Geographic Variation in Healthcare Utilization:

The Role of Physicians

Ivan Badinski — *Amazon*

Amy Finkelstein — *MIT and NBER*

Matthew Gentzkow — *Stanford and NBER*

Peter Hull — *Brown and NBER**

April 2024

Abstract

We study the role of physicians in driving geographic variation in US healthcare utilization. We estimate a model that separates variation in utilization of Medicare beneficiaries due to physicians, non-physician supply side factors, and patient demand. The model is identified by patient and physician migration and within-area matching. We find that at least a third of geographic variation can be explained by differences in the intensity with which physicians treat otherwise similar patients. Around three-fifths of physicians' role comes from differences across areas in practice styles within specialty, while the other two-fifths reflects differences across areas in physician specialty mix.

*We thank the National Institute on Aging (Finkelstein, R01-AG032449) and the Stanford Institute for Economic Policy Research (SIEPR) for support. We are grateful to Yunan Ji, Ken Jung, Isaac Norwich, Tamar Oostrom, Enzo Profili, and Steven Shi for expert research assistance, and to Tim Layton, Jonathan Skinner and seminar participants at the BU-Harvard-MIT Health Economics Seminar, Chicago Booth, and the NBER Economics of Aging Summer Institute for helpful comments. We are especially grateful to Heidi Williams, who played a key role in the genesis of this project and made important contributions throughout its early stages. Keywords: Healthcare spending, geographic variation, physicians. JEL: H51, I1, I11

The most expensive piece of medical equipment, as the saying goes, is a doctor's pen.

Atul Gawande (2009), "The Cost Conundrum"

1 Introduction

This paper studies the role of physicians in driving geographic variation in US healthcare utilization. This variation has been well-documented, particularly in the over-65 Medicare population, where the highest spending areas of the country have annual age, race, and sex-adjusted per-capita spending that is more than double that of the lowest spending areas (Austin et al. 2020). Noting that higher-spending areas do not tend to have better patient outcomes (Skinner 2011), both academics and policymakers have long asked whether spending could be substantially reduced by making healthcare practice in high-spending parts of the country more similar to that in low-spending parts of the country (Congressional Budget Office 2008; Gawande 2009; Skinner 2011).

The impact and feasibility of such changes depends crucially on what factors cause some areas to spend more than others. A natural hypothesis is that physicians play a key role. They are the nucleus of the healthcare system—making diagnoses, guiding treatment decisions, and delivering many treatments themselves. However, evidence on their role in driving geographic variation remains limited, and in some cases suggests conflicting conclusions. Cutler et al. (2019), for example, examine the relationship between areas' per-capita spending and the way their patients and physicians respond to treatment vignettes, concluding that physician beliefs about treatment are a key correlate of cross-area variation. Molitor (2018), in contrast, studies the way cardiologists' treatment decisions change when they move, concluding that cardiologist practice styles explain only a small share of cross-area variation.

Understanding the overall importance of physicians in driving geographic variation in health care utilization is essential for understanding the potential impacts of policies aiming to reduce utilization in high-use areas. If physicians are indeed a key driving force, this could suggest a potentially important role for policies that standardize physician training or incentives. If, on the other hand, other supply-side factors such as hospital organization, ownership, or competition are the key factors, such policies could be ineffective or counterproductive. Prior evidence on the drivers of aggregate utilization suggests that 40 to 50 percent of geographic variation reflects differences in patient demand—such as health or preferences—while the other half reflects supply-side factors

(Finkelstein et al. 2016), but it does not isolate the portion of the supply-side component that is attributable to physicians.

In this paper, we develop and estimate a model of annual patient health care utilization that separates variation due to physicians, non-physician supply factors, and patient demand. The physician component, which we refer to as *physician practice intensity*, includes differences in how physicians of a given specialty practice, as well as differences in the mix of physician specialties across areas. The non-physician supply component, which we refer to as *practice environment*, includes differences in hospital capacity, physical capital, hospital or health care facility ownership, the organization of health care markets, the degree of market competition, and organizational culture and norms (Lee and Mongan 2009). Recent empirical work has shown these practice environment factors to be important in affecting care delivery (e.g. Chan et al. 2022a; Duggan et al. 2022; Bloom et al. 2015; Otero and Munoz 2022; Doyle and Staiger 2022; Frakes et al. forthcoming; Eliason et al. 2020; Ho and Lee 2017, 2019).

We model utilization in two steps. In the first step, patients choose whether to seek care based on their own demand factors (e.g. health and preferences) as well as aspects of the practice environment (e.g. wait times or travel times to nearby hospitals). If a patient does seek care, she matches with a physician in the second step. We refer to a patient-physician match as an *encounter.* The ensuing utilization in the encounter depends on patient demand, the practice environment, and the physician's practice intensity, which is measured by a physician's average effect on healthcare utilization per encounter. We allow for arbitrary matching between patients with different demand and physicians with different practice intensities.

We estimate the model using claims data from a 20-percent random sample of over-65 traditional Medicare beneficiaries from 1998 through 2013. Following the literature, we focus on variation in utilization across hospital referral regions (HRRs). We show how both the model of encounters and the model of per-encounter utilization are identified by patient and physician migration across HRRs, with the per-encounter utilization model also leveraging within-area variation across connected patient-physician pairs (similar to the approach of Abowd et al. 1999).

We build support for our identifying assumptions through event-study analyses of utilization trends in the years before and after cross-HRR moves by both patients and physicians. When physicians move, their average utilization per encounter jumps sharply toward the average in their destination, but closes only around 40 percent of the origin-destination gap. This suggests that a substantial share of geographic variation can be attributed to differences in factors associated with the physician (which do not change on move). When patients move, their average annual utilization also jumps discretely, closing roughly half of the origin-destination gap, consistent with the earlier findings of Finkelstein et al. (2016). In both cases, the event studies show little systematic trend either before or after the move.

Our main results leverage the model estimates of average physician, patient, and practice environment effects in each HRR. Consistent with the motivating event studies, the model estimates indicate that physician practice intensity varies substantially across HRRs. Changing the physician practice intensity of an otherwise-average HRR from the 10th to the 90th percentile of physician practice intensity would increase average utilization per encounter by 33 percent. Using a sequential decomposition of the sources of differences across HRRs in annual per-capita healthcare utilization, we show that physician practice intensity is the major supply-side factor driving these differences. The distribution of physicians with different practice intensities explains at least a third of the difference in utilization between above- and below-median utilization HRRs, while differences in the practice environment explain less than 15 percent. We show that the role of physicians would be even larger if we relax a conservative assumption that physician practice intensity does not affect the number of encounters patients have. Consistent with the findings of Finkelstein et al. (2016), the remaining half of the differences across HRRs is explained by differences in patient demand.

A key finding that underpins this decomposition is that differences across areas in any non-random matching of patients to physicians with different practice intensities plays only a small role in explaining differences in utilization across areas. Systematic differences across areas in the propensity for higher-intensity physicians to attract more patients, conditional on the distribution of physicians in an area and the number of patient encounters with physicians, could contribute to differences in utilization. Since we do not model the sorting process explicitly, we cannot attribute any such differences in sorting separately to patients, physicians, or practice environment. However, we find that the overall role of sorting is small, explaining only about 6 percent of differences in utilization between above- and below-median utilization HRRs, and, as a result, different ways of attributing it do not substantively affect our findings. In our sequential decomposition, we first consider a counterfactual in which we replace actual sorting patterns with random patient-physician matching. The additional counterfactuals that define the decomposition then equalize physician, patient, and practice environment factors in turn, maintaining the assumption of random matching.

We further decompose the role of physician practice intensity into differences in the mix of specialties and differences in practice intensity within specialty. We refer to the latter as variation in *physician practice style*. Prior work shows that higher spending areas of the US tend to have a higher percentage of specialists relative to primary care physicians (e.g. Chernew et al. 2009; Baicker and Chandra 2004); a higher percentage of physicians who are specialists is also a feature of the high-spending US as a whole relative to other OECD countries (Anderson et al. 2019). The literature has also shown that treatment decisions of physicians of the same specialty facing similar patients can vary widely,[1] possibly reflecting underlying differences in physician preferences, skill, or beliefs about the costs and benefits of different medical choices (e.g. Reinhardt 2019; Orszag 2011). We find that about three-fifths of the cross-area differences due to physicians reflects within-specialty differences in practice style, while the remaining two-fifths reflects cross-area differences in specialty mix. Areas with a higher share of specialists relative to PCPs, in particular, tend to have a higher average physician component.

We then provide additional evidence on the correlates of our physician practice intensity estimates. Areas with higher physician practice intensity tend to be in the Southeast and Northeast, while areas with lower physician practice intensity tend to be in the Midwest and Northwest. Survey-based measures of practice intensity from Cutler et al. (2019) are correlated in an intuitive way with our estimated effects: areas where physicians respond to patient vignettes by recommending more follow-up care and/or more aggressive end-of-life care have higher estimated practice intensities. Areas with a higher average physician practice intensity also tend to have higher-quality hospitals and more hospital beds per capita.

We also conduct separate decompositions of the relative roles of physicians, practice environment, and patient demand in driving cross-area differences in utilization for the two largest physician specialties: primary care physicians (PCPs) and cardiologists. Both specialties have attracted

[1] See for example: Chan et al. (2022b); Currie et al. (2016); Currie and MacLeod (2020); Currie and Zhang, forthcoming; Epstein and Nicholson (2009); Fadlon and Van Parys (2020); Kwok 2019; Silver (2021); and Van Parys (2016).

significant prior interest (see e.g. Chandra and Staiger 2007; Cutler et al. 2019; Fadlon and Van Parys 2020; Molitor 2018; Ahammer and Schober 2020). For primary care, we find that the role of physicians in driving cross-area differences in primary care utilization is very similar (19 percent) to the full sample within-specialty estimate (20 percent). In contrast, for cardiology utilization we find, consistent with Molitor (2018), that physicians explain little of the cross-area differences, with the practice environment playing a larger role. This finding may reflect the importance of physical capital, such as facilities for cardiac catheterization, in determining how cardiologists practice.

Finally, while our main focus is on the role of physicians, our findings also shed further light on the roles of the practice environment and of patient demand. We estimate that the role of practice environment comes entirely from factors that determine the number of physician encounters, rather than the amount of utilization per encounter. This is consistent with prior evidence showing that the degree of fragmentation in care—the extent to which patients tend to receive care from many distinct providers—varies systematically across areas and is a key correlate of both costs and outcomes (Agha et al. 2019). The role of patient demand also works mainly through demand for the number of physician encounters. This is consistent with findings from the RAND Health Insurance Experiment that insurance coverage generosity affects the number of visits a patient has, but not the intensity of utilization conditional on a visit (Newhouse and the Insurance Experiment Group 1993).

In addition to contributing to the substantial literature on the causes of geographic variation in healthcare utilization, our analysis adds to a growing empirical literature using quasi-experimental changes in location or matching to separately identify individual heterogeneity from the systematic effects of geography or institutions. Prominent examples in the study of healthcare include Song et al. (2010), Finkelstein et al. (2016), Molitor (2018), and Gottlieb et al. (2023). Outside of healthcare, a similar approach has been used to study determinants of many different phenomena, including wage variation (Abowd et al. 1999), wage inequality (Card et al. 2013; Card et al. 2016; Bonhomme et al. 2019), neighborhood effects Chetty and Hendren 2018), teacher effects (Chetty et al. 2014), brand preferences (Bronnenberg et al. 2012), healthy eating (Allcott et al. 2019), voting behavior (Cantoni and Pons 2022), and tax reporting (Chetty et al. 2013), among other topics.

Relative to this literature, we innovate by explicitly modeling the process that determines the number of encounters between cases (e.g. patients, workers) and agents (e.g. physicians, firms) while allowing for a flexible matching process. From this perspective, our work relates to a number of recent studies in the worker-firm setting that model and estimate flexible matching processes (e.g. Abowd et al. 2019; Bonhomme et al. 2019; Hagedorn et al. 2017). While most of this literature decomposes outcome variation in terms of two sources (e.g. patients and places, workers and firms, teachers and students), we introduce and model a third dimension (physicians). Our approach may be useful in other settings in which multiple dimensions are important. This includes, for example, work on education value-added that has separately studied the relative role of teachers vs students (Chetty et al. 2014) and of students vs schools (Kramarz et al. 2015), or work on wage setting that has separately examined the relative role of workers vs firms (e.g. Card et al. 2013), and workers vs locations (e.g. Card et al. 2023).

2 Model

We start by developing a model of annual per-capita healthcare utilization and the potential roles of patient demand, physician practice intensity, and practice environments. Here we outline the model's general structure; in Section 4 we impose additional restrictions to bring the model to data in a tractable regression framework.

Patients stochastically receive health shocks of varying severity. For each shock, the patient chooses whether or not to seek treatment. This decision reflects the patient's latent demand for healthcare, as well as aspects of the practice environment in her area which may affect the availability or accessibility of physicians. If a patient seeks treatment, she is matched to a physician. The physician then chooses the level of healthcare utilization as a function of patient demand, the practice environment, and the physician's own practice intensity. The model thus has two sequential components: a process that generates encounters (or unique "matches") between a patient and a physician, and a process that determines utilization in each encounter. We develop each of these components in turn.

Consider a set of patients across different years. Patient i receives n_{it} unobserved health shocks in year t; each shock k has a latent severity of ψ_{ikt}. This severity reflects both the objective severity of the health condition as well as the patient's proclivity to seek treatment for a given objective level of severity. Whether or not a health shock generates utilization depends on its severity relative to an area-specific threshold, $\breve{\gamma}_j$. Formally, we let $E_{ikt} = \mathbf{1}[\psi_{ikt} > \breve{\gamma}_{j(it)}]$ be an indicator for whether the shock leads the patient to seek treatment, where $\mathrm{j}(it)$ indexes the area of patient i in year t. The number of treated shocks is then given by $N_{it} = \sum_{k=1}^{n_{it}} E_{ikt}$.

The area-specific threshold for an encounter $\breve{\gamma}_j$ may reflect aspects of area j's practice environment, such as capacity (e.g. the number of physicians per capita) or accessibility (e.g. distance to the nearest healthcare provider). It may also depend on characteristics of an area's physicians, such as their propensity to refer patients to specialists. Some of these characteristics might more appropriately be attributed to the role of the physician; our baseline case conservatively attributes them to the non-physician supply-side factors (i.e. the practice environment). We discuss in Section 5 how much alternative attributions would increase our estimate of the role of physicians in driving geographic variation.

If a shock leads to treatment, the patient is matched to a physician d. For now we leave this matching process unspecified, allowing it to depend arbitrarily on (i,k,t). Throughout, for simplicity, we group all visits that a given patient has with a given physician in a given year into a single "encounter." We thus model the number of different physicians a patient sees in a year, but do not model how utilization is distributed across different visits within a physician-patient-year. Letting $D_{idt} = 1$ if patient i matches to doctor d in year t, we can equivalently write the number of treated shocks as the number of encounters, i.e., $N_{it} = \sum_d D_{idt}$.

We model patient i's utility from an encounter with utilization y as $u_{idt}(y) = a_i y - \frac{1}{2}(y - h_{idt})^2$ where h_{idt} is an objective level of appropriate utilization and a_i is a patient-specific demand parameter. The net cost of utilization to the physician (in utility units) is given by $c_{idt}(y) = (c_{\mathrm{j}(it)} + g_{idt})y$, where higher values of g_{idt} denote higher marginal costs and may include factors such as liability concerns (e.g. Currie and MacLeod 2020), or the opportunity cost of physicians' time. The c_j term denotes costs that are specific to location j; these may reflect available physical capital, the prevalence of nonprofit hospitals, the liability environment, peer effects among doctors, and organizational culture (e.g. Lee and Mongan 2009), among other factors.

Each physician also has an individual parameter δ_d which characterizes her practice intensity. These are our key objects of interest. Differences across physicians in δ_d may both reflect physician specialty and within-specialty differences in practice style, and we will examine the relative role of each. Within-specialty differences in turn may reflect heterogeneous beliefs about appropriate or effective treatment, such as the "cowboy" or "comforter" approaches to care documented in the survey evidence of Cutler et al. (2019), as well as heterogeneity in physician skill or experience (e.g. Chandra and Staiger 2007).

In each encounter, the physician chooses patient utilization, y_{idt}, to maximize the sum of patient utility and physician utility net of costs. Specifically, we assume she chooses utilization in encounter (i,d,t) as:

$$y_{idt} = \arg\max_{y} \{u_{idt}(y) + \delta_d y - c_{idt}(y)\} = a_i + \delta_d - c_{\mathrm{j}(it)} + h_{idt} - g_{idt}. \tag{1}$$

Below we estimate the two sequential components of the model: encounter generation and healthcare utilization conditional on an encounter. We then use the estimates of key model parameters to decompose average utilization in each area into components driven by patient demand, physician practice intensity, and practice environment.

3 Data and Preliminary Evidence

3.1 Data and Variable Definitions

We outline the data and key variables of interest here; Appendix A provides additional detail. Our geographic unit of analysis is a hospital referral region (HRR), as defined by the 1998 Dartmouth Atlas. HRRs are collections of zip codes aggregated to approximate a tertiary hospital market.

We analyze a 20 percent random sample of traditional Medicare beneficiaries ("patients") from 1998–2013. We observe all Medicare claims and associated physicians for each patient in the sample. Our primary outcome of interest is healthcare utilization, which we construct following the existing literature (e.g. Finkelstein et al. 2016; Gottlieb et al. 2010). Specifically, we use the claims data to build an index of healthcare utilization by adjusting claim expenditures for differences across areas in administratively set prices. We also observe basic patient demographics, including age, sex, race, zip code of residence, and Medicaid coverage (a proxy for low income). The claims data also allow us to impute each physician's clinical specialty.

An important issue is how to assign claims to specific physicians. This involves some judgment calls, as Medicare claims may list multiple physicians with varying degrees of involvement in treatment. Our objective is to associate claims with the physician who is most likely responsible for the treatment decision. Most of the medical care one naturally associates with a physician (e.g. an office visit, the ordering of a lab test, or a surgical procedure done in either an inpatient or an outpatient setting) is contained in a file (known as the "carrier file") which lists both a performing physician and a referring physician. We assign the performing physician to the claim if it is an evaluation and management claim—the nomenclature for a typical "office visit"—and we assign the referring physician to all other claims (e.g. claims involving imaging, testing or procedures). In practice, this means that if a physician refers a patient for a blood draw, a CT scan, or a colonoscopy, the additional utilization is attributed back to the referring physician.[2] However, if a physician refers a patient to a different physician for evaluation (e.g. a PCP referring a patient to a cardiologist for follow-up care), the subsequent evaluation is attributed to the new physician, not the referring one. This approach is perhaps least well-suited to primary care physicians, whose biggest influence on their patients' medical spending may come through their tendency to refer their patient to specialists for an evaluation; we therefore also undertake an alternative analysis in Section 5 in which we instead attribute the spending from such referrals back to the referring PCP. The remaining Medicare claims files consist of claims by hospitals for inpatient stays and outpatient facility charges. Here, we assign each claim to the attending physician, since the attending physician is defined as "the individual who has overall responsibility for the patient's medical care and treatment."[3]

As noted above, our analysis distinguishes between the process that generates encounters between physicians and patients and the process that generates utilization conditional on the encounter. We define a patient-physician encounter by aggregating all unique claims attributed to a particular patient-physician pair in each year. Thus, the number of encounters a patient has in a year corresponds to the number of unique physicians she sees that year.

To define a patient's location, we follow Finkelstein et al. (2016) and use the patient's zip code of residence, regardless of where the claim is incurred. We categorize a patient as a "non-mover" if her HRR of residence does not change during our sample period, and as a "mover" if her HRR of residence changes exactly once. To match the timing at which we observe each patient's residence,

[2]The referring physician is almost never missing for testing and imaging claims, but is missing for about 15 percent of procedure claims (see Appendix Table A1 Panel B). If the referring physician is missing, we assign the performing physician (who is virtually never missing). Overall, 2 percent of our encounters (corresponding to 0.29 percent of utilization) have a missing physician identifier; we assign them a missing physician ID unique by HRR.

[3]See https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c25.pdf; downloaded on 03/08/2022. The attending physician is listed for over 99.9 percent of claims.

we define all outcomes for year t to be totals of claims submitted between April 1 of year t and March 31 of year $t+1$.

We define a physician's location based on the location of the patients she treats. We categorize a physician as a "mover" if she exhibits exactly one clear shift in location during our sample—more precisely, when we observe the physician first in an origin location where at least 75 percent of their patients live in a given year and subsequently in a destination location where at least 75 percent of their patients live in a given year. Otherwise, the physician is categorized as a "non-mover."

3.2 Sample Restrictions and Summary Statistics

Starting from the 16.7 million unique patients we observe between 1998–2013, we impose several restrictions to arrive at our baseline analysis sample; these are again discussed in more detail in Appendix A. We first focus on a 25-percent random sample of non-mover patients to simplify computation of the model. We further exclude all patient-years where the patients are younger than 65 or older than 99, where the patients are enrolled in Medicare Advantage, or where patients are not subscribed to Medicare Part A and B for all months in a year. Following Finkelstein et al. (2016), we exclude the small number of patients whose HRR of residence changes more than once, along with a small number of patient movers for whom the location of observed claims does not clearly shift from the origin location to the destination location. Analogously, some physicians in our sample are observed treating patients from multiple HRRs in a given year without satisfying the definition of a mover, likely reflecting the fact that physicians may simultaneously practice at several facilities and that patients may travel outside of their home HRR to seek care. If a physician treats patients located in multiple HRRs within a year (other than in a mover physician's move year), we treat her as a distinct physician in each HRR in that year. We do so to avoid using such within-physician variation in location to identify physician effects. As we discuss in Section 5 below, our results are not sensitive to instead limiting our analysis to encounters that take place in the HRR where the physician had the plurality of their encounters in each year. Finally, we restrict to the largest connected set of physicians, places, and patients; in practice this set includes over 99 percent of encounters.[4]

[4]Formally, our analysis sample is the largest set of patients and physicians who are "connected" by a path of observed encounters within HRRs and across time. This follows Abowd et al. (2002), who study sets of workers and firms connected by employment spells.

Table 1: Sample Summary Statistics

	A. Patients		B. Physicians	
	(1)	(2)	(3)	(4)
	Non-movers	Movers	Non-movers	Movers
Share female	0.56	0.60		
Share white	0.85	0.88		
Mean age first observed	71.00	72.54		
Share first observed residence:				
Northeast	0.20	0.17	0.23	0.21
South	0.39	0.42	0.34	0.36
Midwest	0.25	0.20	0.23	0.23
West	0.16	0.22	2.98	3.27
Number of chronic conditions:				
Mean	2.15	2.04		
S.D.	0.20	0.20		
Annual utilization/utilization per encounter:				
Mean	$7,678	$7,391	$925	$1,186
S.D.	$11,916	$9,599	$2,475	$2,443
Annual number of encounters†:				
Mean	6.44	6.89	209.72	250.44
S.D.	4.75	4.34	638.63	246.62
Number of patients	2,440,041	650,440	1,628,408	74,934

Notes: In Panel A, rows for female, white, age first observed, and first observed residence report the shares of patients with the given characteristics among movers and non-movers. The sample is the baseline sample of all patient-years excluding the move year for patient movers (N = 23,167,425 patient-years; note that the sample includes a 25-percent random sample of non-mover patients). Panel B has the analogous statistics for the sample of physicians (N = 10,765,990 physician-years). Utilization per encounter, the annual number of encounters, and the number of doctors are computed on the original sample of physicians. However, since physicians are often observed performing in multiple HRRs (which may be in multiple regions of the country; see Appendix C.1), we "fragment" physicians into multiple observations (one for each HRR) in order to compute the shares of first observed residence.
†: denotes that statistic has been multiplied by 5 for physicians to account for the fact that we have a 20-percent sample of Medicare patients.

Our baseline analysis sample contains 159 million encounters between 3 million patients and 1.7 million physicians.[5] We characterize about 650,000 patients as movers and about 75,000 physicians as movers. Table 1 provides some summary statistics for this sample.Panel A of Table 1 shows that patient movers and non-movers are broadly similar, although movers tend to be slightly older and less healthy, and more likely to be female, white, and living initially in the South and

[5]However, because as noted above, we assign different "physician IDs" to physicians treating patients from multiple HRRs in the same year, we have 8 million physician IDs in our sample, and we estimate and analyze that many physician practice intensity parameters δ_d. We show below that different ways of handling these physicians yield similar results.

West. Patients have on average around \$7,600 in utilization a year, and see about six or seven different physicians annually.

Panel B shows that physician movers and non-movers have a similar geographic distribution. Non-mover physicians have a slightly lower average utilization per Medicare patient of \$925 compared with \$1,186 for movers.[6] Non-mover physicians see around 210 Medicare patients annually, while mover physicians see 250. Part of these differences are mechanical. Specifically, our algorithm for detecting moving physicans requires them to have at least four encounters per year in the four years before and after their move, but this requirement is not imposed on non-movers. Imposing this restriction on non-movers (i.e. dropping all physician-years with fewer than four encounters) results in a sample that has an average utilization per encounter of \$949 and 222 encounters per year on average, which are closer to the respective averages for movers. Appendix Table A2 further indicates that for both non-mover and mover-physicians, the single largest physician "specialty" category is primary care, but that mover physicians are disproportionately in primary care—33 percent of mover physicians accounting for 43 percent of mover physician utilization are in primary care, compared to 16 percent of physicians and 36 percent of utilization for non-mover physicians. Appendix Table A4 shows that, where we can observe additional physician demographics, mover physicians tend to be younger than non-movers at the time of their move, more likely to be female and have spent fewer years practicing. Such differences are unlikely to pose threats to our empirical strategy, which conditions on patient and physician fixed effects. As discussed more in Section 4.3, our empirical strategy allows the distributions of fixed effects of both physician and patient movers to be arbitrarily different from those of non-movers.

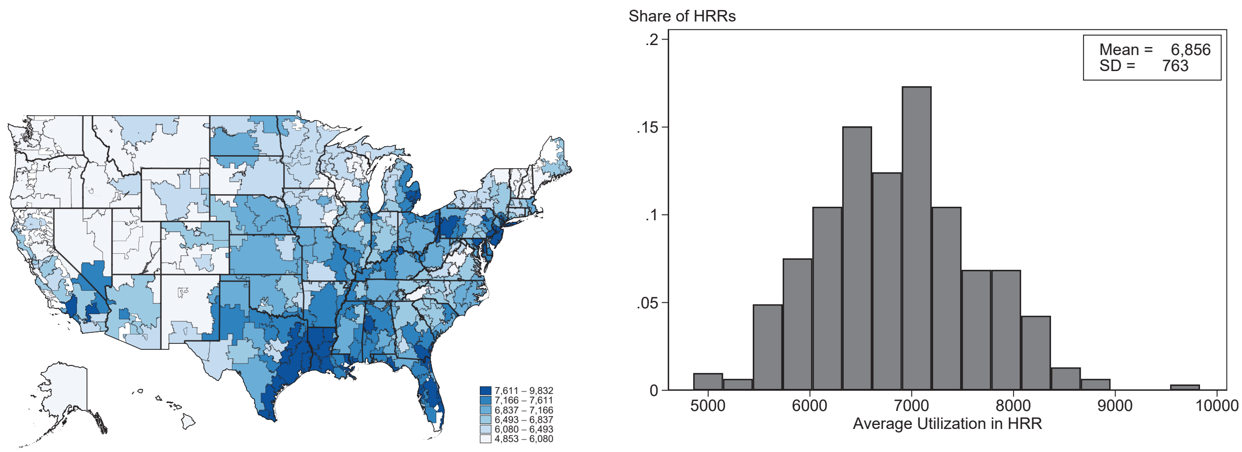
Figure 1 shows the distribution of average annual patient utilization across HRRs. The overall average across HRRs is \$6,856 with a standard deviation of \$763, over 10 percent of the mean. The figure illustrates a high degree of geographic variation, with the South and Midwest outpacing lower-utilization areas in the West and Northeast.[7]

---

[6]We drop from the summary statistics 8 non-mover physician IDs which collect all utilization for multiple disparate physicians either because their recorded ID is missing or because they were assigned a so called "surrogate" Unique Physician Identifier Number (UPIN), used to record utilization for physicians who do not have their own UPIN yet, e.g. because they are still residents ("OTH000", "PHS000", "RES000", "SLF000", "RET000", "INT000", "VAT000"). The utilization associated with these IDs is quite large (since they contain many disparate physicians) and skews the summary statistics for non-movers substantially.

[7]The geographic distribution of utilization remains fairly stable across years of our sample. For example, the rank correlation between an HRR's utilization in the first and second half is 0.9.

Figure 1: Distribution of Annual Patient Utilization Across HRRs



Notes: The map shows the distribution of average annual patient utilization by HRR, in sextiles defined in the legend. The histogram shows the distribution of HRRs' average annual patient utilization. The sample is the baseline sample of all patient-years excluding the move year for patient movers (N = 23,167,425 patient-years).

To decompose the sources of the geographic variation in Figure 1, our empirical strategy leverages the cross-HRR migration of both patients and physicians. Figure 2 illustrates the nature of these moves by showing the gap in average utilization between the origin and destination HRRs of patient movers (Panel A) and physician movers (Panel B). Both distributions appear symmetric, indicating no systematic imbalance in moves from high- to low-utilization areas. The standard deviations are also substantial. For example, the standard deviation of the gap in average annual patient utilization between a patient's origin and destination HRRs is $950, relative to an HRR-average annual patient utilization of $6,856.

A natural question is why patients and doctors move in this sample. For patients, data from the Health and Retirement Survey and the Longitudinal Survey of Aging—both of which ask individuals for their reason to move—lead to similar conclusions. The most frequently reported reason for moves among patients in the relevant age group is to be near/with children or other kin, followed by health reasons, financial reasons, or other amenities (Finkelstein et al. 2016; Choi 1996). For physicians, a 2012 report to the American Medical Group Association found that the most common reasons for migration include financial considerations, dissatisfaction with current practice, and significant personal life changes.[8] Our empirical strategy, as detailed below, allows the timing

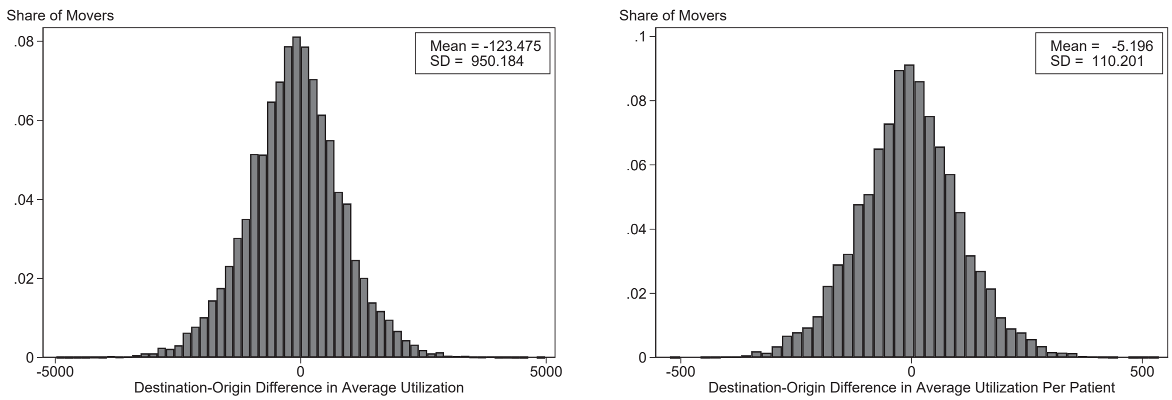
[8] See https://web.archive.org/web/20190516130533/http://www.ericksson.net/surveys.asp.

of these moves to be non-random while assuming that the difference in average utilization between the destination and origin HRR is not systematically related to underlying trends in patient health or physician practice patterns.

