{j \in \mathcal{I}} Q_{ij}} \frac{Q'_{ij}}{Q_{ij}}\right)^{\alpha} = \frac{\hat{R}_i \hat{A}_i}{\hat{w}_i}\left(\sum_{j \in \mathcal{I}} x_{ij} \frac{Q'_{ij}}{Q_{ij}}\right)^{\alpha} \tag{C.1}$$

Second, divide the counterfactual gravity equation by the baseline gravity equation to obtain an expression for relative bilateral flows:

$$\begin{aligned} \frac{Q'_{ij}}{Q_{ij}} &= \frac{\delta'_i}{\delta_i}\left(\frac{\frac{\rho'_{ij}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta'_{i'} \rho'_{i'j}} N'_j}{\frac{\rho_{ij}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j}} N_j}\right) = \frac{\frac{\delta'_i}{\delta_i} \frac{\rho'_{ij}}{\rho_{ij}} \frac{N'_j}{N_j}}{\sum_{i' \in 0 \cup \mathcal{I}} \frac{\delta_{i'} \rho_{i'j}}{\sum_{i' \in 0 \cup \mathcal{I}} \delta_{i'} \rho_{i'j}} \frac{\delta'_{i'} \rho'_{i'j}}{\delta_{i'} \rho_{i'j}}} \\ &= \frac{\hat{\delta}_i \hat{\rho}_{ij} \hat{N}_j}{\sum_{i' \in 0 \cup \mathcal{I}} \frac{Q_{i'j}}{N_j} \hat{\delta}_{i'} \hat{\rho}_{i'j}} = \frac{\hat{\delta}_i \hat{\rho}_{ij} \hat{N}_j}{m_{0j} + \sum_{i' \in \mathcal{I}} m_{i'j} \hat{\delta}_{i'} \hat{\rho}_{i'j}} \end{aligned}$$

Plug this expression for relative bilateral flows into equation (C.1) and rearrange terms to obtain the following system of $\mathcal{I}$ equations with unknowns $\{\hat{\delta}_i\}_{i=1}^{\mathcal{I}}$:

$$\hat{\delta}_i = \left(\hat{R}_i \hat{A}_i / \hat{w}_i\right)^{\frac{1}{1-\alpha}} \left(\sum_{j \in \mathcal{I}} \frac{x_{ij} \hat{\rho}_{ij} \hat{N}_j}{m_{0j} + \sum_{i' \in \mathcal{I}} m_{i'j} \hat{\delta}_{i'} \hat{\rho}_{i'j}} \right)^{\frac{\alpha}{1-\alpha}}. \tag{C.2}$$

C.2 Inferring the number of potential patients

A baseline calibration of our model requires α, x_{ij}, and m_{ij} in order to use equation (C.2) to compute relative counterfactual outcomes. We have estimated α. The export shares $x_{ij} \equiv \frac{Q_{ij}}{\sum_j Q_{ij}}$ are easily computed using the observed trade matrix.[57] The challenge is computing import shares $m_{ij} \equiv \frac{Q_{ij}}{N_j}$ because we do not observe N_j; while we observe the number of Medicare beneficiaries in region j, not all beneficiaries are in the market for all services. This section describes the assumptions we make in order to infer the values of the relevant market size $N_j \; \forall j \in \mathcal{I}$. Specifically, we assume per capita demand is uniform, outside-option quality is constant across regions, and the average outside-option share is 10%, as described below.

We have estimated $\theta_j = N_j/\Phi_j$ in equation (8). We observe the number of beneficiaries enrolled in Traditional Medicare in region j, which we denote S_j^{TM}. By definition, $m_{0j} = \frac{\delta_{0j}}{\Phi_j}$. We assume $\delta_{0j} = \delta_0 \; \forall j$ and $N_j \propto S_j^{\text{TM}}$. This implies

$$m_{0j} = \frac{\delta_{0j}}{\Phi_j} = \frac{\delta_0 \theta_j}{N_j} = \frac{\delta_0 \theta_j}{\mathfrak{s} S_j^{\text{TM}}},$$

where $\mathfrak{s}$ is a constant of proportionality. We set $\frac{\delta_0}{\mathfrak{s}}$ such that the average outside-option share is 10%, $\frac{1}{\mathcal{I}} \sum_j m_{0j} = 0.1$. This requires $\frac{\delta_0}{\mathfrak{s}} = \frac{0.1 \times \mathcal{I}}{\sum_j \theta_j / S_j^{\text{TM}}}$. With m_{0j} in hand, we can infer N_j:

$$m_{0j} = 1 - \sum_{i \in \mathcal{I}} m_{ij} = 1 - \frac{1}{N_j} \sum_{i \in \mathcal{I}} Q_{ij} \implies N_j = \frac{1}{1 - m_{0j}} \sum_{i \in \mathcal{I}} Q_{ij}.$$

With N_j in hand, we can compute all import shares, $m_{ij} = \frac{Q_{ij}}{N_j} \; \forall i \in 0 \cup \mathcal{I}, \forall j \in \mathcal{I}$.

[57] Dingel and Tintelnot (2021) document overfitting problems when calibrating gravity models using noisy observed shares. We obtain similar counterfactual outcomes when calibrating our model using gravity-predicted shares.

We exclude the Anchorage, Alaska and Honolulu, Hawaii HRRs from our counterfactual computations. These geographically isolated and expansive HRRs cover the entire states of Alaska and Hawaii. The average within-Anchorage-HRR procedure incurs more than 53 kilometers of travel. In the gravity regression, these HRRs have the smallest exporter fixed effects: very few patients travel to Alaska or Hawaii for care. Alaska's importer fixed effect is quite large because Alaskans import about 19% of their services and the average import traverses 3,622 kilometers (6% and 5,496 km for Hawaii). As a result, their implied outside-option shares would exceed one when we set the nationwide average to 10%. We therefore exclude the Alaska and Hawaii HRRs from the economy when computing counterfactual outcomes. Given their considerable geographic isolation, Alaska and Hawaii would have little influence on outcomes in other regions.

The qualitative and spatial patterns of counterfactual outcomes are the same if we assume the average outside-option share is 5% or 20% rather than 10%.

C.3 Counterfactual outcomes with multiple patient types

This section describes how to compute counterfactual equilibrium outcomes relative to baseline equilibrium outcomes when there are multiple patient types who face heterogeneous trade costs. The derivation is very similar to that of Section C.1. Define import shares $m_{ij\kappa} \equiv \frac{Q_{ij\kappa}}{N_{j\kappa}}$ in the baseline equilibrium. Define patient-type shares $n_{j\kappa} \equiv \frac{N_{j\kappa}}{N_j}$. We rewrite the system of baseline and counterfactual gravity equations (B.1) and free-entry condition (4) as follows:

$$\begin{array}{ll} \delta_i' = \left(\frac{R_i' A_i'}{w_i'}\right)\left(\sum_j Q_{ij}'\right)^\alpha & \qquad Q_{ij}' = \delta_i' \sum_\kappa \frac{\rho_{ij\kappa}'}{\sum_{i'\in 0\cup\mathcal{I}} \delta_{i'}' \rho_{i'j\kappa}'} N_{j\kappa}' \\ \delta_i = \left(\frac{R_i A_i}{w_i}\right)\left(\sum_j Q_{ij}\right)^\alpha & \qquad Q_{ij} = \delta_i \sum_\kappa \frac{\rho_{ij\kappa}}{\sum_{i'\in 0\cup\mathcal{I}} \delta_{i'} \rho_{i'j\kappa}} N_{j\kappa} \end{array}.$$

As above, dividing the counterfactual free-entry condition by the baseline free-entry condition yields the expression for relative quality in equation (C.1). Second, divide the counterfactual gravity equation by the baseline gravity equation to obtain an expression for relative bilateral flows:

$$\frac{Q'_{ij}}{Q_{ij}} = \frac{\delta'_i}{\delta_i}\left(\frac{\sum_\kappa \frac{\rho'_{ij\kappa}}{\sum_{i'\in 0\cup\mathcal{I}}\delta'_{i'}\rho'_{i'j\kappa}}N'_{j\kappa}}{\sum_\kappa \frac{\rho_{ij\kappa}}{\sum_{i'\in 0\cup\mathcal{I}}\delta_{i'}\rho_{i'j\kappa}}N_{j\kappa}}\right) = \frac{\delta'_i}{\delta_i}\left(\frac{\sum_\kappa \frac{\rho'_{ij\kappa}}{\Phi'_{j\kappa}}N'_{j\kappa}}{\sum_\kappa \frac{\rho_{ij\kappa}}{\Phi_{j\kappa}}N_{j\kappa}}\right) = \hat{\delta}_i\sum_\kappa \frac{n_{j\kappa}}{\frac{m_{ij}}{m_{ij\kappa}}}\frac{\hat{\rho}_{ij\kappa}}{\hat{\Phi}_{j\kappa}}\hat{N}_{j\kappa}$$

Plugging this expression for relative bilateral flows into equation (C.1) and then rearranging terms yields the following system of $\mathcal{I}$ equations with unknowns $\{\hat{\delta}_i\}_{i=1}^{\mathcal{I}}$:

$$\hat{\delta}_i = \left(\hat{R}_i\hat{A}_i/\hat{w}_i\right)^{\frac{1}{1-\alpha}}\left(\sum_{j\in\mathcal{I}} x_{ij}\left(\sum_\kappa \frac{n_{j\kappa}}{m_{ij}}\frac{m_{ij\kappa}\hat{\rho}_{ij\kappa}}{m_{0j\kappa}+\sum_{i'\in\mathcal{I}} m_{i'j\kappa}\hat{\delta}_{i'}\hat{\rho}_{i'j\kappa}}\hat{N}_{j\kappa}\right)\right)^{\frac{\alpha}{1-\alpha}}. \tag{C.3}$$

C.4 Inferring the number of potential patients of each type

Because we do not observe patients who select the outside option, we make assumptions that allow us to infer $N_{j\kappa}$ and thus $n_{j\kappa}$ and $m_{ij\kappa}$, which are needed to compute counterfactual outcomes using equation (C.3). We start from a type-specific variant of the gravity equation (7) with fixed effects, as in the single-type equation (8). The estimating equation is

$$\ln\mathbb{E}\left[S_{ij\kappa}\right] = \ln\delta_i + \ln\left(\frac{N_{j\kappa}}{\Phi_{j\kappa}}\right) + \gamma^\kappa X_{ij} = \ln\delta_i + \ln\theta_{j\kappa} + \gamma^\kappa X_{ij}.$$

This yields an estimate of $\theta_{j\kappa} = N_{j\kappa}/\Phi_{j\kappa}$.

