SOURCES OF GEOGRAPHIC VARIATION IN HEALTH CARE

Author(s): Amy Finkelstein, Matthew Gentzkow and Heidi Williams

Source: *The Quarterly Journal of Economics*, November 2016, Vol. 131, No. 4 (November 2016), pp. 1681-1726

Published by: Oxford University Press

Stable URL: https://www.jstor.org/stable/10.2307/26372676

SOURCES OF GEOGRAPHIC VARIATION IN HEALTH CARE: EVIDENCE FROM PATIENT MIGRATION*

AMY FINKELSTEIN
MATTHEW GENTZKOW
HEIDI WILLIAMS

We study the drivers of geographic variation in U.S. health care utilization, using an empirical strategy that exploits migration of Medicare patients to separate the role of demand and supply factors. Our approach allows us to account for demand differences driven by both observable and unobservable patient characteristics. Within our sample of over-65 Medicare beneficiaries, we find that 40–50% of geographic variation in utilization is attributable to demand-side factors, including health and preferences, with the remainder due to place-specific supply factors. *JEL* Codes: H51, I1, I11.

I. INTRODUCTION

Health care utilization varies widely across the United States (Fisher et al. 2003a, 2003b). Adjusting for regional differences in age, sex, and race, health care spending for the average Medicare enrollee in Miami, FL, was $14,423 in 2010, but just $7,819 for the average enrollee in Minneapolis, MN. The average enrollee in McAllen, TX spent $13,648, compared with $8,714 in nearby and demographically similar El Paso, TX.[1] Similar geographic variation is observed in the frequency of specific treatments (Chandra, Cutler, and Song 2012) and in measures of total

*We are grateful to Daron Acemoglu, Raj Chetty, David Cutler, Rebecca Diamond, Joe Doyle, Liran Einav, Ben Handel, Nathan Hendren, Pat Kline, Matt Notowidigdo, Adam Sacarny, Jesse Shapiro, Jonathan Skinner, Doug Staiger, and seminar participants at the BU/Harvard/MIT health economics seminar, Dartmouth, the NBER Aging Meeting, the NYC health economics seminar, Stanford, the University of Chicago, the University of Maryland, MIT, Utah Winter Business Economics Conference, Wellesley, and the 60th Anniversary Congress of the Yrjo Jahnsson Foundation for comments; to Lizi Chen, Grant Graziani, Yunan Ji, Sara Kwasnick, Tamar Oostrom, Daniel Prinz, Daniel Salmon, and Tony Zhang for excellent research assistance; and to the National Institute on Aging (Finkelstein, R01-AG032449; P01-AG19783) and the National Science Foundation (Williams, 1151497; Gentzkow, 1260411) for financial support. Gentzkow also thanks the Neubauer Family Foundation and the Initiative on Global Markets at the University of Chicago Booth School of Business.

1. Authors' tabulations based on total Medicare Parts A and B reimbursements per enrollee, from Dartmouth Atlas of Health Care, http://www.dartmouthatlas.org/downloads/tables/pa_reimb_hrr_2010.xls.

The Quarterly Journal of Economics (2016), 1681–1726. doi:10.1093/qje/qjw023.
Advance Access publication on July 19, 2016.

health care utilization that adjust for regional variation in administratively set prices (Gottlieb et al. 2010). Higher area-level utilization is not generally correlated with better patient outcomes.[2]

Understanding what drives geographic variation in utilization has important implications for policy. If high-utilization areas like McAllen and Miami are different mainly because their doctors' incentives or beliefs lead them to order excessive treatments with low return, policies that change those incentives or beliefs could result in savings on the order of several percentage points of GDP (Congressional Budget Office 2008; Gawande 2009; Skinner 2011). On the other hand, if patients in high-utilization areas are simply sicker or prefer more intensive care, such policies could be ineffective or counterproductive.

In this article, we exploit patient migration to separate variation due to patient characteristics, such as health, or preferences from variation due to place-specific variables such as doctors' incentives and beliefs, endowments of physical or human capital, and hospital market structure. As a shorthand, we refer to the former as "demand" factors and the latter as "supply" factors.[3] To see the intuition for our approach, imagine a patient who moves from high-utilization Miami to low-utilization Minneapolis. If all of the utilization difference between these cities arises from supply-side differences like doctor incentives or beliefs, we would expect the migrant's utilization to drop immediately following the move, to a level similar to other patients of the low-utilization doctors in Minneapolis. If all of the utilization difference reflects the demand-side reality that residents of Miami are sicker, we would expect the migrant's utilization to remain constant after the move, at a level similar to the typical person in Miami. Where the observed utilization change falls

2. See Skinner (2011) for an extensive discussion. The Congressional Budget Office (2008) concludes that high-spending areas "tend to score no better and, in some cases, score worse than other areas do on process-based measures of quality and on some measures of health outcomes," and that more intensive treatment in high-spending areas "appear[s] to improve health outcomes for some types of patients, but worsen outcomes for others."

3. This corresponds to the usual definitions of demand and supply in most cases, but the correspondence is not perfect. For example, peer effects or social learning will generally be captured in our framework as a place-specific ("supply") factor, since the composition of peers can change when a patient moves, but it would be more natural to think of them as shifters of demand rather than supply.

between these two extremes identifies the relative importance of demand and supply factors.

We implement this strategy using claims data for a 20% sample of Medicare beneficiaries from 1998 to 2008. Our main outcome measure adjusts health care spending for geographic price differences to create a quantity measure of utilization, as in Gottlieb et al. (2010). We introduce a simple model of health care demand and supply, which implies that the log of a patient's annual health care utilization can be written as a combination of a patient fixed effect, a location fixed effect, and a vector of time-varying controls, including indicators for year relative to move for migrants. This specification allows for the possibility that migrants have systematically different utilization levels from nonmigrants and that these levels are correlated with the migrant's origin and destination regions. It also allows for arbitrary differences in utilization trends of migrants relative to nonmigrants. The key identifying assumption is that such differential trends do not vary systematically with the migrant's origin and destination.

We begin with an event-study analysis of changes in log utilization around moves. We observe a sharp change in the year of a move, equal to about half of the difference in average log utilization between the origin and destination. There is little systematic trend premove, and no systematic adjustment postmove. The on-impact effect is similar for moves from low-to-high and high-to-low utilization regions, and is roughly linear in the absolute value of the origin–destination difference in log utilization.

Our estimated model exploits this variation to infer that 47% of the difference in log utilization between above- and below-median areas is due to patient characteristics, with the remainder due to place-specific factors. The shares are similar for differences between the top and bottom quartiles, deciles, or ventiles. The share of the difference in log utilization due to patients is also similar when we isolate differences between the very highest utilization areas, such as McAllen or Miami, and the very lowest utilization areas, such as El Paso or Minneapolis. The results appear inconsistent with patient effects arising primarily from habit formation or persistence of treatments started premove, and instead point toward heterogeneity in health status or preferences that are fixed over the horizon of our data.

Our decomposition can be interpreted in terms of a counterfactual: by what share would the gap in utilization between areas

fall if patients were randomly reallocated between them. Importantly, this is a partial equilibrium experiment in that it holds fixed supply-side characteristics such as stocks of physical and human capital. In the long run, we would expect some of these characteristics to adjust endogenously to the change in patient demand (Chandra and Staiger 2007). If this led to convergence on the supply side, the long-run fall in geographic variation under the counterfactual would be greater than our short-run estimates would suggest.

We replicate our analysis for various components of total utilization. All measures show sharp changes in the year of a move, with magnitudes implying patient shares ranging from 9% to 71%. Consistent with intuition, we find large patient shares for outcomes where we might think patients have significant discretion—preventive care and emergency room visits, for example—and smaller patient shares for outcomes where we might think they have less—diagnostic tests, imaging tests, and inpatient care. We also find some suggestive evidence that the patient share may be lower at higher percentiles of the utilization distribution.

In the final section, we present evidence on the observable area-level correlates of our estimated place and patient effects. The potential correlates we consider include the number, quality, and organizational form of hospitals; survey-based measures of doctor beliefs about appropriate practice style and patient preferences over alternative practice styles; average patient demographics; and average patient health status. An important challenge arises with regard to the latter: because standard measures of underlying health status are derived from claims data, a given condition is more likely to be recorded in a high-intensity area, and standard measures of patient health therefore include a large component of systematic, place-specific measurement error (Song et al. 2010; Welch et al. 2011). To address this, we extend our mover-based strategy to estimate the place-specific measurement error component and derive corrected health measures purged of this measurement error.

The correlations are broadly consistent with intuition from our model and evidence from the existing literature. On the supply side, we find that the place component of utilization is particularly high in areas with a larger share of for-profit hospitals and a larger share of doctors who report a preference for aggressive care; the latter is consistent with recent literature emphasizing the importance of physician practice styles and

beliefs in driving geographic variation (e.g., Cutler et al. 2015). We also find that the place component is higher in areas where patients are sicker, which is consistent with past work arguing that physical and human capital are likely to adjust endogenously to patient demand (Chandra and Staiger 2007). On the demand side, we find that the patient component of utilization is higher where patients are sicker and of higher socioeconomic status, consistent with patient health status and patient preferences playing an important role. Were we to take the correlations between log utilization and corrected health measures as causal, they would imply that about a quarter of the geographic variation in log utilization (or equivalently, about half of our estimate of the patient share of this variation) may be explained by our corrected patient health measures; the remainder may reflect preferences or unmeasured health.

Studying Medicare patients is appealing due to the availability of high-quality, rich data on large numbers of beneficiaries, and the relatively uniform insurance environment. Medicare accounts for a significant share of total U.S. health spending: 20.5% as of 2011 (Moses et al. 2013).[4] Nevertheless, extrapolating our conclusions to other populations requires caution. Although regional variation in utilization appears to be the norm,[5] the relative importance of place and patient factors could differ in other settings. In private insurance markets, moreover, substantial cross-area differences in prices mean the correlates of area spending may differ substantially from the correlates of area utilization (Chernew et al. 2010; Dunn, Shapiro, and Liebman 2013; Cooper et al. 2015).

Our work contributes to a large existing literature seeking to separate the role of demand-side and supply-side factors in driving geographic variation in health care utilization. All of these studies infer the role of demand-side factors from the explanatory power of patient observables. Our main contribution is to develop

4. This number includes under-65 beneficiaries, who we exclude from our study. According to Neuman et al. (2015), beneficiaries under 65 accounted for 22% of spending in traditional Medicare in 2011.

5. Large regional differences have been documented in the US Veterans Affairs system (Ashton et al. 1999; Subramanian et al. 2002; Congressional Budget Office 2008), in private insurance markets (Baker, Fisher, and Wennberg 2008; Rettenmaier and Saving 2009; Chernew et al. 2010; Philipson et al. 2010; Dunn, Shapiro, and Liebman 2013; Cooper et al. 2015), and in other countries including the United Kingdom and Canada (McPherson et al. 1981).

a strategy that exploits migration to capture observed and unobserved patient characteristics.

Taken together, the evidence from the prior literature suggests three conclusions: (i) supply-side factors are a key driver of geographic variation; (ii) patient preferences and characteristics other than health status explain little variation; (iii) differences in health status may be important, but the evidence is inconclusive because of endogenous measurement error.[6] Our findings confirm that supply-side factors are important while revealing that patient preferences and health status together account for a large share of variation. Once we address the endogenous measurement issue with patient health, we find that roughly a quarter of the geographic variation in log health care utilization can potentially be attributed to observable patient health. Whether the remaining patient component reflects preferences or unmeasured health remains an open question.

Like past decompositions, ours is not sufficient to draw strong conclusions about the efficiency of observed geographic variation. Although supply-driven heterogeneity may reflect waste stemming from disagreement among physicians regarding optimal practice styles, it could also reflect an optimal allocation of physical and human capital. Conversely, although our model shows a formal sense in which demand-driven variation is likely to be consistent with efficiency, this need not be the case in a richer model. We view our findings as a first step toward a more welfare-relevant understanding and a clarification of an influential body of existing evidence.

Our empirical strategy relates to past work using changes in residence or employment to separate effects of individual characteristics from geographic or institutional factors. Most closely

6. See Skinner (2011) and Chandra, Cutler, and Song (2012) for reviews, and Cutler et al. (2015) and Baker, Bundorf, and Kessler (2014) for more recent contributions. Chandra, Cutler, and Song (2012) write: "In general, the literature points to the importance of supply-side incentives over demand-side factors in driving treatment choice" (p. 425) and "most of the literature agrees that patient characteristics and preferences do not explain much of the differences across areas" (p. 402). With regard to health status, they write: "Some researchers argue that variation is accounted for by population disease burden . . . but other authors argue that prevalence of diagnosis is itself endogenous across areas" (p. 402). Skinner (2011) writes: "While demand factors are important—health in particular—there remains strong evidence for supply-driven differences in utilization" (p. 46). An exception to this consensus is Sheiner (2014), who argues that patients may explain most or all of the variation.

related are Song et al. (2010), who look at how health measures change around patient moves, and Molitor (2014), who looks at cardiologist behavior changes around their moves. Outside of the health care sector, a number of papers beginning with Abowd et al. (1999) use matched worker-firm data to separately identify worker and firm fixed effects. In this vein, we draw especially on Card, Heining, and Kline's (2013) study of German workers and firms. Other work uses geographic or employment changes to study neighborhood effects on children (Aaronson 1998), cultural assimilation of immigrants (Fernandez and Fogli 2006), brand preferences (Bronnenberg, Dube, and Gentzkow 2012), tax reporting (Chetty et al. 2013), teacher value added (Chetty, Friedman, and Rockoff 2014), and retirement savings decisions (Chetty et al. 2014).

