Appendix A: Validation of MDAC data compared to NVSS

To calculate raw mortality rates using the NVSS data, we first need to obtain an estimate of the population at risk in each demographic group from a different dataset.[1] We proceed by using the IPUMS ACS 2008 to estimate the total U.S. born population at risk in the different age and gender cells. The 2008 ACS is collected throughout the year, which complicates the comparison of death statistics between the MDAC dataset and the NVSS. Based on the sampling design of the 2008 ACS, the matched deaths in 2008 in the MDAC are skewed towards the end of the year. To roughly match the follow-up period of the ACS 2008, we only include the deaths of US born individuals reported in the NVSS for the period July 2008 – December 2015.

We note that individuals in the MDAC are linked to the National Death Index using Social Security Numbers (SSNs). In case information about SSNs is missing or invalid, a combination of other personally identifiable information (first name, last name, date of birth) is used to match individuals to the National Death Index. Overall, less than 0.8% of all the US born respondents ages 50 and older by the time of the 2008 ACS interview did not provide valid SSNs, or complete information on their first name, last name, and date of birth. Furthermore, the proportion of individuals with missing identifying information does not vary substantially across age groups.

We find that the average mortality statistics across the two different datasets are similar. However, there are small differences in the mortality rates across datasets for the youngest age groups in the sample. Overall, the raw mortality rates for the age groups 50-54, 55-59, and 60-64 are smaller in the MDAC compared to the estimates that use the NVSS data. The respondents in the ACS that died in the follow-up period appear to be slightly underrepresented compared to the respondents that were alive by the end of the follow-up period.[2]

---

[1] The calculation of raw death rates with the MDAC sample does not require the use of a secondary dataset. Both the numerator (number of deaths) and denominator (population at risk) are obtained from the same data source, which gets rid of the numerator-denominator bias. For our analysis of life expectancy by state of birth, this is an important advantage of the MDAC dataset with respect to other data sources. Precise intercensal population estimates by state of birth are not readily available. Additionally, in the NVSS data state of birth is missing for close to 2 percent of the sample. Both factors might play a role in increasing the numerator-denominator bias in mortality rates by state of birth.

[2] Another subtle difference between the sampling framework of the MDAC and the NVSS is that the MDAC has a closed design. Only individuals that were residing in the U.S. by the time of the ACS interview are considered in the raw mortality rate calculations. In contrast, the NVSS death records

Appendix Figure 1 shows the scatterplot of the cumulative probability of dying at any point during the follow-up period by five-year age-group and state (either state of residence or state of birth) comparing the observed rates from the MDAC with the equivalent rates from the NVSS data. Even though the raw mortality rates from the MDAC are smaller than the equivalent mortality rates obtained from the NVSS system, the correlation between the cumulative mortality probabilities is equal to 0.98 for both mortality rates by state of residence and mortality rates by state of birth.[3] This provides evidence that the geographical patterns of mortality from the MDAC sample are quite close to the patterns of the U.S. population.[4]

consider individuals that were residing outside of the U.S. in 2008 and that later returned to the country and died at any time in the follow-up period. However, these individuals are not considered in the denominator of population at risk.

[3] However, there are important differences in the survival probabilities by state of birth in the states of Alaska and Nevada. In the ACS 2008, only few individuals report Alaska or Nevada as their state of birth. In some cases, the estimated number of deaths in the follow-up period using the NVSS data is even higher than the estimate of the population at risk using IPUMS 2008 ACS.

[4] The correlation between the mortality rates by state of residence in the MDAC and the NVSS is quite high, considering that the conceptual definition of state of residence is not the same across the two datasets. In the MDAC, the state of residence of an individual corresponds to the state where the individual was interviewed in 2008. In contrast, in the NVSS the state of residence of an individual corresponds to the state of residence of the individual at the time of death.

Appendix B: Construction of life expectancies using the MDAC data

Period life expectancies are typically constructed with the observed one-year mortality rates at different ages. For the purposes of this paper, $m_M^{s,g}(a,t)$ denotes the one-year mortality rate in year t of individuals of gender g and age a, associated to state s using the method of aggregation M, where M corresponds to aggregation by place of residence or by place of birth. The vector of one-year mortality rates $\{m_M^{s,g}(a,t)\}_{a=50}^{\bar{A}}$ is the required input for the construction of life expectancies at age 50 for a given sub-population described by $\{s,g,M\}$ in a given period t, where $\bar{A}$ denotes the maximum years of life a person can live. We fix the subsequent analysis on a given state, gender, and method of aggregation and drop these three indexes to simplify notation.

We do not observe $m(a,t)$ directly in the disclosed data from the MDAC dataset. Instead, we have disclosed the cumulative probability of surviving the 7+ year follow-up period, which starts by the time of the ACS interview in 2008 and ends by December 31, 2015. A second data constraint with the disclosed data from the MDAC is that the cumulative survival probability is aggregated into five-year age groups, based on the age of the individual at the time of the 2008 ACS interview. We will refer to the cumulative survival probability of the individuals in the five-year age group X in 2008 as S_X, where $X \in \mathbb{X} =$ {50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+}.

To estimate life expectancies for this sub-population, we proceed in two steps. In the first stage, we obtain estimates of the one-year mortality rates $\widehat{m(a,t)}$, for $50 \leq a \leq \bar{A}$, from the observed cumulative survival probabilities $S_X, \text{with } X \in \mathbb{X}$. Once that we have estimates of the one-year mortality rates, in the second stage we obtain life expectancy measures using conventional demographic methods (Preston et al., 2001).

Assuming for now that all individuals were surveyed at the same point in the beginning of 2008, the proportion of individuals in age group X that were surveyed in the ACS 2008 and that did not die in the follow-up period, denoted by S_X, can be described in terms of one-year mortality rates $m(a,t)$ as follows:

$$S_X = \sum_{a \in X} \text{Share}_a * \text{S(a)} = \sum_{a \in X} \text{Share}_a * \prod_{y=0}^{7} (1 - m(a+y, 2008+y))$$

Where in the second equality S(a) corresponds to the cumulative probability of surviving after 2015 if you are exactly a years old at the time of the interview and Share_a corresponds to the proportion of people in age group X that is exactly a years old by the 2008 interview. The first equality shows that the cumulative survival probability by five-year age-group that we observe in the data is a weighted average of the cumulative survival probabilities of exact one-year age groups. The second equality comes directly from the relationship between cumulative survival probability in the follow-up period and one-year mortality rates for different years.

We assume that the time variation in mortality rates over the years across this short follow-up period time is substantially less important than the variation in mortality rates across the age dimension. To simplify the estimation, we assume that for a given cell (age x state x gender x method of aggregation), the mortality rates are smooth over time and do not change drastically over the follow-up period. We omit the index on time from $m(a,t)$ in all of what follows and refer to $m(a)$ as the average death rate at age a during the follow-up period.

In order to parametrize the previous equation, we make use of the well-known property of mortality rates discovered by Benjamin Gompertz in 1825 that the growth in the mortality rate can be fitted quite precisely by an exponential function after a certain critical age in adolescence. Our youngest age group corresponds to the 50-54-year-old, so we do not need to consider the age period in childhood and adolescence in which mortality risk temporarily declines with respect to age.[5] By the Gompertz law of mortality, $m(a)$ can be closely approximated with only two parameters as follows:

$$m(a) \approx \exp(\beta_0 + \beta_1 a)$$

Substituting this expression into the original formula of S_X, we approximate S_X a as a nonlinear function of two parameters (β_0, β_1) as follows:

[5] In unreported robustness checks, we explore two alternative mortality models to describe mortality rates as a function of age. We instead assume that mortality rates follow the Kannisto or Logistic models.

$$S_X(\beta_0, \beta_1) \approx \sum_{a \in X} \text{Share}_a * \prod_{y=0}^{7} (1 - \exp(\beta_0 + \beta_1 a))$$

For each cell in our data, we have eight different data points on S_X (which correspond to the eight different five-year age groups) and Share_a is directly observed using the IPUMS version of the 2008 ACS. Finally, we assume that we observe the cumulative survival probabilities with classical measurement error:

$$S_X(\beta_0, \beta_1, \varepsilon) = \sum_{a \in X} \text{Share}_a * \prod_{y=0}^{7} (1 - \exp(\beta_0 + \beta_1 a)) + \varepsilon$$

We run weighted nonlinear least squares to estimate $\hat{\beta}_0$ and $\hat{\beta}_1$.using this equation. We use analytic weights given by the number of individuals in each age group in the 2008 ACS to weigh each of the eight data points. As a final adjustment, we multiply the probability of dying in 2008 by one-half in order to correct the fact that the ACS 2008 was collected continuously throughout 2008, and that the follow-up period varies between 7 and 8 years across individuals.

