NBER WORKING PAPER SERIES

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

Amy Finkelstein
Matthew Gentzkow
Heidi Williams

Working Paper 20789
http://www.nber.org/papers/w20789

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2014

We are grateful to Daron Acemoglu, Raj Chetty, David Cutler, Joe Doyle, Liran Einav, Ben Handel, Nathan Hendren, Pat Kline, Matt Notowidigdo, Adam Sacarny, Jesse Shapiro, Jonathan Skinner, Doug Staiger, and seminar participants at Dartmouth, the NBER Aging Meeting, the NYC health economics seminar, Stanford, the University of Chicago, 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, and Tony Zhang for excellent research assistance and to the National Institute on Aging (Finkelstein, R01-AG032449) 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w20789.ack

Sources of Geographic Variation in Health Care: Evidence from Patient Migration
Amy Finkelstein, Matthew Gentzkow, and Heidi Williams
NBER Working Paper No. 20789
December 2014
JEL No. H51,I1,I11

ABSTRACT

We study the drivers of geographic variation in US 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. We find that 40-50 percent of geographic variation in utilization is attributable to patient demand, with the remainder due to place-specific supply factors. Demand variation does not appear to result from differences in past experiences, and is explained to a significant degree by differences in patient health.

Amy Finkelstein
Department of Economics, E17-228
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
afink@mit.edu

Heidi Williams
Department of Economics, E17-222
MIT
77 Massachusetts Avenue
Cambridge MA 02139
and NBER
heidiw@mit.edu

Matthew Gentzkow
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
gentzkow@chicagobooth.edu

1 Introduction

Health care utilization varies widely across the United States (Fisher et al. 2003*a*; 2003*b*). 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 to $8,714 in nearby and demographically similar El Paso, TX.[1] Similar geographic variation is observed in the frequency of specific treatments (Chandra et al. 2012) and in measures of total health care utilization that adjust for regional variation in prices (Gottlieb et al. 2010). Higher area-level utilization is not generally correlated with better patient outcomes.[2]

Understanding what drives this geographic variation has first-order implications for policy. If high-spending 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). If, on the other hand, patients in high-spending areas are simply sicker or prefer more intensive care, such policies could be ineffective or counterproductive.

In this paper, 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 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-spending Miami to low-spending Minneapolis. If all of the spending difference between these cities arises from supply-side differences like doctor incentives, we would expect the migrant's spending to drop immediately following the

[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.

[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.

move, to a level similar to other patients of the low-spending doctors in Minneapolis. If all of the spending difference reflects the demand-side reality that residents of Miami are sicker, we would expect the migrant's spending to remain constant after the move, at a level similar to the typical person in Miami. Where the observed spending change falls between these two extremes identifies the relative importance of demand and supply factors.

We implement this strategy using claims data for a 20 percent sample of Medicare beneficiaries from 1998 to 2008.[4] In our baseline model, the log of a patient's annual health care utilization arises from 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. 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). The model allows for the possibility that migrants have systematically different utilization levels from non-migrants, and that these levels are correlated with the migrant's origin and destination region. It also allows for arbitrary differences in utilization trends of migrants relative to non-migrants. 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 pre-move, and no systematic adjustment post-move. 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 percent of the difference in log utilization between above- and below-median areas is due to demand-side factors. The share is similar for differences between the top and bottom quartiles, deciles, or ventiles. The share is also

[4]Studying Medicare patients is appealing due to the availability of high-quality, rich data on large numbers of beneficiaries, and the relatively geographically uniform insurance environment. The literature has explored—to the extent feasible given existing data—whether the geographic variation in health care utilization that has been extensively documented in the Medicare setting exists in other settings as well. Regional variation appears to be the norm; geographic variation has been documented in the US Veterans Affairs system (Ashton et al. 1999; Congressional Budget Office 2008; Subramanian et al. 2002), in private insurance markets (Baker et al. 2008; Chernew et al. 2010; Dunn et al. 2013; Philipson et al. 2010; Rettenmaier and Saving 2009), and in other countries including the UK and Canada (McPherson et al. 1981), although the magnitudes of this variation and the correlation with Medicare variation is the subject of some debate in these studies.

similar when we isolate differences between the very highest-spending areas, such as McAllen or Miami, and the very lowest-spending areas, such as El Paso or Minneapolis.

We replicate this 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 percent to 71 percent. Although we do not have a formal framework in which to interpret this heterogeneity, 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, for example.

In the final section of the paper, we consider what underlying economic primitives might drive the differences in patient demand we estimate. We begin by showing that patient demographics (age, race, and sex) predict relatively little of the variation in demand. We also show that a mechanical form of state dependence due to persistence of diagnosis or treatment likewise seems to play at most a small role.

We then turn to considering the role of habit formation, in the sense of Becker and Murphy (1988): to what extent do differences in patients' demand for care today result from differences in the care they consumed in the past? We present two pieces of evidence that suggest habit formation is limited, at least over the 10-year horizon of our data. First, we find no evidence that utilization continues to adjust in the years after a move. Such adjustment is a robust prediction of habit models, and is the key feature of the data that identifies empirical studies of habit formation such as Bronnenberg et al. (2012). Second, we find that older patients actually change their utilization more when they move than younger patients, at odds with many habit models that predict older patients' larger stock of past consumption would cause them to adjust their utilization by less.

The finding that habit formation is limited speaks to questions beyond our demand-supply decomposition. It suggests the demand-side differences we observe in the propensity to consume care are likely to be stable, and not easily affected by policy. On the other hand, the fact that utilization adjusts sharply on move suggests that policies which affect supply-side factors such as doctors' practice patterns can have immediate impacts.

The last mechanism we consider is patient health: are patients in Miami simply sicker than patients in Minneapolis? An important challenge is that measured health status may itself be

endogenous to place (Song et al. 2010): because standard health status measures are derived from claims data, a given condition may be more likely to be recorded in a high-intensity area. To address this, we extend our mover-based empirical strategy to separate patient-specific variation in health from the component of measurement endogenous to place. We then ask how much of the geographic variation in log utilization can be explained by the patient component of health.

Consistent with Song et al. (2010), we find the effect of endogenous measurement to be substantial. Once we adjust for this, we find that patient health can explain 22 to 37 percent of the gap in log utilization between above- and below-median areas—or 47 to 80 percent of our overall patient share—depending on the health measure used. The remainder of the patient share presumably reflects other factors such as information, preferences, or unmeasured dimensions of patient health.

Like our findings on habit formation, our results on patient health have broader relevance. The larger the role played by patient health, the less effective may be policies aimed at other demand-side drivers such as patient preferences or beliefs. Our results also extend the analysis of the endogeneity first identified by Song et al. (2010) to develop a method for separating the endogenous and patient-specific components of these measures; this may have other applications in the large literature that uses these health measures as inputs for risk adjustment. At the same time, there are some important limitations to our health results. Although our measure addresses one form of endogeneity in measured health, it may still pick up the effect of patient characteristics correlated with health, as well as patient-specific factors that affect the likelihood a given health condition is measured in claims.

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.[5] All of these studies infer the role of demand-side factors from the explanatory power of patient observables. The majority of recent work concludes that the role of patients is limited, and that most of the variation likely originates on the supply side. A review by Chandra et al. (2012), for example, concludes: "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

[5]See Skinner (2011) and Chandra et al. (2012) for reviews, and Cutler et al. (2013) and Baker et al. (2014) for more recent contributions.

and preferences do not explain much of the differences across areas." (p. 402) An exception to this consensus is Sheiner (2014), who argues that patients may explain most or all of the variation.

Our strategy has two important advantages relative to this literature. First, we can capture the effect of both observed and unobserved patient characteristics, since both will be reflected in persistence of utilization following a move. Second, our approach correctly isolates variation arising from differences in patient health, even if observable measures of these factors are endogenous to supply-side measurement differences.

Like past decompositions, ours is not sufficient to draw strong conclusions about the efficiency of observed geographic variation. Though it may be tempting to see supply-driven heterogeneity as evidence of waste, such variation could reflect different allocations of physical or human capital, and so be consistent with efficiency (Chandra and Staiger 2007). Conversely, demand-driven heterogeneity could reflect patient misinformation, and so contribute to inefficiency. We view our findings as both 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 related are Song et al. (2010), who looks at how health measures change around patient moves, and Molitor (2014), who looks at physician migration, estimating how 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 et al.'s (2013) study of German workers and firms. Other work uses migration to study neighborhood effects on children (Aaronson 1998), cultural assimilation of immigrants (Fernandez and Fogli 2006), brand preferences (Bronnenberg et al. 2012), tax reporting (Chetty et al. 2013), teacher value added (Chetty et al. 2014*a*), and retirement savings decisions (Chetty et al. 2014*b*).

Section 2 introduces our model and estimation strategy. Section 3 describes our data and presents summary statistics. Section 4 presents our main analysis of the role of demand and supply factors in explaining geographic variation in health care utilization. Section 5 explores potential mechanisms for the role of patients. Section 6 concludes.

2 Empirical Model and Identification

2.1 Model

We index patients by i, geographic areas by j, and years by t. Some patients are "non-movers" who live in one area throughout the sample, while others are "movers" whose area changes exactly once. (In the empirical analysis, we exclude multiple movers.) For a mover i who moves during year t_i^*, we define the year relative to move to be $r(i,t) = t - t_i^*$. The outcome of interest is y, which in our main specifications will be the log of total health care utilization. We discuss the precise definition of these variables in the context of our data in Section 3 below.

We assume outcome y for patient i who lives in area j throughout year t is given by:

$$y_{ijt} = \alpha_i + \gamma_j + \tau_t + \rho_{r(i,t)} + x_{it}\beta + \varepsilon_{ijt}. \tag{1}$$

Here α_i, γ_j, and τ_t are fixed effects for patient, area, and year respectively, and x_{it} is a vector of time-varying patient characteristics, which in our baseline specification is simply a series of indicator variables for five-year age bins. The term $\rho_{r(i,t)}$ is a fixed effect for movers in relative year $r(i,t)$, which we normalize to zero for non-movers. We do not model outcomes for movers in year t_i^*, when (as we show below) they spend part of the year in their origin area and part of the year in their destination; when we estimate equation (1), we omit these observations. We let

$$c_{it} = \alpha_i + \rho_{r(i,t)} + x_{it}\beta$$

denote the combined effect of patient characteristics. We assume that the error term ε_{ijt} satisfies $\mathrm{E}\left(\varepsilon_{ijt} | \alpha_i, \rho_{r(i,t)}, x_{it}, \gamma_j, \tau_t\right) = 0$.

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 $\bar{y}_{jt}$ denote the expectation of y_{it} across patients living in area j in year t, and let $\bar{y}_j$ denote the average of $\bar{y}_{jt}$ across t. Let $\bar{c}_{jt}$ and $\bar{c}_j$ denote the analogous expectations of c_{it}. Then the difference in average log utilization between any two areas j and j' is the sum of the differences of place and patient components: $\bar{y}_j - \bar{y}_{j'} = \left(\gamma_j - \gamma_{j'}\right) + \left(\bar{c}_j - \bar{c}_{j'}\right)$. When we talk

about larger groups R that consist of multiple areas j, we abuse notation by letting $\overline{y}_R$, $\overline{c}_R$, and $\overline{\gamma}_R$ denote the simple averages of $\overline{y}_j$, $\overline{c}_j$, and γ_j across areas in R.

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

$$S_{place}\left(j,j'\right)=\frac{\gamma_j-\gamma_{j'}}{\overline{y}_j-\overline{y}_{j'}} \tag{2}$$

and we define the share attributable to patients to be

$$S_{pat}\left(j,j'\right)=\frac{\overline{c}_j-\overline{c}_{j'}}{\overline{y}_j-\overline{y}_{j'}}.$$

Note that although $S_{pat}\left(j,j'\right)$ and $S_{place}\left(j,j'\right)$ sum to 1, neither need be between 0 and 1, since it is possible that $\left(\gamma_j-\gamma_{j'}\right)$ and $\left(\overline{c}_j-\overline{c}_{j'}\right)$ have opposite signs. We define $S_{pat}\left(R,R'\right)$ and $S_{place}\left(R,R'\right)$ to be the analogous shares for groups R and R'.

