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 $Q_i = Q_i^{\text{TM}} + Q_i^{\text{MA}} + Q_i^{\text{NM}}$ 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:

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

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

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

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

where $s_i^{\text{MA}} \equiv \frac{Q_i^{\text{MA}}}{Q_i^{\text{TM}} + Q_i^{\text{MA}} + Q_i^{\text{NM}}}$ is the MA share of production in region i, $\varrho_i^{\text{MA}} \equiv \frac{\mathrm{d} \ln \frac{Q_i^{\text{MA}}}{Q_i^{\text{TM}}}}{\mathrm{d} \ln Q_i^{\text{TM}}}$ 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 ($\varrho^{\text{MA}} = \varrho^{\text{NM}} = 0$), 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 $\bar{s}^{\text{MA}}$ and $\bar{s}^{\text{NM}}$ 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 $\bar{s}^{\text{MA}} = 0.111$. 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 $\bar{s}^{\text{NM}} = 0.676$.

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 $\hat{\varrho}^{\text{MA}} = 0.089$ and $\hat{\varrho}^{\text{NM}} = 0.075$. 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 $\mathrm{d}\ln\delta_1$:

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

Further tedious algebra delivers the following expression for quality changes:

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

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

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

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 $\bar{\mathcal{Q}} \equiv \frac{\Phi^2}{\bar{N}}\frac{(1-\alpha)}{\alpha}\bar{\delta}^{\frac{1-2\alpha}{\alpha}} > 0$. 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

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

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:

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

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 $S_{j\kappa}^{\text{TM}}$. We assume $\delta_{0j} = \delta_0\ \forall j$ and $N_{j\kappa} = \mathfrak{s}S_{j\kappa}^{\text{TM}}$, where $\mathfrak{s}$ is a constant of proportionality that is common across types. This implies

$$m_{0j\kappa} = \frac{\delta_0}{\Phi_{j\kappa}} = \frac{\delta_0\theta_{j\kappa}}{N_{j\kappa}} = \frac{\delta_0\theta_{j\kappa}}{\mathfrak{s}S_{j\kappa}^{\text{TM}}}.$$

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%, $\frac{1}{\mathcal{I}\mathcal{K}}\sum_{j\kappa} m_{0j\kappa} = 0.1$. This implies

$$m_{0j\kappa} = 0.1 \times \frac{\theta_{j\kappa}/S_{j\kappa}^{\mathrm{TM}}}{\frac{1}{\mathcal{I}\mathcal{K}}\sum_{j'\kappa'}\theta_{j'\kappa'}/S_{j'\kappa'}^{\mathrm{TM}}}.$$

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)
	Cross-sectional PPML			IV: 1940 population	IV: Bedrock	2013–2017 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)	(5)	(6)
	Cross-sectional PPML			IV: 1940 population	IV: Bedrock	2013–2017 panel
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)
	Cross-sectional PPML			IV: 1940 population	2013–2017 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) × 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) × 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 × 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 × 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) × 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) × 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) × 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) × 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

Dependent variable:	(1)	(2)	(3)
	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

Estimation method:	(1) PPML	(2) PPML	(3) PPML	(4) 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

Estimation method:	(1) PPML	(2) PPML	(3) PPML	(4) 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.

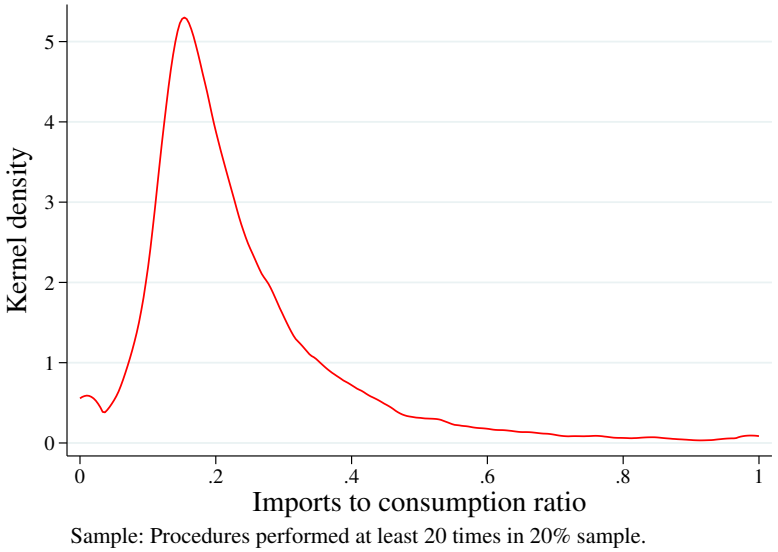
Table D.22: Classification of rare and common procedures in Medicare vs. private insurance data

