Supplemental Appendix

This appendix is divided into five parts. Part I provides comparisons of SSA data on mortality with data from the NCHS. Part II presents a complete description of our methodology for estimating life expectancies. Part III presents further detail on the construction of certain local area characteristics used in the correlational analysis in Figures 8 and 9. Part IV assesses the sensitivity of our life expectancy estimates to alternative assumptions. Part V presents a set of supplemental results, including county level estimates of life expectancy. All of the commuting zone (CZ), county, state, and national level statistics that were constructed for the analysis as well as replication code can be downloaded from www.healthinequality.org.

Part I: Comparison of Mortality Rates in SSA vs. NCHS data

The death records we employ are from the Social Security Administration's (SSA) Death Master File. Most existing analyses of mortality in the United States are based on deaths recorded by the National Center for Health Statistics (NCHS).[1] In eTable 1, we compare mortality counts and rates in the NCHS and SSA data, which help explain why we restrict our analysis to people with positive earnings. The statistics in this table are averages over 2001-2014 for all individuals between the ages of 40 and 63, the set of "working age" individuals in our dataset.

The first panel of eTable 1 compares total population counts across various samples, by sex. Column 1 reports the total number of people aged 40-63 recorded by the NCHS. Column 2 shows the number of individuals aged 40-63 with a valid SSN recorded by the SSA. The SSA data contains 22% more men and 16% more women than the NCHS data, because it includes all individuals who have ever had an SSN in the U.S., whereas NCHS data only includes current U.S. residents.

Column 3 shows the number of people with an SSN who file a tax return or have an information return (such as a W-2 form or social security income form) filed on their behalf. This subsample still contains 10% more men and 5% more women than the NCHS data, partly because non-residents continue to have some taxable income and benefits in the U.S.

In Column 4, we further subset this group to those with positive household earnings two years earlier, i.e. the number of individuals in our primary analysis sample. This sample covers 91% of the U.S. resident population. A small fraction of individuals (representing approximately 1% of the U.S. resident population) report negative household income in the tax data. Hence, roughly 8% (100-91-1) of individuals have zero household earnings (or do not have an SSN). Column 5 shows the number of individuals with zero household earnings in the tax records. This number is larger than 8% because it includes non-residents.

The second panel of eTable 1 compares the total count of deaths in each of the five samples. Column 1 shows the total count of deaths of U.S. residents as recorded by the NCHS. The average number of deaths recorded per year in the NCHS and the SSA are very similar, with a difference of 2% for men and 4% for women. The SSA data should be expected to have slightly fewer deaths than NCHS data because they only include deaths of individuals with a Social Security Number. eFigure 1 shows that the total count of deaths by age and by year (pooling all individuals aged 40-76) are also very similar in the NCHS and SSA data.

Importantly, the total death counts are very similar even though the total population recorded by the SSA is roughly 20% larger. This is because the Social Security Administration does not receive a systematic record of deaths for former residents of the United States. Since we cannot identify non-residents in our data and non-residents typically have zero earnings in tax data, our dataset understates mortality rates for individuals with zero earnings. This is why we focus on individuals with positive earnings in our primary analysis.

Columns 4 and 5 show that individuals with positive earnings account for 62% of the total deaths recorded in NCHS statistics, while those with zero household earnings account for 32% of the deaths. Since approximately 8% of the population has zero household earnings, this implies that the mortality rate of individuals with zero earnings is 4 times larger than the mean mortality rate of individuals with positive earnings. This is consistent with eFigure 2A, which shows that the mortality rate for individuals in the bottom 1% of the income distribution is also nearly 4 times larger than the mean mortality rate for individuals with positive earnings.

The third panel of eTable 1 shows mortality rates per 100,000 individuals, i.e. the total death counts in Panel B divided by the total population counts in Panel A. Mortality rates are about 20% lower in the SSA data (Column 2) than the NCHS data (Column 1) because the denominator in the SSA data include some people who are no longer residents, as noted above. Restricting the sample to those with IRS tax forms (Column 3) reduces the gap in mortality rates to about 12%. The mortality rate for individuals with positive earnings (Column 4) is 68% as large as that in the population as a whole because the mortality rate for those with zero income is much higher than average. The estimated mortality rate for individuals with zero earnings in the tax data (Column 5) is approximately twice as large as in the population as a whole. This two-fold difference is smaller than the factor of four calculated above precisely because the tax data under-counts deaths for individuals with zero earnings.

In sum, the SSA and NCHS datasets are well aligned in their records of deaths of U.S. residents. However, the SSA data include additional individuals in the denominator because it does not exclude non-residents. Excluding individuals with zero earnings rectifies this problem. Since individuals with positive earnings have lower mortality rates than those with zero earnings, mean life expectancies are higher in our sample than in the NCHS data.

Part II: Methodology for Estimation of Life Expectancies

Our goal is to estimate race- and ethnicity-adjusted life expectancy by income percentile at age 40. If we observed race and mortality rates for all ages by income at age 40, it would have been straightforward to calculate life expectancies non-parametrically based on the empirical survival curves.

In practice, we face two missing data problems in conducting this analysis. First, because of the limited time span of the tax data (1999-2014), we do not observe income at age 40 when measuring mortality rates after age 55. Second, tax records do not contain information on race and ethnicity. Here, we describe our methodology for overcoming these two missing data problems and present supplementary evidence for our approach. Our approach consists of three elements: (1) exploiting the invariance of mortality rates to the year in which income is measured, (2) using Gompertz extrapolations to predict mortality rates at older ages, and (3) using data on race shares and mortality differences by race from other sources to make race adjustments. We discuss each element in turn, and finally describe how we calculate standard errors for the life expectancy estimates using a bootstrap procedure.

A. Income Lag Invariance

Let p_{40} denote an individual's income percentile at age 40. Because our data span only 15 years, p_{40} is not observed when measuring mortality at age $a > 55$. Instead, we measured mortality rates at age a conditional on an individual's income percentile at age $a\text{-}2$ for ages $a \leq$

63 and using income at age 61 for ages $a > 63$. For example, the life table for individuals at the 5th percentile was constructed using data for those at the 5th percentile at each age *a-2* (or at age 61) rather than those at the 5th percentile at age 40.

