NBER WORKING PAPER SERIES

PLACE-BASED DRIVERS OF MORTALITY: EVIDENCE FROM MIGRATION

Amy Finkelstein
Matthew Gentzkow
Heidi L. Williams

Working Paper 25975
http://www.nber.org/papers/w25975

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2019

We are grateful to the National Institute on Aging (Finkelstein, R01-AG032449), the National Science Foundation (Williams, 1151497) and the Stanford Institute for Economic Policy Research (Gentzkow) for financial support. 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/w25975.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Amy Finkelstein, Matthew Gentzkow, and Heidi L. Williams. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Place-Based Drivers of Mortality: Evidence from Migration
Amy Finkelstein, Matthew Gentzkow, and Heidi L. Williams
NBER Working Paper No. 25975
June 2019
JEL No. H51,I1,I11

ABSTRACT

We estimate the effect of current location on elderly mortality by analyzing outcomes of movers in the Medicare population. We control for movers' origin locations as well as a rich vector of pre-move health measures. We also develop a novel strategy to adjust for remaining unobservables, based on the assumption that the relative importance of observables and unobservables correlated with movers' destinations is the same as the relative importance of those correlated with movers' origins. We estimate substantial effects of current location. Moving from a 10th to a 90th percentile location would increase life expectancy at age 65 by 1.1 years, and equalizing location effects would reduce cross-sectional variation in life expectancy by 15 percent. Places with favorable life expectancy effects tend to have higher quality and quantity of health care, less extreme climates, lower crime rates, and higher socioeconomic status

Amy Finkelstein
Department of Economics, E52-442
MIT
50 Memorial Drive
Cambridge, MA 02142
and NBER
afink@mit.edu

Heidi L. Williams
Department of Economics, E52-440
MIT
50 Memorial Drive
Cambridge, MA 02142
and NBER
heidiw@mit.edu

Matthew Gentzkow
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
and NBER
gentzkow@stanford.edu

1 Introduction

Mortality rates vary substantially across the US. Focusing on the 100 most populous commuting zones, Chetty et al. (2016b) estimate that life expectancy at age 40 ranges from a high of 85 in San Jose, California to a low of 81 in Las Vegas, Nevada, with a standard deviation across commuting zones of 1.2 years.[1] Murray et al. (2006) estimate that county-level life expectancy at birth in 1999 ranged from 66.6 years in Bennett County, South Dakota to 81.3 years in Summit County, Colorado. Currie and Schwandt (2016) likewise document substantial disparities across county groups in life expectancy at birth as well as in mortality at older ages.

Why do people in some parts of the US live longer than others? The long list of possible causes can be divided into two broad categories: differences in residents' stocks of *health capital* (Grossman 1972), and differences in the environment associated with their current location. Health capital includes genetic endowments, as well as the persistent effects of prior health behaviors (e.g., smoking, diet, exercise), prior medical care, and other past experiences that impact current mortality. Potentially mortality-relevant aspects of residents' current locations include the quality and quantity of available medical care, local climate and pollution, and risk factors such as crime and traffic accidents. Chetty et al. (2016b) find that the main correlates of area mortality in the cross section are health capital factors such as smoking, obesity, and exercise, and that correlations with place factors such as health care spending or local environmental conditions are weak. Neither they nor other past work, however, isolate the causal impact of place effects.

In this paper, we use mortality outcomes of migrants in the elderly Medicare population to separately identify the causal role of health capital and current location on mortality in the U.S. We will refer to the impact of current location by the shorthand "*place effects.*" Our strategy proceeds in two steps. First, we analyze mortality differences among movers to different destinations, controlling for both their origin locations and a rich vector of pre-move observable health measures in Medicare claims data. The idea behind our approach is to take two patients from the same origin (say, Boston), one of whom moves to a low-mortality area (say, Minneapolis), and the other of whom moves to a high-mortality area (say, Houston), and to compare their mortality outcomes

[1] Authors' calculations based on the publicly reported data provided by Chetty et al. (2016b) on life expectancy for each commuting zone reported separately by gender, which we use to calculate overall life expectancy assuming equal shares of men and women in each commuting zone. Note that these data include only individuals with non-zero reported household income.

after they move. If origin location plus pre-move health measures capture all differences in health capital potentially correlated with choice of destination, this would provide a valid estimate of the place effects.

Second, we apply a novel strategy to correct for any remaining selection on unobserved health capital. Our strategy builds on prior work (Murphy and Topel 1990; Altonji et al. 2005; Oster 2016) in using variation in observable characteristics to adjust for variation in unobservables. In our context, this amounts to using the correlation between movers' choice of destination and their observed health capital to adjust for potential correlation between choice of destination and unobserved health capital.

What distinguishes our strategy from prior work is that we do not need to specify *a priori* the overall importance of the unobservables. As Oster (2016) emphasizes, the standard approach requires the researcher to specify this importance in the form of an assumption on the R^2 of a hypothetical regression including both the observables and the unobservables. We weaken this assumption by exploiting an additional moment of our data: the extent to which the origin location from which individuals move is predictive of their mortality after controlling for observed health. If our observable measures capture all relevant dimensions of health capital, movers' origins would have no further predictive power. The key assumption that we make is that the relative importance of observables and unobservables correlated with movers' origins is the same as the relative importance of observables and unobservables correlated with movers' destinations. We present several pieces of evidence consistent with this assumption, and also show that it is implied by a natural class of models of migration decisions.

We use data on all Medicare beneficiaries aged 65 and older from 1999 through 2014. The enrollee-level panel data contain information on zip code of residence and date of death (if any), along with demographic variables such as age, race, sex, and enrollment in Medicaid (a proxy for low income). The claims data provide us with detailed annual measures of health conditions based on recorded diagnoses, as well as measures of health care utilization. Our geographic unit of analysis is a Commuting Zone (CZ), a standard aggregation of counties that partitions the US and is designed to approximate labor markets.

The main outcome we focus on is life expectancy at age 65. We estimate the average of this life expectancy in each location by fitting a standard Gompertz mortality model using observed

age-mortality gradients (Olshansky and Carnes 1997; Chetty et al. 2016b). In our analysis sample, mean life expectancy at age 65 is 83.3 years, with an across-area standard deviation of 0.84 years.

We find that current location has a large causal impact on mortality. Our results imply that moving from an area at the 10th percentile of estimated place effects to an area at the 90th percentile would increase life expectancy at age 65 by 1.1 years, or about half of the 90-10 cross-sectional difference. We estimate that equalizing place effects across areas would reduce the cross-sectional variation in life expectancy at age 65 by 15 percent. By comparison, equalizing health capital across areas would reduce the cross-sectional variation by about 75 percent. Our analyses assume that current location has no direct impact on health capital. While in principle location may influence health behaviors and hence health capital, we think it is a reasonable assumption that the elderly's health capital is fixed in the short run; consistent with this assumption, we provide evidence that the impacts of place show up immediately and do not appear to increase with time spent in the new location.

Our findings suggest that policies which affect short-run determinants of mortality such as medical care or environmental factors can potentially produce large and immediate changes in outcomes, as can policies such as the Moving to Opportunity Project (Ludwig et al. 2012; Chetty et al. 2016a) that relocate vulnerable populations to areas with more favorable conditions. At the same time, our findings suggest that health capital also plays an important role. While our estimated place effects are positively correlated with average area life expectancy, this correlation is far from perfect. Our place-by-place estimates of these components identify areas such as Santa Fe, New Mexico; Denver, Colorado; and El Paso, Texas as having negative causal effects despite relatively high average life expectancy, and other areas such as Charleston, West Virginia as having positive causal effects despite relatively low average life expectancy.

Finally, we present evidence on the observable area-level correlates of our estimated place effects. The results are intuitive. Areas with positive place effects tend to have higher-quality hospitals, more primary care physicians and specialists per capita, and higher health care utilization. The positive correlation between an area's healthcare utilization and its causal impact on life expectancy contrasts with the lack of correlation between utilization and average health outcomes which has been emphasized in the *Dartmouth Atlas* literature and which we replicate here (Fisher et al. 2003a,b; Skinner 2011). Areas with favorable place effects also tend to have less extreme

climates, less pollution, fewer homicides, fewer automobile fatalities, and are more urban. They also tend to have higher socioeconomic status (SES) as measured by income and education, as well as better health behaviors, which may reflect higher willingness to pay for healthcare quality and other favorable place characteristics among such individuals.

Our work contributes to the large literature on the determinants of mortality. McGovern et al.'s (2014) recent review of studies on health determinants concludes that this literature tends to attribute the largest importance for mortality to health capital — specifically to behaviors (35-50%) and to genetics (20-30%). Among potential place effects, it attributes between 5-20% of the determinants of mortality to environment and around 10% to medical care. While the methodologies of the studies underlying these estimates vary, they generally all rely on correlational analyses to quantify the relative importance of these different factors.[2] Our analysis advances this body of descriptive work with a research design that more convincingly isolates causal effects.

Our work is particularly related to prior work on the drivers of geographic variation in mortality. This work has also tended to highlight the importance of health capital, particularly health behaviors. Fuchs (2011) famously attributes the lower mortality rates of clean-living, predominantly Mormon residents of Utah to better health behaviors than their neighbors in the more dissolute state of Nevada.[3] Chetty et al. (2016b) show that geographic variation in life expectancy for low-income individuals is significantly correlated with health behaviors such as smoking, obesity and exercise, but not significantly correlated with measures of health care quality or quantity. This is consistent with the large *Dartmouth Atlas* literature which has found health care utilization to be uncorrelated with mortality (Fisher et al. 2003a,b; Skinner 2011).[4]

Summarizing the state of knowledge on both the determinants of mortality and the determinants of geographic variation in mortality, Cutler (2018) concludes, "Behavior is the key. When we compare geographic regions, the dominant factor driving health differences is how Americans behave. Unhealthy areas smoke more, drink more and eat to excess; healthier areas avoid these be-

[2]The underlying studies included in their review are DHH (1980), McGinnis and Foege (1993), Lantz et al. (1998), McGinnis et al. (2002), Mokdad et al. (2004), Danaei et al. (2009), WHO (2009), Booske et al. (2010), Stringhini et al. (2010), and Thoits (2010).

[3]See also Fuchs (1965) on geographic variation in mortality within the US.

[4]In addition to geographic variation in medical care, a number of studies have examined the correlates of another natural component of place effects — current environmental factors such as air pollution — with regional variation in mortality rates (e.g. Dockery et al. 1993; Samet et al. 2000). For example, Dockery et al. (1993) estimate that across-city variation in air pollution is positively associated with deaths from lung cancer and cardiopulmonary disease.

haviors." The large role we estimate for health capital is consistent with this conventional wisdom. However, our results also show that there is a substantial causal impact of place-based factors that this conventional wisdom may understate.

Finally, our work is related to two recent papers that isolate causal impacts of place factors. Doyle (2011) uses health emergencies of visitors to different areas of Florida to show that hospitals in high-spending areas produce better outcomes than hospitals in low-spending areas. Deryugina and Molitor (2018) document that Medicare survivors of Hurricane Katrina who move to lower-mortality regions experience subsequently lower mortality than those who move to higher-mortality regions.

Our empirical strategy for correcting for selection on unobservables may have applications in other contexts. Oster (2016) emphasizes the sensitivity of the standard approach to assumptions about the overall explanatory power of the observables, and notes that direct information to guide such assumptions is often limited. We propose weaker assumptions under which this decision can be guided by the data. Our approach is most obviously relevant to other contexts in which individuals move across geographies, firms, or other units of analysis, and in which selection on unobserved individual characteristics is a potential confound; this could arise due to data limitations (e.g. Bronnenberg et al. 2012) or because an outcome cannot be measured repeatedly in individual panel-level data (such as mortality in our case or inter-generational mobility in Chetty and Hendren 2016). It may also be applied to other settings where there are auxiliary variables whose relative correlation with observables and unobservables is plausibly similar to that of the treatment of interest.

The rest of the paper proceeds as follow. Sections 2 and 3 describe our model and empirical strategy, and Section 4 presents our data and summary statistics. Section 5 presents evidence on the selection of movers across origins and destinations and describes how our empirical strategy addresses this selection. Section 6 presents our main results on the impact of current environment on life expectancy, and explores some observable correlates of the place effects. Section 7 provides additional support for some of our key assumptions and shows robustness of our main results to alternative specifications. The last section concludes.

2 Model

We consider a set of individuals indexed by i and a set $\mathscr{J}$ of locations indexed by j. The individuals are either (i) movers who live in an origin location $o \in \mathscr{J}$ in years $t < t_i^*$, move in year t_i^* from o to $j \in \mathscr{J}$, and then live in destination location j thereafter; or (ii) non-movers who live in the same location $j \in \mathscr{J}$ throughout the sample, and to whom we assign a reference year t_i^* as discussed below. We abuse notation slightly in using j to denote a generic location and also letting $j(i)$ denote the observed location of individual i (permanent location if i is a non-mover, and destination if i is a mover). Similarly, we use o to denote a generic origin location and $o(i)$ to denote the observed origin of mover i.

We analyze a continuous-time survival model in which the mortality rate of person i at age a depends on her location and her stock of health capital. We follow Chetty et al. (2016b) in adopting a Gompertz specification in which the log of the mortality hazard rate $m_{ij}(a)$ that individual i would experience at age a if she lived in location j is linear in age:

$$\log\left(m_{ij}(a)\right) = \beta a + \gamma_j + \theta_i. \tag{1}$$

Here, θ_i is i's health capital, which we assume is fixed over the horizon of ages observed in our data, but may be endogenous to experiences earlier in life. The term γ_j is a fixed effect capturing the causal effect of living in location j, which we will refer to as the *place effect*.[5] We let $\overline{\theta}_j$ denote the average health capital of non-movers in j. In order to mirror the literature, which focuses on race and sex adjusted mortality rates as the object of interest, in computing $\overline{\theta}_j$ we assign each area j the national average racial and gender composition. We define the mortality rate of an average non-mover in j at age a to be $\overline{m}_j(a) = \exp\left[\beta a + \gamma_j + \overline{\theta}_j\right]$. We refer to the sum $\left(\gamma_j + \theta_i\right)$ as the *mortality index* of individual i, and to $\left(\gamma_j + \overline{\theta}_j\right)$ as the *average mortality index* in area j.

Our main outcome of interest is life expectancy at age 65, hereafter, *life expectancy*. Given a generic continuous mortality hazard rate $m(a)$, the probability the individual survives to age a conditional to surviving to age 65 is given by the survival function $S(a) = e^{-\int_{65}^{a} m(v)dv}$. The life expectancy of an individual who survives until age 65 is $65 + \int_{65}^{\infty} S(a)da$.[6] We define the life

[5] More precisely, $\gamma_j - \gamma_k$ is the causal effect of living in place j rather than place k.

[6] Let $F(a)$ and $f(a)$ denote the distribution and density of age at death conditional on living to age 65, which

expectancy at 65 of an average non-mover in j by substituting $\overline{m}_j(a)$ into these expressions. We will denote this L_j, and refer to it simply as *average life expectancy in area j*.

Our ultimate goal is to estimate the causal effect on life expectancy of living in area j. We define this by considering a thought experiment in which an individual with average health capital is assigned to live counterfactually in each location j beginning at age 65. Letting $\overline{\theta}$ denote the average health capital over the full population of non-movers, this defines a set of counterfactual mortality rates $\overline{m}_j^*(a) = \exp\left[\beta a + \gamma_j + \overline{\theta}\right]$ that differ across j only because of the place effects γ_j. Substituting $\overline{m}_j^*(a)$ into the expression for life expectancy yields the counterfactual life expectancy L_j^*. Letting $\overline{\gamma}$ denote the population-weighted average of the γ_j, and letting $\overline{L}$ denote the life expectancy associated with mortality hazard $\exp\left[\beta a + \overline{\gamma} + \overline{\theta}\right]$, we define the *treatment effect of area j* to be $L_j^* - \overline{L}$.

We assume that health capital θ_i can be further decomposed into a component that depends on demographics X_i, a component that depends on observed health H_i, a series of terms capturing unobserved health capital orthogonal to X_i and H_i but correlated with locations, and an orthogonal residual:

$$\theta_i = X_i\psi + H_i\lambda + \eta_{j(i)}^{nm} + \eta_{o(i)}^{orig} + \eta_{j(i)}^{dest} + \tilde{\eta}_i. \tag{2}$$

Here, both X_i and H_i are measured as of year $t_i^* - 1$. We define $\eta_{j(i)}^{nm}$, $\eta_{o(i)}^{orig}$, and $\eta_{j(i)}^{dest}$ to be the fixed effects from a hypothetical regression of θ_i on X_i, H_i, and fixed effects for non-movers' locations, movers' origins, and movers' destinations respectively. (We fix $\eta_{j(i)}^{nm} = 0$ for movers and $\eta_{o(i)}^{orig} = \eta_{j(i)}^{dest} = 0$ for non-movers.) We define $\tilde{\eta}_i$ to be the residual from this regression. We thus have $\mathrm{E}(\tilde{\eta}_i | X_i, H_i, o(i), j(i)) = 0$ for movers and $\mathrm{E}(\tilde{\eta}_i | X_i, H_i, j(i)) = 0$ for non-movers by construction.

Our definition of $\tilde{\eta}_i$ as a residual that is orthogonal by construction mirrors Altonji et al. (2005) and Oster (2016). It means that the coefficients ψ and λ capture both the causal effects of X_i and H_i and the effects of any unobservables that may be correlated with X_i and H_i. It is natural to assume that such correlations will exist, as unobserved determinants of health capital such as smoking will generally be correlated with observed measures of health capital such as diagnoses

we assume is a continuous random variable. We have $S(a) = 1 - F(a)$. The hazard function is $m(a) = \frac{f(a)}{S(a)} = -\frac{d}{da}\log S(a)$. Integrating both sides of this equation yields $\log S(a) = -\int_{65}^{a} m(v)\,dv$. Life expectancy at age 65 is $\int_{65}^{\infty} a f(a)\,da$. Integrating by parts, and assuming a finite end time, shows this is equal to $65 + \int_{65}^{\infty} S(a)da$.

of hypertension. This means that equation (2) does not define a structural relationship, and the η terms include only the components of the unobservables orthogonal to X_i and H_i.

A key assumption in our model is the additive separability of health capital and the place effects in equation (1) for log mortality. Analogous assumptions are standard in the literature using changes in residence or employment to separate effects of individual characteristics from geographic or institutional factors (e.g. Card et al. 2013; Chetty and Hendren 2018, 2016; Finkelstein et al. 2016).

This is a strong assumption, but we see it as a reasonable one in our setting. It has the intuitive implication that health capital and current location affect the *level* of mortality multiplicatively, and, thus, that the level of mortality of individuals with poor health capital (high θ_i) will vary more across areas than that of individuals who have better health capital (low θ_i); this has indeed been documented by Chetty et al. (2016b). More concretely, suppose that there are two possible levels of health capital, such that in an average location, individuals have either a 0.1% annual mortality hazard or a 10% annual mortality hazard. The additive separability assumption implies that anything about the current environment that reduces mortality — such as the quality of health care or the air quality — will reduce mortality by a constant *proportion* for all individuals, with a larger percentage point effect on individuals with worse health capital. Our specification rules out place effects that cause the same level shift in mortality for all patients regardless of their health capital. For example, if some places have a higher risk of death from auto accidents and this probability is independent of health capital, our assumption would be violated. We present empirical support for additive separability in Section 7 below.