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

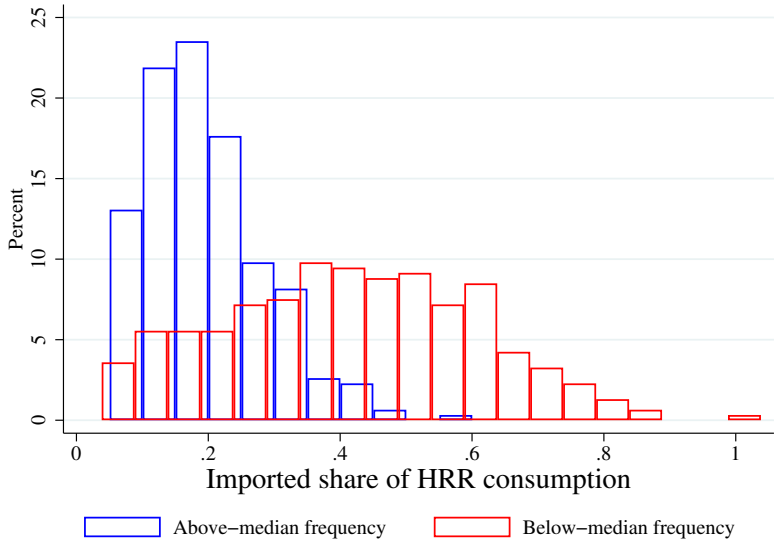
Notes: This table compares the percentage of procedures classified as rare (above median frequency equals one) or common (above median frequency equals zero) in the public Medicare data versus the private insurance data from the Health Care Cost Institute (HCCI). Classifying procedures as rare versus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

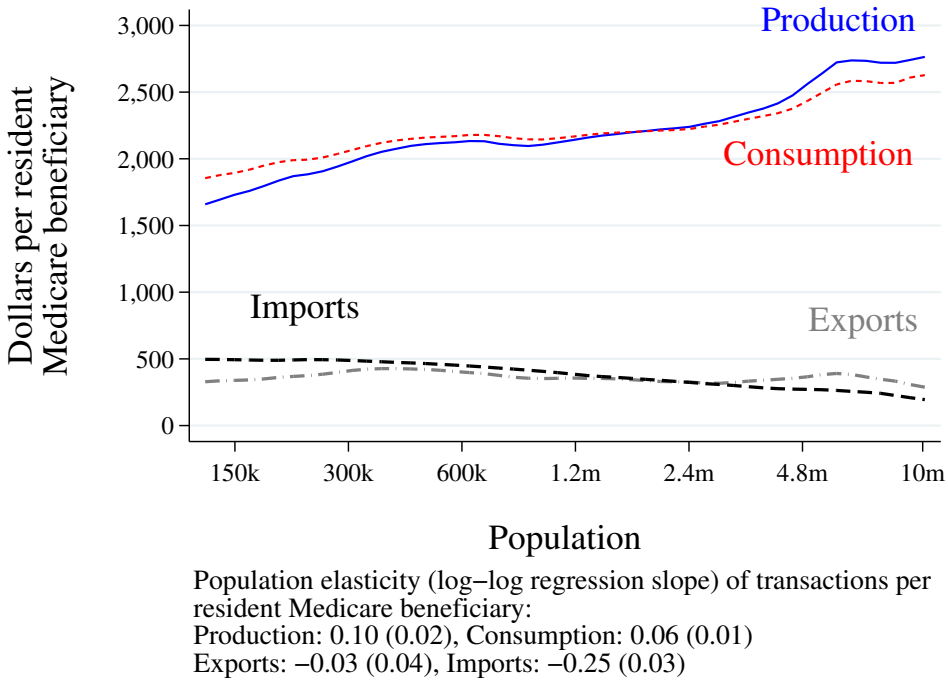


(b) Distributions of import shares for common and rare services



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.