For the previous explanation, we have fixed the explanation on a given sub-population of individuals of gender g, that are associated to state s, using the aggregation method M. In practice, we estimate $\hat{\beta}_{0,M}^{s,g}$ and $\hat{\beta}_{1,M}^{s,g}$ with weighted NLLS for each $\{s, g, M\}$ combination separately. We proceed by substituting $\hat{\beta}_0^{s,g,M}$ and $\hat{\beta}_1^{s,g,M}$ into the expression of the Gompertz mortality function to obtain estimates for $\left\{\widehat{m_M^{s,g}(a)}\right\}_{a=50}^{\bar{A}}$. We further assume that $\bar{A} = 110$, and that $\widehat{m_M^{s,g}(110)} = 1$.

Gompertz approximations of mortality rates tend to overestimate mortality rates at older ages. We thus follow the methodology in Chetty et al. (2016) and assume that the sex-specific mortality rate after age 90 is constant across different states. We use official national life tables from the National Center for Health Statistics (NCHS) to impute mortality rates up to age 99. For ages 100-109, we use life tables from the Social Security Administration (SSA).

In a second stage, with the estimates of mortality rates at each age, we apply standard life table methods to obtain estimates of life expectancies from one-year mortality rates for each sub-population separately (Preston et al., 2001).

Standard Errors

In our preferred specification, we adapt a parametric Bootstrap specification following the methodology in Chetty et al (2016) to compute standard errors of the life expectancies (either by state of birth or state of residence). We assume that the true vector of parameters $(\beta_{0,M}^{s,g}, \beta_{1,M}^{s,g})$ is distributed as a bivariate normal distribution with mean $(\hat{\beta}_{0,M}^{s,g}, \hat{\beta}_{1,M}^{s,g})$ and variance equal to the estimated variance-covariance matrix of $(\hat{\beta}_{0,M}^{s,g}, \hat{\beta}_{1,M}^{s,g})$ in the weighted NLLS estimation. For each sub-population (s, g, M), we draw 100 draws from the normal distribution. For each draw, we estimate the vector of mortality rates by age and then calculate an estimate of life expectancy. Finally, we construct a bootstrapped 95 percent confidence interval for the life expectancy in sub-population using the 25th and 975th ordered values of the simulated life expectancies as lower and upper bounds of the confidence interval, respectively.

As a second alternative to compute standard errors, we use the classical formula in Chiang (1984) which requires the complete distribution of deaths by one-year age group and year. We thus first use the estimates of $\left\{\widehat{m_M^{s,g}(a)}\right\}_{a=50}^{\bar{A}=110}$ to further distribute the total number of deaths observed for each five-year age group in the follow-up period into deaths by one-year age group and year.

We present all the results using Bootstrapped standard errors in the main text, which are slightly bigger that the standard errors using the Chiang method. Nonetheless, the two different sets of standard errors are highly correlated, with a correlation above 0.7 in the standard errors for both types of life expectancy measures (state of residence or state of birth), for both genders and at different ages (at age 50 and at age 65).

The main hypothesis that we test in the paper is whether the life expectancy by state of residence in each sub-population is significantly different than the life expectancy measured by state of birth. In order to test whether this difference is

statistically significant, we would ideally need to obtain an estimate of the standard error of the difference in life expectancies across the two measures. As such, the original formula to obtain the standard of the difference in life expectancies in a fixed state is the following:

$$se\left(\widehat{LE}_{SoR} - \widehat{LE}_{SoB}\right) = \sqrt{\left(se(\widehat{LE}_{SoR})\right)^2 + \left(se(\widehat{LE}_{SoB})\right)^2 - 2Cov(\widehat{LE}_{SoR}, \widehat{LE}_{SoB})}$$

We do not have the required data to compute the last term, which requires disclosed tables of survival rates for all possible combinations of state of birth and state of residence. We instead compute the standard error of the difference in life expectancies across measures in a conservative way using an upper bound. It is very likely that the two different estimators of life expectancy are highly positively correlated, as a substantial proportion of individuals never leave their state of birth and are considered in both measures. Thus, a conservative lower bound on $\widehat{Cov}(\widehat{LE}_{SoR}, \widehat{LE}_{SoB})$ is 0. In turn, this means that

$$se\left(\widehat{LE}_{SoR} - \widehat{LE}_{SoB}\right) = \sqrt{\left(se(\widehat{LE}_{SoR})\right)^2 + \left(se(\widehat{LE}_{SoB})\right)^2 - 2Cov(\widehat{LE}_{SoR}, \widehat{LE}_{SoB})}$$

$$\ll \sqrt{\left(se(\widehat{LE}_{SoR})\right)^2 + \left(se(\widehat{LE}_{SoB})\right)^2}$$

Throughout the paper, we use $\sqrt{\left(se(\widehat{LE}_{SoR})\right)^2 + \left(se(\widehat{LE}_{SoB})\right)^2}$ as an upper bound estimate of the standard error of the difference in life expectancy measures in a given sub-population. For all the set of results, we are conservative when we reject that the hypothesis that the difference in life expectancy measures in a given state is equal to zero at a given significance level.

Appendix C: Assessment of fit and robustness checks

In this section we assess the fit of the weighted NLLS model to the aggregate probabilities of surviving throughout the follow-up period disclosed by the Census. Across all our 204 regressions, the distribution of R^2 has an unweighted mean of 0.9999 and a negligible standard deviation. This shows that the Gompertz mortality model provides a very close empirical approximation to how mortality rates grow with respect to age.[6]

In Appendix Figure 2A we plot the observed and predicted values of the 7+ year cumulative survival probabilities for all the (gender x age groups x state x method of aggregation) combinations. We highlight a handful of data points for which the absolute value of the residual is more than 0.05. Overall, we conclude that the fit of the parsimonious nonlinear model based on the Gompertz mortality model is good.
With respect to the second step of our estimation approach, life expectancy estimates have been shown to be particularly sensitive to the assumptions about the mortality rates of the oldest age group in the life table (Németh and Missov, 2018). In our set-up, the oldest age group corresponds to the 85+ group. We perform three different robustness in order to mitigate concerns that assumptions about the oldest age group are driving the empirical patterns about the discrepancies between life expectancies measures that we documented in the previous section.

In our first two robustness exercises, we vary the age at which we stop fitting mortality rates using the weighted NLLS estimation and start using the gender-specific national mortality rates. We instead use ages 85 and 110 as alternative cut-offs. Overall, the cut-offs affect the levels of the life expectancy by state of residence and state of birth. The usage of the age cut-off of 85 is associated with a decrease of 0.17 and 0.19 years on average for men and women, respectively. In contrast, life expectancy estimates that use the alternative age cut-off of 110 are 0.22 and 0.33 years higher for men and women, respectively.

[6] We also assess the fit with the distribution of the mean absolute error across the 204 regressions. This measure is directly interpretable as the average distance between the observed and predicted survival probabilities. The average of the mean absolute error is equal to 0.01.

In our last robustness check, we drop the sample of 85+ individuals entirely from our main sample to assess how influential these observations are on the estimates of differences in life expectancy measures. We instead fit cumulative survival probabilities from the remaining five-year age groups and derive mortality rates from those estimates as in our preferred specification, up to age 90. Afterwards, we use constant mortality rates by gender as in the main specification. We find that dropping individuals in the oldest age group does not change our estimates. The correlation in the difference between life expectancy by state of residence and life expectancy by state of birth is equal to 0.99 for both men and women.

The comparison between the life expectancy estimates in the baseline scenario with the estimates in the alternative scenarios is shown in the graphs in the left side of Appendix Figure 2b. From both graphs of Appendix Figure 2b, we conclude that different cut-offs shift life expectancy estimates uniformly across states.