3 Empirical Strategy

3.1 Estimation and Identification

Our main goal in estimation is to identify the place effects γ_j. This will in turn allow us to recover the average health capital θ_i of movers and non-movers in each location. Combining equations (1)

and (2) yields the following estimating equation for the realized mortality rate $m_i(a)$:

$$\log(m_i(a)) = \beta a + X_i\psi + H_i\lambda + \tau_{o(i)}^{orig} + \tau_{j(i)}^{dest} + \tau_{j(i)}^{nm} + \tilde{\eta}_i. \tag{3}$$

where $\tau_{o(i)}^{orig}$, $\tau_{j(i)}^{dest}$, and $\tau_{j(i)}^{nm}$ are fixed effects for movers' origins, movers' destinations, and non-movers' locations respectively, and we have $\tau_{o(i)}^{orig} = \eta_{o(i)}^{orig}$, $\tau_{j(i)}^{nm} = \gamma_{j(i)} + \eta_{j(i)}^{nm}$, and $\tau_{j(i)}^{dest} = \gamma_{j(i)} + \eta_{j(i)}^{dest}$.

We estimate this model by maximum likelihood. Given the estimated parameters, we can consistently estimate the area j mortality rate $\overline{m}_j(a)$ by $\hat{m}_j(a) = \exp\left[\hat{\beta}a + \overline{X}_j\hat{\psi} + \overline{H}_j\hat{\lambda} + \hat{\tau}_j^{nm}\right]$, where $\overline{X}_j$ and $\overline{H}_j$ are the averages of X_i and H_i over non-movers in j.[7] Consistent with the definition of $\overline{\theta}_j$ above, when we compute $\overline{X}_j$ we set the elements of the vector associated with race and sex to their national rather than their area averages. We compute estimates $\hat{L}_j$ of average life expectancy L_j in area j by substituting $\hat{m}_j(a)$ for $\overline{m}_j(a)$ in the derivation of L_j in Section 2. All of our reported estimates of average life expectancy in area j are therefore race- and sex-adjusted.

The central challenge is identification of γ_j. Simply comparing average mortality rates across areas in the cross-section does not recover γ_j, because locations may differ in their average health capital $\mathrm{E}(\theta_i | j(i) = j)$. An optimistic assumption would be that X_i and H_i absorb all such differences. In this case, η_j^{nm}, η_o^{orig}, and η_j^{dest} would be equal to zero for all $j, o \in \mathscr{J}$, and we would not need to use movers at all; we could simply estimate equation (3) using non-movers and the $\hat{\tau}_j^{nm}$ would be consistent estimators of γ_j.

A more plausible assumption would be that X_i and H_i do not absorb all area differences in health capital, but that the remaining differences for movers are absorbed by the origin fixed effects η_o^{orig}, so that $\eta_j^{dest} = 0$ for all j. In this case, the estimated destination fixed effects $\hat{\tau}_j^{dest}$ from equation (3) would be consistent estimators of γ_j. This assumption would follow from a model in which the locations where people are born and live up to age 65 or older may be related to their genetic endowments, health behaviors, and other determinants of health capital, but in which late-life moving decisions are driven by idiosyncratic factors.

Our findings below are qualitatively consistent with this intuition, in the sense that conditioning on movers' origins eliminates a significant amount of non-random selection on observables. How-

[7] Note that $\overline{\theta}_j = \overline{X}_j\psi + \overline{H}_j\lambda + \eta_{j(i)}^{nm}$, and so $\overline{X}_j\hat{\psi} + \overline{H}_j\hat{\lambda} + \hat{\tau}_j^{nm}$ converges in probability to $\gamma_j + \overline{\theta}_j$.

ever, our results also suggest that some non-random selection may remain, implying that $\eta_j^{dest} \neq 0$ and thus that $\hat{\tau}_j^{dest}$ may not exactly recover γ_j. The selection correction strategy we develop in the next sub-section is designed to deal with any such remaining selection.

Given consistent estimates $\hat{\gamma}_j$ of γ_j, we can estimate the treatment effects $L_j^* - \overline{L}$ of each area j. To do so, we estimate $\overline{\theta}$ as the mean across all non-movers of $X_i\hat{\psi} + H_i\hat{\lambda} + \hat{\tau}_j^{nm} - \hat{\gamma}_j$, a consistent estimator of $\hat{\theta}_i$. We estimate $\overline{\gamma}$ by the non-mover population-weighted mean of the $\hat{\gamma}_j$. We then substitute these estimates in place of their population counterparts in the definitions of L_j^* and $\overline{L}$ in Section 2.

We will at various points form estimates of variances of CZ-level terms such as γ_j. Unless otherwise noted, all such estimates $\hat{\text{Var}}(z)$ for CZ-level variables z are based on a split sample approach in which we randomly partition our sample into two parts, form separate estimates $\hat{z}^1$ and $\hat{z}^2$ using the two samples, and then define $\hat{\text{Var}}(z) = \text{Cov}\left(\hat{z}^1, \hat{z}^2\right)$. We compute confidence intervals via 200 iterations of the Bayesian bootstrap procedure (Rubin 1981).[8]

When we report individual values of the place effects $\hat{\gamma}_j$ or the life expectancy treatment effects that depend on them, we adjust the $\hat{\gamma}_j$ estimates for sampling error using a standard Empirical Bayes' procedure, producing adjusted estimates we denote $\hat{\gamma}_j^{EB}$. This closely follows the approach of Chetty and Hendren (2016) and Finkelstein et al. (2017). Appendix A.1 provides more detail on this procedure.

3.2 Adjusting for Selection on Unobservables

In this section, we introduce our strategy to allow for the possibility that movers' destinations are correlated with their unobserved health—i.e., that $\eta_j^{dest} \neq 0$. Our approach builds on the now-standard methodology developed by Murphy and Topel (1990) and Altonji et al. (2005), and expanded on by Oster (2016), which uses variation in observables to make inferences about the likely bias due to unobservables.

The standard approach relies on two key assumptions. The first is that the relationship between the treatment of interest and the index of observables is similar to the relationship between the treatment of interest and the index of unobservables. Altonji et al. (2005) and Oster (2016) refer

[8]The Bayesian bootstrap smooths bootstrap samples by reweighting rather than resampling observations. For a recent application see Angrist et al. (2017); their on-line Appendix provides implementation details that we follow.

to this as the *equal selection* assumption. Intuitively, it allows us to learn about the *direction* of bias induced by the unobservables from the bias induced when we omit the observables. In a standard labor economics context where we would attempt to measure returns to education, equal selection would imply that if education is increasing in observed proxies for worker skill, it will be increasing in unobserved skill as well. In our context, equal selection implies that if movers to a particular destination tend to have unusually good observed health capital they will probably have unusually good unobserved health capital as well.

The second assumption pins down the overall importance of the unobservables relative to the observables. Oster (2016) operationalizes this as an assumed value for the R^2 of a hypothetical regression of the outcome on the treatment, the observables, and all the relevant unobservables.[9] We will refer to this as the R^2 *assumption.* Intuitively, specifying this value allows us to determine the *magnitude* of the bias induced by the unobservables. In the labor economics example, the bias would be small if there is very little variation in unobserved skill conditional on the observed proxies, or large if this variation is large. In our context, the bias would be small if observed proxies captured most of the variation in health capital, and so the variance of the unobserved components was small. Oster (2016) emphasizes that the choice of the R^2 value is by necessity arbitrary in typical applications, and suggests some benchmark values researchers could use to obtain conservative bounds.

The main innovation in our approach is to suggest an additional moment of the data that allows us to weaken the R^2 assumption. That moment is the variance of the origin component of unobserved health—η_o^{orig} in equation (2), which we recall is consistently estimated by the origin fixed effect $\hat{\tau}_j^{orig}$ from equation (3). If our observable measures H_i captured all relevant dimensions of health capital, movers' origins would have no further predictive power, and we would have $\eta_j^{orig} = 0$. The extent to which origins remain predictive of mortality after we control for H_i is a gauge of the extent to which important unobserved components remain.

To apply this logic formally, we first introduce some new constructs and notation. First, define a "treatment" indicator $T_{ij} = \mathbf{1}\left(j\left(i\right) = j\right)$ for movers equal to one if i's destination is j. Second, as an input to our selection correction strategy, we will need to estimate the components of observed health capital related to movers' origins and destinations respectively. Let $h_i = H_i\lambda$ (where λ

[9] Altonji et al. (2005) do not name this assumption, but they implicitly assume that the relevant R^2 is one.

is defined in equation (3)) be the index of observed health capital for individual i; we refer to it throughout as "observed health" for short. Define the following regression in the sample of movers:

$$h_i = \beta^h a + X_i \psi^h + h^{orig}_{o(i)} + h^{dest}_{j(i)} + \tilde{h}_i, \tag{4}$$

where $h^{orig}_{o(i)}$ and $h^{dest}_{j(i)}$ are origin and destination fixed effects respectively and $\tilde{h}_i$ is a residual. We refer to $h^{orig}_{o(i)}$ and $h^{dest}_{j(i)}$ as the origin and destination components of observed health respectively. These are by construction the residual components of observed health after partialing out age and demographics. We normalize $h^{orig}_{o(i)}$ so the population mean of $h^{dest}_{j(i)}$ is zero. To estimate these terms, we first form $\hat{h}_i = H_i \hat{\lambda}$ using the estimates $\hat{\lambda}$ from equation (3). We then estimate equation (4) replacing h_i with $\hat{h}_i$.

Our two key assumptions can now be stated as follows.

Assumption 1. *(Equal Selection)* $\text{Corr}\left(T_{ij}, h^{dest}_{j(i)}\right) = \text{Corr}\left(T_{ij}, \eta^{dest}_{j(i)}\right)$ *in the sample of movers for all* $j \in \mathscr{J}$.

Assumption 2. *(Relative Importance)* $\frac{\text{StDev}\left(\eta^{orig}_{j(i)}\right)}{\text{StDev}\left(h^{orig}_{j(i)}\right)} = \frac{\text{StDev}\left(\eta^{dest}_{j(i)}\right)}{\text{StDev}\left(h^{dest}_{j(i)}\right)}$ *in the sample of movers.*

Assumption 1 is a version of the Altonji et al. (2005) and Oster (2016) equal selection assumption applied to our setting. Note that our setting differs from the one they consider in that our "treatment" is multidimensional—a vector of indicators for moving to the various destinations in $\mathscr{J}$. To map this back to the standard case, we imagine a setting where the treatment of interest was the effect of moving to one particular destination j, and so the treatment variable is just the binary indicator T_{ij}. We then assume the assumption applies separately for each possible destination $j \in \mathscr{J}$.[10]

Assumption 2 allows us to use information from origin unobservables in place of the R^2 assumption. Rather than assuming an arbitrary value for the variance of the destination unobservables $\text{Var}\left(\eta^{dest}_{j(i)}\right)$ as the standard approach would dictate, we assume that the variance of these unobservables relative to the variance of the destination observables $\text{Var}\left(h^{dest}_{j(i)}\right)$ is the same as the corresponding ratio for movers' origins. Combining these two assumptions allows us to consistently estimate the key unobservables η^{dest}_j for each j from observed moments of the data.

[10] Our assumption also differs in that we state it in terms of correlations rather than regression coefficients.

These assumptions are strong, but they follow naturally from economic primitives. They will hold in a broad class of models of selective migration so long as selection of locations is related to overall health capital but not differentially to the observed and unobserved components. We show this formally in Appendix A.2. Under some additional structure on the distributions of observables and unobservables, we show that Assumptions 1 and 2 are both implied by the assumption that selection of origins and destinations may depend on the single index $\tilde{\theta}_i = h_i + \eta_i$, where $\eta_i = \eta^{orig}_{o(i)} + \eta^{dest}_{j(i)} + \tilde{\eta}_i$, but that origins and destinations are independent of h_i and η_i conditional on $\tilde{\theta}_i$.

If the dimensions of health capital relevant to selection are not captured by a single index, our assumption requires that the relative importance of unobservable to observable health in determining origin must be the same as the relative importance of unobservable to observable health in determining destination. This could be violated if, for example, observed dimensions of health capital such as diabetes are more strongly related to people's choice of where to live when young, while unobserved dimensions such as physical mobility are more strongly related to their migration decisions when they are elderly. We provide empirical support for the assumptions behind our selection correction approach in Section 7.2 below.

Proposition 1. *Assumption 1 is equivalent to*

$$\eta^{dest}_j = \frac{\text{StDev}\left(\eta^{dest}_{j(i)}\right)}{\text{StDev}\left(h^{dest}_{j(i)}\right)} h^{dest}_j. \tag{5}$$

Proof. Recalling that $h^{dest}_{j(i)}$ and $\eta^{dest}_{j(i)}$ are normalized to have mean zero, it is straightforward to show that $\text{Cov}\left(T_{ij}, h^{dest}_{j(i)}\right) = \frac{N}{N'} h^{dest}_j p(1-p)$ and $\text{Cov}\left(T_{ij}, \eta^{dest}_{j(i)}\right) = \frac{N}{N'} \eta^{dest}_j p(1-p)$, where N is the total number of movers, N' is the number with $T_{ij} = 0$, and $p = \Pr\left(T_{ij} = 1\right)$.[11] Assumption 1

[11] Since T_{ij} is a binary variable, $\text{Cov}\left(T_{ij}, h^{dest}_{j(i)}\right) = \left[\text{E}\left(h^{dest}_{j(i)} | T_{ij} = 1\right) - \text{E}\left(h^{dest}_{j(i)} | T_{ij} = 0\right)\right] p(1-p)$, where $p = \Pr(T_{ij} = 1)$. Let I be the set of all movers and let I' be the set of movers for whom $T_{ij} = 0$. We know $\text{E}\left(h^{dest}_{j(i)} | T_{ij} = 1\right) = h^{dest}_j$ and $\text{E}\left(h^{dest}_{j(i)} | T_{ij} = 0\right) = \frac{1}{N'} \sum_{i \in I'} h^{dest}_{j(i)} = \frac{N}{N'} \left(\frac{1}{N} \sum_{i \in I} h^{dest}_{j(i)} - \frac{1}{N} \sum_{i \in I \setminus I'} h^{dest}_{j(i)}\right)$. Since our normalization implies the population mean $\frac{1}{N} \sum_{i \in I} h^{dest}_{j(i)}$ is zero, and noting that $\frac{1}{N} \sum_{i \in I \setminus I'} h^{dest}_{j(i)} = \frac{N-N'}{N} h^{dest}_j$, we have

$$\begin{aligned} \text{Cov}\left(T_{ij}, h^{dest}_{j(i)}\right) &= h^{dest}_j \left[1 + \frac{N-N'}{N'}\right] p(1-p) \\ &= \frac{N}{N'} h^{dest}_j p(1-p). \end{aligned}$$

is then equivalent to

$$\frac{\frac{N}{N'} h_j^{dest} p(1-p)}{\text{StDev}\left(T_{ij}\right) \text{StDev}\left(h_{j(i)}^{dest}\right)} = \frac{\frac{N}{N'} \eta_j^{dest} p(1-p)}{\text{StDev}\left(T_{ij}\right) \text{StDev}\left(\eta_{j(i)}^{dest}\right)}.$$

Canceling terms yields the desired result. □

This proposition is intuitive. It says that under our equal selection assumption, the destination component η_j^{dest}—i.e., the average unobserved, residual health in destination j—is equal to the observed term h_j^{dest} scaled by a constant. The value of that constant is the ratio of the standard deviations of $\eta_{j(i)}^{dest}$ and $h_{j(i)}^{dest}$, and it can be interpreted as the relative importance of the unobserved and observed components of health capital correlated with destinations. Assumption 2 then allows us to estimate this ratio using the analogous ratio for movers' origins.

Corollary 1. *Under Assumptions 1 and 2,*

$$\hat{\eta}_j^{dest} = \frac{\hat{\text{StDev}}\left(\tau_{j(i)}^{orig}\right)}{\hat{\text{StDev}}\left(h_{j(i)}^{orig}\right)} \hat{h}_j^{dest} \tag{6}$$

is a consistent estimator of η_j^{dest}, and $\hat{\gamma}_j = \hat{\tau}_j^{dest} - \hat{\eta}_j^{dest}$ is a consistent estimator of γ_j (where $\hat{\text{StDev}}\left(\tau_{j(i)}^{orig}\right)$ and $\hat{\text{StDev}}\left(h_{j(i)}^{orig}\right)$ are consistent estimators of the standard deviations of $\tau_{j(i)}^{orig}$ and $h_{j(i)}^{orig}$).

4 Data and Summary Statistics

4.1 Data and Variable Definitions

We use administrative data on Medicare enrollees for a 100% panel of Medicare beneficiaries — both Traditional Medicare and Medicare Advantage — from 1999 to 2014.[12]

The steps for $\eta_{j(i)}^{dest}$ are analogous.

[12] About one-third of Medicare beneficiaries are enrolled in Medicare Advantage, a program in which private insurers receive capitated payments from the government in return for providing Medicare beneficiaries with health insurance. Because insurance claims (and hence healthcare utilization measures) for enrollees in Medicare Advantage are not available, the literature on geographic variation in healthcare spending and health outcomes for Medicare enrollees has focused primarily on Traditional Medicare. However, the Medicare data do contain demographic, health and mortality information for both Traditional Medicare and Medicare Advantage enrollees.

We observe each enrollee's zip code of residence each year. We define a year t for the purposes of our analysis to run from April 1 of calendar year t to March 31 of calendar year $t+1$ since, for most years, we observe residence as of March 31st of that year.

For each enrollee, we observe time-invariant indicators for race and gender. We observe time-varying indicators for age, as well as enrollment in Medicaid (the supplemental public health insurance program for low income elderly), Medicare Parts A and B, and Medicare Advantage. We observe all claims for inpatient and outpatient care for enrollee-years in Traditional Medicare. For individuals who die during our sample, we observe the date of death.

Our primary analysis focuses on a sample of movers and non-movers defined below. We restrict attention to movers whose CZ of residence changes exactly once. For each mover, we define year t_i^* (an individual's "move year") to be the year in which their location changes and t_i^*+1 to be their first full year in the new location. For non-movers, we define t_i^* to be the second year we observe them in the data without any missing covariates, so that we can measure their characteristics in the prior year.

We use the Chronic Conditions segment of the Master Beneficiary Summary File from 1999 to 2014 to define 27 health status indicators for each person-year, with each indicator capturing the presence of a specific chronic condition. Examples include lung cancer, diabetes, and depression; the share of patients with each of these conditions and the estimated coefficients for each from the Gompertz mortality hazard model (equation (3)) can be seen in Appendix Table A.1. The algorithms defining these measures are publicly available[13] and are based on definitions used in the medical literature.[14] Importantly, because we measure observed health H_i pre-move, and equation (3) controls for origin fixed effects, we are not concerned about bias arising in our estimation from the type of place-specific measurement error of health in claims data that prior work has highlighted (Song et al. 2010; Finkelstein et al. 2016, 2017).

We measure total health care utilization for each person-year in Traditional Medicare, defined to be total inpatient and outpatient spending, adjusted for price differences following the procedure of Gottlieb et al. (2010).[15] As discussed below, we restrict our analysis to beneficiaries enrolled in

[13]See `https://www.ccwdata.org/documents/10280/19139421/original-ccw-chronic-condition-algorithms.pdf`.

[14]See `https://www.ccwdata.org/documents/10280/19139421/original-ccw-chronic-condition-algorithms-reference-list.pdf`.

[15]Specifically, we follow the approach from Finkelstein, Gentzkow, and Williams (2016), except that we exclude

Traditional Medicare during year $t_i^* - 1$. This restriction implies that total health care utilization is observed in year $t_i^* - 1$ for all individuals in our analysis sample, even if those individuals may be enrolled in Medicare Advantage (and hence have unobserved health care utilization) during years other than $t_i^* - 1$.