Section II introduces our model and estimation strategy. Section III describes our data and presents summary statistics. Section IV presents our main analysis of the role of demand and supply factors in explaining geographic variation in health care utilization. Section V explores the correlates of our estimated patient and place effects. Section VI concludes. All appendix materials are available in the Online Appendix.

II. Model and Empirical Strategy

II.A. Model

We build a simple model of demand and supply for health care, similar in spirit to the model in Cutler et al. (2015). Our goals are to illustrate the demand and supply factors that drive equilibrium utilization and clarify the underlying assumptions of our empirical specification.

A population of patients i in year t utilizes health care $y_{it} \in \mathbb{R}^{+}$. Patients differ along three dimensions: health status h_{it}, preferences η_i, and geographic area j. Higher values of h_{it} represent worse health; the time-constant scalar preference parameter η_i is defined so that higher values represent tastes for more aggressive care. Some patients are "nonmovers" who live in one area j throughout the sample, whereas others are "movers" whose area changes exactly once.[7] Patients' expected

7. We exclude patients who move more than once from the main analysis for simplicity, but we show that the results are robust to including them.

continuation utility $u(y|h_{it}, \eta_i) = -\frac{1}{2}(y - h_{it})^2 + \eta_i y$ is maximized at $y^*_{it} = h_{it} + \eta_i$, the level of care the patient would choose if they were fully informed and faced a zero out-of-pocket price for care. We assume that the expectation of y^*_{it} given the data observed by the econometrician depends only on a patient fixed effect and a vector of observables x_{it}: $E(y^*_{it}|\{i, j, t, x_{it}\}) = \alpha_i + x_{it}\beta$.

Each patient living in area j in year t is matched to a representative physician who determines the patient's care. We assume that a physician chooses y_{it} to maximize the perceived utility of her patients $\tilde{u}_j(y)$ minus her (net, private) costs of care provision, $PC_{jt}(y)$.[8] Thus we can write:

$$y_{it} = \arg\max_y \tilde{u}_j(y|h_{it}, \eta_i) - PC_{jt}(y). \tag{1}$$

The difference between perceived patient utility $\tilde{u}_j()$ and true patient utility $u()$ captures potentially heterogeneous beliefs that would lead physicians to disagree about the appropriate level of care. We assume $\tilde{u}_j(y|h_{it}, \eta_i) = u(y|h_{it}, \eta_i) + \lambda_j y$, so that higher values of λ_j represent relatively aggressive practice styles. A variety of factors can affect PC_{jt}, including monetary incentives, organizational rewards, and physical and human capital. We assume $PC_{jt}()$ is linear in y and additively separable in j and t.

Maximization of equation (1) yields our main estimating equation for patients i living in area j throughout year t:[9]

$$y_{ijt} = \alpha_i + \gamma_j + \tau_t + x_{it}\beta + \epsilon_{ijt}, \tag{2}$$

where $\gamma_j + \tau_t = \lambda_j - PC_{jt}'()$ and our assumptions imply $E(\epsilon_{ijt}|\{i, j, t, x_{it}\}) = 0$. In our main specification, the quantity y will be the log of total utilization, which we define more precisely in Section III, and x_{it} will consist of dummies for five-year age bins, and fixed effects $\rho_{r(i,t)}$ for movers, where for a mover who moves during year t^*_i the relative year is $r(i, t) = t - t^*_i$. Including these relative year effects allows for the possibility that the decision to move is correlated with shocks to health

8. For simplicity, we assume that net provider costs have been scaled to the same units as patient utility times its weight in the physician's objective function. Less compactly, we can assume the physician maximizes $\Psi\tilde{u}_j() - \tilde{PC}_{jt}()$, where $\tilde{PC}_{jt}()$ is measured in dollars and Ψ is the weight in the physician's objective assigned to patient utility, then define $PC_{jt}() = \frac{\tilde{PC}_{jt}()}{\Psi}$.

9. We do not model outcomes for movers in year t^*_i, when they spend part of the year in their origin area and part of the year in their destination. When we estimate equation (2), we omit these observations.

status—for example, because sick patients sometimes move to seek care, or, at the other extreme, because they are unable to bear the physical costs of moving. We normalize $\rho_{r(i,t)}$ to zero for nonmovers.

Our main goal is to decompose variation in average log utilization across regions into a demand-side component attributable to patients and a supply-side component attributable to place. To define this decomposition formally, let $\overline{y}_{jt}$ denote the expectation of y_{it} across patients living in area j in year t, and let $\overline{y}_j$ denote the average of $\overline{y}_{jt}$ across t. Let $\overline{y}^*_{jt}$ and $\overline{y}^*_j$ denote the analogous expectations of the patient-optimal level of care $y^*_{it} = h_{it} + \eta_i$, which by the foregoing assumptions is equal to $\alpha_i + x_{it}\beta$. Then the difference in average log utilization between any two areas j and j' is the sum of the differences of the place and patient components: $\overline{y}_j - \overline{y}_{j'} = (\gamma_j - \gamma_{j'}) + (\overline{y}^*_j - \overline{y}^*_{j'})$. When we talk about larger groups R that consist of multiple areas j, we abuse notation by letting $\overline{y}_R$, $\overline{y}^*_R$, and $\overline{\gamma}_R$ denote the simple averages of $\overline{y}_j$, $\overline{y}^*_j$, and γ_j across areas in R.

We define the share of the difference between areas j and j' attributable to place to be

$$S_{place}(j,j') = \frac{\gamma_j - \gamma_{j'}}{\overline{y}_j - \overline{y}_{j'}}, \tag{3}$$

and we define the share attributable to patients to be

$$S_{pat}(j,j') = \frac{\overline{y}^*_j - \overline{y}^*_{j'}}{\overline{y}_j - \overline{y}_{j'}}.$$

Note that although $S_{pat}(j,j')$ and $S_{place}(j,j')$ sum to 1, neither need be between 0 and 1, since it is possible that $\gamma_j - \gamma_{j'}$ and $\overline{y}^*_j - \overline{y}^*_{j'}$ have opposite signs. We define $S_{pat}(R,R')$ and $S_{place}(R,R')$ to be the analogous shares for groups R and R'. We let $\hat{y}_j$ denote the sample analogue of $\overline{y}_j$. Given consistent estimates $\hat{\gamma}_j$ of γ_j, we form consistent estimates $\hat{y}^*_j = \hat{y}_j - \hat{\gamma}_j$ of $\overline{y}^*_j$.

II.B. Discussion

Our stylized model clarifies the underlying economic factors that drive the patient and place components we estimate and their relationship to factors previously discussed in the literature.

The patient component $(\overline{y}^*_j - \overline{y}^*_{j'})$ is the difference in the utility-maximizing level of care y^*_{it} for an average patient living in j

and an average patient living in j'. This in turn is the sum of the differences in average health h_{it}, and average preferences η_i. The former will be driven by demographics such as age; behavioral factors such as diet, exercise, or smoking (Xu et al. 2013); and genetic predispositions to disease. The latter captures the way patients trade off the disutility of the pain, suffering, or inconvenience of treatment against the value of improved health, as well as ethical or religious beliefs about the value of prolonging life (Barnato et al. 2007).

The place component $(\gamma_j - \gamma_{j'})$ is the sum of the differences between j and j' in physicians' perceptions of marginal benefits λ_j minus private marginal costs $PC'()$. Each of these nests a variety of factors that have been fleshed out in more detail in the literature. For example, differences in $\tilde{u}'()$ capture 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. (2015). Differences in private marginal costs $PC'()$ capture a number of factors including physicians' disutility or difficulty in delivering a given level of care, which in turn reflects factors such as skill, training, or experience (as in Chandra and Staiger 2007), liability concerns (as explored in Currie and MacLeod 2008), or the opportunity cost of physicians' time. Private marginal costs may also be affected by organizational features such as available physical capital, the prevalence of nonprofit hospitals, nonmonetary career incentives, insurer constraints, peer effects among doctors, and organizational culture (Lee and Mongan 2009).

One way to interpret the patient share $S_{pat}(j, j')$ is in terms of a counterfactual: by what share would the gap in utilization between the two areas fall if patients were randomly reallocated between them? Crucially, this is a partial equilibrium experiment in that it holds fixed the existing perceptions, incentives, and physical and human capital of physicians and organizations embedded in γ_j. In the medium or long run, we would expect all of these factors to potentially adjust: an area facing sicker patients might invest more in physical or human capital, might specialize in more intensive forms of care, or might see shifts in its physicians' perceptions about what care is appropriate. Chandra and Staiger (2007) provide evidence suggesting that such adjustments are quantitatively important, and we stress that our estimates should be interpreted as partial equilibrium effects that shut down adjustment along these margins.

Under the assumptions of our model, we can also relate our patient-place decomposition to conclusions about welfare. Suppose that the social cost of care in location j and year t is $SC_{jt}(y)$, so that a social planner would choose y_{it} to maximize $u(y|h_{it}, \eta_i) - SC_{jt}(y)$. Then maintaining the other assumptions of our model, and assuming that we can write $SC_{jt}'(y) = -(\gamma_j^* + \tau_t^*)$, equation (2) under the social planner solution would be identical, except that $\tau_t + \gamma_j$ would be replaced with $\tau_t^* + \gamma_j^*$. For any two locations j and j', the patient component $(\overline{y}_j^* - \overline{y}_{j'}^*)$ is "efficient" in the sense that it would remain unchanged under the social planner. In contrast, the place component $(\gamma_j - \gamma_{j'})$ may or may not be efficient in this sense: it will differ from the social planner solution to the extent that physicians have inaccurate beliefs $(\lambda_j \neq 0)$, or their net private marginal costs $PC'()$ differ from the social marginal cost $SC'()$. Of course, these welfare implications require strong assumptions, and our patient-place decomposition need not have a tight relationship to welfare in a more general model. For example, if variation in patient preferences (η_i) partly reflects misinformation or distorted beliefs (as in Baicker, Mullainathan, and Schwartzstein 2015), some component of the patient component could be inefficient as well.[10]

II.C. Identification

The model in equation (2) is only identified if the data include movers. If all patients were nonmovers, there would be no way to separate differences in the area fixed effects γ_j from differences in the average patient characteristics $\overline{y}_j^*$. The key to separate identification of these two components is the observed changes in utilization when patients move.[11]

To build intuition, consider a simplified version of our model in which the τ_t and x_{it} are set to zero, so utilization depends only on patient and place fixed effects plus the error term. Normalize

10. The patient component could also fail to be efficient if the difference between the private and social marginal cost of care, $SC'() - PC'()$, varies with patient health or preferences. An example would be if out-of-pocket costs for different kinds of treatments vary across patients and/or areas. Setting this aside seems a reasonable approximation in the Medicare context, since patients have relatively homogeneous insurance, although it may not be strictly true (for example, because rates of supplemental coverage vary across areas).

11. A sufficient condition for identification is that the number of movers between any pair of areas j and j' grows large as the total sample size approaches infinity. Abowd et al. (2002) discuss weaker conditions for identification.

the national mean of α_i to zero (in general, we will not be able to identify the national means of patient, place, and time effects separately, but this will not affect our decompositions of geographic variation). Suppose we observe a large number of patients who move from area j' to area j. Then the difference $\Delta^j_{j'}$ between their average y_{it} in the years after the move and the years before the move is a consistent estimator of $\gamma_j - \gamma_{j'}$. If we observe similar samples of patients moving between the other areas in the sample, along with the overall mean of log utilization $\overline{y}$, we can form consistent estimates $\hat{\gamma}_j$ of each γ_j. The $\overline{y}^*_j$ would then be consistently estimated by $\hat{y}_j - \hat{\gamma}_j$.

Identification in the full model is similar. Identifying the τ_t and age coefficients is standard and does not rely on movers. Adding the relative year effects $\rho_{r(i,t)}$ to x_{it} has a more substantial effect. It allows for arbitrary changes in log utilization for movers pre- and postmove, with the restriction that these changes are the same regardless of the origin and destination. In the full model, therefore, observing only movers from j' to j is not enough to identify $\gamma_j - \gamma_{j'}$, because $\Delta^j_{j'}$ would also depend on the difference between the postmove and premove $\rho_{r(i,t)}$. Identification in this case comes from the differences in the changes across movers with different origins and destinations. If we have movers from j' to j and also movers from j to j', for example, we can estimate $\gamma_j - \gamma_{j'}$ consistently as $\frac{\Delta^j_{j'} - \Delta^{j'}_j}{2}$.