However, the relevant outcome in our analysis is the difference in life expectancy by state of residence and state of birth for a given state. We show how the estimated differences in life expectancy by state of residence and state of birth vary when we modify the assumptions of mortality at very old ages in the graphs in the right side of Appendix Figure 2b. The use of different cut-offs has a negligible effect on the differences between the two life expectancy measures within states. The unweighted correlation between the differences in life expectancy by state of residence and state of birth in our preferred specification and each of the robustness exercises mentioned above is higher than 0.94, for both men and women.

Appendix D. Differences between life expectancy measures under counterfactual migration scenarios

We perform three different counterfactual scenarios in which we reshuffle the final destination of out-migrants in order to assess how the destination choice of out-migrants affects the life expectancy measures by state of residence and state of birth. In these counterfactual exercises, we hold fixed the number and composition of out-migrants from every state but modify their final destinations.

We require counterfactual relocation probabilities conditional on moving to perform our exercises. We construct these counterfactual probabilities separately by five-year age group in 2008 and gender. In our first counterfactual exercise, we consider that the probability to relocate to a given destination is proportional to the relative importance of that state as a final destination for migrants from all other states. Under this counterfactual scenario, in-migration rates by age and gender are closely related to the in-migration rates observed in the data. However, we modify the state of origin and unobserved mortality risk of the in-migrants that each state receives from the observed patterns in the data.

In our second counterfactual exercise, we instead consider probabilities of relocating to destinations conditional on moving that are proportional to the stayer population across states. Thus, in this second reshuffling exercise we additionally equalize the in-migration rates across states. In our final counterfactual exercise, we instead consider probabilities of relocating to destinations conditional on moving that are proportional to the out-migrant population across states. In this counterfactual exercise we equalize the net migration rate in each (age x gender x state) cell to zero. Importantly, in all counterfactual exercises the allocation of out-migrants from a given state of origin to destinations is independent of unobserved mortality risk.[7]

[7] In mathematical terms, we define the relative importance of state u as a final destination for the sub-population of age group a and gender g ($w_u^{a,g}$) as follows: $w_u^{a,g} = \frac{In-Migrants_u^{a,g}}{\sum_{v \neq u} Out-Migrants_v^{a,g}}$ for the first counterfactual exercise, $w_u^{a,g} = \frac{Stayers_u^{a,g}}{\sum_{v \neq u} Stayers_v^{a,g}}$ for the second counterfactual exercise, and $w_u^{a,g} = \frac{Out-Migrants_u^{a,g}}{\sum_{v \neq u} Out-Migrants_v^{a,g}}$ for the third counterfactual exercise. Notice that the sum of $w_u^{a,g}$ across all states in S does not need to add up to one. Given the vector $\{w_u^{a,g}\}_{u \in S}$, we assume that the counterfactual probability that individuals from age group a and gender g move out of state s to state t is a scaled version of $w_t^{a,g}$,

In order to assess how the reshuffling of migrants to destinations affects the life expectancy measures by state of residence and state of birth, we need to take a stance on the relative importance of causal "place effects" in explaining geographical disparities in mortality outcomes across states (Deryugina and Molitor, 2020; Finkelstein et al, 2021). In a first approach ("No Place Effects"), we assume that out-migrants are not affected by the "place effects" of their final destinations. In other words, we assume that the counterfactual average mortality rates of out-migrants from a given state are not affected when we modify their final destinations. Thus, only the life expectancies by state of residence change in our counterfactuals as a result of modifying the composition of in-migrants, while the life expectancies by state of birth remain unchanged.

In a second approach, we instead assume that "place effects" play an important role in explaining differential mortality patterns of out-migrants ("Constant Place Effects"). We run the following auxiliary regression in order to get a reasonable estimate of how mortality patterns of out-migrants across states are affected by the causal place effects of different destinations:

$$m\_OutMigrants_s^{a,g} = \beta_0^{a,g} + (1-\gamma^g)\, m\_Stayers_s^{a,g} + \gamma^g\, m\_Stayers\_RepDest_s^{a,g} + \varepsilon_s^{a,g}$$

where the dependent term is the average probability of dying in the follow-up period for out-migrants from age group a and gender g that were born in state s, while the two independent variables are the average probability of dying in the follow-up period for stayers from state s ($m\_Stayers_s^{a,g}$) and the average mortality of stayers at the representative destination where out-migrants from s relocate to ($m\_Stayers\_RepDest_s^{a,g}$), respectively. [8] In order to construct the second term, the mortality probabilities of stayers from all states other than s are weighted by the proportion of out-migrants from state s that move to each state.

---

given by $p_{s\to t}^{a,g} = \frac{w_t^{a,g}}{\sum_{v\neq s} w_v^{a,g}}$ if $s \neq t$, and $p_{s\to s}^{a,g} = 0$ by construction. The denominator is required in order to construct relocation probabilities that add up to 1 across all possible destination choices for a given state of origin.

[8] This simple regression specification assumes, among other things, that the "place effects" of a given state are constant for stayers and in-migrants, and that the relative importance of the final destination in determining mortality outcomes is constant across age groups.

In this simplified model, the parameter $\gamma^g \in [0,1]$ assesses the extent to which the mortality of out-migrants is associated to the mortality of the stayers from their state of birth and to the mortality of stayers at their representative destination. We allow this coefficient to vary by gender in our main specification. The "No Place Effects" approach corresponds to the limit case in which $\gamma^g = 0$. Meanwhile, the vector of coefficients $\beta_0^{a,g}$ captures the overall mortality advantage of out-migrants with respect to stayers across states for the sub-population of individuals in age group a and gender g.

We run constrained least squares imposing the constraint that the sum of the coefficients in front of $m\_Stayers_s^{a,g}$ and $m\_Stayers\_RepDest_s^{a,g}$ adds up to one. We weigh observations by the number of out-migrants in each cell. In our main specification, we estimate that γ is equal to 0.385 (s.e. = 0.06) for men. We use the estimate $\widehat{\gamma^g}$ to compute counterfactual mortality profiles of out-migrants when we modify their final destinations. Under a counterfactual scenario where the representative destination of out-migrants is given by $m\_\widetilde{Stayers}_s^{a,g}$, we assume that the counterfactual average mortality rate of out-migrants from state s ($m\_\widetilde{OutMigrants}_s^{a,g}$) will be given by:

$$m\_\widetilde{OutMigrants}_s^{a,g} = m\_OutMigrants_s^{a,g} + \widehat{\gamma^{a,g}} \left(m\_\widetilde{Stayers}_s^{a,g} - m\_Stayers_s^{a,g} \right)$$

In Appendix Figure 10 we plot the observed and the three counterfactual differences in male life expectancies at age 50 by state of residence and state of birth for the subset of states where the differences in life expectancy measures are significant at the 10 percent level. Panel A shows the observed and counterfactual differences in life expectancy measures under the assumption of "No Place Effects". In contrast, Panel B shows the observed and counterfactual differences in life expectancy measures under the assumption of "Constant Place Effects", in which the mortality rates of out-migrants are affected by the final destination choice.

From Panel A of Appendix Figure 10 we observe that there are important changes between the observed differences in life expectancy measures and the differences in the first counterfactual scenario where we reshuffle individuals into destinations in a way that preserves in-migration rates but modifies the composition of in-migrants that each state receives. This means that we would expect an absolute difference between life expectancy measures that is below 0.5 for the states of OK, OH,

IL, IN, and LA if in-migrants to these states were selected randomly from the pool of available out-migrants. However, in the data the absolute differences are considerably higher. Thus, our counterfactual exercise provides supportive evidence that these states attract migrants with ex-ante higher mortality risk under the "No Place Effects" assumption. In contrast, our results are suggestive that states like SC, CO, VA, and MD appear to attract in-migrants with lower ex-ante mortality risk. Overall, the mean absolute difference between life expectancy measures for these thirteen states is equal to 0.87 years in the data, but only between 0.51 years in the first counterfactual exercise.

Comparing the results between the first counterfactual scenario and the other counterfactual exercises, we find that the equalization of in-migration or net migration rates across states further contributes to the attenuation of the differences in life expectancy measures across states. However, it plays a quantitatively smaller role than the homogenization of the ex-ante mortality risk of in-migrants for most states. Only in the case of the state of Florida, which has an in-migration rate of more than 80 percent in the sub-population of men ages 50 and above, the further equalization of migration rates modifies the difference in life expectancy measures by more than 0.5 years.