An important concern in interpreting the percentile-specific life tables is that individuals' incomes fluctuate over time; an individual in the 5th percentile at age 50 might not have been in the 5th percentile at age 40. We evaluate this concern in eFigure 2A by plotting mortality rates vs. income percentile measured with a two, five, and ten year lag for men and women aged 50-54 in 2014 (results are similar for other groups). The three series in eFigure 2A are very similar: the relationship between mortality rates at age a and income percentile at age a-t (p_{a-t}) is essentially invariant to the lag t with which income is measured. This lag invariance property implies that the life tables we construct using income at later ages are approximately equivalent to life tables conditional on income percentile at age 40.

The lag with which income is measured does not matter greatly for estimates of mortality rates by income because individuals' income percentiles are very stable over time at the ages we studied. eFigure 2B plots the correlation between an individual's household income percentile in the current year and x years ago, restricting the sample to those who are 61 or younger in the current year. For men, the correlation between an individual's income percentile in year t and t-1 is 0.89, and the correlation between his income percentile in year t and t-10 is 0.70.

Prior work has argued that it may be necessary to average income over several years in order to obtain reliable measures of permanent income. Using averages over several years instead of income in a single year has little impact on our results. In Column 1 of eTable 2, we regress an indicator for death on income percentile in year t-2 (our baseline specification for the working age sample); in columns 2 and 3, we use the average income percentile over the preceding 5 years (years *t-2 to t-6*) and 10 years (years *t-2* to *t-11*). The estimated relationship between mortality rates and income percentiles is virtually identical with all three measures. Averaging does not matter much because of the high degree of stability in an individual's income percentile as measured in tax returns, which have much less measurement error than the surveys used in most prior studies.

B. Gompertz Approximations to Mortality Rates

Using the approach described above, we measure mortality rates up to age 76 as a function of income percentile at or before age 61. To calculate mortality rates beyond age 76, we use extrapolations based on Gompertz distributions.

As far back as 1825, Benjamin Gompertz noted that the logarithm of mortality rates was approximately linear in age a: $\log(\mathrm{m}(a)) = \alpha + \beta a$. We examined the current empirical validity of this equation using NCHS data. eFigure 3 plots log mortality rates by age and gender using NCHS data for the U.S. population in 2001. The linear Gompertz fit is quite accurate overall, with an R-squared exceeding 0.99 for both men and women. However, after age 90, mortality rates rise less rapidly than the Gompertz model predicts, as is well known from prior work.[2] We therefore estimate Gompertz models using mortality rates for the ages we observe and extrapolate to predict mortality rates up to age 90.

After age 90, we assume that all people have the same (gender-specific) mortality as observed in the NCHS and SSA published life tables, independent of income and year. Mortality rates from age 90 to 99 are obtained from NCHS life tables, averaging mortality rates from 2001 to 2011.[1] Because NCHS does not provide estimates of mortality rates beyond age 99, we use estimates provided from SSA data for the year 2000 for ages 100 to 119.[3]

Our methodology assumes that both the group-specific mortality distributions (e.g., by income percentile and area) are Gompertz and that the aggregate distribution of mortality is Gompertz. This assumption is motivated by empirical evidence that the Gompertz model fits not just aggregate mortality rates in the population but also mortality rates in each area and at each income percentile (Figure 1). This empirical regularity is somewhat surprising given that the sum of Gompertz random variables does not follow a Gompertz distribution. However, the sum of Gompertz random variables is approximately Gompertz when the differences in mortality rates across the groups are sufficiently small.

To establish this approximation formally, let $\mathrm{m_{ia}}$ denote the mortality rate in group i at age a and ω_{ia} denote the fraction of individuals at age a who belong to group i. Assume mortality rates in each subgroup follow a Gompertz distribution: $\log(\mathrm{m_{ia}}) = \alpha_{\mathrm{i}} + \beta_{\mathrm{i}}\mathrm{a}$. The aggregate mortality at age a is $\mathrm{m_a} = \sum_i \omega_{\mathrm{ia}}\,\mathrm{m_{ia}}$. Taking a second-order Taylor approximation of $\log(\mathrm{m_{ia}})$ around $\mathrm{m_a}$ yields

$$\log(\mathrm{m_{ia}}) \approx \log(\mathrm{m_a}) + \frac{(\mathrm{m_{ia}} - \mathrm{m_a})^2}{2\mathrm{m_a}^2}.$$

Taking a weighted sum over the groups gives:

$$\sum_i \omega_{\mathrm{ia}} \log(\mathrm{m_{ia}}) \approx \log(\mathrm{m_a}) + \frac{\mathrm{Var}(\mathrm{m_{ia}})}{2\mathrm{m_a}^2}$$

Hence $\log(\mathrm{m_a}) \rightarrow \sum_i \omega_{\mathrm{ia}} \log(\mathrm{m_{ia}})$ as $\mathrm{Var}(\mathrm{m_{ia}}) \rightarrow 0$. Additionally, $\alpha_{\mathrm{i}} \rightarrow \alpha$ and $\beta_{\mathrm{i}} \rightarrow \beta$ as $\mathrm{Var}(\mathrm{m_{ia}}) \rightarrow 0$. Therefore,

$$\log(\mathrm{m_a}) \rightarrow \sum_i \omega_{\mathrm{ia}} \log(\mathrm{m_{ia}}) \rightarrow \alpha + \beta\mathrm{a}$$

as $\mathrm{Var}(m_{ia}) \to 0$. That is, when the difference in mortality rates across groups is small and each subgroup follows a Gompertz distribution, aggregate mortality is approximately Gompertz. The error in the approximation is proportional to the variance in mortality rates across groups.

To assess the quality of this approximation in our application, we calculated aggregate mortality rates in a population that has two subgroups, each of which constitutes 50% of the population at age 40. The first group has mortality rates corresponding to the empirical estimates for men at the 5th percentile in Figure 1 (α_1 = -8.2, β_1 = 0.07), while the second has mortality rates corresponding to those for men at the 95th percentile (α_2 =-11.9, β_2 = 0.10). With these parameters, the actual aggregate mortality rate has a correlation of 0.999 with a Gompertz prediction based on the aggregate mortality rates. The mean relative difference between the actual and predicted aggregate mortality rates is 2.2%. Thus, even when one considers two groups with very different mortality profiles – men at the 5th and 95th percentiles – it is reasonable to model both the aggregate distribution of mortality rates and the subgroup-specific mortality rates as Gompertz. These calculations support our approach of treating the mortality distributions at all levels of aggregation as Gompertz. Intuitively, the variation in mortality rates across income and racial groups is small relative to the variation across ages, making the Gompertz model a good approximation at all levels of aggregation.