Importantly, our model permits movers to differ arbitrarily from nonmovers in both levels of log utilization (via the α_i) and trends in log utilization around their moves (via the $\rho_{r(i,t)}$). The latter would allow for moves to be associated with either positive or negative health shocks, for example. In principle we can allow substantially more flexibility, including area- or individual-specific trends, different fixed effects by subperiods, and interactions between γ_j and patient observables. We can also add flexibility by using data for movers only in the years just before or after their move in the spirit of a regression discontinuity. We explore robustness to specifications along these lines later.

Our model is nevertheless restrictive in several important ways. First, we cannot allow for shocks to utilization that coincide exactly with the timing of the move and that are correlated with utilization in the origin and destination. In the example above, suppose that for movers from j' to j the conditional expectation of

health h_{it} in years just after the move is strictly greater than for movers from j to j'. This would inflate $\Delta^{j}_{j'}$ relative to $\Delta^{j'}_{j}$, and lead $\frac{\Delta^{j}_{j'}-\Delta^{j'}_{j}}{2}$ to be an overestimate of $\gamma_j - \gamma_{j'}$. As a concrete example, this could occur if patients who receive adverse health shocks respond by moving to relatively high-utilization areas. The result in this case would be that we would attribute some of the health shock to the effect of moving and thus overstate the role of places relative to patients.

Although we cannot rule out such bias entirely, the pattern of results provides some comfort that it is unlikely to be large. Deterioration in health status that occurs gradually and is correlated with utilization in the destination and origin would tend to show up as pretrends in our event study analysis (Section IV.A). In fact we do find a positive pretrend, but its magnitude is small, and we show that the results are robust to restricting the data to a small window around the move. Sudden health events that prompt a move to systematically higher utilization places could potentially cause bias without a pretrend. However, such shocks would tend to produce a postmove spike in our event studies that dissipates over time (assuming treatment for acute conditions is most intense immediately after they occur), and this is not the pattern we observe.

Second, our specification assumes that α_i and γ_j are additively separable in the equation for log utilization. We see this as an attractive assumption economically. It has the intuitive implication that patient and place characteristics affect the *level* of utilization multiplicatively, and thus that the (level) utilization of patients who are sick or prefer intensive care (i.e., have high α_i) will vary more across places than that of patients who are healthy or rarely seek care (i.e., have low α_i).[12] We also see the log model as appealing on econometric grounds, given utilization's skewed cross-sectional distribution and large secular trend.

That said, the log specification nevertheless imposes some important restrictions. It rules out, for example, variation across places that causes an equal level shift for all patients

12. To take a concrete example, suppose that patients have either one or two chronic conditions, and that places spend either \$5,000 or \$10,000 per chronic condition. This would imply a model additive in logs, with $\exp(\alpha_i) \in \{1,2\}$, $\exp(\gamma_j) \in \{5,10\}$, and the log of utilization y_{ij} equal to $\alpha_i + \gamma_j$.

regardless of their α_i. This could occur, for example, if some places mandate flu shots or other preventive treatments with similar cost for all patients. More subtly, our decompositions of geographic variation in log utilization give relatively more weight to differences in the bottom part of the utilization distribution than a decomposition in levels would. In Section IV, we present a variety of specification and robustness checks that bear on these issues.

The assumption that patient and place effects are additively separable also rules out the possibility that different types of patients seek out different types of health care within a place. This can be relaxed by allowing interactions between γ_j and patient observables, as we explore later. It can also be partly addressed by examining whether the results vary when area j is defined at higher and lower levels of geography, which we also explore later. But our specification does not capture richer models of behavior in which, within appropriately defined areas, observationally similar patients seek out different types of providers.

Third, our approach relies fundamentally on the assumption that the γ_j that are relevant for movers are the same as those that are relevant for nonmovers. If movers differ in ways that change the relevant place effects, our decompositions would apply only to variation in utilization among movers rather than to the population as a whole.

Finally, our model does not allow for the possibility that α_i in a given period is a function of past values of y_{it}. If, for example, patients in high-utilization areas become accustomed to visiting the doctor frequently and receiving a large number of tests when they do, they might continue to demand these services postmove. In this case, variation across areas in current α_i could partly be caused by the influence of γ_j in the past. We discuss the possibility of such habit formation and evidence that suggests it may be small in Section IV.A. However, the fact that we focus on older patients means that we cannot rule out habit formation over long horizons. If α_i depends on consumption of health care early in life, for example, some of what we attribute to patients may reflect supply-side differences in the past. The long-term effect of supply-side changes could then be larger than our estimates would suggest.

II.D. Event-Study Representation

To visualize the way utilization changes when patients move, we define an alternative "event-study" representation of equation (2).

To build intuition, it again helps to start with the simple case where τ_t and x_{it} are set to zero and where our panel of movers is balanced in the sense that each mover is observed for the same number of years pre- and postmove. If all movers had the same origin j' and destination j, we could construct an event study by simply plotting the average of y for movers by relative year $r(i,t)$. When origins and destinations vary, however, this plot would not be very informative. If the flow from any j' to j were equal to the flow from j to j', for example, we would expect the graph to show no change around the move, even if the absolute values of the underlying changes on move were large.

To produce a more informative plot, we would like to scale y so that the direction and magnitude of the jump on move are informative regardless of the origin and destination. For a mover i whose origin and destination areas are $o(i)$ and $d(i)$, respectively, we denote by δ_i the difference in average log utilization between the mover's destination and origin:

$$\delta_i = \overline{y}_{d(i)} - \overline{y}_{o(i)}, \tag{4}$$

and we let $S^i_{place} = S_{place}(d(i), o(i))$ and $S^i_{pat} = S_{pat}(d(i), o(i))$. Following Bronnenberg, Dube, and Gentzkow (2012), we define for mover i:

$$y^{scaled}_{it} = \frac{y_{it} - \overline{y}_{o(i)}}{\delta_i}.$$

Note that y^{scaled}_{it} will be 0 if the mover's utilization is equal to the average in his origin, 1 if it is equal to the average in his destination, and between 0 and 1 if the mover's utilization falls between the two. If the model is correct, the expectation of y^{scaled}_{it} should be flat before and after move and the jump on move will be equal to the average value of S^i_{place} across movers. Plotting the averages of y^{scaled}_{it} by relative year would thus produce an event-study figure with a direct interpretation in terms of the model quantities of interest. The larger the jump in y^{scaled}_{it} on move, the greater the share of geographic variation we would attribute to place, and the smaller the share we would attribute to patients.

To implement this in the full model, we must deal with three additional complications. First, we need to allow for the controls τ_t and x_{it}. Second, our panel is not balanced and so changes in the composition of movers could introduce pre- or post-trends into the event-study figure. To avoid this, we need to control for the individual fixed effects α_i explicitly. Third, the difference δ_i can be very small in some cases, which would make the simple average of y_{it}^{scaled} poorly behaved. This leads us to prefer a regression implementation that avoids dividing by δ_i.

Observe that we can rewrite equation (2) for movers as:

$$y_{it} = \alpha_i + \gamma_{o(i)} + I_{r(i,t)>0} S^i_{place} \delta_i + \tau_t + x_{it}\beta + \epsilon_{it}, \tag{5}$$

where $I_{r(i,t)>0}$ is an indicator variable for relative year greater than zero. Combining $\alpha_i + \gamma_{o(i)}$ into a single patient fixed effect $\tilde{\alpha}_i$, replacing δ_i with its sample analogue $\hat{\delta}_i$ (calculated based on both movers and nonmovers in the destination and the origin), and parameterizing the interaction with $\hat{\delta}_i$ as a flexible function of relative year yields

$$y_{it} = \tilde{\alpha}_i + \theta_{r(i,t)}\hat{\delta}_i + \tau_t + x_{it}\beta + \epsilon_{it}. \tag{6}$$

This is the event-study equation we take to the data. The relative-year specific coefficients $\theta_{r(i,t)}$ are the parameters of interest: they measure changes in y_{it} in years around the move scaled relative to δ_i. If the sampling error in $\hat{\delta}_i$ is ignorable,[13] and heterogeneity in S^i_{place} is orthogonal to the other variables in the model, the plot of the $\theta_{r(i,t)}$ will have a precise interpretation similar to that of the average y_{it}^{scaled} in the simple case: the plot should be flat before and after move and jump on move by a weighted average of S^i_{place}.

III. Data and Summary Statistics

III.A. Data and Variable Definitions

Our primary data source is a 20% random sample of Medicare beneficiaries ("patients") from 1998 through 2008.[14] These data contain approximately 13 million patients. For each

13. Because the number of nonmovers we observe in each HRR is large, sampling error in $\hat{\delta}_i$ is small. We show in Online Appendix Section 4.5 that accounting explicitly for noise in $\hat{\delta}_i$ has no impact on our event-study results.

14. The sample is a panel defined by taking all Medicare beneficiaries in each year whose Social Security number ends in either 0 or 5. The sample thus varies

patient, we observe information on all Medicare claims for inpatient care, outpatient care, and physician services. For each claim, the data include information on the diagnosis, type and quantity of care provided, and the dollar value reimbursed by Medicare. We also observe demographic information for each patient, including age, sex, race, and zip code of residence, defined as the address on file for Social Security payments as of March 31 of each year. To match the timing with which we observe patients' residence, we define all outcome variables for year t to be aggregates of claims from April 1 of year t through March 31 of year $t+1$.[15]

Our primary outcome variable is based on an index of overall health care utilization by individual by year, which we refer to simply as "utilization." To construct the measure, we follow Gottlieb et al. (2010) in adjusting total annual expenditure for regional variation in prices. Online Appendix Section 2 describes the construction of the measure in detail. We prefer to focus on utilization quantities and set aside variation in administratively set prices because the drivers of the latter are different and better understood.[16]

In our main specifications, we define the outcome y_{it} to be the log of utilization plus 1, which we refer to simply as "log utilization." As discussed in Section II.C, we prefer a log specification both economically and econometrically. We explore other functional forms in the robustness section later. We also examine a number of other outcome measures, including subcategories of utilization and indicators for particular treatments, which are defined in more detail shortly.

Our geographic unit of analysis is a Hospital Referral Region (HRR), as defined by the 1998 Dartmouth Atlas of Health Care. The 306 HRRs are collections of zip codes designed to approximate markets for tertiary hospital care.[17] Consistent with the

from year to year, but a given patient remains in the sample as long as they are enrolled in Medicare.

15. We include data from the first few months of 2009 to compute outcomes for our final sample year (t = 2008) which runs from April 2008 to March 2009.

16. We show in the robustness analysis that our main conclusions are unchanged if we use total annual expenditure in place of utilization.

17. See www.dartmouthatlas.org/downloads/geography/ziphsahrr98.xls and http://www.dartmouthatlas.org/downloads/methods/geogappdx.pdf. Each HRR consists of a collection of zip codes that contain at least one hospital that performs major cardiovascular procedures and neurosurgeries. Zip codes are grouped into an

existing literature, we define average log utilization and other outcomes for an HRR j to include all claims by residents of j, regardless of the location of the claims themselves. On average, about 16% of claims occur outside a patient's HRR of residence.

We define patients to be "nonmovers" if their HRR of residence is the same throughout our sample period. We define patients to be "movers" if their HRR of residence changes exactly once. Our baseline analysis excludes patients whose HRR of residence changes more than once. We show in Online Appendix Section 4.3 that including multiple movers does not substantively change our estimates.

In some of our analyses, we compare movers to a matched subsample of nonmover patient years chosen to match as closely as possible the characteristics of our mover sample. For each mover in our data in each calendar year, we randomly draw a nonmover in the same year in the mover's origin HRR who shares the mover's gender, race, and five-year age bin. The union of the selected nonmover patient-years forms the "matched sample of nonmovers" we refer to below.

III.B. Sample Restrictions and Summary Statistics

From our original sample of 13 million patients, we retain a 25% random sample of nonmovers along with all movers. We then restrict the sample to the 88% of patient-years where patients are between 65 and 99 years old, exclude 20% of the remaining patient-years for patients enrolled in Medicare Advantage (for whom we do not observe claims), and exclude the remaining 7% of patient-years for patients who do not have Medicare Part A or B coverage in all months (including, for example, patients who enroll mid-year in the year they turn 65). Finally, among patients whose HRR of residence changes at least once, we exclude the 18% whose HRR of residence changes more than once, as well as the 35% of the remaining movers whose share of claims in their destination HRR, among claims in either their origin or

HRR based on where the highest proportion of cardiovascular procedures are referred. Each HRR must have a population of at least 120,000. We drop roughly 2% of patient-years whose zip codes do not match the 1998 HRR definitions.

destination HRR is not higher by at least 0.75 in the postmove years relative to the premove years.[18]

When we compute HRR averages, such as the sample analogue $\hat{y}_j$ of $\overline{y}_j$, we omit movers in their move year, and we weight nonmovers by four to account for our sampling procedure. All HRR averages are computed by first averaging across individuals in the HRR in each year and then taking a simple average across years.