Overall, the cross-standard deviation in the difference between life expectancy measures for the thirteen selected states is equal to 0.95. This standard deviation is reduced to 0.61 when we reshuffle in-migrants, holding fixed state immigration rates. Finally, when we further equalize in-migration or net migration rates the cross-state standard deviation lowers to 0.36 and 0.33, respectively. Thus, we find that 36 percent of the cross-state standard deviation in life expectancy measures in the subset of states where the difference between life expectancy measures is significant can be attributed to the non-random sorting of out-migrants to locations. After accounting for the composition of in-migrants, the differences in migration rates across states contributes to a remaining 26 to 30 percent of the observed cross-state standard deviation.[9]

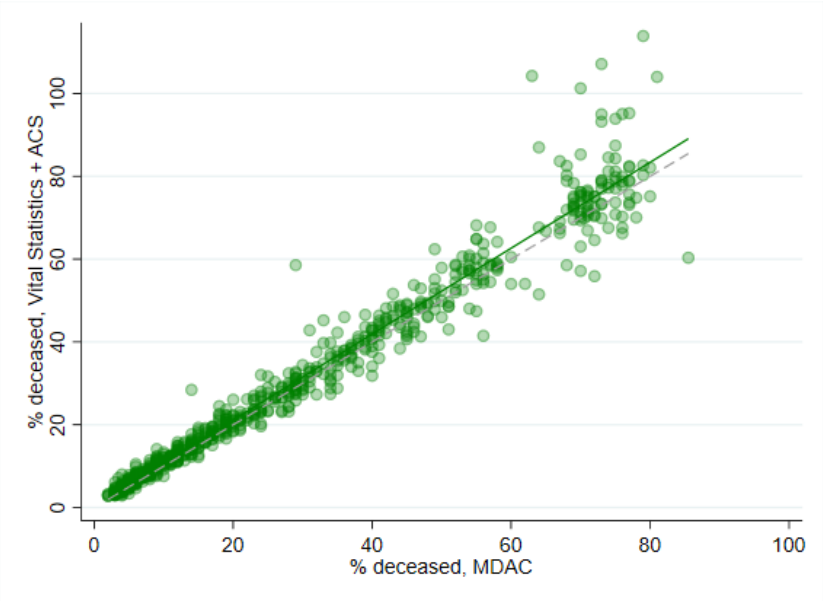
[9] We also perform the decomposition of the cross-state deviation in the difference between life expectancy measures for all 43 states where we can compute male life expectancies of stayers. We find that the non-random sorting of out-migrants to locations accounts for 20 percent of the cross-state standard deviation across states, while the further equalization of migration rates accounts for 6 to 11 percent.

An alternative assumption about "place effects" is that the differences in the mortality risk of in-migrants across destinations is a joint outcome determined by the differential sorting of in-migrants from different states of birth and unobserved mortality risk and cross-state variation in destination place effects. In Panel B of Appendix Figure 10 we show the observed and counterfactual differences in life expectancy measures under the "Constant Place Effects" scenario. The counterfactual differences in Panel B are quite similar to those presented in Panel A. Hence, the key takeaway that the non-random sorting of in-migrants across final destinations is an important driver behind the cross-state variation in the difference of life expectancy measures continues to hold under the "Constant Place Effects" scenario and is not driven by the assumptions about the relative importance of "place effects" in determining mortality outcomes later in life.[10]

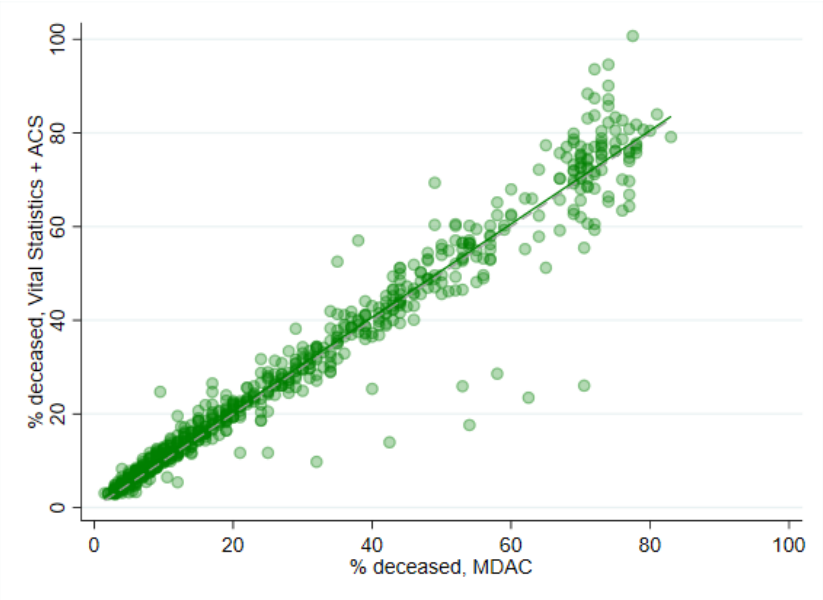
[10] Under the "Constant Place Effects" assumption, the non-random sorting of out-migrants to locations accounts for 49 percent of the cross-state standard deviation across the subset of selected states, while the further equalization of migration rates accounts for 22 to 23 percent.

Appendix Figure 1: Comparison of cumulative mortality probability in the follow-up period, MDAC and NVSS

Panel A: Mortality probabilities, by state of residence

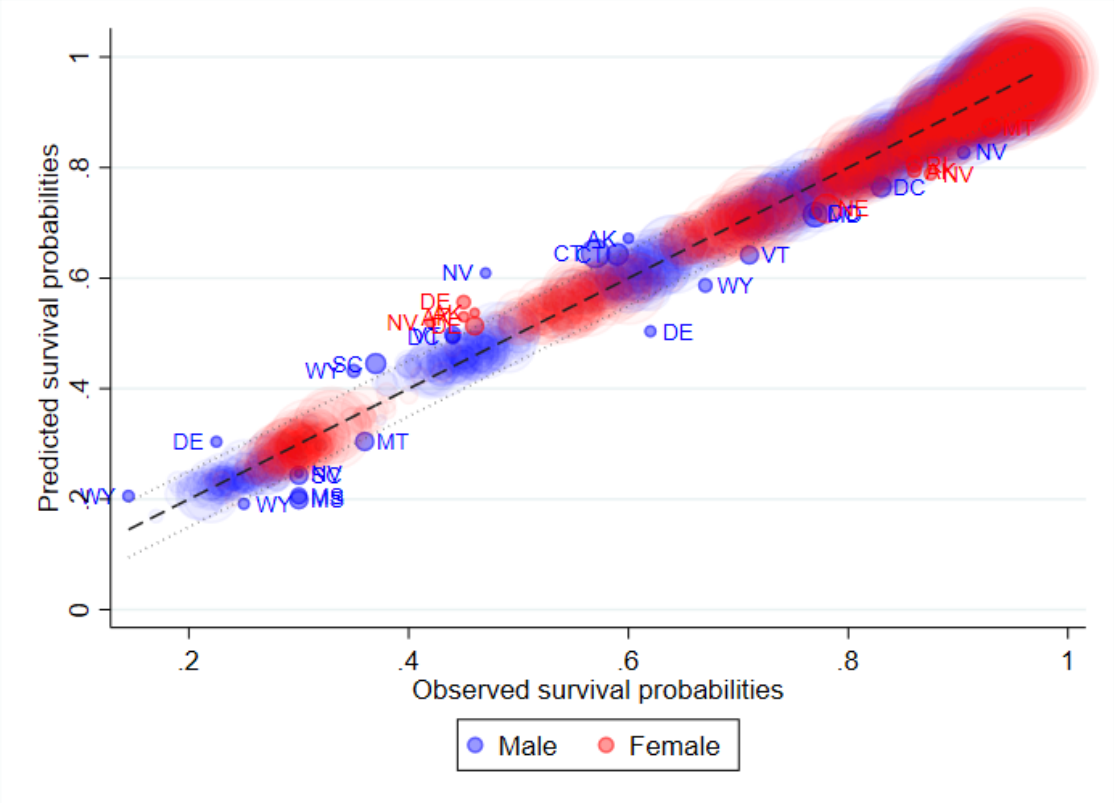


Panel B: Mortality probabilities, by state of birth



Note: Panel A (B) of Appendix Figure 1 shows the comparison between the raw cumulative mortality probabilities by state of residence (state of birth) observed in the MDAC dataset and equivalently constructed cumulative rates using the NVSS data. Each dot corresponds to a (five-year age group x gender x state) combination. To estimate cumulative mortality rates using the NVSS data, all deaths that occurred between July 2008 and December 2015 were clustered by age group of the individual in 2008, as well as gender and state.