As in the single-type case above, we assume per capita demand is uniform and outside-option quality is constant across regions. We observe the number of beneficiaries of type κ enrolled in Traditional Medicare in region j, which we denote $S^{\mathrm{TM}}_{j\kappa}$. We assume $\delta_{0j} = \delta_0\ \forall j$ and $N_{j\kappa} = \mathfrak{s}S^{\mathrm{TM}}_{j\kappa}$, where $\mathfrak{s}$ is a constant of proportionality that is common across types. This implies

$$m_{0j\kappa} = \frac{\delta_0}{\Phi_{j\kappa}} = \frac{\delta_0\theta_{j\kappa}}{N_{j\kappa}} = \frac{\delta_0\theta_{j\kappa}}{\mathfrak{s}S^{\mathrm{TM}}_{j\kappa}}.$$

Let $\mathcal{K} = \sum_\kappa 1$ denote the number of patient types. We set $\frac{\delta_0}{\mathfrak{s}}$ such that the average outside-option share, across all types, is 10%, $\frac{1}{\mathcal{IK}}\sum_{j\kappa} m_{0j\kappa} = 0.1$. This implies

$$m_{0j\kappa} = 0.1 \times \frac{\theta_{j\kappa}/S^{\mathrm{TM}}_{j\kappa}}{\frac{1}{\mathcal{IK}}\sum_{j'\kappa'} \theta_{j'\kappa'}/S^{\mathrm{TM}}_{j'\kappa'}}.$$

Using the resulting $N_{j\kappa} = \frac{1}{1-m_{0j\kappa}}\sum_{i\in\mathcal{I}} Q_{ij\kappa}$ allows us to compute all import shares.

C.5 Comparison of geographic scope between United States and Germany

To summarize differences in geography, we compute a distance-weighted sum of population in other regions, regional "market potential" $\mathrm{MP}_i \equiv \sum_{j\neq i} \rho_{ij} N_j$, for regions in each country. The counterfactual trade costs that make the US distribution of market potential match that of Germany involve scaling the interregional trade costs ρ_{ij} by a symmetric matrix $\hat{\rho}_{ij}$ with an average value of 3.6 and a variance of 1.7.

D Additional exhibits

Table D.1: Summary statistics on HRRs and trade between HRR pairs

Panel A: HRRs (N=306)	Mean	Std. Dev	Min.	Max.
Production (millions \$)	751.1	796.2	66.4	5,204.2
Consumption (millions \$)	751.1	725.0	83.0	4,607.3
Exports (millions \$)	140.7	176.6	6.7	1,570.6
Imports (millions \$)	140.7	99.1	25.6	688.6
Exports as a share of production (%)	19.3	10.9	2.0	70.3
Imports as a share of consumption (%)	24.0	10.7	4.6	59.0
Population (millions)	1.06	1.21	0.13	10.13
Traditional Medicare beneficiaries (millions)	0.13	0.12	0.02	0.77
Log production (\$)	20.04	0.87	18.01	22.37
Log quality $\widehat{\ln \delta_i}$	-0.67	0.89	-5.52	2.18
Panel B: Trade between HRRs (N=93,330)	**Mean**	**Std. Dev**		
Log bilateral trade (\$)	9.4	2.7		
Bilateral trade: facility fees (thousands \$)	314	3,637		
Bilateral trade: professional fees (thousands \$)	148	1,707		
Trade-weighted log distance (km)	4.9	1.2		

Notes: Panel A of this table reports summary statistics describing the 306 HRRs. The level of log quality $\widehat{\ln \delta_i}$ is normalized such that the production-weighted mean is zero. Panel B reports summary statistics describing trade between the $93,330 = 306 \times 305$ HRR pairs. Production, consumption, and trade flows are computed from the Medicare 20% carrier, 100% MedPar, and 100% outpatient Research Identifiable Files. All dollar statistics for professional fees are multiplied by 5 to represent the full Traditional Medicare population. HRR definitions are from the Dartmouth Atlas Project.

Table D.2: Revealed-preference quality measure $\hat{\delta}_i$ is correlated with clinical quality

	(1)	(2)	(3)	(4)
USN sum of inverse rank				0.042
				(0.113)
USN number of mentions in top lists			0.080	0.028
			(0.016)	(0.015)
JC share of hospitals with accreditation				-0.021
				(0.147)
JC log number of accredited hospitals				0.595
				(0.055)
LF average safety grade for HRR				0.140
				(0.053)
CMS overall hospital quality				-0.043
				(0.102)
CMS overall hospital quality (weighted by beds)				0.011
				(0.104)
CMS overall mortality rate across conditions		-0.459		-0.165
		(0.061)		(0.070)
CDS log length of stay (empirical Bayes)				-1.619
				(0.809)
CDS log 30-day inpatient (Part A) costs (empirical Bayes)				0.245
				(0.337)
CDS 30-day mortality rate (empirical Bayes)	-0.213			-0.001
	(0.047)			(0.047)
CDS 30-day readmissions (empirical Bayes)				4.091
				(3.998)
Constant	-0.654	4.539	-0.832	-0.570
	(0.047)	(0.703)	(0.048)	(0.964)
Observations	305	306	306	305
R-squared	0.063	0.138	0.288	0.615
Adj R-squared	0.060	0.136	0.285	0.599

Notes: This table reports regression estimates of our revealed-preference regional quality measure $\hat{\delta}_i$ against external quality measures. Acronyms in each row refer to sources of quality measures. USN: *U.S. News*; JC: Joint Commission; LF: Leapfrog Group; CMS: Centers for Medicare and Medicaid Services; CDS: Chandra, Dalton, and Staiger (2023).

Table D.3: Scale elasticity estimates for CBSAs

All services	Baseline	No Diagonal	Controls
OLS: 2017	0.929	1.201	0.934
	(0.012)	(0.019)	(0.015)
OLS: 2013–2017 difference	0.873	1.176	0.864
	(0.039)	(0.052)	(0.033)
2SLS: population (log)	0.912	1.189	0.913
	(0.014)	(0.019)	(0.018)
	[5671]	[5671]	[3522]
2SLS: population (1940, log)	0.951	1.100	0.940
	(0.021)	(0.025)	(0.026)
	[912]	[912]	[784]
2SLS: bedrock depth	0.843	0.660	0.880
	(0.067)	(0.122)	(0.077)
	[39.4]	[39.4]	[27.5]

Notes: This table reports estimates of α from ordinary least squares (OLS) or two-stage least squares (2SLS) regressions of the form $\widehat{\ln \delta_i} = \alpha \ln Q_i + \ln R_i + \ln w_i + u_i$ using core-based statistical areas (CBSAs) as the geographic units. The dependent variable $\widehat{\ln \delta_i}$ is estimated in equation (8), Q_i is region i's total production for Medicare beneficiaries, R_i is Medicare's Geographic Adjustment Factor, the w_i covariate includes mean two-bedroom property value and mean annual earnings for non-healthcare workers, and u_i is an error term. In the rows labeled "2SLS" we instrument for $\ln Q_i$ using the specified instruments. In the second column, labeled "No Diagonal," S_{ii} observations were omitted when estimating $\widehat{\ln \delta_i}$ in equation (8). In the third column, the $\ln R_i$ and $\ln w_i$ controls are included in the regressions (coefficients not reported). Standard errors (in parentheses) are robust to heteroskedasticity. For 2SLS estimates, first-stage effective F-statistics (Montiel Olea and Pflueger, 2013) are reported in square brackets.

Table D.4: Scale elasticity estimates (commuting zones)

All services	Baseline	No Diagonal	Controls
OLS: 2017	0.936	0.959	0.953
	(0.016)	(0.027)	(0.016)
OLS: 2013–2017 difference	0.937	1.043	0.884
	(0.036)	(0.105)	(0.032)
2SLS: population (log)	0.909	0.967	0.934
	(0.014)	(0.019)	(0.017)
	[2203]	[2895]	[1675]
2SLS: population (1940, log)	0.928	0.899	0.940
	(0.018)	(0.024)	(0.020)
	[984]	[1124]	[838]
2SLS: bedrock depth	0.758	0.804	0.765
	(0.040)	(0.062)	(0.056)
	[76.6]	[80.6]	[35.7]

Notes: This table reports estimates of α from ordinary least squares (OLS) or two-stage least squares (2SLS) regressions of the form $\widehat{\ln \delta_i} = \alpha \ln Q_i + u_i$ using commuting zones (CZs) as the geographic units. They are analogous to Table 1. The dependent variable $\widehat{\ln \delta_i}$ is estimated in equation (8), Q_i is region i's total production for Medicare beneficiaries, and u_i is an error term. In the "no diagonal" column, S_{ii} observations were omitted when estimating $\widehat{\ln \delta_i}$ in equation (8). The first-difference specification estimates $\Delta\widehat{\ln \delta_i} = \alpha \Delta \ln Q_i + \Delta u_i$. In the rows labeled "2SLS," we instrument for $\ln Q_i$ using the specified instruments. The standard errors in parentheses are robust to heteroskedasticity. For 2SLS estimates, first-stage effective F-statistics (Montiel Olea and Pflueger, 2013) are reported in square brackets. All estimates reveal substantial scale economies.