Figure 2: Destination-Origin Gaps in Average Patient Utilization

A. Patients

B. Physicians



Notes: This figure shows the difference in HRR-average annual utilization between the origin and destination HRRs of patient (Panel A) and physician (Panel B) movers. Note that Panel A displays differences in the HRR's average patient-level utilization between destination and origin, while Panel B displays differences in the HRR's average physician utilization per patient, so the scales are naturally quite different. The samples are all patient movers (Panel A, N = 650,440 patients) and physician movers (Panel B, N = 74,934 physicians).

3.3 Illustrative Event Studies

Before parameterizing and estimating the per-encounter utilization model from Section 2, we present some descriptive results on average changes in utilization when physicians or patients move across HRRs. This analysis is meant to illustrate key sources of variation we will use to estimate the model, provide some support for the identifying assumptions, and preview our main findings on the impact of physician practice intensity. The approach we employ here is in the spirit of earlier analysis by Finkelstein et al. (2016) and Molitor (2018), who respectively study utilization changes for Medicare patients and cardiologists who move across HRRs. We show in Appendix B.5 how these event study analyses can be motivated by restricted versions of the model in Section 2.

For the patient mover analysis, we aggregate over the set of doctors each patient sees in a year, $\mathscr{D}_{it}$, to obtain log patient-year utilization y_{it}. We normalize this measure to zero for the small portion (5.2 percent) of patient-years with zero healthcare use. Following Finkelstein et al. (2016), we then estimate the regression

$$y_{it} = \alpha_i^P + \tau_t^P + \theta_{\mathrm{r}(i,t)}^P \Delta_i^P + x_{it}' \beta^P + \eta_{it}^P, \tag{2}$$

where Δ_i^P is the difference in average y_{it} between the patient's destination and origin HRR (normalized to zero for non-movers). We include a vector of patient fixed effects (α_i^P) and a control vector of time-varying patient observables x_{it} consisting of indicators for five-year age bins, as well as indicators for years relative to patient i's move. The relative year $\mathrm{r}(i,t)$ is defined relative to patient i's move year and all of the relative year indicators are set to zero for non-movers. We also include a vector of calendar year t indicators τ_t^P. The main coefficients of interest θ_r^P are on the interactions between relative year indicators and the difference in average log-utilization between a patient's destination and origin HRR; these capture how average annual patient log utilization changes in the years preceding and following a move across HRRs, as a share of the average observed difference in this outcome between the destination and origin HRRs Δ_i^P. For example, a finding of $\theta_r^P \approx 0$ for $r \geq 0$ would indicate that post-move utilization changes are not systematically related to the observed regional differences in utilization, while a finding of $\theta_r^P \approx 1$ would suggest utilization changes closely track the observed regional variation. Intuitively, the former (latter) finding would suggest HRRs play a small (large) role in observed utilization variation.

This specification allows for movers to differ arbitrarily from nonmovers in both levels of log utilization (via α_i^P) and in trends in log utilization around their moves (via the relative year indicators in x_{it}), such as would occur if moves were associated with positive or negative health shocks. The key assumption required to interpret the event-study jump as a causal effect of moving to high- or low-utilization areas is that there are no shocks to utilization that coincide exactly with the timing of the move and that are correlated with utilization in the origin and destination. We can investigate the plausibility of this assumption in the event study results; deterioration in health status that occurs gradually and is correlated with utilization in the destination and origin would tend to show up as differential pre-trends in the event study analysis.

The specification also assumes that patient demand and supply-side factors are additively separable in the equation for log utilization.[9] We see this as a plausible economic assumption. It has the intuitive implication that demand and supply characteristics affect the level of utilization multiplicatively, and thus that the (level) utilization of patients who are sick or prefer intensive care

[9]See Finkelstein et al. (2016) for a formal derivation.

(i.e. have high α_i^P) will vary more across places than that of patients who are healthy or rarely seek care (i.e. have low α_i^P). We also see the log model as appealing on econometric grounds, given the skewed cross-sectional distribution and large secular trend of utilization.

For the physician mover analysis, we likewise aggregate over the set of patients each physician d sees in year t to obtain log doctor-year utilization y_{dt}, normalizing to zero the 0.6 percent of doctor-years with zero healthcare use. We take as the outcome of interest annual average log utilization per patient, or $y_{dt} - \ln N_{dt}$. Following Molitor (2018), we estimate:

$$y_{dt} - \ln N_{dt} = \alpha_d^D + \tau_t^D + \theta_{\mathrm{r}(d,t)}^D \Delta_d^D + w_{dt}' \phi^D + \eta_{dt}^D, \tag{3}$$

where Δ_d^D is the difference in average $y_{dt} - \ln N_{dt}$ between the physician's destination and origin HRR (again normalized to zero for non-movers). We include a vector of doctor fixed effects (α_d^D), calendar year t indicators τ_t^D, and a control vector of time-varying physician observables (w_{dt}) consisting of indicators for years relative to doctor $d's$ move year (relative years are denoted by $\mathrm{r}(d,t)$ and set to zero for non-movers). The main coefficients of interest θ_r^D are on the interactions between the relative year indicators and the difference in average log utilization per patient between a physician's destination and origin; these capture how average annual physician log utilization per patient changes in the years preceding and following a physician's move across HRRs, as a share of the average observed difference in this outcome between the destination and origin HRRs Δ_d^D.

The doctor event study is similar to the above patient analysis. In particular, it allows for mover and non-mover physicians to differ arbitrarily in both levels of average log utilization per patients (via α_d^D) and trends in this utilization around their moves (via w_{dt}). Causal interpretation requires that there are no shocks to physician log per-patient utilization that both coincide exactly with the timing of the move and are correlated with average per-patient utilization in the origin and destination. Furthermore, it assumes that average log utilization per patient is additively separable in physician practice intensity and other factors (reflecting both patient demand and non-physician practice intensity).

Figure 3 shows the results from estimating the patient-mover event study in equation (2) (Panel A) and the physician mover event study in equation (3) (Panel B). Each figure plots estimates of the relative-year coefficients from the respective regression specifications.[10] The patient-mover

[10]Observations for each relative year preceding and including relative year -6 are binned to a single indicator, as are all observations for each relative year including and after relative year 6. This was done because, as Appendix Figure A1 shows, we rarely observe patients in these extreme relative years. Appendix Figure A2 shows corresponding event

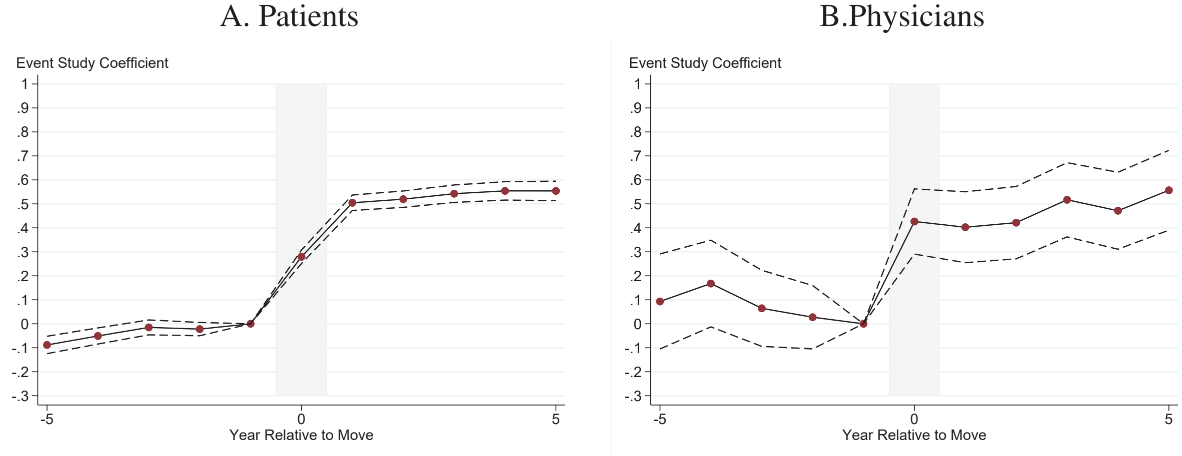
event study shows that the average utilization of movers is stable in the years preceding a move, conditional on the controls, while in the years following a move average patient utilization changes sharply in the direction of the observed difference in average HRR utilization. The lack of pronounced pre-trends is consistent with the identifying assumption we will use to estimate the model below.[11] The jump in the event study after the move provides an initial indication of the importance of patient demand in driving differences across areas in health care utilization. When patients move, they take with them the factors that affect their demand—such as their health and their preferences for medical care. Therefore, if all of the geographic differences in health care utilization were due to differences in patient demand, we would expect patient utilization to remain constant after the move. Conversely, if patient demand played no role and geographic differences were entirely driven by local supply side factors which change discretely on a move (including both geographic differences in physician practice intensity and geographic differences in practice environment), we would expect a jump of 1 as patient utilization resets to the new location's average. Quantitatively, we estimate an event study jump of around 0.5-0.6, similar to the earlier finding of Finkelstein et al. (2016).

The physician-mover event study in Panel B of Figure 3 shows that average per-patient utilization of moving physicians is also relatively stable in the years prior to a move. Following a move, physician utilization per patient changes sharply. Once again, the size of the jump provides an initial indication of the role of physician practice intensity in driving geographic differences, with a smaller jump indicating a larger role for physicians relative to patient demand and practice environment which change discretely on move. The magnitude of the jump is equal to 0.4-0.5. Of course, this speaks to the role of physician practice intensity for a typical physician mover, who may be different from a typical physician.

studies estimated on a balanced panel of patients or physicians that we observe for all relative years between -5 and 1, and a balanced panel of patients or physicians that we observe for all relative years -1 to 5. We see similarly flat pre-trends and qualitatively similar post-event study effects, though confidence intervals for the physician analysis grow large with the substantial reductions in the number of movers. In Appendix C.4, we show that these results are robust to allowing for heterogeneous treatment effects by the timing of the treatment (i.e. move year) in the spirit of recent advances in the two-way fixed effects literature (e.g. Callaway and Sant'Anna 2021; Sun and Abraham 2021).

[11] One possible source of bias is endogenous moves, caused by patients seeking better care due to worsening health status. While we cannot fully rule out this possibility, the patterns we observe in the data suggest it is likely to be small. Gradual worsening of health status that would lead to eventual relocation would tend to show up as pre-trends in our motivating event studies. Although sudden negative health shocks that cause immediate movement to a more intensive area might occur without causing pre-trends, such changes might lead to a spike in utilization immediately following a move. However, we also find relatively flat post-trends in our event study analysis.

Figure 3: Patient and Doctor Mover Event Studies



Notes: Panels A and B show the estimated θ_r^P and θ_r^D coefficients in equations (2) and (3) for patient and doctor movers, respectively. The coefficients for relative year -1 are normalized to 0. The dependent variable in Panel A is log annual patient utilization, and the control vector includes indicator variables for five-year age bins and relative-year main effects for movers. The dependent variable in Panel B is log annual physician utilization per patient, and the control vector includes relative-year main effects for movers. Dashed lines indicate upper and lower bounds of the 95 percent confidence intervals, clustered at the person (i.e. patient or physician) level. The sample is all patient-years (Panel A, N = 23,663,477 patient-years) or physician-years (Panel B, N = 23,788,172 physician-years) binning observations that are more than 6 years before the move year into a single indicator and the observations more than 6 years after the move year into a separate indicator; the coefficients on these indicators are not plotted here.

In Appendix Figure A3, we show that the initial size of this jump varies in an intuitive manner for older and younger physician movers. Specifically, younger movers exhibit a larger jump than older movers in the years immediately after the move. This signals that younger physicians adjust to patient and non-physician supply side factors in their destination more quickly than older physicians do, perhaps because they are less entrenched in certain practice styles. However, the event study jump eventually converges to about 0.5 for older physicians as well. Weighting movers to resemble non-movers on observables such as age, gender, and specialty does not noticeably change our results (see Appendix Figure A4 and Appendix Table A3).

Together, the two event studies in Figure 3 suggest a meaningful role for each driver of geographic differences in utilization (patients, physicians, and non-physician supply-side factors). However, the event studies alone do not allow us to quantify the relative importance of each. For example, while the fact that the jump in the physician-mover event study is far from one is indicative of a role for physician practice intensity, it admits two very different interpretations of this physician effect which have different implications for counterfactuals. To the extent that the physician effects reflect differences across physicians in the amount of utilization they produce conditional on an encounter with a given patient (i.e. the δ_d's in equation (1)), changes in the distribution of physicians across areas would change the distribution of utilization across areas. However, to the extent the physician effects reflect differences across physicians in their propensity to attract high demand patients (i.e. patients with high ψ_{ikt}), changing the distribution of physicians across areas need not affect the distribution of utilization. The patient and physician event studies are also defined at different levels of aggregation—patient-year and physician-encounter, respectively—so their results cannot be directly combined in a single decomposition. Our model, which we estimate next, provides a unified framework to address these issues.

4 Empirical Strategy

4.1 Estimation

We make several parametric restrictions to bring the model in Section 2 to data. We first specify the total number of physician encounters, $N_{it} = \sum_{k=1}^{n_{it}} \mathbf{1}[\psi_{ikt} > \check{\gamma}_{it}]$, of patient i in year t as a Poisson random variable, with mean

$$E[N_{it} \mid x_{it}, \mathrm{j}(it)] = \exp\left(\alpha_i^N + \gamma_{\mathrm{j}(it)}^N + \tau_t^N + x_{it}'\beta^N\right). \tag{4}$$

The area-specific term γ_j^N may reflect both practice environment factors (such as the density of local hospitals) and the geographic distribution of physician practice intensities (such as the propensity of PCPs to refer patients to specialists).[12] Equation (4) specifies a two-way fixed-effect Poisson regression, which can be estimated from the number of encounters N_{it} that a patient has each year, her location $\mathrm{j}(it)$, and the observables x_{it}. We discuss identification from patient migration below.

We will use equation (4) to study the component of encounter quantity attributable to patients, which we denote by $\tilde{\alpha}_{it}^N$, and the component attributable to practice environment and physicians, which we denote $\tilde{\gamma}_j^N$. Both components are derived from the nonlinear model (equation (4)) by setting other parameters to their average, such that differences in the components capture average marginal effects. Thus the patient component of the encounter model $\tilde{\alpha}_{it}^N$ shows how the expected number of encounters changes in a *typical* HRR if we were to change the patient population only, and the place component of the encounter model $\tilde{\gamma}_j^N$ shows how a *typical* patient's expected number

[12] Appendix B.1 microfounds this assumption via parameterizations of the stochastic process that determines a patient's number of health shocks n_{it} and the latent severity of each shock ψ_{ikt}.

of encounters changes if only the intensity of her HRR is allowed to vary.[13] These are objects that we will equalize in various counterfactuals below.

We next parameterize the per-encounter utilization model. We assume the combination of patient health and physician cost of providing care (i.e. $h_{idt} - g_{idt}$ in equation (1)) can be forecasted by a time effect, a patient effect, a place effect, and sets of time-varying patient and physician observables x_{it} and w_{dt} given encounter locations. This implies that we can express utilization y_{idt} among realized encounters (with $D_{idt} = 1$) as

$$y_{idt} = \underbrace{\alpha_i + \tau_t + x_{it}'\beta}_{\equiv \tilde{\alpha}_{it}} + \underbrace{\delta_d + w_{dt}'\phi}_{\equiv \tilde{\delta}_{dt}} + \underbrace{\gamma_{\mathrm{j}(it)}}_{\equiv \tilde{\gamma}_{\mathrm{j}(it)}} + \varepsilon_{idt}, \tag{5}$$

with $E[\varepsilon_{idt} \mid x, w, \mathrm{j}(it), D_{idt} = 1] = 0$.[14] Equation (5) specifies a linear fixed effects regression for per-encounter utilization y_{idt} in terms of a patient component $\tilde{\alpha}_{it}$, a physician component $\tilde{\delta}_{dt}$, and a practice environment component (capturing non-physician supply factors) $\tilde{\gamma}_j$. Again, these are objects that we will equalize in various counterfactuals below.[15]

We measure per-encounter utilization y_{idt} as the log healthcare utilization of patient i with physician d in year t. This is only observed (or defined) for the subset of patient-physician matches that actually take place that year. Our specification (5) assumes that log utilization is additively separable in patient demand, physician practice intensity, and the practice environment.[16]

Estimation of the per-encounter utilization model in equation (5) requires the assumption that $E[\varepsilon_{idt} \mid x, w, \mathrm{j}(it), D_{idt} = 1] = 0$. This will be violated if this residual from the utilization equation is correlated with the health-shock intensity residual in the encounter model (ξ_{idt}), after conditioning on observables (specifically, on the fixed effects, x, w, and the vector of $\mathrm{j}(it)$). The assumption

[13] Specifically, the patient component of the encounter model $\tilde{\alpha}_{it}^N$ refers to the expected number of encounters in a Poisson model with mean $\tilde{\alpha}_{it}^N \equiv \exp\left(\alpha_i^N + x_{it}'\beta^N + \tau_t^N + \bar{\gamma}^N\right)$, where $\bar{\gamma}^N$ is the sample average of the place effect estimates $\gamma_{\mathrm{j}(it)}^N$ from equation (4). Similarly, the practice environment component of the encounter model $\tilde{\gamma}_j^N$ refers to the expected number of encounters in a Poisson model with mean $\tilde{\gamma}_j^N \equiv \exp\left(\bar{\alpha}^N + \bar{x}'\beta^N + \bar{\tau}^N + \gamma_{\mathrm{j}(it)}^N\right)$, where $(\bar{\alpha}^N + \bar{x}'\beta^N + \bar{\tau}^N)$ is the sample average of the patient effect and observable estimates $\alpha_i^N + x_{it}'\beta^N + \tau_t^N$ from equation (4).

[14] Once again, Appendix B.1 provides additional details on this derivation.

[15] The non-linearity of the Poisson encounter model requires us to transform the estimated fixed effects into average marginal effects in order to report their variation on a meaningful scale. However, the linear per-encounter utilization model allows for the effects to be reported directly, with the standard deviations directly interpretable as standard deviations in average effects on log per-encounter utilization.

[16] As discussed in the context of similar assumptions for the event study analyses in Section 3.3, this has the intuitive implication that these factors affect the level of utilization multiplicatively. Thus, for example, the level of utilization for patients who are sicker or otherwise prefer more intensive care (i.e. have higher α_i) will vary more across physicians with different practice intensities (δ_d) than for patients who are healthy or dislike intensive treatment (i.e. have low α_i).

thus requires that any such dependence is fully accounted for by time, patient, and place effects, along with the time-varying patient and physician controls. Our model furthermore imposes that the patient-physician matching process depends systematically on only these same factors. As we discuss in the next section, identification of equation (5) follows under these assumptions, given quasi-experimental movement of patients and doctors across areas.

In our baseline specification of the encounter model and the per-encounter utilization model (equations (4) and (5)), the patient observables x_{it} consist of indicator variables for five-year age bins and relative-year fixed effects $\rho_{\mathrm{r}(i,t)}$ for patients who move between HRRs (recall we normalize $\rho_{\mathrm{r}(i,t)}$ to zero for non-movers). The physician observables w_{dt} include similar relative-year fixed effects for physician movers, again normalized to zero for non-movers. We estimate the parameters of the encounter model (equation (4)) by a two-way fixed effect Poisson regression in the full sample of patient-years, both those with and without physician encounters.[17] We estimate the parameters of the per-encounter utilization model (equation (5)) by a three-way fixed effects linear regression on the full set of physician-patient matches that occur each year.

4.2 Aggregation

Our goal is to use the estimates from the encounter model in equation (4) and the per-encounter utilization model in equation (5) to write HRR-average annual patient log utilization in terms of the components attributable to patients, physicians, and practice environment. This will then allow us to decompose differences in utilization across areas into shares due to each of these factors.

To this end, let $\mathscr{D}_{it}$ denote the set of physicians d which patient i sees in year t with $N_{it} = |\mathscr{D}_{it}|$. Aggregating over this set, we write realized annual log patient utilization in terms of the model as

$$y_{it} = \ln\left(\sum_{d\in\mathscr{D}_{it}} \exp y_{idt}\right) = \tilde{\alpha}_{it} + \tilde{\gamma}_j + \ln N_{it} + \ln\left(\frac{1}{N_{it}}\sum_{d\in\mathscr{D}_{it}} \exp\left(\tilde{\delta}_{dt} + \varepsilon_{idt}\right)\right), \tag{6}$$

where we normalize annual log utilization y_{it} to zero when $N_{it} = 0$.[18]

To characterize the role of physicians in per-encounter utilization, we further decompose the final term of this equation. Let $\mathscr{D}^*(n,j,t)$ denote a random set of physicians of size n practicing in

[17]Hausman et al. (1984) establish the consistency of conditional maximum likelihood estimation of such models, which we implement using the algorithm of Guimaraes (2014).

[18]The model of annual log patient utilization (equation (6)) admits an event study representation, which can be used to visualize the components of the decomposition and assess the identifying restrictions via conventional pre-trend checks. This complementary event study analysis, which builds on the simpler event study analysis in Section 3.3, is described in Appendix B.6. The results build confidence in the model's key identifying assumptions by showing flat pre- and post-trends around patient moves.

area j in year t and define

$$\overline{\delta}_{it} = E\left[\ln\left(\frac{1}{N_{it}}\sum_{d\in\mathscr{D}^*(N_{it},\mathrm{j}(it),t)}\exp\left(\tilde{\delta}_{dt}+\varepsilon_{idt}\right)\right) \mid N_{it}\right] \tag{7}$$

as the typical contribution of physicians to patient i's utilization in time t if she were to select N_{it} physicians at random from her area $\mathrm{j}(it)$. The expectation in $\overline{\delta}_{it}$ is taken both with respect to the random sets of physicians $\mathscr{D}^*(N_{it},\mathrm{j}(i,t),t)$ and the unforecastable contribution of utilization ε_{idt}. Thus, $\overline{\delta}_{it}$ captures the typical utilization due to the availability in the area of physicians with different practice intensities, removing differences in how patients select different physicians from an area. To capture the importance of such physician selection, we further define

$$\sigma_{it} = E\left[\ln\left(\frac{1}{N_{it}}\sum_{d\in\mathscr{D}_{it}}\exp\left(\tilde{\delta}_{dt}+\varepsilon_{idt}\right)\right) \mid N_{it}\right] - \overline{\delta}_{it} \tag{8}$$

as the patient's expected difference in physician-driven utilization given her actual chosen set of physicians $\mathscr{D}_{it}$ and a random set of the same size. We then can rewrite equation (6) as

$$y_{it} = \tilde{\alpha}_{it} + \tilde{\gamma}_j + \ln N_{it} + \overline{\delta}_{it} + \sigma_{it} + \nu_{it}. \tag{9}$$

where ν_{it} is a mean-zero residual.

Finally, aggregating equation (9) across patients and years, we obtain a model of the average annual patient log utilization $\overline{y}_j \equiv E[y_{it} \mid \mathrm{j}(it) = j]$ in HRR j:

$$\overline{y}_j = p_j(\overline{\alpha}_j + \tilde{\gamma}_j + \overline{N}_j + \overline{\delta}_j + \overline{\sigma}_j) \tag{10}$$

where $p_j = Pr(N_{it} > 0 \mid \mathrm{j}(i,t) = j)$ denotes the probability of positive utilization among patient-years in area j, $\overline{\alpha}_j$ is the average patient component ($\tilde{\alpha}_{it}$) among those with positive utilization in area j, $\overline{N}_j$ is the average number of log physician encounters among those with positive utilization in area j, $\overline{\delta}_j$ is the average physician component ($\overline{\delta}_{it}$) for patients in area j, and $\overline{\sigma}_j$ is the average selection component for patients in area j. Following our definition in equation (8), $\overline{\sigma}_j$ captures the extent to which physicians who practice more intensively in HRR j have more encounters than what would be expected under a benchmark of random patient assignment. For example, higher values of $\overline{\sigma}_j$ could reflect high-utilization physicians in HRR j tending to see more patients than low-utilization physicians.

We estimate $\overline{\alpha}_j$ by averaging the estimated $\tilde{\alpha}_{it}$ from equation (9) to the HRR-level; $\tilde{\gamma}_j$ is already constant within HRRs. To estimate $\overline{\delta}_j$ and $\overline{\sigma}_j$, we average simulation-based estimates of average physician utilization $\overline{\delta}_{it}$ in equation (7) and the selection term σ_{it}, respectively. Specifically, we use estimates from the per-encounter utilization model (equation (5)) to form $\tilde{\delta}_{dt}$ as defined in equation (5), and then use the estimates of $\tilde{\delta}_{dt}$ to form simulation-based estimates of average physician utilization $\overline{\delta}_{it}$ in equation (7) and the selection term σ_{it}.[19] Finally, we use estimates of the encounter model (equation (4)) to compute p_j and $\overline{N}_j$.[20] Below, we use these estimates to decompose average HRR utilization into its constituent components.

4.3 Identification

Identification of the per-encounter utilization model from equation (5) leverages the variation from patient and physician moves across HRRs, as well as the within-HRR variation in utilization across patient-physician pairs. Relative patient and physician utilization effects within each HRR are identified by how utilization for a given patient varies across different doctors and how utilization for a given doctor varies across patients (similar to the assumption in Abowd et al. 1999). Quasi-experimental movement of physicians and patients across HRRs identifies the average patient and physician effects in each HRR and thus the practice environment effects.

To build intuition for the roles that movers can play in identification, consider a special case with no time effects or time-varying controls and where a group of patients in each area sees a representative non-moving physician $\mathrm{d}(j)$. Identification of (relative) combined physician-practice-environment effects $\delta_{\mathrm{d}(j)} + \gamma_j$ is then given by a "parallel trends" assumption on the utilization of patient movers: that the unobserved trends in patient health and cost of care ε_{idt} for movers between different origin-destination pairs of HRRs are similar. Formally, the observed utilization trend among patients who move from HRR k to HRR j between time $t-1$ and t can be written

$$\begin{aligned} T_{t,k\to j} \equiv & E[y_{idt} \mid \mathrm{j}(it) = j, \mathrm{j}(i,t-1) = k] - E[y_{id,t-1} \mid \mathrm{j}(it) = j, \mathrm{j}(i,t-1) = k] \\ = & \delta_{\mathrm{d}(j)} + \gamma_j - (\delta_{\mathrm{d}(k)} + \gamma_k) + E[\varepsilon_{idt} - \varepsilon_{id,t-1} \mid \mathrm{j}(it) = j, \mathrm{j}(i,t-1) = k]. \end{aligned} \tag{11}$$

When trends in the unobserved ε_{idt} are comparable across different origin-destination pairs, such that the final term in this expression does not depend on (j,k), the relative aggregate place effect is identified by comparing the utilization trend of movers from HRR k to HRR j to the utilization trend of movers from HRR j to HRR k:

$$\frac{1}{2}(T_{t,k\to j} - T_{t,j\to k}) = \delta_{\mathrm{d}(j)} + \gamma_j - (\delta_{\mathrm{d}(k)} + \gamma_k). \tag{12}$$

[19] More specifically, for each patient and year, we take a random draw of physicians from her HRR with the number of physicians drawn equaling her actual number of encounters for this patient and year. When we randomly draw physicians, we draw them with probabilities equal to the share of their encounters in that HRR-year. We use these simulated encounters averaged over 100 random draws to form estimates of $\overline{\delta}_{it}$ and σ_{it}.

[20] We explore the role of estimation error in our analysis in Appendix C.2.

This common trends assumption would be violated if, for example, patients select into different practice environments in response to an anticipated change in their healthcare needs. For example, estimates of $\overline{\gamma}_j - \overline{\gamma}_k$ could be biased upwards if the patients moving from HRR k to HRR j expect steeper increases in unobserved healthcare needs, relative to movers from HRR j to HRR k.

Just as patient migration can separate the contribution of patient utilization effects from other factors, physician migration can separate the contribution of physicians. To again see this simply, consider a group of physicians who move from HRR k to HRR j between time $t-1$ and t, treating a representative group of non-moving patients in each period. By the same logic as above, so long as these physicians are similar to movers from HRR j to HRR k in terms of their unobserved trends in ε_{idt}, a comparison of average physician utilization before and after different moves identifies the difference in $\alpha_{\mathrm{i}(j)} + \gamma_j$, where $\alpha_{\mathrm{i}(j)}$ denotes the average α_i of non-moving patients in HRR j. With multiple patient (physician) movers seeing distinct groups of physicians (patients) we can thus disentangle distinct average differences to fully separate the variation in α_i, δ_d, and γ_j.

In practice, identification of equation (5) is assisted by the inclusion of time-varying patient- and physician-level controls x_{it} and w_{dt} and by within-HRR variation in the matching of patients to physicians. Including time and patient age effects weakens the key common trends assumptions to allow movers and matching to vary across these dimensions. Similarly, including relative year effects for movers allows for arbitrary differences in utilization before and after a move and thus for the above parallel trends assumption to not compare trends in ε_{idt} across movers and non-movers. The lack of pre-trends in Figure 3 provides some support for these identifying assumptions.

While allowing for persistent unobserved patient, physician, and area heterogeneity, the assumption of conditionally idiosyncratic matching of patients and physicians imposes important restrictions on the data. Most notably, it requires there to be no sudden changes in per-encounter utilization demand that coincide with the systematic matching of patients to particular physicians. This assumption would be violated if, for example, patients systematically respond to a negative health shock such as a cancer diagnosis by seeking out specialists who tend to practice more intensively than a randomly drawn physician. Below, we present within-specialty results that are not affected by any systematic correlation between patient health shocks and sorting to particular specialties. In addition, we show below that our results are robust to controlling for time-varying patient health conditions.

Importantly, identification of the per-encounter utilization model does not require either patient movers or physician movers to be representative of the broader patient and physician populations. When parallel trends holds, all patient, place, and physician effects are identified within the connected set which—as noted above—includes over 99 percent of encounters. This follows by the additivity of effects in the utilization model, which is consistent with the lack of clear post-trends in Figure 3.