Our final sample includes 2.5 million patients, of whom approximately 0.5 million are movers. Table I reports summary statistics separately for movers and nonmovers. The characteristics of the two groups are broadly similar, although there are some differences. Relative to nonmovers, movers are slightly more likely to be female, white, and older and more likely to live initially in the South or West, rather than the Midwest or Northeast. Average annual utilization in both groups is roughly $7,500 per year, with a standard deviation of about $10,000, and 6% of observations equal to 0. Health care utilization is notoriously right-skewed: the median across both groups is about $4,300 and the 90th percentile is almost $18,000.

There are a variety of reasons that individuals may enter or exit the sample, including death, entering or exiting Medicare Advantage, and entering or exiting our 65–99 age window. The average nonmover in our sample is observed for 6.3 years (out of a possible 11), and the average mover for 7.5 years. The difference is partly mechanical, due to the fact that we must observe a patient for at least two years to classify them as a mover. About a third of patients die during our sample period, and about 20% enter or exit at some point due to enrollment in Medicare

18. The change in claim share is not defined for movers who do not have at least one claim both pre- and postmove. We exclude these cases if (i) they have no postmove claims and a premove destination claim share greater than 0.05; (ii) they have no premove claims and a postmove destination claim share less than 0.95. The claims data suggest several explanations for why some movers do not satisfy our change in claim share criterion. In a large share of cases, the geographic distribution of claims remains roughly the same before and after the recorded move, suggesting that the patient changed the address on file with Social Security without changing their residence. This could occur if they decided to have their Social Security checks sent to a child who was handling their finances, for example. In other cases, patients appear to have multiple residences both before and after the move, with the share of claims in the destination increasing postmove by an amount less than our 0.75 threshold. We show in Online Appendix Section 4.3 that our results are robust to alternative ways of defining movers.

TABLE I
SUMMARY STATISTICS

	(1) Nonmovers	(2) Movers
Female	0.57	0.60
White	0.86	0.88
Age first observed:		
65–74	0.67	0.59
75–84	0.24	0.31
≥85	0.09	0.09
First observed residence:		
Northeast	0.20	0.17
South	0.39	0.41
Midwest	0.26	0.19
West	0.16	0.23
Annual utilization:		
Mean	$7,796	$7,399
Std. dev.	$12,690	$9,567
Share of patient-years with zero	0.06	0.06
Number of chronic conditions:		
Mean	2.98	3.30
Std. dev.	2.15	2.06
Share of patient-years with zero	0.18	0.15
Average # of years observed	6.26	7.45
Share who die during sample	0.35	0.32
Share of patient-years excluded because patient is in Medicare Advantage that year	0.18	0.20
# of patients	2,033,096	497,097
# of patient-years	12,730,766	3,702,189

Notes. Rows for female, white, age first observed, and first observed residence report the shares of patients with the given characteristics among movers and nonmovers. Patient-years in Medicare Advantage are excluded from the baseline sample. The denominator for the row "Share of patient-years excluded because patient is in Medicare Advantage that year" is the sample of all movers and 25% of nonmovers, before any other sample restrictions. In all other rows, the sample is the baseline sample of all movers and 25% of nonmovers ($N = 16{,}432{,}955$ patient-years).

Advantage. In our robustness analysis, we discuss possible biases due to selective attrition and show that our results are robust to some alternative ways of handling it. Mortality rates are broadly similar for movers and nonmovers, alleviating concern that our analysis of movers might miss end-of-life expenditures.

Figure I shows the distribution of average annual utilization across HRRs. The mean HRR has average utilization of $6,629 per person per year, with a standard deviation of $779. The ranking of HRRs by utilization is reasonably stable over time: the correlation between an HRR's rank in the first half of our

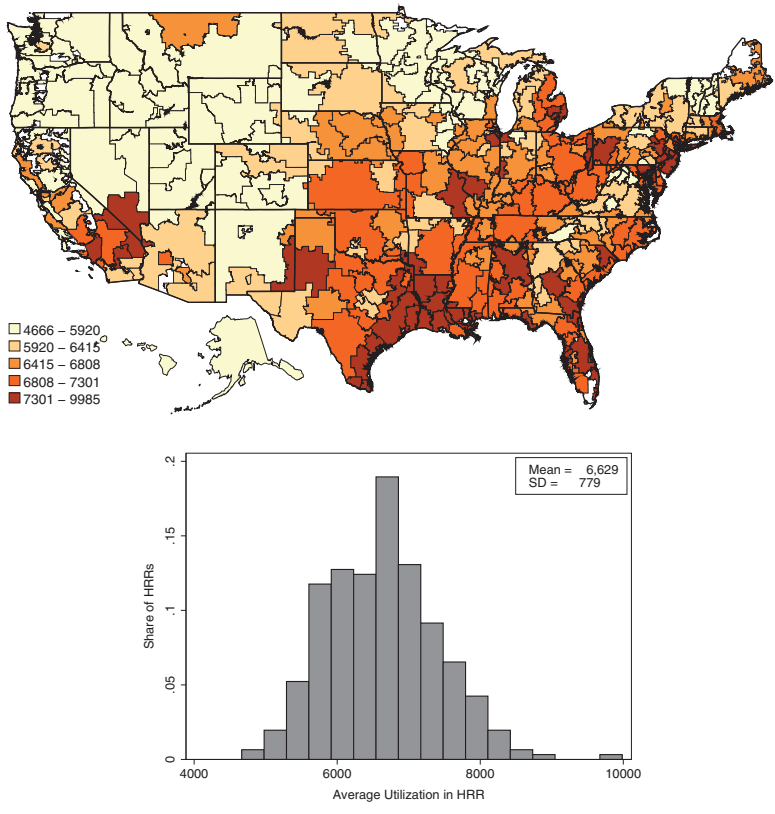

FIGURE I

Distribution of Utilization across HRRs

Map shows the distribution of level utilization in quintiles. Lower and upper limit of each quintile are displayed in the legend. The sample is all movers and nonmovers ($N = 16{,}432{,}955$ patient-years). Histogram displays the distribution of average utilization by HRR. We first average utilization across individuals within each HRR-year, upweighting nonmovers by four, and then take a simple average within HRR across years.

sample (1998–2003) and the second half of our sample (2004–2008) is 0.9. We show in Online Appendix Figure 9 that if we divide HRRs into quintiles by utilization, the evolution of utilization for the different quintiles is roughly parallel. These facts are consistent with prior literature showing patterns of geographic variation in health care utilization have been relatively stable since the early 1990s (Weinstein et al. 2004; Rettenmaier and Saving 2009; Chandra, Sabik, and Skinner 2011).

Online Appendix Section 3.1 presents additional summary statistics for movers. The average distance moved is 588 miles, with a median of 357 miles and a standard deviation of 616 miles. Roughly 68% of moves cross state boundaries, and 50% cross census division boundaries. Moves to Florida account for 12% of all moves, and moves to Arizona or California account for an additional 12%; we show in Online Appendix Table 9 that our results are robust to excluding moves to Florida, Arizona, and California. We also show the distribution of movers across different destination HRRs. The median HRR receives 1,133 movers; the range of movers into an HRR is from 135 to 12,797.

Finally, we examine the time-varying correlates of moving. Online Appendix Figure 4 shows that moving is correlated with an increase in utilization, including a spike in utilization in the year of move. We also report evidence from the Health and Retirement Study (HRS) on the reasons older Americans move. The study is a nationally representative (approximately biannual) longitudinal survey of Americans over the age of 50. We limit the HRS sample to individuals aged 65 and over, and define movers as individuals who move across HRRs. The most common self-reported reasons for moving are to be "Near/with children" (31%), "Health problems or services" (13%), and to be "Near/with relatives or friends" (10%). Analysis of the HRS panel data shows that significant predictors of moving include being widowed and retiring. Declines in self-reported health status do not predict moving in the panel.

IV. Main Results: Patient versus Place

IV.A. Event Study

We begin with two figures that illustrate the variation driving our event study. Figure II shows a mover's claims in her destination HRR as a share of those in either her origin or her destination, by relative year. The figure shows a sharp change in the year of the move, with only a small share of claims in the destination premove or in the origin postmove.[19] The claim share in the year of the move (relative year 0) is close to 0.5, consistent with moves

19. In Online Appendix Section 4.3, we show that our results are robust to adjusting for the small amount of apparent measurement error in the timing of moves and to a range of alternative definitions of movers.

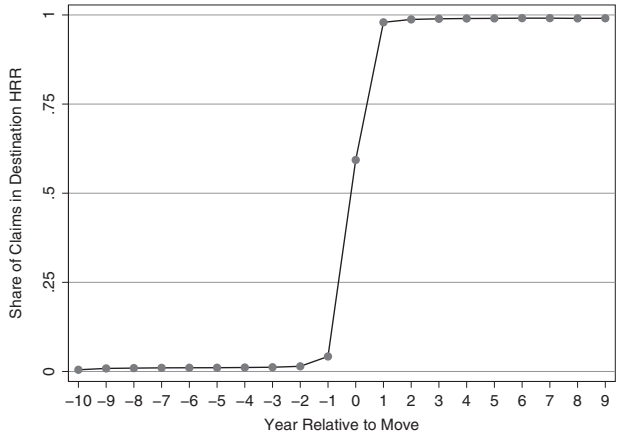

FIGURE II

Share of Claims in Destination by Relative Year

Figure shows the share of a mover's claims located in their destination HRR, among those in either their origin or their destination HRR. The sample is all movers (N = 3,702,189 patient-years).

being roughly uniform throughout the year. Figure III shows the distribution of $\hat{\delta}_i$, the average log utilization in a mover's destination minus the average log utilization in her origin. The mean value of $\hat{\delta}_i$ is close to zero and the distribution is roughly symmetric, implying that moves from low- to high-utilization HRRs are as common as moves from high to low. The standard deviation is 0.25, and there are a significant number of moves for which the absolute value of the difference is greater than 0.5.

As a first look at the way utilization changes around moves, Figure IV plots the change in log utilization (the average two to five years postmove minus the average two to five years premove) against the destination–origin difference in log utilization $\hat{\delta}_i$. If all geographic variation were due to place effects, we would expect this plot to have a slope of 1. If all variation were due to patient effects, we would expect this plot to have a slope of 0. One minus the actual slope is an estimate of a weighted average of the patient share S^i_{pat}.

Figure IV shows that the slope is in fact 0.63, suggesting an average patient share of roughly 0.37. The relationship is symmetric above and below zero and strikingly linear. This provides strong support for our additively separable model, which implies

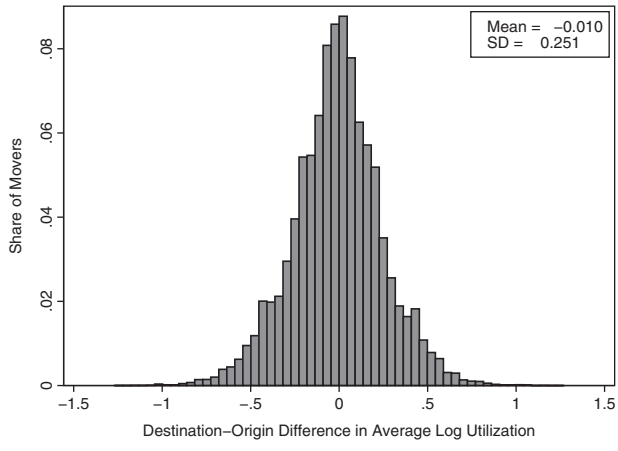

FIGURE III

Distribution of Destination-Origin Difference in Log Utilization

Figure shows the distribution across movers of the difference $\hat{\delta}_i$ in average log utilization between their origin and destination HRRs. The sample is all movers (N = 3,702,189 patient-years).

that the absolute change in log utilization when patients move from j to j' should be the same as when patients move from j' to j. These patterns are also consistent with the relative importance of patients being similar across origin-destination pairs.

We also plot with an $\times$ the average change in log utilization over the same period for our matched sample of nonmovers, to whom we assign $\hat{\delta}_i = 0$.[20] That this point and all points for movers have y values greater than zero reflects the positive time and age trends in utilization. That the point for nonmovers lies below the ones for movers with $\hat{\delta}_i \approx 0$ shows that moving is associated with an increase in utilization on average. This main effect of moving will be absorbed by our relative year indicators $\rho_{r(i,t)}$. We present additional descriptive evidence on the main effects of moving in Online Appendix Section 3.1.