We define areas j to be Commuting Zones (CZs). Specifically, we use the 709 CZs defined by the Census Bureau in 2000 as an aggregation of counties; CZs are designed to approximate local labor markets and have been used previously to analyze geographic variation in life expectancy (e.g. Chetty et al. 2016b).[16]

All of the enrollee-level covariates in our analysis (i.e. H_i and X_i) are measured as of year $t_i^* - 1$. In our baseline specification, observable health (H_i) is a series of indicator variables for each of the 27 chronic conditions in the Chronic Conditions segment of the Master Beneficiary Summary file and log(utilization + 1). X_i is a set of indicators for race (white or non-white), gender, and their interaction; we also include an indicator variable for Medicaid status (as a proxy for low income), a series of indicator variables for the calendar year corresponding to t_i^*, and a constant.

4.2 Sample Restrictions and Summary Statistics

Our data contain approximately 81 million people and over 665 million person-years. We drop from this sample person-years in which the enrollee is younger than 65 or older than 99.[17] This leaves us with a core sample of about 69 million beneficiaries; we exclude a few hundred thousand beneficiaries with incomplete data.

To define our non-mover sample, we begin with the 62 million enrollees whose CZ of residence does not change over the years we observe them. We need to assign each non-mover a valid reference year t_i^* such that we are able to see observable health characteristics in year $t_i^* - 1$. We therefore eliminate all non-movers who do not have a pre-2012 year t_i^* such that they are 99 or younger and alive until the end of that year, and also on Traditional Medicare during year $t_i^* - 1$. We take a random 10% sample of the remaining 43 million non-movers and define their t_i^* to be

physician services ("carrier files") because these files are only available for a 20 percent subsample.

[16] See `https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/` for more details.

[17] Individuals younger than 65 appear in our data if they are disability-eligible (through Social Security disability benefits) rather than age-eligible for Medicare.

the second year they are in the sample. When we estimate equation (3) using the pooled sample of movers and non-movers, we upweight the non-movers by ten.

To define our mover sample, we begin with the 7 million enrollees whose CZ of residence changes at least once during our sample period. To ensure changes in address reflect real changes in location, we define a mover's "claim share" in a particular year to be the ratio of the number of claims located in their destination to the number located in either their origin or their destination. We then follow Finkelstein, Gentzkow, and Williams (2016) in excluding those for whom the claim share does not increase by at least 0.75 in their post-move years relative to their pre-move years. Appendix A.3 provides more detail.

We further restrict the sample to movers who are not on Medicare Advantage in the year immediately prior to or immediately after the move (since we need to measure claim shares in those years) and who moved in years 2000-2012 (so that we can observe pre-move characteristics and post-move mortality). We also exclude those who move at age 99 or later or do not survive through the end of their move year (t_i^*). Our final sample contains 6.3 million individuals, of whom 2 million are movers. Appendix A.3 provides more detail on the sample restrictions. By construction, we are able to observe mortality for all beneficiaries for at least one year following t_i^*. We are able to observe mortality at least 7 years after t_i^* for 63 percent of movers and at least 10 years after t_i^* for 35 percent of movers.

Because our strategy for estimating place effects requires that we observe a significant number of movers to each area, we aggregate CZs that receive small numbers of movers to form larger areas within states. Specifically, we first collect the bottom quartile of CZs by the number of incoming movers. Then, in any case where a state contains two or more such CZs, we consolidate those CZs into a single area. Appendix Figure A.1 shows the locations of the bottom quartile of CZs; they are predominantly in the Great Plains. The number of movers to these CZs ranges from 2 to 359, with a median of 155. Our final sample has 528 CZs and 35 aggregated CZs; these are the areas corresponding to the j index in our model and we refer to these simply as "CZs" in what follows.[18] Appendix Table A.2 shows summary statistics on the number of movers to each CZ; the minimum number of movers to a CZ is 48, and the median is about 1,500.

[18]Note that 11 of these bottom quartile CZs are within a single state and therefore remain disaggregated. This procedure causes us to omit roughly 3,000 movers who move across small CZs within the same state.

Table 1 reports summary statistics for comparable samples of movers and non-movers. The first row shows our full sample, which consists of roughly 2 million movers and 4.3 million non-movers. The remainder of the table shows characteristics of a sub-sample of movers and non-movers with reference year $t_i^* = 2006$. We focus on this subset to facilitate comparison of movers' and non-movers' characteristics.[19] Movers tend to be older than non-movers, are slightly more likely to be female and white, and slightly less likely to be on Medicaid. Not surprisingly given the age differences, movers are also less healthy as measured by their count of chronic conditions and their one and four year mortality.

5 Preliminary Evidence

5.1 Patterns of Mortality and Migration

Figure 1 shows our estimates $\hat{L}_j$ of average non-mover life expectancy by area, constructed from the estimated model of equation (3) as described in Section 3.1. The average life expectancy across areas is 83.3 years, with a standard deviation of 0.84 years. Our life expectancy estimates are highly correlated with the life expectancy estimates at age 40 by Chetty et al. (2016b), as shown for the 100 largest CZs in Appendix Figure A.2.

Since moves will be key to identifying place effects, we briefly discuss the characteristics of moves in our sample. There is substantial variation across moves in the destination-origin difference in non-mover life expectancy (L_j). The standard deviation of this gap is roughly one year, and the share of movers to higher life expectancy destinations (48 percent) is similar to the share of moves to lower life expectancy destinations (52 percent); Appendix Figure A.3 shows more detail on the destination-origin differences in average life expectancy. Conditional on origin, the average standard deviation of destination life expectancy across CZs is 0.74, quite close to the

[19] For completeness Appendix Table A.3 reports the same summary statistics on the full set of 2 million movers and 4 million non-movers used to estimate equation (3), but the two sets of statistics are not directly comparable given the differences in how the two samples are defined.

cross sectional standard deviation of 0.84.[20]

We next examine the extent to which the observed health of movers differs systematically according to their destinations. In Panel A of Figure 2, we compare the average observed health of movers to different destinations adjusted for age and demographics (X_i). For each area j, we compute the mean across movers to j of the residuals from a regression of our observed health index $\hat{h}_i = H_i\hat{\lambda}$ on age in year $t_i^* - 1$ and demographics X_i. The left-hand figure shows the distribution of these average values across destinations. If movers were randomly assigned to destinations, these averages should vary little; this is not the case. The right-hand figure is a binned scatterplot showing how these average observed health values for movers to different destinations are correlated with the average estimated mortality index $\gamma_j + \theta_i$ of non-movers in each destination. The relationship is significant and positive, suggesting that low-mortality destinations tend to attract healthier movers.

In Panel B of Figure 2, we partial out fixed effects for movers' origins (in addition to the age and demographics that were already partialed out in Panel A). These values capture the extent to which healthier movers from a given origin select systematically different destinations. The results indicate that conditional on origin, mover observed health is still correlated with destination mortality, but conditioning on origin lowers the slope from 0.24 to 0.15. While the selection on observed health shown here will be accounted for by the explicit H_i controls in our model, it suggests that there may be remaining selection on unobserved health which we will need to address with our selection correction strategy.

5.2 Inputs to Selection Correction

Table 2 shows the standard deviations of the components of health capital that enter our selection correction. For each component, we report the standard deviation across CZs, estimated using our split-sample strategy, as well as 95-percent confidence intervals based on our Bayesian bootstrap. The magnitudes are not easily interpretable, as they are in units of the log mortality rate $\log(m_i)$,

[20] Appendix Table A.4 shows the full transition matrix of movers by decile of life expectancy in the origin and decile of life expectancy in the destination. While moves are more common between closer deciles and from higher life expectancy origins, we find that all the cells contain a significant number of observations. There are at least several thousand people in each cell of the transition matrix, including moves from the highest life expectancy decile to the lowest, and vice versa. Appendix A.3 provides additional information on migration patterns.

but to get a sense, note that a 65-year old with average health capital $\overline{\theta}$ and sample-wide average place effect $\bar{\gamma}$ (which is 0 by construction) has an annual mortality rate of $m = 0.013$, and increasing her health capital by one standard deviation (among 65-year-olds) would increase her mortality rate by 0.005.

The first two rows report the estimated standard deviations of the components $h_{j(i)}^{orig}$ and $\eta_{j(i)}^{orig}$ correlated with movers' origins. Recall that our estimators of these terms are the origin fixed effects from equations (3) and (4) respectively. We find that the standard deviation of the unobservable component $\eta_{j(i)}^{orig}$ is 0.061, and the standard deviation of the observable component $h_{j(i)}^{orig}$ is 0.037. This suggests that, despite the richness of our observable health measures, the remaining systematic variation in health capital correlated with locations is substantial. The ratio of these terms $\frac{0.061}{0.037} = 1.65$ is the key conversion factor that is used in Corollary 1 to pin down the relative importance of unobservables and observables.

The last two rows report the estimated standard deviations of the components $h_{j(i)}^{dest}$ and $\eta_{j(i)}^{dest}$ correlated with movers' destinations. The $h_{j(i)}^{dest}$ components are estimated by the destination fixed effects in equation (4); we find that their standard deviation is 0.024. The $\eta_{j(i)}^{dest}$ components cannot be directly estimated, and are the key objects our selection correction is designed to infer. Applying Corollary 1, we estimate that the standard deviation of $\eta_{j(i)}^{dest}$ is $0.024 \times \frac{0.061}{0.037} = 0.040$.

6 Main Results

6.1 Place Effects

Table 3 reports our decomposition of the area average mortality index $\gamma_j + \overline{\theta}_j$. As shown in the first row, the standard deviation across CZs of this index is 0.105.

The following three rows report the decomposition of this index when we do not apply our selection correction—i.e., when we assume $\eta_j^{dest} = 0$ for all j. In this case, our estimate of the place effects γ_j is simply the destination fixed effects $\hat{\tau}_j^{dest}$ from equation (3), and average health capital $\overline{\theta}_j$ is given by the average value of the remaining terms in that equation (excluding the age term $a_i\beta$, and taking the national average of race and sex as discussed in Section 3.1). In this case, we estimate that the standard deviation of the place effects is 0.077, or three-quarters of the

standard deviation of the overall index. The standard deviation of average health capital is 0.080, and the correlation between the two components is slightly negative.

The bottom three rows report our preferred estimates applying the selection correction. Here, our estimate of the place effects γ_j is the difference $\hat{\tau}_j^{dest} - \hat{\eta}_j^{dest}$, where the unobservable component $\hat{\eta}_j^{dest}$ is inferred following the steps broken out in Table 2. Average health capital $\overline{\theta}_j$ is again given by the average value of the remaining terms in equation (3) (excluding the age term $a_i\beta$, and taking the national average of race and sex as discussed in Section 3.1). The standard deviation of the selection-corrected place effects is 0.054, about one-third smaller than the uncorrected version, and roughly half the standard deviation of the overall index. The standard deviation of average health capital is 0.094, and the correlation between the two components is now positive.

Figure 3 shows a map of our estimated treatment effects $(L_j^* - \overline{L})$. These are defined in Section 2 and capture the impact of moving to an area on life expectancy for a mover with average health capital. The most favorable effects are found in the Northeast and along the eastern seaboard, down through parts of Florida. The most adverse effects, meanwhile, are concentrated in the deep south (Alabama, Arkansas, Georgia, Louisiana, and parts of Florida) and in the Southwest (Texas, Oklahoma, New Mexico, and Arizona).

Figure 4 shows a scatterplot of these treatment effects against estimated average life expectancy $\hat{L}_j$ in each place. The two are positively correlated: a one unit increase in average life expectancy is associated with a 0.23 year increase in the treatment effect. Interestingly, for Medicare survivors of Hurricane Katrina, Deryugina and Molitor (2018) estimate somewhat larger effects. They find that moving to a place with a 1 percentage point higher mortality rate is associated with an increase in migrant mortality of approximately 1 percentage point. The fact that they find larger effects could reflect the fact that our estimates are adjusted for selection, as well as the specific sub-sample of destinations that their migrants move to.

Figure 4 also shows a number of examples that highlight how average life expectancy and treatment effects can diverge. For example, Charleston, West Virginia is a place that in the cross-section has low average life expectancy, despite a relatively favorable treatment effect. The gap reflects Charleston's unusually poor average health capital. At the other extreme, Sante Fe is an example of a place with relatively high average life expectancy despite a negative treatment effect. The gap reflects the unusually good health capital of Sante Fe residents.

Table 4 reports the CZs with the ten most favorable and ten most adverse treatment effects. For comparison, we also report average life expectancy in each place. The treatment effects of the ten most favorable places range from 0.85 to 1.26 years, with the five best CZs all in New York and Florida. The treatment effects of the ten least favorable places range from -0.75 to -0.70 years; the two worst CZs are Lake Charles, LA and Beaumont, TX.

Table 5 summarizes our estimated treatment effects. The top panel reports the standard deviation across CZs of average life expectancy, which is 0.84 years. The second row shows the standard deviation of our estimated treatment effects, which is 0.44, or roughly half of the cross-sectional variation in life expectancy.

To translate these estimates into the impact on life expectancy from moving from a place at one part of the distribution of treatment effects to another, we assume the treatment effects are normally distributed with a standard deviation equal to our preferred estimate in row (2) of the table. This provides a simple summary that is not sensitive to adjusting individual estimates for sampling error. This exercise suggests that moving from a 25th percentile area to a 75th percentile area would increase life expectancy by 0.60 years; moving from a 10th to a 90th percentile area would increase life expectancy by 1.1 years, or half the cross sectional 90-10 gap in life expectancy.

The final rows of the table show how much of the cross-sectional variation in life expectancy can be explained by our treatment effects. We find that about 15 percent of the cross-CZ variance in life expectancy would be eliminated if place effects were made equal across areas (with the observed variation in health capital remaining the same). Conversely, we find that about 75 percent of the variation would be eliminated if health capital were equalized (with the observed variation in the causal effects of place remaining the same).[21]

6.2 Heterogeneity

Previous work has found that geographic variation in life expectancy is higher for lower-income individuals (Chetty et al. 2016b). We replicate this result here, and examine to what extent it results from different variances of place effects and health capital respectively. We restrict attention to the 100 largest CZs (which constitute about half of the non-mover population) to ensure sufficient

[21] Note that these shares need not sum to 1, both because of the non-zero correlation between average health capital and place effects and because of the non-linear translation into life expectancy.

sample sizes to estimate treatment effects for each subgroup.

Table 6 summarizes the results. The first column shows that our main results are similar in this restricted sample. The remaining columns re-estimate the model separately by race and by Medicaid enrollment (an indicator of low socio-economic status), partitioning both movers and non-movers. Row (2) is consistent with the prior Chetty et al. (2016b) finding: the standard deviation of life expectancy is larger for individuals on Medicaid compared to those not on Medicaid, and larger for non-white individuals compared to white individuals. We estimate that the standard deviation of health capital effects is larger for Medicaid enrollees compared to non-Medicaid (see row 4), while the standard deviation of treatment effects is quite similar (row 3). Similar patterns also are apparent for non-whites compared to whites, although the results are less precise.

These estimates suggest that the greater geographic variation in life expectancy for low-income populations may be primarily driven by greater variation in their health capital, rather than by greater variation in treatment effects of place. This is consistent with evidence in Chetty et al. (2016b) suggesting that variation in area life expectancy for low-income individuals is strongly correlated with health behaviors such as smoking and exercise.

6.3 Dynamics

We consider an alternative binary Logit model of mortality, in which the outcome is mortality within a fixed window of n years. This allows us to estimate effects separately for different window lengths n, providing insight into the path of mortality effects in the years following moves. It also provides a check on the robustness of our results to the Gompertz functional form assumed in our main model.

We replace estimating equation (3) with a binary Logit model of n-year mortality. All covariates are the same as in equation (3) except that we include in the X_i a fully interacted set of five year age bins, race, and sex, rather than including age linearly and interacting race and gender. We estimate the Logit model for 1-year, 2-year, 3-year, and 4-year mortality.

Table 7 reports the results. The first row reports our baseline estimates of the standard deviation of the mortality index $(\gamma_j + \overline{\theta}_j)$ and the standard deviation of the selection-corrected place effects γ_j from Table 3. In our baseline, the standard deviation of γ_j is about half the standard deviation

of $\gamma_j + \overline{\theta}_j$. The last four rows show the results of the Logit model for different horizons.

There is no evidence that the impact of place attenuates with time. These results are consistent with our interpretation of γ_j as the on-impact effect of current location. While it is possible that location may influence health behaviors and other determinants of health capital over longer time horizons, the sharp on-impact effect of place we observe suggests such a channel is unlikely to be a source of substantial bias in our results. It also suggests that the way that place affects life expectancy for the elderly is primarily by affecting the arrival rate of health shocks (e.g. via temperature or pollution) or the response to those shocks (e.g. via healthcare).

6.4 Correlates of Treatment Effects

To provide some suggestive evidence on what may drive the treatment effects we estimate, we explore their correlation with various observable place characteristics. In keeping with the existing literature, we focus primarily on observables that proxy for the environment and for medical care. We present detailed definitions, data sources, and summary statistics for these measures in Appendix A.4.

Figure 5 reports bivariate correlations of both average life expectancy and our estimated treatment effects with various area level characteristics. Each place characteristic has been normalized to have mean zero and standard deviation one. We emphasize that these are simply correlations and need not reflect causal effects. Still, most of the results follow intuitive patterns.

The top panel shows that places with favorable treatment effects tend to have higher quality and quantity of health care. Treatment effects are significantly positively correlated with hospital quality (as measured by the Hospital Compare score), primary care physicians per capita, and specialists per capita. Areas with favorable treatment effects have fewer hospital beds per capita.

Measures of utilization – including utilization itself, along with imaging tests and diagnostic tests – are also positively correlated with our treatment effects, though the magnitudes are smaller than they are for hospital quality or physician quantity. Our finding of a positive correlation between an area's health care utilization and its estimated impact on life expectancy is intriguing in light of the well-documented finding that places with higher healthcare utilization do not have better average health outcomes (Fisher et al. 2003a,b; Skinner 2011). Figure 5 shows that we replicate

this finding, with the correlation between utilization and average life expectancy estimated to be precisely zero.

The bottom panel examines correlates with various non-healthcare area characteristics. Areas with favorable place effects on life expectancy tend to have less pollution, less extreme summer and winter temperatures, fewer homicides, and fewer automobile fatalities. They also tend to have higher income and education, which could reflect either greater demand for quality health care and amenities that reduce mortality or sorting of people with higher incomes and more education to high-treatment-effect areas. These areas also tend to exhibit better health behaviors (more exercise, less smoking, and lower obesity), which may similarly reflect either demand or sorting. Places with higher shares of urban populations tend to have more favorable treatment effects. The share of people over the age of 60 is uncorrelated with our treatment effects.

In general, the correlation of the characteristic with the estimated place component of life expectancy is smaller (in absolute value) than the correlation with the cross-sectional life expectancy. This difference is particularly pronounced for health behaviors and demographics, consistent with the raw correlations reflecting not only the causal effects but also the direct impacts of these variables on health capital.

7 Validation and Robustness

7.1 Additive Separability

Equation (1) assumes that health capital and current place have additively separable effects on log mortality. As discussed above, we consider this a strong assumption but one that is attractive economically since it has the intuitive implication that health capital and current location affect the *level* of mortality multiplicatively. Thus, the level of mortality of individuals with poor health capital (high θ_i) will vary more across areas than that of individuals who have better health capital.