Appendix Figure 2a: Assessment of fit in cumulative survival probabilities in 7+ follow-up period

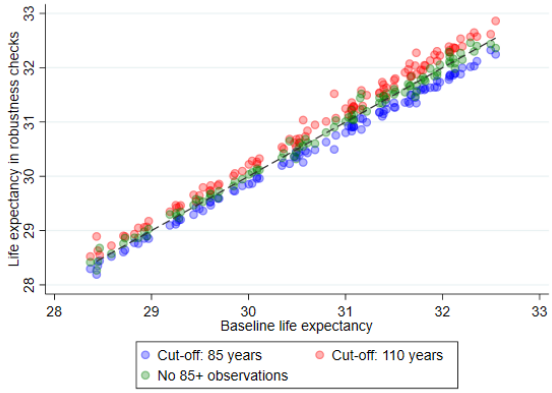


Note: Appendix Figure 2a shows the fit of the mortality models to the observed data. The average survival probability of being alive after the 7+ follow-up period by five-year age group, state, gender, and method of aggregation is shown on the x-axis. The predicted probabilities generated from weighted NLLS regressions that are run separately by state, gender, and method of aggregation are depicted on the y-axis. The size of each bubble is proportional to the number of individuals surveyed in the 2008 ACS in each cell. Cells in which the difference between the actual and predicted value is higher than 0.05 are highlighted.

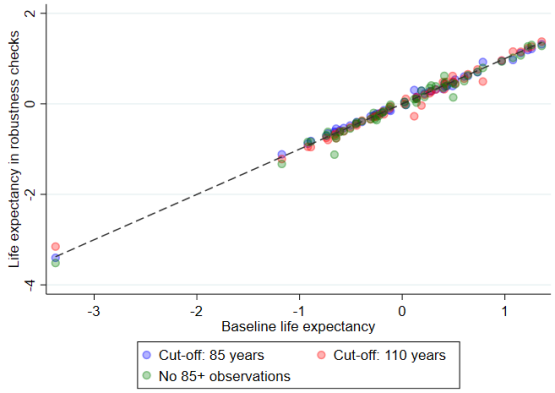
Appendix Figure 2b: Life expectancies at age 50 by state of residence and state of birth under different assumptions of mortality at very old ages

Panel A: Men

Levels

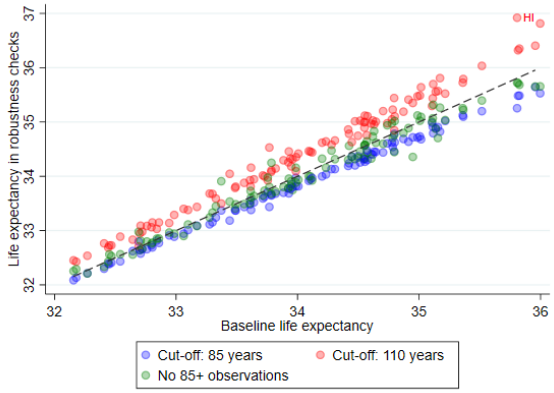


Differences

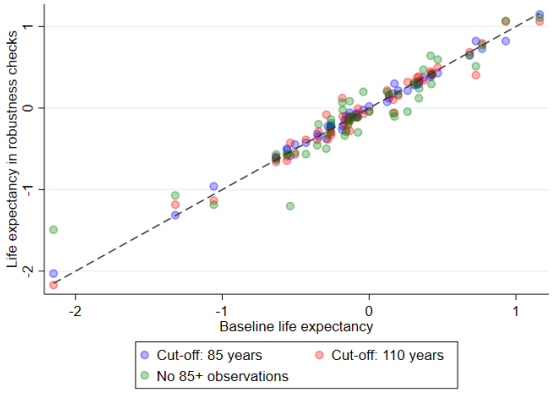


Panel B: Women

Levels



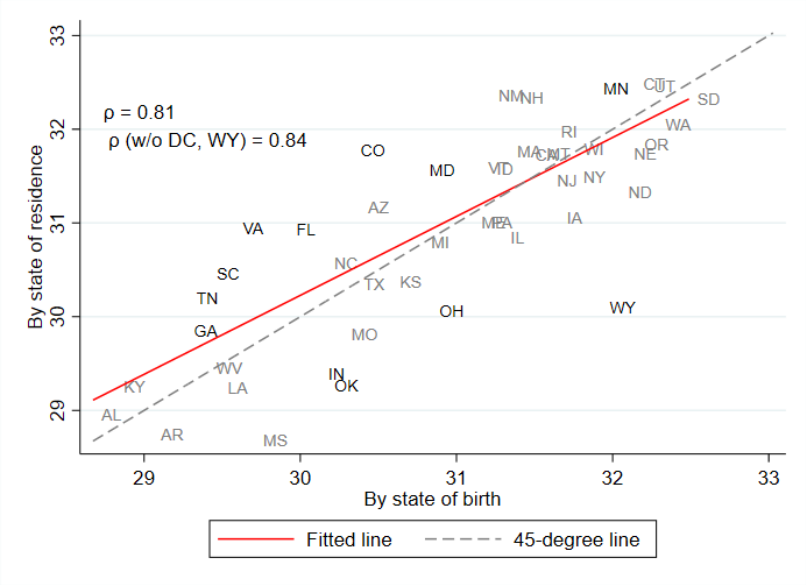
Differences



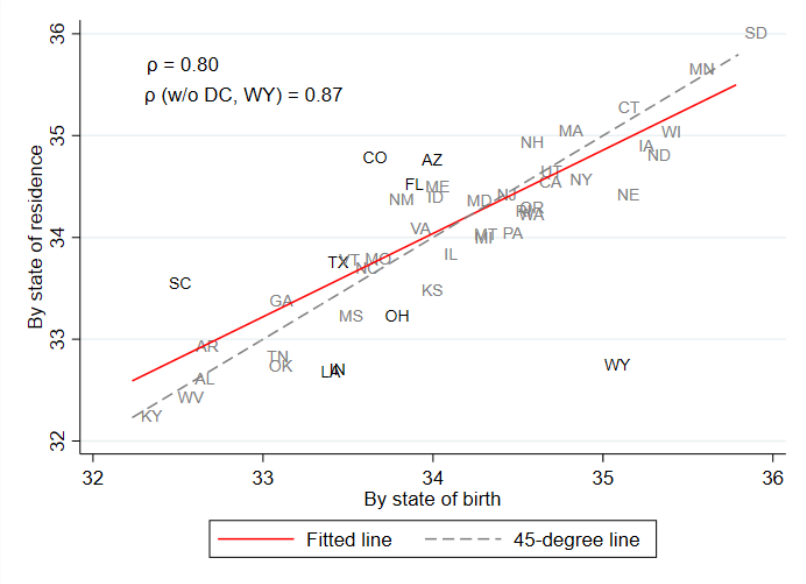
Note: Appendix Figure 2b shows the life expectancy estimates under three different robustness checks. Panel A shows the estimates for men, while Panel B shows the estimates for women. On the x-axis of the graphs in the left side, we plot the life expectancy estimates from our preferred specification, in which we stop using the mortality rates derived from the weighted NLLS estimation at age 90. On the y-axis, we include the life expectancy estimates from models that use different assumptions about mortality patterns at very old ages. The dots in blue color correspond to estimates where we use the fitted mortality rates from the Gompertz model to estimate mortality rates up to age 85. The dots in red correspond to life expectancy estimates where we use the fitted mortality probabilities from the weighted NLLS estimation up to age 110. The dots in green correspond to life expectancy estimates that entirely discard the observed survival probabilities for the age group 85+ in the data. Each dot within a series corresponds to a sub-population characterized by a (state x method of aggregation) combination. Observations where the difference between the baseline life expectancy estimate and the alternative life expectancy measure is bigger than one year are highlighted. In the graphs on the right hand side, we instead plot in the x-axis the difference between the life expectancy by state of residence and life expectancy by state of birth in our preferred specification and in the y-axis the difference between the life expectancy measures using alternative assumptions about mortality patterns at very old ages.