We let $\hat{y}_j$ denote the sample analogue of $\overline{y}_j$. Given estimates $\hat{\gamma}_j$ of the γ_j, we form a consistent estimate $\hat{c}_j=\hat{y}_j-\hat{\gamma}_j$ of $\overline{c}_j$.

2.2 Identification

The model in equation (1) is only identified if the data include movers. If all patients were non-movers, there would be no way to separate differences in the area fixed effects γ_j from differences in the average patient characteristics $\overline{c}_j$. The key to separate identification of these two components is the observed changes in utilization when patients move.[6]

To build more precise intuition, consider a simplified version of our model in which the τ_t, x_{it}, and $\rho_{r(i,t)}$ are all set to zero, and so utilization depends only on patient and place fixed effects plus the error term. 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 $\left(\gamma_j-\gamma_{j'}\right)$. 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 $\overline{c}_j$ would then be consistently estimated by $\hat{y}_j-\hat{\gamma}_j$.

Identification in the full model is similar. Identifying the τ_t and β is standard and does not

[6] 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.

rely on movers. Adding the $\rho_{r(i,t)}$ has a more substantial effect. It allows for arbitrary changes in log utilization for movers pre- and post-move, 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 $\left(\gamma_j - \gamma_{j'}\right)$, because $\Delta^j_{j'}$ would also depend on the difference between the post-move and pre-move $\rho_{r(i,t)}$. 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 $\left(\gamma_j - \gamma_{j'}\right)$ consistently as $\left(\Delta^j_{j'} - \Delta^{j'}_j\right)/2$.

Importantly, our model permits movers to differ arbitrarily from non-movers 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 example, for moves to be associated with either positive or negative health shocks. We can in principle allow substantially more flexibility, including area- or individual-specific trends, different fixed effects by sub-periods, 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 below.

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 ε_{ijt} 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 $\left(\Delta^j_{j'} - \Delta^{j'}_j\right)/2$ to be an overestimate of $\left(\gamma_j - \gamma_{j'}\right)$. As a concrete example, this could occur if a subset of movers move in order to seek intensive treatment in their destination, and they are differentially likely to move to relatively high- or low-spending areas. The pattern of pre-move trends documented in the event study analysis in Section 4.1 argues against this specific story, but we cannot in general rule out such correlated shocks.

Second, our model 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 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).[7] 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 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 distribution than a decomposition in levels would. In Section 4, we present a variety of specification and robustness checks that bear on these issues.

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 post-move. 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 bears on it at length in Section 5.3 below.

2.3 Event Study Representation

To visualize the way utilization changes when patients move, we define an alternative "event study" representation of equation (1).

To build intuition, it again helps to start with the simple case where τ_t, x_{it}, and $\rho_{r(i,t)}$ are all 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 post-move. 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

[7] To take a concrete example, suppose that patients have either one or two chronic conditions, and that places spend either five or ten thousand dollars 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 $\alpha_i + \gamma_j$.

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:

$$\delta_i = \overline{y}_{d(i)} - \overline{y}_{o(i)}, \tag{3}$$

and we let $S^i_{place} = S_{place}(d(i), o(i))$ and $S^i_{pat} = S_{pat}(d(i), o(i))$. Following Bronnenberg et al. (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 both 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, x_{it}, and $\rho_{r(i,t)}$. 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^{scaled}_{it} poorly behaved. This leads us to prefer a regression implementation that avoids dividing by δ_i.

Observe that we can rewrite equation (1) for movers as:

$$y_{it} = \alpha_i + \gamma_{o(i)} + I_{r(i,t)>0} S^i_{place} \delta_i + \tau_t + \rho_{r(i,t)} + x_{it}\beta + \varepsilon_{it}, \tag{4}$$

where $I_{r(i,t)>0}$ is an indicator variable for relative year greater than 0. 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 non-movers in the destination and the origin), and parameterizing the interaction with $\hat{\delta}_i$ as a flexible function of relative year yields

$$y_{it} = \tilde{\alpha}_i + \theta_{r(i,t)}\hat{\delta}_i + \tau_t + \rho_{r(i,t)} + x_{it}\beta + \varepsilon_{it}. \tag{5}$$

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, 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^{scaled}_{it} 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}.

3 Data and Summary Statistics

3.1 Data and Variable Definitions

Our primary data source is a 20 percent random sample of Medicare beneficiaries ("patients") from 1998 through 2008.[8] These data contain approximately 13 million patients. For each 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, the type and quantity of care provided, and the dollar value reimbursed by Medicare. We also observe demographic information for each patient, including age, gender, race, and zip code of residence, defined as the address on file for Social Security payments as of March 31st 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 $t+1$.[9]

Our primary outcome variable is based on an index of overall health care utilization by individ-

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

[9]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.

ual by year, constructed by adjusting an individual's total annual expenditure for regional variation in prices, following Gottlieb et al. (2010). We refer to this throughout as simply "utilization." Online Appendix Section 1 describes the construction of the measure in detail. 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 we discussed in Section 2.2, we prefer a log specification both economically and econometrically. We explore other functional forms in the robustness section below. We also examine a number of other outcome measures, including subcategories of utilization and indicators for particular treatments, which are defined in more detail below.

Our geographic unit of analysis is a Hospital Referral Region (HRR), as defined by the 1998 Dartmouth Atlas of Health Care.[10] The 306 HRRs are collections of zip codes designed to approximate markets for tertiary hospital care. Consistent with the 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 percent of claims occur outside a patient's HRR of residence.

We define patients to be "non-movers" 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. We exclude patients whose HRR of residence changes more than once.

In some of our analysis below, we compare movers to a matched subsample of non-mover 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 non-mover 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 non-mover patient-years forms the "matched sample of non-movers" we refer to below.

[10] 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 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 percent of patient-years whose zip codes do not match the 1998 HRR definitions.

3.2 Sample Restrictions and Summary Statistics

From our original sample of 13 million patients, we retain a 25 percent random sample of non-movers along with all movers. We then restrict the sample to patient-years where patients are between 65 and 99 years old, exclude the 20 percent of patient-years for patients enrolled in Medicare Advantage (for whom we do not observe claims), and exclude the remaining 7 percent 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 percent whose HRR of residence changes more than once, as well as the 35 percent of the remaining "movers" whose share of claims in their destination HRR, among claims in either their origin or destination HRR, is not higher by at least 0.75 in the post-move years relative to the pre-move years.[11]

When we compute HRR averages, such as the sample analogue $\hat{y}_j$ of $\bar{y}_j$, we omit movers in their move year and we weight non-movers by 4 to account for our sampling procedure. All HRR averages are computed by first averaging across individuals in the HRR 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 1 reports summary statistics separately for movers and non-movers. The characteristics of the two groups are broadly similar, although there are some differences. Relative to non-movers, 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 percent of observations equal to zero. 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

[11] The claims data suggest several explanations for why some movers do not satisfy this last 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 post-move by an amount less than our 0.75 threshold. We show in Online Appendix Section 2.3 that our results are robust to alternative ways of defining movers.

non-mover 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 at least 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 percent enter or exit at some point due to enrollment in Medicare Advantage. In Section 4.4 below, we discuss possible biases due to selective attrition and show that our results are robust to some alternative ways of handling this attrition.

Figure 1 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 sample (1998-2003) and the second half of our sample (2004-2008) is 0.9. We show in Online Appendix Figure 7 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 (Chandra et al. 2009; Rettenmaier and Saving 2009; Weinstein et al. 2004).

Online Appendix Section 2.1 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 percent of moves cross state boundaries, and 50 percent cross census division boundaries. Moves to Florida account for 12 percent of all moves, and moves to Arizona or California account for an additional 12 percent; we show in Online Appendix Table 8 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 $12,797$.[12]

Finally, we examine the time-varying correlates of moving. Online Appendix Figure 3 shows that moving is correlated with an increase in utilization, including a spike up in utilization in the year of move. We also report evidence from the Health and Retirement Study (HRS) on the reasons why 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

[12]Not surprisingly given these sample sizes, we show in Online Appendix Section 2.4 that our event study results are not affected by adjusting our estimates of $\hat{\delta}_i$ for noise.

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.

4 Main Results: Patient vs. Place

4.1 Event Study

We begin with two figures that illustrate the variation driving our event study. Figure 2 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 pre-move or in the origin post-move.[13] The claim share in the year of the move (relative year zero) is close to 0.5, consistent with moves being roughly uniform throughout the year. Figure 3 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, 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 4 plots the change in log utilization (the average two to five years post-move minus the average two to five years pre-move) 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 one. If all variation were due to patient effects, we would expect this plot to have a slope of zero. One minus the actual slope is an estimate of a weighted average of the patient share S^i_{pat}.

Figure 4 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 that the absolute change in log

[13] In Online Appendix Section 2.3, 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.

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 "$\times$" in the same figure the average change in log utilization over the same period for our matched sample of non-movers, to whom we assign $\hat{\delta}_i = 0$.[14] 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 non-movers 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 2.1.

Figure 5 shows how pre-move utilization of movers compares to utilization of non-movers in their origin HRR's. The plot is identical to Figure 4 except that the variable on the y axis is now the average difference between log utilization of movers two to five years pre-move and that of their matched non-movers in the same years. The plot has a small upward slope, suggesting that patients who will move to a high-utilization HRR have relatively higher pre-move utilization than those who will move to a low-utilization HRR. This slope is an order of magnitude smaller than the slope in Figure 4. Any systematic differences of this kind in the average utilization of movers will be absorbed by our patient fixed effects α_i.

Our main event study results are shown in Figure 6, which plots estimated coefficients $\hat{\theta}_{r(i,t)}$ from equation (5).[15] 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 above, one 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.

Under the assumptions of our model, the plot should be flat in the years before and after the move. In reality, it shows no post trend but a small but statistically significant pre-trend. This trend could reflect systematic changes in utilization of movers relative to non-movers. Because our model restricts both HRR and patient effects to be time constant, it could also pick up HRR-specific

[14] See notes to Figure 4 for details on this matching.

[15] For computational ease, all of the event studies we report are estimated on the sample of movers only. We show in Online Appendix Figure 12 that including non-movers does not affect the analysis.

trends that are the same for movers and non-movers but happen to be correlated with movers' $\hat{\delta}_i$. In Section 4.4 below, we explore extensions that allow our fixed effects to change over time. We also allow arbitrary pre- and post-move trends for movers by using data only from the years just before or after the move, in the spirit of a regression discontinuity.

Figure 7 presents three alternative event study plots using balanced panels. Panel (a) restricts the sample to early movers for whom we have data for relative years -1 through 7, using only these years in estimation. Panels (b) and (c) are analogous, restricting the sample to movers with data for relative years -4 through 4 and -7 through 1 respectively. The balanced panel figures suggest if anything a slightly larger patient share, and confirm the finding of a small pre-trend and no post-trend.

4.2 Model Estimates

We exploit the variation captured in Figure 6 to estimate equation (1). 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 2, which we consider the central set of results in the paper, 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 $S_{pat}(R,R') = (\hat{c}_R - \hat{c}_{R'}) / (\hat{y}_R - \hat{y}_{R'})$, as well as the components $(\hat{y}_R - \hat{y}_{R'})$, $(\hat{\gamma}_R - \hat{\gamma}_{R'})$, and $(\hat{c}_R - \hat{c}_{R'})$.

Column (1) decomposes the difference between above-median and below-median HRRs. We find that 47 percent 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 percent or less than about 41 percent.