Figure V shows how premove utilization of movers compares to utilization of nonmovers in their origin HRR's. The plot is identical to Figure IV except that the variable on the y-axis is now the average difference between log utilization of movers two to five years premove and that of their matched nonmovers in the same

20. See notes to Figure IV for details on this matching.

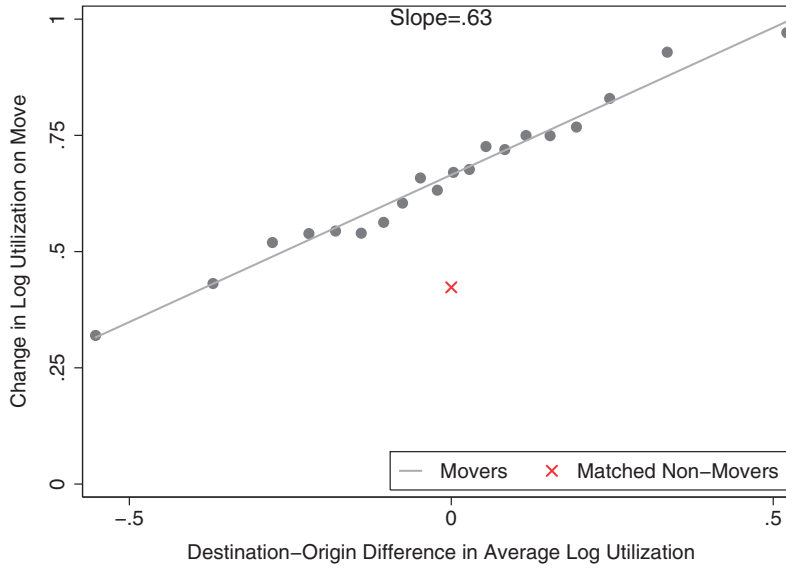

FIGURE IV

Change in Log Utilization by Size of Move

Figure shows the change in log utilization before and after move. For each mover, we calculate the difference $\hat{\delta}_i$ in average log utilization between their origin and destination HRRs, then group $\hat{\delta}_i$ into ventiles. The x-axis displays the mean of $\hat{\delta}_i$ for movers in each ventile. The y-axis shows, for each ventile, average log utilization two to five years postmove minus average log utilization two to five years premove. The line of best fit is obtained from simple OLS regression using the 20 data points corresponding to movers, and its slope is reported on the graph. The sample is all mover years between two and five years premove and between two and five years postmove (N = 1,919,137 patient-years). For comparison, we also compute the average change in log utilization for a sample of matched nonmovers, which we show with the × marker on the graph. Specifically, for each mover in our data in each calendar year, we randomly draw a nonmover in the same year in the mover's origin HRR who shares the mover's sex, race, and five-year age bin; the union of the selected nonmover patient-years forms the matched sample.

years. The plot has a small upward slope, suggesting that patients who will move to a high-utilization HRR have relatively higher premove utilization than those who will move to a low-utilization HRR. This slope is an order of magnitude smaller than the slope in Figure IV. Any systematic differences of this kind in the average utilization of movers will be absorbed by our patient fixed effects α_i.

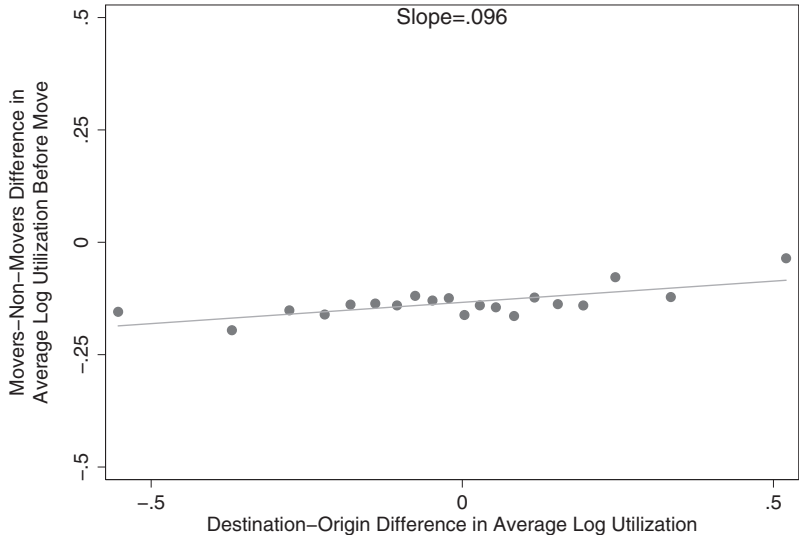

FIGURE V

Premove Differences in Log Utilization

Figure shows the level of premove log utilization for movers relative to nonmovers by the size of their subsequent move $\hat{\delta}_i$. For each mover, we calculate the difference $\hat{\delta}_i$ in average log utilization between their origin and destination HRRs, then group $\hat{\delta}_i$ into ventiles. The x-axis displays the mean of $\hat{\delta}_i$ for movers in each ventile. The y-axis shows for each ventile the average of difference in log utilization between mover and matched nonmover patient-years two to five years premove. In Figure IV we describe the construction of the matched sample of nonmovers. The line of best fit is obtained from simple OLS regression using the 20 data points, and its slope is reported on the graph. The sample is all mover years between two and five years premove (N = 1,048,843 patient-years).

Our main event-study results are shown in Figure VI, which plots estimated coefficients $\hat{\theta}_{r(i,t)}$ from equation (6).[21] Since these coefficients are only identified up to a constant term, we normalize the value for $r(i, t) = -1$ to 0. The figure shows a sharp, discontinuous jump at the time of the move, from 0 to approximately 0.5. As discussed, 1 minus the size of this jump can also be interpreted as an estimate of a weighted average of S^i_{pat}. This figure thus implies a patient share of roughly a half.

21. For computational ease, all of the event studies we report are estimated on the sample of movers only. We show in Online Appendix Figure 16 that including nonmovers does not affect the analysis.

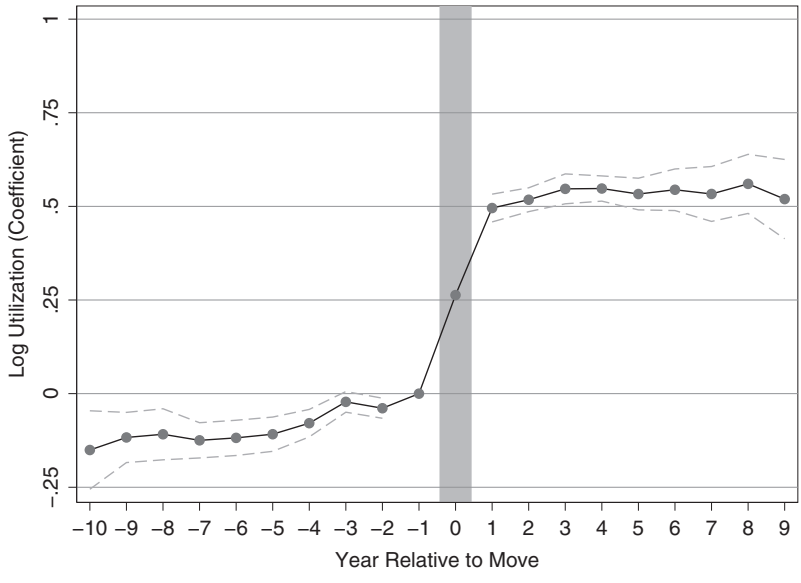

FIGURE VI

Event Study

Figure shows the coefficients $\tilde{\theta}_{r(i,t)}$ estimated from equation (6). The coefficient for relative year −1 is normalized to 0. The dependent variable y_{it} is log utilization; x_{it} consists of indicator variables for five-year age bins. The dashed lines are upper and lower bounds of the 95% confidence interval. We construct this confidence interval using a two-step procedure. In the first step, for each HRR j, we construct the asymptotic distribution of $\bar{y}_j$, which is a normal distribution with mean μ_j and standard deviation σ_j calculated from the data. In the second step, we bootstrap equation (6) with 50 repetitions drawn at the patient level, making a random draw from the distribution of $\bar{y}_j$ for each mover's origin and destination to construct their $\hat{\delta}_i$ for each repetition. The sample is all movers (N = 3,702,189 patient-years).

Under the assumptions of our model, the plot should be flat in the years before and after the move. In practice, the plot shows no post-trend and a small but statistically significant pretrend. This trend could reflect systematic changes in log utilization of movers relative to nonmovers. Because our model restricts both HRR and patient effects to be time constant, it could also pick up HRR-specific trends that are the same for movers and nonmovers but happen to be correlated with movers' $\hat{\delta}_i$. In our robustness analysis, we explore extensions that allow our fixed effects to change over time. We also allow arbitrary pre- and postmove trends for movers by using data only from the years just before or after the move, in the spirit of a

regression discontinuity. Finally, we show that the event studies look similar when estimated on various balanced panels.

One source of the patient heterogeneity we measure could be habit formation in the sense of Becker and Murphy (1988): patient preferences today could be a function of patients' utilization in the past. For example, patients who build a habit of getting regular checkups or flu shots may continue to do so wherever they go. More mechanically, patients who receive diagnoses or begin treatment in high-utilization areas may continue their treatment even after they move to low-utilization areas. This would affect the interpretation of our results, since patient characteristics today would partly reflect the impact of place characteristics in the past.

Several features of Figures IV and VI suggest that the role of habit formation may be limited. First, stories such as continuing aggressive treatments started premove would tend to predict a lot of persistence for those moving from high- to low-utilization areas, and less persistence for those moving from low to high. In fact, Figure IV shows that for any given magnitude $|\hat{\delta}_i|$ of the difference between origin and destination log utilization, changes in log utilization look symmetric for moves up and down.[22] Second, a signature of most models of habit formation is that utilization should continue to adjust toward average behavior in the destination in the years following a move. This is the key pattern that identifies habit formation in Bronnenberg, Dube, and Gentzkow (2012), for example. However, Figure VI shows remarkably little evidence of postmerge convergence.[23] Log utilization jumps discretely on move but remains almost perfectly flat for up to nine years thereafter, and this remains true whether we look at moves from low to high-utilization areas or moves from high to low (see Online Appendix Figure 11). Finally, many habit-formation models would predict less adjustment for older patients, since they have accumulated larger stocks of past experience; we show in Online Appendix Figure 12 that there is no evidence of such systematic differences by age. As already mentioned, however, none of this evidence rules out habit formation occurring at younger ages or over longer time spans than we observe in our data.

22. As further evidence, Online Appendix Figure 11 shows similar changes in utilization on move in event-study plots separately for moves up ($\hat{\delta}_i > 0$) and moves down ($\hat{\delta}_i < 0$).

23. Online Appendix Figure 8 shows that this remains true when we estimate our event study using a balanced panel.

TABLE II

ADDITIVE DECOMPOSITION OF LOG UTILIZATION

	(1) Above/ below median	(2) Top & bottom 25%	(3) Top & bottom 10%	(4) Top & bottom 5%	(5) McAllen & El Paso	(6) Miami & Minneapolis
Difference in average log utilization						
Overall	0.283	0.456	0.664	0.817	0.587	0.667
Due to place	0.151	0.271	0.406	0.461	0.374	0.466
Due to patients	0.132	0.185	0.258	0.356	0.213	0.200
Share of difference due to						
Patients	0.465	0.405	0.388	0.435	0.363	0.300
	(0.027)	(0.029)	(0.026)	(0.025)	(0.161)	(0.088)
Place	0.535	0.595	0.612	0.565	0.638	0.700

Notes. Table is based on estimation of equation (2), where the dependent variable y_{ijt} is log utilization and the controls x_{it} are indicators for age in five-year bins. The adjusted R-squared from estimating equation (2) is 0.503. Each column defines a set of areas R and R'. In columns (1)–(4) these are based on percentiles of average utilization $\bar{y}_j$. The first row reports the difference in average utilization overall between the two areas $(\hat{y}_R - \hat{y}_{R'})$; the second row reports the difference due to place $(\hat{\gamma}_R - \hat{\gamma}_{R'})$; the third row reports the difference due to patients $(\hat{y}^*_R - \hat{y}^*_{R'})$. The fourth row reports the share of the difference in average utilization between the two areas due to patient $(\hat{S}_{pat}(R, R'))$, which is the ratio of the third row to the first row. The last row reports the share of the difference in average utilization between the two areas due to place $(\hat{S}_{place}(R, R'))$, which is the ratio of the second row to the first row. Standard errors (in parentheses) are calculated using a bootstrap with 50 repetitions at the patient level. The sample is movers and nonmovers, excluding relative year 0 ($N = 16{,}031{,}875$ patient-years).

IV.B. Model Estimates

We exploit the variation captured in Figure VI to estimate equation (2). We use the estimates to quantify the roles of patients and of places in explaining geographic variation in log utilization. We present three main types of decompositions.

Table II, which we consider the central set of results in the article, presents an additive decomposition of the difference between high- and low-utilization areas. For different sets of high- and low-utilization HRRs R and R', we report the sample analogue of the patient share $S_{pat}(R, R') = \left(\frac{\hat{y}^*_R - \hat{y}^*_{R'}}{\hat{y}_R - \hat{y}_{R'}}\right)$, as well as the components $\hat{y}_R - \hat{y}_{R'}$, $\hat{\gamma}_R - \hat{\gamma}_{R'}$, and $\hat{y}^*_R - \hat{y}^*_{R'}$.

Column (1) decomposes the difference between above-median and below-median HRRs. We find that 47% of the difference in average log utilization is due to patients. This estimate is fairly precise; we can reject a role for patients of more than about 52% or less than about 41%.

Other partitions of HRRs result in a similar patient share. Patients account for 41% of the difference between the top and

bottom quartiles (column (2)), 39% of the difference between the top and bottom deciles (column (3)), and 44% of the difference between the top and bottom 5% (column (4)).[24] The final two columns look at two cases discussed in the introduction: McAllen relative to El Paso, and Miami relative to Minneapolis. Here, we find that patients account for 36% and 30% of the differences, respectively, though precision naturally falls with these smaller samples.