One way to assess the validity of the assumption that place effects are separable from health capital is to test whether these place effects differ across subsets of enrollees. We construct four partitions of our mover sample based on move year, gender, age at move, and individual health at move. Each partition results in two groups with approximately the same number of movers; we

estimate the model separately for movers in each group. For each partition, we use two summary statistics to evaluate the stability of place effects across the two groups. Appendix Table A.5 shows the results.

First, we analyze the standard deviation of place effects for each group. For five of the eight groups the estimated standard deviations fall within the confidence interval $[0.038, 0.067]$ of our baseline estimates. The three exceptions are "young movers" (standard deviation $= 0.075$), movers in "good health" (standard deviation $= 0.101$), and male movers (standard deviation = 0.068).

Second, we examine the correlation of place effects between the two groups. The correlation of the place effects between the two subsamples ranges from 0.16 (when we partition by individual health) to about 0.24 (when we partition by gender or move year). To assess these correlations, we need to adjust for the role of sampling error, as it reduces the correlation between any two independent subsamples even if the true place effects are the same. Appendix Figure A.4 compares the estimated correlations to the distribution of correlation coefficients produced by randomly partitioning the mover sample into two equally sized groups and re-estimating the model 200 times. The median correlation of place effects between two random partitions is 0.29. For partitions based on age, move year, and gender, the correlation coefficients are within the 95% confidence interval formed from the distribution of correlation coefficients from the random partitions. Only the correlation coefficient for the partition based on individual health is outside of this interval. Overall, this evidence supports the view that any deviations from additive separability may be modest.

7.2 Selection Correction Assumptions

The key novel assumption in our selection correction strategy is Assumption 2: that the relative importance of the unobserved and observed components of health capital correlated with movers' destinations is the same as the relative importance of the components correlated with movers' origins.

One way to provide support for this assumption is to ask whether the analogous condition would hold if some of our observed health measures had in fact been unobserved. That is, suppose we divide H_i into K subsets H_i^k. For each subset, we imagine a hypothetical world where the elements of H_i^k are the unobservables and the elements of $H_i^{-k} = H_i \setminus H_i^k$ are the observables, so

the analogues of h_i and η_i would be $h_i = H_i^{-k}\lambda^{-k}$ and $\eta_i = H_i^k\lambda^k$ (where λ^{-k} and λ^k are the appropriate sub-vectors of λ). Denote the associated origin and destination components by $h_{j,k}^{dest}$, $h_{j,k}^{orig}$, $\eta_{j,k}^{dest}$, $\eta_{j,k}^{orig}$. We would like to confirm that

$$\frac{\text{StDev}\left(\eta_{j(i),k}^{orig}\right)}{\text{StDev}\left(h_{j(i),k}^{orig}\right)} \approx \frac{\text{StDev}\left(\eta_{j(i),k}^{dest}\right)}{\text{StDev}\left(h_{j(i),k}^{dest}\right)} \forall k.$$

To implement this test, we define 10 different subsets H_i^k, each of which is a random draw of 13 of the 27 total conditions. In each case we include log utilization in H_i^{-k}. For each subset, we estimate Equation (3) and compute $\hat{\eta}_{j,k}^{orig} = \hat{\tau}_{j,k}^{orig}$, $\hat{h}_i = H_i^{-k}\hat{\lambda}^{-k}$, and $\hat{\eta}_i = H_i^k\hat{\lambda}^k$. We then compute the implied $\hat{h}_{j,k}^{dest}$ and $\hat{h}_{j,k}^{orig}$ by re-estimating equation (4), and compute and $\hat{\eta}_j^{dest}$ from equation (6).

Panel (a) of Figure 6 shows the results. This figure plots $\frac{\text{StDev}\left(\eta_{j(i),k}^{orig}\right)}{\text{StDev}\left(h_{j(i),k}^{orig}\right)}$ on the x axis and $\frac{\text{StDev}\left(\eta_{j(i),k}^{dest}\right)}{\text{StDev}\left(h_{j(i),k}^{dest}\right)}$ on the y axis. If these ratios vary proportionately for any subset of health measures k, they should lie on a line that goes through the origin. The results support this; the points have a clear monotonic relationship and we estimate an intercept of -0.01.

Panel (b) of Figure 6 directly examines how our key estimates vary if we re-estimate the entire model using the different subsets of observables H_i in panel (a). It plots the distribution of these ten estimates for the standard deviation of treatment effects (left hand panel) and the correlation of the estimated treatment effects with our baseline estimates (right hand panel). The results indicate that the standard deviation of treatment effects is lowest in our baseline model, suggesting it is conservative, and that the correlation of treatment effects with the baseline is high.

Relaxing the assumptions

These results provide broad support for our key assumption, while suggesting that the true constant of proportionality in Assumption 2 may be somewhat larger than one. Our baseline assumption is that the ratios $\frac{\text{StDev}\left(\eta_{j(i),k}^{orig}\right)}{\text{StDev}\left(h_{j(i),k}^{orig}\right)}$ and $\frac{\text{StDev}\left(\eta_{j(i),k}^{dest}\right)}{\text{StDev}\left(h_{j(i),k}^{dest}\right)}$ are not just proportional, but in fact are equal. This would imply that the points in panel (a) of Figure 6 should have a slope of one. The observed slope of 1.72 is larger, but given the large standard error we cannot reject that it is equal to one. To look at this another way, Appendix Figure A.5 shows the distribution of the ratio of $\frac{\text{StDev}\left(\eta_{j(i),k}^{dest}\right)}{\text{StDev}\left(h_{j(i),k}^{dest}\right)}$ to

$\frac{\text{StDev}\left(\eta^{orig}_{j(i),k}\right)}{\text{StDev}\left(h^{orig}_{j(i),k}\right)}$ across the 10 draws. This ratio is always larger than one with a median value of 1.58.

To assess robustness to such deviations, we can consider relaxing both Assumptions 1 and 2 as follows.

Assumption 3. *(Relaxed Equal Selection)* $\text{Corr}\left(T_{ij}, \eta^{dest}_{j(i)}\right) = C_1 \text{Corr}\left(T_{ij}, h^{dest}_{j(i)}\right)$ *in the sample of movers for all* $j \in \mathscr{J}$*, where* C_1 *is a constant.*

Assumption 4. *(Relaxed Relative Importance)* $\frac{\text{StDev}\left(\eta^{dest}_{j(i)}\right)}{\text{StDev}\left(h^{dest}_{j(i)}\right)} = C_2 \frac{\text{StDev}\left(\eta^{orig}_{j(i)}\right)}{\text{StDev}\left(h^{orig}_{j(i)}\right)}$ *in the sample of movers, where* C_2 *is a constant.*

Under these relaxed assumptions, the consistent estimator of η^{dest}_j is now scaled by the factor $C_1 C_2$.

Corollary 2. *Under Assumptions 3 and 4,*

$$\hat{\eta}^{dest}_j = C_1 C_2 \frac{\widehat{\text{StDev}}\left(\tau^{orig}_{j(i)}\right)}{\widehat{\text{StDev}}\left(h^{orig}_{j(i)}\right)} \hat{h}^{dest}_j \tag{7}$$

is a consistent estimator of η^{dest}_j*, and* $\hat{\gamma}_j = \hat{\tau}^{dest}_j - \hat{\eta}^{dest}_j$ *is a consistent estimator of* γ_j *(where* $\widehat{\text{StDev}}\left(\tau^{orig}_{j(i)}\right)$ *and* $\widehat{\text{StDev}}\left(h^{orig}_{j(i)}\right)$ *are consistent estimators of the standard deviations of* $\tau^{orig}_{j(i)}$ *and* $h^{orig}_{j(i)}$*).*

We consider the sensitivity of our results to values of $C = C_1 C_2$ not equal to one. We focus on the implied variability of the place effects γ_j and of the treatment effects $L^*_j - \overline{L}$ as summary outcomes in this exercise.

The implied $\text{StDev}\left(\gamma_j\right)$ will not be monotonic in C. In fact, it is straightforward to show that $\text{StDev}\left(\gamma_j\right)$ is minimized when $C = \frac{\text{Cov}\left(\tau^{dest}_j, \eta^{baseline}_j\right)}{\text{Var}\left(\eta^{baseline}_j\right)}$, where $\eta^{baseline}_j = \frac{\text{StDev}\left(\tau^{orig}_{j(i)}\right)}{\text{StDev}\left(h^{orig}_{j(i)}\right)} h^{dest}_j$ is the population value of our baseline estimator of $\hat{\eta}^{dest}_j$ in Corollary 1.[22] In our data $\frac{\text{Cov}\left(\tau^{dest}_j, \eta^{baseline}_j\right)}{\text{Var}\left(\eta^{baseline}_j\right)} =$

[22] Since $\eta^{dest}_j = C\eta^{baseline}_j$, we have $\gamma_j = \tau^{dest}_j - C \cdot \eta^{baseline}_j$ and thus

$$\text{Var}(\gamma_j) = \text{Var}\left(\tau^{dest}_j\right) + \text{Var}\left(C \cdot \eta^{baseline}_j\right) - 2 \cdot \text{Cov}\left(\tau^{dest}_j, C \cdot \eta^{baseline}_j\right).$$

Minimizing with respect to C yields the desired result.

1.26, suggesting that assuming $C = 1$ will imply conservative estimate of the importance of place effects relative to alternatives $C < 1$ or $C > 1.26$.

Our baseline estimate of the standard deviation of life expectancy treatment effects is 0.44. Appendix Table A.6 reports this estimate for other values of C. If we choose the variance-minimizing value $C = 1.26$, the implied standard deviation falls to 0.43. If we set C equal to the median value 1.58 from Figure A.5, the implied standard deviation is 0.45. If we set C equal to 2, the implied standard deviation is 0.53. We conclude that our results are not sensitive to modest deviations from our baseline assumption $C = 1$, and that this assumption is if anything conservative in the sense that the alternatives imply even larger causal effects of place.

Furthermore, it turns out that we can derive both lower and upper bounds on the StDev (γ_j) without imposing the Equal Selection Assumption (Assumption 1) at all. These bounds turn out to be fairly tight. Our point estimate (see Table 3) is 0.054 for StDev (γ_j). As we derive in Appendix A.5, without the Equal Selection assumption we can still derive a lower bound for the StDev (γ_j) of 0.041 (and an upper bound of 0.113).

7.3 Panel vs. Cross-Section Comparison

Another way to assess validity of our method is to apply it to outcomes which, unlike mortality, are observed repeatedly for the same individual. For such outcomes, we can follow Finkelstein et al. (2016) and adjust for selection directly by including individual fixed effects. We can then compare the estimates we obtain for these outcomes using the cross-sectional selection correction of the current paper to the "gold standard" estimates obtained from the panel.

The panel regression of Finkelstein et al. (2016) is:

$$y_{ijt} = \alpha_i + \gamma_j + \omega_t + x_{it}\beta + \varepsilon_{ijt} \tag{8}$$

where y_{ijt} is an outcome observed in a panel, such as a particular measure of health care utilization; α_i, γ_j, and ω_t are individual, CZ, and calendar year fixed effects; and x_{it} consists of dummies for five-year age bins as well as fixed effects for relative year for movers.

We consider three panel outcomes y_{ijt} that we can construct using the inpatient and outpatient claims data: an indicator for any hospital admission, an indicator for any emergency room visit,

and an indicator for any outpatient visit. For each of these outcomes, we first assume that we only observe the outcome once post-move (as we do for mortality), and estimate equation (3) for the binary outcome measured one year post-move. We report results both with and without the selection correction. We then estimate equation (8) and compare.

The results are shown in Appendix Table A.7. In all cases, the selection correction moves the estimates closer to the panel estimates. For both any hospital admission and any emergency room visit, this is a substantial change, closing more than half the gap between the naive uncorrected estimates and the panel estimates. For any outpatient visit, the effect of the selection correction is smaller, though in the right direction. These results provide independent variation that our selection correction succeeds in reducing bias due to unobservables.

7.4 Robustness

Appendix Table A.8 reports a suite of additional robustness checks. For each, we report a number of key results: the standard deviation of average life expectancy $\left(L_j\right)$, the standard deviation of area treatment effects $\left(L_j^* - \overline{L}\right)$, the correlation between the treatment effects estimated in that row and the baseline treatment effects, and the correlation between average life expectancy and the treatment effects $(\text{Corr}\left(L_j, L_j^*\right))$. The first row repeats our baseline estimates for reference; once again, we focus on the 100 largest CZs.

In row (2), we estimate a variant of our baseline model that allows the coefficients on age, demographics, and health (β, ψ, and λ, respectively in equation (3)) to differ for movers and nonmovers.

In row (3), we interact the components of observed health H_i with an enrollee's age in the year prior to their reference year t_i^*. Since we define H_i as of $t_i^* - 1$ for all enrollees, our baseline specification assumes that the coefficients that relate specific chronic conditions to log mortality are independent of age. This robustness check relaxes that assumption in a limited way.

In row (4), we add average race- and sex-adjusted mortality rates in a mover's origin *county* as a control variable. This adjusts for selection of movers across different areas within origin CZs.

In row (5), we restrict the sample of moves to those of more than 100 miles, as measured between the centroids of the mover's origin and destination zip codes.

In rows (6) and (7), we focus on moves in which the gap between life expectancy in the mover's origin and in her destination is either above or below median respectively.

In row (8), we exclude any moves in which the origin CZ is geographically adjacent to the destination CZ.

In row (9), we exclude moves to Florida, Arizona, and California. This provides a check that patterns of selection specific to these popular retirement destinations are not biasing our results.

Rows (10) and (11) restrict the sample to moves occurring in 1999-2003 or to moves occurring in 2004-2012 respectively. In the latter case we define the reference year t_i^* for non-movers to be the second year they appear in the data in the 2004-2012 period.

In all of these cases, the results are qualitatively unchanged. The correlation between the estimated treatment effects and our baseline treatment effects is above 0.9 in all but two cases, and above 0.8 in all cases.

8 Conclusion

This paper documents a substantial causal impact of current locations on mortality. We estimate that moving from a 10th percentile area in terms of impact on life expectancy to the 90th percentile area would increase life expectancy at 65 by 1.1 years, or about 5 percent of average remaining life expectancy at 65. Equalizing place effects would reduce the cross-sectional variation in life expectancy at 65 by 15 percent.

The findings have two key implications for further work. First, it is important to better understand what aspects of current environments are important for life expectancy. We presented suggestive, cross-sectional evidence that places that are more favorable for life expectancy tend to have higher quantity and better quality of health care; they also have lower pollution and less extreme temperatures. More work is needed to understand the causal mechanisms.

Second, although place matters, our results also suggest an important role for heath capital in affecting life expectancy. We estimate that equalizing health capital across places would reduce the cross-sectional variation in life expectancy by three-quarters. Once again, more work is needed to understand what aspects of health capital are important causal determinants of life expectancy. Relatedly, understanding the determinants of mortality from an earlier vantage point could yield

further insights; it might well be that current environment in childhood or adulthood affects health capital (via e.g. medical investments) which in turn affects elderly mortality.

References

Ten Leading Causes of Death in the United States, 1977. Technical report, Center for Disease Control, July 1980. URL https://stacks.cdc.gov/view/cdc/7655/cdc_7655_DS1.pdf.

Joseph Altonji, Todd Elder, and Christopher Taber. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 2005.

Joshua D. Angrist, Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, 132(2): 871–919, 2017.

Bridget C. Booske, Jessica K. Athens, David A. Kindig, Hyojun Park, and Patrick L. Remington. Different perspectives for assigning weights to determinants of health. Technical report, University of Wisconsin: Population Health Institute, 2010.

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

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

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility II: County-level estimates. Working Paper 23002, National Bureau of Economics Research, 2016.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162, 2018.

Raj Chetty, Nathaniel Hendren, and Lawrence F. Katz. The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review*, 106(4):855–902, 2016a. doi: 10.1257/aer.20150572.

Raj Chetty, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. The association between income and life expectancy in the United States, 2001-2014. *Journal of the American Medical Association*, 315(16):1750–1766, 2016b. doi: 10.1001/jama.2016.4226.

Janet Currie and Hannes Schwandt. Mortality inequality: the good news from a county-level approach. *Journal of Economic Perspectives*, 30(2):29–52, 2016.

David M Cutler. The school-first solution. *Politico*, Jan 2018. URL https://www.politico.com/agenda/story/2018/01/10/cost-effective-health-care-education-000607.

Goodarz Danaei, Eric L. Ding, Dariush Mozaffarian, Ben Taylor, Jurgen Rehm, Christopher J. L. Murray, and Majid Ezzati. The preventable causes of death in the United States: Comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Medicine*, 6(4):e1000058, 2009. doi: 10.1371/journal.pmed.1000058.

Tatyana Deryugina and David Molitor. Does when you die depend on where you live? Evidence from Hurricane Katrina. Working paper 24822, National Bureau of Economics Research, 2018.

Douglas W. Dockery, C. Arden Pope, Xiping Xu, John D. Spengler, James H. Ware, Martha E. Fay, Benjamin G. Ferris, and Frank E. Speizer. An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine*, 329(24):1753–1759, 1993. doi: 10.1056/NEJM199312093292401.

Joseph J. Doyle. Returns to local-area health care spending: Evidence from health shocks to patients far from home. *American Economic Journal: Applied Economics*, 3(3):221–43, 2011.

Amy Finkelstein, Matthew Gentzkow, and Heidi Williams. Sources of geographic variation in health care: Evidence from patient migration*. *The Quarterly Journal of Economics*, 131(4): 1681, 2016. doi: 10.1093/qje/qjw023.

Amy Finkelstein, Matthew Gentzkow, Peter Hull, and Heidi Williams. Adjusting risk adjustment–accounting for variation in diagnostic intensity. *The New England journal of medicine*, 376(7): 608, 2017.

Elliott S. Fisher, David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. The implications of regional variations in medicare spending. Part 1: the content, quality, and accessibility of care. *Annals of Internal Medicine*, 138(4):273–287, 2003a.

Elliott S. Fisher, David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. The implications of regional variations in medicare spending. Part 2: the content, quality, and accessibility of care. *Annals of Internal Medicine*, 138(4):288–299, 2003b.

Victor R. Fuchs. Some economic aspects of mortality in the United States. Draft of Study Paper, National Bureau of Economic Research, 1965.

Victor R. Fuchs. *Who Shall Live? Health, Economics and Social Choice*. World Scientific Publishing Company, 2011.

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

Michael Grossman. On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255, 1972. doi: 10.1086/259880.

Paula M. Lantz, James S. House, James M. Lepkowski, David R. Williams, Richard P. Mero, and Jieming Chen. Socioeconomic factors, health behaviors, and mortality: Results from a nationally representative prospective study of US adults. *Journal of the American Medical Association*, 279(21):1703–1708, 1998. doi: 10.1001/jama.279.21.1703.

Jens Ludwig, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. Neighborhood effects on the long-term well-being of low-income adults. *Science*, 337(6101):1505, September 2012. doi: 10.1126/science.1224648.

J. Michael McGinnis and Williams H. Foege. Actual causes of death in the United States. *Journal of the American Medical Association*, 270(18):2207–2212, 1993. doi: 10.1001/jama.1993.03510180077038.

J. Michael McGinnis, Pamela Williams-Russo, and James R. Knickman. The case for more active policy attention to health promotion. *Health Affairs*, 21(2):78–93, 2002. doi: 10.1377/hlthaff.21.2.78.

Laura McGovern, George Miller, and Paul Hughes-Cromwick. The relative contribution of multiple determinants to health outcomes. *Health Affairs*, August:1–9, 2014.

Ali H. Mokdad, James S. Marks, and Donna F. Stroup. Actual causes of death in the United States, 2000. *Journal of the American Medical Association*, 291(10):1238–1245, 2004. doi: 10.1001/jama.291.10.1238.