Other partitions of HRRs result in a similar patient share. Patients account for 41 percent of the difference between the top and bottom quartiles (column 2), 39 percent of the difference between the top and bottom deciles (column 3), and 44 percent of the difference between the top and bottom 5 percent (column 4). 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 percent and 30 percent 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 percent based on the jump in log utilization from relative year -1 to 1, as well as with the slope of Figure 4. 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 post-move years rather than the on-impact effect of the move, and that the model is estimated on both movers and non-movers. The stability of the patient share across different partitions is consistent with the linear relationship shown in Figure 4, which implies that $S_{pat}(j,j')$ is not strongly correlated with $\bar{y}_j - \bar{y}_{j'}$.

We present a second, alternative decomposition in Table 3. Here, we ask what share of the cross-HRR variance in log utilization would be eliminated in a counterfactual where average patient characteristics $\bar{c}_j$ were equalized across HRRs. This is

$$S^{var}_{pat} = 1 - \frac{\text{Var}\left(\gamma_j\right)}{\text{Var}\left(\bar{y}_j\right)}. \tag{6}$$

Similarly, the change if area fixed effects were equalized is

$$S^{var}_{place} = 1 - \frac{\text{Var}\left(\bar{c}_j\right)}{\text{Var}\left(\bar{y}_j\right)}.$$

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 one so long as $Cov\left(\bar{c}_j, \gamma_j\right)$ is nonzero.

We find that 55 percent of variance would be eliminated if patient effects were equalized. We find that 72 percent of variance would be eliminated if place effects were equalized. We also find that there is a positive correlation between $\bar{c}_j$ and γ_j, with patients with a high demand for health care tending to sort to slightly higher-spending areas. Because this correlation is positive, S^{var}_{pat} and S^{var}_{place} sum to more than one.

The results thus far decompose geographic variation in log utilization. As discussed in Section 2.2 above, 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 utilization—for example, because patient preferences were either more or less important in big ticket end of life expenditures compared to 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 $\bar{c}_j$ or γ_j were equalized across places. The details of this exercise and a complete set of results are presented in Online Appendix Section 2.2. 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 one. The results suggest that equalizing patient characteristics across areas would reduce geographic differences by 27 percent, while equalizing place characteristics would reduce them by 72 percent. As a separate exercise, we show in Online Appendix Table 7 that simply estimating the model in levels also yields a somewhat lower patient share (23 percent). Comparing these results directly to our main estimates is difficult, but they suggest that the relative importance of patients may be somewhat smaller when we focus on levels.

4.3 Other Outcomes

In Table 4, 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 one 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 1.2. For each row, we reestimate equation (1) 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. Although we do not have a formal framework in which to interpret this heterogeneity, 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, while 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.[16]

In addition to looking at components of utilization, we can also consider alternative functional forms for our measure of overall utilization. In Online Appendix Table 7, we present results for models in which y_{ijt} is defined to be the level of utilization and a patient's percentile rank in the national distribution of utilization respectively; the corresponding event studies are shown in Online Appendix Figure 10. These specifications yield patient shares of 23% and 30%. The same table also presents specifications where y_{ijt} is a dummy variable for the patient being in the top X percent of the national distribution of utilization for different definitions of X. These shares range from 17% to 51%, with some trend toward lower patient shares at the top of the distribution.

4.4 Robustness

Table 5 explores the sensitivity of our main results to relaxing a variety of identifying assumptions. For each specification, the table reports the patient share of the difference between above- and below-median HRRs, analogous to column (1) of Table 2. Event study figures for some of these specifications are shown in Online Appendix Figure 9.

A central assumption of our baseline model is that there are no differential trends in the log utilization of movers that vary systematically with their origin or destination. The event study in

[16] Online Appendix Figure 8 shows event study graphs parallel to Figure 6 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 post-move. The size and direction of the pre- and post- trends vary somewhat across outcomes, and so as a robustness check Online Appendix Table 6 shows the results are similar in magnitude when we limit the estimation sample for movers to one year pre- or post-move.

Figure 6 suggests that this assumption is almost but not perfectly satisfied: there is no meaningful post-trend, but there is a small positive pre-trend. Rows (2)-(4) of Table 5 relax the assumption of no differential trends by using movers' data only in progressively smaller windows around the move year. As we would expect given the positive pre-trend, the estimated patient share increases with smaller windows, rising from 0.47 at baseline to 0.56 when use data only in the year before or after the move. Identification in this latter case is analogous to a regression discontinuity, requiring only the assumption that there are no shocks to utilization that vary systematically with the origin and destination and coincide exactly with the timing of the move.

A second important assumption is that place and patient differences in log utilization are constant over time, up to the variation allowed by the age controls in x_{it} and by the relative year fixed effects $\rho_{r(i,t)}$. The simplest way to relax this assumption is to estimate our model separately for different sub-periods of the data, effectively allowing all of the place and patient parameters to vary between them. Rows (5) through (7) of Table 5 report results using the sub-periods 1998-2001, 2002-2005, and 2006-2008 respectively. The estimated share due to patients ranges from 0.49 in the first period to 0.62 in the latest period.

Another way to allow more flexible changes over time is to estimate equation (1) in first differences, allowing for patient and place-specific *trends* in log utilization:

$$\Delta y_{ijt} = \alpha_i^{FD} + \gamma_j^{FD} + \Delta\gamma_j + \Delta\tau_t + \Delta\rho_{r(i,t)} + \Delta x_{it}\beta + \Delta\varepsilon_{ijt}. \tag{7}$$

Here α_i^{FD} and γ_j^{FD} are new parameters added to the model, and the remaining terms are simply the differenced version of equation (1). The term $\Delta\gamma_j$ is zero if the patient is in the same HRR in periods $t-1$ and t, and $\left(\gamma_j - \gamma_{j'}\right)$ for a patient who moves from j' to j. Results from this model are presented in row (8) of Table 5. They imply a patient share of 0.58, somewhat higher than our baseline estimate.

A third important assumption is that c_{it} and γ_j enter the equation for log utilization additively. Violations of this assumption that lead the γ_j to be relatively more important for some patients and relatively less important for others would mean our estimate of the patient share is local to the characteristics of our patient movers, and not necessarily generalizable to the full population. As a first step in relaxing this assumption, row (9) of Table 5 reports results from a model that allows

different γ_j by quartile of patient age. This seems like a reasonable diagnostic for more general violations where the γ_j differ for high- and low-spending patients, since age is one of the largest observable patient predictors of the level of spending. We find a patient share of 0.44 in the more flexible model, very close to our baseline estimate. That changes in log utilization for patients moving from low to high-utilization HRRs are similar to those for patients moving from high to low (Figure 4 and Section 5.3 below) provides further support for additivity.

Finally, a version of our model with a fully saturated set of HRR-patient fixed effects has an adjusted R^2 of 0.515, compared to 0.503 for our preferred specification. The relatively small increase in explanatory power puts some bound on the scope for violations of additivity, as emphasized by Card et al. (2013).[17]

A fourth important assumption is that the errors in our model are not correlated with entering or exiting the sample due to death, HMO status, or failure to enroll for a complete year in Medicare Part A or B. To get a feel for the importance of this assumption, rows (10)-(12) of Table 5 present results excluding all observations for patients who die in sample, are ever in an HMO, or enter or exit the sample for any of the above reasons, respectively. The associated patient shares are all somewhat higher than our baseline estimate.

A final assumption is that HRRs are an adequate market definition. We need not assume that all patients receive care within their home HRR, but we do need to assume that the geographic distribution of care received by movers in a given a HRR is similar to that of non-movers in that HRR. Otherwise, this could lead the effective γ_j for movers to be different, and so violate our assumption of additivity. To address this, rows (13) and (14) of Table 5 show results for alternative market definitions. In row (13), we define markets to be US states, which yields a patient share of 0.45. In row (14), we define markets to be Hospital Service Areas (HSAs), geographic units that are subsets of HRRs. (There are 3,436 HSAs in the US, compared to 306 HRRs.) This yields a patient share of 0.56, somewhat higher than our main estimate.

As a further robustness check related to market definition, the final two rows of Table 5 show specifications where we only include movers who cross state lines or who cross census region boundaries respectively. Among other things, this alleviates the concern that patients might have

[17] If we estimate the models using only movers, the adjusted R^2 values are 0.554 for the fully saturated model and 0.490 for the baseline specification.

been seeking care in their destination (origin) before (after) their move. The patient shares for both of these subsamples are very close to our baseline estimate.

Finally, Online Appendix Table 8 presents a number of additional robustness checks. We estimate the model using only movers, without the age controls in x_{it}, and without the relative year controls in $\rho_{r(i,t)}$. We use alternative dependent variables: expenditure rather than utilization, and the log of 10 plus utilization or 0.1 plus utilization rather than 1 plus utilization. We drop moves to Florida, Arizona, and California. We estimate equation (1) using the balanced panel samples from Figure 7. In all these cases, the results remain similar in magnitude.

5 Mechanisms

Both the patient and place components of utilization could reflect a range of underlying economic primitives, and understanding these underlying mechanisms is both interesting in its own right and important for policy. In this section, we take a first step in this direction, discussing evidence on possible drivers of the patient component $\overline{c}_j$.

5.1 Patient Demographics

The simplest explanation for the patient component is that it reflects differences in observable demographics such as age, sex, or race. As discussed in the introduction, the consensus of the past literature is that such observables explain only a small share of geographic variation in utilization. We ask whether this remains true in our analysis.

Building on the standard approach in the literature (see, e.g., Zuckerman et al. 2010), we estimate a predictive relationship between our individual-level patient component c_{it} (estimated in equation 1) and a vector of observable characteristics z_{it} using only within-area and within-year variation, then use the coefficients from this relationship to predict between-area differences. We assume that for non-movers

$$\mathrm{E}\left[c_{it}|z_{it},\gamma_j^{obs},\tau_t^{obs}\right]=\gamma_j^{obs}+\tau_t^{obs}+z_{it}\phi, \tag{8}$$

where γ_j^{obs} and τ_t^{obs} are area and year fixed effects distinct from those in equation (1), and ϕ is a

vector of coefficients which we estimate by OLS on the sample of non-movers. We define the component of c_{it} explained by z_{it} to be $c_{it}^{obs} = z_{it}\hat{\phi}$ for both movers and non-movers, and we define the averages $\overline{c}_{jt}^{obs}$, $\overline{c}_{j}^{obs}$, and $\overline{c}_{R}^{obs}$ analogously to $\overline{c}_{jt}$, $\overline{c}_{j}$, and $\overline{c}_{R}$. Finally, we define the patient share explained by demographics to be $S_{pat}^{obs}(R,R') = \left(\overline{c}_{R}^{obs} - \overline{c}_{R'}^{obs}\right) / \left(\overline{y}_{R} - \overline{y}_{R'}\right)$. Like the past literature, we view the projection on z_{it} to be a predictive rather than causal relationship. Variation explained by race, for example, could reflect the effect of characteristics such as poverty or education correlated with race as well as the causal effect of race *per se*.

The results, presented in Table 6, confirm that relatively little geographic heterogeneity is explained by patient demographics. The share of the difference between above- and below-median HRRs explained by differences in age alone is 0.04, or 9 percent of the overall patient share of 0.47. The share explained by age, race, and sex together is 0.05, or 11 percent of the overall patient share. Of course, more variation might be explained by richer demographics not observed in the Medicare data.

5.2 Persistent Treatment

Our patient component reflects differences in utilization that persist regardless of where patients move. A possible explanation for this persistence is that patients who receive diagnoses and aggressive treatments in high-utilization areas will continue their treatment even after they move to low-utilization areas. This mechanism, or related mechanisms in which pre-move treatments are continued post-move, would introduce a form of state dependence in utilization, with the "patient component" partly picking up the stock of existing diagnoses or treatments, which would itself partly be due to place.

A signature of this mechanism is that utilization should adjust less for patients moving from high to low-utilization areas than for patients moving from low to high. A patient from Miami who has been diagnosed with diabetes and is on a treatment regimen may request continuing treatment after moving to Minneapolis, resulting in a relatively small change in utilization. A patient from Minneapolis who rarely visits the doctor and whose diabetes is undiagnosed or untreated, on the other hand, might be likely to receive such a diagnosis and/or new treatment regimen after moving to Miami, leading to a large change in utilization.