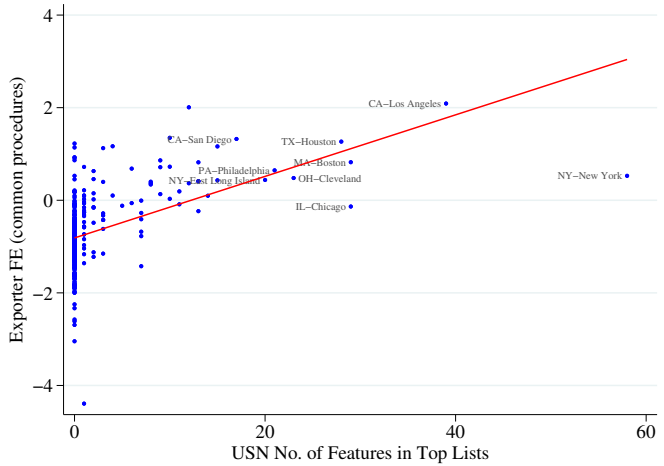
Figure D.2: Production and consumption of medical care across regions (professional fees)



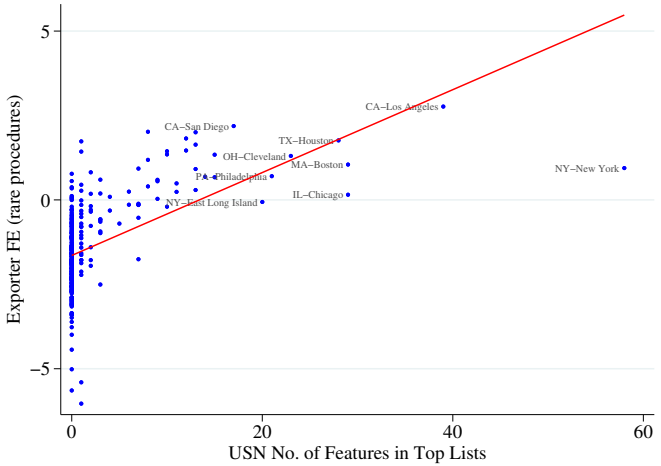
Notes: This figure shows production, consumption, and trade per capita of Medicare services across hospital referral regions (HRRs) of different sizes, all smoothed via local averages. We use the Medicare 20% carrier Research Identifiable Files to compute the dollar value of services, excluding emergency-room 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

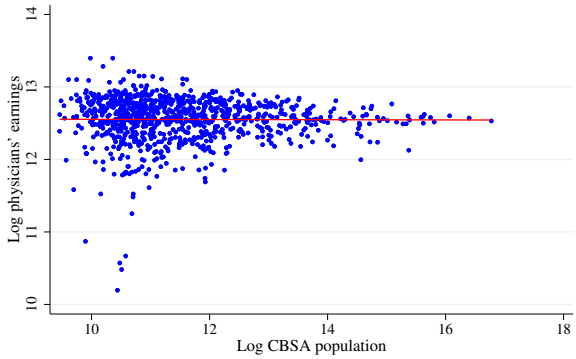
Notes: This figure shows the relationships between the exporter fixed effects (our revealed-preference measure of quality), estimated separately for common and rare services, and external quality measures. The vertical axis shows the exporter fixed effects for each HRR estimated using trade in common 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)