Appendix Figure 3: Life expectancies at age 50 by state of birth and state of residence, White population only

Panel A: Men



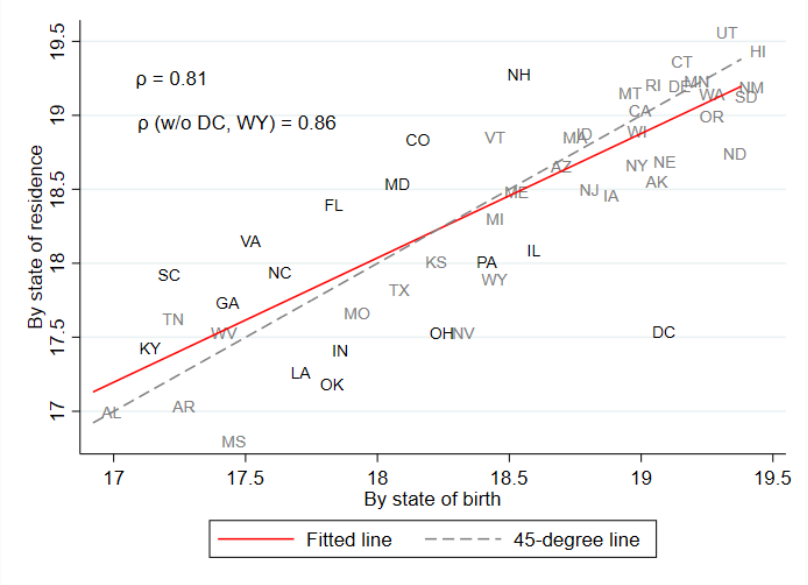
Panel B: Women



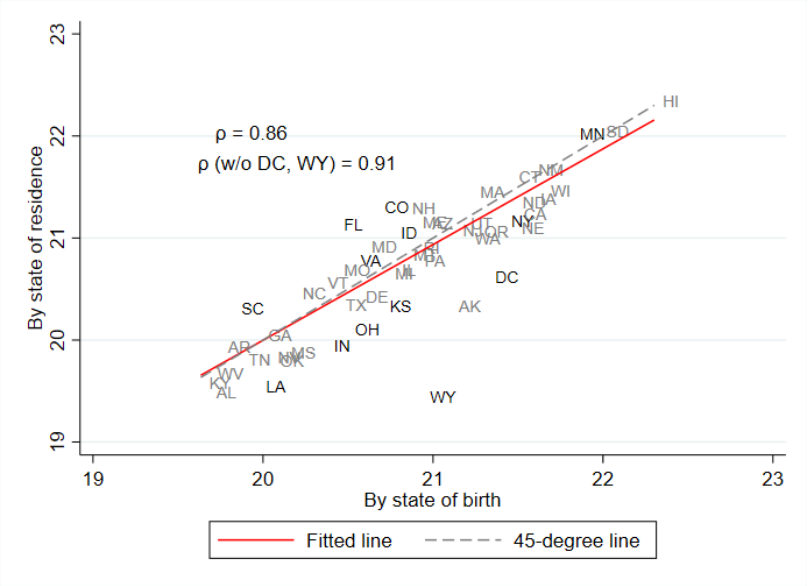
Note: Appendix Figure 3 shows the relationship between life expectancy at age 50 excluding Black and Latino individuals by state of birth and state of residence separately by gender. Life expectancies were constructed using data disclosed from the MDAC dataset with Census disclosure numbers CBDRB-FY20-CES004-022, using the methods explained in the paper and further detailed in Technical Appendix A. Panel A shows the relationship between the two alternative measures of life expectancy for men. Panel B shows the relationship for women. States that have a significant difference between life expectancy measures at the 10% level are marked in black. The rest of the states are depicted in gray.

Appendix Figure 4: Life expectancies at age 65 by state of birth and state of residence

Panel A: Men



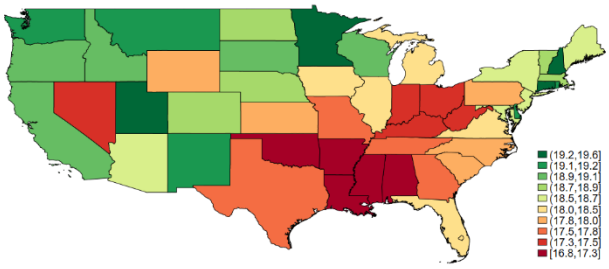
Panel B: Women



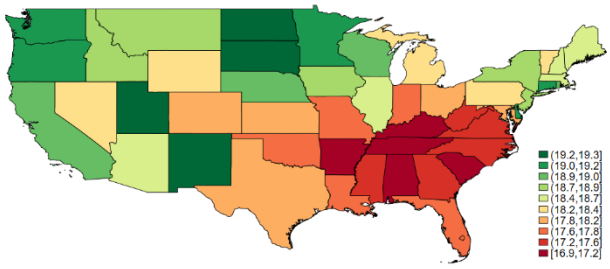
Note: Appendix Figure 4 shows the relationship between life expectancy at age 65 by state of birth and state of residence separately by gender. Life expectancies were constructed using data disclosed from the MDAC dataset with Census disclosure numbers CBDRB-FY19-304 and CBDRB-FY20-092, using the methods explained in the paper and further detailed in Technical Appendix A. Panel A shows the relationship between the two alternative measures of life expectancy for men. Panel B shows the relationship for women. States that have a significant difference between life expectancy measures at the 10% level are marked in black. The rest of the states are depicted in gray.

Appendix Figure 5: Male life expectancy at age 65, 2008-2015

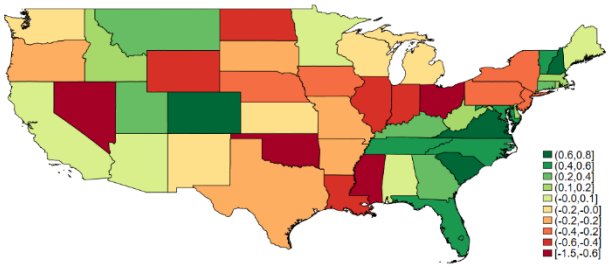
Panel A. Life expectancy by state of residence



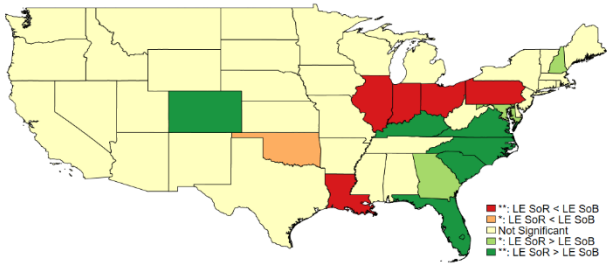
Panel B. Life expectancy by state of birth



Panel C. Difference in life expectancy measures State of residence – State of birth (all)



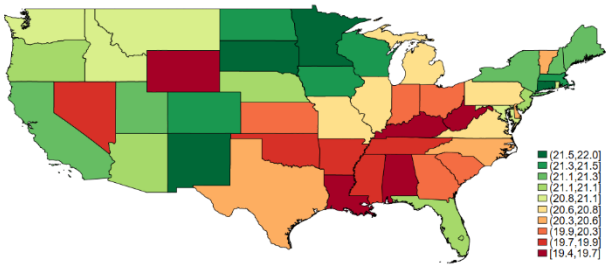
Panel D. Difference in life expectancy measures State of residence – State of birth (significance)



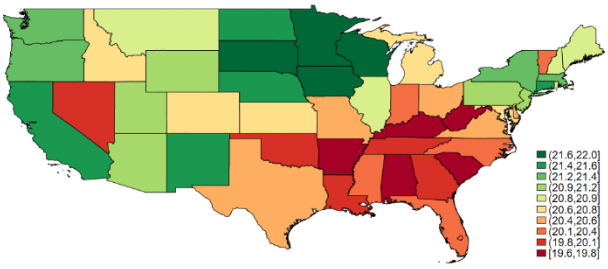
Note: Panel A of Appendix Figure 5 presents male life expectancies at age 65 grouping individuals by their state of residence in 2008, while Panel B of Appendix Figure 5 presents life expectancies at age 65 grouping individuals by their state of birth. Panel C of Appendix Figure 5 shows the differences between life expectancies by state of residence and life expectancies by state of birth for each of the states. Panel D of Appendix Figure 5 shows in red states in which the life expectancy by state of residence is significantly lower than the life expectancy by state of birth at the 5 and 10 percent significance levels. States in which the life expectancy by state of residence is significantly higher than the life expectancy by state of birth at the 5 and 10 significance levels are shown in green.