Such asymmetry does not appear to be a first-order feature of our data. Figure 4 provided some initial evidence: for any given magnitude $\left|\hat{\delta}_i\right|$ of the difference between origin and destination log utilization, changes in log utilization look symmetric for moves up and down. As further evidence, Figure 8 shows event study plots separately for moves up ($\hat{\delta}_i > 0$) and moves down ($\hat{\delta}_i < 0$) . The figure suggests no significant difference in the magnitude of the utilization change at move, nor in the pattern of post-move adjustment.

This evidence suggests that relatively little of the persistence that drives our measured patient component is explained by past diagnoses and treatment regimens. Of course, there are some versions of this mechanism that would not necessarily show up as asymmetry between moves up and moves down. It could be that certain low-intensity treatments could persist as well. For example, a prostate cancer patient in a low-utilization area whose doctor recommends "watchful waiting" might continue with this approach even after moving to a high-utilization area where it is not usually prescribed. This would amount to a more general form of dependence between current and past utilization, a possibility we consider next.

5.3 Habit Formation

A third mechanism that could underlie our patient component is habit formation in the sense of Becker and Murphy (1988): patient preferences today are a function of patients' past utilization. Patients who build a habit of getting regular checkups or flu shots may continue to do so wherever they go. A patient who has had successful low-intensity outpatient treatment for several past conditions may prefer to do the same for subsequent conditions. The persistent treatment discussed in the previous subsection can be viewed as a special case of habit formation; the general version, however, does not necessarily imply asymmetry between moves up and moves down.

A closely related mechanism is learning: past utilization affects current beliefs rather than current preferences. Learning will typically look similar to habit formation in our setting, so we consider it a version of habit formation for the purposes of this discussion.

A signature of many forms of habit formation is that utilization should continue to adjust toward average behavior in the destination in the years following a move. To see the intuition, suppose that current utilization depends on a weighted average of a patient's current environment and their

average utilization in the past. Consider a patient who moves from low-utilization Minneapolis to high-utilization Miami at the end of year t. Utilization will jump discretely in year $t+1$ because the current environment has changed, but the jump will be muted by the fact that the average of past utilization remains relatively small. In year $t+2$, the current environment is still Miami, but the stock of past utilization is now greater, since the average includes year $t+1$; utilization will increase further. Utilization will continue to increase in years $t+3$, $t+4$, and so forth, as the average of past utilization is weighted more and more toward the patient's post-move experience. This intuition is formalized in Bronnenberg et al. (2012), and it is the key pattern in the data they exploit to identify habit formation in brand preferences.

As illustrated in Figures 6 and 7, our data show remarkably little evidence of this kind of post-move convergence. Log utilization jumps discretely on move but remains almost perfectly flat for up to nine years thereafter. As shown in Figure 8, this remains true whether we look at moves from low to high-utilization areas or moves from high to low.

A second prediction of many habit formation models is that persistence should be greater the larger the stock of past experience a patient has accumulated. In particular, we might expect utilization to be more persistent for older patients than for younger patients. This prediction is also formalized and confirmed in the context of brand preferences by Bronnenberg et al. (2012). To test this prediction in our data, Figure 9 shows event studies separately by quartiles of age. We see no systematic tendency toward smaller jumps for older patients; if anything, the figures suggest the reverse. We confirm in Online Appendix Table 9 that our main estimate of the patient share is actually lower for subsamples of older patients than for subsamples of younger patients.

Together, these facts suggest that a relatively small share of heterogeneity in the patient component of utilization results from habit formation. Of course, we are limited by the fact that our sample is all age 65 or older, and we are only able to observe a maximum of nine years post-move. It is possible that the patient component is endogenous to early life experience with health care, but that it is relatively fixed by the time someone is age 65. It is also possible that we would see more post-move convergence over longer time horizons.

Finding limited evidence of habit formation in health care is of independent interest beyond trying to understand our patient-place decomposition. The lack of post-move adjustment in utilization suggests that individual differences in propensity to consume health care are likely to be relatively

stable, and that the scope for policy to change them may be limited, at least in the short run. At the same time, the fact that the jump in utilization on moving is large and takes place within a two-year window suggests that policies affecting supply-side factors such as doctors' practice patterns are likely to have immediate impacts. These findings may not apply to younger patients, but they seem robustly true for the elderly, who account for about one third of total annual health care spending (Moses et al. 2013).

5.4 Patient Health

A final hypothesis which we investigate is that our patient component reflects variation in health. Patients in Miami may simply be sicker than patients in Minneapolis, due to differences in lifestyle, environment, or genetics, for example.

Past literature has shown that observable measures of patient health status can explain some though not all of the geographic variation in spending. Zuckerman et al. (2010), for example, show that a rich vector of health status indicators explain roughly a third of the spending gap between top and bottom quintile HRRs. Past literature has also emphasized, however, that such correlations are difficult to interpret because the measurement of health conditions is itself likely to be endogenous to place. A given HRR's estimated rate of hypertension, for example, is based on the number of patients who have had recent Medicare claims that included a code for hypertension diagnosis. Such codes are typically only recorded when a patient visits a doctor and receives a billable treatment related to her hypertension.[18] A high-utilization and a low-utilization HRR that had the same underlying rates of hypertension might therefore have very different recorded rates: patients in the high-utilization area may visit the doctor more often, and be more likely to receive billable treatment for their hypertension conditional on visiting. Consistent with this hypothesis, Song et al. (2010) use an empirical strategy similar to ours to show that patients who move across quintiles of the HRR spending distribution experience large, discrete changes in health status as measured by standard proxies, consistent with a higher probability of diagnoses being recorded in claims in more intensive areas.

In order to create a health status measure purged of this endogeneity, we extend our mover-based empirical strategy to separate the underlying patient-specific component of measured health

[18] See chapter 23 of the Medicare Claims Processing Manual (Centers for Medicare and Medicaid Services 2014)

status from the component endogenous to place. Formally, we assume that measured health h_{ijt}^{meas} is a function of true health h_{it} and a measurement error whose distribution depends on place and year:

$$h_{ijt}^{meas} = h_{it} + \xi_{ijt}, \tag{9}$$

where

$$h_{it} = \alpha_i^h + \rho_{r(i,t)}^h + x_{it}\beta^h \tag{10}$$

$$\xi_{ijt} = \gamma_j^h + \tau_t^h + \varepsilon_{ijt}^h. \tag{11}$$

The parameters in these equations are distinct from those in equation (1), but we use the same variable names with "h" superscripts to emphasize the fact that the functional form we assume is ultimately the same, with the same vector of age controls x_{it}. We use the same strategy based on patient movers to identify the patient component of health (h_{it}) separately from the place and year-specific measurement error (ξ_{ijt}), and assume that the analogous identifying conditions hold. Note that we interpret the differential trends for movers captured by $\rho_{r(i,t)}^h$ and the variation correlated with age captured by β^h as changes in true health.[19] We estimate equation (9) by OLS, and form estimates $\hat{h}_{it}$ of the patient component of health for each patient-year.

We consider four standard health status measures. All of them take diagnosis codes as inputs. They differ in which diagnoses they use (although there is considerable overlap), the weights assigned to them, and the lookback period. First, we use the log of the patient's Hierarchical Condition Categories (HCC) score. The HCC score is defined by the Centers for Medicare and Medicaid Services (CMS) for use in computing Medicare payments, and is designed to approximate predicted spending given demographics (including age, gender, and Medicaid eligibility) and diagnoses coded in the previous year.[20] Second, we use the log of one plus the number of 27 possible chronic conditions. These conditions are defined by CMS and are based on diagnoses coded

[19] These components may in fact represent a mix of true health and measurement error. For example, older patients may both have more chronic conditions and be more likely to have a given chronic condition recorded in claims. We include these terms in h_{it} for simplicity; this should be borne in mind in interpreting the results.

[20] Our HCC score derivation is based on Pope et al. (2004)

in the past 1-3 years depending on the condition.[21] Third, we use the log of one plus the patient's Charlson Comorbidity Index (Charlson et al. 1987), a weighted count of diagnoses coded in that year that is designed to predict ten-year mortality.[22] Fourth, we use the log of one plus the number of Iezzoni Chronic Conditions (Iezzoni et al. 1994) that the patient has had diagnoses coded for in that year.[23]

Figure 10 presents an event study illustrating how one of our health measures—the log of the number of chronic conditions—changes around moves. The event study specification is derived from equation (9) analogously to the way equation (5) (and corresponding Figure 6) is derived from equation (1). We replicate the result of Song et al. (2010) that there are sharp changes in measured health status when patients move. There is no meaningful pre-trend and a small post-trend. The size of the jump in these figures is roughly 0.5, implying that the share of variation in measured health status due to the patient component h_{it} is roughly 0.5. In Online Appendix Figure 11 we present event study figures for our other health measures; Online Appendix Table 10 presents decompositions of the variation in health status analogous to Table 2 that confirm a patient share of measured health of about 0.4 to 0.5.

With estimates of equation (9) in hand, we can now ask how much of the geographic variation in log utilization is explained by differences in true patient health h_{it}, which we measure by the estimated patient component of health, $\hat{h}_{it}$ from equation 10. For comparison, we also report how much of the geographic variation in log utilization is explained by differences in measured patient

[21] See https://www.ccwdata.org/web/guest/condition-categories. The conditions are Acquired Hypothyroidism (reference time period: 1 year), Acute Myocardial Infarction (1 year), Alzheimer's Disease and Related Disorders or Senile Dementia (3 years), Alzheimer's Disease (3 years), Anemia (1 year), Asthma (1 year), Atrial Fibrillation (1 year), Benign Prostatic Hyperplasia (1 year), Breast Cancer (1 year), Cataract (1 year), Chronic Kidney Disease (2 years), Chronic Obstructive Pulmonary Disease (1 year), Colorectal Cancer (1 year), Depression (1 year), Diabetes (2 years), Endometrial Cancer (1 year), Glaucoma (1 year), Heart Failure (2 years), Hip/Pelvic Fracture (1 year), Hyperlipidemia (1 year), Hypertension (1 year), Ischemic Heart Disease (2 years), Lung Cancer (1 year), Osteoporosis (1 year), Prostate Cancer (1 year), Rheumatoid Arthritis / Osteoarthritis (2 years), and Stroke / Transient Ischemic Attack (1 year).

[22] The conditions are Acute Myocardial Infarction, AIDS/HIV, Cancer, Cerebrovascular Disease, Chronic Pulmonary Disease, Congestive Heart Failure, Dementia, Diabetes with chronic complications, Diabetes without complications, Hemiplegia or Paraplegia, Metastatic Carcinoma, Mild Liver Disease, Moderate or Severe Liver Disease, Peptic Ulcer Disease, Peripheral Vascular Disease, Renal Disease, and Rheumatologic Disease (Connective Tissue Disease).

[23] The conditions are Chronic Pulmonary Disease, Congestive Heart Failure, Coronary Artery Disease, Dementia, Diabetes With End Organ Damage, Malignant Cancer, Leukemia, Peripheral Vascular Disease, Renal Failure, and Severe Chronic Liver Disease.

health, h_{ijt}^{meas}. We follow the procedure defined in Section 5.1.[24] As with demographics, we stress that the resulting estimates cannot be given a firm causal interpretation. In addition, the patient-specific health component we isolate could partly pick up patient-specific differences that affect the likelihood of health conditions being recorded—for example, because some patients prefer to visit the doctor frequently—in addition to differences in health per se.