Kevin M Murphy and Robert H Topel. Efficiency wages reconsidered: Theory and evidence. In *Advances in the Theory and Measurement of Unemployment*, pages 204–240. Springer, 1990.

Christopher Murray, Sandeep Kulkami, Catherine Michaud, Niels Tomijima, Maria Bulzacchelli, Terrell Iandiorio, and Majid Ezzati. Eight Americas: Investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Medicine*, 3(9):1513–1524, 2006.

S. Jay Olshansky and Bruce A. Carnes. Ever since Gompertz. *Demography*, 34(1):1–15, Feb 1997. doi: 10.2307/2061656.

Emily Oster. Unobservable selection and coefficient stability: Theory and evidence. Working Paper 19054, National Bureau of Economics Research, 2016.

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

Jonathan M. Samet, Francesca Dominici, Frank C. Curriero, Ivan Coursac, and Scott L. Zeger. Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. *New England Journal of Medicine*, 343(24):1742–1749, 2000. doi: 10.1056/NEJM200012143432401.

Jonathan Skinner. Causes and consequences of regional variations in health care. In Thomas G. McGuire Mark V. Pauly and Pedro P. Barros, editors, *Handbook of Health Economics*, volume 2, pages 45–93. Elsevier, 2011.

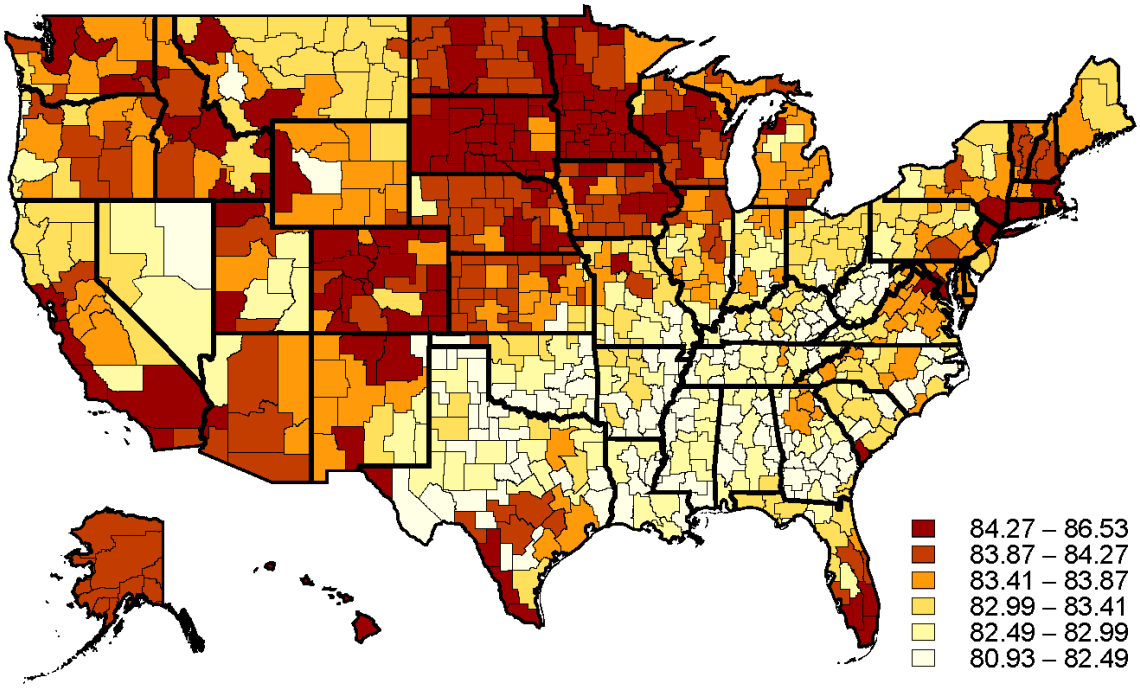
Yunjie Song, Jonathan Skinner, Julie Bynam, Jason Sutherland, and Elliott Fisher. Regional variations in diagnostic practices. *New England Journal of Medicine*, 363(1):45–53, 2010.

Silvia Stringhini, Severine Sabia, and Martin Shipley. Association of socioeconomic position with health behaviors and mortality. *Journal of the American Medical Association*, 303(12):1159–1166, 2010. doi: 10.1001/jama.2010.297.

Peggy A. Thoits. Stress and health: Major findings and policy implications. *Journal of Health and Social Behavior*, 51(1_suppl):S41–S53, 2010. doi: 10.1177/0022146510383499. PMID: 20943582.

WHO. Global health risks: Mortality and burden of disease attributable to selected major risks. Technical report, World Health Organization, 2009. URL https://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf?ua=1.

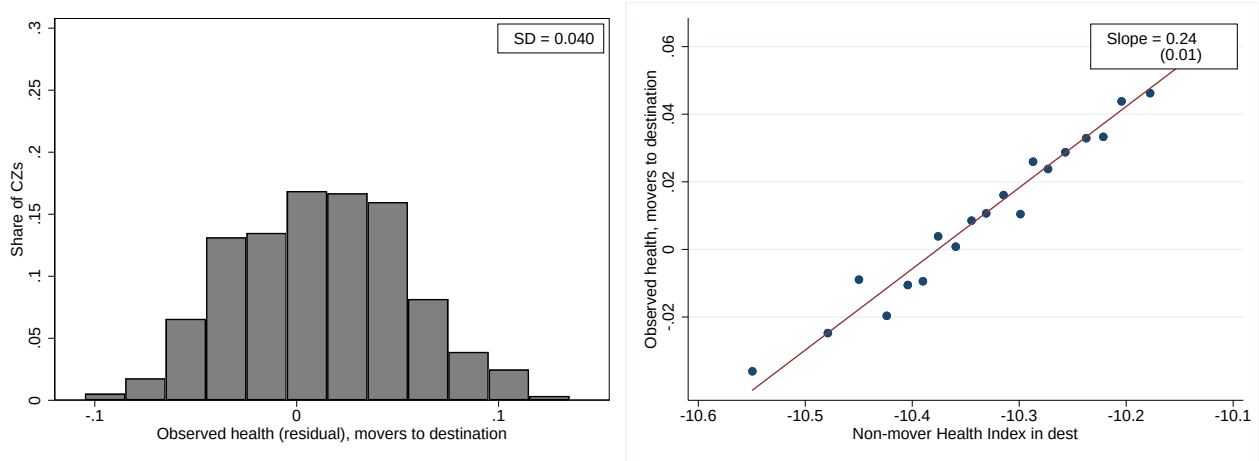
Figure 1: Age 65 Life Expectancy



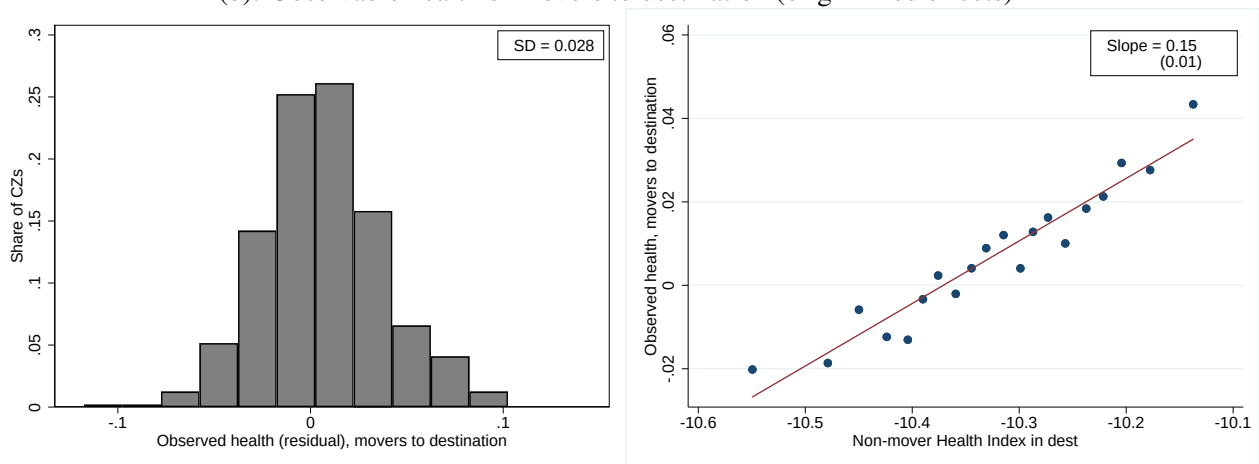
Notes: Figure reports estimated life expectancy at age 65 for non-movers in each CZ (L_j). Average life expectancy at 65 in each CZ is computed assuming a non-mover with the average characteristics in that CZ, except for race and sex for which national averages are used. Note that small CZs have been aggregated within each state (see Appendix Figure A.1) and a single life expectancy estimate is reported for each aggregate CZ.

Figure 2: Observable Health and Non-Mover Mortality

(a): Observable health of movers to destination

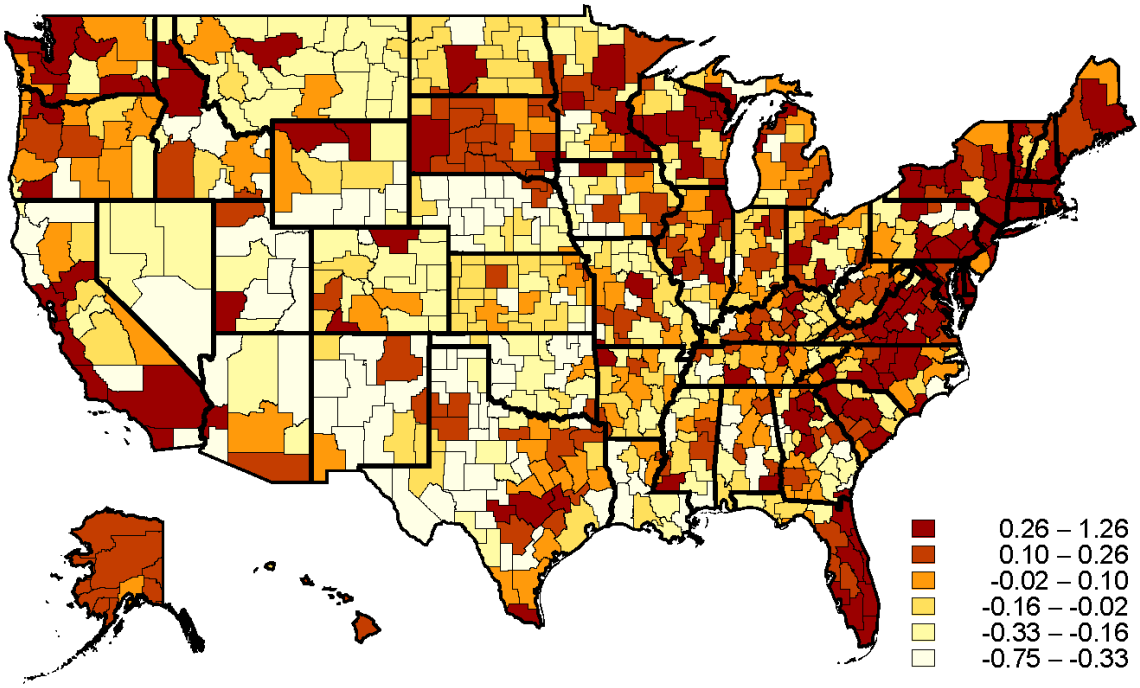


(b): Observable health of movers to destination (origin fixed effects)



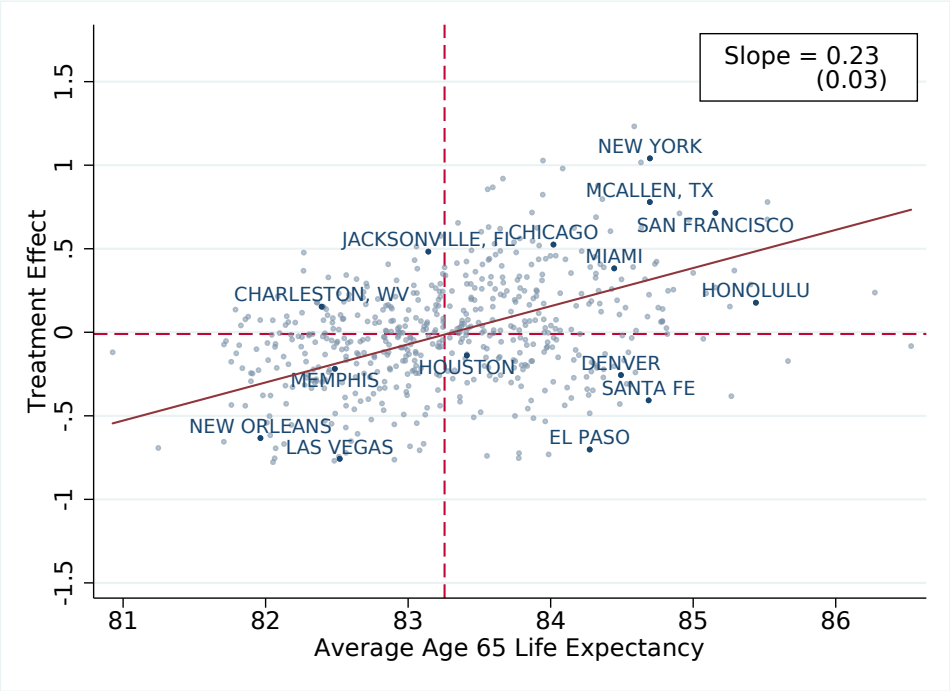
Notes: The left panels plot the distribution across CZs of the average observable, residualized health of movers to CZ j. Specifically, the top left panel plots average residual observed health ($\hat{h}_i$), conditional on X_i and age. Bottom left plots $\hat{h_j^{dest}}$ as defined in the text, and so also conditions on origin fixed effects in addition to X_i and age. All estimates are normalized such that the mean (across movers) of each is zero; both panels also report the cross-CZ standard deviation. The right panels show binned scatter plots of these two measures of average, residualized observable health of movers to CZ j against the average mortality index in CZ j ($\overline{m}_j$). The average mortality index estimates come from the sample and model estimates of only non-movers (i.e. the same estimates as are used in Figure 1). Regression line and standard error are both estimated using the CZ level data.

Figure 3: Life Expectancy Treatment Effects



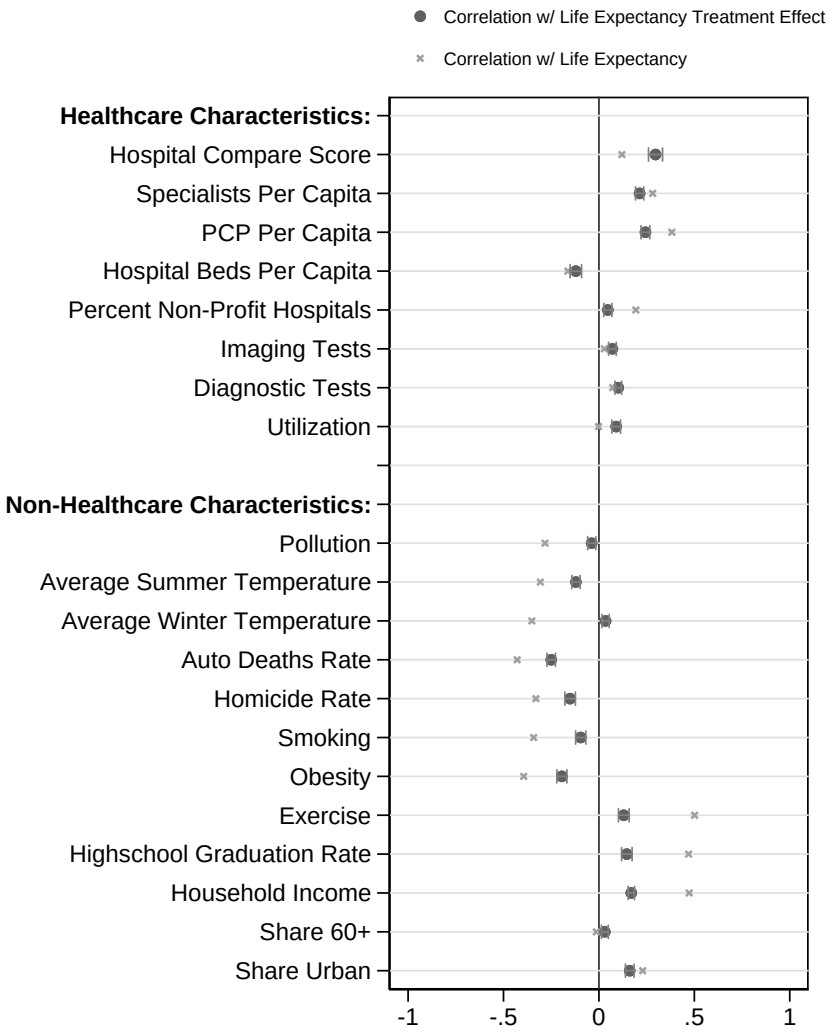
Notes: The map shows the Empirical Bayes-adjusted estimates of life-expectancy treatment effects ($L_j^* - \overline{L}$). Note that small CZs have been aggregated within state (see Appendix Figure A.1) and a single life expectancy estimate is reported for this aggregate CZ.

Figure 4: Life Expectancy Treatment Effects vs. Life Expectancy



Notes: The plot shows a scatterplot of the Empirical Bayes (EB)-adjusted age 65 life expectancy treatment effects for CZ j ($L_j^* - \overline{L}$) on the average age 65 non-mover life expectancy (L_j). The line of best fit comes from a regression of non-EB-adjusted treatment effects on average non-mover life expectancy. The horizontal and vertical dashed lines show the medians of treatment effects and life expectancy, respectively, over all CZs.

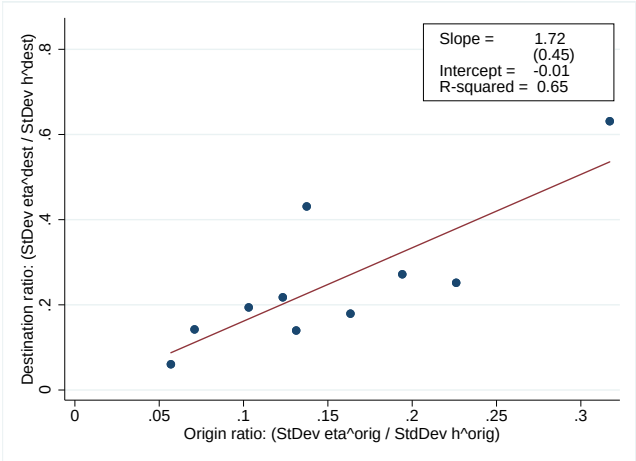
Figure 5: Correlations with Place Characteristics



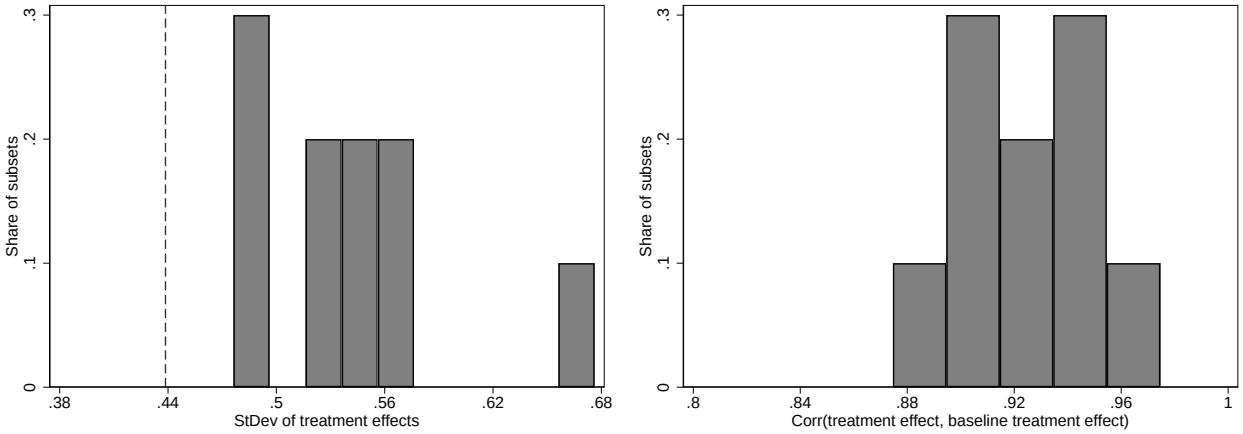
Notes: The dots in this panel report bivariate OLS regression results of our life expectancy treatment effects ($L_j^* - \overline{L}$) on z-scores of the indicated place characteristic; Appendix A.4 provides more detail on their definitions. The x marks in this panel report bivariate OLS regression results of our age 65 life-expectancy estimates (L_j) on z-scores of the indicated place characteristic. All regressions are at the CZ level, and the treatment effect regressions are weighted by the inverse variance of the treatment effects. In this figure, the sample for each bivariate regression is all CZs for which that place characteristic is defined (see Appendix Table A.10 column 3), although the results are nearly identical if we instead use the 554 CZs for which every place characteristic (except homicide rates) is defined.