Identification of the encounter model (equation (4)) similarly follows from exogenous patient migration. Specifically, as we show in Appendix B.2, conditional maximum likelihood estimates of the practice environment effects γ_j can be obtained from contrasts in log growth rates of encounters across movers with different origin and destination pairs. Figure A5 (Panel B) provides some support for this assumption.

5 Results

5.1 Model Estimates

Tables 2 and 3 report, respectively, the cross-HRR standard deviations and the within-HRR correlation matrix of the key model components. Broadly, these estimates confirm the findings of the motivating event studies that physicians, patients, and practice environments each play an important role in driving geographic differences in healthcare utilization.

For the encounter model, the results indicate that patients and practice environments exhibit similar dispersion across HRRs in their impacts on encounters, and that they are positively correlated within HRRs. Specifically, we estimate that moving a typical patient to an HRR with a one standard deviation higher practice environment component increases her expected number of different physician encounters each year by 0.35 (relative to a mean number of annual encounters of about 6.5), and increasing the average patient component by one standard deviation in a typical practice environment increases expected number of physician encounters by 0.34 (see Table 2). The practice environment and average patient components of the encounter model have a 0.47 positive correlation within HRR (Table 3), indicating that patients who demand more physician encounters tend to be located in HRRs that induce higher numbers of encounters.

Table 2: Standard Deviations of Model Components

	HRR-Average of:	
Encounter Model	Practice environment ($\tilde{\gamma}_j^N$)	0.346
		(0.005)
	Patients ($\tilde{\alpha}_{it}^N$)	0.335
		(0.004)
Per-Encounter Utilization Model	Practice environment ($\tilde{\gamma}_j$)	0.134
		(0.002)
	Patients ($\tilde{\alpha}_{it}$)	0.023
		(0.001)
	Physicians ($\tilde{\delta}_{dt}$)	0.133
		(0.002)

Notes: This table is based on estimation of our Poisson encounter model (equation 4) (rows 1 and 2) and per-encounter utilization model (equation 5) (rows 3-5) in the baseline sample and lists the standard deviations for the individual components in the two models averaged to the HRR level. These individual components are defined in footnote 13 for the encounter model and in equation (5) for the per-encounter utilization model. The estimation sample for the first two rows is the baseline sample of all patient-years (N = 23,663,477 patient-years), and for the next three rows is the baseline sample of all encounters (N = 159 million encounters). Standard errors (in parentheses) are clustered at the patient level and calculated using a Bayesian bootstrap as described in Rubin (1981), with 50 repetitions. Specifically, for each patient in each dataset, we draw 50 weights coming from a Dirichlet distribution. We then repeat our estimation procedure 50 times, weighting each observation by its respective Dirichlet weight. We bootstrap the encounter-level connected set used in the per-encounter utilization regression (equation (5)) and patient-year-level dataset used in the Poisson encounter model regression (equation (4)) separately. We combine the estimates from the bootstrapped-sample regressions to produce the estimates above for each draw. The reported standard errors are the standard deviation of the resulting bootstrap estimates.

For per-encounter utilization, we find that while all three factors exhibit a high degree of geographic variation, the practice environment and physician components have substantially higher geographic variation (with a cross-HRR standard deviation of about 0.13 for each) than the patient component, whose cross-HRR standard deviation is only 0.02 (see bottom panel of Table 2). The relatively low variation in average HRR patient components might at first seem surprising in light of prior findings—which we replicate below—that the patient component explains about half of overall geographic variation in utilization (Finkelstein et al. 2016). However, as we show in our counterfactual results below, most of the patient contribution comes through patient demand for physician encounters rather than through per-encounter utilization quantity.

The model estimates also reveal that physicians display substantial differences in practice intensities across HRRs. For example, we estimate that, all else equal, increasing an HRR's average physician practice intensity by one-standard deviation will increase utilization per encounter by about 13 percent. A patient in an HRR in the 90th percentile of average physician practice intensity will receive 33 percent more utilization per encounter than a similar patient in an HRR in the 10th percentile.

Table 3: Correlation Matrix of Model Components

		Encounter Model		Per-Encounter Utilization Model		
	HRR-Average of:	$\tilde{\gamma}_j^N$	$\tilde{\alpha}_{it}^N$	$\tilde{\gamma}_j$	$\tilde{\alpha}_{it}$	$\tilde{\delta}_{dt}$
Encounter Model	Practice environment ($\tilde{\gamma}_j^N$)	1.000				
	Patients ($\tilde{\alpha}_{it}^N$)	0.470	1.000			
		(0.022)				
Per-Encounter Utilization Model	Practice environment ($\tilde{\gamma}_j$)	-0.512	-0.264	1.000		
		(0.015)	(0.025)			
	Patients ($\tilde{\alpha}_{it}$)	-0.001	-0.125	-0.041	1.000	
		(0.015)	(0.023)	(0.024)		
	Physicians ($\tilde{\delta}_{dt}$)	0.445	0.357	-0.852	0.028	1.000
		(0.018)	(0.029)	(0.005)	(0.031)	

Notes: This table is based on estimation of our Poisson encounter model (equation 4) (rows 1 and 2) and per-encounter utilization model (equation 5) (rows 3-5) in the baseline sample and presents the correlation matrix for the individual components in the two models averaged to the HRR level. These individual components are defined in footnote 13 for the encounter model and in equation (5) for the per-encounter utilization model. The estimation sample for the first two rows is the baseline sample of all patient-years (N = 23,663,477 patient-years), and in the next three rows is the baseline sample of all encounters (N = 159 million encounters). Standard errors (in parentheses) are clustered at the patient level and calculated using a Bayesian bootstrap as described in Rubin (1981), with 50 repetitions. Specifically, for each patient in each dataset, we draw 50 weights coming from a Dirichlet distribution. We then repeat our estimation procedure 50 times, weighting each observation by its respective Dirichlet weight. We bootstrap the encounter-level connected set used in the per-encounter utilization regression (equation (5)) and patient-year-level dataset used in the Poisson encounter model regression (equation (4)) separately. We combine the estimates from the bootstrapped-sample regressions to produce the estimates above for each draw. The reported standard errors are the standard deviation of the resulting bootstrap estimates.

This large variation in physician components is consistent with recent findings documenting substantial variation in practice patterns across physicians within a particular specialty, such as variation in prescribing anti-depressants (Currie and MacLeod 2020) and interpreting chest x-rays (Chan et al. 2022b). We expand on these findings by documenting substantial variation in physician practice intensities both within and across specialties (Table 4, Panel A).[21] Likewise, we

[21] Appendix Table A5 reports our estimates of the average physician effect for each of the 73 specialties in our data; they generally accord with what we would have expected. For example, we estimate that cardiac surgery and thoracic surgery have the two most intensive physician effects, while primary care physicians rank 21st, and podiatrists rank 55th.

Table 4: Physician Practice Intensity Variation

A. Physician Level	
	Standard Deviation
Overall (δ_d)	0.904
	(0.000)
Within-Specialty	0.724
	(0.000)
Between-Specialty	0.540
	(0.000)
B. HRR Level	
	Standard Deviation
Overall ($\bar{\delta}_j$)	0.135
	(0.002)
Within-Specialty ($\bar{\delta}_j - \bar{\delta}_j^S$)	0.104
	(0.002)
Between-Specialty ($\bar{\delta}_j^S$)	0.054
	(0.000)

Notes: The overall standard deviation in Panel A is computed using the physician effects δ_d, as defined in equation (5). For the "within-specialty" standard deviation, we subtract the mean of δ_d in each specialty and compute the standard deviation of the difference; the "between-specialty" standard deviation is the standard deviation of these specialty-average fixed effects. All standard deviations are weighted by the number of encounters each physician has during the entire sample period (1998-2013). In Panel B, we compute the standard deviations of the HRR-level physician components $\bar{\delta}_j$ defined as the HRR average of $\overline{\delta}_{it}$ from equation (7). We report $\bar{\delta}_j$, $\bar{\delta}_j - \bar{\delta}_j^S$, and $\bar{\delta}_j^S$ as measures of the overall, within-specialty, and between-specialty variation, respectively, where $\bar{\delta}_j^S$ is defined in Appendix Section B.3. In Panel A, the sample size is N = 8,292,034 physicians. In Panel B, the sample size is 306 HRRs. Standard errors (in parentheses) are clustered at the patient level and calculated using a Bayesian bootstrap as described in Rubin (1981), with 50 repetitions. Specifically, for each patient in each dataset, we draw 50 weights coming from a Dirichlet distribution. We then repeat our estimation procedure 50 times, weighting each observation by its respective Dirichlet weight. The reported standard errors are the standard deviation of the resulting bootstrap estimates.

find that the cross-HRR variation in the average physician component reflects both differences in the physician component within specialty and differences in the mix of specialties across HRRs (see Table 4, Panel B). [22]

Finally, we note that the correlation matrix across the various components in Table 3 will be important for understanding our counterfactual results on the role of physicians in driving geo-

[22] Note that we report the standard deviations of the HRR-average simulated physician component ($\bar{\delta}_j$) in Table 4 rather than the HRR-average of the actual physician components ($\tilde{\delta}_{dt}$) as in Table 2. This allows us to difference out the HRR-level within-specialty component, $\bar{\delta}_j^S$, as defined in our counterfactual analysis described in Section 5.2 and formalized in Appendix B.3. For the purposes of Table 4, these variables are nearly identical because the simulated physician component is formed by taking the average over random draws of actual physician components for physicians practicing in the same HRR, as discussed in Section 4.2.

graphic variation in utilization. In particular, it is noteworthy that while areas where physicians practice more intensively—i.e. have higher average physician components—are areas with higher patient demand (particularly on the encounter margin) and higher average practice environment components on the encounter margin, they also exhibit a strong negative correlation with the HRR-average practice environment component in the per-encounter utilization model.[23] In other words, HRRs where physicians practice with higher intensity per encounter tend to have lower intensity per encounter practice environments but also a larger number of encounters for a given patient. The large negative correlation between the physician and practice environment components of the per-encounter utilization model suggests that while physician practice intensity varies widely across HRRs, a patient who moves from a low- to a high-intensity HRR need not see a large increase in her utilization—since changes in average physician practice intensity may be offset by changes in other supply-side factors. We now directly investigate this possibility.

5.2 Geographic Variation Counterfactuals

We use estimates of the different components in equation (10) to decompose differences in HRR-average log patient utilization between groups of high and low-utilization HRRs in seven incremental steps. These decompositions augment the analysis in Tables 2 and 3 by incorporating both the standard deviation and correlation terms for the individual components in the per-encounter utilization model, and by combining these estimates with estimates from the encounter model. Each step of this decomposition is defined formally in Appendix B.3.

Table 5 presents the results.[24] Specifically, it shows the decomposition for HRRs above and below the median level of average patient utilization (columns 1-3), the top and bottom quartile (columns 4-6), and the top and bottom decile (columns 7-9). In each set of columns, the first column shows how the absolute difference in average log utilization changes as we restrict different parts of the model, the second column shows the implied percentage change, and the third column shows the cumulative percentage change in utilization. The results are qualitatively similar across the comparisons, so we focus our discussion on the above/below median HRR comparison (columns 1-3), where, as row 1 indicates, the difference in average log patient utilization is 0.253.

[23] Appendix C.2 conducts a sample-splitting exercise that confirms that estimation error plays very little role in explaining this large and negative correlation estimate, given the large number of patient and doctor movers which underpin our identification strategy.

[24] We present standard errors for these estimates in Appendix Table A6.

We begin in row 2 by first setting $\overline{\sigma}_j = 0$ to simulate a counterfactual in which, holding fixed the number of encounters each patient has and the set of patients and physicians in each HRR, patients are randomly matched to physicians in each HRR. This shuts down any utilization differences across HRRs in the tendency of higher- or lower-intensity physicians to disproportionately attract more patients. Our model does not allow us to isolate the relative roles of the physician, patient, and practice environment in driving $\overline{\sigma}_j$. In practice, however, eliminating this systematic patient-physician selection has little effect, reducing the above/below median gap by 6 percent (to 0.238).

Turning to our main research question, we next evaluate the role of physicians in driving geographic variation by equalizing average physician components $\overline{\delta}_j$ across areas. To better understand the role of physicians, we conduct this counterfactual in two steps. In row 3, we first simulate a counterfactual in which we set the average physician component in each HRR equal to the within-specialty average physician component. This allows us to examine the role of within-specialty differences in practice styles, while accounting for the fact that certain specialties (e.g. cardiac surgery) are naturally more intensive than other specialties (e.g. primary care). In row 4, we then further equalize the across-specialty physician component across HRRs, to eliminate variation in utilization due to differential sorting of different types of specialists—with different average practice intensities—to different HRRs. Row 3 shows that eliminating differences across areas in within-specialty practice styles further decreases the above/below median difference by 20 percent of the original difference, to 0.187. Row 4 shows that when we further eliminate differences across areas in the mix of physician specialties, the difference decreases by another 15 percent, to 0.149. Over half of the effect of this specialty mix in turn reflects differences in the share of specialists relative to primary care physicians across HRRs.[25] Thus, differences in physician practice intensities together explain 35 percent of the difference in utilization between high- and low-utilization HRRs. In the remainder of Table 5, we sequentially eliminate variation in patient demand and in practice environment.

We eliminate differences across HRRs in patient demand in two steps (rows 5 and 6). We first equalize $\overline{\alpha}_j$ to simulate a counterfactual in which patient demand for per-encounter utilization is equalized across areas. We then use the encounter model to equalize patient effects on the average

[25] Specifically, Appendix Table A7 shows that equalizing the mix of non-PCP specialists (e.g. cardiologists vs dermatologists vs urologists etc.) across markets eliminates 6 percent of the difference in utilization; the remaining 8 percent is eliminated when we equalize the mix of PCPs relative to all other specialists.

number of physicians seen, affecting both p_j and $\overline{N}_j$. This step simulates a counterfactual in which there is additionally no systematic sorting across HRRs of patients with different demand for physician encounters. Together, these two steps reduce the difference in utilization across high- and low-utilization HRRs by 46 percent of the original difference. Our estimate of the role of patient demand is consistent with Finkelstein et al. (2016), who find that around half of geographic variation in utilization arises from the sorting of patients with different demand for care.

Our analyis further indicates that essentially all of the patient component reflects differences in patient demand for the number of unique physicians seen in a year (row 6); eliminating patient differences in demand for per-encounter utilization has virtually no effect on differences in utilization across HRRs (row 5). This is consistent with the small standard deviation in HRR-average patient components from the per-encounter utilization model, documented in Table 2. The large role for patient-driven differences in the number of encounters is also consistent with the large overall patient demand share in Figure 3. More broadly, our findings on the demand side are consistent with existing work pointing to the potential importance of fragmentation in the delivery of healthcare—defined as a patient receiving care from a large number of distinct providers—in increasing health care spending (Frandsen et al. 2015), and to a large role for what we would call the practice environment in driving fragmentation (Agha et al. 2019). They are also consistent with estimates from the RAND Health Insurance Experiment that increases in patient demand—via increased generosity of insurance coverage—affect the number of visits a patient has but not spending conditional on the visit (Newhouse and the Insurance Experiment Group, 1993).

Finally, we equalize practice environment components in an analogous two-step fashion: first equalizing practice environment effects on per-encounter utilization first (row 7) and then equalizing practice environment effects on the number of physicians seen (row 8). Variation in practice environment accounts for the remaining 13 percent of cross-geographic variation in utilization. Interestingly, eliminating practice environment effects on utilization per encounter actually increases utilization differences by 20 percent (row 7). This stems from the strong negative correlation of practice environment with average physician practice intensities shown in Table 3. However, this is offset by eliminating practice environment effects on the number of encounters, which reduces utilization differences by 32 percent (row 8).

Table 5: Geographic Variation Counterfactuals

	Above/below median			Top/bottom 25%			Top/bottom 10%		
	Absolute Differ- ence	% decline (increment)	% decline (cumulative)	Absolute Differ- ence	% decline (increment)	% decline (cumulative)	Absolute Differ- ence	% decline (increment)	% decline (cumulative)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Observed (1)	0.253			0.405			0.594		
Patient-physician selection (2)	0.238	-6%	-6%	0.378	-7%	-7%	0.564	-5%	-5%
Physicians	0.149	-35%	-41%	0.222	-39%	-45%	0.345	-37%	-42%
Of which: within-specialty effects in utilization per encounter (3)	0.187	-20%	-26%	0.281	-24%	-31%	0.422	-24%	-29%
Of which: between-specialty effects in utilization per encounter (4)	0.149	-15%	-41%	0.222	-14%	-45%	0.345	-13%	-42%
Patients	0.032	-46%	-87%	0.044	-44%	-89%	0.081	-44%	-86%
Of which: patient effects in utilization per encounter (5)	0.148	0%	-42%	0.223	0%	-45%	0.350	1%	-41%
Of which: patient effects in # encounters (6)	0.032	-46%	-87%	0.044	-44%	-89%	0.081	-45%	-86%
Practice Environment	0.000	-13%	-100%	0.000	-11%	-100%	0.000	-14%	-100%
Of which: practice environment effects in utilization per encounter (7)	0.082	20%	-68%	0.126	20%	-69%	0.198	20%	-67%
Of which: practice environment effects in # encounters (8)	0.000	-32%	-100%	0.000	-31%	-100%	0.000	-33%	-100%

Notes: This table is based on estimation of equation (5), equation (4), and the counterfactuals described in Section 5.2. Each set of three columns partitions HRRs into two groups based on percentiles of average log annual patient utilization. First, we report the observed difference in average log annual patient utilization between the two areas at the top of each panel (row 1). Each successive row reports this difference under a particular counterfactual, along with the incremental and cumulative percentage change relative to this baseline. Row (2) reports the counterfactual difference if there were no differential physician selection within regions. Row (3) reports the difference if additionally there were no variation in average physician intensity in healthcare within an encounter across regions, holding fixed the clinical specialty of the physician. Row (4) reports the difference if there were also no differential sorting of clinical specialties across regions. Rows (5) and (6) report the difference if additionally there were no differential sorting of patients' demand for healthcare across regions, breaking this change into two separate sequential steps eliminating patient effects on the demand for care within an encounter and for healthcare encounters, respectively. The last two rows report the difference if additionally there were no variation in practice environment effects on healthcare utilization, breaking this change into two separate sequential steps eliminating practice environment effects on care within an encounter and number of encounters across regions, respectively. For details on how we define each counterfactual, see Appendix Section B.3. The sample is all encounters (159 million encounters of 3 million patients with 1.7 million physicians).

As discussed in Section 2, by assuming that physicians have no effect on the encounter margin, we may understate the role of the physician in contributing to geographic variation in utilization. Namely, one way physicians may affect utilization is through their propensity to refer their patients to other physicians, generating additional encounters. If physician tendencies to make referrals differ across areas, this will load onto the practice environment effect in the encounter margin. We can therefore get an upper bound on the role of the physician by counting the entire practice environment effect on utilization through the encounter margin toward the physician effect. To do this, we use an alternative decomposition method detailed in Appendix B.4, for which results are displayed in Appendix Table A8. In practice, we find that this results in 68% of utilization differences being attributable to physicians, which is virtually identical to the result if we simply add the share of variation due to physicians in our baseline decomposition (35%) to the share driven by practice environment effects on the number of encounters (32%).

Sensitivity Analyses

We conduct a number of checks to these baseline decomposition results and summarize many of them in Appendix Table A9. Column 1 shows our baseline decomposition for above/below median HRRs from Table 5, and the other columns show results from alternative specifications. These all show broadly similar estimates for the role of physician practice intensity in explaining differences in utilization across areas. In column 2 we confirm that our decomposition results are not sensitive to the way we handle physicians who practice in multiple HRRs. In our baseline analysis, we assign separate physician IDs to physicians who practice in multiple HRRs within the same year. As an alternative, in column 2 we assign physicians who practice across multiple HRRs in the same year to a "primary" HRR, defined as the HRR in which she sees the most patients that year; Appendix C.1 provides more details. We find that 99 percent of the observed differences in utilization across HRRs can be explained by decomposing only the utilization associated with a physician's primary HRR. Not surprisingly, therefore, when we focus only on the drivers of differences in primary HRR utilization, we find broadly similar shares of the utilization difference attributed to physicians, patients, and places as in our main analysis.

In column 3, we add controls for time-varying patient health. Specifically, we add to the x_{it} in both the encounter model (equation (4)) and the per-encounter utilization model (equation (5)) a series of indicator variables for whether the patient has each of 21 different chronic conditions.

This is designed to address the potential concern that unmeasured, time-varying patient health shocks are spuriously loading onto our estimates of the physician practice intensity, as patients who become sicker may seek out physicians who practice more intensely. We do not make this our baseline specification because of concerns that the measures of these health conditions may partly reflect the intensity with which health care is practiced within an area (Song et al. 2010; Finkelstein et al. 2016), but we consider it a useful robustness check. The results are reassuring: including these time-varying patient health characteristics has essentially no effect on the share of the geographic variation in health care utilization attributed to physician practice intensity, although it does noticeably increase the share attributed to the patient at the expense of the role of the practice environment. Columns 4 and 5 show that the results are similar—or if anything suggest a somewhat larger role for physician practice intensity—when we estimate the model separately on the first half of the data (1998-2005) and on the second half (2006-2013). The final column shows that, as in Finkelstein et al. 2016, the results are robust to eliminating patient or physician moves to Florida, California, and Arizona, which account for about 24.4 percent of patient destinations in our data. We also confirm that we obtain qualitatively similar results with alternative sequencings of the sequential decomposition (see Appendix C.3).

5.3 Results by Physician Specialty

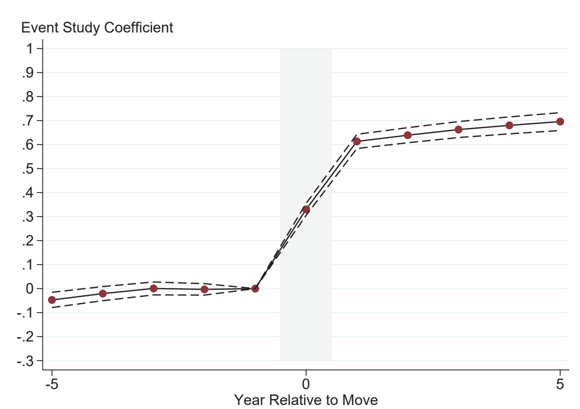
Thus far we have focused on the drivers of differences across areas in overall healthcare utilization. Here, we repeat our main descriptive analyses and decompositions for healthcare utilization for two specific physician specialties: primary care physicians (PCPs) and cardiologists. These are the two largest specialties (see Appendix Table A2) and they have been the focus of prior analyses of the role of physicians in driving geographic variation (Molitor 2018; Cutler et al. 2019).

Our findings suggest substantial heterogeneity across specialties in the role of physicians relative to other factors in driving geographic variation in utilization. An initial indication of this result can be seen in the physician-mover event studies in Panels C and D of Figure 4. They show a much larger—and more immediate—jump upon move for cardiologists than for PCPs, suggesting that cardiologists may play less of a role in driving geographic variation in cardiology utilization than PCPs do in PCP utilization. One possible explanation is that cardiologists' practice styles are more constrained by available physical capital (e.g., facilities for cardiac catheterization) than practice styles of a primary care physician. Another possibility is that patient demand plays a larger role in

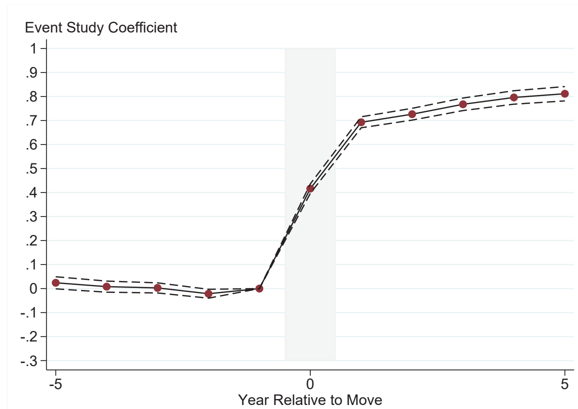
Figure 4: Specialist Patient and Doctor Mover Event Studies

Patients

A. PCPs

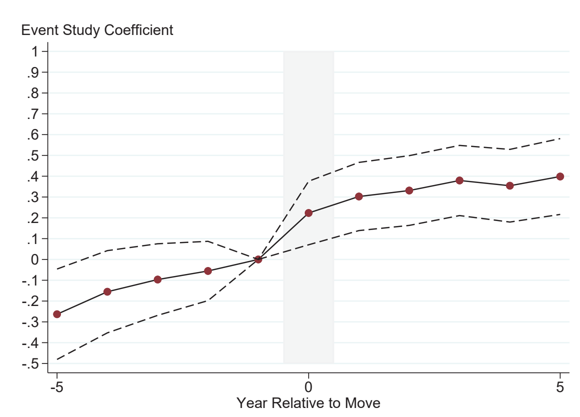


B. Cardiologists

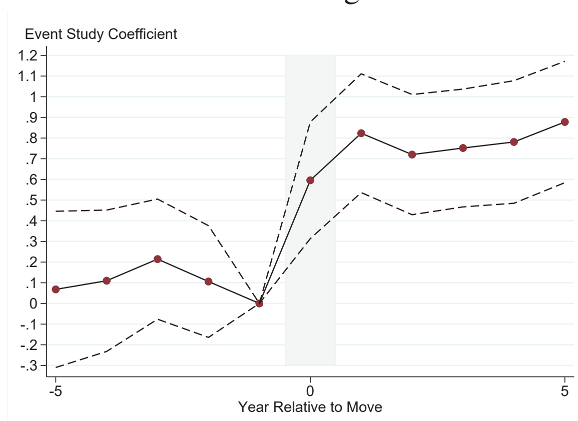


Physicians

C. PCPs



D. Cardiologists



Notes: Panels A and B show the estimated θ_r^P coefficients in equation (2) based on patient movers; the dependent variables are log annual patient utilization from encounters with primary care physicians (Panel A) and cardiologists (Panel B), respectively. Panels C and D show the estimated θ_r^D coefficients in equation (3) based on physician movers; the dependent variables are log annual physician utilization per patient for primary care physicians (Panel C) and cardiologists (Panel D), respectively. The coefficients for relative year -1 are normalized to 0. Observations before and including relative year -6 are binned into a single indicator, as are all observations in relative year 6 and beyond; the coefficients on these indicators are not plotted here. The included covariates are the same as those in the patient or physician event studies in Figure 3. Dashed lines indicate upper and lower bounds of the 95 percent confidence intervals, clustered at the person (i.e. patient or physician) level. The sample is N = 23,636,464 patient-years in Panel A, N = 23,658,474 patient-years in Panel B, N = 5,954,845 physician-years in Panel C, and N = 1,161,572 physician-years in Panel D.

Table 6: Geographic Variation Counterfactuals, by Specialty

	Above/below median					
	Cardiologists		PCPs			
			Original Assignment		Including Specialist Referrals	
	Absolute Differ-ence	% decline (increment)	Absolute Differ-ence	% decline (increment)	Absolute Differ-ence	% decline (increment)
	(1)	(2)	(3)	(4)	(5)	(6)
Observed (1)	0.523		0.384		0.400	
Patient-physician selection (2)	0.514	-2%	0.369	-4%	0.382	-5%
Physicians (3)†	0.499	-3%	0.296	-19%	0.307	-19%
Patients	0.431	-13%	0.099	-51%	0.107	-50%
Of which: patient effects in utilization per encounter (4)	0.505	1%	0.291	-1%	0.298	-2%
Of which: patient effects in # encounters (5)	0.431	-14%	0.099	-50%	0.107	-48%
Practice Environment	0.000	-82%	0.000	-26%	0.000	-27%
Of which: practice environment effects in utilization per encounter (6)	0.406	-5%	0.115	4%	0.107	0%
Of which: practice environment effects in # encounters (7)	0.000	-78%	0.000	-30%	0.000	-27%

Notes: This table is based on estimation of equation (5), equation (4), and the counterfactuals described in Section 5.2. First, we report the observed difference in average log annual patient utilization between HRRs above and below the median (row 1). Each successive row reports this difference under a particular counterfactual, along with the incremental percentage change relative to this baseline. Row (2) reports the counterfactual difference if there were no differential physician selection within regions. Row (3) reports the difference if additionally there were no variation in average physician intensity in healthcare within an encounter across regions. Rows (4) and (5) report the difference if additionally there were no differential sorting of patients' demand for healthcare across regions, breaking this change into two separate sequential steps eliminating patient effects on the demand for care within an encounter and for healthcare encounters respectively. Rows (6) and (7) report the difference if additionally there were no variation in practice environment effects on healthcare utilization, breaking this change into two separate sequential steps eliminating practice environment effects on care within an encounter and number of encounters across regions respectively. For details on how we define each counterfactual, see Appendix Section B.3. The sample is all encounters with cardiologists in columns 1 and 2 under the original utilization algorithm (28 million encounters between 3 million patients and 29 thousand physicians). In columns 3 and 4, the sample is all encounters with PCPs in columns 1 and 2 under the original utilization assignment algorithm (44 million encounters between 3 million patients and 267 thousand physicians). In columns 5 and 6, we repeat the analysis on PCPs with an alternative method of assigning some claims for robustness. The sample size is 44 million encounters between 3 million patients and 278 thousand PCPs. †: the total physician share for PCPs and cardiologists is equivalent to the within-specialty component for the full sample (row 3 of Table 5).

cardiology utilization than in PCP utilization. However, the patient event studies in Panels A and B of Figure 4 suggest that if anything, patient demand plays a smaller role in cardiology utilization than in PCP utilization.