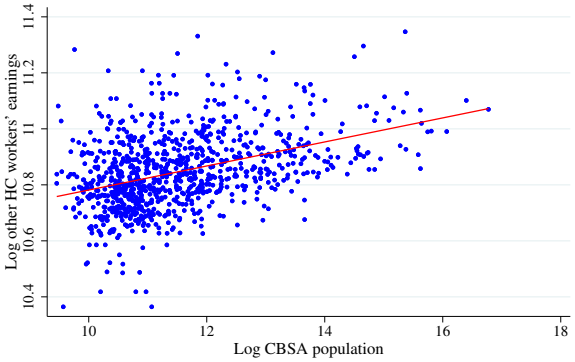
Y = 12.5794 – 0.0098 [0.0082] Log CZ population + ε
N = 111; R^2 = 0.0127

(b) Physicians' earnings (CBSAs)



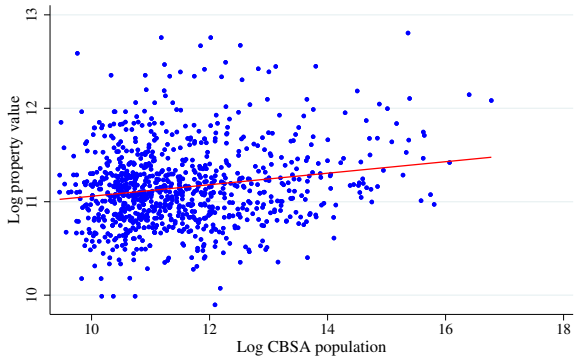
Y = 12.5670 – 0.0014 [0.0082] Log CBSA population + ε
N = 923; R^2 = 0.0000

(c) Other healthcare workers' earnings



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

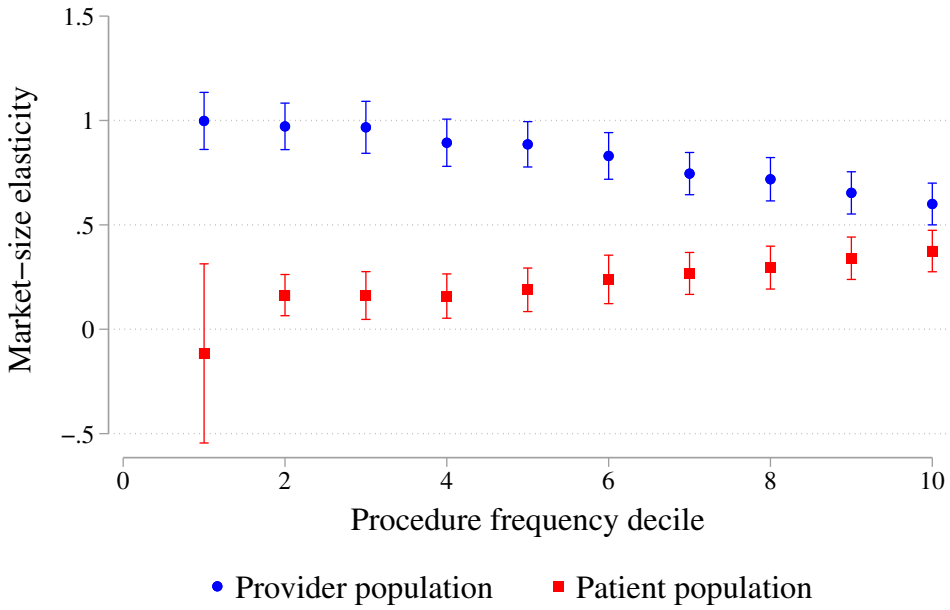
(d) Median house value



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.