Figure 6: Support for Selection-Correction Assumptions

(a): Destination and origin ratios of $\frac{StDev(\eta_{j(i)})}{StDev(h_{j(i)})}$



(b): Alternate Observable Health Characteristics



Notes: Panel (a) plots $\frac{\text{StDev}\left(\eta^{dest}_{j(i),k}\right)}{\text{StDev}\left(h^{dest}_{j(i),k}\right)}$ against $\frac{\text{StDev}\left(\eta^{orig}_{j(i)},k\right)}{\text{StDev}\left(h^{orig}_{j(i),k}\right)}$ for 10 different subsets H_i^k; each point in the scatter plot represents a different defintion of k. For each k, H_i includes log(overall utilization) and a random subset of 13 of the 27 chronic conditions. Panel (b) reports various summary statistics about the treatment effects $(L_j^* - \bar{L})$ produced by each of the 10 different definitions of k in panel (a). The left figure in panel (b) plots the standard deviation across CZs of the treatment effects from each of these alternate specifications; the dotted line shows the standard deviation across CZs of the treatment effects in the baseline specification (Table 5). All standard deviations are computed using the split-sample approach. The right figure in panel (b) plots the correlation of the treatment effects in each of the alternate specifications with the baseline treatment effects.

Table 1: Summary Statistics

	(1) Movers	(2) Non-movers
Estimation sample (# of individuals)	2,033,263	4,312,726
2006 comparison sample (# of individuals)	168,853	168,853
Age:		
65-74	0.48	0.54
75-84	0.35	0.35
85+	0.17	0.11
Female	0.60	0.57
White	0.89	0.86
Region:		
Northeast	0.19	0.20
South	0.43	0.37
Midwest	0.19	0.27
West	0.19	0.16
On Medicaid	0.10	0.12
Avg. # of chronic conditions	3.08	2.82
One year mortality	0.09	0.05
Four year mortality	0.29	0.20

Notes: The first row shows the sample size for the full estimation sample. The summary statistics on movers in the comparison sample are restricted to those who moved in the year 2006 (8.30% of movers in the whole sample). A random subset of non-movers that meet sample restrictions in 2006 are included, with their reference years set to 2006, such that the number of movers is equal to the number of non-movers. Rows for female, white, age, and region report the share of individuals with the given characteristics. Time-varying characteristics are measured in the year prior to each enrollee's reference year.

Table 2: Inputs to Selection Correction

	Standard Deviation
Origin Components:	
Observed health ($h^{orig}_{j(i)}$)	0.037 [0.036, 0.037]
Unobserved health ($\eta^{orig}_{j(i)}$)	0.061 [0.058, 0.064]
Destination Components:	
Observed health ($h^{dest}_{j(i)}$)	0.024 [0.024, 0.025]
Unobserved health ($\eta^{dest}_{j(i)}$), *inferred based on Corollary* 1	0.040 [0.038, 0.043]

Notes: Standard deviations are computed using the split-sample approach, and are weighted by the number of movers with each CZ as their destination. Confidence intervals are computed using 200 replications of the Bayesian bootstrap.

Table 3: Mortality Components

	Standard Deviation
Mortality Index $\left(\gamma_j + \overline{\theta}_j\right)$	0.105
Unadjusted:	
Place Effects (γ_j)	0.077 [0.068, 0.086]
Health Capital $\left(\overline{\theta}_j\right)$	0.080 [0.069, 0.088]
Correlation of γ_j and $\overline{\theta}_j$	-0.008 [-0.090, 0.074]
Selection Corrected:	
Place Effects (γ_j)	0.054 [0.038, 0.067]
Health Capital $\left(\overline{\theta}_j\right)$	0.094 [0.083, 0.103]
Correlation of γ_j and $\overline{\theta}_j$	0.020 [-0.082, 0.106]

Notes: These standard deviations across CZ give equal weight to each CZ and use the split-sample approach. 95% confidence intervals are computed using 200 replications of the Bayesian bootstrap. For the "unadjusted" results in the top panel, γ_j is defined as the destination fixed effects $\hat{\tau}_j^{dest}$ from equation (3), and average health capital $\overline{\theta}_j$ is given by the average value of the remaining terms in that equation (excluding the age term $a_i\beta$). For the "selection corrected" results in the bottom panel, γ_j is defined as the difference $\hat{\tau}_j^{dest} - \hat{\eta}_j^{dest}$, where $\hat{\tau}_j^{dest}$ is the destination fixed effect from equation (3) and the unobservable component $\hat{\eta}_j^{dest}$ is inferred following the steps broken out in Table 2; average health capital $\overline{\theta}_j$ is then calculated using the same approach as in the unadjusted results.

Table 4: Largest and Smallest Life Expectancy Treatment Effects

(1)	(2)	(3)
CZ Name	Treatment Effect	Average Life Expectancy
	Ten Largest	
Yonkers, NY	1.26	84.59
New York, NY	1.07	84.70
Syracuse, NY	1.05	83.95
Port St. Lucie, FL	1.04	84.63
North Port, FL	1.01	84.08
Harrisonburg, VA	0.95	83.66
Bridgeport, CT	0.90	84.33
Wilmington, NC	0.89	83.59
Asheville, NC	0.88	83.56
Lancaster, PA	0.85	83.94
	Ten Smallest	
Shreveport, LA	-0.70	82.56
Gallup, NM	-0.70	83.78
Salt Lake City, UT	-0.71	83.99
Joplin, MO	-0.72	82.51
Emporia, KS	-0.73	83.78
Gulfport, MS	-0.73	82.06
Las Vegas, NV	-0.73	82.52
Bakersfield, CA	-0.74	82.90
Beaumont, TX	-0.74	82.48
Lake Charles, LA	-0.75	82.05

Notes: This table reports the Empirical Bayes-adjusted life-expectancy treatment effect ($L_j^* - \overline{L}$) for the CZs with the ten highest and ten lowest Empirical Bayes-adjusted estimates.

Table 5: Life Expectancy Decompositions

Cross-CZ standard deviation of:		
(1)	Age 65 Life Expectancy (L_j)	0.84
(2)	Treatment Effects ($L_j^* - \overline{L}$)	0.44 [0.30, 0.54]
(3)	Health Capital Effects	0.78 [0.70, 0.85]
(4)	Correlation of Treatment and Health Capital Effects	-0.04 [-0.13, 0.08]
Share variance would be reduced if:		
(5)	Place Effects were Made Equal	0.13 [-0.02, 0.31]
(6)	Health Capital was Made Equal	0.73 [0.59, 0.87]

Notes: All objects are computed at the CZ level using the split-sample approach described in Section 3.1 and give equal weight to each CZ; 95% confidence intervals are computed via 200 replications of the Bayesian bootstrap. In row (2), we compute the standard deviation of life expectancy if health capital were held constant; specifically, for each CZ j, we compute the counterfactual age 65 life expectancy if each CZ had its own γ_j but the nationally representative health capital $\bar{\theta}$ as defined in the text. In row (3), we compute the standard deviation in life expectancy if the place effects were held constant; specifically, we define the nationally representative place effect as the median of γ_j among matched non-movers, and for each CZ j, compute the counterfactual age 65 life expectancy where the CZ has its own $\bar{\theta}_j$, but a nationally representative place effect. Row (4) reports the correlation between the health capital component of life expectancy (whose standard deviation is shown in row 3) and the place component of life expectancy (whose standard deviation is shown in row 2). In row (5) we show the share of the variance that would be reduced if place effects were made equal; this is computed by calculating the variance of life expectancy with place effects held constant (i.e. the square of row 3) and the variance in life expectancy (i.e. the square of row 1), and taking 1 minus the ratio of these numbers. Row (6) is computed in an analogous fashion.

Table 6: Heterogeneity by Medicaid Status and Race

		Medicaid Status		**Race**	
	Baseline	Non-Medicaid	Medicaid	White	Non-White
(1) Number of movers	710,990	650,246	60,744	629,126	81,864
(2) StDev of life expectancy (L_j)	0.67	0.64	1.57	0.57	1.44
Cross-CZ standard deviation of:					
(3) Treatment effects ($L_j^* - \bar{L}$)	0.47 [0.40, 0.53]	0.46 [0.38, 0.54]	0.72 [0.37, 1.00]	0.48 [0.41, 0.54]	0.73 [0.00, 1.17]
(4) Health capital effects	0.55 [0.47, 0.61]	0.49 [0.40, 0.56]	1.55 [1.31, 1.83]	0.46 [0.38, 0.53]	1.01 [0.47, 1.35]

Notes: This table summarizes the decompositions for the largest 100 CZs by population in 2000, estimated separately by race and Medicaid status during the year prior to the reference year. Both non-mover and mover samples are partitioned by race or Medicaid satus. Sample sizes in row (1) exclude movers to or from any CZ outside of the 100 largest CZs; this leaves us with about one-third of the baseline mover sample. Row (2) shows the cross-CZ standard deviation of life expectancy at 65 among non-movers in the indicated sample. The standard deviations of treatment effects and health capital effects in rows (3) and (4) are computed using the split-sample approach, giving equal weight to each CZ. Brackets show the 95% confidence intervals computed via 100 iterations of the Bayesian bootstrap. Since standard deviations cannot be negative, any split-sample approach that produces a negative result we set to 0.00.

Table 7: Logistic Model

	(1)	(2)
	Standard Deviation of Mortality Index $\left(\gamma_j + \bar{\theta}_j\right)$	Standard Deviation of Place Effects $\left(\gamma_j\right)$
Baseline	0.105	0.054 [0.038, 0.067]
Logistic Model:		
1-year mortality	0.063	0.081 [0.072, 0.091]
2-year mortality	0.069	0.073 [0.061, 0.079]
3-year mortality	0.078	0.083 [0.078, 0.088]
4-year mortality	0.088	0.082 [0.075, 0.089]

Notes: Column (1) reports the cross-CZ standard deviation of our mortality index. Row (1) reports results for the baseline Gompertz specification (See Table 3). For the baseline specification, the mortality index is age, race, and sex adjusted; for the logit specifications it is not. Furthermore, in the logit specifications, rather than a $\beta \cdot t$ term, five-year age bins are fully interacted with race and sex. Column (2) reports the cross-CZ standard deviation of our place effects. 95% confidence intervals are reported underneath the point estimate in Column (2), and are computed with 20 replications of the Bayesian bootstrap. All standard deviations of γ_j are computed using the split-sample approach, and all standard deviations in both columns give equal weight to each CZ.

A Appendix

A.1 Empirical Bayes Adjustment

Our Empirical Bayes adjustment follows Chetty and Hendren (2016). This appendix describes the approach in more detail.

Let γ_j be the true life expectancy treatment effects with mean 0. Let M be the average causal place effect which, by construction, is also 0. There is no measurement error in M. We assume that γ_j is a normally distributed random variable, so that

$$\gamma_j = M + \eta_j \tag{9}$$

with $\eta_j \sim N\left(0, \chi^2\right)$.

Further, assume that the unbiased estimates of γ_j are subject to idiosyncratic measurement error:

$$\hat{\gamma}_j = \gamma_j + \upsilon_j \tag{10}$$

where the estimation error $\upsilon_j \sim N\left(0, s_j^2\right)$ and s_j is the standard error of γ_j from the bootstrap.

Combining equations (9) and (10) implies:

$$\hat{\gamma}_j = M + \eta_j + \upsilon_j \tag{11}$$

and using OLS we are able to estimate $\text{Var}\left(\eta_j + \upsilon_j\right)$ as $\widehat{\text{Var}\left(\eta_j + \upsilon_j\right)}$. Note that in our setting, $\widehat{\text{Var}\left(\eta_j + \upsilon_j\right)} = \text{Var}\left(\hat{\gamma}_j\right)$.

With these assumptions, we are able to compute:

$$\chi^2 = \text{Var}\left(\eta_j\right) = \text{Var}\left(\eta_j + \upsilon_j\right) - E\left[s_j^2\right] \tag{12}$$

Optimal linear predictions We compute forecasts γ_j^{EB} of each CZ's true causal effect γ_j that minimize the mean squared prediction error:

$$\sum_{j=1}^{J} \left(\gamma_j^{EB} - \gamma_j\right)^2 \tag{13}$$

Note that the (unobserved) true causal effect of moving to j can be written as:

$$\gamma_j = \beta_{1,j} \cdot M + \beta_{2,j} \cdot \hat{\gamma}_j \tag{14}$$

A hypothetical OLS regression across the 563 CZs to estimate the 563 $\beta_{1,j}$ coefficients and the 563 $\beta_{2,j}$ coefficients allows us to form predictions of the true causal effects, γ_j, using M and $\hat{\gamma}_j$, which we call γ_j^{EB}.

$$\gamma_j^{EB} = \hat{\beta}_{1,j} \cdot M + \hat{\beta}_{2,j} \cdot \hat{\gamma}_j \tag{15}$$

Note that these predictions, γ_j^{EB}, would minimize the objective function in equation (13). Given a way to estimate these coefficients, we can directly compute the optimal forecasts. However,

because γ_j is unobserved, we cannot simply estimate the coefficients in an OLS regression. Instead, we use the derivation of these coefficients as in Chetty and Hendren (2016):

$$\gamma_j^{EB} = \frac{\chi^2}{\chi^2 + s_j^2} \cdot \hat{\gamma}_j + \frac{s_j^2}{\chi^2 + s_j^2} M \tag{16}$$

Because in our setting $M = 0$, this simplifies to:

$$\gamma_j^{EB} = \frac{\chi^2}{\chi^2 + s_j^2} \cdot \hat{\gamma}_j \tag{17}$$

A.2 Microfoundation for Assumptions 1 and 2

In this section, we show a natural set of assumptions on the selection process under which Assumptions 1 and 2 are guaranteed to hold. The key condition is that selection works only through the single index of overall health capital.

We begin with an underlying population of movers with health capital $\theta_i = h_i + \eta_i$, where h_i and η_i are observed and unobserved components respectively. For simplicity, we ignore the role of demographics and set $X_i\beta = 0$. Following the approach of Section 2, we define η_i to be a residual orthogonal to h_i, so that any unobserved determinants of health capital correlated with the observed measures are absorbed in h_i, and η_i only includes the components not predictable from observables. We go beyond the structure imposed above to assume h_i and η_i are independently normally distributed in the population, with $h_i \sim N(0, \sigma_h)$ and $\eta_i \sim N(0, \sigma_\eta)$. We assume $\mathrm{E}(\eta_i | o(i), j(i)) = \eta_{o(i)}^{orig} + \eta_{j(i)}^{dest}$ and $\mathrm{E}(h_i | o(i), j(i)) = h_{o(i)}^{orig} + h_{j(i)}^{dest}$.

There is an unmodeled selection process under which each mover i is assigned an origin $o(i) \in \mathscr{J}$ and a destination $d(i) \in \mathscr{J}$. These assignments are potentially correlated with health capital. Such correlation could arise because health capital changes the relative appeal of living in different locations, because determinants of location choices are correlated with determinants of health capital, and/or because origin locations exert a causal effect on health capital as of the time of move.

The key assumption we impose on the selection process is that all such correlation operates only through overall health capital index $\theta_i = h_i + \eta_i$ and not differentially through h_i or η_i on their own. Formally, we assume that once we condition on overall health capital θ_i, origin and destination locations provide no further information about the values of h_i and η_i.

Assumption 5. *(Single index)* $\mathrm{E}(h_i | \theta_i, o(i), j(i)) = \mathrm{E}(h_i | \theta_i)$

Note that since $\eta_i = \theta_i - h_i$, Assumption 5 implies $\mathrm{E}(\eta_i | \theta_i, o(i), j(i)) = \mathrm{E}(\eta_i | \theta_i)$. This single index assumption naturally constrains the selection on h_i to be tightly related to selection on η_i, and we show in the following proposition that it actually implies both of the key assumptions required for our selection correction above.

Proposition 2. *Assumption 5 implies Assumptions 1 and 2.*

Proof. Normality of h_i and η_i as well as the fact that $\theta_i = h_i + \eta_i$ imply

$$\theta_i | h_i \sim N(h_i, \sigma_\eta).$$

Standard conjugate prior results for the normal distribution with known variance imply

$$\begin{aligned}\mathrm{E}\left(h_i|\theta_i\right) &= \frac{\frac{1}{\sigma_h}}{\frac{1}{\sigma_h}+\frac{1}{\sigma_\eta}}\cdot 0+\frac{\frac{1}{\sigma_\eta}}{\frac{1}{\sigma_h}+\frac{1}{\sigma_\eta}}\cdot\theta_i \\ &= \frac{\sigma_h}{\sigma_h+\sigma_\eta}\theta_i.\end{aligned}$$

It then follows that for any $o(i)$ and $j(i)$,

$$\begin{aligned}h_{o(i)}^{orig}+h_{j(i)}^{dest} &= \mathrm{E}\left(h_i|o(i),j(i)\right) \\ &= \mathrm{E}_{\theta_i}\left[\mathrm{E}\left(h_i|\theta_i,o(i),j(i)\right)|o(i),j(i)\right] \\ &= \mathrm{E}_{\theta_i}\left[\mathrm{E}\left(h_i|\theta_i\right)|o(i),j(i)\right] \\ &= \mathrm{E}_{\theta_i}\left[\frac{\sigma_h}{\sigma_\eta+\sigma_h}\theta_i|o(i),j(i)\right] \\ &= \frac{\sigma_h}{\sigma_\eta+\sigma_h}\left(h_{o(i)}^{orig}+h_{j(i)}^{dest}+\eta_{o(i)}^{orig}+\eta_{j(i)}^{dest}\right),\end{aligned}$$

where the third line uses Assumption 5. We therefore have $h_{o(i)}^{orig}+h_{j(i)}^{dest}=\frac{\sigma_h}{\sigma_\eta}\left(\eta_{o(i)}^{orig}+\eta_{j(i)}^{dest}\right)$. The fact that this must hold for all $o(i)$ and $j(i)$ implies

$$\begin{aligned}h_{o(i)}^{orig} &= \frac{\sigma_h}{\sigma_\eta}\eta_{o(i)}^{orig} \\ h_{j(i)}^{dest} &= \frac{\sigma_h}{\sigma_\eta}\eta_{j(i)}^{dest}.\end{aligned}$$

We therefore have $\frac{\mathrm{StDev}\left(h_{o(i)}^{orig}\right)}{\mathrm{StDev}\left(\eta_{o(i)}^{orig}\right)}=\frac{\mathrm{StDev}\left(h_{j(i)}^{dest}\right)}{\mathrm{StDev}\left(\eta_{j(i)}^{dest}\right)}=\frac{\sigma_h}{\sigma_\eta}$, which implies Assumption 2. We also have $\frac{\mathrm{StDev}\left(h_{j(i)}^{dest}\right)}{\mathrm{StDev}\left(\eta_{j(i)}^{dest}\right)}=\frac{\sigma_h}{\sigma_\eta}=\frac{h_{j(i)}^{dest}}{\eta_{j(i)}^{dest}}$, so Assumption 1 follows by Proposition 1. □

A.3 Sample Restrictions, Mover Definition, and Characteristics of Moves

Appendix Table A.9 details the number of observations excluded by each of our sample criteria. Our analysis sample consists of almost 69 million Medicare enrollees whom we observe between the ages of 65 and 99. Of these, almost 62 million are non-movers; their zip code of residence does not change at any point over the years we observe them. The remaining 7 million are "potential movers," in that their zip code of residence changes at least once. To the extent possible, we impose a parallel set of restrictions to the non-mover and mover samples.