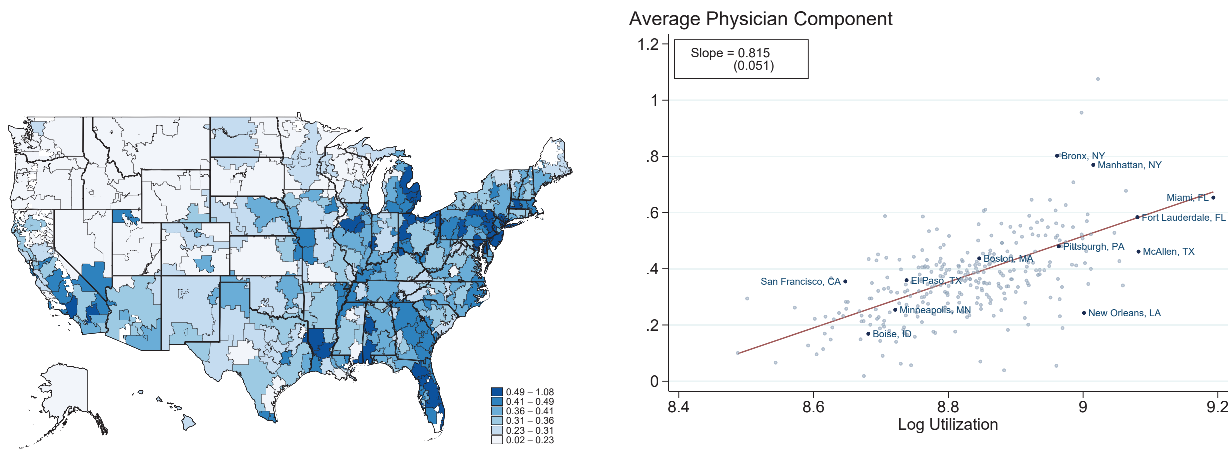
The decomposition results in Table 6 are consistent with what the event studies would lead us to expect. In particular, they indicate that for PCPs the relative roles of different factors in driving geographic variation in utilization are similar to what we find in the full sample of physicians, while for cardiologists the relative roles are quite different. When focusing on cardiologists, the practice environment plays a much larger role (and physician practice style and patient demand play smaller roles) in driving differences in utilization across HRRs than when looking at overall utilization. Specifically, the practice environment accounts for 82 percent of cardiology utilization differences across HRRs, compared to only 13 percent of overall utilization differences across HRRs in Table 5. At the same time, physician practice style accounts for only 3 percent of cardiology utilization differences across HRRs (compared to a within-specialty role of physicians in the full sample of 20 percent), and patient demand accounts for only 13 percent (compared to 46 percent for overall utilization). By contrast, the role of physicians in driving geographic differences in healthcare utilization in the PCP subsample is very similar to the within-specialty role of physicians in thefull sample: 19 percent and 20 percent, respectively. These results are consistent with the relatively small role for physicians found in the Molitor (2018) analysis of cardiologist movers, but also suggest that the role of cardiologists in driving geographic variation is not representative of the role of the larger population of physicians.

The last two columns of Table 6 explore how our findings regarding the role of PCP practice style in driving geographic differences in PCP-related utilization is affected if we instead "credit" utilization generated by PCP referrals to specialists back to the referring PCP. As discussed in Section 3, our baseline approach is to assign utilization for all evaluation and management claims to the performing physician rather than the referring physician. The idea is that an office visit with a cardiologist (the performing physician) may produce different amounts of utilization depending on that cardiologist's tendency to order various procedures and tests. But a downside to our approach is that it excludes a potentially important way that PCPs in particular influence health care spending, namely through their propensity to refer their patients to specialists for further evaluation and treatment. To explore how this affects our analysis of the role of PCP practice style in driving geographic variation in PCP-related utilization, we also show results from an alternative approach in which we attribute back to the PCP any evaluation and management visits from specialists to whom they referred the patient for further evaluation. We find that this does not meaningfully affect the main conclusions of our analysis.

5.4 Correlates of the Physician Component Across HRRs

Given our findings that physicians are the key supply-side factor driving geographic variation in health care utilization, we explore the distribution of the physician component across HRRs, and its correlates with other factors.

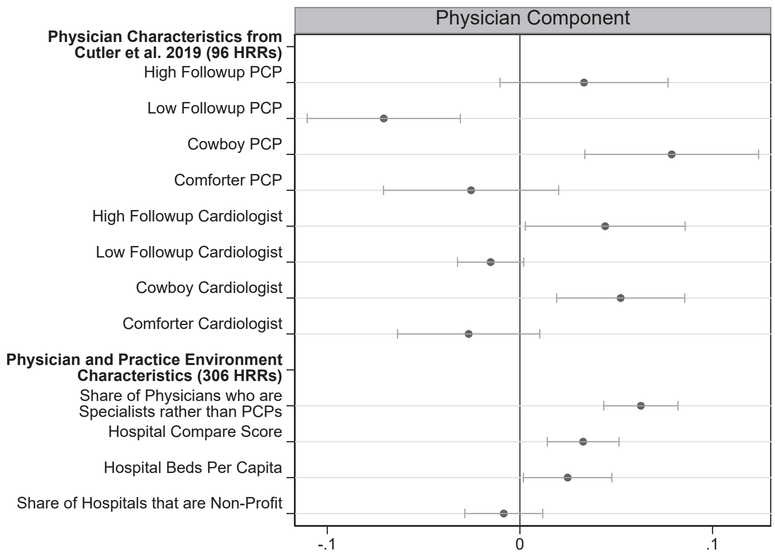
Figure 5: Distribution of Physician Component Across HRRs



Notes: The map shows the distribution of the average physician component in each HRR ($\bar{\delta}_j$ as defined in Section 4.2), in sextiles defined in the legend. The scatterplot plots this against the average annual log patient utilization in each HRR. The regression line is weighted by the number of encounters in each HRR throughout the entire sample period (1998-2013).

Figure 5 shows the distribution of the physician component across HRRs, as well as their correlation with HRR-average annual utilization per capita. Areas with more intensive physicians tend to be in the Southeast and Northeast of the country, while areas with lower-intensity physicians tend to be in the Midwest and Northwest of the country. Areas with higher average physician components tend to have higher average annual utilization, but there is dispersion around this general trend. New Orleans and Manhattan have very similar utilization per patient, for example, but New Orleans has a very low physician component while Manhattan has a very high one. Interestingly, McAllen, TX (made famous for its high spending by Gawande 2009) has a lower physician component than is typical for an area with its spending level, while low-spending Minneapolis and high-spending Miami—which feature heavily in the literature on geographic variation (see e.g. Skinner 2011)—are both right on the regression line.

Figure 6: HRR-Level Correlates With Physician Component



Notes: This figure plots bivariate, HRR-level regression coefficients from a regression of the average physician component in each HRR ($\bar{\delta}_j$ as defined in Section 4.2) against various HRR-level covariates, along with 95% confidence intervals constructed using heteroskedasticity-robust standard errors. All covariates are standardized to have mean 0 and standard deviation 1. The first eight measures are computed on a sample of 96 HRRs for which physicians were surveyed in Cutler et al. (2019); these regressions are weighted by the number of PCPs surveyed for the PCP measures and the number of cardiologists surveyed for the cardiologist measures. For the last five measures, we use the sample of all 306 HRRs, and the regressions are weighted by the number of Medicare patients we observe throughout the entire sample period (1998-2013). Hospital Compare Score approximates hospital quality using timely and effective care measures publicly reported by CMS. Hospital beds per capita counts hospital beds per thousand residents. Non-profit hospitals is the percent of hospitals that are nonprofit. More detail on the construction of these variables can be found in Finkelstein et al. (2016).

Figure 6 shows the correlation of the physician component with various HRR-level characteristics of physicians and of the practice environment. Each row represents the coefficient from a separate bivariate regression. All variables are standardized so that the coefficients report the association between a one standard deviation change in the covariate and the average physician component of the HRR. The top panel examines the relationship between the average physician component in the HRR and measures of physician beliefs about appropriate practice style. To capture physician beliefs, we draw on survey-based measures from Cutler et al. (2019).[26] They present

[26] We are grateful to the authors for sharing these data.

a sample of physicians with patient vignettes and ask them to rate the likelihood they would recommend different courses of action. We use the shares of primary care physicians and the shares of cardiologists in each HRR who recommend levels of follow-up care greater ("high follow-up") or less ("low follow-up") than clinical guidelines suggest, as well as the respective shares who recommend aggressive ("cowboy") or less aggressive ("comforter") end-of-life care. The results are shown for the 96 HRRs for which these measures are available.

The correlations are all of the expected sign: areas with a higher share of doctors who are "high follow up" or "cowboys" have a higher average physician component, while areas with a higher share of doctors who are "low follow up" or "comforters" have a lower average physician component. For example, the results indicate that an HRR with a share of "cowboy" PCPs that is one standard deviation higher would have, on average, about 7.5% higher utilization per encounter in that HRR. About half of these relationships are statistically distinguishable from zero. By contrast, the relationship between these measures of physician beliefs and either the average patient component or the average practice environment component is much smaller in magnitude and almost always statistically insignificant (Appendix Figure A6).

The bottom panel of Figure 6 examines other area measures that are available for all 306 HRRs. We find that areas with a higher share of physicians who are specialists (rather than PCPs) have a higher average physician component. We also see that places with a higher average physician component tend to have higher quality hospitals and more hospital beds per capita; there is no relationship, however, between the average physician component and the share of hospitals in the area that are non-profit.

6 Conclusion

Physicians play a unique role in healthcare markets, guiding most key diagnosis and treatment decisions that affect patients. Yet evidence remains limited on the extent to which physician practice differences are a quantitatively important factor behind the substantial geographic variation in US healthcare utilization. We fill this gap by leveraging migration of Medicare patients and physicians to estimate a model of encounters and of per-encounter utilization, allowing for variation in patient demand, physician practice intensity, and other supply side differences across areas.

Our findings indicate that physicians are a key driver of geographic differences in utilization.

We estimate that differences in physician practice intensities across areas explain at least a third of geographic variation in utilization—roughly three times the importance of other supply side factors. The role of the physician reflects both differences across areas in within-specialty practice style (about three-fifths of the physician component) as well as differences across areas in the physician specialty mix (about two-fifths of the physician component).

Taken together, these findings suggest that policies that change physicians' preferences or skills could substantially reduce healthcare utilization in high-utilization parts of the country.

Of course, our counterfactual analyses of changing the distribution of patient, physician, or practice environment characteristics across areas do not account for potential general equilibrium effects of such changes; these are outside of our model and the scope of our analysis. Likewise, policies that may affect physician practice intensity could also have general equilibrium consequences which we have not considered. Nonetheless, our results suggest the importance of further work to understand the determinants of physician differences in practice intensity and the effects of potential policies that might affect that practice intensity. Here, the role of physician training—both training within specialty and the availability of training for different specialties—seems particularly important to examine, as differences in training programs may contribute not only to differences across physicians in their practice intensity but also to differences across areas in average physician practice intensity, as physicians tend to practice in the same area where they trained (Association of American Medical Colleges 2019).

References

Abowd, J. M., Creecy, R. H., and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, U.S. Census Bureau Center for Economic Studies.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Abowd, J. M., McKinney, K. L., and Schmutte, I. M. (2019). Modeling endogenous mobility in earnings determination. *Journal of Business and Economic Statistics, 37(3): 405-418.*

Agha, L., Frandsen, B., and Rebitzer, J. B. (2019). Fragmented division of labor and healthcare costs: Evidence from moves across regions. *Journal of Public Economics*, 169:144–159.

Ahammer, A. and Schober, T. (2020). Exploring variations in health-care expenditures: What is the role of practice styles? *Health Economics*, 29(6):683–699.

Allcott, H., Diamond, R., Dubé, J.-P., Handbury, J., Rahkovsky, I., and Schnell, M. (2019). Food deserts and the causes of nutritional inequality. *The Quarterly Journal of Economics*, 134(4):1793–1844.

Anderson, G. F., Hussey, P., and Petrosyan, V. (2019). It's still the prices, stupid: why the us spends so much on health care, and a tribute to Uwe Reinhardt. *Health Affairs*, 38(1):87–95.

Association of American Medical Colleges (2019). Report on residents: Executive summary. https://www.aamc.org/data-reports/students-residents/data/report-residents/2019/executive-summary.

Austin, A. M., Gottlieb, D. J., Charmichael, D., Chakraborti, G., Sutherland, J. M., Andrews, K. G., Raymond, S. R., Tomlin, S., Zhou, W., Song, Y., and Skinner, J. (2020). A standardized method for adjusting medicare expenditures for regional differences in prices. *Dartmouth Atlas Project Technical Report*.

Baicker, K. and Chandra, A. (2004). The productivity of physician specialization: Evidence from the medicare program. *American Economic Review*, 94(2):357–361.

Bloom, N., Propper, C., Seiler, S., and Reenan, J. V. (2015). The impact of competition on management quality: Evidence from public hospitals. *The Review of Economic Studies*, 82(2):457–489.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739.

Bronnenberg, B. J., Dube, J.-P. H., and Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, 102(6):2472–2508.

Callaway, B. and Sant'Anna, P. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Cantoni, E. and Pons, V. (2022). Does context outweigh individual characteristics in driving voting behavior? Evidence from relocations within the United States. *American Economic Review*, 112(4):1226–72.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*.

Card, D., Heining, J., and Kline, P. M. (2013). Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics*, 128(3):967–1015.

Card, D., Rothstein, J., and Yi, M. (2023). Location, location, location. *NBER Working Paper No. 31587*.

Chan, D. C., Card, D., and Taylor, L. (2022a). Is there a VA advantage? Evidence from dually eligible veterans. Technical report.

Chan, D. C., Gentzkow, M., and Yu, C. (2022b). Selection with variation in diagnostic skill: Evidence from radiologists. *Quarterly Journal of Economics*, 132(2):729–783.

Chandra, A. and Staiger, D. O. (2007). Productivity spillovers in health care: Evidence from the treatment of heart attacks. *Journal of Political Economy*, 115(1):103–140.

Chernew, M. E., Sabik, L., Chandra, A., and Newhouse, J. P. (2009). Would having more primary care doctors cut health spending growth? *Health Affairs*, 28(5):1327–1335.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chetty, R., Friedman, J. N., and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *American Economic Review*, 103(7):2683–2721.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *Quarterly Journal of Economics*.

Choi, N. G. (1996). Older persons who move: Reasons and health consequences. *Journal of Applied Gerontology*, 15(3):325–344.

Congressional Budget Office (2008). Geographic variation in health care spending. Washington: Government Printing Press.

Currie, J., MacLeod, W. B., and Van Parys, J. (2016). Provider practice style and patient health outcomes: The case of heart attacks. *Journal of Health Economics*, 47:64–80.

Currie, J. and Zhang, J. (forthcoming). Doing more with less: Predicting primary care provider effectiveness. *Review of Economics and Statistics*.

Currie, J. M. and MacLeod, W. B. (2020). Understanding doctor decision making: The case of depression treatment. *Econometrica*, 88(3):847–878.

Cutler, D., Skinner, J., Stern, A. D., and Wennberg, D. (2019). Physician beliefs and patient preferences: A new look at regional variation in spending. *American Economic Journal: Economic Policy*, 11(1):191–221.

DesRoches, C. M., Barrett, K. A., Harvey, B. E., Kogan, R., Reschovsky, J. D., Landon, B. E., Casalino, L. P., Shortell, S. M., and Rich, E. C. (2015). The results are only as good as the sample: Assessing three national physician sampling frames. *Journal of General Internal Medicine*, 30(3):595–601.

Doyle, J. J. and Staiger, B. (2022). Physician group influences on treatment intensity and health: Evidence from physician switchers. *NBER Working Paper No. 29613*.

Duggan, M., Gupta, A., Jackson, E., and Templeton, Z. (2022). The impact of privatization: Evidence from the hospital sector. *NBER Working Paper No. 30824*.

Eliason, P. J., Heebsh, B., McDevitt, R. C., and Roberts, J. W. (2020). How acquisitions affect firm behavior and performance: Evidence from the dialysis industry. *The Quarterly Journal of Economics*, 135(1):221–267.

Epstein, A. and Nicholson, S. (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*, 28:1126–1140.

Fadlon, I. and Van Parys, J. (2020). Primary care physician practice styles and patient care: Evidence from physician exits in medicare. *Journal of Health Economics*, 71.

Finkelstein, A., Gentzkow, M., and Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *Quarterly Journal of Economics*, 131(4):1681–1726.

Frakes, M. D., Gruber, J., and Justicz, T. (forthcoming). Public and private options in practice: The military health system. *American Economic Journal: Economic Policy*.

Frandsen, B., Joynt, K., Rebitzer, J., and Jha, A. (2015). Care fragmentation, quality, and costs among chronically ill patients. *American Journal of Managed Care*, 21(5):355–362.

Gawande, A. (2009). The cost conundrum: What a Texas town can teach us about health care. *The New Yorker*.

Gottlieb, D. J., Zhou, W., Song, Y., Andrews, K. G., Skinner, J. S., and Sutherland, J. M. (2010). Prices don't drive regional medicare spending variations. *Health Affairs*, 29(3):537–543.

Gottlieb, J. D., Polyakova, M., Rinz, K., Shiplett, H., and Udalova, V. (2023). Who values human capitalists' human capital? The earnings and labor supply of U.S. physicians. *NBER Working Paper No. 31469*.

Guimaraes, P. (2014). Poi2hdfe: Stata module to estimate a poisson regression with two high-dimensional fixed effects. Technical report, Statistical Software Components S457777, Boston College Department of Economics.

Hagedorn, M., Law, T. H., and Manovskii, I. (2017). Identifying equilibrium models of labor market sorting. *Econometrica 85(1): 29-65*.

Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52(4):909–938.

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

Ho, K. and Lee, R. S. (2019). Equilibrium provider networks: Bargaining and exclusion in health care markets. *American Economic Review*, 109(2):473–522.

Kletke, P. R., Polsky, D., Wozniak, G. D., and Escarce, J. J. (2000). The effect of HMO penetration on physician retirement. *Health Services Research 35(5):17-31*.

Kramarz, F., Machin, S., and Ouazad, A. (2015). Using compulsory mobility to identify school quality and peer effects. *Oxford Bulletin of Economics and Statistics*, 77(4):566–587.

Kwok, J. H. (2019). How do primary care physicians influence healthcare? Evidence on practice styles and switching costs from Medicare. *Working Paper*.

Lee, T. H. and Mongan, J. J. (2009). *Chaos and Organization in Health Care*. MIT Press, Cambridge, MA.

Medicare Payment Advisory Comission (2021). A data book: Healthcare spending and the Medicare program. Medicare Payment Advisory Comission, Washington, DC.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy*, 10(1):326–356.

Newhouse, J. P. and Garber, A. M. (2013). Geographic variation in Medicare services. *New England Journal of Medicine*, 368(16):1465–1468.

Newhouse, J. P. and the Insurance Experiment Group (1993). *Free for all? Lessons from the RAND health insurance experiment*. Harvard University Press.

Orszag, P. R. (2011). How health care can save or sink America: The case for reform and fiscal sustainability. *Foreign Affairs*.

Otero, C. and Munoz, P. (2022). Managers and public hospital performance. *Working Paper*.

Reinhardt, U. (2019). *Priced Out: The Economic and Ethical Costs of American Health Care*. Princeton University Press, Princeton, New Jersey.

Rubin, D. B. (1981). The Bayesian bootstrap. *Annals of Statistics 9(1): 130-134*.

Silver, D. (2021). Haste or waste? Peer pressure and productivity in the emergency department. *Review of Economic Studies*, 88(3):1385–1417.

Skinner, J. (2011). Causes and consequences of regional variations in health care. In Mark V. Pauly, T. G. M. and Barros, P. P., editors, *Handbook of Health Economics*, volume 2, pages 45–93. Elsevier.

Song, Y., Skinner, J., Bynam, J., Sutherland, J., and Fisher, E. (2010). Regional variations in diagnostic practices. *New England Journal of Medicine*, 363(1):45–53.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

U.S. Department of Health and Human Services (HHS) (2014). 2014 CMS statistics 80146-23597.

Van Parys, J. (2016). Variation in physician practice styles within and across emergency departments. *PloS One*, 11(8).

White, D. P., Enewold, L., Geiger, A. M., Banks, R., and Warren, J. L. (2020). Comparison of physician data in two data files available for cancer health services research. *Journal of the National Cancer Institute Monographs*, 55:66–71.

Online Appendices

A Baseline Analysis Sample

A.1 Patients

This appendix details the restrictions which lead to our baseline sample of encounters. As summarized in Section 3.2, we exclude a random 75 percent of patient non-movers for ease of computation; we weight all analyses by this sampling probability. We also exclude patient-years in which the patient is on Medicare Advantage, younger than 65 or older than 99, or not subscribed to Medicare Part A and Part B for all months of a year. We further exclude patients who moved more than once and whose share of claims in the destination HRR did not exceed that in the origin HRR by at least 0.75 after the move. We exclude patients whose HRR of residence changes multiple times. Lastly, we restrict the sample of encounters to the largest connected set of patients and physicians. Appendix Table A10 shows the effects of these sequential restrictions; after excluding 75 percent of patient non-movers, the next most impactful restriction is the exclusion of patients enrolled in Medicare Advantage, which excludes about 4 percent of the sample. The connected set restriction is the least impactful restriction, reducing the sample size by 1 percent relative to the original raw claims data.

Patients can exit our sample for three primary reasons: death, switching to Medicare Advantage, and exiting the 65-99 age window. About a third of patients die in our sample window; mortality is similar for movers and non-movers. About a fifth exit by switching to Medicare Advantage. We observe the average non-mover for 7.2 years and the average mover for 9.2 years; part of this difference is mechanical, since the mover label is contingent on observing a patient for at least 2 years. Correspondingly, we observe the average non-mover physician for 6.6 years and the average mover physician for 9.9 years.

A.2 Physicians

Assigning Each Physician a Unique ID

We assign each physician a unique ID through a combination of their National Provider Number (NPI) and Unique Provider Identification Number (UPIN). Until 2006, physicians were identified in claim forms by their UPIN only. However, starting in 2006, CMS transitioned from the UPIN to the NPI. Thus, in order to assign each physician a unique ID throughout our 1998-2013 time period, we must match UPINs to NPIs.

To do so we rely on two main sources of information: a crosswalk produced by the National Bureau of Economic Research (NBER) that matches UPINs to NPIs[27] and has been used by other studies (e.g. Molitor 2018; Kwok 2019), and our own construction of a claims-based crosswalk matching UPINs to NPIs. This section details how this matching is performed.

The NBER crosswalk is based on files from the NPI Registry, which is maintained by CMS. We exclude from the NBER crosswalk matches with group UPINs and organization NPIs, in order to ensure that matches only occur among individuals.We also exclude cases in which (1) a UPIN is mapped to two or more NPIs, which may occur as a product of typos in the NPI application, and (2) an NPI is mapped to two or more UPINs, which may occur if a physician lists another identifier as an additional UPIN.

The NBER crosswalk does not produce a comprehensive set of UPIN-NPI matches, because matches are based on the voluntary provision of UPINs in the NPI application. Similar to Molitor (2018), we therefore supplement the NBER crosswalk with a separate crosswalk that we build using the claims data, which contains both UPIN and NPI information.

To build this claims-based crosswalk, we make use of the fact that, during the UPIN-NPI transition period (2006-2008), Medicare encouraged physicians to include both identifiers in claims, so that providers would gain experience using the new number. As a result, as can be seen in Figure A7, in 2007 over 50 percent of utilization was associated with claims that included both a UPIN and an NPI. However, after the transition to NPIs was complete in 2008, UPINs were no longer inspected for accuracy, and the inclusion of an incorrect UPIN no longer led to a claim denial.[28] Not surprisingly, therefore, there is ample evidence in the data that UPINs were often inputted incorrectly—UPINs were frequently listed with typos (e.g. E00000 entered as F00000), and NPIs were listed in conjunction with another physician's UPIN from the same location. Furthermore, Kwok (2019) notes that physician group practices "may have systematically entered the incorrect

[27] See https://www.nber.org/research/data/national-provider-identifier-npi-unique-physician-identification-number-upin-crosswalk.

[28] In fact, Medicare's expectations were that clearinghouses would not even pass the legacy identifier to Medicare after the NPI was fully implemented in 2008. See https://www.cms.gov/newsroom/fact-sheets/national-provider-identifier-npi-may-23-2008-implementation and https://www.hhs.gov/guidance/sites/default/files/hhs-guidance-documents/JA4320.pdf.

physician UPIN when it was not tied to reimbursement (e.g., selecting the physician at the top of an alphabetical list rather than the actual performing physician)".

With these data shortcomings in mind, we produce a claims-based crosswalk that matches UPINs to NPIs based on a majority of claims, instead of requiring a one-to-one match (as in Molitor 2018). This allows a larger set of IDs to be matched while still producing accurate matches. The steps below summarize the algorithm:

1. Let A be the set of all UPIN-NPI combinations observed in the claims data. Consistently with our process of assigning physicians to claims (see Appendix A.2), we utilize referring and performing physicians for carrier claims, and attending physicians for outpatient and inpatient claims. Thus, the set A represents the collection of potential UPIN-NPI matches.

2. Exclude from set A combinations that contain (i) a missing identifier (UPIN or NPI); (ii) a group UPIN;[29] (iii) a surrogate UPIN;[30] or (iv) an organization NPI.[31] This is because we are only interested in UPIN-NPI combinations that are associated with individual physicians.

3. Define C_U as the number of claims in which UPIN U is observed within set A, and define C_N analogously for NPIs. Finally, define C_{UN} as the number of claims in which we observe UPIN U combined with NPI N.

4. Let $s_{UN} = \frac{C_{UN}}{C_U}$ be the share of UPIN U's claims that are associated with NPI N, and let $s_{NU} = \frac{C_{UN}}{C_N}$ be the share of NPI N's claims that are associated with UPIN U.

5. If both s_{UN} and s_{NU} are larger than 0.75, UPIN U and NPI N are matched. These matches will constitute the claims-based crosswalk.

To produce a finalized crosswalk that draws on both the NBER crosswalk and the claims-based crosswalk, we retain UPIN-NPI matches that agree between crosswalks, as well as matches that only appear in either crosswalk. Some matches may be associated with conflicts, meaning that either an NPI is matched to a UPIN in the claims-based crosswalk and to another UPIN in the

[29] Group UPINs are identified by their first characters (W-Z). See https://www.nber.org/research/data/national-provider-identifier-npi-unique-physician-identification-number-upin-crosswalk.

[30] Surrogate UPINs were used when a physician did not have a UPIN, and were used instead of leaving the UPIN field blank. The most common ones are "OTH000" and "RES000".

[31] These are identified from NPPES data, assembled by the NBER. See https://data.nber.org/data/nppes/.

NBER crosswalk, or a UPIN is matched to an NPI in the claims-based crosswalk and to a different NPI in the NBER crosswalk; we do not link UPINs and NPIs that are associated with conflicts.

Table A11 displays UPIN-NPI matching results. Panel A displays statistics for NPIs that are part of the finalized crosswalk. Among these NPIs, most matches agree between crosswalks and, since the claims-based crosswalk is more complete, a large share of matches are only found in the claims-based crosswalk. Overall, we can produce 921,000 UPIN-NPI matches, accounting for 87 percent of total 2009 utilization. Panel B, in turn, displays statistics for NPIs that are not matched to a UPIN. We see that only a small number of NPIs are associated with conflicts,[32] while a large number of NPIs are not matched because they (and presumably the physician associated with that NPI) enter the sample after the UPIN-NPI transition.[33,34] Altogether, non-matched NPIs correspond to only 13 percent of total 2009 utilization.

To evaluate the accuracy of the 920,000 matches in the finalized crosswalk, we validate them against physician identifier information in the AMA Physician Masterfile, which contains both NPIs and UPINs for the physicians in the data. Like the NBER crosswalk, the AMA data are not comprehensive. Still, it allows us to evaluate accuracy where there is overlap. Table A12 displays such a comparison for the NPIs that were either matched to the finalized crosswalk or associated with a non-missing UPIN in the AMA masterfile. We see that for cases accounting for 80 percent of total utilization in 2009, both crosswalks produce the same matching, while conflicts are rare (accounting for less than 0.5 percent of utilization).

Using the finalized crosswalk, we assign physician IDs to each claim based on the year and information provided (similarly to Kwok 2019). Between 1998 and 2005, we assign physician IDs based on the listed UPIN. Between 2009 and 2013, we assign physicians based on the listed NPI. Finally, during the transition period in 2006-2008, we also assign physician IDs based on the listed NPI. We do so because CMS guidance indicates that claims are denied if the listed NPI cannot be located or does not meet certain criteria, indicating that the NPI functions as a primary identifier.[35]

[32] A large share of these conflicts are due to typos, likely in the NPI registry.

[33] We consider NPIs to be unmatched if they are listed with at least one non-missing UPIN in the 20 percent claims data, and we consider NPIs to have entered the sample after the transition ("post transition NPIs") if they are not listed with any UPINs.

[34] The large number of post-transition NPIs can be explained by an expansion in the number of Medicare providers (1.05 million providers in 2006 to 1.23 million providers in 2013; see 2006 CMS statistics and 2013 CMS statistics) and by the gradual replacement of the workforce between 2007 and 2013.

[35] See https://www.hhs.gov/guidance/sites/default/files/hhs-guidance-documents/JA4320.pdf

If the NPI is missing during the transition period, we assign IDs based on the UPIN. Table A13 below shows an illustration of the UPIN-NPI crosswalk and some ID assignment examples.

Assigning Physicians to Patient Claims

We would like to match each claim to the physician who is most likely the responsible "decider" for that care. In practice, this requires some judgment calls, since Medicare claims may list multiple physicians that were involved in the patient care that generated the claim.

More specifically, we use three types of Medicare claim files: carrier, inpatient, and outpatient. Carrier claims make up 34 percent of utilization, while inpatient and outpatient claims are 48 and 18 percent of utilization, respectively. Inpatient claims are those billed by hospitals under Part A of Medicare for inpatient stays. Outpatient claims are filed by institutions that also bill for Part A services, notably hospitals. Carrier claims are filed by physicians and by institutions that do not provide inpatient services (such as clinical laboratories, or certain outpatient clinics). Most of the medical care one naturally associates with a physician (e.g. an office visit, a surgical procedure—whether done inpatient or outpatient—the ordering of a lab test, etc.) is contained on the carrier file.[36]

Although the distinction between carrier and outpatient claims is not always clear, as a general rule physician charges are billed under carrier claims (whether they occur in an inpatient or outpatient setting), while other (non-inpatient) services may be billed either in the carrier or outpatient files. Non-physician charges (also called facility charges) are billed either to the outpatient or carrier file; as a rough rule of thumb, hospitals and other inpatient facilities will bill outpatient facility charges to the outpatient file, while physicians and free-standing clinics will bill these same outpatient facility charges to the carrier file.[37] As a concrete example, consider a patient visiting an Emergency Room for a broken leg. She is treated by a physician and has an x-ray. The ER physician will file a carrier claim for the physician care, the radiologist will file a carrier claim for the x-ray, and the hospital will file an outpatient claim for the x-ray machinery and facilities.

In the inpatient and outpatient files, we observe up to three physicians for each claim—attending,

[36]As in prior work (e.g. Finkelstein et al. 2016) our measure of health care utilization excludes several dimensions of care, including durable medical equipment, home health agency care, hospice care, skilled nursing facility care, inpatient rehabilitation facility care, and claims filed through Medicare Part D (prescription drug coverage); there is also substantial variation in these additional measures of care (Newhouse and Garber 2013).

[37]See https://resdac.org/videos/using-carrier-and-outpatient-files and https://healthcaredelivery.cancer.gov/seermedicare/medicare/claims.html.

operating, and other. According to the Medicare Claims Processing Manual,[38] the attending physician "is the individual who has overall responsibility for the patient's medical care and treatment reported in this claim/encounter," and the operating physician—which is only filled in if there is a surgical procedure listed on the claim—is the "individual with the primary responsibility for performing the surgical procedure(s)." Practically speaking, therefore, although over 99.9 percent of claims have an attending physician, about 80 percent of outpatient claims and 40 percent of inpatient claims are missing an operating physician (the "other" physician is also frequently missing). Thus, we assign the attending physician to each claim.