C. Estimation of Gompertz Parameters

We estimate the Gompertz parameters α and β for each income percentile, gender, and year using maximum likelihood estimation (MLE). Under the Gompertz model, the number of deaths at each age follows a Binomial distribution with a probability of death given by $e^{\alpha + \beta age}$. Hence, the Gompertz model is equivalent to a generalized linear model (GLM) with a binomial probability distribution and a log link function with a single predictor (age). We estimate this GLM separately for each income percentile, gender, and year using Stata's glm function.

An alternative, computationally simpler method of estimating the Gompertz parameters is to calculate sample mortality rates at each age $\hat{m}(age)$ and estimate α and β using an ordinary least squares (OLS) regression of log mortality rates on age: $\log(\hat{m}(age)) = \alpha + \beta age$. Under the Gompertz model, both MLE and OLS yield consistent estimates of α and β, as the sample estimate of the mortality rate $\hat{m}(age)$ converges to the true mortality rate $m(age)$ as the sample grows large. However, MLE is asymptotically efficient while OLS is not.

At the national level, MLE and OLS yield virtually identical estimates because the samples are very large. However, the two estimators differ at the local level, particularly in smaller CZs and counties. To evaluate the finite-sample properties of the two estimators, we ran Monte Carlo simulations under the assumption that all CZs had the same true parameters as the national estimates. These simulations reveal that OLS generally yields life expectancy estimates with significantly higher bias and variance than MLE, especially in CZs with populations below 500,000. We therefore use MLE as our primary estimator.

D. Race and Ethnicity Adjustments

Black adults have lower life expectancies than white adults controlling for income, while Hispanic and Asian adults have higher life expectancies than white adults in the United States.[4] Raw differences in life expectancy across income groups and areas are therefore partly driven by differences in racial and ethnic composition. We control for racial and ethnic differences by answering the question, "What would life expectancy for each gender in each area and income group be if that group had black, Asian, and Hispanic population shares corresponding to the national averages at age 40?"

Race and ethnicity are not recorded in tax records, so we cannot directly measure the life expectancies of people with different racial or ethnic backgrounds. We therefore draw on data from the U.S. Census and the National Longitudinal Mortality Study (NLMS) to adjust for differences in life expectancy by race in three steps. First, we use data from the NLMS to estimate differences in mortality rates between black, Asian, Hispanic, and all other Americans, controlling for income. Second, we use Census data to estimate the share of black, Asian, and Hispanic individuals in every group for which we observe mortality rates. Third, combining the adjustment factors estimated from the NLMS and the race shares from the Census data with our estimates of aggregate mortality rates from the tax data, we calculate race-specific life expectancies and the mean life expectancy that would prevail in each group if it had nationally representative black, Asian, and Hispanic population shares at age 40. The details of these steps are explained below.

Step 1: Estimation of Racial Differences in Mortality Rates using the NLMS

The first step in our race adjustment procedure is to estimate differences in mortality rates by race and age using data from the NLMS. The NLMS contains information on self-reported race and income as well as subsequent mortality for a nationally representative panel of individuals. We use the PUMS Release 5 File 11, which contains 1.8 million people who were initially surveyed between the 1970s and the early 2000s and followed for 11 years. The public use file weights these records to be representative of the U.S. population on April 1, 1990. A limitation of these data is that many of these cohorts were observed before the period we study using the tax data, which spans 2001-2014.

We restrict the NLMS sample to individuals whose income is measured between ages 38 and 61 and who survive into the second year after their income is observed. This matches the sample restrictions applied to the tax records. Following individuals' mortality for up to 11 years after their incomes are measured, we are able to observe mortality rates up to age 72. We define income as inflation-adjusted family income, and group individuals into income quartiles based on their ranks in the income distribution within the NLMS sample. Because each person's income is only recorded upon entry into the panel, we measure income with a lag that varies between two and eleven years. The results on income lag invariance in the tax data imply that these differences in when income is measured are unlikely to affect the results.

Our methodology for estimating racial differences in mortality is motivated by the fact that mortality rates in each racial group are well approximated by Gompertz relationships. eFigure 5 plots log mortality rates (in five year age bins) vs. age for black, Hispanic, and Asian individuals, as well as members of all other races/ethnicities, who are labeled "white" (the modal group). To control for differences in income across the four groups, we compute mortality rates by income quartile, racial group, and age bin, and report an unweighted average across the four income quartiles within each age bin and racial group. The figure shows that log mortality rates increase linearly with age for all four groups, with a different slope and intercept for each race. Consistent with evidence from the NCHS, black mortality rates are higher than white mortality rates at all ages, while Hispanic and Asian mortality rates are lower at all ages.

Under the Gompertz approximation, the mortality rate for a person of race $r \in \{B, H, A, W\}$ in income quartile y at age a is described by $\log(\mathrm{m}_{ry}(a)) = \alpha_{ry} + \beta_{ry}a$. We further assume that the difference across racial/ethnic groups in log mortality rates does not vary with income – an assumption we validate below. These assumptions allow us to summarize the differences in black, Hispanic, Asian and white mortality rates at all ages and income levels using two parameters: a shift in the Gompertz intercept ($\delta_r = \alpha_{ry} - \alpha_{Wy}$) and the Gompertz slope ($\Delta_r = \beta_{ry} - \beta_{Wy}$) relative to white mortality rates.

We estimate the racial mortality shifters δ_r and Δ_r, separately for each gender, using Maximum Likelihood Estimation with the following Gompertz survival model:

$$
\begin{aligned}
\log\left(m_{ryd}(a)\right) = \; & \sum_{y} (\gamma_y + \Gamma_y \cdot a) \cdot 1\{\text{income quartile} = y\} \\
& + \sum_{d} (\gamma_d + \Gamma_d \cdot a) \cdot 1\{\text{census division} = d\} \\
& + \delta_B \cdot 1\{\text{race} = B\} + \Delta_B \cdot 1\{\text{race} = B\} \cdot a \\
& + \delta_H \cdot 1\{\text{race} = H\} + \Delta_H \cdot 1\{\text{race} = H\} \cdot a \\
& + \delta_A \cdot 1\{\text{race} = A\} + \Delta_A \cdot 1\{\text{race} = A\} \cdot a
\end{aligned}
$$

By controlling for income and census division fixed effects, we isolate the effect of race on mortality after accounting for income disparities between races and geographic differences in racial composition.

The differences between races in the Gompertz slopes and intercepts are approximately constant across income groups and areas. eFigures 6 and 7 plot estimates of the racial mortality shifters δ_r and Δ_r, estimated separately by income quartile (eFigure 6) and Census region (eFigure 7). The solid horizontal line shows the estimates pooling all income quartiles and Census regions. There is no evidence that racial differences in mortality differ substantially by area or income. This result allows us to simplify our analysis significantly by estimating a single set of racial shifters that we use in all areas and for all income groups.