The magnitudes are consistent with the event-study analysis, which suggested a patient share of 50% based on the jump in log utilization from relative year −1 to 1, as well as with the slope of Figure IV. That the estimates are not identical reflects the fact that the additive decomposition is a slightly different experiment—analyzing differences between two groups of HRRs rather than averaging S^i_{pat} across all movers i—that the model uses all pre- and postmove years rather than the on-impact effect of the move, and that the model is estimated on both movers and nonmovers. The stability of the patient share across different partitions is consistent with the linear relationship shown in Figure IV, which implies that $S_{pat}(j,j')$ is not strongly correlated with $\overline{y}_j - \overline{y}_{j'}$.

We present a second, alternative decomposition in Table III. Here we ask what share of the cross-HRR variance in log utilization would be eliminated in a counterfactual where average patient characteristics $\overline{y}^*_j$ were equalized across HRRs. This is

$$S^{var}_{pat} = 1 - \frac{\mathrm{Var}(\gamma_j)}{\mathrm{Var}(\overline{y}_j)}. \tag{7}$$

Similarly, the change if area fixed effects were equalized is

$$S^{var}_{place} = 1 - \frac{\mathrm{Var}\left(\overline{y}^*_j\right)}{\mathrm{Var}(\overline{y}_j)}.$$

Note that unlike S_{pat} and S_{place}, this is not an additive decomposition; the sum of S^{var}_{pat} and S^{var}_{place} will not be 1 so long as $Cov(\overline{y}^*_j, \gamma_j)$ is nonzero. In estimating the relevant variances and covariances, we correct for sampling error using a split-sample approach.[25]

24. Online Appendix Figure 19 shows the corresponding event studies for the various partitions shown in columns (1) through (4).

25. We randomly assign movers within each origin-destination pair and nonmovers within each HRR to two approximately equal-sized subsamples and

TABLE III

VARIANCE DECOMPOSITION OF LOG UTILIZATION

	(1)
Cross-HRR variance of average:	
Log utilization	0.035
HRR effects	0.015
Patient effects	0.010
Correlation of average HRR and patient effects	0.353
	(0.052)
Share variance would be reduced if:	
HRR effects were made equal	0.717
	(0.014)
Patient effects were made equal	0.558
	(0.013)

Notes. Results based on estimates of equation (2). The first row reports variance of $\hat{y}_j$, which is estimated using the same specification as in Table II. The second, third, and fourth rows report the variance of $\hat{\gamma}_j$, variance of $\hat{y}_j^*$, and the correlation between $\hat{\gamma}_j$ and $\hat{y}_j^*$, respectively, using a split-sample approach to correct for the (correlated) measurement error in $\hat{\gamma}_j$ and $\hat{y}_j^*$. The last two rows report the share of the variance in cross-HRR utilization that would be reduced if HRR effects were made equal across areas (S_{place}^{var}) and if patient effects were made equal across areas (S_{pat}^{var}). Standard errors (in parentheses) are calculated using a bootstrap with 50 repetitions at the patient level. The sample size is the same as in Table II.

We find that 56% of variance would be eliminated if patient effects were equalized. We find that 72% of variance would be eliminated if place effects were equalized. We also find that there is a positive correlation between $\overline{y}_j^*$ and γ_j, with patients with a high demand for health care tending to sort to higher-utilization areas. Because this correlation is positive, S_{pat}^{var} and S_{place}^{var} sum to more than 1.

The results thus far decompose geographic variation in log utilization. As discussed in Section II.C, we believe modeling utilization in logs is appealing both economically and econometrically. However, even if our model is correctly specified, quantifying the drivers of geographic variation in logs is a different exercise than quantifying the drivers of variation in levels, as the implicit weights on low- and high-utilization observations will be different. If the relative importance of demand and supply factors varied substantially across the distribution of

estimate equation (2) separately on each subsample. We compute the variance of $\hat{\gamma}_j$ (or $\hat{y}_j^*$) as the covariance between the $\hat{\gamma}_j$'s (or $\hat{y}_j^*$'s) estimated from the two subsamples. The correlation between $\hat{\gamma}_j$ and $\hat{y}_j^*$ is computed from the variances of $\hat{\gamma}_j$ and $\hat{y}_j^*$, and the covariance between $\hat{\gamma}_j$ and $\hat{y}_j^*$, which we estimate as the average of the covariances between $\hat{\gamma}_j$ from one subsample and $\hat{y}_j^*$ from the other subsample.

utilization—for example, because patient preferences were either more or less important in big-ticket end-of-life expenditures compared with low-cost routine care—log and level decompositions could give very different answers.

To assess the importance of this issue, we present a third decomposition. Here, we ask what our estimated (log) model implies about the drivers of geographic variation measured in levels—how differences in level utilization would change if either $\overline{y}_j^*$ or γ_j were equalized across places. The details of this exercise and a complete set of results are presented in Online Appendix Section 4.2. A limitation to this exercise is that our "additive decomposition" into $S_{pat}(R,R')$ and $S_{place}(R,R')$ is no longer additive: the difference between the high- and low-utilization areas in levels is the product rather than the sum of the patient and place components, and so the percentage changes when we equalize one or the other need no longer sum to 1. The results suggest that equalizing patient characteristics across areas would reduce geographic differences by 27%, whereas equalizing place characteristics would reduce them by 72%. As a separate exercise, we show in Online Appendix Table 8 that simply estimating the model in levels also yields a somewhat lower patient share (23%). We also show in the same table that if we define the outcome to be whether the patient is in the top X% of the national distribution of utilization, the patient share ranges from 17% to 51%, with some trend toward lower patient shares at the top of the distribution. Comparing these results directly to our main estimates is difficult, but they suggest that the relative importance of patients may be somewhat lower at higher percentiles of the utilization distribution.

Robustness. We explore the robustness of our results along a number of dimensions in Online Appendix Section 4. Here, we briefly summarize the main conclusions.

First, we show that our results are robust to using observations for movers only in small windows around the move year, in the spirit of a regression discontinuity. This suggests that our results are not driven by differential utilization trends for movers that are related systematically to the origin and destination.

Second, consistent with our assumption that patient and place effects are time-constant, we show that our conclusions are similar if we estimate the model separately on subperiods of

the data, or allow explicitly for patient and place-specific trends in utilization.

Third, we relax our additive separability assumption by allowing different place effects γ_j for each quartile of patient age. We see this as a step toward a model where γ_j and α_i interact, since patient age is the one of the strongest observable predictors of patient demand. This model yields similar results. As further support for additive separability, we follow Card, Heining, and Kline (2013) and show that the increase in R^2 when we fully saturate the model with patient-place fixed effects is relatively small.

Fourth, we show that the qualitative conclusions are robust to excluding all observations for patients who exit or enter the sample due to death or HMO status, suggesting that any bias from selective attrition is likely small.

Fifth, to assess robustness to our market definition, we show that results are similar if we define markets at higher or lower levels of geography and if we include only movers who cross state lines or census region boundaries.

Finally, we explore robustness to other implementation decisions. This includes using alternative definitions of movers (such as including individuals who move multiple times within our sample period or varying the criteria used to define valid moves), using alternative dependent variables (such as expenditure rather than utilization or other functional forms for utilization), excluding nonmovers from the estimation altogether, dropping age and relative year as covariates, and excluding moves to Florida, Arizona, and California.

IV.C. Other Outcomes

In Table IV, we replicate our main decomposition results for the following alternative components of annual utilization: dummies for whether a patient has (i) seen a primary care physician, (ii) seen a specialist, (iii) been hospitalized, or (iv) visited the emergency room; the log of 1 plus (i) the number of diagnostic tests the patient received, (ii) the number of imaging tests the patient received, (iii) the number of preventive care measures the patient received, (iv) the number of different doctors the patient saw, (v) inpatient utilization, (vi) outpatient utilization, (vii) emergency room utilization, and (viii) other utilization. Detailed definitions of these measures are provided in Online Appendix Section 2.2. For each row, we reestimate equation (2) with y_{ijt} defined to be the

measure in question. We then report the sample mean of the outcome measure, the difference in the mean of the outcome measure between above- and below-median HRRs, and the share of this difference due to patients, where the partitions into above- and below-median are defined based on the outcome measure in question and so vary across rows.

The results suggest that the patient share varies from a low of 0.09 for diagnostic tests to a high of 0.71 for emergency room visits.[26] A natural hypothesis is that this variation reflects the degree of patient involvement in decision making: the outcomes for which we find a large patient share—preventive care and emergency room visits, for example—tend to be ones where we might think patients have a significant amount of discretion, whereas the outcomes for which we find a smaller patient share—diagnostic tests, imaging tests, and inpatient care, for example—tend to be ones where we might think more discretion lies with physicians.

We can make this hypothesis more precise through a slight revision of our model. In equation (1), we assume that physicians unilaterally choose the level of care y_{it} based on their perception $\tilde{u}_j()$ of the patient's utility. Suppose instead that for a particular outcome m (diagnostic tests, emergency room visits, etc.), the quantity chosen maximizes a combination of the physician's and the patient's incentives:

$$(8)\quad y_{it}^{m} = \arg\max_{y} \kappa_m u(y|h_{it},\eta_i) + (1-\kappa_m)\left[\tilde{u}_j(y|h_{it},\eta_i) - PC_{jt}(y)\right].$$

Here the parameter κ_m reflects the weight of the patient relative to the physician in decision making or bargaining. The patient share $S^{pat}(j,j')$ is increasing in κ_m holding other parameters of the model constant, as long as $S^{pat}(j,j') \in [0,1]$. For outcomes such as emergency room visits, where the patient may decide whether to go unilaterally, we might expect κ_m to be high, consistent with the high observed patient share; for outcomes such as diagnostic or imaging tests, we would expect κ_m to be lower,

26. Online Appendix Figure 10 shows event-study graphs parallel to Figure VI for each of these outcomes. As with our main utilization measure, we observe in each case large discontinuous changes on move and relatively small trends pre- or postmove. The size and direction of the pre- and post-trends vary somewhat across outcomes, and so as a robustness check Online Appendix Table 7 shows the results are similar in magnitude when we limit the estimation sample for movers to one year pre- or postmove.

consistent with the low observed patient share. None of these modifications substantively change our conclusions.

V. Correlates of Patient and Place Effects

Both the patient and place components of utilization could reflect a range of underlying economic primitives. In Section IV.A, we argued that the evidence looks inconsistent with patient effects being primarily driven by either habit formation or persistence of specific treatments started premove. Here, we provide additional evidence on mechanisms by exploring the observable correlates of our estimated place effects ($\hat{\gamma}_j$) and average patient effects ($\hat{y}_j^*$).

We focus on observables that proxy for the main demand and supply factors suggested by the model in Section II.A. We present detailed definitions, data sources, and summary statistics for these measures in Online Appendix Section 3.2.

For places, we are interested in proxies for net private costs (PC_{jt}) and physician beliefs (λ_j). The literature discussed in Section II.B suggests a number of economic drivers of the former, including physical capital, human capital, and organizational form. As proxies, we use hospital beds per capita, primary care physicians per capita, specialists per capita, hospital quality of care scores, and the share of hospitals that are nonprofit. To capture physician beliefs about appropriate practice style, we draw on survey-based measures from Cutler et al. (2015).[27] They present 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 cardiologists, respectively, 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.

For patients, we are interested in proxies for average treatment preferences (η_i) and health status (h_{it}) in each HRR. We include survey-based measures of preferences collected by Cutler et al. (2015): the shares of patients in each HRR who would request additional tests or specialist referrals even if not

27. We are grateful to the authors for sharing these data, as well as the patient survey data discussed below.

recommended by their primary care physician ("have unneeded tests" and "see unneeded cardiologist") and the shares that would choose relatively more or less aggressive end-of-life care ("aggressive" and "comforter" patient shares). We also use average patient age, race, and sex from the Medicare claims data, as well as median household income and high school completion rates from census data. These demographics could proxy for both preferences and health. Finally, we include standard health measures derived from the diagnoses recorded in Medicare claims. These are the log of the Hierarchical Condition Categories (HCC) score, and the log of 1 plus the count of a patient's chronic conditions, the Charlson Comorbidity Index, and the count of a patient's Iezzoni Chronic Conditions.

An important challenge arises with interpreting these standard health measures. Although they are intended to capture the underlying health status of a patient, Song et al. (2010) document evidence that they in fact include a large measurement error component that varies systematically by place. In places that treat more aggressively, a given underlying condition is more likely to be diagnosed, and more likely to be recorded in claims conditional on being diagnosed. This makes interpreting the correlations of the raw health measures with patient and place components difficult, and it has led the literature on geographic variation to be cautious about inferring the role of patient health.

The empirical strategy developed in this article gives us a way to extend the approach developed by Song et al. (2010) and purge these measures of the place-specific measurement error component. We model a given observed health measure h_{ijt}^{meas} as the sum of true health h_{it} and a place-specific measurement error ξ_{ijt} that in turn depends on place and year fixed effects and an orthogonal mean-zero error term. This yields a model with the same functional form as equation (2), where the dependent variable is now h_{ijt}^{meas}. Estimates of this model allow us to difference out the place-specific measurement error and recover a corrected estimate $\hat{h}_{it}$ of patient health. Details of this exercise, and the associated event-study figure showing changes in measured health around moves, are in Online Appendix Section 1; we estimate that about 50% of the geographic variation in measured health reflects place-specific measurement error. We use the corrected health measures throughout the analysis that follows.