Table D.5: Higher-income patients are less sensitive to distance: Procedure-level estimates

	(1) 25min visit	(2) cataract removal	(3) knee joint repair	(4) heart artery bypass	(5) gallblader removal
Distance (log)	-1.929	-2.202	-2.228	-2.177	-2.022
	(0.0800)	(0.104)	(0.0837)	(0.0866)	(0.0819)
Distance (log) × income tercile 2	0.118	0.152	0.214	0.185	0.249
	(0.0640)	(0.0884)	(0.0713)	(0.0856)	(0.0723)
Distance (log) × income tercile 3	0.208	0.202	0.293	0.455	0.367
	(0.0842)	(0.110)	(0.0855)	(0.0973)	(0.0931)
Observations	271,422	268,974	263,772	245,632	252,144
Patient market-income FE & Provider market FE	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the coefficient on log distance for each income tercile from gravity regressions estimated separately for five procedures varying in frequency: 25 min office visit (HCPCS 99214), cataract removal (66984), knee joint repair (27447), heart artery bypass (33533), and gallblader removal (47562). The dependent variable in all regressions is the number of procedures traded. Each regression includes log distance interacted with an income tercile indicator, an indicator for same-HRR observations ($i = j$), an exporting HRR fixed effect, and an income-tercile-importing-HRR fixed effect. The coefficients for higher income terciles are positive, indicating that patients residing in higher-income ZIP codes are less sensitive to distance. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.6: Estimates of a strong home-market effect by CBSA

	(1)	(2)	(3)	(4)	(5)	(6)
				IV: 1940	IV:	2013–2017
	Cross-sectional PPML			population	Bedrock	panel
λ_X Provider-market population (log)	0.806	0.813	0.753	0.807	1.273	0.864
	(0.0261)	(0.0252)	(0.0237)	(0.0237)	(0.377)	(0.178)
λ_M Patient-market population (log)	0.242	0.237	0.287	0.257	-0.115	0.184
	(0.0322)	(0.0316)	(0.0299)	(0.0295)	(0.465)	(0.148)
Distance (log)	-2.372	-3.565		-3.571	-5.121	
	(0.0509)	(0.338)		(0.296)	(1.468)	
Distance (log, squared)		0.114		0.115	0.240	
		(0.0341)		(0.0290)	(0.114)	
Distance (log) × 2017						-0.00255
						(0.00653)
p-value for H_0: $\lambda_X \leq \lambda_M$	<0.001	<0.001	<0.001	<0.001	0.046	0.010
Observations	857,476	857,476	857,476	857,476	781,456	472,374
Fixed effects						ij
Distance elasticity at mean		-2.46		-2.46	-2.79	
Distance deciles			Yes			

Notes: This table reports estimates of equations (10) and (11), which evaluate the presence of weak or strong home-market effects, using CBSAs as the geographic units. They are analogous to the HRR results reported in Table 2. The dependent variable in all regressions is the value of trade computed by assigning each procedure its national average price. The independent variables are patient- and provider-market log population, log distance between CBSAs, and an indicator for same-CBSA observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments. Column 5 presents estimates using panel variation from 2013–2017 and includes ij fixed effects. Trade flows are computed from the Medicare 20% carrier, 100% MedPar, and 100% outpatient Research Identifiable Files. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.7: Estimates of a strong home-market effect (commuting zones)

	(1)	(2)	(3)	(4) IV: 1940 population	(5) IV: Bedrock	(6) 2013–2017 panel
	Cross-sectional PPML					
λ_X Provider-market population (log)	0.857	0.858	0.790	0.841	1.197	1.048
	(0.0281)	(0.0283)	(0.0259)	(0.0310)	(0.316)	(0.166)
λ_M Patient-market population (log)	0.118	0.116	0.212	0.144	-0.140	-0.0618
	(0.0272)	(0.0270)	(0.0270)	(0.0306)	(0.206)	(0.145)
Distance (log)	-2.458	-3.310		-3.257	-3.483	
	(0.0752)	(0.514)		(0.528)	(0.566)	
Distance (log, squared)		0.0788		0.0751	0.0835	
		(0.0468)		(0.0472)	(0.0472)	
Distance (log) × 2017						-0.0134
						(0.00664)
p-value for H_0: $\lambda_X \leq \lambda_M$	<0.001	<0.001	<0.001	<0.001	0.004	<0.001
Observations	390,625	390,625	390,625	390,625	390,625	326,118
Fixed effects						ij
Distance elasticity at mean		-2.51		-2.50	-2.64	
Distance deciles			Yes			

Notes: This table reports estimates of equations (10) and (11), which evaluate the presence of weak and strong home-market effects, using commuting zones (CZs) as the geographic units. They are analogous to the HRR results reported in Table 2. The dependent variable in all regressions is the value of trade computed by assigning each procedure its national average price. The independent variables are patient- and provider-market log population, log distance between CZs, and an indicator for same-CZ observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments. Trade flows are computed from the Medicare 20% carrier, 100% MedPar, and 100% outpatient Research Identifiable Files. CZ definitions are from Fowler and Jensen (2020). Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.8: Estimates of a strong home-market effect for professional fees

	(1)	(2)	(3)	(4)	(5)
				IV: 1940	2013–2017
	Cross-sectional PPML			population	panel
λ_X Provider-market population (log)	0.618	0.622	0.621	0.566	0.951
	(0.0516)	(0.0506)	(0.0383)	(0.0673)	(0.161)
λ_M Patient-market population (log)	0.360	0.357	0.380	0.369	-0.119
	(0.0519)	(0.0507)	(0.0374)	(0.0482)	(0.173)
Distance (log)	-1.621	-0.721		-0.737	
	(0.0495)	(0.274)		(0.239)	
Distance (log, squared)		-0.0903		-0.0892	
		(0.0269)		(0.0234)	
Distance (log) × 2017					0.0144
					(0.00693)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.005	0.004	<0.001	0.031	<0.001
Observations	93,636	93,636	93,636	93,636	129,202
Fixed effects					*ij*
Distance elasticity at mean		-1.60		-1.60	
Distance deciles			Yes		

Notes: This table reports estimates of equation (10) and (11), which estimates the presence of weak or strong home-market effects. The sample is all HRR pairs ($N = 306^2$), and the dependent variable in all regressions is the value of trade for professional fees. The independent variables are patient- and provider-market log population, log distance between HRRs, and an indicator for same-HRR observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments (GMM). Column 5 estimates equation (10), which uses panel variation in population by controlling for importer-exporter-pair fixed effects. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files, using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.9: Healthcare labor costs by occupation

Category	Number of Workers	Average Earnings ($)	Total Spending ($ Millions)
Other Healthcare workers	7,642,379	51,843	396,200
Physicians	677,431	294,560	199,544

Notes: This table reports the number of physicians and other healthcare workers and their average earnings using American Community Survey (ACS) 2015-2019 5-year estimates for public-use microdata areas that we assign to core-based statistical areas. Total spending is the product of number of workers and average earnings. Note that Gottlieb et al. (2023) report higher physician earnings, averaging $350,000, than shown in this table due to under-reporting and top-coding in the ACS.

Table D.10: Contrasting geographies of colonoscopies and LVAD implantation

	Colonoscopy	LVAD Implant
Code	G0121	33979
N	58,785	333
Physicians	13,469	177
$\hat{\beta}_p^{\text{production}}$	-0.01	0.87
$\hat{\beta}_p^{\text{consumption}}$	-0.01	0.03
Share traded (HRR)	0.10	0.48
Share traded (CBSA)	0.11	0.48
Median distance traveled (km)	13.83	65.27
Share > 100km	0.04	0.37

Notes: This table reports statistics for two HCPCS codes: screening colonoscopy (G0121) and LVAD implantation (33979). We report the number of times the procedure is performed in 2017 in our 20% sample of Medicare patients and the number of distinct physicians performing it. The population elasticities of production and consumption are estimated using the Poisson models in equation (13) based on production HRR and patients' residential HRR, respectively. We also report the shares of procedures in which the patient and service location are in different HRRs or CBSAs, the median distance traveled for all care, and the share in which the patient and service location are more than 100 kilometers apart.