When estimating trends in life expectancy over time, our methodology also assumes that the differences in mortality rates by race (conditional on income) are constant over time. Recent work by Case and Deaton using NCHS data shows that for certain subgroups (in particular, men aged 45-54), the black-white mortality gap narrowed in the 2000s.[5] This result does not directly contradict our assumption of constant racial gaps conditional on income and area, as the differential trends for blacks and whites at the national level could be driven by differences in mortality trends across income groups and areas. In practice, the correlation between the race-unadjusted CZ-level trend estimates and the race-adjusted trend estimates exceeds 0.99 for all income quartile by gender groups in our data because changes in race shares are small in comparison to trends in life expectancy. Hence, we expect that further adjustments for differential trends in mortality rates by race will have little impact on our trend estimates.

Step 2: Estimation of Race Shares using Census Data

The second step in our race adjustment procedure is to estimate the white, black, Hispanic, and Asian shares for each group we use in our analysis. We obtain population counts by race, gender, age (5-year age bins at the county level), year, and county from the 2001-2014 Intercensal and Postcensal estimates.[6,7] These data do not provide breakdowns by income, so we estimate racial income distributions using the 2000 Census long form.

At the national level, we estimate income distributions in percentiles by race, age and gender using the 5% microdata sample of the 2000 Census.[8] Since the exact distribution of population by race, age, gender and income percentile is noisy due to sampling error, we estimate smooth distributions using a Lowess estimator with a kernel bandwidth of 0.2.

At the local level, we estimate income distributions in quartiles by race, county, and 10-year age bin using the 2000 Census SF3 tables.[9] We aggregate the 16 income bins provided in the SF3 tables to quartiles of the national income distribution. Since these data are available at the level of households, not individuals by gender as at the national level, we assume that the national distributions of household income are the same for men and women — an approximation that holds well in the tax data. The set of adults in most households consists of either one woman, one man, or a man and a woman. Therefore, the household-weighted income distribution is very similar to the person-weighted household income distribution for men and women under the assumption that the distributions do not differ by gender.

We use these income distributions combined with the Intercensal and Postcensal population counts by race to estimate counts by race and income group. Since the Census is a cross-sectional survey, we cannot measure income at age 61 for individuals at older ages as we do with mortality rates in the tax data. We therefore apply the distribution of income at age 61 for all subsequent ages. At the national level, we apply the income distributions by race, age and gender to each year of the annual population data by race, age and gender, obtaining corresponding population counts by income percentile. We do the same at the local level by county, obtaining population counts by income quartile (or by ventile, using the same procedure). We aggregate these county-level estimates to get CZ- and state-level data. The local Intercensal and Postcensal population data are available by 5-year age bins and the local income distributions are available by 10-year age bins; at each individual age, we use the population data and income distribution from the corresponding bin.

We then calculate the share of the population in each group that is white, black, Hispanic, or Asian. As with the mortality data, we pool individuals from 2001-2014 when studying levels of life expectancy, and use the individual years of data when studying trends in life expectancy.

Step 3: Constructing Race-Specific Mortality Rates and Life Expectancies

In the final step of our race adjustment procedure, we combine the estimates of the race shifters from the NLMS and the Census data on race shares with our estimates of aggregate mortality rates by age from the tax data to estimate race-specific mortality rates and life expectancies.

Let ω_{rg} denote the fraction of individuals who belong to race r in a given group g (e.g., men in the bottom quartile in the New York CZ). Let $m_g(a)$ denote the average mortality rate at age a in group g, estimated from the tax data using Maximum Likelihood with a Gompertz model as described above. The average mortality rate $m_g(a)$ can be decomposed as

$$\log\left(\mathrm{m}_g(a)\right) \approx \sum_{\mathrm{r}} \omega_{rga} \log\left(\mathrm{m}_{rg}(a)\right)$$

where the approximation follows from the result in the preceding section on the aggregation of Gompertz distributions. Applying the results on racial differences in mortality from the NLMS, the difference in mortality rates between race r and whites at age a is:

$$\log\left(\mathrm{m}_{rg}(a)\right) = \log\left(\mathrm{m}_{Wg}(a)\right) + \delta_r + \Delta_r a$$

Therefore, the average mortality rate at age a is approximately

$$\log\left(\mathrm{m}_g(a)\right) \approx \log\left(\mathrm{m}_{Wg}(a)\right) + \omega_{Bga}(\delta_B + \Delta_B a) + \omega_{Hga}(\delta_H + \Delta_H a) + \omega_{Aga}(\delta_A + \Delta_A a)$$

Using estimates of the race shares ω_{rga} from the Census, the race shifters (δ_r, Δ_r) from the NLMS, and the average mortality rate $\mathrm{m}_g(a)$ from the tax data, we can solve for the mortality rate of white individuals at age a in group g using the equation above. Since the Gompertz aggregation result is an approximation, the estimated white mortality rates are approximately but not exactly Gompertz. We find the Gompertz parameters that best summarize white mortality rates using an OLS regression, $\log\left(\mathrm{m}_{Wg}(a)\right) = \alpha_{Wg} + \beta_{Wg} a$, which penalizes the squared error in the aggregation approximation. We then compute Gompertz parameters for black, Hispanic, and Asian mortality rates by applying the NLMS race shifters, e.g. $\alpha_{Bg} = \alpha_{Wg} + \delta_B$.

Using the Gompertz parameters for each race, we calculate life expectancy at age 40 (L_{rg}) following the procedure described above. Finally, we compute mean race-adjusted life expectancy (L_g) as the average life expectancy across racial groups using national race shares $\omega_{r,40}$ at age 40 (which are 13% black, 11% Hispanic, and 4% Asian for women; 12% black, 12% Hispanic, and 4% Asian for men):

$$L_g = \omega_{W,40} L_{Wg} + \omega_{B,40} L_{Bg} + \omega_{H,40} L_{Hg} + \omega_{A,40} L_{Ag}.$$

E. Standard Errors

There are two primary sources of sampling error in our estimates: sampling error in the Gompertz parameters estimated using the tax data and sampling error in the Gompertz race adjustment shifters estimated using the NLMS data. We account for these two sources of uncertainty by drawing bootstrap samples from each dataset as described below.