Before turning to the results, we note that because the correlations we are looking at are cross-sectional, they are likely to

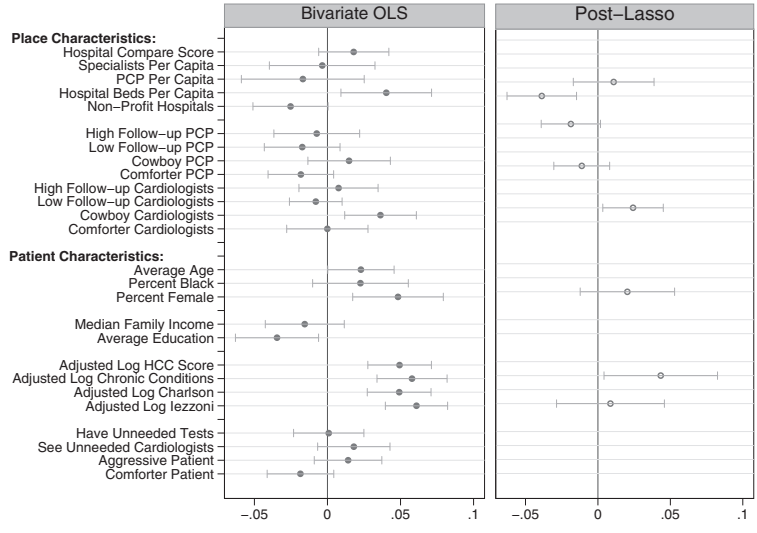

FIGURE VII

Correlates of Average Place Effects

Figure shows bivariate OLS regression results (left panel) and post-Lasso multivariate regression results (right panel) of HRR-level place effects on a set of HRR-level characteristics. All covariates have been standardized to have mean 0 and standard deviation 1. To obtain the post-Lasso estimates, we first run a Lasso regression on the full set of covariates, with the penalty level chosen by a 10-fold cross-validation to minimize mean squared error. We then run an OLS regression on the set of covariates chosen by the Lasso regression. The sample in both panels is the 96 HRRs for which all covariates are available. Horizontal bars show 95% confidence intervals. Hospital Compare Score approximates hospital quality using timely and effective care measures publicly reported by CMS. Specialists per capita, PCP per capita, and hospital beds per capita count specialists, primary care physicians, and hospital beds per thousand residents, respectively. Non-profit hospitals is the percent of hospitals that are nonprofit. Physician preference measures are drawn from survey responses of PCPs and cardiologists from Cutler et al. (2015); physicians classified as high follow-up or low follow-up recommend follow-up visits more (or less) frequently than clinical guidelines suggest; physicians classified as cowboy recommend care more intensive than guidelines suggest, and those classified as comforter recommend palliative care for severely ill patients. Average age, percent black, and percent female are computed among all patients in our baseline sample of Medicare beneficiaries. Median family income is the median income of households across zip codes in each HRR taken from census data. Average education is the percent of the 25 and older population with a high school degree as computed from census data. The health variables are all the estimated patient components of a series of health measures as described in Online Appendix Section 1. The patient preferences variables are drawn from Cutler et al. (2015) and detail Medicare beneficiaries' survey responses to desired care in hypothetical cases; have unneeded tests and see unneeded cardiologists are the fraction of patients who would desire such treatment regimens; aggressive patient provides the fraction of patients who would like aggressive end-of-life care; and comforter patient provides the fraction of patients who would like palliative end-of-life care even if it shortens their life.

pick up the kind of long-term endogenous response of physical and human capital to patient demand documented in Chandra and Staiger (2007). For example, places that start out with sicker patients may build more hospitals and hire doctors trained in aggressive treatments. This would imply that the place component of utilization could be correlated with patient proxies such as health (as seen in Figure VII), and that the patient component could be correlated with place proxies such as number of hospitals or share of cowboy physicians (as we will now see).

Figure VII summarizes the correlates of the estimated place effects ($\hat{\gamma}_j$). Each row represents a different proxy variable. The points in the left panel are coefficients from separate bivariate OLS regressions. The points in the right panel are coefficients from post-Lasso multivariate OLS, where a subset of the variables shown in the left panel were selected in a first-stage Lasso regression (Belloni and Chernozhukov 2013).[28] All variables are standardized so that the coefficients report the association between a 1 standard deviation change in the covariate and the respective outcome. The sample is limited to the 96 HRRs (representing about 60% of our baseline sample) for which the measures of patient preferences and physician beliefs are available from Cutler et al. (2015). We present results using the full set of HRRs and omitting the Cutler et al. (2015) measures in Online Appendix Table 13.

The results show that HRRs with more hospital beds per capita, a higher share of cowboy cardiologists, and a lower share of nonprofit hospitals are all associated with statistically significantly higher place effects. HRRs with more female patients, less educated patients, and sicker patients by any measure are also associated with higher place effects. The post-Lasso multivariate regressions show that the shares of cowboy cardiologists and the number of nonprofit hospitals remain significant, as does patient health. The evidence is therefore consistent with past literature highlighting physician beliefs (Cutler et al. 2015) and endogenous responses to patient demand (Chandra and Staiger 2007) as key drivers.

28. In the first stage we select variables using Lasso regression with a penalty chosen by 10-fold cross-validation to minimize the mean squared error; in the second stage we report the coefficients and standard errors from multivariable OLS on the selected covariates. Online Appendix Figures 17 and 18 show the set of covariates that would have been chosen for alternative values of the penalty.

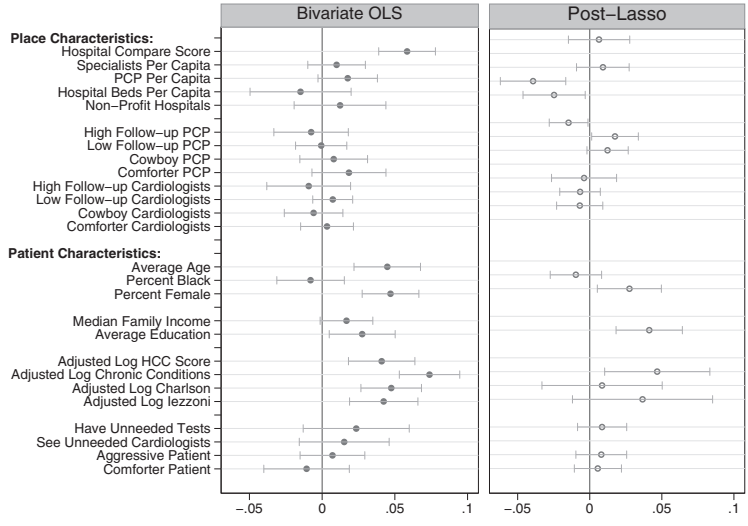

FIGURE VIII

Correlates of Average Patient Effects

Figure shows bivariate OLS regression results (left panel) and post-Lasso multivariate regression results (right panel) of HRR-level patient effects on a set of HRR-level characteristics. Procedure and explanatory variables are the same as in Figure VII.

Figure VIII shows analogous results for the estimated average patient effects ($\hat{y}_j^*$). Consistent with intuition, we see that HRRs with older patients, sicker patients (by any of the health measures), female patients, and higher socioeconomic status patients (as measured by income or education) are associated with statistically significant higher average patient demand. The first two of these seem clearly related to health status; the latter two could reflect differences in preferences. We also see that places where hospitals with higher quality of care ("Compare") scores are also associated with statistically significant higher average patient effects, an association that might reflect reverse causality. The post-Lasso multivariate regression confirms the intuitive pattern for the patient characteristics, with sex, education, and health all remaining significant; several place characteristics also have significant coefficients, perhaps again partly reflecting reverse causality.

One of the clearest takeaways from this analysis is that patient health status is a strong predictor of both components of utilization. Were we to take the leap and view these relationships

as causal, the bivariate correlation in Figure VIII would suggest that equalizing the chronic conditions measure across areas would reduce the gap in the patient component of log utilization between above- and below-median areas from 0.13 to 0.04.[29] Recalling that the overall gap in utilization is 0.28 (Table II), this implies that $\frac{0.13-0.04}{0.28}$ or 32% of the overall gap could be attributed to the demand-side differences in observable patient health.

Following recent literature (e.g., Zuckerman et al. 2010), we can refine this estimate by using individual-level rather than cross-area variation to determine the coefficient linking health status to utilization. That is, rather than using the bivariate OLS coefficient from Figure VIII, we can use the coefficient from an individual-year-level panel regression of log utilization of nonmovers on the adjusted health measure in question along with HRR and year fixed effects.[30] Here, the main difference relative to past work based on similar regressions is the ability to correct for measurement error in observed health.

The results of this decomposition are shown in Table V. To emphasize the importance of our measurement error correction, the first four rows show results using the raw, unadjusted health measures. They suggest that failing to account for measurement error would lead us to attribute between 40% and 80% of the log utilization difference between above- and below-median areas to patient health. (The bottom end of this range is close to the estimate in Zuckerman et al. 2010.) The bottom four rows show results using our adjusted measures. We find that between 22% and 37% of the log utilization difference can be explained by patient health.

29. The standardized bivariate coefficient on log chronic conditions in Figure VIII is 0.074. Converting this to unstandardized units by dividing by the standard deviation of the health measure (0.042) yields a coefficient of 1.76. The average of the log chronic conditions measures is 1.25 in above-median HRRs and 1.20 in below median HRRs, so we predict that this gap falls by $1.76 \times (1.25 - 1.20)$ or 0.088.

30. That is, for the sample of nonmovers, we estimate:

$$y_{it} = \gamma_j^{obs} + \tau_t^{obs} + \hat{h}_{it}\phi + \epsilon_{ijt}^{obs},$$

where γ_j^{obs}, τ_t^{obs}, and ε_{ijt}^{obs} are, respectively, area fixed effects, year fixed effects, and an error term, distinct from those in equation (2), and $\hat{h}_{it}$ is one of our corrected health measures constructed as described in Online Appendix Section 1. We then estimate the share of the utilization gap between two areas R and R' attributable to patient health as $\frac{\varphi\left(\overline{h}_R - \overline{h}_{R'}\right)}{\overline{y}_R - \overline{y}_{R'}}$ where $\overline{h}_R$ is the average of $\hat{h}_{it}$ in R, with the average computed analogously to $\overline{y}_R$.

TABLE IV

COMPONENTS OF UTILIZATION

	Utilization measure	(1) Mean of utilization measure	(2) Above/below median difference in utilization measure	(3) Share due to patients	
(1)	Baseline: log(utilization)	7.193	0.283	0.465	(0.027)
(2)	Seen a primary care physician	0.884	0.042	0.452	(0.027)
(3)	Seen a specialist	0.815	0.051	0.322	(0.024)
(4)	Any hospitalization	0.226	0.037	0.410	(0.034)
(5)	Any emergency room visit	0.346	0.045	0.714	(0.031)
(6)	Log(# of diagnostic tests)	1.449	0.550	0.092	(0.008)
(7)	Log(# of imaging tests)	0.842	0.220	0.142	(0.014)
(8)	Log(# of preventive care measures)[a]	1.376	0.098	0.611	(0.018)
(9)	Log(# of different doctors seen)	1.525	0.113	0.392	(0.016)
(10)	Log(inpatient utilization)[b]	2.004	0.340	0.242	(0.035)
(11)	Log(outpatient utilization)[b]	6.890	0.193	0.358	(0.031)
(12)	Log(emergency room utilization)[b]	2.296	0.352	0.639	(0.031)
(13)	Log(other utilization)[b]	3.430	0.957	0.124	(0.010)

Notes. Table reports the share of the difference in utilization between above and below median HRRs due to patients, analogous to column (1) of Table II, with the dependent variable y_{ijt} defined to be various components of utilization. The partition of HRRs into above and below median groups is based on the utilization of individuals in the baseline sample and differs in each row according to the definition of utilization used. Column (1) reports the mean of the utilization measure for the given sample. Column (2) reports the difference in the average utilization measure between above and below median HRRs ($\hat{y}_R - \hat{y}_{R'}$). Column (3) reports the share of the difference in column (2) that is due to patients ($S_{pat}(R, R')$). All log outcome measures are the log of the outcome plus one. Online Appendix Table 11 shows the percent with zero for each of these outcomes. Standard errors (in parentheses) are calculated using a bootstrap with 50 repetitions at the patient level. The sample size is the same as in Table II.

[a]# of preventive care measures is a count of the number of the following preventive treatments the patient received in the past year: ambulatory care, eye screening, hemoglobin test, lipid screen, cardio screen, diabetes management, pelvic screen, bone mass test, colorectal cancer screening, and flu shot, or in the past two years: mammogram, Pap test, and prostate cancer screening.

[b]These four measures are mutually exclusive and exhaustive.