In the carrier files, we observe both a performing physician and a referring physician. The performing physician is rarely missing, but the referring physician is missing on about 15 percent of claims. The carrier files include claims for many different types of services, and whether the performing or referring physician is the one who "decided" on the treatment likely varies across these different types of services. For example, for an office visit with a primary care physician (PCP) or a specialist, we suspect that the performing physician has considerable discretion over treatment decisions. However, for other services such as an MRI or an electrocardiogram, the decision to get this service was likely made by the referring physician, while the physician who actually "performs" is merely implementing, with little if any discretion. Naturally, other cases may fall somewhere in between these extremes.

To get a better sense of this, we categorize carrier file services using Berenson-Eggers Type of Service (BETOS) codes, which divide claim lines into seven groups: Evaluation and Management (E&M),[39] Procedures, Imaging, Tests, Durable Medical Equipment, Other, and Exceptions/Unclassified.[40] Claims for ambulance services also appear in the carrier files, but we exclude these from our data and analysis. Table A1 displays carrier claims summary statistics by BETOS code groups. Panel A indicates that about 40 percent of carrier utilization is related to E&M services and about 30 percent is related to procedures. 20 percent of utilization is related to testing or imaging, and the remaining 10 percent is related to other services.

[38]See https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c25.pdf; downloaded on 03/08/2022.

[39]E&M BETOS code services include office visits, hospital visits, ER visits, nursing home visits and specialist consultations.

[40]For a full list of BETOS codes, see https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/downloads/BETOSDescCodes.pdf.

Table A1 Panel B displays, by BETOS group, the share of claims in which (i) the performing and referring physicians match, (ii) the referring physician is different from the performing (but not missing), (iii) only the referring physician is missing, (iv) only the performing physician is missing and (v) both referring and performing physician are missing. For our baseline analysis, we assign the performing physician for E&M codes, and use the referring physician for all other codes, unless it is missing, in which case we assign the performing physician (who is virtually never missing). As seen in Table A1, the referring physician is almost never missing for testing and imaging claims, but is missing for about 15 percent of procedure claims. If both performing and referring physicians are missing, we assign a missing physician ID unique by HRR; these missing physician IDs account for only 0.29 percent of utilization.

This approach is relatively clear cut for testing and imaging claims and for Evaluation and Management (E&M claims) which together make up about three-fifths of utilization. For testing and imaging, the referring physician is missing in less than 0.5 percent of claims, and the referring physician differs from the performing physician in 70 to 80 percent of claims. Moreover, over half of testing claims are "performed" by physicians with a clinical laboratory specialty code (not shown). This suggests that these services are performed by providers with little discretion, at the request of the referring physician. By comparison, for E&M claims, in about 20 percent of claims the referring physician is missing, and in another 40 percent the referring physician and the performing physician are the same; this suggests that performing physicians for E&M claims often have significant discretion in treatment.

There is more ambiguity regarding the appropriate physician assignment when the BETOS group is for procedures (which are about 30 percent of utilization) or other (about 10 percent). To see this, consider two procedures we observe in the data: anesthesia and pacemaker insertion. For anesthesia claims, it is reasonable to assume that the referring physician (who is typically the surgeon performing the surgery) is responsible for treatment decisions, instead of the anesthetist that administers the anesthesia. However, for pacemaker insertions, we see that PCPs often refer patients to cardiologists. In these cases, there is more ambiguity in which physician is most responsible for treatment decision—the PCP who referred for pacemaker or the cardiologist who performs the insertion. In practice, we would attribute the pacemaker utilization to the PCP.

Note that, in this baseline definition, if a PCP refers a patient for testing, the tests will be assigned to the PCP. But if a PCP refers a specialist for an E&M visit, that specialist care will be assigned to the specialist. This makes sense as the specialist likely has considerable discretion at this point. However, one way that a PCP can be a "high utilization" PCP is by having a high propensity to refer to a specialist, or to refer to high intensity specialists, and this will not be attributed to the PCP in our baseline approach.

Identifying Physician Movers and Physician Location

In each year, we begin by defining a physician's potential location(s) as any HRR in which a patient with whom they have an encounter resides (see the previous subsection for how we assign physicians to patient claims). Note that because physicians may treat patients residing in different HRRs in a given year, they may have multiple potential locations in a given year. We define a physician to be a mover if they exhibit exactly one clear shift in location during our sample. Our objective is to define movers as physicians who can be reliably traced to an origin HRR and subsequently traced to a destination HRR, with a clear shift in utilization between these two locations at some point in time. Furthermore, our mover algorithm is designed to err on the side of false negatives, to increase our confidence that any physician whom we label a mover is very likely to be one.

Physician Movers We begin by restricting our attention to physicians who are mainly located in exactly two HRRs over time. For each physician-year, we define their "focal HRR" as the HRR that accounts for over 75 percent of the physician's encounters that year. If no HRR accounts for at least 75 percent of encounters, no focal HRR is assigned to that physician-year. Our initial step is to restrict the set of potential movers to physicians with exactly two focal HRRs between 1998 and 2013.

The next step is to ensure that each potential mover is initially located in the origin HRR and then moved to the destination HRR, without moving back to the origin. To do so, let a "focal HRR event" be the range of years in which an HRR is assigned as the focal HRR for a given physician. We restrict to physicians whose two focal HRR events do not overlap—that is, the last year of one focal HRR event must come before the first year of the other focal HRR event. The HRRs associated with the first and second focal events (in calendar time) are considered the origin and destination HRRs, respectively.

Given the origin and destination focal HRR events, we can determine the move year. Let t_{pre} be the last year in which the origin HRR is the focal HRR, and let t_{post} be the first year in which the destination HRR is the focal HRR. In the set of years between (and including) t_{pre} and t_{post}, our objective is to set the move year as the year in which the physician provided care most evenly across origin and destination. Hence we define the move year as the year in which the number of encounters in the destination, as a share of the number of encounters in both the origin and destination, is closest to 50 percent. Ties are broken by assigning the move year to the latest year,[41] and the move year is assigned relative year 0.

As a last step, we try to ensure that movers are clearly settled in the origin and destination HRRs before and after the move, respectively. To do so, we restrict to physicians with more than 50 percent of encounters associated with the origin in each of the final four pre-move years, or starting from when the physician is first observed if the physician enters the sample three or fewer years before the move year. Similarly, we restrict to physicians with more than 50 percent of encounters associated with the destination in each of the first four post-move years, or until the physician is last observed.[42] We also restrict to physicians that, in the move year, had at least 75 percent of their encounters in either the origin or destination HRR.

Overall, this algorithm designates movers as physicians that (1) are reliably settled in the origin HRR for at least four years before moving, (2) move to the destination HRR, (3) are reliably settled in the destination for at least four years, and (4) do not subsequently move to any other HRR (including moving back to the origin). Figure A8 displays the average share of encounters a mover physician has in the origin HRR (Panel A) and in the destination HRR (Panel B), by relative year. It confirms that movers are well-anchored in the origin and in the destination in the pre-move and post-move years, respectively.

Physician Locations Any physician who is not a mover is a non-mover. We restrict each physician to have only one location in each year (unless it is a mover physician in their mover year). To accomplish this, we assign a separate physician ID to each non-mover physician-HRR combi-

[41] Ties are decided this way because most of them consist of total utilization in the origin HRR in t_{pre} and total utilization in the destination HRR in $t_{post} = t_{pre} + 1$.

[42] Physicians that are not observed in one of the pre-move or post-move years (after entering and before leaving the sample) are assigned as non-movers, since we cannot confidently anchor them in the origin or destination HRRs.

nation. For movers, we assign a single physician ID for all encounters in the origin HRR during pre-move years, in origin and destination HRRs during the move year, and in the destination HRR during post-move years. Other encounters are assigned a separate, physician-HRR specific physician ID. These encounters occur either (i) in an HRR other than the origin or destination, (ii) in the destination HRR before the physician moved, or (iii) in the origin HRR after the physician moved.[43]

Assigning Physician Specialties

To determine a physician's specialty, we use Health Care Finance Administration (HCFA) codes, which are present for each claim in the carrier file. Specifically, we first assign each claim in the carrier files to the performing physician. Next, we proceed as follows:

1. Drop all claims filed under "non-physician" specialty codes, such as "clinical laboratory".[44]

2. Following Fadlon and Van Parys (2020), classify all claims filed under HCFA codes for (i) general practice, (ii) family practice, or (iii) internal medicine as belonging to the same specialty, which we define to be the PCP specialty.

3. Determine which specialty code (or group of specialty codes, in the case of the PCP specialty) is responsible for the plurality of each physician's claim amounts. This is defined to be the physician's specialty.

4. If any physicians only filed claims under a "non-physician" specialty code, they are assigned a single catch-all specialty. This would be assigned, for example, to a physician who only files claims under the HCFA code for "clinical laboratory" as well as a physician who only files claims under the HCFA code for "mass immunization roster". This "non-physician" specialty accounts for just 0.27 percent of total utilization.

5. Finally, if a physician is only present in the inpatient and/or outpatient files and not in the

[43]Note that since each patient-physician match constitutes a single encounter per year, the encounter "occurs" where the patient lives that year.

[44]The full list of non-physician specialty codes is as follows: hospice and palliative care, mammography, independent diagnostic testing facility, ambulatory surgical center, other medical supply company, medical supply company with registered pharmacist, ambulance service, public health or welfare agency, voluntary health or charitable agency, portable X-ray supplier, clinical laboratory, single or multi-specialty clinic or group practice, mass immunization roster, radiation therapy center, slide preparation facilities, all other suppliers, unknown provider, unknown supplier, unknown physician specialty, hospital, pharmacy, and centralized flu clinic.

carrier files, they are assigned a “missing” specialty (since this assignment algorithm is based on carrier file claims only). This “missing” specialty code accounts for 1.15 percent of total utilization.

This results in a unique specialty for every physician in the sample.

Assigning Additional Physician Demographics From the AMA Masterfile

To get additional demographic information on physicians—in particular, age, gender, and years of experience—we use the American Medical Association (AMA) Physician Masterfile. The AMA Masterfile aims to be a census of all allopathic and ostheopathic physicians in the US, including both AMA members and non-members, going back to 1906. Information for the Masterfile comes from a variety of primary sources, including medical schools, post-graduate medical programs, state licensing agencies, state and federal disciplinary actions, the Educational Commission for Foreign Medical Graduates, the American Board of Medical Specialties, the Federal Drug Enforcement Administration, and post-graduate surveys of individual physicians. The institutions involved in the data collection effort provide all relevant information directly to the AMA.[45] As a result, the Masterfile is considered to be a near-census of physicians in the US and is frequently used in designing sampling frames for physician surveys (DesRoches et al. 2015). At the end of each year, the Masterfile is “frozen” so as to provide a historical snapshot of US physicians in that year (Kletke et al. 2000).

We match the 2014 AMA Masterfile records to our baseline analysis sample based on physician identifiers as defined in Section A.2. Our baseline sample consists of about 1.7 million physicians (Table 1). We match 43 percent of them (about 726,000) to their records in the AMA Masterfile.

Table A4 presents summary statistics on this matched sample. It shows that mover physicians tend to be, on average, younger and more likely to be female. We measure age and experience at the year of the move for mover physicians and a randomly generated “move” year for non-movers, where the “move” years for non-movers are sampled, so as to preserve the probability distribution of move years for movers.

There are two reasons why we are sometimes unable to match physicians in our baseline file to their corresponding AMA data. First, 51

[45]Information based on the AMA’s own description is available at https://www.ama-assn.org/about/masterfile/ama-physician-masterfile.

of the physicians in our baseline sample have neither NPIs nor UPINs that can be found in the AMA Masterfile. These "no-match" physicians are likely a combination of a few disparate groups. First, our baseline sample likely includes non-physician medical providers, e.g. nurse practitioners, anaesthesiologists, and institutions with their own NPI or UPIN. These entities are not part of the Masterfile, but appear in the claims data. Second, the "no-match" physicians likely include some mistakenly entered NPIs and UPINs. For instance, while the median physician in our analysis sample is observed for five years, the median for the "no-match physicians" is 3 years, with 28 percent appearing for a single year. This strongly suggests that many of the identifiers among the "no-match" physicians are used infrequently or even sporadically, so some of them may be mistakenly entered. Also, the "no-match" physicians likely include some physicians who are not included in the Masterfile. For instance, previous studies comparing various US physician databases have found that the Masterfile includes a very high, but not complete fraction of the physicians identified in other databases (e.g. the CMS's National Provider Enumeration System) (DesRoches et al. 2015; White et al. 2020).

Benchmarking our Implied Aggregate Statistics on Physicians

We compare the number of physicians and their annual spending in our data against public information on Medicare spending patterns (Appendix Table A14). Specifically, we compare estimates in our data from 2013 against roughly comparable information on physicians in Medicare in 2013 published by the Medicare Payment Advisory Commission (MedPAC) and the U.S Department of Human and Health Services (HHS).[46]

In our data, average annual physician spending in 2013 is approximately $219,000. This average reflects total annual physician spending of $204 billion and a total number of physicians of around 930 thousand. The published Medicare report shows slightly lower average annual physician spending of $209,500, with a higher total annual physician spending of $257 billion partly offset by a higher number of physicians (of around 1.2 million). The higher total annual physician spending figure and lower physician count in our data likely stem from our patient sample restrictions (e.g. the age-based restrictions, the exclusion of patients who move multiple times or who do not experience a clear shift in claims upon move, and other restrictions described in Section 3.2).

[46]Specifically, we benchmark our estimates against the statistics in Medicare Payment Advisory Comission (2021) and U.S. Department of Health and Human Services (HHS) (2014), which contain information on overall spending on physician fees and number of physicians billing Medicare fee-for-service claims in 2013.

Given the complexity of our data construction procedure and the closeness of the estimates reported in the two columns of Table A14, we take these numbers as evidence that the physician spending measure we obtain is reasonable and consistent with external sources.

B Econometric Appendix

B.1 Derivation of the Parametric Encounter Quantity and Per-Encounter Utilization Models in Section 4.1

We can derive our Poisson specification for the number of physician encounters per patient-year in equation (4) from an underlying model of the stochastic process that determines a patient's number of health shocks n_{it} and the latent severity of each shock ψ_{ikt}. Specifically, we model the number of health shocks n_{it} that patient i receives in year t as a Poisson random variable: $n_{it} \mid x_{it}, \mathrm{j}(it) \sim Poisson\left(\exp(\alpha_i^n + x_{it}'\beta^n)\right)$, with an arrival rate that depends on a patient-specific time invariant effect α_i^n and a set of time-varying patient observables x_{it}. We next model the severity of each shock as $\psi_{ikt} = \breve{\alpha}_i + x_{it}'\breve{\beta} + \zeta_{ikt}$, with a patient component that again includes a fixed effect $\breve{\alpha}_i$ and the time-varying controls, and assume the residual satisfies $\zeta_{ikt} \mid x_{it}, \mathrm{j}(it) \sim \exp(1)$. The total number of encounters $N_{it} = \sum_{k=1}^{n_{it}} \mathbf{1}[\psi_{ikt} > \breve{\gamma}_{it}]$ is then conditionally distributed as a Poisson random variable, with mean

$$E[N_{it} \mid x_{it}, \mathrm{j}(it)] = \exp\left(\alpha_i^N + \gamma_{\mathrm{j}(it)}^N + \tau_t^N + x_{it}'\beta^N\right) \tag{13}$$

for $\alpha_i^N = \breve{\alpha}_i + \alpha_i^n$, $\gamma_j^N = -\breve{\gamma}_j$, and $\beta^N = \breve{\beta} + \beta^n$.

Likewise, we can derive the per-encounter utilization model in equation (5) by assuming that the sum of patient health and the physician cost of providing care (i.e. $h_{idt} + g_{idt}$ in the model of per-encounter utilization in equation (1)) can be forecasted by a time effect, a patient effect, a place effect, and sets of time-varying patient and physician observables x_{it} and w_{dt} given encounter locations. That is, we define $\xi_{idt} \equiv h_{idt} + g_{idt}$ and assume $E\left[\xi_{idt} | x, w, \mathrm{j}(it), D_{idt} = 1\right] = \dot{\tau}_t + \dot{\alpha}_i + \dot{\gamma}_{j(it)} + x_{it}'\dot{\beta} + w_{dt}'\dot{\phi}$. This implies that we can express utilization y_{idt} among realized encounters (with $D_{idt} = 1$) as

$$y_{idt} = \underbrace{\alpha_i + \tau_t + x_{it}'\beta}_{\equiv\tilde{\alpha}_{it}} + \underbrace{\delta_d + w_{dt}'\phi}_{\equiv\tilde{\delta}_{dt}} + \underbrace{\gamma_{\mathrm{j}(it)}}_{\equiv\tilde{\gamma}_{\mathrm{j}(it)}} + \varepsilon_{idt}, \tag{14}$$

with $E[\varepsilon_{idt} \mid x, w, \mathrm{j}(it), D_{idt} = 1] = 0$, where $\alpha_i = a_i + \dot{\alpha}_i$, $\tau_t = \dot{\tau}_t$, $\gamma_j = c_j + \dot{\gamma}_j$, $\beta = \dot{\beta}$, and $\phi = \dot{\phi}$.

Note that $\dot{\alpha}_i$ and $x_{it}'\dot{\beta}$ can be arbitrarily correlated with the corresponding components of shock severity, $\breve{\alpha}_i$ and $x_{it}'\breve{\beta}$.

B.2 Mover Identification of the Poisson Encounter Model

This appendix shows how the Poisson fixed effects model identifies causal effects under a "common growth rates" assumption, similar to the common trends assumption identifying effects in linear regression models. Suppose

$$y_{it} \sim Poisson(\lambda_{it}) \tag{15}$$

where $\ln \lambda_{it} = x_{it}'\beta$. Suppose $t \in \{0,1\}$ and $x_{it}'\beta = \alpha_i + \tau T_{it} + \gamma D_{it}$, for binary D_{it} and where $T_{it} = \mathbf{1}\{t = 1\}$. Let $c_i = \sum_t y_{it}$. Following Hausman et al. (1984), the population log-likelihood first-order condition for τ and γ can be written

$$\begin{aligned}
0 =& E\left[c_i \sum_t \begin{bmatrix} T_{it} \\ D_{it} \end{bmatrix} \left(\frac{y_{it}}{c_i} - \frac{\exp(\tau^* T_{it} + \gamma^* D_{it})}{\sum_s \exp(\tau^* T_{is} + \gamma^* D_{is})}\right)\right] \\
=& \sum_{c=1}^{\infty} c p_c E\left[\sum_t \begin{bmatrix} T_{it} \\ D_{it} \end{bmatrix} \left(\frac{y_{it}}{\sum_t y_{it}} - \frac{\exp(\tau^* T_{it} + \gamma^* D_{it})}{\sum_s \exp(\tau^* T_{is} + \gamma^* D_{is})}\right) \mid c_i = c\right] \\
=& \sum_{c=1}^{\infty} c p_c \begin{bmatrix} p_{0\to1|c}\left(S^1_{0\to1,c} - \frac{\exp(\tau^*)}{1+\exp(\tau^*+\gamma^*)}\right) + p_{1\to0|c}\left(S^0_{1\to0,c} - \frac{\exp(\gamma^*)}{\exp(\tau^*)+\exp(\gamma^*)}\right) \\ p_{0\to0|c}\left(S^1_{0\to0,c} - \frac{\exp(\tau^*)}{1+\exp(\tau^*)}\right) + p_{0\to1|c}\left(S^1_{0\to1,c} - \frac{\exp(\tau^*+\gamma^*)}{1+\exp(\tau^*+\gamma^*)}\right) \end{bmatrix} \\
&+ \sum_{c=1}^{\infty} c p_c \begin{bmatrix} p_{1\to1|c}\left(S^0_{1\to1,c} - \frac{\exp(\gamma^*)}{\exp(\gamma^*)+\exp(\tau^*+\gamma^*)}\right) + p_{1\to1|c}\left(S^1_{1\to1,c} - \frac{\exp(\tau^*+\gamma^*)}{\exp(\gamma^*)+\exp(\tau^*+\gamma^*)}\right) \\ p_{1\to0|c}\left(S^1_{1\to0,c} - \frac{\exp(\tau^*)}{\exp(\tau^*)+\exp(\gamma^*)}\right) + p_{1\to1|c}\left(S^1_{1\to1,c} - \frac{\exp(\tau^*+\gamma^*)}{\exp(\beta^*)+\exp(\tau^*+\gamma^*)}\right) \end{bmatrix},
\end{aligned} \tag{16}$$

where $p_c = Pr(c_i = c)$, $p_{j\to k,c} = Pr(D_{i0} = j, D_{i1} = k \mid c_i = c)$, and

$$S^t_{j\to k,c} = E\left[\frac{y_{it}}{c} \mid D_{i0} = j, D_{i1} = k, c_i = c\right]. \tag{17}$$

First suppose, as in a canonical difference-in-differences setting, that all individuals are untreated in period $t = 0$ and some individuals switch into treatment in period $t = 1$. Then equation (16) simplifies to

$$0 = \begin{bmatrix} \sum_{c=1}^{\infty} c p_c p_{0\to1|c}\left(S^1_{0\to1,c} - \frac{\exp(\tau^*)}{1+\exp(\tau^*+\gamma^*)}\right) \\ \sum_{c=1}^{\infty} c p_c p_{0\to0|c}\left(S^1_{0\to0,c} - \frac{\exp(\tau^*)}{1+\exp(\tau^*)}\right) + \sum_{c=1}^{\infty} c p_c p_{0\to1|c}\left(S^1_{0\to1,c} - \frac{\exp(\tau^*+\gamma^*)}{1+\exp(\tau^*+\gamma^*)}\right) \end{bmatrix}. \tag{18}$$

Note that $\sum_{c=1}^{\infty} c p_c p_{0\to1|c} S^1_{0\to1,c} = E[y_{i1} \mid D_{i1} = 1]$, $\sum_{c=1}^{\infty} c p_c p_{0\to0|c} S^1_{0\to0,c} = E[y_{i1} \mid D_{i0} = 1]$, and

$$\sum_{c=1}^{\infty} c p_c p_{0\to1|c} = Pr(D_{i1} = 1) = 1 - Pr(D_{i1} = 0) = 1 - \sum_{c=1}^{\infty} c p_c p_{0\to0|c}. \tag{19}$$

Solving out for the estimands thus yields

$$\tau^* = \ln\left(\frac{E[y_{i1} \mid D_{i1}=0]}{E[y_{i0} \mid D_{i1}=0]}\right) \tag{20}$$

$$\gamma^* = \ln\left(\frac{E[y_{i1} \mid D_{i1}=1]}{E[y_{i0} \mid D_{i1}=1]}\right) - \ln\left(\frac{E[y_{i1} \mid D_{i1}=0]}{E[y_{i0} \mid D_{i1}=0]}\right). \tag{21}$$

The treatment coefficient γ^* is the difference in log growth rates among those treated and untreated in period 1. Let $y_{it}(0)$ and $y_{it}(1)$ denote untreated and treated potential outcomes of individual i in time t, respectively, and consider an assumption of common log growth rates:

$$\frac{E[y_{i1}(0) \mid D_{i1}=1]}{E[y_{i0}(0) \mid D_{i1}=1]} = \frac{E[y_{i1}(0) \mid D_{i1}=0]}{E[y_{i0}(0) \mid D_{i1}=0]}. \tag{22}$$

Under this assumption, we have

$$\gamma^* = \ln\left(\frac{E[y_{i1}(1) \mid D_{i1}=1]}{E[y_{i1}(0) \mid D_{i1}=1]}\right). \tag{23}$$

This shows that in the simple difference-in-differences setting, the Poisson fixed effect regression identifies the log percentage effect of treatment on the treated under common log growth rates.

A similar result holds for the simplest mover design, in which individuals either switch out of or into treatment in $t=1$. Specifically, it can be shown that the treatment coefficient satisfying equation (16) simplifies in this case to

$$\gamma^* = \frac{1}{2}\left(\ln\left(\frac{E[y_{i1} \mid D_{i1}=1]}{E[y_{i0} \mid D_{i1}=1]}\right) - \ln\left(\frac{E[y_{i1} \mid D_{i0}=1]}{E[y_{i0} \mid D_{i0}=1]}\right)\right). \tag{24}$$

Thus, under the same common log growth rate assumption as above,

$$\begin{aligned}
\gamma^* =& \frac{1}{2}\left(\ln\left(\frac{E[y_{i1}(1) \mid D_{i1}=1]}{E[y_{i0}(0) \mid D_{i1}=1]}\right) - \ln\left(\frac{E[y_{i1}(0) \mid D_{i0}=1]}{E[y_{i0}(1) \mid D_{i0}=1]}\right)\right) \\
=& \frac{1}{2}\left(\ln\left(\frac{E[y_{i1}(1) \mid D_{i1}=1]}{E[y_{i0}(0) \mid D_{i1}=1]}\right) - \ln\left(\frac{E[y_{i1}(0) \mid D_{i1}=1]}{E[y_{i0}(0) \mid D_{i1}=1]}\right)\right) \\
&+ \frac{1}{2}\left(\ln\left(\frac{E[y_{i1}(0) \mid D_{i0}=1]}{E[y_{i0}(0) \mid D_{i0}=1]}\right) - \ln\left(\frac{E[y_{i1}(0) \mid D_{i0}=1]}{E[y_{i0}(1) \mid D_{i0}=1]}\right)\right) \\
=& \frac{1}{2}\left(\ln\left(\frac{E[y_{i1}(1) \mid D_{i1}=1]}{E[y_{i1}(0) \mid D_{i1}=1]}\right) + \ln\left(\frac{E[y_{i0}(1) \mid D_{i0}=1]}{E[y_{i0}(0) \mid D_{i0}=1]}\right)\right)
\end{aligned} \tag{25}$$

which is the average log percentage treatment-on-the-treated effect across the two time periods.

B.3 Decomposition of Geographic Variation

This appendix formalizes our counterfactual analysis of how differences in average annual log patient utilization in each HRR, as represented in equation (10), change as we equalize the various underlying sources of utilization differences. This analysis proceeds in seven incremental steps: first by shutting down patient-physician selection and then exploring sequentially the effect of eliminating variation due to physicians, patients, and the practice environment.

Our first counterfactual sets $\bar{\sigma}_j = 0$:

$$\bar{y}_j^{(1)} = p_j(\bar{\alpha}_j + \gamma_j + \overline{N}_j + \bar{\delta}_j) \tag{26}$$

This captures how the geographic distribution of healthcare utilization would change if there were no systematic differences in the allocation of patient encounters to physicians with different practice intensities, holding fixed the number of physicians each patient sees and the set of patients and physicians in each HRR. The mean of $\bar{y}_j^{(1)}$ is displayed in row (2) of Tables 5 and 6.

We then remove geographic variation in average physician intensity due to different physician practice styles within the same specialty by setting $\bar{\delta}_j$ to the average value of $\bar{\delta}_{it}^S$ across HRRs $\bar{\delta}_j^S$, where $\bar{\delta}_{it}^S$ is defined as:

$$\bar{\delta}_{it}^S = \mathbb{E}\left[\ln\left(\frac{1}{N_{it}}\sum_{d\in\mathscr{D}^{*S}\left(N_{it}^1,\ldots,N_{it}^S\right)}\exp(\delta_d + w'_{dt}\beta + \varepsilon_{idt}\right)\right]$$

where $\mathscr{D}^{*S}(n^1,\ldots,n^S)$ gives a random set of physicians drawn from the entire country consisting of n^s physicians from each specialty $s = 1,\ldots,S$, and N_{it}^s denotes the number of physicians in specialty s that individual i sees in year t. The distribution of

$$\bar{y}_j^{(2)} = p_j(\bar{\alpha}_j + \gamma_j + \overline{N}_j + \bar{\delta}_j^S) \tag{27}$$

captures how the geographic distribution of healthcare utilization would change if there were no systematic differences in the average within-specialty physician practice styles across areas. The mean of $\bar{y}_j^{(2)}$ is displayed in row (3) of Tables 5 and 6.

The next counterfactual eliminates differential practice intensity variation of physicians across specialties. Define

$$\bar{\delta}_{it}^U = \mathbb{E}\left[\ln\left(\frac{1}{N_{it}}\sum_{d\in\mathscr{D}^{*}(N_{it})}\exp(\delta_d + w'_{dt}\beta + \varepsilon_{idt}\right)\right]$$

where $\mathscr{D}^*(n)$ gives a random set of n physicians drawn from the entire country and N_{it} denotes the number of physicians individual i sees in year t. Letting $\bar{\delta}^U$ be the average of $\bar{\delta}_{it}^U$ across all patient-years in the sample, write

$$\bar{y}_j^{(3)} = p_j(\bar{\alpha}_j + \gamma_j + \overline{N}_j + \bar{\delta}^U). \tag{28}$$

The difference between $\bar{y}_j^{(3)}$ and $\bar{y}_j^{(2)}$ captures how much utilization differences across HRRs are affected by the sorting of physicians of different specialties across HRRs. The mean of $\bar{y}_j^{(3)}$ is displayed in row (4) of Table 5 (it is not applicable to Table 6).

The next two counterfactuals eliminate geographic variation due to patients:

$$\bar{y}_j^{(4)} = p_j(\bar{\alpha} + \gamma_j + \overline{N}_j + \bar{\delta}^U) \tag{29}$$

$$\bar{y}_j^{(5)} = \ddot{p}_j(\bar{\alpha} + \gamma_j + \ddot{N}_j + \bar{\delta}^U). \tag{30}$$

Here $\bar{y}^{(4)}$ sets $\bar{\alpha}_j$ to its average $\bar{\alpha}$, thus eliminating geographic variation due to patient effects on utilization per encounter. Its mean across HRRs is displayed in row (5) of Table 5 and row (4) of Table 6. Then, $\bar{y}_j^{(5)}$ leverages the encounter model (equation 4) to eliminate geographic variation coming from patient effects on the number of encounters. Specifically we define, in contrast to p_j,

$$\ddot{p}_j = 1 - E[\exp(-\alpha_i^N - \gamma_j^N - x_{it}'\beta^N)], \tag{31}$$

where here $E[\cdot]$ is understood as averaging over all individuals. Thus $\ddot{p}_j$ captures the share of individuals with any healthcare utilization in area j, given a random geographic reallocation of patients. This is a known function of the encounter model parameters α_i^N, γ_j^N, and β^N. We similarly define $\ddot{N}_j$ as the average log number of physicians seen in area j, when non-zero, under random patient reallocation. The mean of $\bar{y}_j^{(5)}$ across HRRs is displayed in row (6) of Table 5 and row (5) of Table 6.