To generate bootstrap estimates of the Gompertz parameters from the tax data, we draw new Gompertz parameters from a multivariate normal distribution with the means and covariance matrix obtained directly from the maximum likelihood estimation procedure. We use this parametric bootstrap procedure rather than directly resampling the 1.4 billion observations in the tax data at the individual level for computational simplicity; comparisons of the parametric approach to the individual-level resampling in selected CZs shows that the two approaches yield very similar results.

To generate bootstrap estimates of the race adjustment shifters, we directly resample the NLMS microdata at the individual level with replacement and replicate our baseline procedure to obtain a new estimate of the race shifters.

We pair each bootstrap estimate of the Gompertz parameters from the tax data with a bootstrap estimate of the race shifters from a bootstrap sample of the NLMS data to compute a bootstrap estimate of race-adjusted life expectancy following our standard methodology. We then compute the bootstrap standard error of life expectancy as the standard deviation of life expectancy across 1,000 bootstrap samples. We compute the bootstrap standard error of the time trends by state and CZ in the same manner, estimating the trend in each bootstrap sample and then taking the standard deviation across samples to calculate the standard error. We estimate bootstrapped 95% confidence intervals using the basic bootstrap formula

$$(2\theta - \theta^*_{0.975}; 2\theta - \theta^*_{0.025})$$

where θ is the point estimate, and θ^* is the bootstrap 2.5$^{\text{th}}$ or 97.5$^{\text{th}}$ percentile.

Part III: Construction of Local Area Characteristics

eTable 3 gives definitions of all of the variables used in the correlational analysis in Figures 8 and 9 as well as the sources used to construct them. Here we present further detail on the construction of certain variables used in the analysis, particularly indices that combine multiple measures.

We proxy for the quality of inpatient care using risk-adjusted 30-day hospital mortality for acute myocardial infarction, congestive heart failure, and pneumonia patients admitted to a hospital between 2002 to 2013, taken from Joynt et al.[10] We form a single quality index as the average of the z-statistics for the three conditions.[11]

The quality of primary and preventive care was proxied by the mean of several variables drawn from 2010 Medicare claims: the share of people with at least one visit to a primary care physician in a year; the share of diabetic patients who received an annual Hemoglobin A1c test, retinal exam, and LDL screen; the share of women aged 67-69 who received a mammogram in a year; and the hospitalization rate for ambulatory care sensitive admissions per 1,000 Medicare beneficiaries. Each measure is defined so that higher numbers are good, and the index was formed as the mean of the z-statistics for each variable.

Inequality was measured using a Gini index, constructed using tax records by Chetty et al. (2014). For comparability with prior work we excluded individuals in the top 1 percent of the income distribution, whose incomes are typically top-coded in survey-based measures of inequality, when computing the Gini index.

Part IV: Sensitivity Analysis of Life Expectancy Estimates

A. Sensitivity of National Life Expectancy Estimates to Alternative Assumptions

Our baseline estimates of life expectancy by income percentile (shown in Figure 2) include race and ethnicity adjustments, use Gompertz extrapolations to age 90, and assign individuals income percentiles based on household income. We assess sensitivity to the estimates to each of these choices in eFigure 9.

The first panel of the figure compares life expectancy with and without race and ethnicity adjustments. The two series are virtually identical, for two reasons. First, differences in life expectancy by race are small relative to differences by income. Controlling for differences in income, our estimates imply that life expectancy at age 40 for black men and women is 1.7 and 1.5 years lower than white men and women, respectively. Life expectancies for Hispanic men and women are 4.2 and 3.7 years higher than for white men and women, after controlling for income. Given that approximately 10% of the U.S. population is black and 10% is Hispanic, race and ethnicity adjustments are on the order of approximately 0.1 to 0.4 years. These adjustments are small relative to the 10-15 year gap in life expectancy between the richest and poorest Americans. Second, black and Hispanic individuals are approximately the same share of the overall population, they share similar income distributions, and the black mortality disadvantage is similar to the Hispanic mortality advantage. Thus, the two adjustments roughly cancel out at the national level. At the local level, race and ethnicity adjustments play a larger role, as race and ethnicity shares vary substantially across areas.

The second panel of the figure assesses the implications of extending the Gompertz extrapolation to age 100 instead of age 90. The greatest change is at high income levels, where life expectancy would be even greater extending the Gompertz assumption to later ages. This is because mortality rates for high income groups are lower than mortality rates for other income groups even at age 90. Thus, allowing the Gompertz extrapolation to continue to age 100 instead of switching to uniform mortality rates includes more years of lower mortality for that group.

The third panel of the figure uses individual earnings (excluding spousal income) instead of household earnings to define each individual's income percentile. Results are very similar using the two income measures.

B. Sensitivity of National Trend Estimates to Controlling for Differential Income Growth

Because we measure income by percentile ranks within each year rather than income levels, the trends in life expectancy by quartile could be driven in part by changes in the income level associated with each percentile. In particular, increasing income inequality implies that the incomes of the rich increased relative to the incomes of the poor between 2001-2014, which might be related to the larger growth in life expectancy at higher income percentiles.

To control for the effects of differential changes in income levels, we regress race-adjusted life expectancy in each gender-percentile-year cell on year and the mean income level at age 40 (measured in 2012 dollars) associated with that gender by percentile by year observation. We estimate this regression separately for each income quartile and gender to obtain estimates of the trend in life expectancy within each quartile, netting out the change due to trends in income levels.

The resulting estimates are reported in eTable 4. These estimates are very similar to our baseline estimates in Figure 3; in particular, we continue to find much larger gains in life expectancy in upper income quartiles. Hence, relatively little of the divergence in life expectancy across income groups can be attributed to the growth in income inequality between 2001 and 2014.

C. Sensitivity of Local Life Expectancy Estimates to Alternative Assumptions

In eTable 6, we assess the sensitivity of our CZ-level estimates of life expectancy to alternative assumptions. Each column of the table shows the population-weighted correlation between our baseline estimates of life expectancy by CZ and an alternative measure, by income quartile and gender.

The first alternative uses estimates of life expectancy not adjusting for differences in race and ethnicity across areas. The correlations between the race-adjusted and unadjusted estimates across CZs ranges from 0.71-0.94 for men and 0.40-0.84 for women. These correlations show that the majority of the variance in unadjusted life expectancy across areas is not due to variation in racial and ethnic composition; however, accounting for race and ethnicity does have a significant impact on the spatial patterns.