Appendix Figure 6: Female life expectancy at age 65, 2008-2015

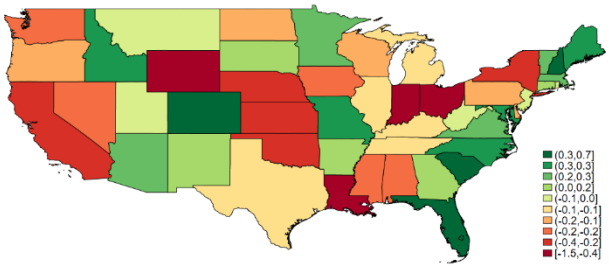
Panel A. Life expectancy by state of residence



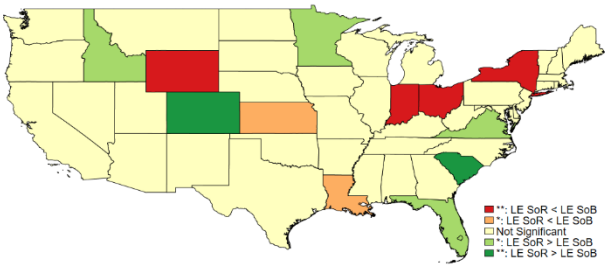
Panel B. Life expectancy by state of birth



Panel C. Difference in life expectancy measures State of residence – State of birth (all)

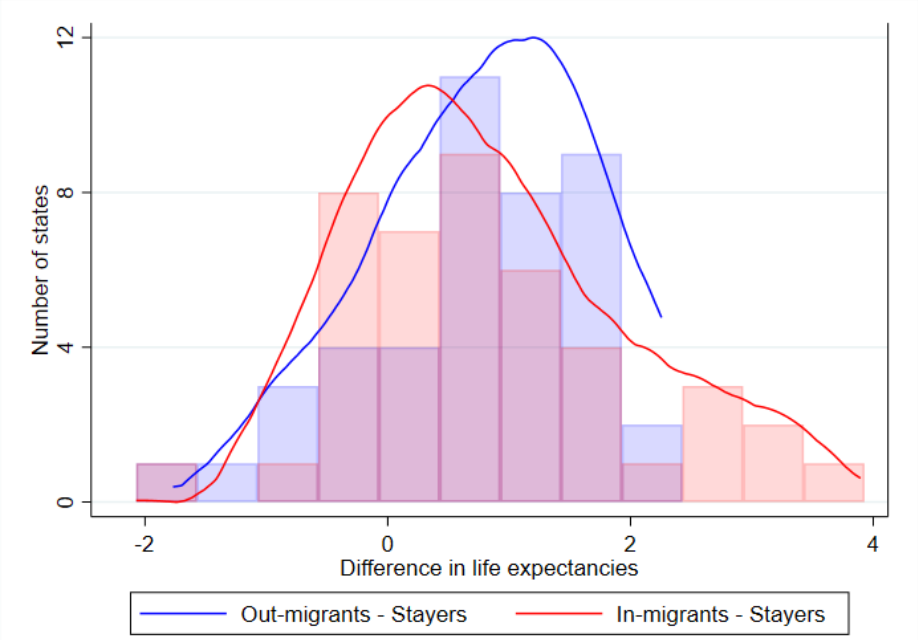


Panel D. Difference in life expectancy measures State of residence – State of birth (significance)



Note: Panel A of Appendix Figure 6 presents female life expectancies at age 65 grouping individuals by their state of residence in 2008, while Panel B of Appendix Figure 6 presents life expectancies at age 65 grouping individuals by their state of birth. Panel C of Appendix Figure 6 shows the differences between life expectancies by state of residence and life expectancies by state of birth for each of the states. Panel D of Appendix Figure 6 shows in red states in which the life expectancy by state of residence is significantly lower than the life expectancy by state of birth at the 5 and 10 percent significance levels. States in which the life expectancy by state of residence is significantly higher than the life expectancy by state of birth at the 5 and 10 significance levels are shown in green.

Appendix Figure 7: Distribution of ex-post relative mortality advantage of male in-migrants and out-migrants across states

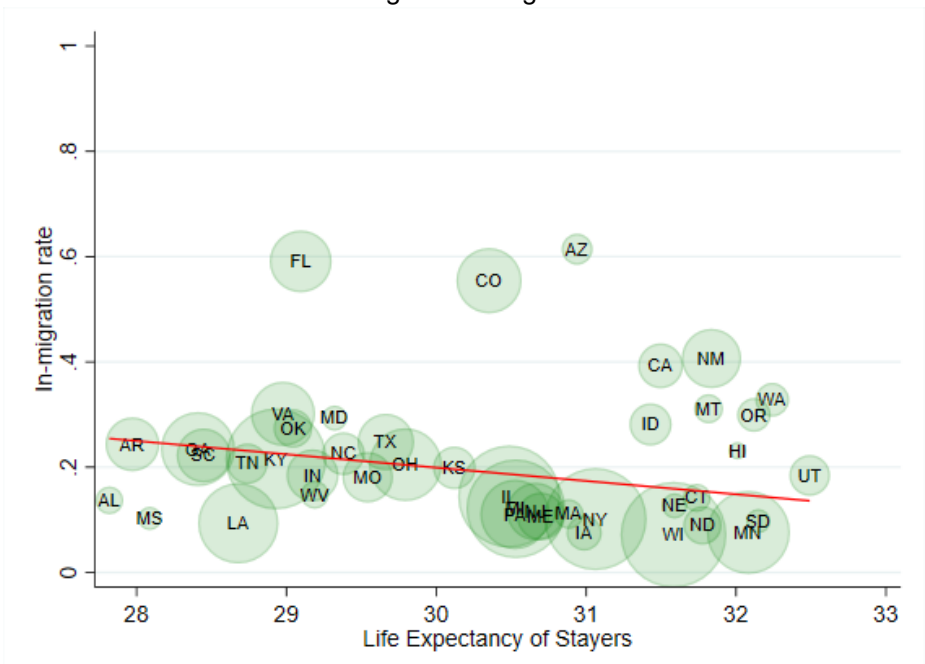


Note: Appendix Figure 7 shows the histogram and smoothed distribution of the relative health advantage of male in-migrants and out-migrants relative to stayers across states.

Appendix Figure 8: Relationship between life expectancy of male stayers at age 50 and regional migration rates

Panel A: Regional out-migration rates

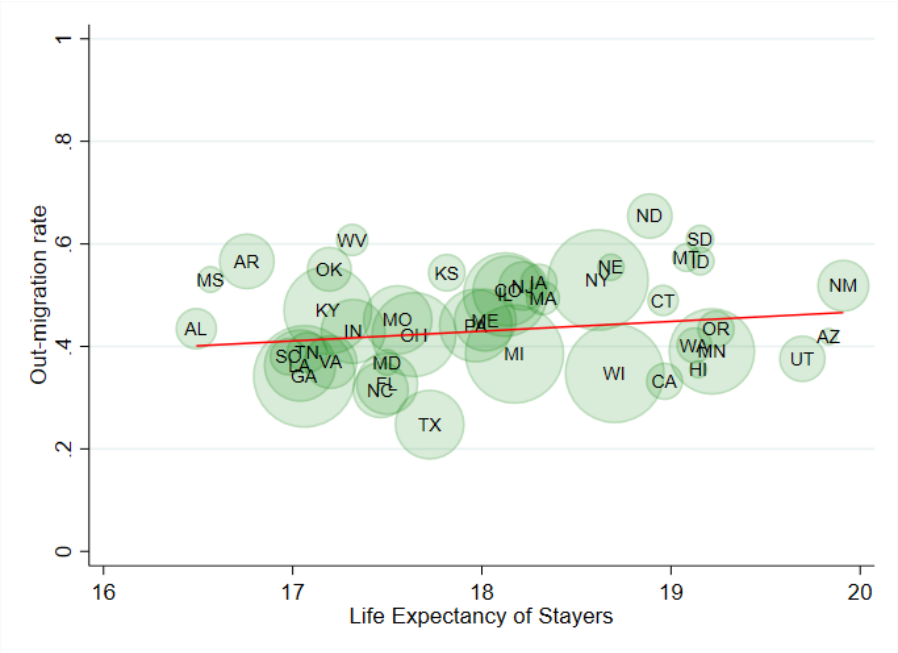
Panel B: Regional in-migration rates



Note: Panel A (B) of Appendix Figure 8 shows the relationship between male life expectancy of stayers at age 50 and regional out-migration (in-migration) rates at the state level. We define out-migration rate of state s as the proportion of 50-64 year-old men that were born in state s that are out of their region of birth by the time of the 2008 ACS interview. Similarly, the in-migration rate of state s is calculated as the proportion of 50-64 year-old men that are observed in s by the time of the ACS interview that were born in a different region. In both panels, states are weighted by the inverse variance of the male life expectancy of stayers at age 50.