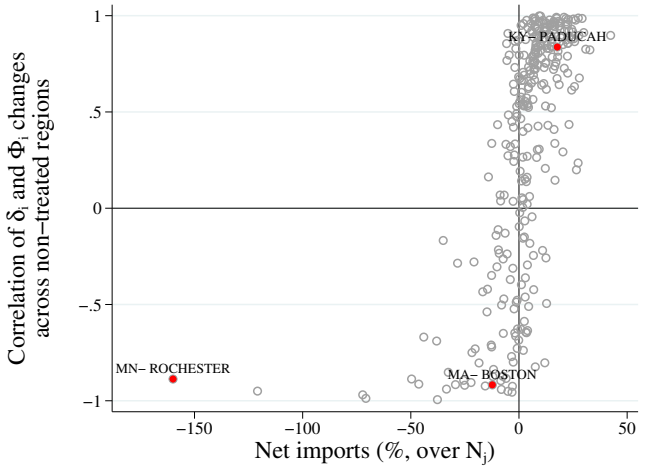
Figure D.5: The home-market effect is stronger for rare procedures



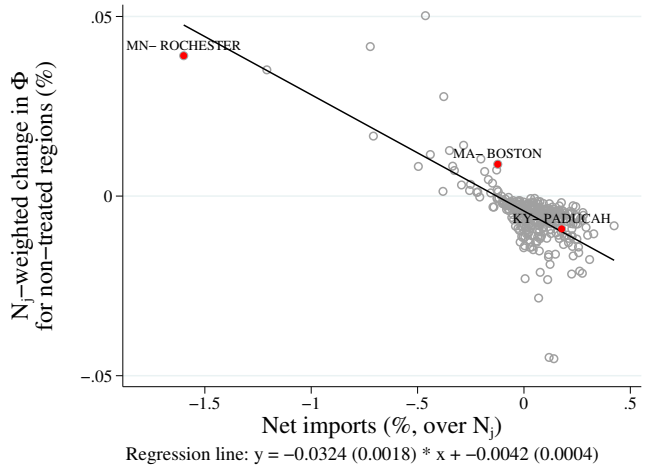
Notes: This figure groups non-emergency physician-provided services in the Medicare claims data into deciles based on the national frequency of each procedure. For each decile, we estimate equation (10) and plot the estimated coefficients on provider- and patient-market log population with their 95% 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

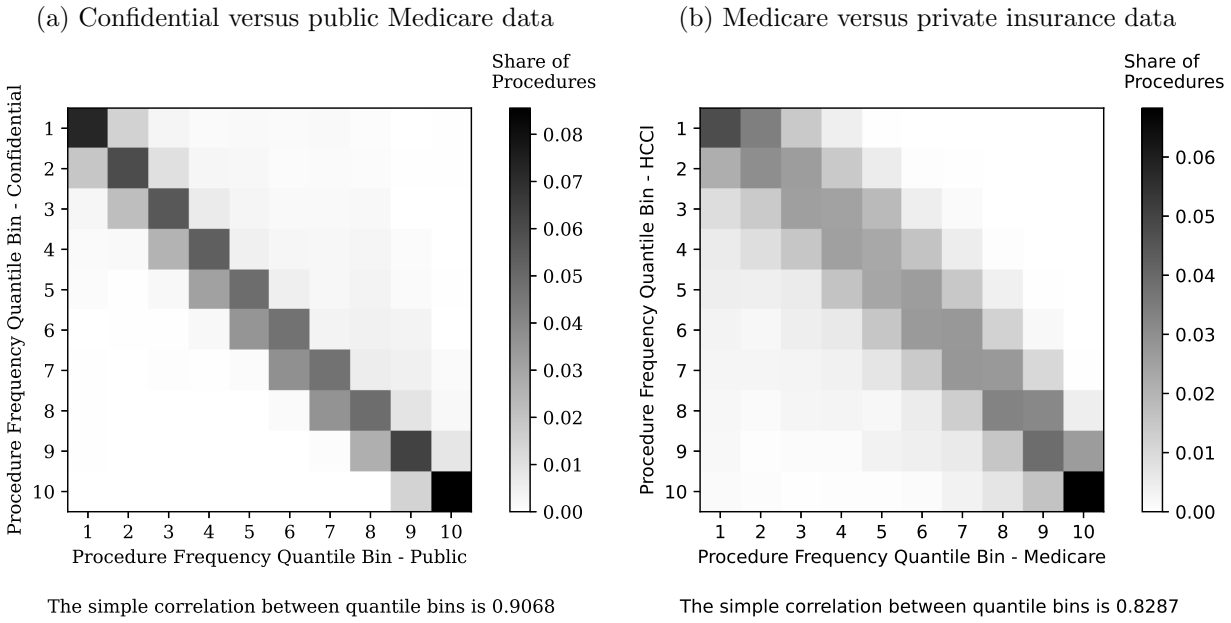


(b) Change in non-treated regions' aggregate market access



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.