Table 7 shows the results. The first four rows present the variation in log utilization explained by the raw health measures, without adjusting for endogenous measurement. The shares of the difference in log utilization between above and below median utilization HRRs explained by the HCC, Charlson, and Iezzoni measures are 0.44, 0.48, and 0.48 respectively, or between 94 and 104 percent of our overall patient share. The share explained by the chronic conditions measure is significantly higher, at 0.79. This is 171 percent of our overall patient share. If there were no endogenous measurement, this would suggest unobserved patient characteristics would have to be negatively correlated with the chronic condition measure across these regions.

In the following four rows, we show that correcting for endogenous measurement has a large effect, reducing the share of variation explained by health by roughly half. After the correction, we find that the shares of the above-below median gap explained by the HCC, Charlson, and Iezzoni measures are 0.22, 0.24, and 0.26 respectively, or 47 to 55 of our overall patient share. The log chronic conditions measure explains share 0.37 of the gap, or 80 percent of the patient share.

Like our findings on habit formation, these results on patient health speak to issues beyond our patient-place decomposition. The large role for patient health suggests that demand-side differences may not be easily affected by policies aimed at changing patients' information or beliefs. In addition, the estimates of equation (9) provide a new way to look at the magnitude of the endogeneity of measured health first identified by Song et al. (2010), showing that the endogenous component accounts for the majority of geographic variation in measured health. The findings of limited post-move adjustment in measured health, and symmetric adjustment for moves up and moves down, shed more light on the nature of the measurement process.[25] Finally, our method

[24]Specifically, we define the observable z_{it} of interest to be either the raw health measure h_{ijt}^{meas} or the estimated patient component of health, $\hat{h}_{it}$, estimate equation (8), and use the results to define the patient share explained by the resultant c_{it}^{obs}.

[25]For example, this pattern may suggest that the endogenous component is mainly related to the recording of diagnoses in claims rather than to diagnoses per se. If the primary force were endogeneity of diagnoses (that is, patients in some places learn they have hypertension while the hypertension elsewhere goes undiagnosed), we would

for separating the endogenous and patient-specific components of health status may have other applications in the large literature that uses health status measures as inputs into risk adjustments.

6 Conclusion

We find robust evidence that 40 to 50 percent 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. Our examination of mechanisms suggests that a large part of this demand-side heterogeneity may be due to patient health. The remaining 50 to 60 percent of variation is due to place-specific factors, possibly including doctor practice patterns and characteristics of health care organizations.

These results suggest that demand-side factors play a larger role in geographic variation than conventional wisdom might suggest. This does not translate immediately into conclusions about efficiency. The correlation of utilization with demand-side factors (and with patient health in particular) may reflect differences in the marginal impact of treatment or the marginal utility from a given impact, and so be consistent with efficiency. But it could also reflect differences in other demand drivers, such as patient information or beliefs. A more careful examination of the efficiency implications of the geographic variation is an important direction for further work.

Our findings have implications beyond our patient-place decomposition. The fact that habit formation seems limited implies that demand-side differences in utilization are unlikely to change quickly in response to policy, at least among the 65 and over population, a population that accounts for about a third of total annual health care spending (Moses et al. 2013). The fact that a large part of demand-side geographic variation reflects variation in patient health may also point to limits to the effectiveness of demand-side policies aimed at changing patient beliefs or preferences. At the same time, the sharp adjustment we observe around moves suggests policies that affect the supply-side can have immediate impacts.

While we have taken a first step toward understanding the origins of the patient component we measure, it remains for future work to better understand the mechanisms behind the place component. Particularly interesting questions concern the role of physicians' training and practice patterns, and the role of health care organizations.

---

have expected to see more adjustment up than adjustment down, based on the intuition developed in Section 5.2.

References

Aaronson, Daniel. 1998. "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes." *Journal of Human Resources*, 33(4): 915–946.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333.

Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. "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.

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. 1999. "Geographic Variations in Utilization Rates in Veterans Affairs Hospitals and Clinics." *New England Journal of Medicine*, 340(1): 32–39.

Baker, Laurence C., Elliott S. Fisher, and John E. Wennberg. 2008. "Variations in Hospital Resource Use for Medicare and Privately Insured Populations in California." *Health Affairs*, 27(2): w123–134.

Baker, Laurence C., M. Kate Bundorf, and Daniel P. Kessler. 2014. "Patients' Preferences Explain a Small but Significant Share of Regional Variation in Medicare Spending." *Health Affairs*, 33(6): 957–963.

Becker, Gary S., and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96(4): 675–700.

Bronnenberg, Bart J., Jean-Pierre H. Dube, and Matthew Gentzkow. 2012. "The Evolution of Brand Preferences: Evidence from Consumer Migration." *American Economic Review*, 102(6): 2472–2508.

Card, David, Joerg Heining, and Patrick M. Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics*, 128(3): 967–1015.

Centers for Medicare and Medicaid Services. 2014. "Medicare Claims Processing Manual." *Centers for Medicare and Medicaid Services*.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy*, 115(1): 103–140.

Chandra, Amitabh, David Cutler, and Zirui Song. 2012. "Who Ordered That? The Economics of Treatment Choices in Medical Care." In *Handbook of Health Economics*. Vol. 2, eds. Thomas G. McGuire Mark V. Pauly and Pedro P. Barros, 397–432. Elsevier.

Chandra, Amitabh, Lindsay Sabik, and Jonathan S. Skinner. 2009. "Cost Growth in Medicare: 1992 to 2006." In *Explorations in the Economics of Aging*. eds. David A. Wise, 133–157. University of Chicago Press.

Charlson, Mary E., Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases*, 40(5): 373–383.

Chernew, Michael, Lindsay Sabik, Amitabh Chandra, Theresa Gibson, and Joseph Newhouse. 2010. "Geographic Correlation between Large-Firm Commercial Spending and Medicare Spending." *American Journal of Managed Care*, 16(2): 131–138.

Chetty, Raj, John N. Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review*, 103(7): 2683–2721.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014*a*. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9): 2633–2679.

Chetty, Raj, John N. Friedman, Soren Leth-Petersen, Torben Nielsen, and Tore Olsen. 2014*b*. "Active vs. Passive Decision and Crowd-out in Retirement Savings Accounts: Evidence from Denmark." *Quarterly Journal of Economics*, 129(3): 1141–1219.

Congressional Budget Office. 2008. "Geographic Variation in Health Care Spending." *Washington: Government Printing Press*.

Cutler, David, Jonathan Skinner, Ariel Dora Stern, and David Wennberg. 2013. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Spending." NBER Working Paper 19320.

Dunn, Abe, Adam Hale Shapiro, and Eli Liebman. 2013. "Geographic Variation in Commercial Medical-Care Expenditures: A Framework for Decomposing Price and Utilization." *Journal of Health Economics*, 32(6): 1153–1165.

Fernandez, Raquel, and Alessandra Fogli. 2006. "Fertility: the Role of Culture and Family Experience." *Journal of the European Economic Association*, 4(2-3): 552–561.

Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003*a*. "The Implications of Regional Variations in Medicare Spending: the Content, Quality, and Accessibility of Care. Part 1." *Annals of Internal Medicine*, 138(4): 273–287.

Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003*b*. "The Implications of Regional Variations in Medicare Spending: the Content, Quality, and Accessibility of Care. Part 2." *Annals of Internal Medicine*, 138(4): 288–299.

Gawande, Atul. 2009. "The Cost Conundrum: What a Texas Town Can Teach Us about Health Care." *The New Yorker*.

Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn G. Andrews, Jonathan S. Skinner, and Jason M. Sutherland. 2010. "Prices Don't Drive Regional Medicare Spending Variations." *Health Affairs*, 29(3): 537–543.

Iezzoni, Lisa I., Timothy Heeren, Susan M. Foley, Jennifer Daley, John Hughes, and Gerald A. Coffman. 1994. "Chronic Conditions and Risk of in-Hospital Death." *Health Services Research*, 29(4): 435–460.

McPherson, Klim, P.M. Strong, Arnold Epstein, and Lesley Jones. 1981. "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. Part A: Medical Sociology*, 15(3 Part 1): 273–288.

Molitor, David. 2014. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *University of Illinois at Urbana-Champaign Working Paper: 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. 2013. "The Anatomy of Health Care in the United States." *Journal of the American Medical Association*, 310(18): 1947–1964.

Philipson, Tomas J., Seth A. Seabury, Lee M. Lockwood, Dana P. Goldman, and Darius N. Lakdawalla. 2010. "Geographic Variation in Health Care: The Role of Private Markets." *Brookings Papers on Economic Activity*, 41(1).

Pope, Gregory C., John Kautter, Randall P. Ellis, Arlene S. Ash, John Z. Ayanian, Lisa I. Iezzoni, Melvin J. Ingber, Jesse M. Levy, and John Robst. 2004. "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model." *Health Care Financing Review*, 25: 119–141.

Rettenmaier, Andrew J., and Thomas R. Saving. 2009. "Perspectives on the Geographic Variation in Health Care Spending." *Special Publication National Center for Policy Analysis*.

Sheiner, Louise. 2014. "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*.

Skinner, Jonathan. 2011. "Causes and Consequences of Regional Variations in Health Care." In *Handbook of Health Economics*. Vol. 2, eds. Thomas G. McGuire Mark V. Pauly and Pedro P. Barros, 45–93. Elsevier.

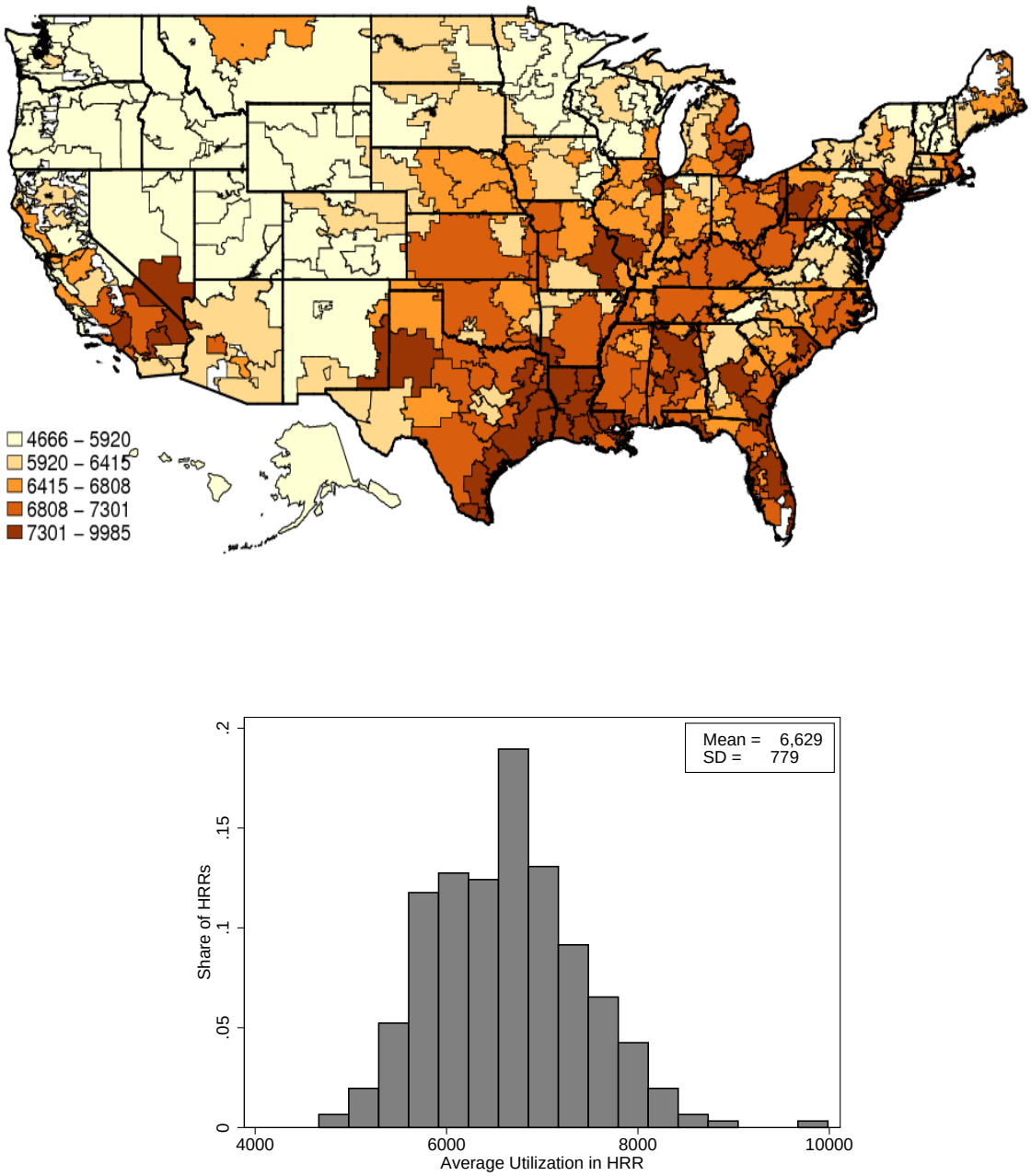
Song, Yunjie, Jonathan Skinner, Julie Bynam, Jason Sutherland, and Elliott Fisher. 2010. "Regional Variations in Diagnostic Practices." *New England Journal of Medicine*, 363(1): 45–53.