Non-mover sample

To define our non-mover sample, we begin with the 62 million enrollees whose CZ of residence does not change over the years we observe them. We make several further restrictions that bring the non-mover sample down to just over 43 million. For each non-mover, we need to be able to define a year t_i^* as a counterfactual move year. Most importantly, this requires that they have a year $t_i^* - 1$ in which the non-mover was enrolled in Traditional Medicare, so that we can measure their healthcare utilization in year $t_i^* - 1$. We also exclude non-movers who do not have a $t_i^* - 1$ in which they are younger than 98 and that is before 2012. These restrictions decrease the number of eligible non-movers from 62 million to 52 million. We exclude non-movers who do not have a year $t_i^* - 1$ such that they survive through the end of year t_i^*, so that we are able to observe their mortality in year $t_i^* + 1$. This eliminates another 9 million non-movers. Finally, we exclude the small number of non-movers who do not have a remaining year $t_i^* - 1$ with data on controls of health utilization and chronic conditions. For the remaining, non-movers, t_i^* is defined as their second year in the sample. In all of our analyses we work with a random 10% sample of these remaining 43 million non-movers.

Mover sample

To define our mover sample, we begin with the 7 million "potential movers" - i.e. individuals whose zip code of residence changes at least once. We make several further restrictions to the mover sample that brings the number of movers down to just over 2 million. First, we exclude individuals whose CZ residence changes more than once; this brings the 7 million potential movers down to 5.6 million. Second, we exclude movers who are enrolled in Medicare Advantage the year before move $(t_i^* - 1)$ or the year after move $(t_i^* + 1)$ since, as discussed, we cannot observe healthcare claims for Medicare Advantage enrollees and we need to observe the location of healthcare claims to define movers. Following the approach of Finkelstein et al. (2016), we exclude "movers" for whom the ratio of the number of claims located in their destination to the number located in either their origin or their destination does not increase by at least 0.75 in their post-move years relative to their pre-move years; these are individuals who, despite having a change of official address on file, do not appear to have really changed CZs based on their claims pattern.[23]

The exclusion of movers who are on MA in $(t_i^* - 1)$ or $(t_i^* + 1)$ brings the number of movers down from 5.6 million to 4.2 million. The exclusion of "false" movers (i.e. those whose claims

[23]The change in claim share is not defined for movers who do not have at least one claim both pre- and post-move. Following Finkelstein et al. (2016), we exclude these cases if: (i) they have no post-move claims and a pre-move destination claim share greater than 0.05; (ii) they have no pre-move claims and a post-move destination claim share less than 0.95.

share does not increase by at least 0.75), further reduces the number of movers to 2.6 million. A few other exclusions for data reasons bring our final mover sample down to 2 million movers. Of the 2 million movers in our final sample, about 18% of them are on MA in at least one year. Given the number of total enrollee-years we observe, we estimate an average annual cross-CZ move rate for Medicare enrollees of about 0.5 percent.

Appendix Figure A.6 shows a mover's claims in her destination CZ, as a share of those in either her origin or her destination, by relative year. There is a sharp change in the year of the move, and only a very small share of claims in the destination pre-move or in the origin post-move. The share of claims in the destination in the year of the move (relative year 0) is close to 0.5, suggesting that moves are made roughly uniformly throughout the year.

Characteristics of moves

We examined some of the characteristics of moves. The average distance between origin and destination zip code centroids of movers in our sample is 547 miles, with a median of 305 miles and a standard deviation of 601 miles. Roughly 66 percent of moves cross state boundaries, and 48 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 10 percent.

In our previous work (Finkelstein et al. 2016), we also used data from the Health and Retirement Survey to explore some of the time-varying correlates of moving in the Medicare population; widowhood and retirement were significant predictors of moving, and the most common self-reported rationale for moving was to be near one's children

A.4 Data and Definitions for Place Characteristics

Here we describe the data and definitions used for the place characteristic measures that we correlate with treatment effects in Figure 5. Summary statistics for all of these measures can be found in Appendix Table A.10.

A.4.1 Healthcare Utilization

We follow Finkelstein et al. (2016) to construct our health care utilization measures. The utilization measure we use as a pre-period control in our estimation is created by aggregating care provided to Medicare beneficiaries as recorded in the inpatient and outpatient claims data. For the healthcare place characteristics in Figure 5, we use a 20% random sample of data from the inpatient, outpatient, and carrier files from Medicare year 2010. See Finkelstein et al. (2016) Online Appendix for more details on how utilization is computed. Our definitions of diagnostic tests and imaging tests also follow directly from Finkelstein et al. (2016) and detailed definitions of these variables can also be found in that paper's Online Appendix.

A.4.2 Other Healthcare Characteristics

Share of hospitals that are non-profit and Hospital beds per capita are defined as in Finkelstein et al. (2016) using the 1998-2008 American Hospital Association's annual survey of hospitals.

Specialists per thousand residents and PCPs per thousand residents are also defined as in Finkelstein et al. (2016) using counts of physicians from the 2011 AMA Physician Masterfile. CZ populations are computed by first aggregating county-level populations from the 2000 Census and 2007-2011 ACS, and then taking the simple average across the two.

Hospital Compare Score, a measure that reports the quality of hospitals, is derived from "process of care" measures that are publicly reported by CMS and uses quarterly data from 2005 to 2011. For a given measure (e.g., share of heart attack patients given aspirin at arrival or share of pneumonia patients given oxygenation assessment), we standardize the score by first taking a simple average across the quarterly measures within a year for a given hospital to get an annual measure. We then construct z-scores for each measure across hospitals in a given year. Lastly, for each hospital we take the simple average of the z-scores across measures within a year and then the simple average over years.

A.4.3 Non-healthcare Characteristics

Measures derived from Centers for Disease Control (CDC) data Many of our CZ-level measures of non-healthcare characteristics are derived from data downloaded from the CDC (`https://wonder.cdc.gov/`) and cover the years 2001-2011 (except pollution, for which records are only available beginning in 2003).

• Homicides and Auto Deaths are defined by the National Center for Health Statistics using ICD-9 and ICD-10 mortality codes and are reported per 100,000 people from 2001-2011 at the county level. We take the simple average across counties to aggregate to the CZ level.

• Pollution is a measure of fine particulate matter and is reported in micrograms per cubic meter. For each county we have the daily average across all days from 2003-2011. We aggregate these single county-level measures to the CZ level by taking the simple average across counties within each CZ.

• Average Winter Temperature is defined as the average daily minimum air temperature during the months of January, February, and March for each county. For each county we take a simple average across all winter months from 2001-2011, and then aggregate these single county-level measures to the CZ level by taking the simple average across counties within each CZ.

• Average Summer Temperature is defined similarly to Average Winter Temperature, but uses the average daily maximum air temperature during the months of June, July, and August.

Measures derived from Chetty et al. (2016b) data Our health behavior measures are derived from the health behavior data posted by Chetty et al. (2016b) (`https://healthinequality.org/data`), as originally drawn from the Behavioral Risk Factor Surveillance Survey (BRFSS). The data cover 1996-2008 and are reported at the CZ level separately for each income quartile. We take the simple average of the four quartiles to get the average measure in each CZ.

• Smoking is the fraction of respondents who report currently smoking in each CZ of the pooled BRFSS sample over years 1996-2008.

• Obesity is the fraction of respondents who are obese (BMI$\geq$30) in each CZ of the pooled BRFSS sample over years 1996-2008.

• Exercise is the fraction of respondents who have exercised in the past 30 days in each CZ of the pooled BRFSS sample over years 1996-2008.

Measures derived from Census data Our other CZ level measures of non-healthcare characteristics are derived from Census data.

• Share Urban is derived from the 2000 and 2010 Census data. Urban and total populations are available at the county level, and we aggregate these values within each CZ and compute the share of that population that is urban. We then take the simple average of these values across the two census surveys.

• Share over 60, Median household income, and High school graduation rate are computed similarly using the 2000 Census survey and 2007-2011 American Community Survey. Median household income and high school graduation rates are computed for people 25 and older.

A.5 Bounding $\text{StDev}\left(\gamma_j\right)$ without Equal Selection

As formulated, Assumption 1 requires that equal selection holds for each individual location (all $j \in J$). This assumption is necessary to compute η_j, and therefore to compute each $\gamma_j = \tau_j^{dest} - \eta_j^{dest}$. However, even in the absence of this equal selection assumption, we are able to recover a subset of our main results. Importantly, we can estimate bounds on the standard deviation of γ_j. All that is required is to restate the Relative Importance Assumption with j rather than $j(i)$ indexing in order to estimate the standard deviation, though not the actual quantities, of η_j^{dest}. That is:

Modified Assumption 2: $\frac{\text{StDev}\left(\eta_j^{orig}\right)}{\text{StDev}\left(h_j^{orig}\right)} = \frac{\text{StDev}\left(\eta_j^{dest}\right)}{\text{StDev}\left(h_j^{dest}\right)}$

Recall that $\text{Var}\left(\gamma_j\right) = \text{Var}\left(\tau_j^{dest}\right) - \text{Var}\left(\eta_j^{dest}\right) - 2 \cdot \text{Cov}\left(\gamma_j, \eta_j^{dest}\right)$. Because correlations are bounded on [-1,1], we know

$$\text{Cov}\left(\gamma_j, \eta_j^{dest}\right) \leq \sqrt{\text{Var}\left(\gamma_j\right) \cdot \text{Var}\left(\eta_j^{dest}\right)} \tag{18}$$

and

$$\text{Cov}\left(\gamma_j, \eta_j^{dest}\right) \geq -\sqrt{\text{Var}\left(\gamma_j\right) \cdot \text{Var}\left(\eta_j^{dest}\right)} \tag{19}$$

Therefore,

$$\text{Var}\left(\gamma_j\right) \geq \text{Var}\left(\tau_j^{dest}\right) - \text{Var}\left(\eta_j^{dest}\right) - 2 \cdot \sqrt{\text{Var}\left(\gamma_j\right) \cdot \text{Var}\left(\eta_j^{dest}\right)} \tag{20}$$

With some algebraic manipulations we see:

$$\begin{aligned}
\text{Var}\left(\gamma_j\right) + \text{Var}\left(\eta_j^{dest}\right) + 2 \cdot \sqrt{\text{Var}\left(\gamma_j\right) \cdot \text{Var}\left(\eta_j^{dest}\right)} &\geq \text{Var}\left(\tau_j^{dest}\right) \\
\left(\sqrt{\text{Var}\left(\gamma_j\right)} + \sqrt{\text{Var}\left(\eta_j^{dest}\right)}\right)^2 &\geq \text{Var}\left(\tau_j^{dest}\right) \\
\sqrt{\text{Var}\left(\gamma_j\right)} + \sqrt{\text{Var}\left(\eta_j^{dest}\right)} &\geq \sqrt{\text{Var}\left(\tau_j^{dest}\right)} \\
\sqrt{\text{Var}\left(\gamma_j\right)} &\geq \sqrt{\text{Var}\left(\tau_j^{dest}\right)} - \sqrt{\text{Var}\left(\eta_j^{dest}\right)} \\
\text{StDev}\left(\gamma_j\right) &\geq \text{StDev}\left(\tau_j^{dest}\right) - \text{StDev}\left(\eta_j^{dest}\right)
\end{aligned}$$

The upper bound follows from symmetry:

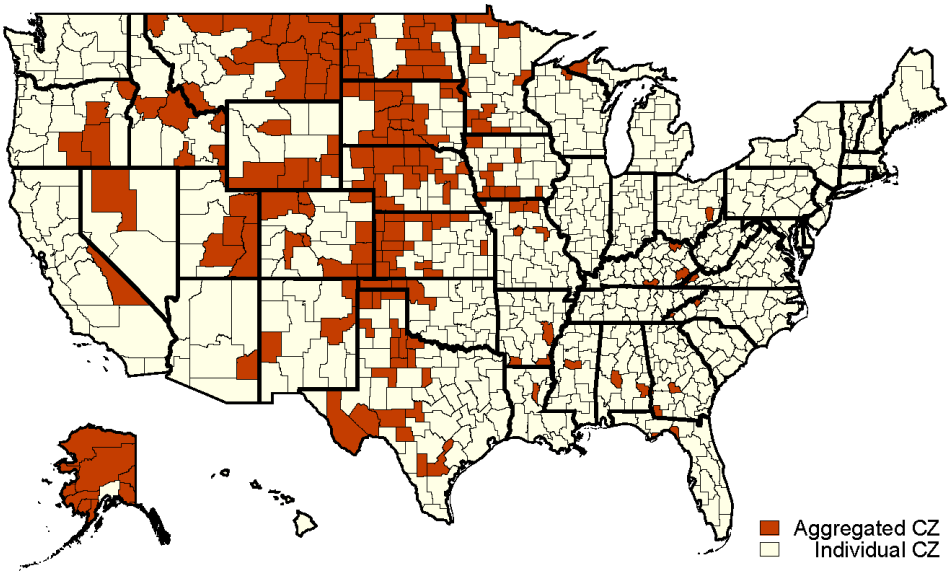
$$\text{StDev}\left(\gamma_j\right) \leq \text{StDev}\left(\tau_j^{dest}\right) + \text{StDev}\left(\eta_j^{dest}\right)$$

In our data, we have $\text{StDev}\left(\tau_j^{dest}\right) = 0.077$ and, from Modified Assumption 2, $\text{StDev}\left(\eta_j^{dest}\right) =$

0.036. Therefore, we estimate the lower and upper bounds on the standard deviation of γ_j without the equal selection assumption as 0.041 and 0.113.

Appendix Figures

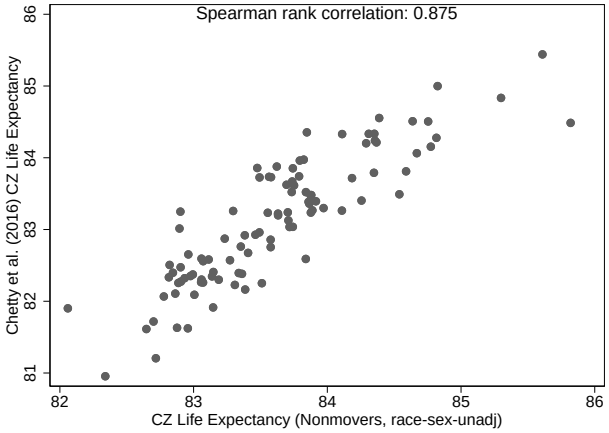
Figure A.1: Location of Small CZs



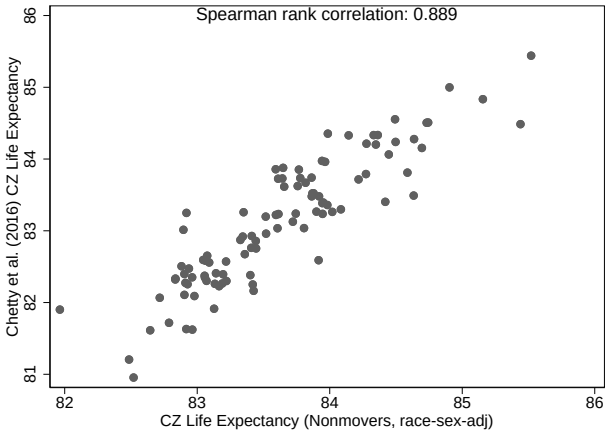
Notes: Figure shows the location of small CZs. Small CZs within the same state are combined and considered a single location, resulting in 35 aggregated CZs.

Figure A.2: Life Expectancy Correlations with Chetty et al. (2016b)

(a): Unadjusted Life Expectancy

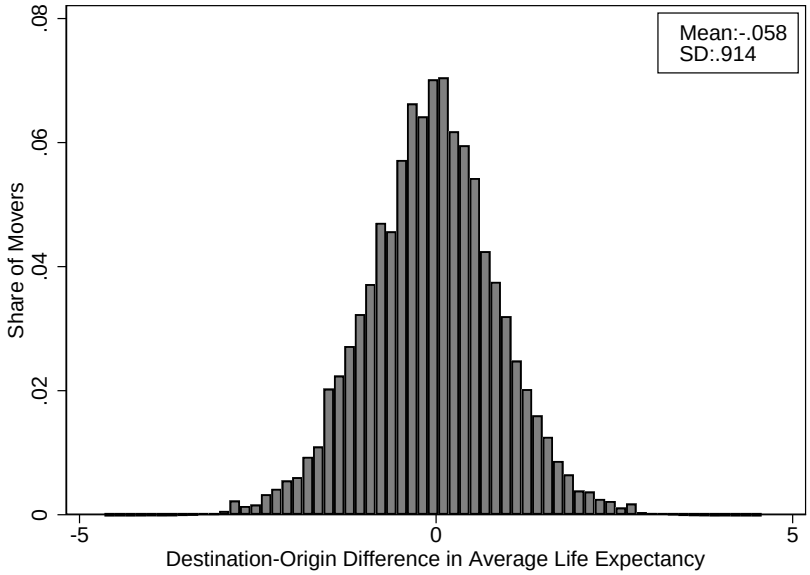


(b): Race-Sex Adjusted Life Expectancy (L_j)



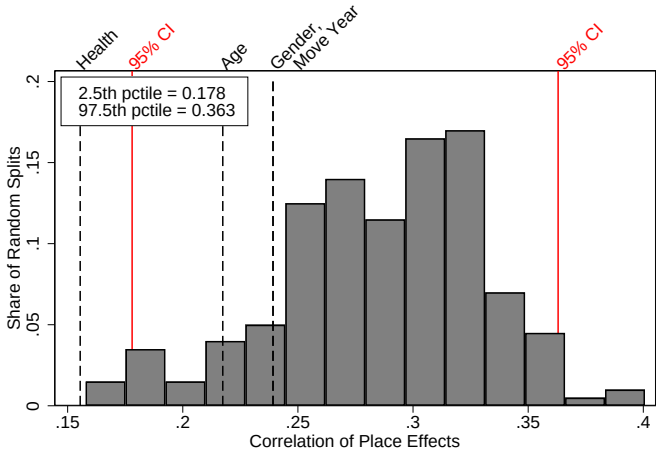
Notes: These figures compare our non-mover life expectancy at age 65 (L_j) to life expectancy estimates at age 40 from Chetty et al. (2016b), among the 100 largest CZs by population in 2000. Panel (a) uses a version of L_j that is not adjusted for race and sex; specifically in panel (a), we set the elements of the vector associated with race and sex to the CZ average rather than the national average for each CZ. Panel (b) uses our race- and sex-adjusted L_j from Figure 1. These figures use the life expectancies from Chetty et al. (2016b) that are not adjusted based on race or sex. Since their life expectancies are based on CZs as of 1990, we convert their estimates to CZs as of 2000 by taking an average of the 1990 CZ life expectancies, weighted by the proportion of the population in each CZ in 2000 who lived in the CZ in 1990. Correlation coefficients are based on the Spearman rank correlation, although results are similar when comparing life expectancies using the Pearson correlation coefficient.