In the second column, we eliminate the use of extrapolations of mortality rates beyond observed ages by measuring the expected number of life years up to age 77. We implement this by estimating life expectancy at age 40 assuming a 100% mortality rate at age 77. Life years up to age 77 are very highly correlated with our baseline estimates, showing that the Gompertz extrapolations play a limited role in generating the geographical variation we document.

In the third column, we use Gompertz extrapolations to age 100 instead of age 90. This change has also little impact on the estimates, with correlations above 0.98 in all cases.

In the fourth column, we evaluate the sensitivity of our estimates to using OLS to estimate the Gompertz parameters instead of maximum likelihood. The two estimators yield very similar results overall, with correlations above 0.95.

The fifth specification assesses the sensitivity of our quartile-specific estimates to differences in income distributions *within* each quartile across areas. For instance, higher income areas will generally have fewer people in the lowest percentiles of the income distribution, and thus mean life expectancy in the bottom quartile may be higher in such areas simply because average income for those in the bottom quartile is higher. To control for these within-quartile differences, we first calculate *unweighted* means of mortality rates across the 25 percentiles within each quartile in the working-age sample (up to age 63) and means weighted to account for differences in survival probabilities by income percentile in the retired sample (above age 63). We then estimate race- and ethnicity-adjusted life expectancy using these mean mortality rates (which put uniform weight on all percentiles within each quartile) using OLS. We use OLS rather than MLE here because the unweighted average of the mortality rates in each quartile no longer follows a Binomial distribution. We are able to construct these income-controlled estimates only for the 207 CZs in which there are at least 10 observations in every income percentile by gender and age, in order to estimate mortality rates for every income percentile separately. In this subsample of CZs, the correlation between the income-controlled estimates and our baseline estimates exceeds 0.89 in all subgroups.

Finally, in the last column, we assess the sensitivity of our results to differences in costs of living across areas. We measure local costs of living using a CZ-level price index constructed by Chetty et al. using information from the ACCRA price survey and local house prices.[12] We use this price index to define a measure of "real income" for each household, rank households in the national income distribution based on real income, and replicate our baseline analysis using these measures. This procedure effectively increases the income percentile of individuals living in low cost-of-living CZs, such as rural areas. Using real vs. nominal income to define income quartiles generally has little effect on the results, with correlations of approximately 0.9 with the baseline estimates.

Part V: Supplementary Results

A. Life Expectancy vs. Income Levels

eFigure 8 shows the relationship between the level of income and life expectancy. This figure plots the same race- and ethnicity-adjusted life expectancies by percentile shown in Figure 2 vs. the mean level of household income in each percentile bin, instead of the percentile rank number. In levels, the relationship between income and life expectancy is non-linear and concave: at higher income levels, a given dollar increase in income is associated with a smaller increase in life expectancy.

B. Additional CZ-Level Estimates of Life Expectancy

eFigure 10 shows maps of race- and ethnicity-adjusted life expectancy by CZ for the second and third income quartiles, analogous to the maps for the bottom and top quartiles shown in Figure 5. The second quartile looks similar to the first quartile, and the third quartile looks more similar to the fourth quartile.

C. Standard Errors of CZ-Level Estimates of Life Expectancy

In eFigure 11, we present summary statistics to gauge the standard errors of the typical estimate we report. The first panel plots standard errors of the life expectancy estimates at the CZ level for men and women in the bottom and top quartiles of the income distribution, restricting the sample to CZs with populations above 25,000. The key determinant of the standard error across CZs is the sample size, which is directly related to the population of the area. Therefore, we divide CZs into twenty ventiles (5% bins) and plot the mean standard error vs. the average CZ population size in each bin (on a log scale).

As expected, the standard errors vary inversely with the square root of population. The mean (population-weighted) standard error of the CZ by quartile life expectancy estimates is 0.29. In CZs with a population above 100,000 – which account for 95% of the U.S. population – the standard errors on the life expectancy estimates are typically less than 1 year. The standard errors are larger in the top quartile especially in small CZs because there are fewer high income individuals in small rural areas and because mortality rates are much lower for high-income individuals in the age range we observe, leading to greater uncertainty in the life expectancy estimates based on the Gompertz extrapolation.

An alternative way to assess the precision of the life expectancy estimates across CZs is to ask what fraction of variance across CZs is driven by differences in actual life expectancy (signal) vs. estimation error (noise). The noise variance can be estimated as the average (population-weighted) standard error squared. The signal variance is then the total variance of the life expectancy estimates across CZs minus the noise variance. eTable 5 presents the signal standard deviations (i.e., the square root of the signal variance estimates) by income quartile and gender. In the bottom quartile, signal accounts for 97% of the variance in life expectancy across CZs for men and 90% for women. The corresponding figures in the top quartile are 71% for men and 62% for women.

The second panel of eFigure 11 replicates the first panel at the county level, restricting the sample to counties with more than 25,000 people. The mean (population-weighted) standard error of the county-level life expectancy estimates is 0.48 years. In the bottom quartile, signal accounts for 93% of the variance in life expectancy across counties for men and 83% for women. The corresponding figures in the top quartile are 60% for men and 47% for women.

The third panel of eFigure 11 shows estimates of standard errors for the annual trends by CZ, again as a function of CZ population, restricting the sample to the 100 most populated CZs, with populations above 590,000. The mean (population-weighted) standard error of the CZ-level trend estimates is an annual change of 0.08 years. In the bottom quartile, signal accounts for 58% of the variance in trends across CZs for men and 41% for women. The corresponding figures in the top quartile are 22% for men and 36% for women.

Finally, the mean (population-weighted) standard error of the state-level trend estimates mapped in Figure 6 is an annual change of 0.05 years. In the bottom quartile, signal accounts for 79% of the variance in trends across states for men and 70% for women. The corresponding figures in the top quartile are 23% for men and 41% for women. The lower panel of eTable 5 reports estimates of signal standard deviations in trends across states.

D. County-Level Estimates of Life Expectancy

eFigure 12 shows life expectancy for men and women in the bottom quartile of the income distribution by county in the New York and Detroit metro areas. There is considerable spatial variation across nearby counties within metro areas. For example, among counties in the New York area, male life expectancy in the bottom income quartile varies from 75.3 in Carbon County, PA to 80.3 in Kings County, NY. New York exhibits a spatial pattern that is common in many areas: the nearer people live to the city center (Manhattan and Queens), the higher their life expectancy. However, in Detroit, the pattern is the opposite: life expectancy in Wayne County (the county encompassing the city of Detroit) is lower than in surrounding counties.

eTable 7 lists the top 10 and bottom 10 counties in terms of life expectancy among the 100 largest counties by population in the U.S. for the bottom and top quartiles. In the bottom quartile, life expectancy ranges from 74.1 and 80.1 years for men and women in Wayne County, MI to 80.2 and 85.0 years for men and women in Queens County, NY.