Table D.11: The stronger home-market effect for rare procedures is robust to instrumenting for population

	(1)	(2)	(3)	(4)	(5)	(6)
Geography:	HRR	HRR	CBSA	CBSA	CBSA	CBSA
Instrument:	1940 pop	1940 pop	1940 pop	1940 pop	Bedrock	Bedrock
Procedure Sample:	Common	Rare	Common	Rare	Common	Rare
Provider-market population (log)	0.562	1.127	0.753	0.915	1.242	1.790
	(0.0674)	(0.0871)	(0.0215)	(0.0408)	(0.344)	(0.551)
Patient-market population (log)	0.372	0.00372	0.315	0.288	0.0831	-0.617
	(0.0482)	(0.107)	(0.0273)	(0.0372)	(0.437)	(0.602)
Distance (log)	-0.746	0.845	-4.734	-1.611	-6.157	-5.017
	(0.239)	(0.403)	(0.257)	(0.930)	(1.143)	(2.608)
Distance (log, squared)	-0.0882	-0.244	0.232	-0.0515	0.345	0.215
	(0.0234)	(0.0457)	(0.0251)	(0.0877)	(0.0882)	(0.203)
Observations	93,636	93,636	857,476	857,476	781,456	781,456
Distance elasticity at mean	-1.60	-1.53	-2.44	-2.12	-2.74	-2.89

Notes: This table reports estimates of equation (10), when separating procedures into those above- and below-median frequency and instrumenting for log population. The dependent variable in all regressions is the value of trade. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files, using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. We report coefficients on provider market population, patient market population, log distance, and log distance squared. Every specification also includes a same-market ($i = j$) indicator variable. The odd-numbered columns are trade in above-median-frequency procedures; the even-numbered columns are trade in below-median-frequency procedures. In columns 1 and 2, the sample is all HRR pairs ($N = 306^2$). In columns 3 and 4, the sample is all CBSA pairs ($N = 926^2$). In columns 5 and 6, the sample is all CBSA pairs for which the bedrock-depth instrumental variable is available ($N = 844^2$). We use 1940 population counts to produce two instrumental variables: 1940 population in the patient market and 1940 population in the provider market are instruments for log population in the patient market and log population in the provider market, respectively. Similarly, we use bedrock depth to produce two instrumental variables for CBSAs. Both the strong home-market effect and its larger magnitude for rare procedures are robust to instrumenting for population, estimating by GMM. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.12: The home-market effect is stronger for rare procedures (commuting zones)

	Procedure		Procedure		Procedure		Diagnosis	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
λ_X Provider-market population (log)	0.813	0.783	0.782		0.783		0.782	
	(0.0278)	(0.0279)	(0.0279)		(0.0289)		(0.0291)	
λ_M Patient-market population (log)	0.219	0.237	0.238		0.230		0.230	
	(0.0264)	(0.0266)	(0.0265)		(0.0260)		(0.0262)	
μ_X Provider-market population (log) $\times$ rare			0.221	0.266	0.219	0.252	0.0565	0.0682
			(0.0351)	(0.0343)	(0.0355)	(0.0356)	(0.0137)	(0.0152)
μ_M Patient-market population (log) $\times$ rare			-0.118	-0.127	-0.114	-0.107	-0.0224	-0.0263
			(0.0377)	(0.0276)	(0.0378)	(0.0305)	(0.0123)	(0.0116)
p-value for H_0: $\lambda_X \leq \lambda_M$	<0.001	<0.001	<0.001		<0.001		<0.001	
p-value for H_0: $\mu_X \leq \mu_M$			<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	781,250	157,924	157,924	157,924	157,924	157,924	156,206	156,206
Distance [linear] controls	Yes	Yes	Yes	Yes				
Distance [quadratic] controls					Yes	Yes	Yes	Yes
Patient-provider-market-pair FEs				Yes		Yes		Yes

Notes: This table reports estimates of equation (12), which introduces interactions with an indicator for whether a procedure is "rare" (provided less often than the median procedure, when adding up all procedures provided nationally), using commuting zones (CZs) as the geographic units. The interactions with patient- and provider-market population reveal whether the home-market effect is larger for rare procedures. The unit of observation is {rare indicator, exporting CZ, importing CZ} so the number of observations is 2×625^2 in column 1, and the dependent variable in all regressions is the value of trade. All columns reflect estimates by procedure. Columns 2 onwards drop CZ pairs with zero trade in both procedure groups, and column 2 shows that this restriction has a negligible impact on the estimated log population coefficients. Columns 3 onwards include the rare indicator interacted with patient- and provider-market populations and distance covariates. Columns 1–4 control for distance using the log of distance between CZs. Columns 5–6 add a control for the square of log distance. Columns 4 and 6 introduce a fixed effect for each ij pair of patient market and provider market, so these omit all covariates that are not interacted with the rare indicator. The positive coefficient on provider-market population $\times$ rare across all columns indicates that the home-market effect is stronger for rare than for common services. The negative coefficient on patient-market population $\times$ rare across all columns indicates that the *strong* home-market effect has a larger magnitude for rare services. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files, using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. CZ definitions are from Fowler and Jensen (2020). Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.13: Gravity regression by procedure: individual procedures exhibit a strong home-market effect

	(1)	(2)	(3)	(4)	(5)	(6)
Procedure:	Colonoscopy	Cataract surgery	Brain tumor	Brain radiosurgery	LVAD	Colon removal
HCPCS code:	G0121	66982	61510	61798	33979	44155
Provider-market population (log)	0.454	0.349	0.932	1.107	1.332	0.871
	(0.0600)	(0.0819)	(0.0936)	(0.126)	(0.158)	(0.180)
Patient-market population (log)	0.391	0.521	0.162	0.191	0.115	-0.0729
	(0.0609)	(0.0711)	(0.0767)	(0.0943)	(0.136)	(0.159)
Distance (log)	-0.497	-0.0800	1.033	1.125	2.042	6.620
	(0.339)	(0.479)	(0.513)	(0.608)	(0.942)	(2.978)
Distance (log, squared)	-0.116	-0.170	-0.266	-0.272	-0.352	-0.850
	(0.0338)	(0.0474)	(0.0553)	(0.0628)	(0.0940)	(0.307)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.280	0.891	$<$0.001	$<$0.001	$<$0.001	0.001
Observations	93,636	93,636	93,636	93,636	93,636	93,636
Distance elasticity at mean	-1.62	-1.66	-1.56	-1.54	-1.50	-1.58
Total count	58,785	43,547	1,922	754	333	112

Notes: This table reports estimates of equation (10) for procedure-level trade for six selected HCPCS codes, which vary in how common they are. For all procedures, the sample is all HRR pairs ($N = 306^2$). The dependent variable in all regressions is the value of trade in the procedure (computed using each procedure's national average price). The independent variables are patient- and provider-market log population, log distance and square of log distance between HRRs, and an indicator for same-HRR observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market. The bottom row reports the total national count of the procedure in our sample. Common procedures include screening colonoscopy (column 1) and cataract surgery (column 2). In a screening colonoscopy, the physician visualizes the large bowel with a camera to look for cancer. In a cataract surgery, the surgeon removes a cloudy lens from the eye to improve vision. Relatively rare procedures include brain radiosurgery (column 3), brain tumor removal (column 4), left ventricular assist device (LVAD) implantation (column 5) and colon removal (column 6). In brain radiosurgery, an area of the brain is irradiated, often to kill a tumor. In an LVAD implantation, a pump is implanted in the chest to assist a failing heart in pumping blood. Brain tumor and colon removals involve surgical removal of the respective structures. The rare procedures have larger $\hat{\lambda}_X - \hat{\lambda}_M$ differences.

Table D.14: Home-market effect is stronger for rare services controlling for engagement

	(1)	(2)
Provider-market population (log) × common × high engagement	-0.0370	-0.0359
	(0.0177)	(0.0175)
Provider-market population (log) × rare × low engagement	0.272	0.275
	(0.0454)	(0.0359)
Provider-market population (log) × rare × high engagement	0.514	0.389
	(0.0777)	(0.142)
Patient-market population (log) × common × high engagement	0.0381	0.0369
	(0.0102)	(0.00989)
Patient-market population (log) × rare × low engagement	-0.172	-0.145
	(0.0338)	(0.0221)
Patient-market population (log) × rare × high engagement	-0.503	-0.589
	(0.0808)	(0.242)
Distance (log) × common × high engagement	-0.0655	-0.239
	(0.0141)	(0.0792)
Distance (log) × rare × low engagement	0.0122	1.049
	(0.0339)	(0.200)
Distance (log) × rare × high engagement	-0.155	2.348
	(0.0774)	(2.519)
Distance (log, squared) × common × high engagement		0.0165
		(0.00712)
Distance (log, squared) × rare × low engagement		-0.0953
		(0.0177)
Distance (log, squared) × rare × high engagement		-0.265
		(0.293)
Observations	220,804	220,804
Distance controls	Linear	Quadratic
Patient-provider-market-pair FEs	Yes	Yes
Additional distance elasticity at mean for high engagement: common procedures	-0.07	-0.00
Additional distance elasticity at mean for high engagement: rare procedures	-0.17	-1.13

Notes: This table reports estimates of a variant of equation (12), which adds interactions with indicators for whether a procedure is "rare" (provided less often than the median procedure) and for whether a procedure is "high engagement" (median number of distinct claims per patient for the procedure in a given year is above one) or low engagement. The unit of observation is {rare indicator, high-engagement indicator, exporting HRR, importing HRR}, and the dependent variable is the value of trade. Each column includes fixed effects for each ij pair of patient market and provider market, rare versus common procedures, and high- versus low-engagement procedures, plus indicators for three categories (common × high-engagement, rare × low-engagement, and rare × high-engagement) interacted with patient- and provider-market populations and distance covariates. Covariates for common × low-engagement procedures are omitted, since they would lead to collinearity with the ij fixed effects. Column 2 adds a control for the square of log distance and its interactions. The negative coefficient on provider-market population and the positive coefficient on patient-market population for common and high-engagement procedures indicate that the home-market effect is slightly less *strong* compared to common and low-engagement procedures, even though these effects are not all statistically different from zero. The positive coefficient on provider-market population × rare and the negative coefficient on patient-market population × rare for both high- and low-engagement procedures indicates that the *strong* home-market effect is stronger for rare services, whether they are high- or low-engagement. The distance elasticity is more negative for high-engagement procedures (both rare and common). Trade flows are computed from the Medicare 20% carrier Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.15: Home-market effect is stronger for procedures requiring more resources