Figure A.3: Distribution of Destination-Origin Difference in Average Life Expectancy



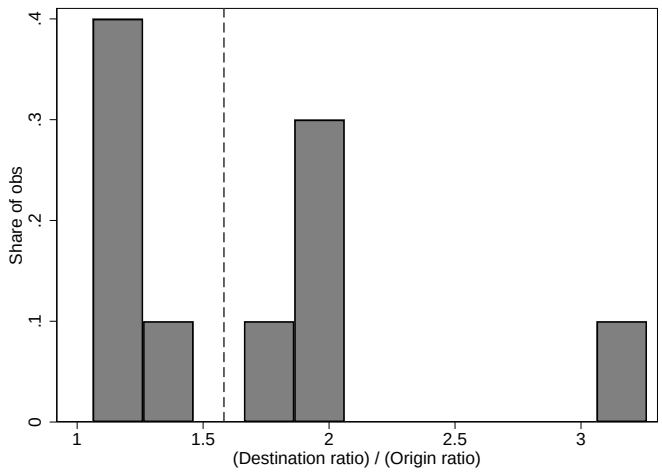
Notes: Figure shows the distribution across movers of the difference in average non-mover life expectancy at age 65 (L_j) between their origin and destination CZs. The sample is all movers (N = 2,033,263 movers).

Figure A.4: Heterogeneity in Place Effects



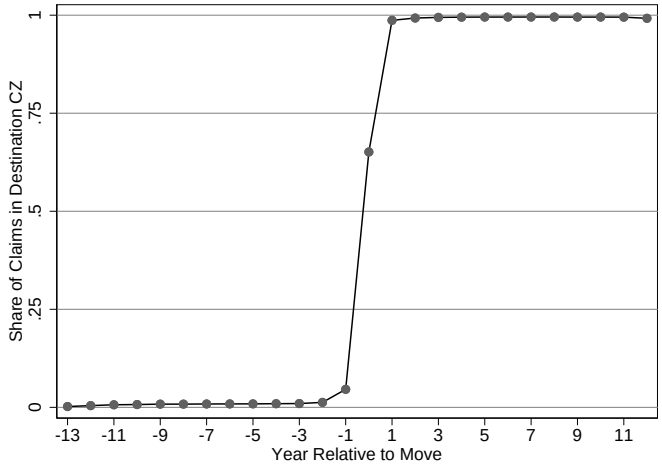
Notes: This histogram shows the distribution of correlation coefficients between place effects (γ_j) resulting from 200 random partitions that split the data into two equally sized groups, with separate estimation of the Gompertz model for each group. The place effects are corrected using the selection correction procedure. Red lines indicate the locations of the 2.5th and 97.5th percentiles; for values outside of this range, we reject the null hypothesis that the place effects are equal in the two groups.

Figure A.5: Histogram of (Destination ratio)/(Origin ratio)



Notes: This histogram plots the distribution of the ratio $\left(\frac{\text{StDev}\left(\eta_{j(i)}^{dest}\right)}{\text{StDev}\left(h_{j(i)}^{dest}\right)}\right) / \left(\frac{\text{StDev}\left(\eta_{j(i)}^{orig}\right)}{\text{StDev}\left(h_{j(i)}^{orig}\right)}\right)$ for 10 different subsets H_i^k of chronic conditions, using the same subsets as in Figure 6. For each k, H_i^k includes log(overall utilization) and a random subset of thirteen of the twenty-seven chronic conditions. The dotted line shows the median of the distribution. All standard deviations are computed using the split-sample approach.

Figure A.6: Claims Share Graph



Notes: This figure shows the share of a mover's claims located in their destination CZ, among those in either their origin or their destination CZ. The sample is all enrollee-years ($N = 17,443,789$) in the 100% Denominator file for all movers in our baseline sample.

Appendix Tables

Table A.1: Predicting Mortality from Observables

	(1) Coefficient	(2) (s.e.)	(3) Average
Log(Utilization + 1)	0.028	(0.000)	3.67
Chronic Conditions:			
Acquired Hypothyroidism	-0.008	(0.003)	0.03
Acute Myocardial Infarction	-0.069	(0.009)	0.00
Alzheimer's	0.214	(0.009)	0.01
Alzheimer's and Related Disorders or Senile Dementia	0.474	(0.006)	0.03
Anemia	0.152	(0.002)	0.09
Asthma	-0.054	(0.005)	0.02
Atrial Fibrillation	0.222	(0.003)	0.03
Benign Prostatic Hyperplasia	-0.214	(0.004)	0.03
Breast Cancer	0.087	(0.005)	0.01
Cataract	-0.095	(0.002)	0.15
Chronic Kidney Disease	0.413	(0.006)	0.02
Chronic Obstructive Pulmonary Disease	0.484	(0.003)	0.05
Colorectal Cancer	0.048	(0.008)	0.01
Depression	0.188	(0.004)	0.04
Diabetes	0.342	(0.002)	0.08
Endometrial Cancer	0.115	(0.014)	0.00
Glaucoma	-0.063	(0.003)	0.05
Heart Failure	0.327	(0.003)	0.06
Hyperlipidemia	-0.221	(0.002)	0.15
Hypertension	0.042	(0.002)	0.25
Hip/Pelvic Fracture	0.042	(0.008)	0.00
Ischemic Heart Disease	0.099	(0.003)	0.13
Lung Cancer	0.772	(0.015)	0.00
Osteoporosis	0.032	(0.004)	0.03
Prostate Cancer	0.035	(0.003)	0.02
Rheumatoid Arthritis	-0.066	(0.003)	0.08
Stroke / Transient Ischemic Attack	0.205	(0.005)	0.02
N	6,345,989		

Notes: This table reports the coefficients of the components of H_i in our main estimating equation, equation (3). Standard errors are computed with 20 replications of the bootstrap. Column (3) reports the sample mean of log(Utilization + 1) in row (1) and, for all other rows, the share of beneficiaries with the indicated chronic condition in year $t_i^* - 1$. Utilization excludes physician services ("carrier files") because these files are only available for a 20 percent subsample. As in the estimation, when computing the sample-wide shares, non-movers are upweighted by ten to account for our sampling procedure.

Table A.2: Number of Movers Received by CZ or Aggregate CZ

Statistics	# Movers to CZ
Minimum	48
10th Percentile	468
25th Percentile	781
Median	1,522
75th Percentile	3,534
90th Percentile	9,241
Maximum	45,360

Notes: This table summarizes the number of movers received by each of the 563 CZs or aggregated CZs.

Table A.3: Summary Statistics on Estimation Sample

	(1) Movers	(2) Non-movers
Age:		
65-74	0.48	0.75
75-84	0.35	0.19
85+	0.18	0.06
Female	0.60	0.55
White	0.90	0.85
Region:		
Northeast	0.19	0.20
South	0.41	0.38
Midwest	0.21	0.25
West	0.19	0.17
On Medicaid	0.10	0.11
Avg. # of chronic conditions	3.05	1.33
One year mortality	0.09	0.04
Four year mortality	0.27	0.15
Life expectancy at age 65	82.10	83.65
Number of individuals	2,033,263	4,312,726

Notes: These summary statistics are computed on all movers and nonmovers in our Gompertz estimation sample. The reference year for movers is their move year, and the reference year for non-movers is set to be their second year in the sample. Rows for female, white, age, and region report the shares of individuals with the given characteristics. The life expectancy measure is conditional on surviving until age 65, and is calculated for 1,000 random 65-year-old enrollees within the sample indicated by each column. Time-varying characteristics are measured in the year prior to each enrollee's reference year.

Table A.4: Transition Matrix of Moves

Origin Decile	Destination Decile										
	1	2	3	4	5	6	7	8	9	10	Origin total
1	10%	13%	16%	17%	13%	12%	6%	5%	4%	4%	74,983
2	7%	10%	11%	14%	14%	13%	8%	8%	7%	9%	110,370
3	7%	9%	10%	16%	14%	14%	9%	8%	6%	6%	107,716
4	5%	7%	9%	12%	15%	13%	10%	9%	9%	10%	159,424
5	3%	6%	6%	12%	17%	14%	11%	11%	9%	10%	219,967
6	3%	6%	6%	9%	13%	14%	12%	14%	11%	13%	238,606
7	2%	4%	5%	8%	12%	15%	9%	16%	14%	15%	184,239
8	1%	4%	4%	7%	11%	12%	12%	14%	18%	17%	220,596
9	1%	3%	3%	6%	9%	11%	12%	21%	14%	19%	305,532
10	1%	4%	2%	6%	9%	10%	11%	15%	16%	27%	411,830

Notes: Table reports the percentage of moves in each row to each destination. The "origin total" column reports the total number of moves in each row. Each row is a (population-weighted) decile of CZ origin life expectancy. Each column is decile of CZ destination life expectancy. Q1 is the lowest life expectancy and Q10 is the highest. The sample is all movers (N = 2,033,263 movers).

Table A.5: Heterogeneity in Place Effects

Sample		
Baseline Standard Deviation of γ_j	0.054	[0.038, 0.067]
Move Year		
Standard Deviation of γ_j		
(i) Late	0.064	[0.000, 0.096]
(ii) Early	0.056	[0.034, 0.071]
Correlation of γ_j from (i) and (ii)	0.239	
Gender		
Standard Deviation of γ_j		
(i) Female	0.056	[0.031, 0.073]
(ii) Male	0.068	[0.034, 0.100]
Correlation of γ_j from (i) and (ii)	0.239	
Age		
Standard Deviation of γ_j		
(i) Young Movers	0.075	[0.050, 0.099]
(ii) Old Movers	0.038	[0.000, 0.067]
Correlation of γ_j from (i) and (ii)	0.217	
Individual Health		
Standard Deviation of γ_j		
(i) Good Health	0.101	[0.074, 0.117]
(ii) Poor Health	0.058	[0.024, 0.081]
Correlation of γ_j from (i) and (ii)	0.156	

Notes: The first row replicates baseline results (See Table 3) and the rest of the table summarizes splits of the main sample that approximately divide the number of movers into two equal groups. Each group includes all non-movers and the Gompertz estimation for each group controls for the same covariates as in the main estimation. "Late movers" includes all movers with a move year of 2005 or later (N = 909,901) and "early movers" includes all movers with a move year before 2005 (N = 1,123,362). There are 1,229,235 female movers and 804,028 male movers. Young movers move when they are 75 or younger (N = 1,038,585) and old movers move when they are older than 75 (N = 994,678). Movers with good health have a value of $\hat{h}_i$ less than or equal to the median value among all movers (N = 1,016,631) and movers with poor health have a value of $\hat{h}_i$ greater than the median value among all movers (N = 1,016,632). Standard deviations are calculated using the split-sample approach. 95% confidence intervals are computed using 100 replications of the Bayesian bootstrap. Since standard deviations cannot be negative, any split-sample approach that produces a negative result we set to 0.000.

Table A.6: Alternative Selection Assumptions

(1)	(2)	(3)
C_1C_2	StDev of Place Effects (γ_j)	StDev of Treatment Effects ($L_j^* - \bar{L}$)
1.0	0.054	0.44
0.5	0.063	0.51
1.26	0.054	0.43
1.58	0.056	0.45
2.0	0.065	0.53
3.0	0.099	0.80

Notes: This table reports the cross-CZ standard deviations of our place effects and treatment effects for various values of C_1C_2 as defined in equation (7).

Table A.7: Panel vs. Cross-Section

	(1) Any hospital admission	(2) Any emergency room visit	(3) Any outpatient visit
(1) Mean of outcome	0.196	0.263	0.617
Cross-section standard deviations:			
(2) Outcome	0.027	0.031	0.098
(3) Place effect, unadjusted (τ_j^{dest})	0.027	0.028	0.084
(4) Place effect, adjusted (γ_j)	0.023	0.025	0.086
(5) Panel standard deviation: place effect (γ_j)	0.020	0.025	0.101

Notes: Each column reports results for a different outcome. Row (1) reports the mean of the dependent variable**.** Row (2) reports the cross-CZ standard deviation of the outcome. The estimates are reported in rows (3) through (5). All estimates are linear probability models; for the cross-sectional estimates (rows 3 and 4) we estimate equation (3) on the outcome one-year post move. For the panel (row 5) we estimate the panel equation (8). The sample is different from our baseline sample ($N = 5,258,502$ enrollees instead of $6,345,989$) because, to be consistent with our panel analysis, we exclude enrollees who do not have 12 months of Parts A and B coverage in relative year 1. For the panel estimation, we further exclude all enrollee-year observations for which an enrollee does not have 12 months of Parts A and B coverage as well as relative year 0 for movers. We also restrict our analysis to years 1999-2011. These exclusions together ensure that each outcome variable in this analysis always includes twelve months of Parts A and B coverage. In column (1), any hospital admission is defined as non-zero inpatient utilization. In column (2), any emergency room visit is defined as non-zero emergency room utilization. In column (3), any outpatient visit is defined as non-zero outpatient utilization. Each of these utilization measures is defined as in Finkelstein et al. (2016), except emergency room utilization, which is defined using only the inpatient and outpatient files (rather than the measure that includes the carrier files in Finkelstein et al. 2016).The mean of the outcome is the average over all enrollee-years in the sample. We compute the CZ level measure by first taking the average over non-movers within each CZ, then the simple average over years.

Table A.8: Robustness Checks

Specification	(1) Movers	(2) StDev of life expectancy (L_j)	(3) StDev of treatment effects ($L_j^* - \bar{L}$)	(4) $Corr(L_j^* - \bar{L}$, baseline)	(5) $Corr(L_j^* - \bar{L}, L_j)$
(1) Baseline	710,990	0.67	0.47	1.00	0.41
			[0.40, 0.53]		[0.34, 0.48]
(2) Heterogeneity by mover status	710,990	0.66	0.47	0.98	0.43
			[0.41, 0.52]	[0.97, 0.98]	[0.37, 0.50]
(3) Interacting H components with age	710,990	0.64	0.42	0.99	0.38
			[0.35, 0.47]	[0.97, 0.99]	[0.31, 0.45]
(4) Controlling for origin county mortality rates	710,990	0.68	0.49	0.98	0.45
			[0.43, 0.55]	[0.97, 0.99]	[0.38, 0.51]
(5) Move distance greater than 100 miles	558,367	0.67	0.45	0.96	0.41
			[0.36, 0.53]	[0.79, 0.97]	[0.33, 0.48]
Destination-origin difference in L_j:					
(6) Greater than median difference	341,469	0.67	0.53	0.91	0.48
			[0.39, 0.65]	[0.83, 0.90]	[0.37, 0.58]
(7) Less than median difference	369,521	0.67	0.49	0.84	0.42
			[0.00, 0.89]	[0.59, 0.83]	[0.11, 0.74]
Excluding moves to:					
(8) Adjacent CZs	554,420	0.67	0.48	0.97	0.41
			[0.40, 0.56]	[0.93, 0.97]	[0.34, 0.49]
(9) Florida, Arizona, and California	485,389	0.62	0.47	1.00	0.37
			[0.40, 0.54]	[0.85, 1.00]	[0.34, 0.48]
(10) Years 1999-2003	325,041	0.62	0.46	0.92	0.44
			[0.36, 0.55]	[0.83, 0.92]	[0.32, 0.54]
(11) Years 2004-2012	385,949	0.71	0.48	0.86	0.34
			[0.36, 0.61]	[0.57, 0.76]	[0.20, 0.44]

Notes: Table reports results for alternative specifications. Estimates in all rows are computed on the 100 largest CZs by total population in 2000. The first row reports the baseline estimates, and each additional row represents a single deviation from the baseline, which are described in Section 7. All treatment effects are treatment effects on life expectancy, and are not adjusted using an empirical Bayes correction. Column (1) shows the number of movers who remain in each specification. In all rows, we estimate treatment effects for 100 CZs, other than row (9) (79 CZs). Column (2) shows the raw cross-CZ standard deviation of age-65 non-mover life expectancy. In column (3), cross-CZ standard deviation of treatment effects are computed using the split-sample. Column (4) shows the cross-CZ correlation between treatment effects and the baseline treatment effects. The cross-CZ correlation of treatment effects and life expectancy in column (5) is computed as the coefficient of the regression of the non-adjusted treatment effect on age-65 non-mover life expectancy. 95% confidence intervals are computed using 100 iterations of the Bayesian bootstrap.

Table A.9: Sample Restrictions

	(1) Enrollees	(2) Enrollee-years
Original sample	80,708,181	665,131,064
Excluding enrollee-years with age < 65 or age > 99	69,330,956	560,057,853
Excluding enrollee-years with incomplete data[1]	68,935,110	556,340,988
Number of non-movers after sample-wide drops	61,899,201	
Excluding non-movers without a valid relative year -1 with Traditional Medicare	52,448,582	
Excluding non-movers without a relative year -1 with 1-year mortality observed	43,147,931	
Excluding non-movers without a relative year -1 with pre-period controls[2]	43,145,670	
Number of movers after sample-wide drops	7,035,909	
Excluding movers with more than one move	5,609,064	
Excluding movers on MA during relative years -1 or 1	4,204,679	
Excluding "false" movers[2]	2,564,376	
Excluding movers for whom we cannot observe one-year mortality	2,033,333	
Excluding movers with missing pre-period controls	2,033,263	

Notes: (1) Data is incomplete if the CZ is missing for an enrollee-year, or an enrollee has gaps in the years they are observed. (2) Pre-period controls consist of health utilization and chronic conditions. (3) False movers are those movers for whom the ratio of the number of claims located in their destination to the number located in either their origin or their destination does not increase by at least 0.75 in their post-move years relative to their pre-move years.

Table A.10: CZ Summary Statistics

	(1) Mean	(2) S.D.	(3) N
Healthcare Characteristics:			
Hospital Compare Score	-0.13	0.50	559
Specialists per capita	1.47	1.08	563
PCPs per capita	0.90	0.30	563
Hospital beds per capita	2.70	1.17	560
Share non-profit hospitals	0.82	0.20	560
Imaging tests	1.84	0.92	563
Diagnostic tests	4.12	2.82	563
Mean utilization	4150.60	1396.57	563
Non-healthcare Characteristics:			
Pollution (μg per cubic meter)	11.84	1.72	559
Average Summer Temperature ($^{\circ}F$)	84.52	6.83	559
Average Winter Temperature ($^{\circ}F$)	30.57	10.05	559
Auto Deaths (per 100,000)	22.30	7.14	563
Homicides (per 100,000)	5.31	3.00	490
Smoking	0.20	0.04	557
Obesity	0.28	0.06	557
Exercise	0.74	0.05	557
Highschool Graduation Rate	0.82	0.06	563
Household Income	39,277	7,782	563
Share 60+	0.20	0.04	563
Share Urban	0.56	0.22	563

Notes: This table reports the simple average across the (aggregated) CZs of the place characteristics in Figure 5. See Appendix A.4 for detailed definitions of these place characteristics.