Following these five counterfactuals, the only geographic variation left is that due to practice environment effects on utilization per encounter and number of encounters. We separate these two factors through our final counterfactual, which sets γ_j to its average $\bar{\gamma}$:

$$\bar{y}_j^{(6)} = \ddot{p}_j(\bar{\alpha} + \bar{\gamma} + \ddot{N}_j + \bar{\delta}^U) \tag{32}$$

The only geographic variation in $\bar{y}_j^{(6)}$ (displayed in row 7 of Table 5 and row 6 of Table 6) is due to practice environment effects on the number of physicians seen, via $\ddot{p}_j$ and $\ddot{N}_j$ (which we then eliminate in a final step, $y^{(7)}$ (row 8 of Table 5 and row 7 of Table 6). Taken together, $\bar{y}_j$ and $\bar{y}_j^{(1)} - \bar{y}_j^{(6)}$ thus provide a full accounting of the partial equilibrium role that each of the primary factors of interest (physicians, practice environment, and patients) play in the geographic variation in utilization.

To disentangle the role of between-specialty sorting of PCPs and non-PCPs from between-specialty sorting of non-PCP specialists in Appendix Table A7, we add an additional counterfactual between $\bar{y}_j^{(2)}$ and $\bar{y}_j^{(3)}$. Specifically, we define

$$\bar{y}_j^{(2.5)} = p_j\left(\bar{\alpha}_j + \gamma_j + \bar{N}_j + \bar{\delta}_j^{\mathrm{PCP}}\right) \tag{33}$$

where $\bar{\delta}_j^{\text{PCP}}$ is the within-HRR average of

$$\bar{\delta}_{it}^{\text{PCP}} = \mathbb{E}\left[\ln\left(\frac{1}{N_{it}} \sum_{d \in \mathscr{D}^{*\text{PCP}}(N_{it}^{\text{PCP}}, N_{it}^{S})} \exp\left(\delta_d + w_{dt}'\beta + \varepsilon_{idt}\right)\right)\right]$$

Here, $\mathscr{D}^{*\text{PCP}}(n_1, n_2)$ gives a random set of $n_1 + n_2$ physicians drawn from the entire country, of which n_1 are PCPs and n_2 are non-PCPs. N_{it}^{PCP} represents the number of PCPs patient i sees in year t, while N_{it}^{S} represents the number of non-PCPs they see in year t. This equalizes the distribution of specialists (i.e. non-PCPs) across areas, so that $\bar{y}_j^{(3)} - \bar{y}_j^{(2.5)}$ gives the role of sorting of non-PCP specialists across areas; $\bar{y}_j^{(3)} - \bar{y}_j^{(2)}$ continues to give the role of physician sorting of all specialties. We display results from the decomposition in full detail (i.e. including $\bar{y}_j^{(2.5)}$, given in row 3) in Appendix Table A7.

The results from the counterfactual calculations are naturally not invariant to the order in which the steps are performed. In Section C.3 below, we show that our results are robust to different orderings of these steps.

B.4 Alternative Decomposition of Geographic Variation

This appendix formalizes an alternative counterfactual analysis of how differences in average annual log patient utilization in each HRR, as represented in equation (10), change as we equalize the various underlying sources of utilization differences. Each object is defined as in Appendix B.3. The key departure from our baseline analysis is that we now label the practice environment effect on the number of encounters as part of the physician component.

The first step of the decomposition is the same: starting with the average annual log patient utilization $\bar{y}_j$ in each HRR, we equalize the effect of selection by setting $\bar{\sigma}_j = 0$:

$$\bar{y}_j^{(1)} = p_j\left(\bar{\alpha}_j + \gamma_j + \bar{N}_j + \bar{\delta}_j\right) \tag{34}$$

Note, however, that it no longer makes sense to decompose the effects of physicians, patients, and then the practice environment if we want to count the practice environment effect on the number of encounters as part of the physician role. Instead, we first equalize the effect of patients on the utilization per encounter by setting $\bar{\alpha}_j$ to its average $\bar{\alpha}$:

$$\bar{y}_j^{(2)} = p_j\left(\bar{\alpha} + \gamma_j + \bar{N}_j + \bar{\delta}_j\right) \tag{35}$$

Next, we equalize the patient effect on the number of encounters by replacing p_j and $\bar{N}_j$ with $\ddot{p}_j$ and $\ddot{N}_j$, respectively:

$$\bar{y}_j^{(3)} = \ddot{p}_j\left(\bar{\alpha} + \gamma_j + \ddot{N}_j + \bar{\delta}_j\right) \tag{36}$$

We then find the within-specialty physician effect on utilization by simultaneously equalizing the within-specialty physician effect on utilization per encounter by setting $\bar{\delta}_j$ to $\bar{\delta}_j^S$ and the practice environment effect on the number of encounters by setting $\ddot{p}_j$ and $\ddot{N}_j$ to their respective averages $\bar{p}$ and $\bar{N}$:

$$\bar{y}_j^{(4)} = \bar{p}\left(\bar{\alpha} + \gamma_j + \bar{N} + \bar{\delta}_j^S\right) \tag{37}$$

Next, we compute the between-specialty physician effect on utilization per encounter by setting $\bar{\delta}_j^S$ to $\bar{\delta}^U$:

$$\bar{y}_j^{(5)} = \bar{p}\left(\bar{\alpha} + \gamma_j + \bar{N} + \bar{\delta}^U\right) \tag{38}$$

Lastly, we equalize the practice environment effect on utilization per encounter by setting γ_j to its average across HRRs, $\bar{\gamma}$:

$$\bar{y}_j^{(6)} = \bar{p}\left(\bar{\alpha} + \bar{\gamma} + \bar{N} + \bar{\delta}^U\right) \tag{39}$$

Comparing $\bar{y}_j^{(5)}$ to $\bar{y}_j^{(3)}$ gives the role of physicians if we count what was originally the practice environment effect on the number of encounters toward the physician effect. The results of this exercise are displayed in Table A8. Compared to the original decomposition in Table 5, the results are intuitive: the role of patients is about the same, while the role of physicians jumps from 35% to 68%—the approximate result if we added the original physician effect (35%) to the practice environment effect on the encounter margin (32%). The practice environment share is now negative 20%, consistent with the original practice environment effect on utilization per encounter.

B.5 Event Study Decomposition

This appendix first shows how the event studies estimated in Section 3.3 can be derived from restricted versions of the model in Section 2. We then present a more general event study representation obtained from estimates of the model components.

We derive equation (2) under the assumption that the final two terms in our patient-level utilization model (equation (6)) are additively separable in patient and practice environment effects; i.e., that

$$\ln N_{it} + \ln\left(\frac{1}{N}\sum_{d\in D_{it}} \exp\left(\delta_d + w'_{dt}\phi + \varepsilon_{idt}\right)\right) = \underline{\alpha}_i + \underline{\gamma}_j^D + \eta_{it}^P \tag{40}$$

where η_{it}^P is a mean-zero residual. Note that then equation (6) for patient movers can be rewritten

$$y_{it} = \alpha_i^P + \tau_t^P + \left(\mathbf{1}[\mathrm{r}(i,t) > 0]S_i^P\right)\Delta_i^P + x'_{it}\beta^P + \rho_{\mathrm{r}(i,t)}^P + \eta_{it}^P, \tag{41}$$

where, with $o(i)$ and $d(i)$ indexing the origin and destination HRR of patient i, respectively, we define $\alpha_i^P = \alpha_i + \underline{\alpha}_i + \gamma_{o(i)}$, $\tau_t^P = \tau_t$, and $S_i^P = (\gamma_{d(i)} + \underline{\gamma}_{d(i)} - (\gamma_{o(i)} + \underline{\gamma}_{o(i)}))/\Delta_i^P$. This S_i^P denotes the share of the observed difference in utilization between a patient mover's destination and origin HRR both due to the practice environment effects γ_j and the component of doctor-driven utilization due to practice environment, $\underline{\gamma}_j^D$. Equation (41) shows that the patient event study jump estimated in Section 3.3 captures an average of this S_i^P share under the restriction (equation (40)). We also explicitly include the main relative-year effects $\rho_{(i,t)}^P$ in the equation above by splitting it off from the vector of time-varying patient observables x_{it} for the purpose of clarity.

A similar restriction motivates the physician event study regression (equation (3)). Aggregating the model of utilization for each physician in each year, we obtain

$$y_{dt} = \delta_d + \tau_t + \gamma_{\mathrm{j}(d,t)} + w'_{dt}\phi + \ln N_{dt} + \ln\left(\frac{1}{N_{dt}}\sum_{i\in P_{dt}} \exp\left(\alpha_d + x'_{it}\beta + \varepsilon_{idt}\right)\right). \tag{42}$$

Suppose the final term of this expression is additively separable in physician and practice environment effects; i.e., that

$$\ln\left(\frac{1}{N_{dt}}\sum_{i\in P_{dt}} \exp\left(\alpha_d + x'_{it}\beta + \varepsilon_{idt}\right)\right) = \underline{\delta}_d + \underline{\gamma}_j^P + \eta_{dt}^D \tag{43}$$

where η_{dt}^D is again a mean-zero residual. Now, we obtain for physician movers

$$y_{dt} - \ln N_{dt} = \alpha_d^D + \tau_t^D + \left(\mathbf{1}[\mathrm{r}(i,t) > 0]S_d^D\right)\Delta_d^D + \rho_{\mathrm{r}(d,t)}^D + \eta_{dt}^D, \tag{44}$$

where, with $o(d)$ and $d(d)$ indexing the origin and destination HRR of doctor d, respectively, we define $\alpha_d^D = \delta_d + \underline{\delta}_d + \gamma_{o(d)}$, $\tau_t^D = \tau_t$, and $S_d^D = (\gamma_{d(d)} + \underline{\gamma}_{d(d)} - (\gamma_{o(d)} + \underline{\gamma}_{o(d)}))/\Delta_d^D$. This S_d^D denotes the share of the observed differences in utilization between a physician mover's destination and origin HRR due to both the practice environment effects γ_j and the component of patient-driven utilization due to practice environment $\underline{\gamma}_j^P$. Equation (44) shows that the doctor event study jump estimated in Section 3.3 captures an average of this S_d^D share under the restriction in equation (43). We again explicitly re-label the main relative-year effects $\rho_{\mathrm{r}(d,t)}^D$ from w_{dt} in the specification above for the purpose of clarity.

B.6 Full Event Study Decomposition

This appendix derives an enriched event study decomposition of the different drivers of geographic variation in patient healthcare utilization, which does not leverage the restriction in Appendix B.5.

Consider equation (9). Note that $\ln N_{it}$ is directly observed and that $\overline{\delta}_{it}$ and σ_{it} may be estimated by first-step estimates of the physician-level parameters δ_d and ϕ and an assumption on the distribution of residual utilization variation ε_{idt} (for example, that ε_{idt} is *iid* given (x, w, j)). Then, by subtracting estimates of $\ln N_{it} + \overline{\delta}_{it} + \sigma_{it}$ from observed utilization y_{it}, we obtain a model like equation (41) which does not impose the restriction in equation (40). In particular, this suggests that a patient-level event study using this adjusted $y_{it} - (\ln N_{it} + \overline{\delta}_{it} + \sigma_{it})$ as an outcome may capture a weighted average of practice environment effect shares $S_i^{P*} = (\gamma_{d(i)} - \gamma_{o(i)})/\Delta_i^P$, by taking into account the potentially non-additive predicted change in physician availability, sorting, and number of encounters. The difference between this adjusted event study and the motivating patient-level event study in Figure 3 can furthermore be evaluated by performing event studies on each of the subtracted components $\ln N_{it}$, $\overline{\delta}_{it}$, and σ_{it}. These auxiliary event studies replace y_{it} in equation (2) with each of these components, with the sum of event study coefficients θ_r across specifications equaling, by construction, the difference between the adjusted and original event study jumps. A large event study jump in a regression of $\overline{\delta}_{it}$ would, for example, suggest that a sizable proportion of the aggregate event study jump in Figure 3 is due to differences in the availability of physicians with different utilization effects across different HRRs. Similarly, event study jumps in $\ln N_{it}$ or σ_{it} would suggest that some of the aggregate event study jump in Figure 3 arises from systematic differences in the number of encounters or matching of patients to physicians across HRRs.

We use the parameters from our encounter-level estimation to perform this decomposition. We first form simulation-based estimates of the average physician utilization and selection terms $\overline{\delta}_{it}$ and σ_{it}.[47] We then use these estimates as components in estimating the enriched event study decomposition (equation 41).[48]

The results from this event study decomposition are presented in Appendix Figure A5. In Panel A, we reproduce the patient-level event study using adjusted annual log utilization $y_{it} - (\ln N_{it} + \overline{\delta}_{it} + \sigma_{it})$ as the outcome. As discussed above, the height of the resulting event study jump estimates a weighted average of the share of practice environment effects in average log utilization differences, net of any potential contribution of physicians to geographic utilization differences. The figure shows that netting out the contribution of physicians to annual utilization differences alters the event study substantially. While the unadjusted event study shows that the place share including physician utilization differences is around 50 percent (Figure (3), Panel A) netting out

[47]Specifically, for each patient and year, we take a random draw of physicians from her HRR with the number of physicians equaling her actual number of encounters for this patient and year. We use these simulated encounters averaged over 100 random draws to form estimates of $\overline{\delta}_{it}$ and σ_{it}.

[48]For the 6 percent of patient-years that have no utilization, we set $\overline{\delta}_{it}$ and σ_{it} equal to 0 since there is no actual patient-physician encounter.

the contribution of physicians decreases the estimated jump to around 10 percent (Appendix Figure A5, Panel A), which is very similar to the estimated role of the practice environment in Table 5.

Panels B-C of Appendix Figure A5 decompose the 40 percentage point difference in the size of the jump between the simple patient-level event study in Figure 3 (Panel A) and the enriched analysis in Panel A of Appendix Figure A5. Around half of this difference in the event study jump is due to an effect on the increased number of physicians seen (Panel B), with the half due to more intense practice of the available physician stock (Panel C). The event study on the residual selection term σ_{it} (Panel D) is flat, suggesting that none of the variation in utilization across HRRs is due to the differential sorting of patients to physicians with different demand and practice intensities.

C Robustness Analysis

C.1 Physicians Practicing in Multiple HRRs

This appendix investigates the sensitivity of our findings to how we handle physicians who practice in multiple HRRs. As described in our sample construction procedure, we assign a separate physician ID to each physician-HRR combination for physician non-movers, and for physician-movers for utilization outside of either their origin or their destination. Since this procedure creates multiple fixed effects for each physician in the sample, a natural concern is that the high degree of variation in physician practice intensities we find, as well as our estimates of its contribution to geographic variation in utilization, are driven by this fragmentation.

To evaluate the impact of physician fragmentation, we divide physician-HRR combinations into "primary" and "non-primary" categories each year (where the primary HRR is the one where the physician sees the majority of her patients that year) and perform separate utilization decompositions by these groups. Reassuringly, we find that physician utilization in the primary HRR accounts for almost all of the difference in utilization across HRRs. Thus, unsurprisingly, we find broadly similar shares of the utilization differences attributed to physicians, patients, and places in the subsample of primary HRR utilization as in our main analysis.

Primary and Non-Primary HRRs

We define a physician's primary HRR in a given year as the HRR from which she sees the most patients (with ties broken by total utilization in that HRR-year). Any other HRR from which she sees patients is considered non-primary. We expect that a physician's primary HRR is the HRR where she is most likely to reside and to see the majority of her patients. A non-primary HRR is likely to arise from a combination of factors such as data entry errors, lags in updating patient address information, and patients crossing HRR boundaries to seek care.

While fragmentation is relatively common, most utilization occurs in a physician's primary HRR. As noted above, we have 1.7 million original physicians in our sample (defined by a combination of UPIN and NPI identifiers). HRR-based fragmentation results in 8 million estimated physician IDs in our empirical analysis. About 66 percent of the original 1.7 million physicians are "fragmented" (i.e. are non-movers who see a patient from more than one HRR in some year) due to our sample construction procedure. The average physician among the original 1.7 million has 80 percent and 82 percent of her annual patients and utilization, respectively, in her primary HRR for that year.[49]

Decomposition of Differences in Primary HRR Utilization

Our fragmentation procedure could present a problem for our analysis if it induces a correlation between the independent variables in our per-encounter utilization regression and the unobserved component of per-encounter utilization. For instance, if a physician's encounters with patients from high-utilization HRRs tend to occur when the unobserved utilization shock for those encounters is also high, our procedure would create additional high-utilization physician IDs and the contribution of physicians to cross-HRR variation in utilization would be overestimated. If, on the other hand, fragmentation occurs largely due to, e.g., data entry errors, this would tend to create physician IDs roughly equivalent to the primary ID whose distribution is orthogonal to cross-HRR utilization patterns, and whose influence on our counterfactual results should be negligible.

To evaluate these possibilities, we conduct a robustness check which repeats our decomposition of per-encounter utilization just on encounters in primary HRRs. Specifically, we first estimate a version of the encounter model which takes as its outcome the number of encounters a patient has each year with physicians in their primary HRR. We next decompose HRR-average utilization $\bar{y}_j$ into a component due to utilization with physicians in their primary HRR $\bar{y}_j^p$ and physicians not in their primary HRR $\bar{y}_j^{np}$. Finally, we apply to $\bar{y}_j^p$ a decomposition analogous to equation (10),

[49] For the purposes of our analysis, we assign the unique HRR in which we observe the remaining 34 percent of the original 1.7 million physicians as their primary HRR.

$$\bar{y}_j^p = p_j^p(\overline{\alpha}_j^p + \gamma_j^p + \overline{N}_j^p + \overline{\delta}_j^p + \overline{\sigma}_j^p) \tag{45}$$

where p_j^p now denotes the probability of positive utilization with a physician in their primary HRR, $\overline{\alpha}_j^p$ is the average patient-year component $(\alpha_i + \tau_t + x_{it}'\beta)$ among those with positive utilization with a physician in their primary HRR, $\overline{N}_j^p$ is the average number of log primary HRR physician encounters among those with positive utilization, $\overline{\delta}_j^p$ is the average physician component $(\overline{\delta}_{it})$ among physicians in their primary HRR, and $\overline{\sigma}_j^p$ is the corresponding average selection component. As in Table 5, we use the model estimates to sequentially eliminate cross-HRR differences in these components of $\bar{y}_j^p$.

Column (2) of Appendix Table A9 shows the results. The sum of the rows indicate that 99 percent of the observed differences in utilization across HRRs can be explained by factors driving utilization by physicians in their primary HRR $\bar{y}_j^p$. Therefore, not surprisingly, it shows that if we focus only on primary HRR utilization, we find broadly similar shares of the utilization difference attributed to physicians, patients, and places as in our main analysis.

The first row shows that the observed difference in utilization between above and below median HRRs is roughly the same with our alternative decomposition. Specifically, the observed difference in $\bar{y}_j^p + \bar{y}_j^{np}$ between HRRs with above- and below-median utilization is 0.206, which is similar to the corresponding difference in $\bar{y}_j$ in Table 5 (at 0.253).[50] The next rows of the table show that counterfactuals affecting $\bar{y}_j^p$ drive most of the observed difference in utilization across HRRs, with only 1 percent of this difference being due to utilization by physicians in their non-primary HRRs. In other words, physician utilization in their non-primary HRR seems to be independently distributed "noise".

Not surprisingly, therefore, focusing only on encounters in primary HRRs yields a similar decomposition of the drivers of differences in utilization across HRRs as our baseline analysis. For example, we find that differences in $\overline{\delta}_j^p$ explain 33 percent of the difference in utilization between above- and below-median utlization HRRs, similar to the 35 percent in Table 5.

[50]The sum of $\bar{y}_j^p$ and $\bar{y}_j^{np}$ is approximately $\bar{y}_j$. Approximation error comes from the fact that we model log patient-year utilization, which is not additive in log primary and non-primary per-encounter utilization.

C.2 Split-Sample Estimates

We construct split-sample standard deviations and correlations in Appendix Table A15 in order to purge any potential mechanical biases from correlated estimation error across the patient and practice environment components in the encounter model as well as the patient, physician, and practice environment components in the per-encounter utilization model. Comparing Appendix Table A15 with Table 3 suggests such bias is minimal. Qualitatively, these tables are extremely similar; almost all of the split-sample correlations are within two bootstrapped standard errors of the full-sample correlations.

To construct Appendix Table A15, we first randomly split the sample of patients in half, stratifying on origin and destination HRRs (or home HRR, for non-movers). We then find the largest connected set of patients and physicians for each split sample and fit the encounter and per-encounter utilization models on both connected sets. We then compute the average patient and place components in each HRR for the encounter model, as well as the average patient, physician, and place components in each HRR for the per-encounter utilization model. The connected sets in the first and second half contain 1,514,920 and 1,480,093 patients, respectively.

The across-sample covariance of each of the HRR-level estimates yields an unbiased estimate of the true HRR-level covariance under *iid* patient sampling. Formally, with $\hat{\kappa}_c^{(n)}$ denoting the estimate of HRR-level parameter κ_c for

$c \in \{\text{patient}_{\text{encounter}}, \text{environment}_{\text{encounter}}, \text{patient}_{\text{utilization}}, \text{physician}_{\text{utilization}}, \text{environment}_{\text{utilization}}\}$

in sample $n \in \{1,2\}$, the sample covariance of $\hat{\kappa}_c^{(1)}$ and $\hat{\kappa}_c^{(2)}$ yields an unbiased estimate of the variance of κ_c while the sample covariances of $\hat{\kappa}_{c_1}^{(1)}$ and $\hat{\kappa}_{c_2}^{(2)}$ or $\hat{\kappa}_{c_1}^{(2)}$ and $\hat{\kappa}_{c_2}^{(1)}$ for $c_1 \neq c_2$ yield unbiased estimates of the covariance of κ_{c_1} and κ_{c_2}; we average the latter two covariance estimates. We then translate our unbiased estimates of the full covariance matrix of

$$\left(\kappa_{\text{patient}_{\text{encounter}}}, \kappa_{\text{environment}_{\text{encounter}}}, \kappa_{\text{patient}_{\text{utilization}}}, \kappa_{\text{physician}_{\text{utilization}}}, \kappa_{\text{environment}_{\text{utilization}}}\right)$$

to standard deviations and correlations, to match Table 3.

C.3 Counterfactual Order Robustness Check

This appendix investigates the robustness of the decomposition results shown in Table 5 to changes in the order in which we perform the utilization counterfactual steps detailed in Section B.3. Table A16 shows the result from varying the counterfactual step order. In all cases, we always eliminate the physician selection term first and, for the purposes of brevity and clarity, combine the patient and practice environment effects on number of encounters and utilization per encounter into single patient and practice environment effects respectively. Similarly, we combine the within- and between-specialty physician effects into a single physician effect.

Each row of the table corresponds to a utilization factor in the order in which they were presented in Table 5. Each column then shows the incremental contribution of each factor (in percentage terms) to overall geographic variation in utilization (defined as the difference in average patient utilization for HRRs above and below the median as in Table 5) if we were to perform the counterfactual steps in the order indicated by the column heading. For example, row 2 of column (1) shows that physician practice intensity is responsible for 35 percent of overall geographic variation in utilization if we were to eliminate variation coming from physicians first, followed by patients and practice environment respectively (the same order as in the baseline analysis). Row 2 of column (2) then shows that this contribution remains unchanged if we were to eliminate variation from physicians first and then practice environment and patients respectively.

Overall, the table indicates that the results from our counterfactual utilization analysis are remarkably consistent across different orderings of the counterfactual steps. Physicians account for about 35 percent of variation in most cases, patients for about 45 percent, and the practice environment for about 20 percent. While there is some variation across columns in these numbers, the basic conclusion that patients account for about half of all utilization variation and that there is a significant role for both physician practice intensity and for practice environment in accounting for the rest remains unchanged and consistent with the baseline analysis of the paper.

C.4 Heterogeneous Treatment Effects in Patient/Physician Event Studies

Recent literature on two-way fixed effect regressions have shown that in the presence of staggered treatment and heterogeneous treatment effects, event study coefficients may recover a weighted average of these heterogeneous effects that involve negative weights (see e.g. Sun and Abraham 2021). We therefore explored the sensitivity of our event study analyses to allowing for heterogeneous effects by the timing of move, and found that this does not affect our results.

Recall that for the patient event study, we estimated the equation

$$y_{it} = \alpha_i^P + \tau_t^P + \theta^P_{\mathrm{r}(i,t)}\Delta_i^P + x_{it}'\beta^P + \rho^P_{\mathrm{r}(i,t)} + \eta_{it}^P \tag{46}$$

where y_{it} represents log patient-year utilization, Δ_i^P is the difference in average y_{it} between patient

i's destination and origin HRR (equal to zero for non-movers), and x_{it} contains indicators for 5-year age bins. Also, α_i^P, τ_t^P, and $\rho_{\mathrm{r}(i,t)}^P$ denote fixed effects for patients, calendar years, and years relative to the move, respectively. We observe movers throughout our sample period. As a result, for each move year $t^M \in \{1999, 2000, ..., 2013\}$, we can obtain a set of move-year specific event study coefficients $\theta_{\mathrm{r}(i,t)}^{P,t^M}$ by estimating the equation

$$y_{it} = \alpha_i^{P,t^M} + \tau_t^{P,t^M} + \theta_{\mathrm{r}(i,t)}^{P,t^M} \Delta_i^P + x_{it}' \beta^{P,t^M} + \eta_{it}^{P,t^M} \tag{47}$$

on the sample of patients who move in year t^M and non-movers only. Note that unlike equation (46), equation (47) does not contain relative year fixed effects ($\rho_{\mathrm{r}(i,t)}^P$). This is because all movers move in year t^M, meaning that relative years and calendar years are collinear.

By stacking all of the samples for each move year and interacting each covariate in equation (47) with a sample indicator, we can estimate the equation

$$y_{it} = \sum_{t^M=1999}^{2013} \check{\alpha}_i^P \times I_{t^M(i)} + \sum_{t^M=1999}^{2013} \check{\tau}_t^P \times I_{t^M(i)} + \check{\theta}_{\mathrm{r}(i,t)}^P \Delta_i^P + x_{it}' \check{\beta}^P + \mu_{t^M(i)}^P + \check{\eta}_{it}^P \tag{48}$$

where $I_{t^M(i)}$ is an indicator for whether patient i belongs to the sample used to estimate equation (47) for calendar year t^M, and $\mu_{t^M(i)}^P$ denotes a set of sample indicators. This guarantees that for each relative year r, the event study coefficient $\check{\theta}_{\mathrm{r}(i,t)}^P$ will be a convex, regression-weighted average of the coefficients $\theta_{\mathrm{r}(i,t)}^{P,t^M}$ in equation (47) estimated using the individual move year samples.

In practice, this exercise is extremely computationally intensive. Since there are over 15 million patient-year observations belonging to non-movers and 15 separate move years, stacking all 15 samples results in a dataset with over 200 million observations. Furthermore, with 3 million patients, the interaction terms in equation (48) result in an extremely large number of covariates. To alleviate this issue, we remove non-moving patients from the sample used to estimate equation (48).[51] Since there are movers in every move year, the original equation (46) is identified off of the sample of moving patients only. As a result, we can exploit this feature of the data so that equation (48) is identified as well. Panel A of Figure A9 displays the point estimates and 95% confidence intervals for three sets of event study coefficients: (1) our original estimates of $\theta_{\mathrm{r}(i,t)}^P$ in equation (46), which we presented in Figure 3, (2) estimates of $\theta_{\mathrm{r}(i,t)}^P$ in equation (46), restricting the sample to consist of moving patients only, and (3) estimates of $\check{\theta}_{\mathrm{r}(i,t)}^P$ in equation (48) using the sample of moving patients only. All three sets of coefficients are similar for each year relative to the move. In particular, including the sample of non-movers leads to almost no difference in the estimates of

[51] Note that this implies that $\mu_{t^M(i)}^P$ is equivalently interpreted as a set of move-year fixed effects.

equation (46), suggesting that our estimates of $\theta^{P,t^M}_{\mathrm{r}(i,t)}$ would have been similar had we been able to stack the non-mover sample as well. This indicates that our patient event study results are robust to the possibility of negatively-weighted heterogeneous treatment effects across move years.

On the physician side, we follow an extremely similar approach. Recall that we estimated the equation

$$y_{dt} - \ln N_{dt} = \alpha^D_d + \tau^D_t + \theta^D_{\mathrm{r}(d,t)}\Delta^D_d + \rho^D_{\mathrm{r}(d,t)} + \eta^D_{dt} \tag{49}$$

where $y_{dt} - \ln N_{dt}$ represents physician d's log utilization per patient, and Δ^D_d is the difference in average $y_{dt} - \ln N_{dt}$ between physician d's destination and origin HRR (equal to zero for non-movers). As before, α^D_d, τ^D_t, and $\rho^D_{\mathrm{r}(d,t)}$ denote fixed effects for physicians, calendar years, and years relative to the move, respectively. Panel B of Figure A9 presents estimates of $\theta^D_{\mathrm{r}(d,t)}$ in equation (49) with and without including non-movers in the sample. The figure also presents estimates of $\check{\theta}^D_{\mathrm{r}(d,t)}$ in our stacked event study

$$y_{dt} - \ln N_{dt} = \sum_{t^M=1999}^{2013} \check{\delta}^D_d \times I_{t^M(d)} + \sum_{t^M=1999}^{2013} \check{\tau}^D_t \times I_{t^M(d)} + \check{\theta}^D_{\mathrm{r}(d,t)}\Delta^D_d + \mu^D_{t^M(d)} + \check{\eta}^D_{dt} \tag{50}$$

where $I_{t^M(d)}$ is an indicator for whether physician d moves in calendar year t^M, and $\mu^D_{t^M(d)}$ denotes a set of move year fixed effects.[52] Again, non-movers are excluded to make estimation feasible. Since all three sets of coefficients are nearly identical, this implies that our physician event study is robust to heterogeneous treatment effects as well.

C.5 Reweighted Physician Event Study Robustness Check

The by-age physician event study results in Panel B of Figure A3 indicate that if there is heterogeneity in age between physician movers and non-movers—which Table A3 suggests is the case—our event study estimates may not generalize to the broader population of physicians. In this appendix, we check robustness of our doctor event study results to this possibility by re-weighting physician movers so that greater weight is placed on movers who resemble non-movers on observables. We construct weights based on four sets of observables: (1) indicators for physician gender, (2) indicators for 5-year bins of each physician's age at move, (3) indicators for physician specialty, and (4) all three sets of indicators combined. Specifically, we run a probit regression of an indicator

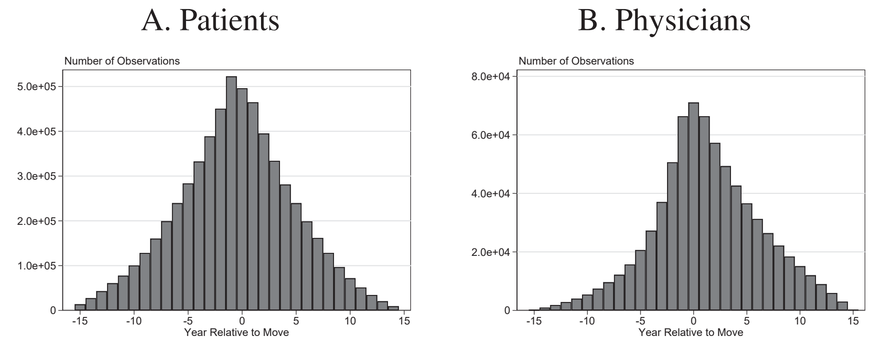
[52] Note that our algorithm for detecting moving physicians technically allows a very small number of physicians to move in 1998, the first year of our data. However, we exclude these physicians from this analysis since relative year -1 cannot be normalized to zero for such movers.

for being a mover on each of these four sets of observables.[53] We then use the resulting coefficients to predict a probability that each physician is a mover. We then re-estimate equation (3), weighting movers by the inverse of these predicted probabilities and non-movers by 1. The resulting event study coefficients are displayed in Figure A4, and the averages of the five post-move coefficients under each weighting scheme are displayed in Table A3. The results indicate that the size of the event study jump remains at about 0.5 with all four weighting schemes. This is evidence that our results are robust to movers being unrepresentative of the general physician population.