	(1)	(2)	(3)	(4)	(5)	(6)
λ_X Provider-market population (log)	0.623	0.609	0.596		0.598	
	(0.0512)	(0.0487)	(0.0490)		(0.0483)	
λ_M Patient-market population (log)	0.366	0.370	0.391		0.390	
	(0.0512)	(0.0484)	(0.0482)		(0.0476)	
μ_{X2} Provider-market population (log) $\times$ intensity tercile 2			0.00378	0.00325	0.00613	0.00133
			(0.0116)	(0.0107)	(0.0131)	(0.0110)
μ_{M2} Patient-market population (log) $\times$ intensity tercile 2			-0.0452	-0.0454	-0.0463	-0.0425
			(0.0100)	(0.0100)	(0.0114)	(0.0101)
μ_{X3} Provider-market population (log) $\times$ intensity tercile 3			0.0848	0.0824	0.0929	0.0777
			(0.0186)	(0.0160)	(0.0216)	(0.0161)
μ_{M3} Patient-market population (log) $\times$ intensity tercile 3			-0.106	-0.106	-0.111	-0.0986
			(0.0172)	(0.0172)	(0.0188)	(0.0168)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.005	0.006	0.015		0.013	
p-value for H_0: $\mu_{X2} \leq \mu_{M2}$			0.004	0.003	0.008	0.007
p-value for H_0: $\mu_{X3} \leq \mu_{M3}$			$<$0.001	$<$0.001	$<$0.001	$<$0.001
Observations	280,908	164,670	164,670	164,670	164,670	164,670
Distance controls	Yes	Yes	Yes	Yes		
Distance [quadratic] controls					Yes	Yes
Patient-provider-market-pair FEs				Yes		Yes

Notes: This table reports estimates of a variant of equation (12) using terciles of the intensity of resources needed to produce the procedure rather than frequency. Procedure intensity is total relative value units (RVUs) in the October 2017 posting of the Medicare physician fee schedule. The unit of observation is {tercile intensity, exporting HRR, importing HRR}, and the dependent variable is the value of trade. Columns 2 onwards drop HRR pairs with zero trade, and column 2 shows that this restriction has a negligible impact on the estimated log population coefficients. Columns 1–4 control for distance using the log of distance between HRRs. Columns 5 and 6 add a control for the square of log distance. Columns 4 and 6 introduce a fixed effect for each ij pair of patient market and provider market, so these omit the patient- and provider-market population covariates. Coefficients on provider-market population are statistically significantly larger than coefficients on patient-market population, indicating a strong home-market effect for all procedures and for procedures in the first tercile of intensity for columns 3 and 5. The positive coefficients on provider-market population and negative coefficients on patient-market population for intensity terciles 2 and 3 across all columns indicates that the *strong* home-market effect is stronger for procedures requiring more resources. Trade flows are computed from the Medicare 20% carrier Research Identifiable Files, using the dollar value of physician services, excluding emergency-room care and assigning each procedure its national average price. Procedures associated with a zero total RVU are excluded, excluding 23.4% of spending. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.16: Scale elasticity estimates: all versus rare services

All services	Baseline	No Diagonal	Controls
OLS: 2017	0.751	0.777	0.763
	(0.028)	(0.047)	(0.040)
OLS: 2013–2017 difference	0.989	1.030	0.983
	(0.086)	(0.086)	(0.093)
2SLS: population (log)	0.723	0.801	0.727
	(0.031)	(0.053)	(0.037)
	[2073]	[2073]	[1634]
2SLS: population (1940, log)	0.449	0.671	0.467
	(0.072)	(0.103)	(0.067)
	[90.3]	[90.3]	[131]
Rare services	**Baseline**	**No Diagonal**	**Controls**
OLS: 2017	0.972	1.119	0.938
	(0.035)	(0.048)	(0.041)
OLS: 2013–2017 difference	1.326	0.859	1.348
	(0.264)	(0.542)	(0.278)
2SLS: population (log)	0.941	1.074	0.897
	(0.041)	(0.053)	(0.053)
	[1581]	[1575]	[1143]
2SLS: population (1940, log)	0.857	1.078	0.797
	(0.065)	(0.089)	(0.072)
	[129]	[128]	[164]

Notes: This table reports estimates of $\hat{\alpha}$ from ordinary least squares (OLS) or two-stage least squares (2SLS) regressions of the form $\widehat{\ln \delta_i} = \alpha \ln Q_i + \ln R_i + \ln w_i + u_i$, where $\widehat{\ln \delta_i}$ is estimated in equation (8) using professional fees, Q_i is region i's total production of non-emergency-room physician services for Medicare beneficiaries, R_i is Medicare's Geographic Adjustment Factor, the w_i covariate includes mean two-bedroom property value and mean annual earnings for non-healthcare workers, and u_i is an error term. In the rows labeled "2SLS" we instrument for $\ln Q_i$ using the specified instruments. In the columns labeled "no diag", Q_{ii} observations were omitted when estimating $\widehat{\ln \delta_i}$ in equation (8). In the third column, the $\ln R_i$ and $\ln w_i$ controls are included in the regressions (coefficients not reported). Standard errors (in parenthesis) are robust to heteroskedasticity. For 2SLS estimates, first-stage effective F-statistics (Montiel Olea and Pflueger, 2013) are reported in square brackets. The lower panel reports results for professional fees for rare procedures. In all specifications, we estimate substantial scale economies. The estimated scale elasticity is larger for rarer services.

Table D.17: Specialization earnings and frequency

	(1)	(2)	(3)
Dependent variable:	Per capita population elasticity		
Number of physicians in specialization (log, national)	-0.0716		-0.0677
	(0.0139)		(0.0137)
Mean earnings (log)		-0.245	-0.174
		(0.0697)	(0.0543)
Observations	209	209	209
R-squared	0.199	0.050	0.223

Notes: This table reports estimates of a regression of per capita population elasticity of physician count on the national count of physicians and mean earnings. Each observation is an NPPES taxonomy code. Earnings (wage and business income) data from Gottlieb et al. (2023) are reported by Medicare specialty groups. We use a crosswalk to map Medicare specialty groups to NPPES taxonomy codes. The estimation sample excludes 11 taxonomy codes that are not mapped to any Medicare specialty. Standard errors (in parentheses) are robust to heteroskedasticity.

Table D.18: Regression of $\hat{\Phi}_{j\kappa}$ on tercile dummies and trade shares

Counterfactual scenario:	Boston Reimbursement Increase		
	(1)	(2)	(3)
Income tercile = 2	0.0101	0.0133	-0.00859
	(0.0173)	(0.0175)	(0.00351)
Income tercile = 3	0.103	0.0979	-0.0140
	(0.0469)	(0.0477)	(0.00470)
$m_{\text{Boston},j\kappa}$			42.10
			(0.823)
Constant	0.0856	0.0860	-0.128
	(0.119)	(0.0193)	(0.00633)
Observations	874	874	874
R-squared	0.000	0.971	1.000
HRR fixed effects	No	Yes	Yes

Notes: This table uses linear regressions to summarize how market access changes across HRRs j and income terciles κ in response to a counterfactual policy: a 30% reimbursement increase in Boston. The dependent variable in all columns is the percentage change in market access, $100 \times (\hat{\Phi}_{j\kappa} - 1)$. Standard errors (in parentheses) are clustered by market. The constant in the first regression reports the percentage change for the lowest income terciles, and the coefficients on the other terciles are the additional percentage point gain for those terciles relative to the lowest. When we control for Boston's market share $m_{\text{Boston},j\kappa}$. the tercile-difference coefficients are much smaller, indicating that differences in baseline trade patterns drive the distributional impacts.

Table D.19: Estimates of a strong home-market effect excluding AZ, FL, CA

	(1)	(2)	(3)	(4)
Estimation method:	PPML	PPML	PPML	IV
Provider-market population (log)	0.706	0.702	0.692	0.752
	(0.0749)	(0.0652)	(0.0431)	(0.0620)
Patient-market population (log)	0.238	0.249	0.284	0.291
	(0.0729)	(0.0624)	(0.0389)	(0.0453)
Distance (log)	-1.715	2.046		2.094
	(0.0587)	(0.438)		(0.372)
Distance (log, squared)		-0.390		-0.394
		(0.0445)		(0.0374)
p-value for H_0: $\lambda_X \leq \lambda_M$	0.001	<0.001	<0.001	<0.001
Observations	67,600	67,600	67,600	67,600
Distance elasticity at mean		-1.61		-1.60
Distance deciles			Yes	

Notes: This table reports estimates of equation (10), which estimates the presence of weak or strong home-market effects, excluding snowbird states. The sample is all HRR pairs, excluding those in Arizona, Florida, or California. The dependent variable in all regressions is the value of trade computed by assigning each procedure its national average price. The independent variables are patient- and provider-market log population, log distance between HRRs, and an indicator for same-HRR observations ($i = j$). The positive coefficient on provider-market log population implies a weak home-market effect, and the fact that this coefficient exceeds that on patient-market population implies a strong home-market effect. Column 2 makes the distance coefficient more flexible by adding a control for the square of log distance. Column 3 replaces parametric distance specifications with fixed effects for each decile of the distance distribution. Column 4 uses the provider-market and patient-market log populations in 1940 as instruments for the contemporaneous log populations when estimating by generalized method of moments. Trade flows are computed from the Medicare 20% carrier, 100% MedPar, and 100% outpatient Research Identifiable Files. HRR definitions are from the Dartmouth Atlas Project. Standard errors (in parentheses) are two-way clustered by patient market and provider market.

Table D.20: Estimates of a strong home-market effect excluding HRRs with high second-home share

	(1)	(2)	(3)	(4)
Estimation method:	PPML	PPML	PPML	IV
Provider-market population (log)	0.683	0.696	0.684	0.789
	(0.0593)	(0.0555)	(0.0386)	(0.0567)
Patient-market population (log)	0.257	0.247	0.281	0.305
	(0.0598)	(0.0551)	(0.0365)	(0.0513)
Distance (log)	-1.639	0.643		0.683
	(0.0503)	(0.316)		(0.272)
Distance (log, squared)		-0.233		-0.236
		(0.0315)		(0.0269)
p-value for H_0: $\lambda_X \leq \lambda_M$	<0.001	<0.001	<0.001	<0.001
Observations	76,176	76,176	76,176	76,176
Distance elasticity at mean		-1.57		-1.55
Distance deciles			Yes	

Notes: This table reports estimates of equation (10), which estimates the presence of weak or strong home-market effects, excluding HRRs with a high second-home share. The sample is all HRR pairs excluding those in the top 10% based on the share of housing units that are vacant for seasonal/recreational purposes in the 2013–2017 American Community Survey. See Table D.19 notes on the variables, instruments, geographic units, and standard errors.