E. Local Area Variation in Mortality by Cause of Death

A first step in understanding the geographic variation in life expectancy is to decompose mortality into two components: medical causes (such as heart disease and cancer) and external causes (such as motor vehicle accidents, suicide, and homicide). The death records we use for our primary analysis do not contain information on cause of death. Instead, we aggregated 2004 NCHS data by cause to construct age- and gender-adjusted medical and external death rates by

CZ for individuals between the ages of 40 and 64.[13] The NCHS does not report data for counties with populations below 100,000; these counties are excluded from the CZ averages, and CZs that have no counties with uncensored data are excluded entirely. We constructed separate measures for people with no college education (high school degree or less) and those with some college education or more. We then related these mortality rates to age- and gender-adjusted total mortality rates across CZs, derived from the same NCHS data.

In eTable 10, we regress mortality rates by cause on total mortality rates. For those with no college education, 87.3% of the marginal deaths come from medical causes, while 12.7% come from external causes. Medical causes similarly account for the majority of the variation in deaths for those with college education.

F. Maps of Health Behaviors

As shown in Figure 8, life expectancy for low-income individuals is very highly correlated with rates of smoking, obesity, and exercise. eFigure 13 shows maps for the bottom income quartile of these three risk factors by CZ along with a map of life expectancy as a reference. These figures are constructed using data from the BRFSS for individuals in the bottom quartile of the income distribution (pooling men and women).[14] The three maps highlight slightly different patterns of risky health behaviors, which together closely resemble the map of life expectancy for low-income individuals. Smoking is very common in Nevada and the Kentucky/Ohio/West Virginia area. Obesity is particularly high in the deep South, and exercise is low in both the deep South and industrial Midwest.

G. Correlations with Local Area Characteristics by CZ and County

In Figures 8 and 9, we reported correlations between life expectancy and a selected set of local area characteristics that were relevant for evaluating leading theories or had the strongest empirical correlations with life expectancy. For completeness, we report correlations for all the local area characteristics we examined in eTable 8 (for CZs) and eTable 9 (for counties) by gender. The factors that are most strongly associated with differences in life expectancy across CZs are typically most predictive of differences across counties.

eFigure 14 plots life expectancy in each income quartile vs. the Gini index of inequality in the CZ. To create this plot, we first group CZs into ventiles based on their Gini index, then for each income quartile, we plot the mean life expectancy against the mean Gini coefficient taken across CZs within each Gini ventile. There is little or no relationship between inequality and life expectancy for individuals in the bottom quartile of the income distribution. Inequality is more negatively correlated with life expectancy for higher income individuals.

H. Correlations with Trends by CZ

eFigure 15 reports correlations of the same set of covariates used in Figure 8 with the trends in life expectancy across CZs for individuals in the bottom quartile of the income distribution. Since we estimate trends only for the 100 most populated CZs (with populations above 590,000), these correlations are based on that subset of CZs. Mean *levels* of smoking, obesity, and exercise (between 1996-2008) are not highly correlated with *changes* in life expectancy, as one might expect. Most of the other correlations are also relatively small and insignificant, except for the covariates in the final group. Areas with higher population density and more college graduates had larger increases in life expectancy for low-income individuals.

I. Comparisons to Mean Life Expectancies in Other Countries

eFigure 16 provides international benchmarks for our life expectancy estimates. It shows mean life expectancy at age 40 by country in 2013 based on World Health Organization statistics in comparison to U.S. life expectancies at the 1st, 25th, 50th, and 100th percentiles (among individuals with positive income) based on the data shown in Figure 2.

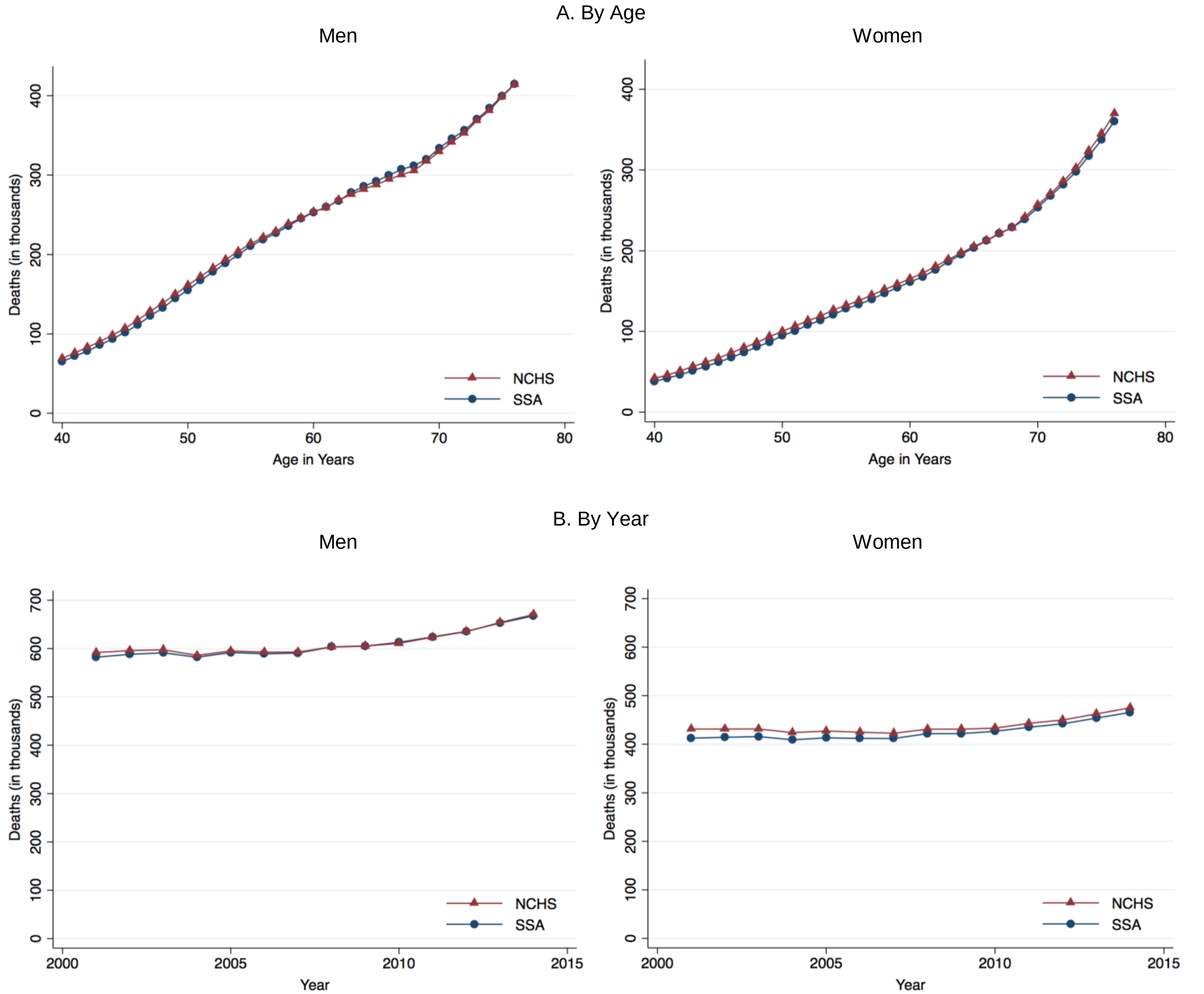
J. Maps of Life Expectancy Pooling All Income Groups