[53]To ensure that the predicted probabilities of being a mover are not too small, we bin all physicians who move before age 30 into a single indicator and all physicians who move after age 60 into a separate indicator. Furthermore, we bin together all specialty indicators for which the share of physicians in that specialty who move is below 6%. These specialties are geriatric psychiatry, the non-physician specialty, speech language pathology, oral surgery, registered dietician, the missing specialty, licensed clinical social worker, audiology, maxillofacial surgery, chiropractics, certified nurse midwife, clinical psychology, anesthesiologist assistant, certified registered nurse anesthetist, occupational therapy, psychology, nuclear medicine, physical therapy, certified clinical nurse specialist, diagnostic radiology, addiction medicine, preventative medicine, optometry, pediatric medicine, allergy immunology, pathology, anesthesiology, plastic and reconstructive surgery, interventional radiology, podiatry, neuropsychiatry, psychiatry, obstetrics/gynecology, nurse practitioners, and sleep medicine.

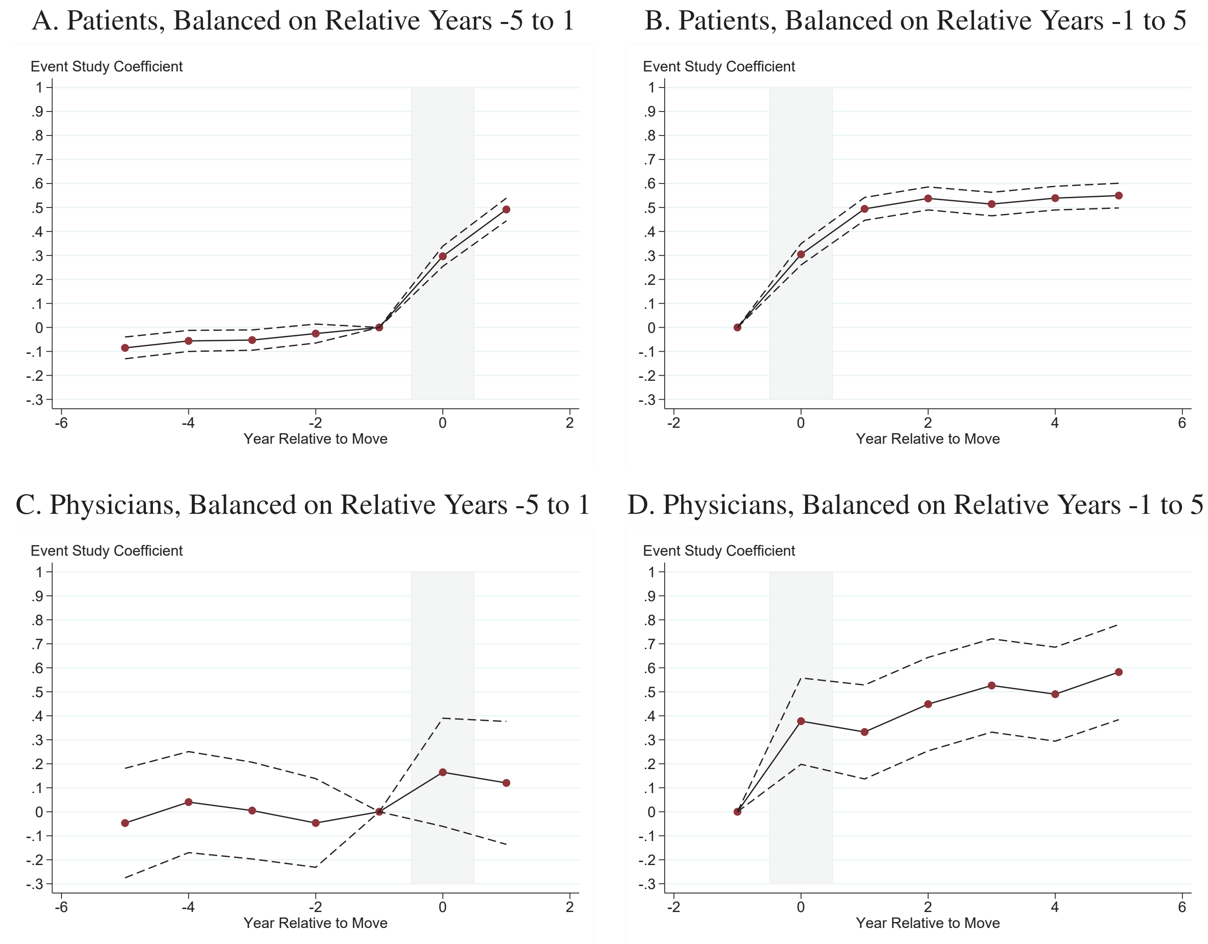
D Appendix Figures and Tables

Figure A1: Distribution of Mover Relative Years



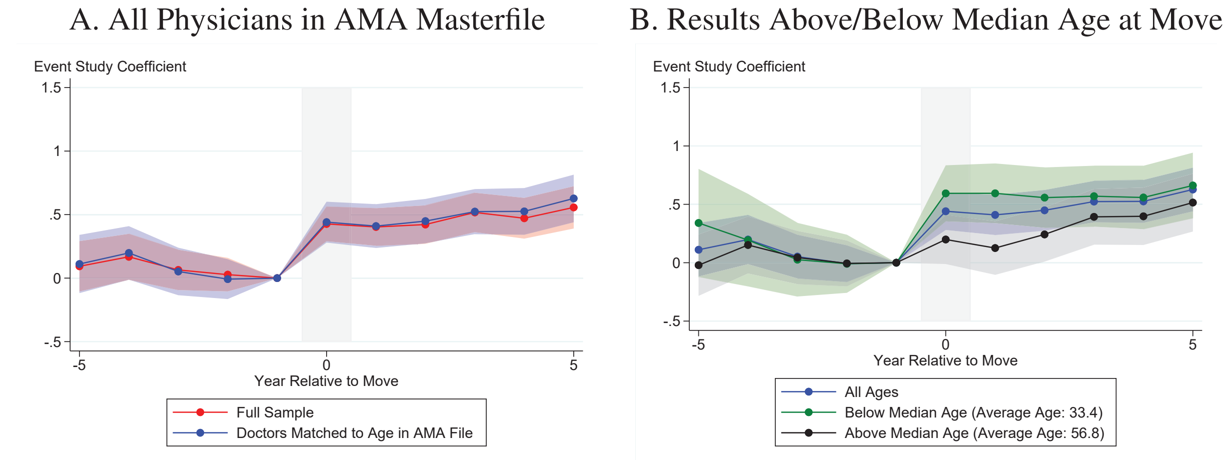
Notes: Panels A and B of this figure show the distributions of relative years for which we observe moving patients and physicians, respectively. The sample size is 6,011,508 patient-years (Panel A) and 728,907 physician-years (Panel B).

Figure A2: Patient and Physician Mover Event Studies (Balanced Panels)



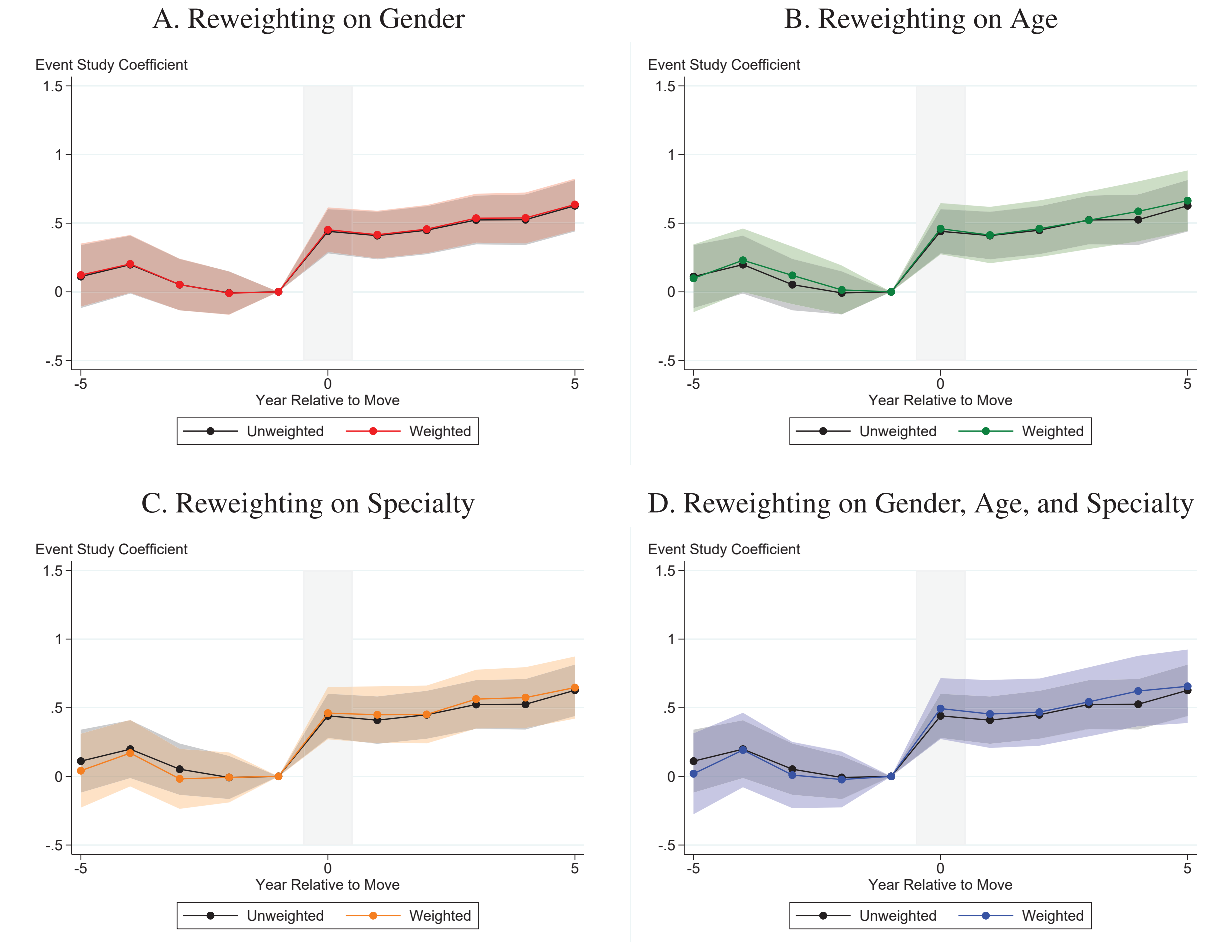
Notes: This figure shows the estimated θ_r^P (Panels A and B) and θ_r^D (Panels C and D) coefficients in equations (2) and (3) for patient and doctor movers, respectively. The coefficients for relative year -1 are normalized to 0. In Panels A and C, only patients and physicians who are observed in every relative year between -5 and 1 are included in the sample if they are movers; all non-movers are included. Similarly, the sample used in Panels B and D only contain mover patients and physicians who are observed in every relative year between -1 and 5, as well as all non-movers. Observations outside of these relative years are dropped. The dependent variable in Panels A and B is log annual patient utilization, and the control vector includes indicator variables for five-year age bins and relative-year main effects for movers. The dependent variable in Panels C and D is log annual physician utilization per patient, and the control vector includes relative-year main effects for movers. Dashed lines indicate upper and lower bounds of the pointwise 95 percent confidence intervals, clustered at the person (i.e. patient or physician) level. The sample size is 18,941,719 patient-years in Panel A, 18,945,198 patient-years in Panel B, 23,176,816 physician-years in Panel C, and 23,274,858 physician-years in Panel D.

Figure A3: Doctor Event Study by Age at Move



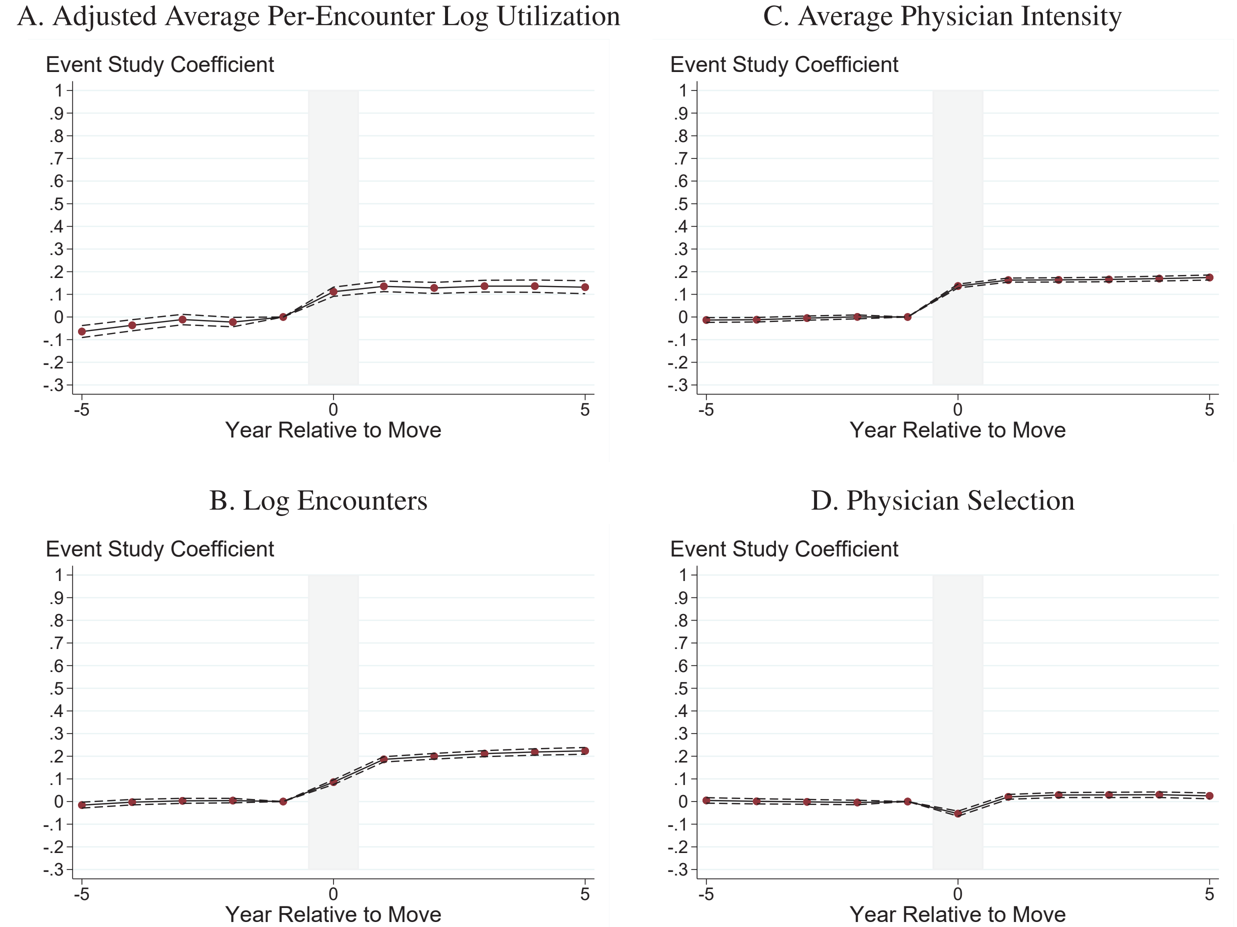
Notes: This figure displays estimates of θ_r^D in equation (3) computed on various samples. Panel A shows that the event study results are very similar when estimated on the full sample of physicians (N = 23,788,172 physician-years) and on the sample of physicians who can be matched to ages in the AMA Masterfile (N = 16,277,075 physician years). Panel B compares the latter result to the coefficients estimated using the sample of movers who move when they are below the median age at move (N = 15,928,782 physician years) and the sample of movers who move when they are above the median age at move (N = 16,055,616 physician-years). All three samples contain all non-movers who are present in the AMA Masterfile. The coefficients for relative year -1 are normalized to 0. Observations before and including relative year -6 are binned into a single indicator, as are all observations in relative year 6 and beyond; the coefficients on these indicators are not plotted here. The dependent variable is log annual physician spending per patient and the control vector includes relative-year main effects for movers. The shaded regions indicate upper and lower bounds of 95 percent confidence intervals computed using standard errors clustered at the physician level.

Figure A4: Reweighted Doctor Event Studies



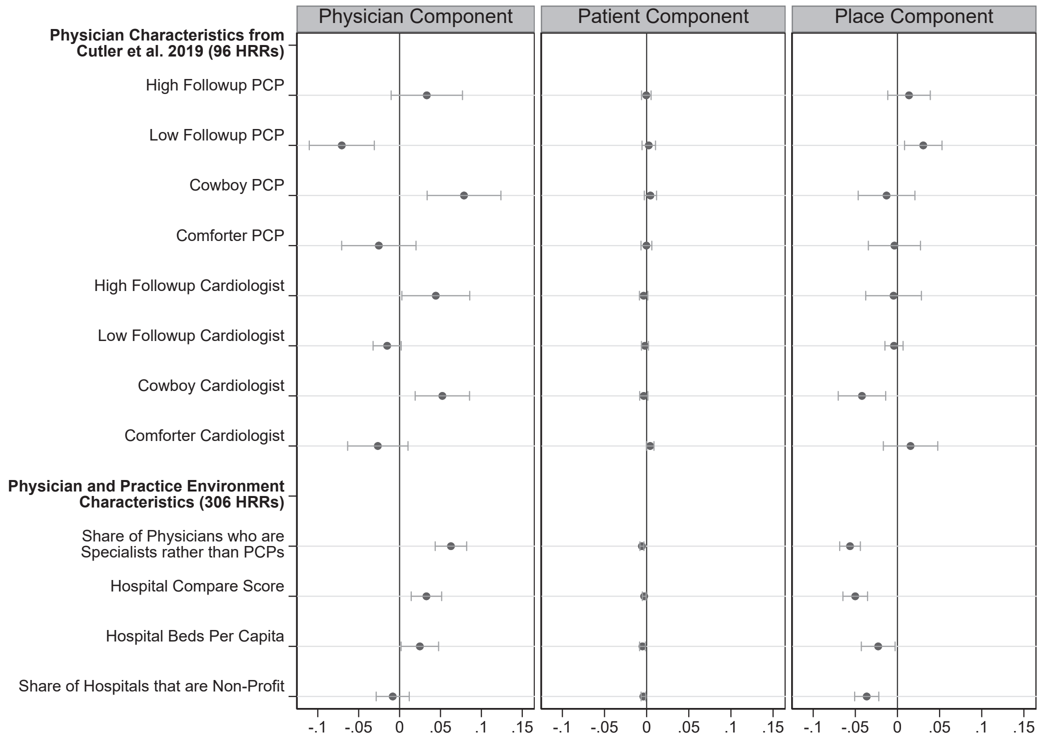
Notes: This figure displays estimates of θ_r^D in equation (3), where movers are weighted to resemble non-movers along various dimensions (for more detail on the reweighting scheme, see Appendix Section C.5). Panel A displays results reweighting movers so that they resemble nonmovers based on gender. Panel B presents results reweighting movers so that they resemble non-movers based on their age at move. Panel C presents results reweighting movers so that they resemble non-movers based on specialties. Panel D presents results reweighting movers so that they resemble non-movers along all three dimensions. The coefficients for relative year -1 are normalized to 0. Observations before and including relative year -6 are binned into a single indicator, as are all observations in relative year 6 and beyond; the coefficients on these indicators are not plotted here. The dependent variable is log annual physician spending per patient and the control vector includes relative-year main effects for movers. The shaded regions indicate upper and lower bounds of 95 percent confidence intervals computed using standard errors clustered at the physician level. The sample is all physicians matched to the AMA Masterfile (N = 16,277,075 physician-years).

Figure A5: Mover Event Study Decomposition



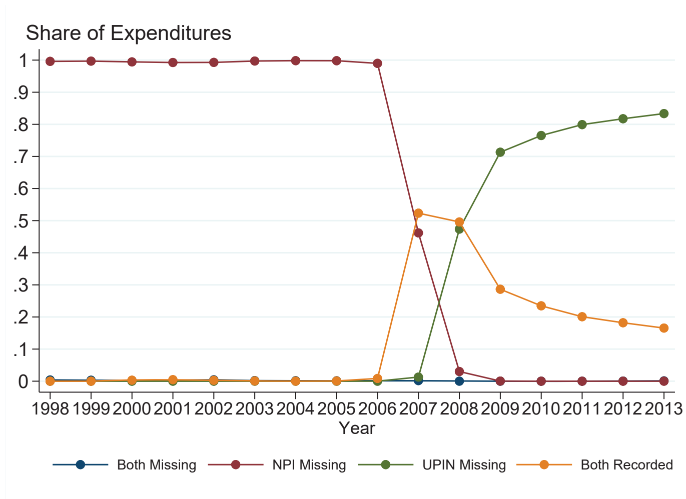
Notes: This figure shows the coefficients θ_r estimated from the enriched event study decomposition described in Section B.6 for patient movers. The coefficient for relative year -1 is normalized to 0. The dependent variable is adjusted annual utilization $y_{it} - (\ln N_{it} + \overline{\delta}_{it} + \sigma_{it})$ in Panel A, the log number of encounters $\log N_{it}$ in Panel B, average physician component $\bar{\delta}_{it}$ in Panel C, and average physician-patient selection σ_{it} in Panel D; x_{it} consists of indicator variables for five-year age bins (Panel A) and relative-year effects (Panel A and Panel B). The dashed lines are upper and lower bounds of the pointwise 95 percent confidence intervals computed using standard errors clustered at the patient level. Observations six years before the move and six years after the move are binned into separate indicators. The sample size is (N = 23,672,671 patient-years).

Figure A6: HRR-Level Correlates With Physician, Patient, and Place Components



Notes: This figure plots bivariate, HRR-level regression coefficients from a regression of the average physician, patient, and place component in each HRR ($\bar{\delta}_j$, $\bar{\alpha}_j$, and γ_j as defined in Section 4.2, respectively) against various HRR-level covariates, along with 95% confidence intervals constructed using heteroskedasticity-robust standard errors. All covariates are standardized to have mean 0 and standard deviation 1. The first eight measures are computed on a sample of 96 HRRs for which physicians were surveyed in Cutler et al. (2019); these regressions are weighted by the number of PCPs surveyed for the PCP measures and the number of cardiologists surveyed for the cardiologist measures. For the last five measures, we use the sample of all 306 HRRs, and the regressions are weighted by the number of Medicare patients we observe throughout the entire sample period (1998-2013). Hospital Compare Score approximates hospital quality using timely and effective care measures publicly reported by CMS. Hospital beds per capita counts hospital beds per thousand residents. Non-profit hospitals is the percent of hospitals that are non-profit. More detail on the construction of these variables can be found in Finkelstein et al. (2016).

Figure A7: Share of Utilization by Year and Missing Identifier Status

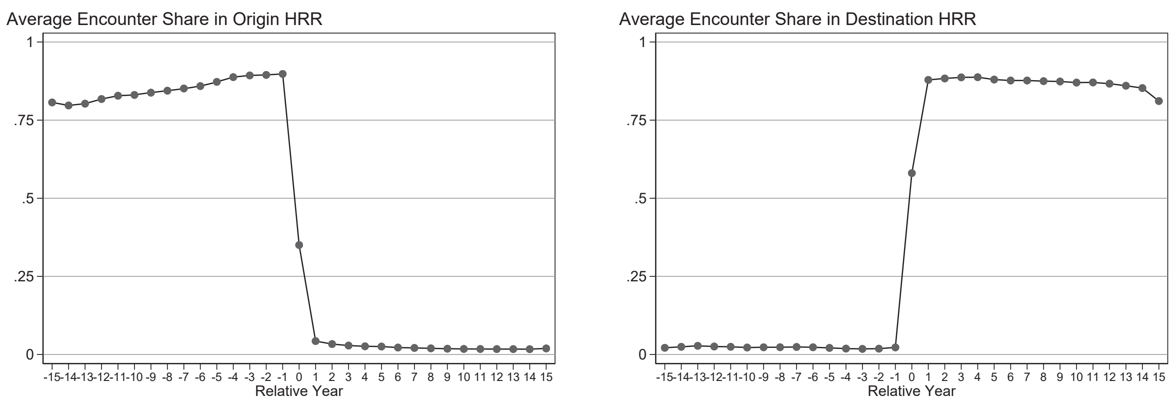


Notes: This figure shows the share of utilization by UPIN and NPI missing status between 1998 and 2013. The utilization sample reflects the baseline encounter sample in the 20-percent random share of Medicare claims, before applying any patient restrictions.

Figure A8: Average Mover Physician Encounter Shares, by Relative Year and HRR Type

A. Origin HRR

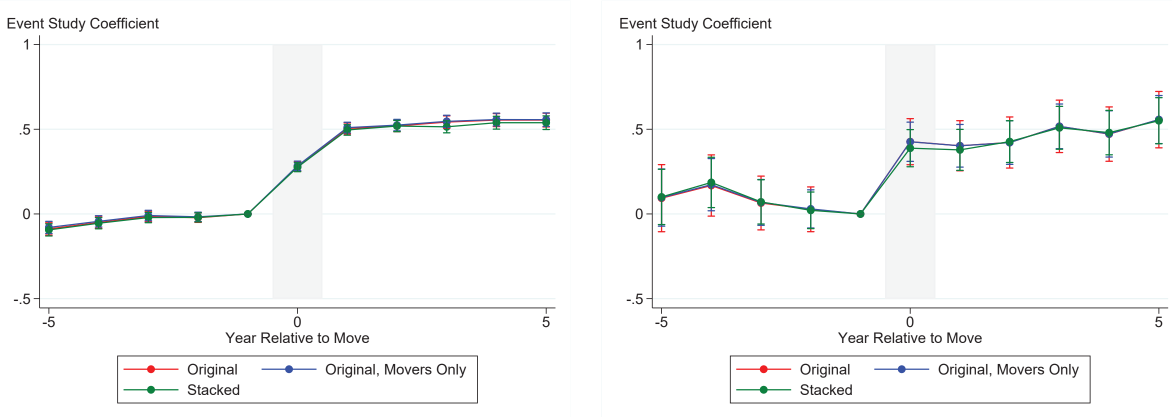
B. Destination HRR



Notes: This figure displays the average physician-year share of encounters in the origin HRR (Panel A) and destination HRR (Panel B), by relative year for mover physicians. Movers are assigned according to the algorithm described in Appendix Section A. Move years are assigned relative year 0, and these shares are computed on a sample of 74,934 movers.

Figure A9: Stacked Patient and Doctor Mover Event Studies

A. Patients B. Physicians



Notes: Panel A of this figure shows estimates of θ_r^P in equation (2) with and without non-moving patients in the sample, as well as estimates of $\check{\theta}_r^P$ in equation (48). Likewise, Panel B shows estimates of θ_r^D in equation (3) estimated with and without non-moving physicians, as well as estimates of $\check{\theta}_r^D$ in equation (50). The coefficients for relative year -1 are normalized to 0. Observations before and including relative year -6 are binned into a single indicator, as are all observations in relative year 6 and beyond; the coefficients on these indicators are not plotted here. Vertical bars denote upper and lower bounds of the pointwise 95 percent confidence intervals clustered at the individual (patient or physician) level. In Panel A, the sample size is N = 23,663,477 patient-years for the original event study and N = 6,011,508 patient-years for the stacked/movers only event studies. In Panel B, the sample size is N = 23,788,172 physician-years for the original event study, 728,907 physician-years for the movers only event study, and 696,277 physician-years for the stacked event study.

Table A1: Utilization Summary Statistics by BETOS Code Group

A. Carrier claim spending share

BETOS Code Group	Spending Share
Evaluation and Management	39.5%
Procedures	28.9%
Imaging	12.0%
Other	10.5%
Testing	8.7%
Durable Medical Equipment	0.3%
Exceptions/Unclassified	0.1%
Total	100%

B. Referring/performing physician match statistics (percent of claim lines)

	Performing and Referring Match	Different Referring Physician	Only Referring Physician Missing	Only Performing Physician Missing	Both Physicians Missing	Total
Evaluation and Management	37.9%	38.5%	21.9%	1.2%	0.5%	100%
Procedures	23.9%	58.1%	15.1%	2.2%	0.8%	100%
Imaging	14.6%	83.0%	0.3%	2.0%	0.1%	100%
Other	39.0%	26.3%	32.1%	1.3%	1.4%	100%
Testing	25.0%	69.7%	0.1%	5.1%	0.0%	100%
Durable Medical Equipment	19.9%	61.7%	14.4%	3.3%	0.7%	100%
Exceptions/Unclassified	26.4%	53.5%	8.5%	10.0%	1.6%	100%

Notes: BETOS code groups are defined as in https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/downloads/BETOSDescCodes.pdf. Ambulance services are excluded from the calculations. Columns in Panel B are mutually exclusive; "Performing and Referring Match" and "Different Referring Physician" only contain non-missing matches/mismatches. The sample is all carrier claims between 1998 and 2013. Note that claims can be further subdivided into claim lines. For example, in a claim for visiting a doctor's office, two separate procedures (e.g. a blood draw and a vaccination) would constitute distinct claim lines. Each claim line may have its own BETOS code; as a result, our statistics in Panel B cover the share of claim lines rather than the share of claims.