Table D.21: Travel for dialysis

Distance (km)	Share of output				
	All (Professional)	All (Facility)	All (Dialysis)	No snowbird states (Dialysis)	Snowbird states (Dialysis)
$[0, 50)$	0.85	0.77	0.94	0.94	0.93
$[50, 100)$	0.08	0.12	0.03	0.03	0.03
$[100, .)$	0.07	0.11	0.03	0.02	0.04

Notes: For the care described in each column and the distance intervals in each row, the entries in this table report the share of patients traveling that distance from their residential ZIP code to the service location's ZIP code. The first column shows professional claims (from Medicare's "carrier" file), the second column shows facility (hospital) claims, and the third column shows dialysis claims. The remaining columns split dialysis claims between "snowbird" states (AZ, CA, and FL, following Finkelstein, Gentzkow, and Williams 2016) and other states. In non-snowbird states, the table shows that 94% of patients travel less than 50 km from their home for dialysis, and only 2% more than 100 km. This is less than one-fifth as much as for other facility or professional care, suggesting that residential location is recorded correctly for almost all patients.

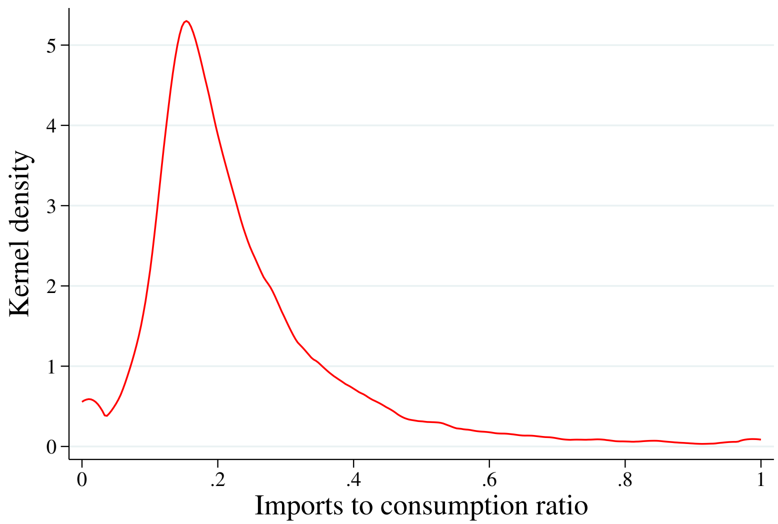
Table D.22: Classification of rare and common procedures in Medicare vs. private insurance data

Above median HCCI Above median CMS	0	1	total
0	82	18	100
1	12	88	100

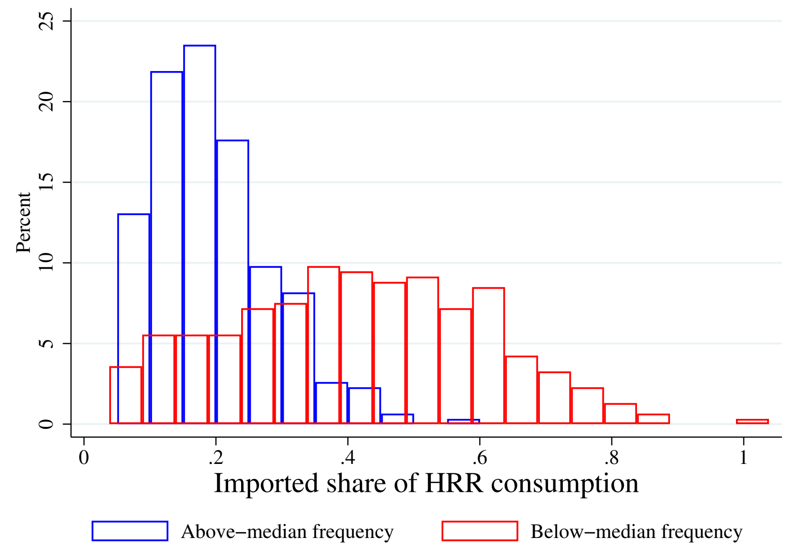
Notes: This table compares the percentage of procedures classified as rare (above median frequency equals one) or common (above median frequency equals zero) in the public Medicare data versus the private insurance data from the Health Care Cost Institute (HCCI). Classifying procedures as rare versus common is consistent when using Medicare or privately insured data.

Figure D.1: Variation in trade shares across procedures and regions

(a) Distribution of import share by procedure

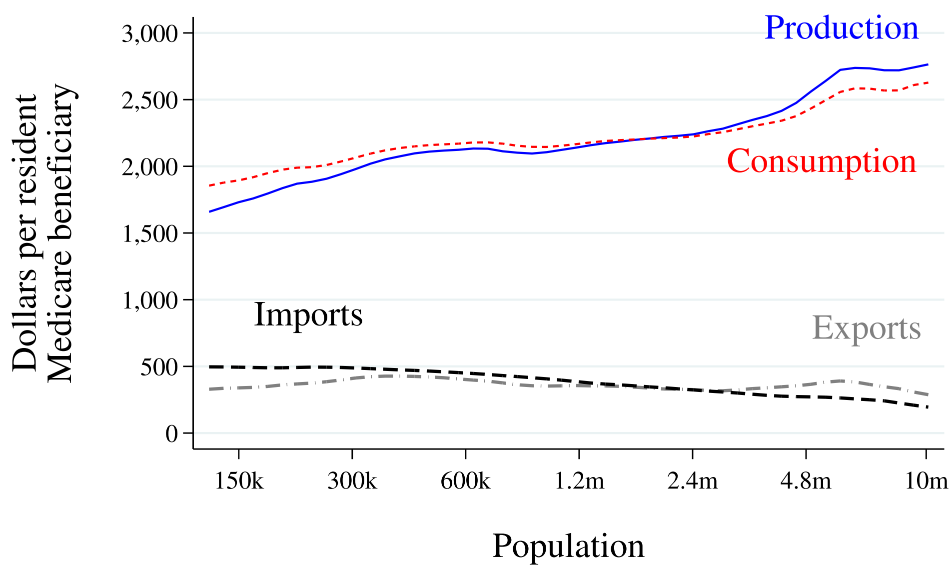


(b) Distributions of import shares for common and rare services



Notes: Panel (a) shows the distribution of the imported consumption share across procedures for procedures performed at least 20 times (in our 20% sample of Medicare claims). Imports are defined as care provided to a patient who lives in one HRR at a service location in a different HRR. Panel (b) splits all services into two groups based on how often they are performed nationally. Those performed less often than the median are shown in red, and those performed more often than the median service are shown in blue. Import shares are substantially higher for the rare services.

Figure D.2: Production and consumption of medical care across regions (professional fees)

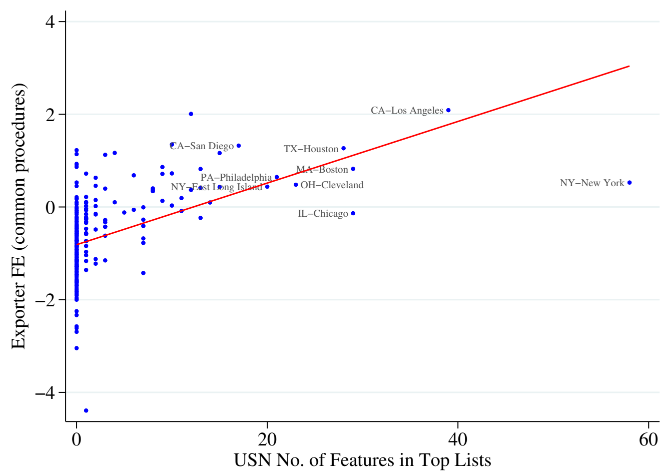


Population elasticity (log−log regression slope) of transactions per resident Medicare beneficiary:
Production: 0.10 (0.02), Consumption: 0.06 (0.01)
Exports: −0.03 (0.04), Imports: −0.25 (0.03)

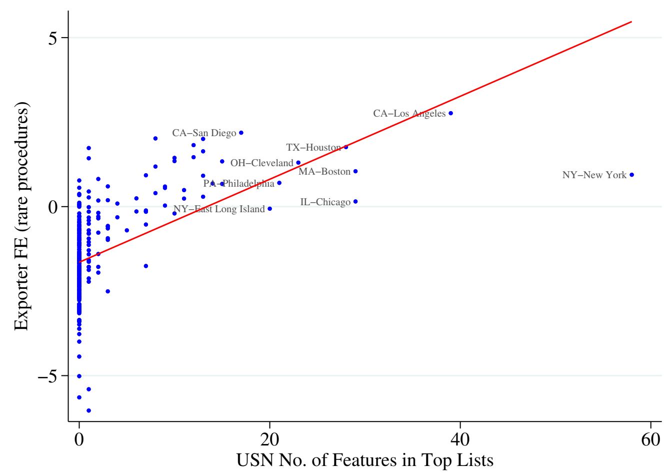
Notes: This figure shows production, consumption, and trade per capita of Medicare services across hospital referral regions (HRRs) of different sizes, all smoothed via local averages. We use the Medicare 20% carrier Research Identifiable Files to compute the dollar value of services, excluding emergency-room care, at national average prices. The blue series shows production of medical care per Medicare beneficiary residing in the HRR of production. The red series shows consumption of medical care per Medicare beneficiary residing in the HRR of consumption. The dashed gray series shows interregional "exports" of medical care and the dashed black series shows interregional "imports" of medical care, again per Medicare beneficiary. HRR definitions are from the Dartmouth Atlas Project.