Subramanian, Usha, Morris Weinberger, George J. Eckert, Gilbert J. L'Italien, Pablo Lapuerta, and William Tierney. 2002. "Geographic Variation in Health Care Utilization and Outcomes in Veterans with Acute Myocardial Infarction." *Journal of General Internal Medicine*, 17(8): 604–611.

Weinstein, James N., Kristen K. Bronner, Tamara Shawver Morgan, and John E. Wennberg. 2004. "Trends and Geographic Variations in Major Surgery for Degenerative Diseases of the Hip, Knee, and Spine." *Health Affairs*, no.(2004): var81–89.

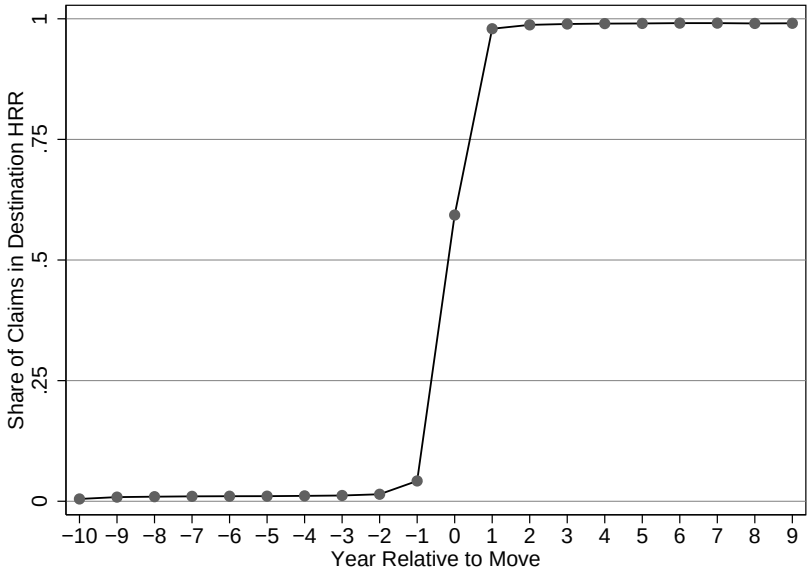
Zuckerman, Stephen, Timothy Waldman, Robert Berenson, and Jack Hadley. 2010. "Clarifying Sources of Geographic Differences in Medicare Spending." *New England Journal of Medicine*, 363(1): 54–62.

Figure 1: Distribution of Utilization Across HRRs



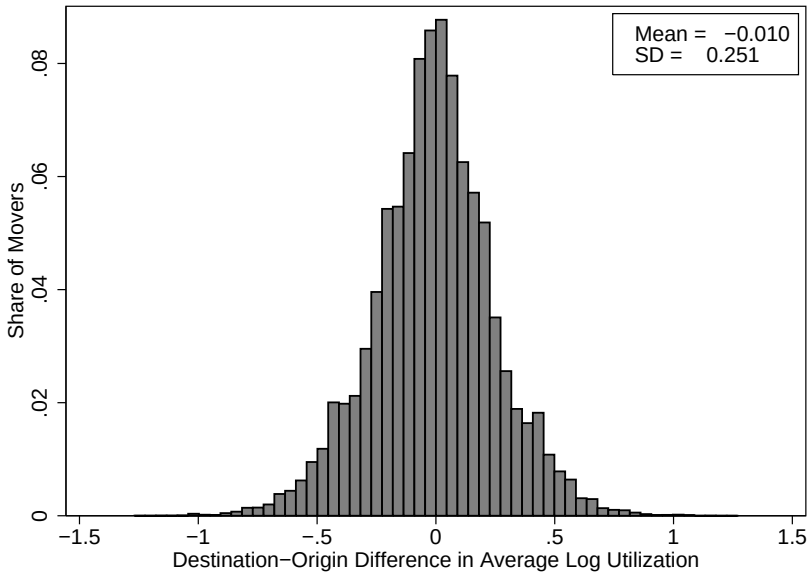
Notes: Figure displays the distribution of average utilization by HRR. We first average utilization across individuals within each HRR-year, upweighting non-movers by four, and then take a simple average within HRR across years. Map shows the distribution of level utilization in quintiles. Lower and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

Figure 2: Share of Claims in Destination by Relative Year



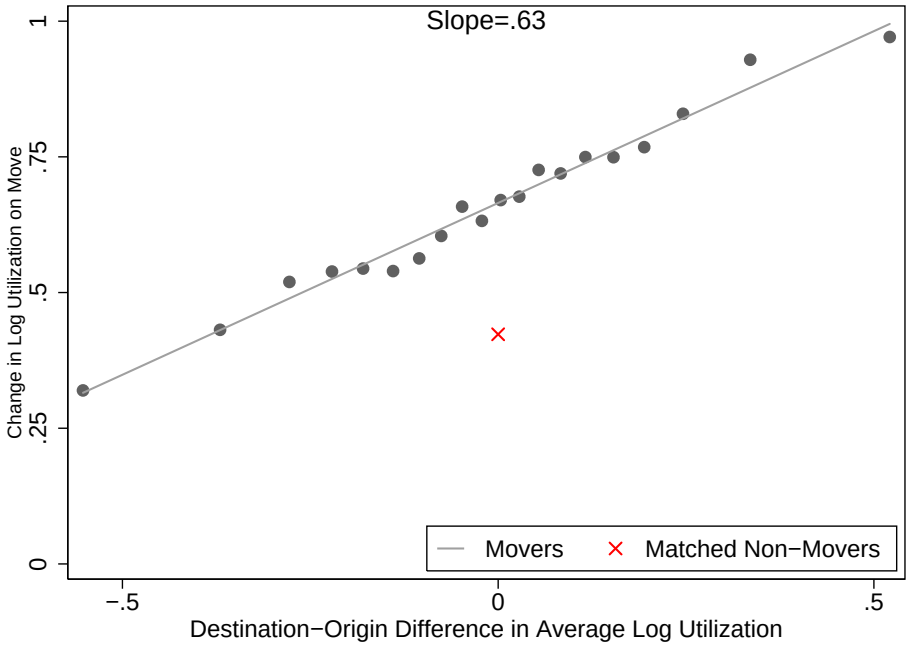
Notes: Figure shows the share of a mover's claims located in their destination HRR, among those in either their origin or their destination HRR. The sample is all movers ($N = 3,702,189$ patient-years).

Figure 3: Distribution of Destination-Origin Difference in Log Utilization



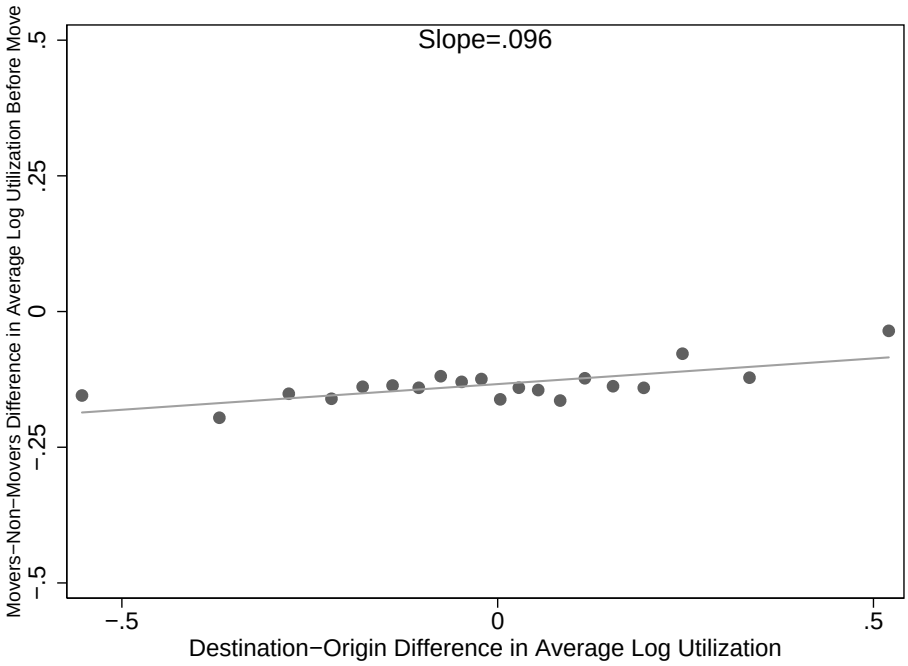
Notes: Figure shows the distribution across movers of the difference $\hat{\hat{\delta}}_i$ in average log utilization between their origin and destination HRRs. The sample is all movers ($N = 3,702,189$ patient-years).

Figure 4: Change in Log Utilization By Size of Move



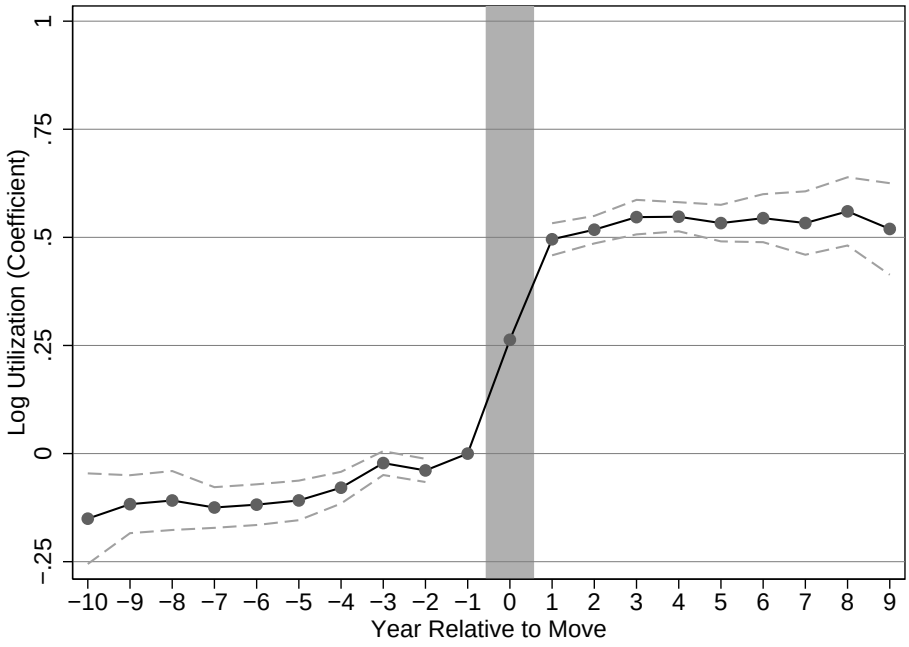
Notes: Figure shows the change in log utilization before and after move. For each mover, we calculate the difference $\hat{\delta}_i$ in average log utilization between their origin and destination HRRs, then group $\hat{\delta}_i$ into ventiles. The x-axis displays the mean of $\hat{\delta}_i$ for movers in each ventile. The y-axis shows, for each ventile, average log utilization two to five years post-move minus average log utilization two to five years pre-move. 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 pre-move and between two and five years post-move ($N = 1,919,137$ patient-years). For comparison, we also compute the average change in log utilization for a sample of matched non-movers, which we show with the red cross on the graph. Specifically, for each mover in our data in each calendar year we randomly draw a non-mover 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 non-mover patient-years forms the matched sample.