Table A2: Distribution of Physician Specialties

Specialty Group	Non-movers		Movers	
	Share of Utilization	Share of Physicians	Share of Utilization	Share of Physicians
Primary Care	36.21%	15.50%	42.61%	33.19%
Cardiology	9.42%	1.69%	9.23%	3.08%
Orthopedic Surgery	6.53%	1.75%	5.93%	2.87%
General Surgery	5.02%	1.84%	5.03%	3.19%
Ophthalmology	3.90%	1.45%	2.52%	2.07%
Nephrology	3.10%	0.54%	2.35%	0.90%
Hematology/Oncology	3.01%	0.55%	2.70%	0.89%
Podiatry	2.62%	1.20%	0.41%	1.28%
Urology	2.41%	0.76%	1.96%	1.13%
Pulmonary Disease	2.24%	0.66%	2.44%	1.33%
Emergency Medicine	2.18%	2.59%	4.64%	10.05%
Gastroenterology	1.68%	0.87%	1.33%	1.40%
Thoracic Surgery	1.57%	0.20%	1.30%	0.26%
Physical Medicine and Rehabilitation	1.39%	0.58%	1.94%	1.09%
Neurosurgery	1.36%	0.37%	1.32%	0.56%
Cardiac Surgery	1.33%	0.13%	1.81%	0.25%
Missing Specialty	1.24%	18.80%	0.11%	1.74%
Dermatology	1.15%	0.84%	0.77%	1.20%
Medical Oncology	1.13%	0.20%	1.00%	0.28%
Vascular Surgery	1.08%	0.18%	1.46%	0.40%
Neurology	1.07%	1.00%	1.10%	1.61%
Psychiatry	0.83%	2.47%	0.58%	2.41%
Obstetrics/gynecology	0.81%	2.70%	0.55%	2.94%
Radiation Oncology	0.77%	0.29%	0.89%	0.55%
Otolaryngology	0.76%	0.73%	0.64%	1.10%
Optometry	0.61%	2.58%	0.22%	1.58%
Nurse Practitioner	0.59%	5.78%	0.46%	5.20%
Rheumatology	0.58%	0.30%	0.41%	0.48%
Physician Assistant	0.44%	4.20%	0.52%	6.18%
Chiropractic	0.39%	4.22%	0.09%	1.07%
Anesthesiology	0.38%	2.65%	0.26%	1.30%
Endocrinology	0.36%	0.35%	0.28%	0.61%
Infectious Disease	0.36%	0.36%	0.33%	0.53%
Diagnostic Radiology	0.33%	2.11%	0.26%	0.88%
Geriatric Medicine	0.31%	0.12%	0.45%	0.32%
Critical Care	0.29%	0.13%	0.30%	0.28%
Non-physician	0.29%	4.52%	0.01%	0.12%
Plastic and Reconstructive Surgery	0.28%	0.45%	0.15%	0.35%
Colorectal Surgery	0.28%	0.09%	0.25%	0.14%
Hematology	0.19%	0.05%	0.21%	0.07%
Gynecological/Oncology	0.18%	0.06%	0.15%	0.08%
Pathology	0.13%	0.98%	0.06%	0.56%
Surgical Oncology	0.13%	0.05%	0.12%	0.07%
Clinical Psychology	0.13%	1.84%	0.03%	0.38%
Interventional Pain Management	0.11%	0.09%	0.15%	0.18%
Pediatric Medicine	0.10%	0.81%	0.10%	0.26%
Allergy Immunology	0.10%	0.27%	0.05%	0.20%
Hand Surgery	0.09%	0.07%	0.08%	0.12%
Physical Therapy	0.08%	3.91%	0.03%	1.81%
Licensed Clinical Social Worker	0.08%	1.91%	0.01%	0.26%
Pain Management	0.05%	0.06%	0.07%	0.12%
Interventional Radiology	0.05%	0.08%	0.04%	0.06%
Cardiac Electrophysiology	0.05%	0.02%	0.10%	0.07%
Osteopathic Manipulative Therapy	0.04%	0.06%	0.08%	0.10%
Certified Registered Nurse Anesthetist	0.04%	1.77%	0.01%	0.41%
Peripheral Vascular Disease	0.04%	0.01%	0.05%	0.01%
Oral Surgery	0.02%	0.40%	0.00%	0.04%
Certified Clinical Nurse Specialist	0.02%	0.21%	0.01%	0.11%
Maxillofacial Surgery	0.02%	0.14%	0.00%	0.04%
Nuclear Medicine	0.01%	0.05%	0.00%	0.02%
Preventative Medicine	0.01%	0.04%	0.01%	0.02%
Occupational Therapy	0.01%	0.49%	0.00%	0.11%
Neuropsychiatry	0.01%	0.01%	0.00%	0.01%
Addiction Medicine	0.01%	0.01%	0.00%	0.01%
Audiology	0.00%	0.42%	0.00%	0.05%
Sports Medicine	0.00%	0.01%	0.00%	0.02%
Psychology	0.00%	0.04%	0.00%	0.01%
Certified Nurse Midwife	0.00%	0.14%	0.00%	0.03%
Speech Language Pathology	0.00%	0.05%	0.00%	0.00%
Anesthesiologist Assistant	0.00%	0.03%	0.00%	0.01%
Registered Dietitian	0.00%	0.14%	0.00%	0.01%
Geriatric Psychiatry	0.00%	0.00%	0.00%	0.00%
Sleep Medicine	0.00%	0.00%	0.00%	0.00%

Notes: This table shows the share of utilization and physicians that belong to each specialty for non-movers (columns 1 and 3) and movers (columns 2 and 4). Physicians are assigned to specialties based on the HCFA code that corresponds to the plurality of their claims; see Appendix Section A.2 for details.

Table A3: Average of Reweighted Post-Period Doctor Event Study Coefficients

Variables Used to Create Weights	Average of Post-Period Coefficients
None	0.507
	(0.080)
Gender	0.517
	(0.081)
Five-Year Age Bins	0.529
	(0.094)
Specialties	0.537
	(0.094)
Gender, Age Bins, and Specialty	0.549
	(0.110)

Notes: This table displays results from re-estimating the doctor event study (equation (3)), reweighting moving physicians so that they are better matched to non-movers on observables based on the procedure described in Appendix section C.5. The first column gives the observables used to construct each set of weights, while the second column displays the value and standard error (in parentheses) of the estimates of $\frac{\sum_{r=1}^{5} \theta_r^D}{5}$. The sample is all physician-years for physicians matched to the AMA Masterfile (N = 16,277,075 physician-years).

Table A4: Sample Summary Statistics (AMA Doctors)

	Non-movers	Movers
Share first observed residence:		
Northeast	0.23	0.20
South	0.34	0.37
Midwest	0.23	0.23
West	0.20	0.20
Annual utilization:		
Mean	\$1,075	\$1,308
S.D.	\$2,156	\$1,660
Annual number of encounters:		
Mean	194.87	238.63
S.D.	250.13	189.26
Doctor age in move year:		
Mean	49.13	41.50
S.D.	14.88	10.27
Doctor experience in move year:		
Mean	21.78	13.72
S.D.	15.14	9.76
Share female	0.26	0.30
Number of doctors	674,973	51,213

Notes: This table displays doctor summary statistics for the subsample of physician movers and non-movers that can be matched to the 2014 AMA Physician Masterfile. Share first observed residence, annual utilization, and annual number of encounters are defined as in Panel B of Table 1. Doctor age and doctor experience in the move year are derived by using the year of birth and year of graduation from medical school in the Masterfile. The move year for non-movers is randomly generated, where the "move" years for non-movers are sampled, so as to preserve the probability distribution of move years for movers. Share female is derived by using the gender indicator in the Masterfile. The sample matched to the 2014 AMA Masterfile has 726,000 physicians.

Table A5: Average Physician Fixed Effect By Specialty

Specialty	Rank	Mean	Standard Error	Share of Utilization	Share of Physicians
Cardiac Surgery	1	1.073	0.008	1.37%	0.13%
Thoracic Surgery	2	0.905	0.007	1.54%	0.21%
Neurosurgery	3	0.723	0.004	1.35%	0.38%
Hematology/Oncology	4	0.611	0.002	2.98%	0.57%
Hematology	5	0.599	0.009	0.19%	0.05%
Medical Oncology	6	0.588	0.004	1.12%	0.21%
Gynecological/Oncology	7	0.547	0.011	0.17%	0.06%
Physical Medicine and Rehabilitation	8	0.535	0.003	1.44%	0.60%
Surgical Oncology	9	0.495	0.011	0.13%	0.05%
Colorectal Surgery	10	0.477	0.006	0.28%	0.09%
General Surgery	11	0.475	0.002	5.02%	1.90%
Orthopedic Surgery	12	0.453	0.001	6.48%	1.80%
Nephrology	13	0.421	0.002	3.04%	0.56%
Licensed Clinical Surgical Worker	14	0.407	0.005	0.07%	1.83%
Geriatric Medicine	15	0.402	0.005	0.33%	0.13%
Interventional Pain Management	16	0.379	0.007	0.12%	0.10%
Radiation Oncology	17	0.357	0.005	0.78%	0.31%
Vascular Surgery	18	0.340	0.004	1.11%	0.19%
Pain Management	19	0.329	0.009	0.05%	0.07%
Rheumatology	20	0.311	0.003	0.57%	0.31%
Primary Care	21	0.287	0.001	36.73%	16.29%
Gastroenterology	22	0.279	0.001	1.65%	0.90%
Hand Surgery	23	0.242	0.006	0.09%	0.07%
Plastic and Reconstructive Surgery	24	0.222	0.004	0.27%	0.45%
Critical Care	25	0.215	0.005	0.29%	0.14%
Pulmonary Disease	26	0.211	0.002	2.25%	0.69%
Peripheral Vascular Disease	27	0.148	0.015	0.04%	0.01%
Osteopathic Manipulative Surgery	28	0.137	0.011	0.05%	0.06%
Endocrinology	29	0.133	0.003	0.36%	0.36%
Neurology	30	0.105	0.001	1.08%	1.02%
Neuropsychiatry	31	0.105	0.025	0.01%	0.01%
Psychologist	32	0.070	0.027	0.00%	0.03%
Psychiatry	33	0.068	0.002	0.81%	2.47%
Clinical Psychologist	34	0.031	0.004	0.12%	1.77%
Sports Medicine	35	0.027	0.037	0.00%	0.01%
Ophthalmology	36	0.019	0.001	3.79%	1.48%
Urology	37	0.013	0.001	2.37%	0.78%
Infectious Disease	38	0.004	0.003	0.35%	0.36%
Cardiology	39	-0.037	0.001	9.40%	1.75%
Geriatric Psychiatry	40	-0.051	0.072	0.00%	0.00%
Interventional Radiology	41	-0.124	0.015	0.05%	0.08%
Cardiac Electrophysiology	42	-0.187	0.016	0.05%	0.02%
Dermatology	43	-0.205	0.001	1.12%	0.85%
Sleep Medicine	44	-0.221	0.102	0.00%	0.00%
Obstetrics/Gynecology	45	-0.265	0.002	0.79%	2.71%
Speech Language Pathologist	46	-0.275	0.039	0.00%	0.05%
Allergy Immunology	47	-0.299	0.005	0.09%	0.27%
Emergency Medicine	48	-0.305	0.001	2.38%	2.92%
Otolaryngology	49	-0.335	0.002	0.75%	0.75%
Addiction Medicine	50	-0.339	0.055	0.00%	0.01%
Anesthesiology	51	-0.419	0.002	0.37%	2.59%
Certified Clinical Nurse Specialist	52	-0.445	0.007	0.02%	0.21%
Nuclear Medicine	53	-0.457	0.018	0.01%	0.05%
Chiropractic	54	-0.485	0.003	0.37%	4.08%
Podiatry	55	-0.501	0.001	2.44%	1.21%
Optometry	56	-0.545	0.001	0.58%	2.54%
Nurse Practitioner	57	-0.656	0.001	0.58%	5.75%
Maxillofacial Surgery	58	-0.656	0.011	0.01%	0.14%
Oral Surgery	59	-0.727	0.008	0.02%	0.39%
Physician Assistant	60	-0.752	0.001	0.45%	4.29%
Anesthesiologist Assistant	61	-0.778	0.027	0.00%	0.03%
Pediatric Medicine	62	-0.788	0.007	0.10%	0.78%
Certified Registered Nurse Anesthetist	63	-0.835	0.003	0.04%	1.71%
Certified Nurse Midwife	64	-0.931	0.015	0.00%	0.14%
Missing Specialty	65	-0.978	0.001	1.15%	18.04%
Occupational Therapist	66	-0.986	0.005	0.01%	0.47%
Physical Therapist	67	-1.034	0.002	0.08%	3.82%
Diagnostic Radiology	68	-1.053	0.003	0.32%	2.06%
Pathology	69	-1.320	0.001	0.13%	0.96%
Preventative Medicine	70	-1.407	0.024	0.01%	0.04%
Non-Physician Specialty	71	-2.076	0.003	0.27%	4.33%
Registered Dietitian	72	-2.161	0.029	0.00%	0.14%
Audiologist	73	-2.472	0.015	0.00%	0.40%

Notes: This table displays the average physician fixed effect, δ_d as defined in equation (5), by specialty. The means and standard errors are computed by weighting each physician by their number of encounters throughout the entire sample period (1998-2013).

Table A6: Geographic Variation Counterfactuals (with Standard Errors)

	Above/below median			Top/bottom 25%			Top/bottom 10%		
	Absolute Differ-ence	% decline (increment)	% decline (cumulative)	Absolute Differ-ence	% decline (increment)	% decline (cumulative)	Absolute Differ-ence	% decline (increment)	% decline (cumulative)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Observed (1)	0.253			0.405			0.594		
	(0.000)			(0.000)			(0.001)		
Patient-physician selection (2)	0.238	-6%	-6%	0.378	-7%	-7%	0.564	-5%	-5%
	(0.000)	(0%)	(0%)	(0.001)	(0%)	(0%)	(0.001)	(0%)	(0%)
Physicians	0.149	-35%	-41%	0.222	-39%	-45%	0.345	-37%	-42%
	(0.003)	(1%)	(1%)	(0.005)	(1%)	(1%)	(0.008)	(1%)	(1%)
Of which: within-specialty effects in utilization per encounter (3)	0.187	-20%	-26%	0.281	-24%	-31%	0.422	-24%	-29%
	(0.003)	(1%)	(1%)	(0.005)	(1%)	(1%)	(0.008)	(1%)	(1%)
Of which: between-specialty effects in utilization per encounter (4)	0.149	-15%	-41%	0.222	-14%	-45%	0.345	-13%	-42%
	(0.003)	(0%)	(1%)	(0.005)	(0%)	(1%)	(0.008)	(0%)	(1%)
Patients	0.032	-46%	-87%	0.044	-44%	-89%	0.081	-44%	-86%
	(0.004)	(1%)	(2%)	(0.006)	(1%)	(2%)	(0.010)	(1%)	(2%)
Of which: patient effects in utilization per encounter (5)	0.148	0%	-42%	0.223	0%	-45%	0.350	1%	-41%
	(0.004)	(1%)	(1%)	(0.005)	(1%)	(1%)	(0.009)	(1%)	(1%)
Of which: patient effects in # encounters (6)	0.032	-46%	-87%	0.044	-44%	-89%	0.081	-45%	-86%
	(0.004)	(1%)	(2%)	(0.006)	(1%)	(2%)	(0.010)	(1%)	(2%)
Practice Environment	0.000	-13%	-100%	0.000	-11%	-100%	0.000	-14%	-100%
	(0.000)	(2%)	(0%)	(0.000)	(2%)	(0%)	(0.000)	(2%)	(0%)
Of which: practice environment effects in utilization per encounter (7)	0.082	20%	-68%	0.126	20%	-69%	0.198	20%	-67%
	(0.002)	(1%)	(1%)	(0.004)	(1%)	(1%)	(0.006)	(2%)	(1%)
Of which: practice environment effects in # encounters (8)	0.000	-32%	-100%	0.000	-31%	-100%	0.000	-33%	-100%
	(0.000)	(1%)	(0%)	(0.000)	(1%)	(0%)	(0.000)	(1%)	(0%)

Notes: Refer to the notes in Table 5 for the estimation procedure. The observed difference between above/below median HRRs is in average log patient annual utilization. Relative to Table 5, we include the standard errors of our estimates. Standard errors (in parentheses) are clustered at the patient level and calculated using a Bayesian bootstrap with 50 repetitions. We bootstrap the encounter-level connected set used in the per-encounter utilization regression (equation (5)) and patient-year-level dataset used in the Poisson encounter model regression (equation (4)) separately. We combine the estimates from the bootstrapped-sample regressions to produce the counterfactuals above for each draw. The reported standard errors are the standard deviation of the resulting bootstrap estimates.

Table A7: Geographic Variation Counterfactuals (Separating PCPs and Non-PCPs)

	Above/below median			Top/bottom 25%			Top/bottom 10%		
	Absolute Difference (1)	% decline (increment) (2)	% decline (cumulative) (3)	Absolute Difference (4)	% decline (increment) (5)	% decline (cumulative) (6)	Absolute Difference (7)	% decline (increment) (8)	% decline (cumulative) (9)
Observed (1)	0.253			0.405			0.594		
Patient-physician selection (2)	0.238	-6%	-6%	0.379	-6%	-6%	0.564	-5%	-5%
Physicians	0.149	-35%	-41%	0.222	-39%	-45%	0.345	-37%	-42%
Of which: within-specialty effects in utilization per encounter (3)	0.183	-22%	-28%	0.276	-25%	-32%	0.419	-24%	-29%
Of which: between-specialist effects in utilization per encounter (4)	0.168	-6%	-33%	0.250	-7%	-38%	0.390	-5%	-34%
Of which: specialist vs PCP effects in utilization per encounter (5)	0.149	-8%	-41%	0.222	-7%	-45%	0.345	-8%	-42%
Patients	0.032	-46%	-87%	0.044	-44%	-89%	0.081	-44%	-86%
Of which: patient effects in utilization per encounter (6)	0.148	0%	-42%	0.223	0%	-45%	0.350	1%	-41%
Of which: patient effects in # encounters (7)	0.032	-46%	-87%	0.044	-44%	-89%	0.081	-45%	-86%
Practice Environment	0.000	-13%	-100%	0.000	-11%	-100%	0.000	-14%	-100%
Of which: practice environment effects in utilization per encounter (8)	0.082	20%	-68%	0.126	20%	-69%	0.198	20%	-67%
Of which: practice environment effects in # encounters (9)	0.000	-32%	-100%	0.000	-31%	-100%	0.000	-33%	-100%

Notes: This table is based on estimation of equation (5), equation (4), and the counterfactuals described in Section 5.2. Each set of three columns partitions HRRs into two groups based on percentiles of average log annual patient utilization. First, we report the observed difference in average log annual patient utilization between the two areas at the top of each panel (row 1). Each successive row reports this difference under a particular counterfactual, along with the incremental and cumulative percentage change relative to this baseline. Row (2) reports the counterfactual difference if there were no differential physician selection within regions. Row (3) reports the difference if additionally there were no variation in average physician intensity in healthcare within an encounter across regions, holding fixed the clinical specialty of the physician. Row (4) reports the difference if there were also no differential sorting of specialists (i.e. non-PCPs) and PCPs across regions. Row (5) reports the difference if there were additionally no sorting of specialties at all. Rows (6) and (7) report the difference if additionally there were no differential sorting of patients' demand for healthcare across regions, breaking this change into two separate sequential steps eliminating patient effects on the demand for care within an encounter and for healthcare encounters, respectively. The last two rows report the difference if additionally there were no variation in practice environment effects on healthcare utilization, breaking this change into two separate sequential steps eliminating practice environment effects on care within an encounter and number of encounters across regions, respectively. For details on how we define each counterfactual, see Appendix Section B.3. The sample is all encounters (159 million encounters of 3 million patients with 1.7 million physicians).

Table A8: Geographic Variation Counterfactuals Counting Practice Environment Encounter Margin Toward Physician Effect

	Above/below median		
	Absolute Difference (1)	% decline (increment) (2)	% decline (cumulative) (3)
Observed (1)	0.253		
Patient-physician selection (2)	0.238	-6%	-6%
Patients	0.122	-46%	-52%
Of which: patient effects in utilization per encounter (3)	0.237	0%	-6%
Of which: patient effects in # encounters (4)	0.122	-45%	-52%
Physicians	-0.050	-68%	-120%
Of which: within-specialty effects in utilization per encounter (5)	-0.016	-54%	-106%
Of which: between-specialty effects in utilization per encounter (6)	-0.050	-14%	-120%
Practice Environment (7)	0.000	20%	-100%

Notes: This table is based on estimation of equation (5), equation (4), and the counterfactuals described in Section 5.2. First, we report the observed difference in average log annual patient utilization between HRRs above and below the median (row 1). Each successive row reports this difference under a particular counterfactual, along with the incremental and cumulative percentage change relative to this baseline. Row (2) reports the counterfactual difference if there were no differential physician selection within regions. Rows (3) and (4) report the difference if additionally there were no differential sorting of patients' demand for healthcare across regions, breaking this change into two separate sequential steps eliminating patient effects on the demand for care within an encounter and for healthcare encounters respectively. Row (5) reports the difference if additionally there were no variation in average physician intensity in healthcare within an encounter across regions, holding fixed the clinical specialty of the physician. Row (6) reports the difference if there were also no differential sorting of clinical specialties across regions. Row (7) reports the difference if additionally there were no variation in practice environment effects on healthcare utilization. For details on how we define each counterfactual, see Appendix Section B.4. The sample is all encounters (159 million encounters of 3 million patients with 1.7 million physicians).

Table A9: Robustness Checks for Geographic Variation Counterfactuals

	Above/below median					
	Baseline	Primary HRR	Time-Varying Health	1998-2005	2006-2013	No Moves to/from AZ/CA/FL
	(1)	(2)	(3)	(4)	(5)	(6)
Observed difference (1)	0.253	0.206	0.253	0.251	0.242	0.246
Patient-physician selection (2)	-6%	-5%	-6%	-5%	-7%	-7%
Physicians	-35%	-33%	-35%	-43%	-62%	-41%
Of which: within-specialty effects in utilization per encounter (3)	-20%	-23%	-22%	-28%	-47%	-26%
Of which: between-specialty effects in utilization per encounter (4)	-15%	-10%	-13%	-15%	-15%	-15%
Patients	-46%	-38%	-60%	-51%	-59%	-46%
Of which: patient effects in utilization per encounter (5)	0%	-1%	-3%	-1%	-2%	-2%
Of which: patient effects in # encounters (6)	-46%	-37%	-57%	-49%	-57%	-44%
Practice Environment	-13%	-23%	1%	-1%	29%	-6%
Of which: practice environment effects in utilization per encounter (7)	20%	24%	22%	20%	50%	23%
Of which: practice environment effects in # encounters (8)	-32%	-47%	-21%	-21%	-21%	-29%

Notes: This table is based on estimation of equation (5), equation (4), and the counterfactuals described in Section 5.2. Each column of this table presents the percent decline in the difference in average log utilization for HRRs above and below the median utilization as we equalize components of the model. Column (1) displays the baseline increments, column (2) separates utilization from each physician's primary HRR and non-primary HRRs, decomposing $\bar{y}_j^p$ as defined in equation (45), and column (3) adds indicators for having each of 21 chronic conditions, as well as an additional indicator for having just entered the sample (in which case the chronic condition indicators are set to zero). Columns (4) and (5) restrict the sample to encounters occurring between 1998-2005 and 2006-2013, respectively. Column (6) removes patient and physician movers who moved to or from Arizona, California, or Florida. The sample size is 159 million encounters between 3 million patients and 1.7 million physicians in columns (1)-(3), 77 million encounters between 2.2 million patients and 1 million physicians in column (4), 82 million encounters between 2.2 million patients and 1.4 million physicians in column (5), and 124 million encounters between 2.4 million patients and 1.5 million physicians in column (6).

Table A10: Patient Sample Restrictions (1998-2013)

	Patients		Patient-Years		Encounters	
	N	Reduction (%)	N	Reduction (%)	N	Reduction (%)
	16,700,441		140,819,152		763,713,408	
Unique patients in the 20% Medicare claims sample	5,303,053	68	48,019,228	66	266,927,568	65
Exclude 75% random sample of non-mover patients						
Exclude patient-years where patients are:	4,722,835	3	42,270,648	4	234,264,928	4
Younger than 65 or older than 99						
Enrolled in Medicare Advantage	4,101,594	4	32,674,782	7	217,079,424	2
Not subscribed to Medicare Part A & B for all months in a year	3,741,866	2	29,681,748	2	204,799,712	2
Exclude multiple movers	3,466,888	2	26,950,282	2	184,376,992	3
Exclude movers with insufficient destination claim share	3,093,133	2	23,672,672	2	160,367,168	3
In connected set of observations	2,996,555	1	22,311,618	1	159,081,232	0

Notes: This table shows the impact of each of our restrictions on the sample size in terms of patients, patient-years, and encounters, respectively. In each case we show two columns: one ("N") displaying the sample size after applying the restriction and another ("Reduction (%)") displaying the marginal reduction in percentage terms due to the restriction. The encounter count includes patient-year zeros.

Table A11: Claims-Based Crosswalk Match Comparison with NBER Crosswalk

	A. Finalized Crosswalk Matched NPIs	
	Number of NPIs (thousands)	Share of Total Utilization (2009)
Matched in Both Crosswalks	517	61.1%
Only Matched in NBER Crosswalk	88	10.0%
Only Matched in Claims-Based Crosswalk	316	15.7%
Total	**921**	**86.7%**
	B. Non-Matched NPIs	
	Number of NPIs (thousands)	Share of Total Utilization (2009)
Associated With Conflicts	13	1.2%
Unmatched NPIs	120	7.4%
Organization NPIs	41	1.1%
Post-transition NPIs	716	3.6%
Total	**890**	**13.3%**

Notes: This table compares all 1,810,474 NPIs observed in the 20 percent Medicare claims data between 1998 and 2013. The total 2009 utilization is computed from the baseline encounter sample, before applying patient restrictions. Conflicts represent cases in which an NPI is associated with more than one UPIN between crosswalks, or a UPIN is associated with more than one NPI. We consider NPIs to be unmatched if they are listed with at least one non-missing UPIN in the claims data, and we consider NPIs to have entered the sample after the transition ("post-transition NPIs") if they are not listed with any UPINs. Organization NPIs are determined from NPPES data entity type information.

Table A12: Finalized Crosswalk Match Comparision With AMA Data

	Number of NPIs (thousands)	Share of Total Utilization (2009)
Matched to the Same UPIN	562	79.8%
Only Matched in AMA Data	55	3.1%
Only Matched in Finalized Crosswalk	356	6.6%
Associated With Conflicts	3	0.3%
Total	**976**	**89.8%**

Notes: The sample is the 976,445 individual NPIs that were either matched by the finalized crosswalk or associated with a non-missing UPIN in the AMA Physician Masterfile. Conflicts represent cases in which an NPI is associated with more than one UPIN between crosswalks, or a UPIN is associated with more than one NPI. The total 2009 utilization is computed from the baseline encounter sample, before applying patient restrictions. The utilization sample includes organization NPIs, NPIs that entered the sample after the UPIN-NPI transition, and individual NPIs that were not matched by either crosswalk; these NPIs correspond to 9.2 percent of utilization.

Table A13: Example Crosswalk and Claims

UPIN	NPI	Physician ID
A	1	A_1
B	2	B_2
C	3	C_3

(a) Crosswalk Illustration

Year	UPIN	NPI	Preferred Identifier	Physician ID
2003	D		D	D_
2007	A		A	A_1
2007	A	1	1	A_1
2007		1	1	A_1
2010	A	2	2	B_2
2010	B	2	2	B_2
2010		3	3	C_3
2010		4	4	_4

(b) ID Assignment to Claims Example

Notes: this table presents examples of how we construct physician IDs from UPINs and NPIs (Panel A) as well as how we match UPIN/NPI pairs in the claims data to these IDs (Panel B). For more detail on this process, see Appendix Section A.2.

Table A14: Physician Sample Comparison

Statistic	Estimates from our 2013 sample (1)	Published reports from 2013 Medicare data (2)
Mean annual spending per physician	$219,088	$209,500
Total annual physician spending (billions)	$204	$257
Number of doctors	930,446	1,226,728

Notes: This table compares our sample summary statistics on physician spending to public information on Medicare spending patterns. Note that "spending" is equivalent to "utilization" in this context because it is aggregated across all patients in the entire country, meaning that our purging of geographic variation in administratively set prices to construct "utilization" is no longer relevant. In column (1), row 1 we show average physician spending for 2013. Average physician spending is the ratio of total spending and the total number of physicians in our sample in 2013. We show these separate components in rows 2 and 3 respectively, accounting for the 20 percent patient sampling in our data. We then show comparable estimates for the same three numbers coming from information published by the Medicare Payment Advisory Commission and the Centers for Medicare & Medicaid Services (CMS). Annual physician spending in column (2) is taken from Chart 1-13 in Medicare Payment Advisory Comission (2021), calculated as the sum of spending on inpatient hospitals, outpatient hospitals, and physician fees. The number of doctors in column (2) is taken from Table II.8 in U.S. Department of Health and Human Services (HHS) (2014). All values are for 2013. Non-physician practitioners account for 308,994 of the 1,226,728 providers.

Table A15: Split-Sample Standard Deviation and Correlation Matrices of Utilization Components

	HRR-Average of:	Standard Deviation	Correlation Matrix				
Encounter Model	Practice environment ($\tilde{\gamma}_j^N$)	0.336	1.000				
	Patients ($\tilde{\alpha}_{it}^N$)	0.318	0.580	1.000			
Per-Encounter Utilization Model	Practice environment ($\tilde{\gamma}_j$)	0.132	-0.536	-0.291	1.000		
	Patients ($\tilde{\alpha}_{it}$)	0.019	0.020	-0.174	-0.009	1.000	
	Physicians ($\tilde{\delta}_{dt}$)	0.131	0.464	0.392	-0.852	0.056	1.000

Notes: This table displays the adjusted correlation matrix between utilization components from the encounter model (rows 1-2, equation 4) and per-encounter utilization model (rows 3-5, equation 5) across split samples. Details of how these correlations are computed are reported in Appendix Section C.2.

Table A16: Utilization Counterfactuals (Rearranged Step Order)

	Physicians → Patients → Practice Environment	Physicians → Practice Environment → Patients	Patients → Physicians → Practice Environment	Patients → Practice Environment → Physicians	Practice Environment → Physicians → Patients	Practice Environment → Patients → Physicians
	(1)	(2)	(3)	(4)	(5)	(6)
Patient-Physician Selection ($\bar{y}_j^{(1)}$)	-6%	-6%	-6%	-6%	-6%	-6%
Physicians ($\bar{y}_j^{(2)}$, $\bar{y}_j^{(3)}$)	-35%	-35%	-16%	-36%	-36%	-36%
Patients ($\bar{y}_j^{(4)}$, $\bar{y}_j^{(5)}$)	-46%	-39%	-46%	-46%	-39%	-39%
Practice Environment ($\bar{y}_j^{(6)}$, $\bar{y}_j^{(7)}$)	-13%	-20%	-32%	-13%	-19%	-19%

Notes: This table shows the relative contribution of each factor in overall geographic utilization variation under different orderings of the utilization counterfactual steps. In all cases, we eliminate physician selection first. For simplicity, we combine patient and practice environment effects on the number of encounters and utilization per encounter into single patient and practice environment effects, respectively. We also combine within- and between-specialty physician effects into a single physician effect. Each row corresponds to a utilization factor listed in the row heading. Each column then shows the incremental contribution of said factor (in percentage terms) to overall geographic variation in utilization (defined as the difference in average patient log utilization for HRRs above and below the median) if we were to perform the counterfactual steps in the order indicated by the column heading. The sample is the baseline sample of all encounters (N = 159 million encounters).