Figure D.3: Revealed-preference vs. external quality measures: common and rare

(a) *U.S. News* vs. quality estimated for common services

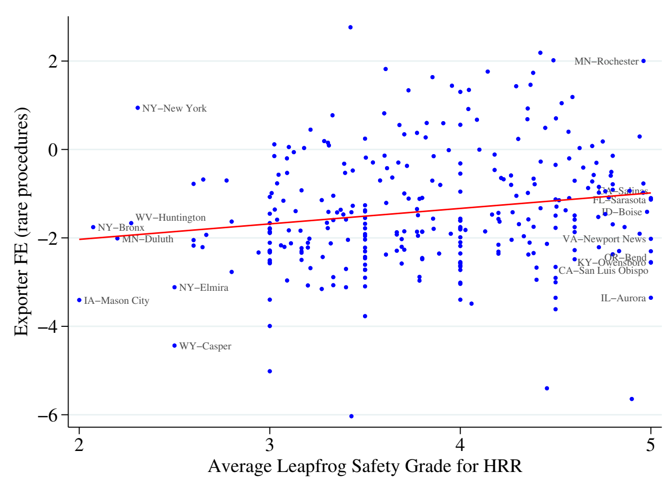


(b) *U.S. News* vs. quality estimated for rare services



(c) Leapfrog Safety Grade vs. quality for common services

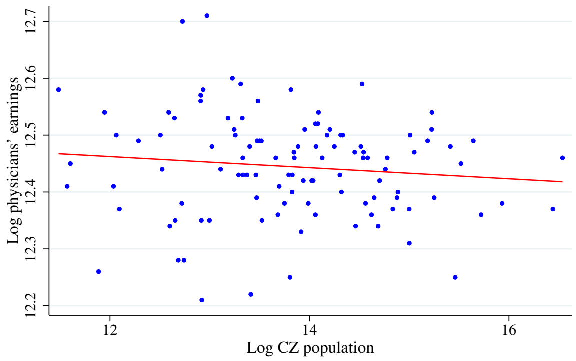
(d) Leapfrog Safety Grade vs. quality for rare services



Notes: This figure shows the relationships between the exporter fixed effects (our revealed-preference measure of quality), estimated separately for common and rare services, and external quality measures. The vertical axis shows the exporter fixed effects for each HRR estimated using trade in common services in Panels (a) and (c), and using trade in rare services in Panels (b) and (d). The horizontal axis in Panels (a) and (b) is a count of the number of times each region's hospitals appear on the *U.S. News* list of best hospitals. *U.S. News* produces an overall ranking as well as rankings for 12 particular specialties. We count the number of times each HRR's hospitals appear on any of these 13 lists. Both panels show a positive relationship, indicating that patients travel farther to obtain care from regions highly ranked by *U.S. News*. The relationship is stronger for rare services, as the slope is nearly double that for common services. The horizontal axis in Panel (c) and (d) is the average safety grade for hospitals in an HRR, determined by the Leapfrog Group. The Leapfrog Safety Grades range from A to F, which we scale as integers from 1 (for F) to 5 (for A). We then compute the mean score for all hospitals in the HRR. The Safety Grades are positively associated with the exporter fixed effects for both rare and common procedures.

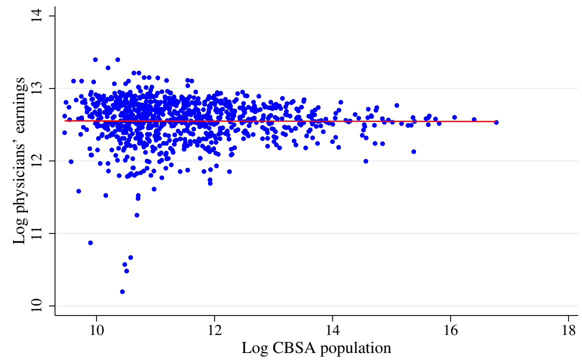
Figure D.4: Population elasticities of input costs

(a) Physicians' earnings (commuting zones)



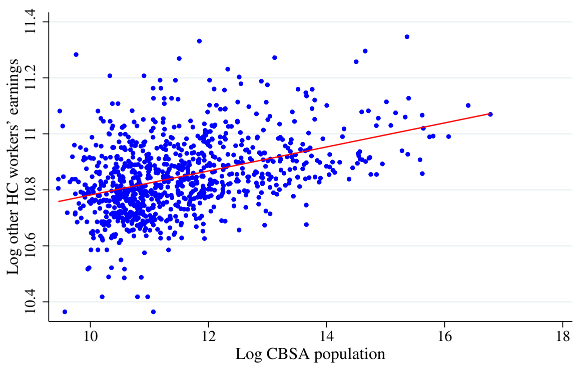
Y = 12.5794 − 0.0098 [0.0082] Log CZ population + ε
N = 111; R^2 = 0.0127

(b) Physicians' earnings (CBSAs)



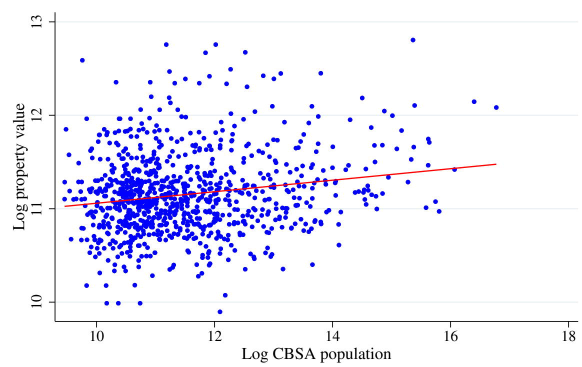
Y = 12.5670 − 0.0014 [0.0082] Log CBSA population + ε
N = 923; R^2 = 0.0000

(c) Other healthcare workers' earnings



Y = 10.3523 + 0.0429 [0.0032] Log CBSA population + ε
N = 926; R^2 = 0.1628

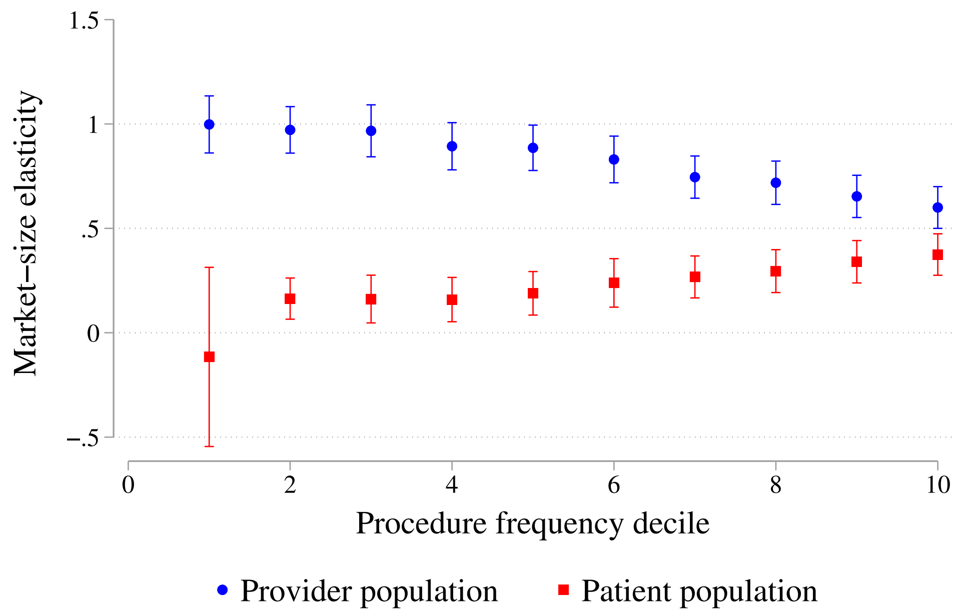
(d) Median house value



Y = 10.4431 + 0.0616 [0.0110] Log CBSA population + ε
N = 926; R^2 = 0.0331

Notes: This figure depicts relationships between input costs and population sizes. Panel (a) shows physicians' earnings across 111 commuting zones using data from Gottlieb et al. (2023). Panels (b), (c), and (d) show variation across CBSAs in physicians' earnings, other healthcare workers' earnings, and median house values (a proxy for real estate and other locally priced inputs) using data from the 2015–2019 American Community Survey.

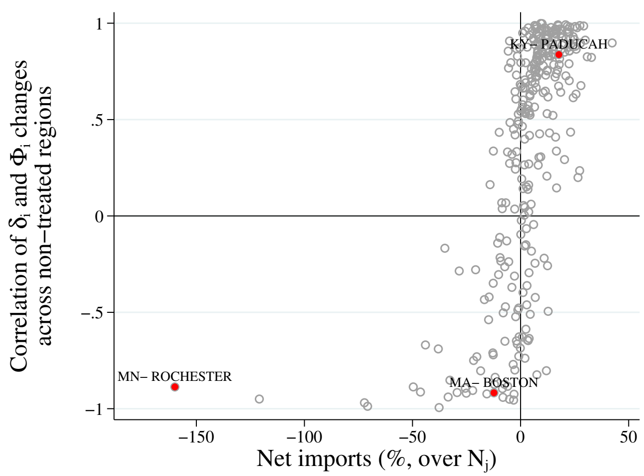
Figure D.5: The home-market effect is stronger for rare procedures



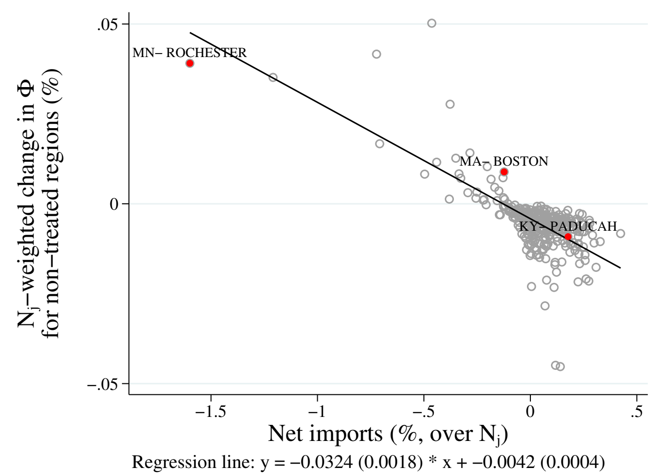
Notes: This figure groups non-emergency physician-provided services in the Medicare claims data into deciles based on the national frequency of each procedure. For each decile, we estimate equation (10) and plot the estimated coefficients on provider- and patient-market log population with their 95% confidence intervals. The blue circles show estimated provider-market population elasticities; the red squares show patient-market population elasticities. The coefficients on provider-market size always exceed the respective coefficients on patient-market size, indicating a strong home-market effect. The coefficients on provider-market size monotonically decrease across the deciles. The coefficients on patient-market size monotonically increase across the deciles. Together, these two patterns show that the home-market effect is stronger the less common the procedure.