Appendix Figure 9: Relationship between life expectancy of male stayers at age 65 and migration rates

Panel A: Out-migration rates

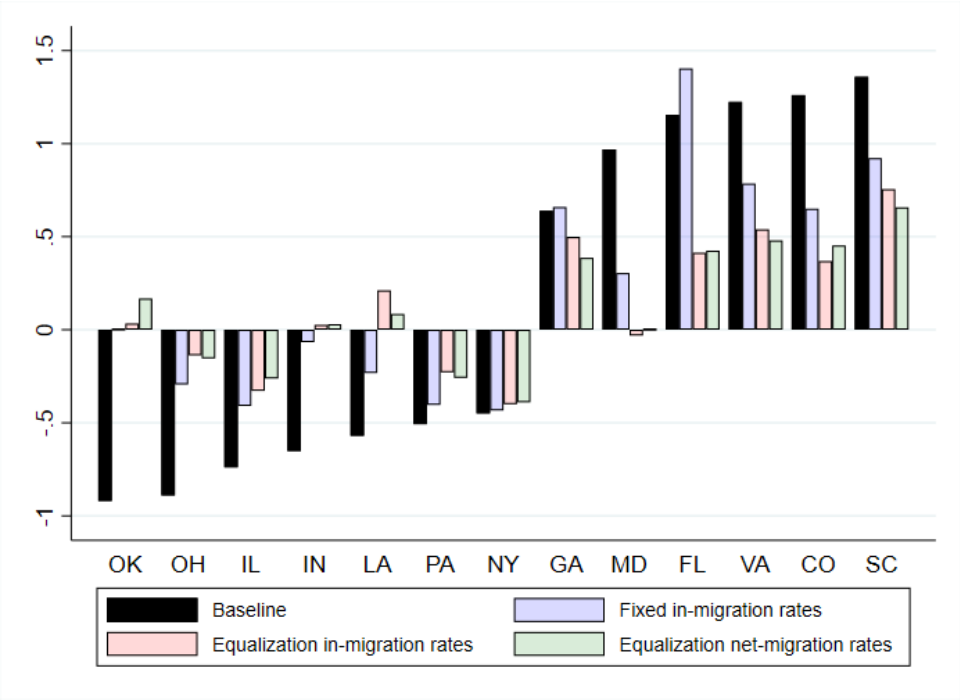


Panel B: In-migration rates

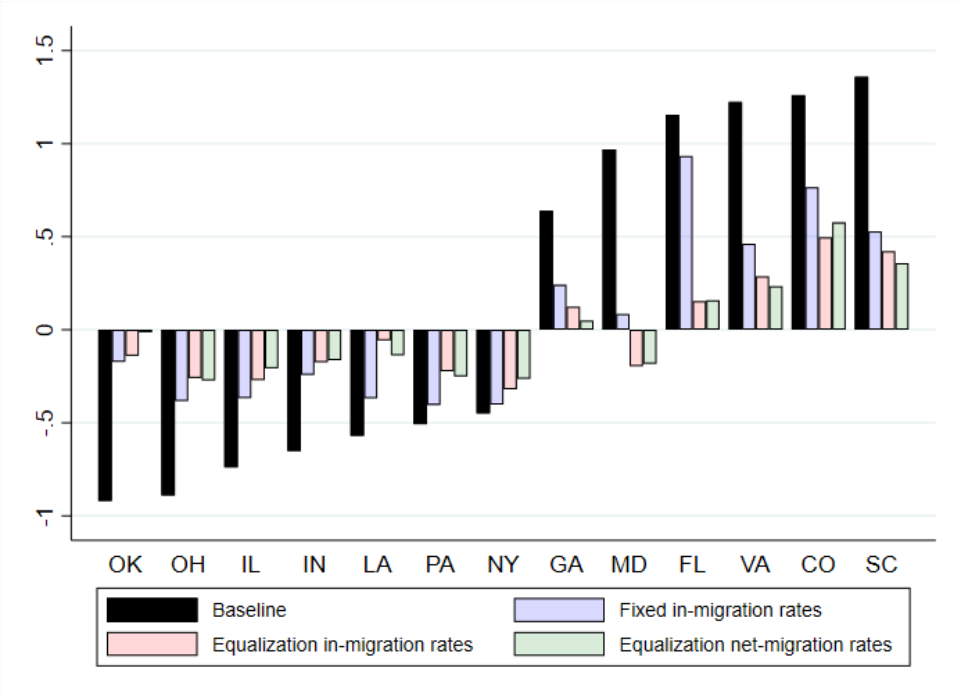
Note: Panel A (B) of Appendix Figure 9 shows the relationship between male life expectancy of the stayer sub-population at age 65 and out-migration (in-migration) rates at the state level. The out-migration rate of state s is calculated as the proportion of 65-79-year-old men that were born in state s that are out of their state of birth by the time of the 2008 ACS interview. Similarly, the in-migration rate of state s is calculated as the proportion of 65-79-year-old men that are observed in s by the time of the ACS interview that were born in a different state. In both panels, states are weighted by the inverse variance of the male life expectancy of stayers at age 65.

Appendix Figure 10: Counterfactual differences in male life expectancies at age 50 in selected states

Panel A: No Place Effects



Panel B: Constant Place Effects



Note: Appendix Figure 10 shows the observed and counterfactual differences in male life expectancies by state of residence and state of birth in the three different reshuffling exercises for the subset of states where the difference between life expectancy by state of residence and state of birth was significant at the 10 percent level. Panel A (B) shows the observed and counterfactual differences under the assumption of "No Place Effects" ("Constant Place Effects"). Appendix D explains how the counterfactual migrations probabilities and mortality rates were constructed.

Appendix Table 1: Relative importance of ex-post relative mortality advantage of in-migrants and out-migrants for differences in male life expectancies at age 50

Dependent variable: $1(LE^{SoR} - LE^{SoB} > 0)$	(1)	(2)
$LE^{In} - LE^{Stay}$	0.290	0.235
	(0.031)	(0.062)
$LE^{Out} - LE^{Stay}$	-0.065	-0.192
	(0.110)	(0.171)
LE^{Stay}	-	-0.101
		(0.100)
R^2	0.887	0.896
Number of observations	13	13

Note: Appendix Table 1 shows the relative importance of the ex-post relative mortality advantage of out-migrants and ex-post relative mortality advantage of in-migrants in determining the sign of the difference in the life expectancy by state of residence and life expectancy by state of birth. The dependent variable is equal to one if the difference between life expectancy by state of residence and state of birth is positive. We included the set of 13 states with significant differences. The results show that the ex-post relative health advantage of out-migrants is only weakly related to the sign of the difference in life expectancy measures. In contrast, an additional year in the differential health advantage of in-migrants is associated with a significant 24 to 29 percentage point increase in the probability that the difference is positive (p-value: 0.000). Column (1) only includes the ex-post relative mortality advantage of in-migrants $(LE^{In} - LE^{Stay})$ and out-migrants $(LE^{Out} - LE^{Stay})$ as regressors. Column (2) also includes the life expectancy of stayers (LE^{Stay}) as an additional regressor. In both columns, states are weighted by the inverse variance of the male life expectancy of stayers at age 50. This table was constructed from information disclosed by the Census with DRB review numbers CBDRB-FY19-304 and CBDRB-FY20-092.