Figure 5: Pre-move Differences in Log Utilization



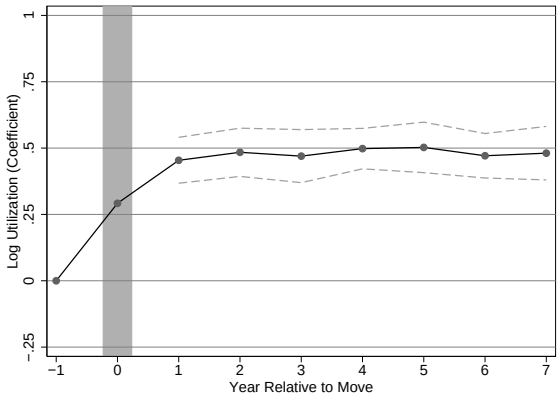
Notes: Figure shows the level of pre-move log utilization for movers relative to non-movers by the size of their subsequent move $\hat{\delta}_i$. For each mover, we calculate the difference $\hat{\delta}_i$ in average log utilization between their origin and destination HRRs, then group $\hat{\delta}_i$ 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 non-mover patient-years two to five years pre-move. In Figure 4 we describe the construction of the matched sample of non-movers. 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 pre-move ($N = 1,048,843$ patient-years).

Figure 6: Event-Study

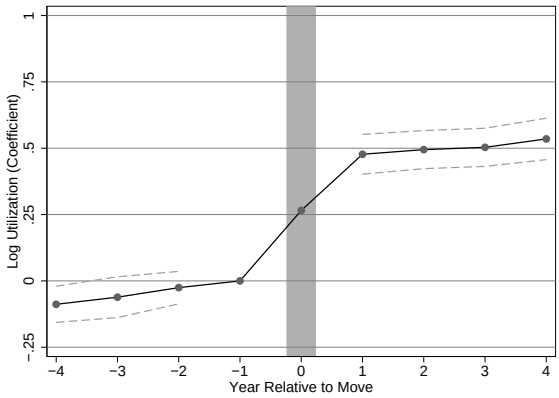


Notes: Figure shows the coefficients $\tilde{\theta}_{r(i,t)}$ estimated from equation (5). The coefficient for relative year -1 is normalized to 0. The dependent variable y_{it} is log utilization; x_{it} consists of indicator variables for five-year age bins. The dashed lines are upper and lower bounds 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 (5) 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 $\hat{\tilde{\delta}}_i$ for each repetition. The sample is all movers ($N = 3,702,189$).

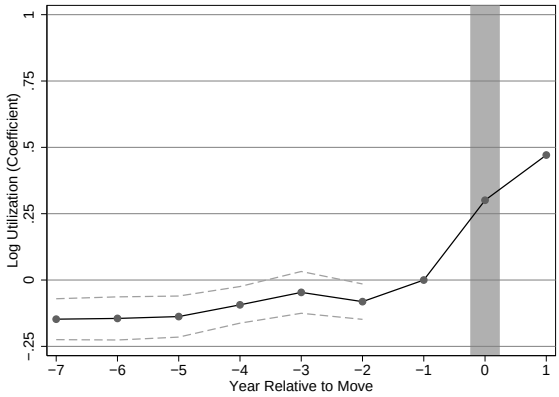
Figure 7: Balanced-Panel Event-Study



(a) Early Moves



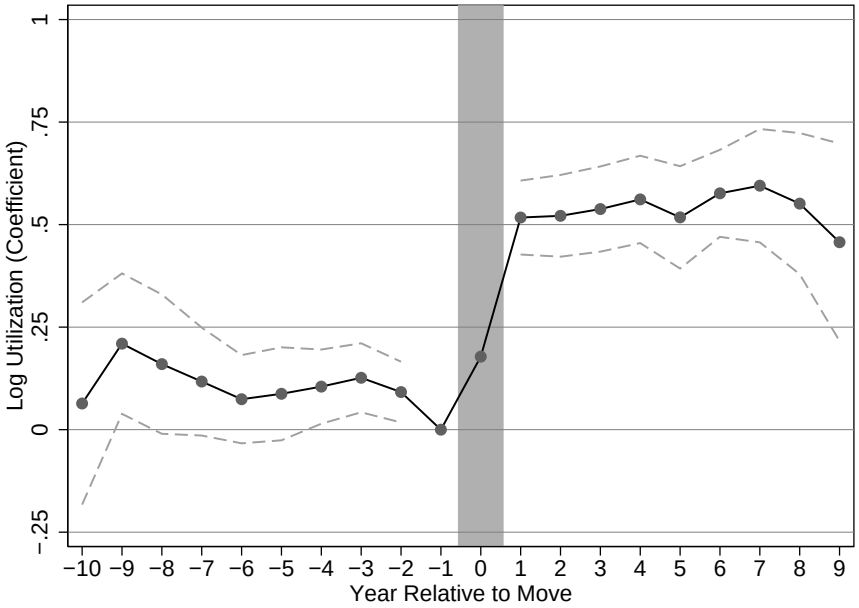
(b) Middle Moves



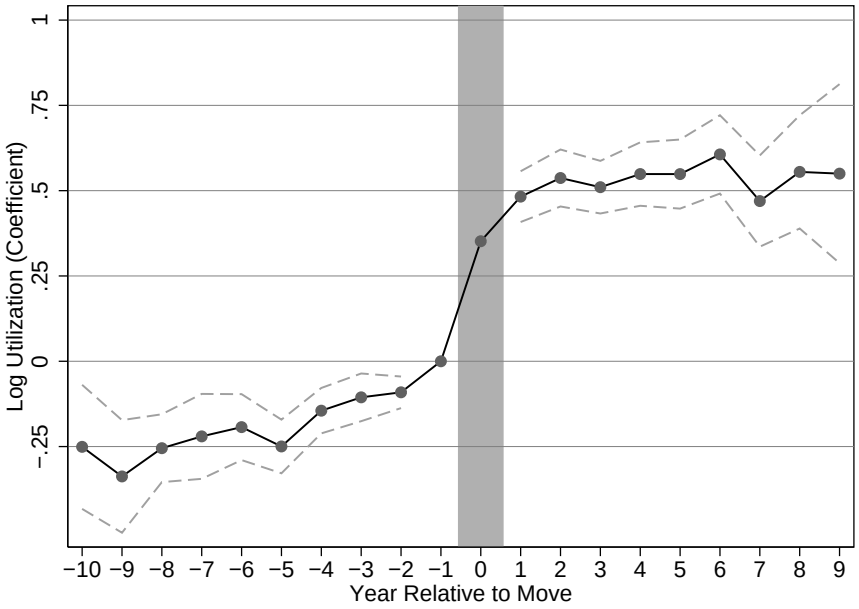
(c) Late Moves

Notes: These figures are constructed in the same manner as Figure 6 above, except they are estimated on balanced-panel subsamples of movers whom we observe in each of a given set of relative years. The dashed lines show the 95% confidence interval, constructed using the same bootstrap approach as in Figure 6. Panel (a) restricts to movers whom we observe in every relative year in [-1,7] ($N = 422,226$ patient-years). Panel (b) restricts to movers whom we observe in every relative year in [-4,4] ($N = 474,462$ patient-years). Panel (c) restricts to movers whom we observe in every relative year in [-7,1] ($N = 544,221$ patient-years).

Figure 8: Event-Study, Moves Up and Moves Down



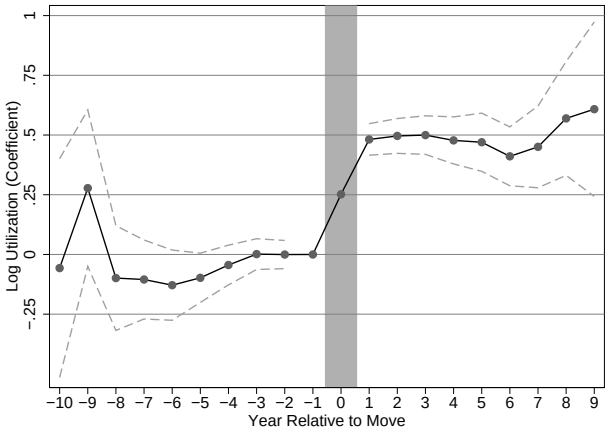
(a) Moves from Low to High-Utilization HRRs



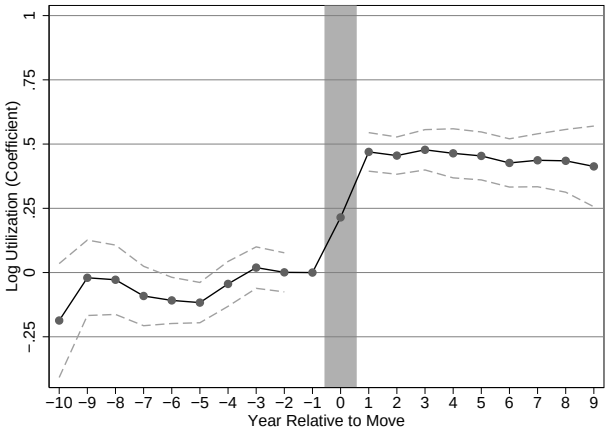
(b) Moves from High to Low-Utilization HRRs

Notes: These figures are constructed in the same manner as Figure 6, except they are estimated on moves up in panel (a) and on moves down in panel (b). A move up is defined to be a move to a destination HRR with higher mean log utilization than the mean log utilization of the origin. A move down is defined to be a move to a destination HRR with lower mean log utilization than the mean log utilization of the origin. The dashed lines show the 95% confidence interval, constructed using the same bootstrap approach as in Figure 6. The sample in panel (a) is all movers who move up ($N = 1,792,033$ patient-years). The sample in panel (b) is all movers who move down ($N = 1,910,156$ patient-years).