Figure D.6: Spillovers from higher reimbursements in one region depend on that region's net imports

(a) Correlation of $\hat{\delta}_i$ and $\hat{\Phi}_i$ across non-treated regions



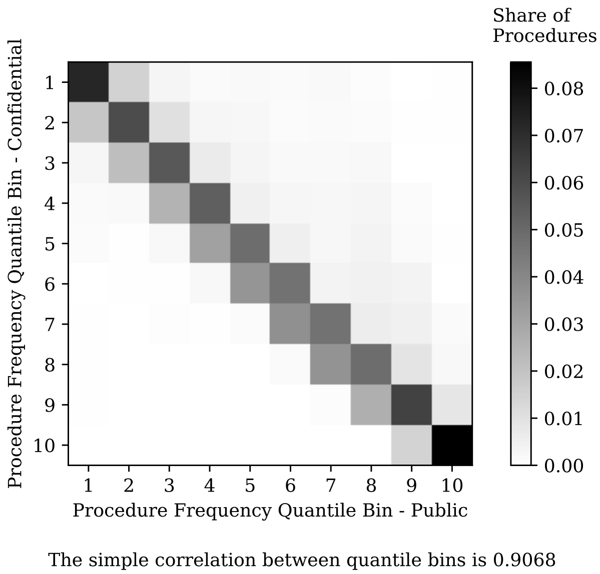
(b) Change in non-treated regions' aggregate market access



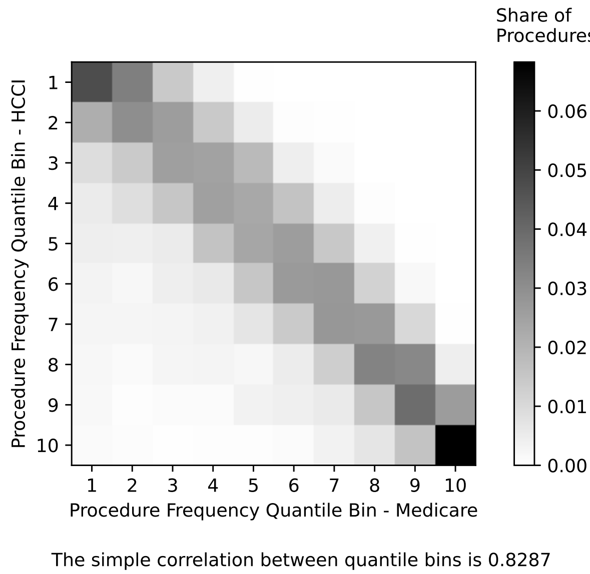
Notes: This figure characterizes counterfactual outcomes when raising reimbursements by 30 percent in one HRR. We conduct this exercise for each region, one at a time, and each observation in each panel represents one such counterfactual scenario. Panel (a) illustrates the contrast in spillovers as a function of net imports of the treated region. The vertical-axis value for each observation reports the correlation—across *all regions other than the treated one* for the exercise in question—between the counterfactual changes $\hat{\delta}_i$ and $\hat{\Phi}_i$. The scatterplot relates these correlations to the *treated* region's net import share, which is plotted on the horizontal axis. When the treated region is a net exporter, changes in quality δ_i and in market access Φ_i for non-treated regions move in opposite directions: a region whose output quality declines experiences an increase in market access through imports from the treated region. However, increasing reimbursements in a net-importing region often has the opposite effect: neighboring regions with quality reductions also experience lower market access, (changes in δ_i and Φ_i are positively correlated). For each counterfactual, the vertical-axis value in Panel (b) shows the aggregate impact on patient market access *excluding the treated region*. The panel relates this impact to the *treated* region's net imports, shown on the horizontal axis. When the treated region is a net importer, the aggregate impact on market access for non-treated regions tends to be smaller or even negative.

Figure D.7: Procedure frequencies are similar across data sources

(a) Confidential versus public Medicare data



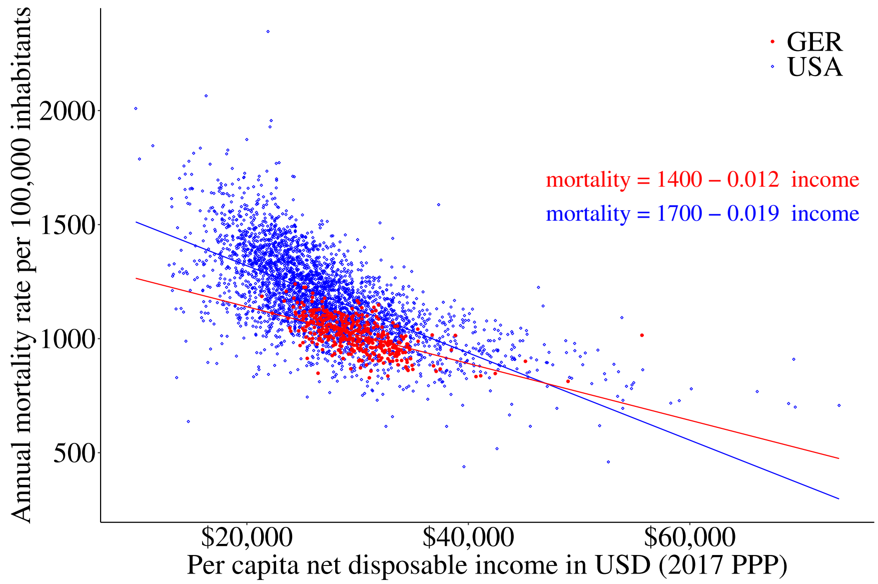
(b) Medicare versus private insurance data



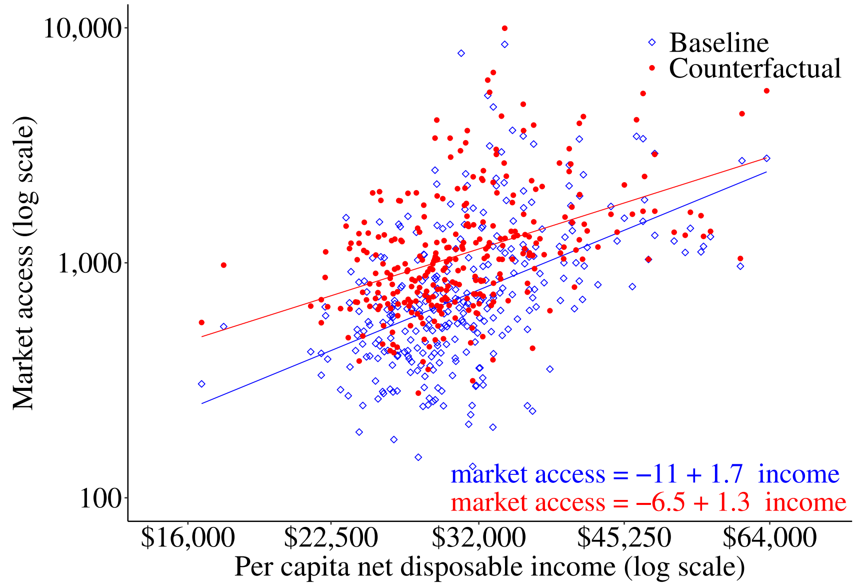
Notes: The first panel shows the share of procedures in each frequency decile in the Medicare public data compared to the Medicare confidential data. The classification of procedures by frequency deciles appears largely consistent between the two data sources for Medicare patients. The second panel shows the share of procedures in each frequency decile in the Medicare versus data on privately insured patients from the Health Care Cost Institute (HCCI). The classification of procedures by frequency deciles appears largely consistent when comparing public Medicare data with the privately insured.

Figure D.8: Counterfactual health-income gradients if US geography were like Germany

(a) Mortality-income gradient for the United States and Germany



(b) Change in US market access gradient for counterfactual with German trade costs



Notes: Panel (a) shows the mortality-income gradients across regions in the United States and in Germany. For the United States, we use county-level data on mortality by age and income (obtained from the Institute for Health Metrics and Evaluation at the University of Washington). For Germany, we use NUTS3 ("Kreise")-level data on mortality by age (from EuroStat) and income (from the German Federal Statistical Office). We compute average mortality from age-bracket-specific mortality rates using the 2011–2030 European Standard Population as the reference distribution. We see a steeper gradient in the United States. Panel (b) considers to what extend this can be explained by the geography of the United States, in particular the remoteness of many regions to big markets. To do this, we conduct a counterfactual analysis, described in Section 5.2 and Appendix C.5. We use our model and estimates to simulate the market access (Φ_i) that would emerge if US geography looked more like Germany, and relate it to areas' incomes. Panel (b) shows that transformation reduces the Φ_i-income elasticity by 21%.