We have already stressed that this is a predictive rather than causal relationship. A further important caveat is that our measurement error correction only adjusts for place-level sources of error. Our corrected measures could still include other forms of measurement error. In particular, patients with a preference for visiting the doctor more often may be more likely to have their conditions recorded in claims, leading our adjusted measure to partly reflect differences in preferences.

Our findings on health may be of some interest beyond the study of geographic variation. We provide a way to quantify and correct for the measurement error identified by Song et al. (2010)

TABLE V

VARIATION IN LOG UTILIZATION EXPLAINED BY PATIENT HEALTH

		(1) Share of above/below median utilization difference due to patient health
	Raw health measure	
(1)	Log(HCC score)	0.435
(2)	Log(Charlson Comorbidity Index)	0.483
(3)	Log(# of Iezzoni chronic conditions)	0.483
(4)	Log(# of chronic conditions)	0.794
	Corrected health measure	
(5)	Log(HCC score)	0.220
(6)	Log(Charlson Comorbidity Index)	0.242
(7)	Log(# of Iezzoni chronic conditions)	0.256
(8)	Log(# of chronic conditions)	0.371

Notes. Table reports shares of the difference in average log utilization between above-median and below-median utilization HRRs explained by observable patient health ($\hat{S}^{obs}_{pat}(R, R')$). Rows (5)–(8) use the patient component (h_{it}) of health measures estimated from Online Appendix equation (1). All log outcome measures are the log of the outcome plus 1, except the HCC score, which is simply the log of the outcome (there are no zeros). Online Appendix Table 11 shows the percent with zero for each of these outcomes. The sample size is the same as in Table II in rows (1)–(7). In row (8), the sample also excludes the year 1998, as chronic conditions are not observed in that year ($N = 14{,}598{,}443$ patient-years).

and show that the endogenous error component accounts for a large share of the geographic variation in measured health. Our corrected measures may have other applications in the large literature that uses health status measures as inputs into risk adjustments.

VI. CONCLUSION

Looking at over-65 Medicare beneficiaries, we find robust evidence that 40–50% of geographic variation in the log of health care utilization is due to fixed characteristics of patients that they carry with them when they move. The remaining 50–60% of variation is due to place-specific factors. Patients matter more for outcomes such as emergency room visits, where they have substantial discretion, and they matter less for outcomes such as diagnostic and imaging tests, where the physician is the main decision maker.

Our analysis of mechanisms suggests that a substantial share of the patient component can in turn be attributed to differences across areas in average health status. The remainder

likely reflects a combination of both unmeasured health and preferences. We find intriguing correlations with our estimated place effects, which suggest, among other things, that areas with higher place components of log utilization also have physicians who believe in a more aggressive practice style, fewer nonprofit hospitals, and a sicker population.

These findings both reinforce and refine the conclusions of the existing literature on geographic variation. On the one hand, they confirm that supply-side variation plays an important role and argue against claims that the role of the supply side is negligible. On the other hand, they clarify that a large component of variation is in fact due to differences in patient health and suggest that both patient preferences and unmeasured health differences may play a larger role than conventional wisdom would suggest.

Our results do not permit us to draw strong conclusions about welfare. Variation on the supply side may be, but need not be, inefficient. The correlation of our place component with the health of the local population suggests that at least some of the former may be due to endogenous responses of human and physical capital. In the presence of such adjustments, we would expect nontrivial supply-side variation even in a first-best world. Conversely, demand-side variation need not be fully efficient; patient demand may reflect misinformation or behavioral biases that could potentially be corrected by policy. Continuing to drill down on the efficiency implications of geographic variation remains an important goal for future work.

Even without direct welfare implications, some of the patterns of changes we observe have potential implications for policy. For example, the event-study evidence that place effects have immediate rather than gradual effects on movers implies that changing supply-side factors such as doctor practice styles could have large effects in the short run. At the same time, the lack of postmove convergence suggests that policies aimed at changing demand-side factors such as patient preferences may have at best very gradual effects, at least among the 65 and older population (Moses et al., 2013).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY AND NATIONAL BUREAU OF ECONOMIC RESEARCH
STANFORD UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH

MASSACHUSETTS INSTITUTE OF TECHNOLOGY AND NATIONAL BUREAU OF ECONOMIC RESEARCH

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at QJE online (qje.oxfordjournals.org).

REFERENCES

Aaronson, Daniel, "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes," *Journal of Human Resources*, 33 (1998), 915–946.

Abowd, John M., Robert H. Creecy, and Francis Kramarz, "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," U.S. Census Bureau Center for Economic Studies Longitudinal Employer-Household Dynamics Technical Papers 2002-06, 2002.

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 67 (1999), 251–333.

Ashton, Carol M., Nancy J. Peterson, Julianne Souchek, Terri J. Menke, Hong-Jen Yu, Kenneth Pietz, Marsha L. Eigenbrodt, Galen Barbour, Kenneth Kizer, and Nelda P. Wray, "Geographic Variations in Utilization Rates in Veterans Affairs Hospitals and Clinics," *New England Journal of Medicine*, 340 (1999), 32–39.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein, "Behavioral Hazard in Health Insurance," *Quarterly Journal of Economics*, 130 (2015), 1623–1667.

Baker, Laurence C., M. Kate Bundorf, and Daniel P. Kessler, "Patients' Preferences Explain a Small but Significant Share of Regional Variation in Medicare Spending," *Health Affairs*, 33 (2014), 957–963.

Baker, Laurence C., Elliott S. Fisher, and John E. Wennberg, "Variations in Hospital Resource Use for Medicare and Privately Insured Populations in California," *Health Affairs*, 27 (2008), w123–w134.

Barnato, Amber R., M. Brooke Herndon, Denise L. Anthony, Patricia M. Gallagher, Jonathan S. Skinner, Julie P. W. Bynum, and Elliott S. Fisher, "Are Regional Variations in End-of-Life Care Intensity Explained by Patient Preferences? A Study of the US Medicare Population," *Medical Care*, 45 (2007), 386–393.

Becker, Gary S., and Kevin M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy*, 96 (1988), 675–700.

Belloni, Alexandre, and Victor Chernozhukov, "Least Squares after Model Selection in High-Dimensional Sparse Models," *Bernoulli*, 19 (2013), 521–547.

Bronnenberg, Bart J., Jean-Pierre H. Dube, and Matthew Gentzkow, "The Evolution of Brand Preferences: Evidence from Consumer Migration," *American Economic Review*, 102 (2012), 2472–2508.

Card, David, Joerg Heining, and Patrick M. Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128 (2013), 967–1015.

Chandra, Amitabh, David Cutler, and Zirui Song, "Who Ordered That? The Economics of Treatment Choices in Medical Care," in *Handbook of Health Economics*, vol. 2, Mark V. Pauly, Thomas G. McGuire, and Pedro P. Barros, eds. (Amsterdam: Elsevier, 2012).

Chandra, Amitabh, Lindsay Sabik, and Jonathan S. Skinner, "Cost Growth in Medicare: 1992 to 2006," in *Explorations in the Economics of Aging*, David A. Wise, ed. (Chicago: University of Chicago Press, 2011).

Chandra, Amitabh, and Douglas O. Staiger, "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy*, 115 (2007), 103–140.

Chernew, Michael, Lindsay Sabik, Amitabh Chandra, Theresa Gibson, and Joseph Newhouse, "Geographic Correlation between Large-Firm Commercial Spending and Medicare Spending," *American Journal of Managed Care*, 16 (2010), 131–138.
Chetty, Raj, John N. Friedman, Soren Leth-Petersen, Torben Nielsen, and Tore Olsen, "Active vs. Passive Decision and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark," *Quarterly Journal of Economics*, 129 (2014), 1141–1219.
Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104 (2014), 2633–2679.
Chetty, Raj, John N. Friedman, and Emmanuel Saez, "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, 103 (2013), 2683–2721.
Congressional Budget Office, "Geographic Variation in Health Care Spending," Washington, DC, Congress of the United States, CBO, 2008.
Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen, "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured," NBER Working Paper 2015.
Currie, Janet, and Bentley MacLeod, "First Do No Harm? Tort Reform and Birth Outcomes," *Quarterly Journal of Economics*, 123 (2008), 795–830.
Cutler, David, Jonathan Skinner, Ariel Dora Stern, and David Wennberg, "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," Harvard Business School Working Paper 15-090, 2015.
Dunn, Abe, Adam Hale Shapiro, and Eli Liebman, "Geographic Variation in Commercial Medical-Care Expenditures: A Framework for Decomposing Price and Utilization," *Journal of Health Economics*, 32 (2013), 1153–1165.
Fernandez, Raquel, and Alessandra Fogli, "Fertility: The Role of Culture and Family Experience," *Journal of the European Economic Association*, 4 (2006), 552–561.
Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder, "The Implications of Regional Variations in Medicare Spending: The Content, Quality, and Accessibility of Care. Part 1," *Annals of Internal Medicine*, 138 (2003a), 273–287.
———, "The Implications of Regional Variations in Medicare Spending: The Content, Quality, and Accessibility of Care. Part 2," *Annals of Internal Medicine*, 138 (2003b), 288–298.
Gawande, Atul, "The Cost Conundrum: What a Texas Town Can Teach Us about Health Care," The New Yorker, 2009.
Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn G. Andrews, Jonathan S. Skinner, and Jason M. Sutherland, "Prices Don't Drive Regional Medicare Spending Variations," *Health Affairs*, 29 (2010), 537–543.
Lee, Thomas H., and James J. Mongan, Chaos and Organization in Health Care (Cambridge, MA: MIT Press, 2009).
McPherson, Klim, P. M. Strong, Arnold Epstein, and Lesley Jones, "Regional Variations in the Use of Common Surgical Procedures: Within and Between England and Wales, Canada and the United States of America," *Social Science & Medicine A: Medical Psychology & Medical Sociology*, 15 (1981), 273–288.
Molitor, David, "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration," University of Illinois at Urbana-Champaign Working Paper, 2014, available at http://www.business.illinois.edu/dmolitor/movers.pdf.
Moses, Hamilton, David H. M. Matheson, E. Ray Dorsey, Benjamin P. George, David Sadoff, and Satoshi Yoshimura, "The Anatomy of Health Care in the United States," *Journal of the American Medical Association*, 310 (2013), 1947–1964.
Neuman, Tricia, Juliette Cubanski, Jennifer Huang, and Anthony Damico, "The Rising Cost of Living Longer: Analysis of Medicare Spending by Age for Beneficiaries in Traditional Medicare," Kaiser Family Foundation Report, 2015.

Philipson, Tomas J., Seth A. Seabury, Lee M. Lockwood, Dana P. Goldman, and Darius N. Lakdawalla, "Geographic Variation in Health Care: The Role of Private Markets," *Brookings Papers on Economic Activity*, 41 (2010), 325–361.
Rettenmaier, Andrew J., and Thomas R. Saving, "Perspectives on the Geographic Variation in Health Care Spending," *National Center for Policy Analysis*, 2009.
Sheiner, Louise, "Why the Geographic Variation in Health Care Spending Can't Tell Us Much about the Efficiency or Quality of Our Health Care System," *Brookings Papers on Economic Activity* (2014), 1–72.
Skinner, Jonathan, "Causes and Consequences of Regional Variations in Health Care," in *Handbook of Health Economics*, vol. 2, Mark V. Pauly, Thomas G. McGuire , and Pedro P. Barros, eds. (Amsterdam: Elsevier, 2011).
Song, Yunjie, Jonathan Skinner, Julie Bynam, Jason Sutherland, and Elliott Fisher, "Regional Variations in Diagnostic Practices," *New England Journal of Medicine*, 363 (2010), 45–53.
Subramanian, Usha, Morris Weinberger, George J. Eckert, Gilbert J. L'Italien, Pablo Lapuerta, and William Tierney, "Geographic Variation in Health Care Utilization and Outcomes in Veterans with Acute Myocardial Infarction," *Journal of General Internal Medicine*, 17 (2002), 604–611.
Weinstein, James N., Kristen K. Bronner, Tamara Shawver Morgan, and John E. Wennberg, "Trends and Geographic Variations in Major Surgery for Degenerative Diseases of the Hip, Knee, and Spine," *Health Affairs*, 23 (2004), var81–var89.
Welch, H. Gilbert, Sandra M. Sharp, Dan J. Gottleib, Jonathan S. Skinner, and John E. Wennberg, "Geographic Variation in Diagnosis Frequency and Risk of Death among Medicare Beneficiaries," *Journal of the American Medical Association*, 305 (2011), 1113–1118.
Xu, Fang, Machell Town, Lina Balluz, William Bartoli, Wilmon Murphy, Pranesh Chowdhury, William Garvin, Carol Pierannunzi, Yuna Zhong, Simone Salandy, Candace Jones, and Carol Crawford, "Surveillance for Certain Health Behaviors Among States and Selected Local Areas—United States, 2010," Morbidity and Mortality Weekly Report (MMWR), Surveillance Summaries, Centers for Disease Control and Prevention, 2013.
Zuckerman, Stephen, Timothy Waldman, Robert Berenson, and Jack Hadley, "Clarifying Sources of Geographic Differences in Medicare Spending," *New England Journal of Medicine*, 363 (2010), 54–62.