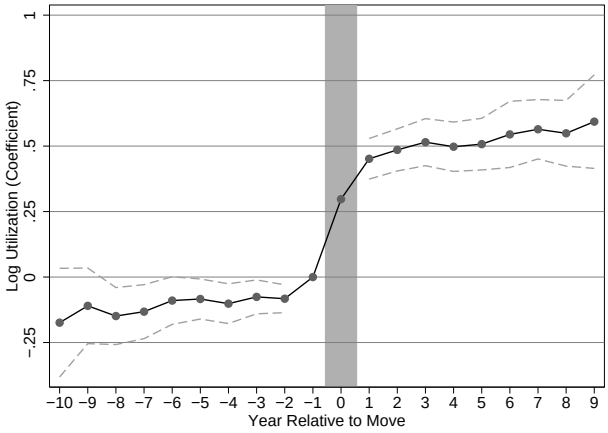
Figure 9: Event-Study, Results By Age Quartile



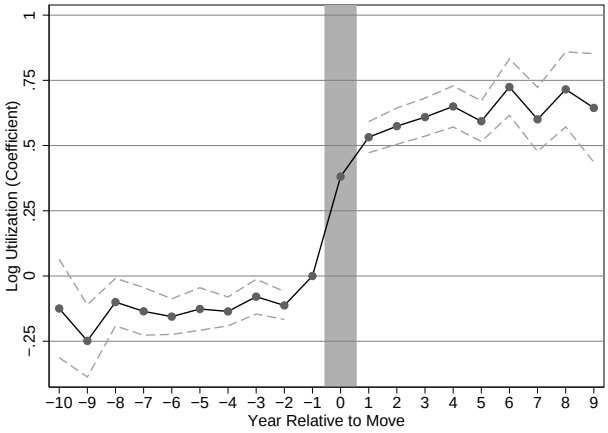
(a) Age Quartile 1



(b) Age Quartile 2



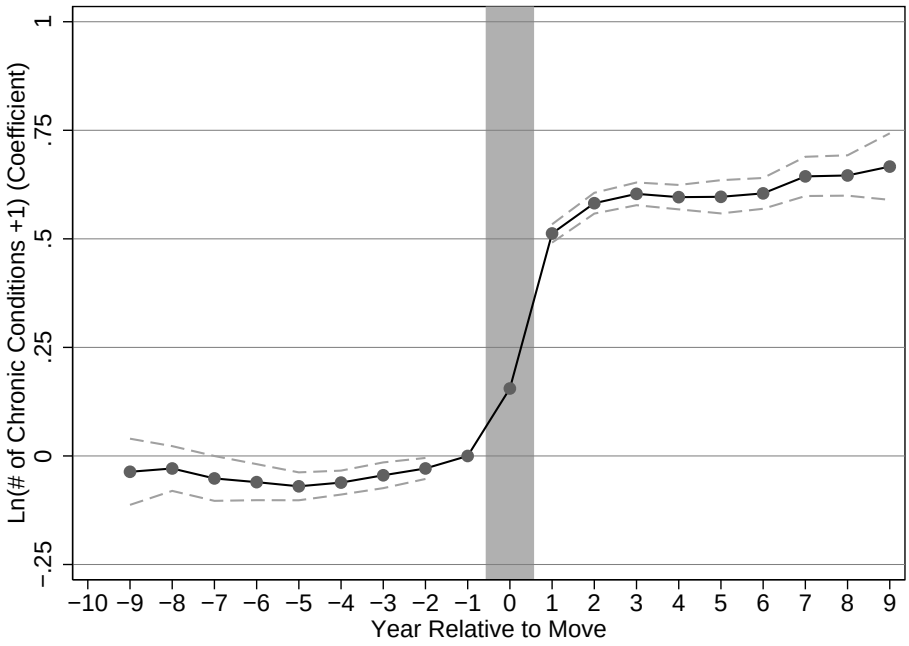
(c) Age Quartile 3



(d) Age Quartile 4

Notes: These figures are constructed in the same manner as Figure 6, except that they are estimated on subsamples of all movers divided by age quartiles. Quartiles of age are determined based on the mean age over the years observed for each patient. Panel (a) provides estimates for the first quartile of age (mean age 68.5), panel (b) provides estimates for the second quartile of age (mean age 72.8), panel (c) provides estimates for the third quartile of age (mean age 78.3), and panel (d) provides estimates for the fourth quartile of age (mean age 86.0). The dashed lines show the 95% confidence interval, constructed using the same bootstrap approach as in Figure 6. The sample in panel (a) includes movers in the first quartile of age ($N = 746,132$ patient-years); panel (b) includes movers in the second quartile ($N = 868,531$ patient-years); panel (c) includes movers in the third quartile ($N = 977,512$ patient-years); panel (d) includes movers in the fourth quartile ($N = 1,110,014$ patient-years).

Figure 10: Event-Study Analysis of Log Number of Chronic Conditions



Notes: Figure is constructed in the same manner as Figure 6, except that it uses the log number of chronic conditions as the dependent variable. The dashed lines show the 95% confidence interval, constructed using the same bootstrap approach as in Figure 6. The sample includes all mover-years except 1998, as chronic conditions are not observed in that year ($N = 3,407,590$ patient-years).

Table 1: Summary Statistics

	(1)	(2)
	Non-movers	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
S.D.	\$12,690	\$9,567
Share of patient-years with zero	0.06	0.06
Number of chronic conditions:		
Mean	2.98	3.30
S.D.	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. 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 non-movers ($N = 16,432,955$ patient-years).

Table 2: Additive Decomposition of Log Utilization

	(1)	(2)	(3)	(4)	(5)	(6)
	Above / below median	Top & bottom 25%	Top & bottom 10%	Top & bottom 5%	McAllen & El Paso	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 based on estimation of equation (1), 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 (1) 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 $(\hat{c}_R - \hat{c}_{R'})$. The fourth row reports the share of the difference in average utilization between the two areas due to patient $\left(\hat{S}_{pat}(R,R')\right)$ 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 $\left(\hat{S}_{place}(R,R')\right)$ 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 non-movers, excluding relative year zero ($N = 16,031,875$ patient-years).

Table 3: 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	0.549
	(0.013)

Notes: Results based on estimates of equation (1). The first row reports variance of $\hat{y}_j$, which is estimated using the same specification as in Table 2. The second, third, and fourth rows report the variance of $\hat{\gamma}_j$, variance of $\hat{c}_j$, and the correlation between $\hat{\gamma}_j$ and $\hat{c}_j$, respectively, using a split sample approach to correct for the (correlated) measurement error in $\hat{\gamma}_j$ and $\hat{c}_j$. Specifically, we randomly assign movers within each origin-destination pair and non-movers within each HRR to two approximately equal-sized subsamples and estimate equation (1) separately on each subsample. We compute the variance of $\hat{\gamma}_j$ (or $\hat{c}_j$) as the covariance between $\hat{\gamma}_j$'s (or $\hat{c}_j$'s) estimated from the two subsamples. The correlation between $\hat{\gamma}_j$ and $\hat{c}_j$ is computed from the variances of $\hat{\gamma}_j$ and $\hat{c}_j$, and the covariance between $\hat{\gamma}_j$ and $\hat{c}_j$, which we estimate as the average of the covariances between $\hat{\gamma}_j$ from one subsample and $\hat{c}_j$ from the other subsample. The last two rows of the table report the share of the variance in cross-HRR utilization that would be reduced if HRR effects were made equal across areas ($\hat{S}^{var}_{place}$) and if patient effects were made equal across areas ($\hat{S}^{var}_{pat}$). 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 2.

Table 4: Various 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
(2)	Seen a primary care physician	0.884	0.042	0.452
(3)	Seen a specialist	0.815	0.051	0.322
(4)	Any hospitalization	0.226	0.037	0.410
(5)	Any emergency room visit	0.346	0.045	0.714
(6)	Log (# of diagnostic tests)	1.449	0.550	0.092
(7)	Log(# of imaging tests)	0.842	0.220	0.142
(8)	Log(# of preventive care measures)[a]	1.376	0.098	0.611
(9)	Log(# of different doctors seen)	1.525	0.113	0.392
(10)	Log(inpatient utilization)[b]	2.004	0.340	0.242
(11)	Log(outpatient utilization)[b]	6.890	0.193	0.358
(12)	Log(emergency room utilization)[b]	2.296	0.352	0.639
(13)	Log(other utilization)[b]	3.430	0.957	0.124

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 2, 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 ($\hat{S}_{pat}(R, R')$). All log outcome measures are the log of the outcome plus 1. Online Appendix Table 11 shows the percent with zero for each of these outcomes. The sample size is the same as in Table 2.

[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, Flu Shot, or in the past two years: Mammogram, Pap Test, Prostate Cancer Screening.

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

Table 5: Robustness

	Specification	(1) N	(2) Mean of log utilization	(3) Above / below median utilization difference	(4) Share due to patients
(1)	Baseline	16,031,875	7.193	0.283	0.465
(2)	Relative years -5 to 5	15,430,835	7.193	0.283	0.469
(3)	Relative years -3 to 3	14,689,929	7.193	0.283	0.499
(4)	Relative years -1 to 1	13,511,698	7.192	0.284	0.557
(5)	First third of sample only (1998-2001)	4,857,799	6.936	0.284	0.490
(6)	Second third of sample only (2002-2005)	5,238,278	7.252	0.290	0.519
(7)	Third third of sample only (2006-2008)	3,599,208	7.452	0.303	0.621
(8)	First differences with fixed effects	16,432,955	7.197	0.281	0.583
(9)	HRR fixed effects interacted with age quartiles	16,031,875	7.193	0.283	0.441
(10)	Patients who never die	10,999,832	6.904	0.292	0.527
(11)	Patients never in an HMO	13,432,817	7.224	0.284	0.468
(12)	Patients never missing outcomes	8,135,140	6.921	0.287	0.509
(13)	Using states as geographic unit	16,029,246	7.146	0.282	0.446
(14)	Using HSAs as geographic unit	16,031,875	7.201	0.391	0.561
(15)	Cross state movers only	14,974,181	7.192	0.283	0.451
(16)	Cross census region movers only	13,967,660	7.190	0.284	0.451

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 2, for alternative samples and specifications. Columns report the sample size, mean of log utilization, difference in average utilization between above and below median utilization HRRs ($\hat{y}_R - \hat{y}_{R'}$) and patient share ($\hat{S}_{pat}(R, R')$). Rows (2) to (4) narrow the sample of years for movers to relative years -5 to 5, relative years -3 to 3, and relative years -1 to 1, respectively. Rows (5)-(7) limit the sample to patient years in 1998-2001, 2002-2005, and 2006-2008, respectively, excluding movers whose move year falls outside the time window in question. In row (8) we estimate the model in first differences, allowing for patient and place-specific trends in log utilization (see equation 7); here we do not drop the year of the move (relative year 0), but use the adjustment technique described in Online Appendix Section 2.3 assuming that there is no misreporting in move timing other than a 50% chance of a patient being in their origin in relative year 0 and a 50% chance of a patient being in their destination in relative year 0. In row (9), we add interaction terms between HRR dummies and dummies for age quartiles. Rows (10), (11), and (12) restrict the sample to patients who do not die in sample, who are never in an HMO, and who never have a missing value of utilization for any reason, respectively. Rows (13) and (14) change the geographic of unit of analysis to states and Hospital Service Areas (HSAs), respectively. We do not change our sample selection criteria for movers when we vary the definition of the geographic unit j in the model. In Row (13), the sample size falls slightly because there is a small number of patients for whom we do not have a valid state code. Rows (15) and (16) restrict the sample of movers to those who cross state or census region boundaries, respectively.

Table 6: Variation in Log Utilization Explained by Patient Demographics

		(1)	(2)
		Share of above / below median utilization difference due to patient demographic(s)	Fraction of patient share
(1)	Age	0.040	0.086
(2)	Race	0.002	0.005
(3)	Sex	0.012	0.027
(4)	Age, race, sex	0.051	0.110

Notes: Column (1) reports the share of the difference in average log utilization between above-median and below-median utilization HRRs due to patient observables ($\hat{S}^{obs}_{pat}(R,R')$). Results based on the estimation of equation (8), letting z_{it} be the patient observables indicated in the first column. The age, race, and sex dummies are fully interacted in row (4). The sample size is the same as in Table 2.

Table 7: Variation in Log Utilization Explained by Patient Health

		(1)	(2)
		Share of above / below median utilization difference due to patient health	Fraction of patient share
	Raw health measure		
(1)	Log(HCC score)	0.435	0.935
(2)	Log(Charlson Comorbidity Index)	0.483	1.037
(3)	Log(# of Iezzoni chronic conditions)	0.483	1.037
(4)	Log(# of chronic conditions)	0.794	1.707
	Patient component of health measure		
(5)	Log(HCC score)	0.220	0.473
(6)	Log(Charlson Comorbidity Index)	0.242	0.520
(7)	Log(# of Iezzoni chronic conditions)	0.256	0.550
(8)	Log(# of chronic conditions)	0.371	0.797

Notes: Table reports shares of the difference in average log utilization between above-median and below-median utilization HRRs explained by patient health. Results based on the estimation of equation (8), letting z_{it} be the measure of health indicated in the first column. Rows (5)-(8) use the patient component (h_{it}) of health measures estimated from equation (9). Column (1) reports the share of the difference in average log utilization between these areas due to observable patient health ($\hat{S}^{obs}_{pat}(R,R')$). Column (2) shows this as a fraction of the overall patient share estimated in column (1) of Table 2. 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 2 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).