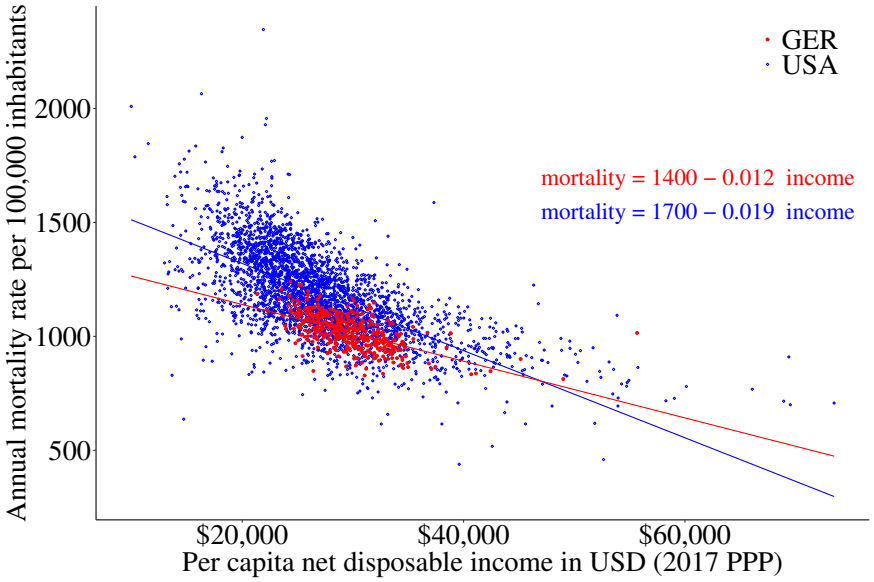
Figure D.7: Procedure frequencies are similar across data sources



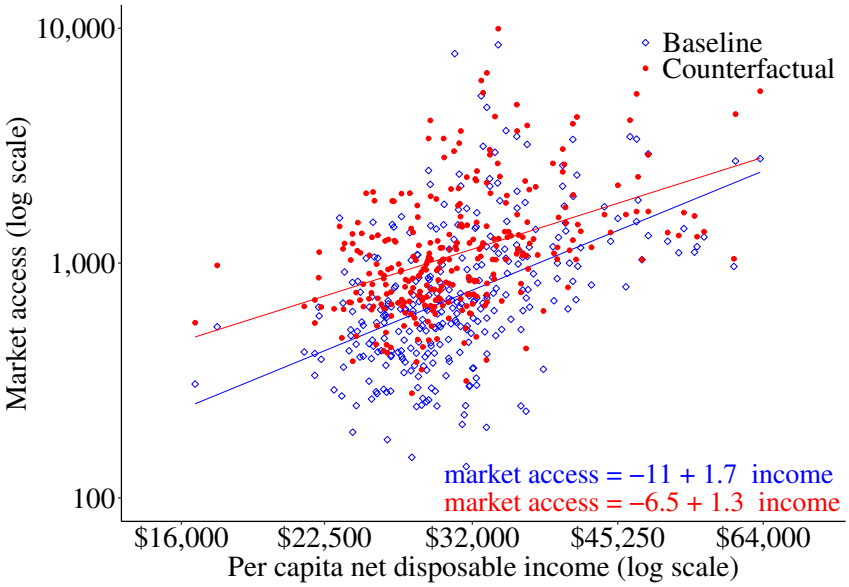
Notes: The first panel shows the share of procedures in each frequency decile in the Medicare public data compared to the Medicare confidential data. The classification of procedures by frequency deciles appears largely consistent between the two data sources for Medicare patients. The second panel shows 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%.