eFigure 17 presents maps of race- and ethnicity-adjusted expected age at death, pooling all income groups (i.e., not controlling for income). For comparability with prior work, we include individuals with zero income in these maps, with the caveat that mortality rates for those with zero income are underestimated in our sample for the reasons discussed in the first section of this Appendix. Life expectancy is lower in the South in these maps relative to the maps in Figure 5 (which show differences in life expectancy conditional on income), especially in the lowest-income states such as Arkansas and Mississippi.

List of Supplementary Figures

List of Supplementary Tables

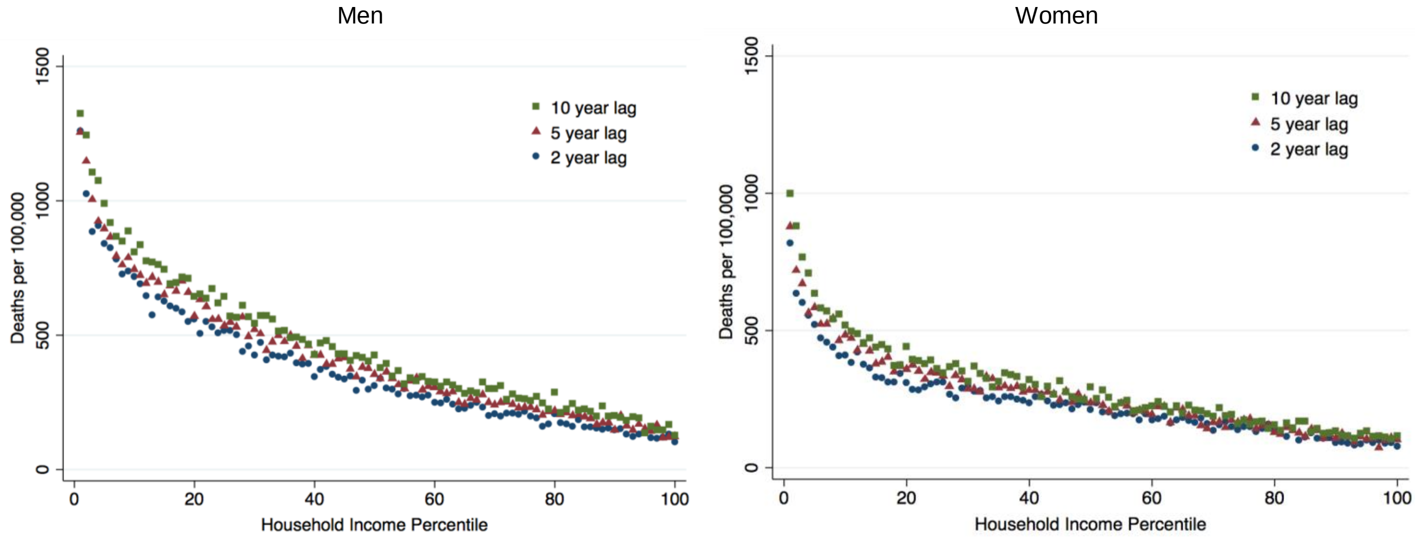
eFigure 1. Comparison of Number of Deaths in NCHS and SSA Data, 2001-2014



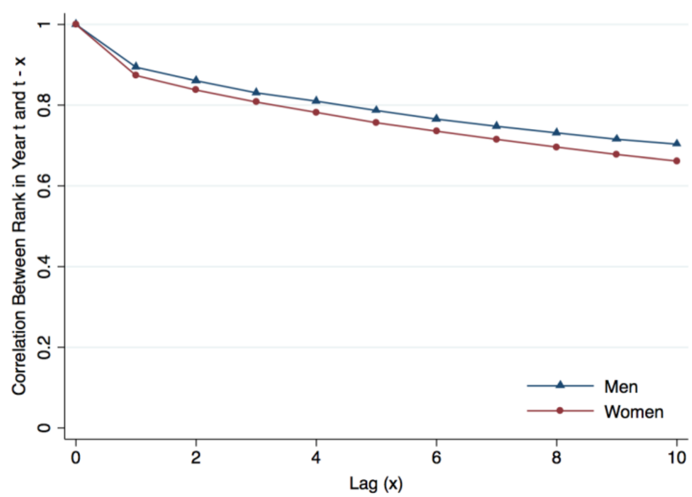
Panels show total count of deaths reported in CDC NCHS data and SSA data, for individuals between the ages of 40 and 76 over the years 2001-2014. Top panel shows total counts by age; bottom panel shows total counts by year. Left column shows estimates for men, right column for women.

eFigure 2. Sensitivity of Mortality-Income Relationship to Year of Income Measurement, 2001-2014

A. Annual Mortality Rates vs. Household Income Percentile by Gender, Ages 50-54, 2014



B. Correlation Between Current and Lagged Income, 2001-2014



Top panel plots mortality rates vs income percentile for men and women 50-54 years of age in 2014. Sample consists of individuals with non-zero income at the time of income measurement. Mortality rates are reported as the mean number of deaths per 100,000 individuals in each percentile of the national household income distribution. Bottom panel shows correlation of household income percentiles in years *t* and *t*-x by gender. Estimates in bottom panel are based on a random sample of individuals between the ages of 40 and 61 with non-zero household income.

eFigure 3. Gompertz Approximation to Mortality Rates in NCHS Data, 2001

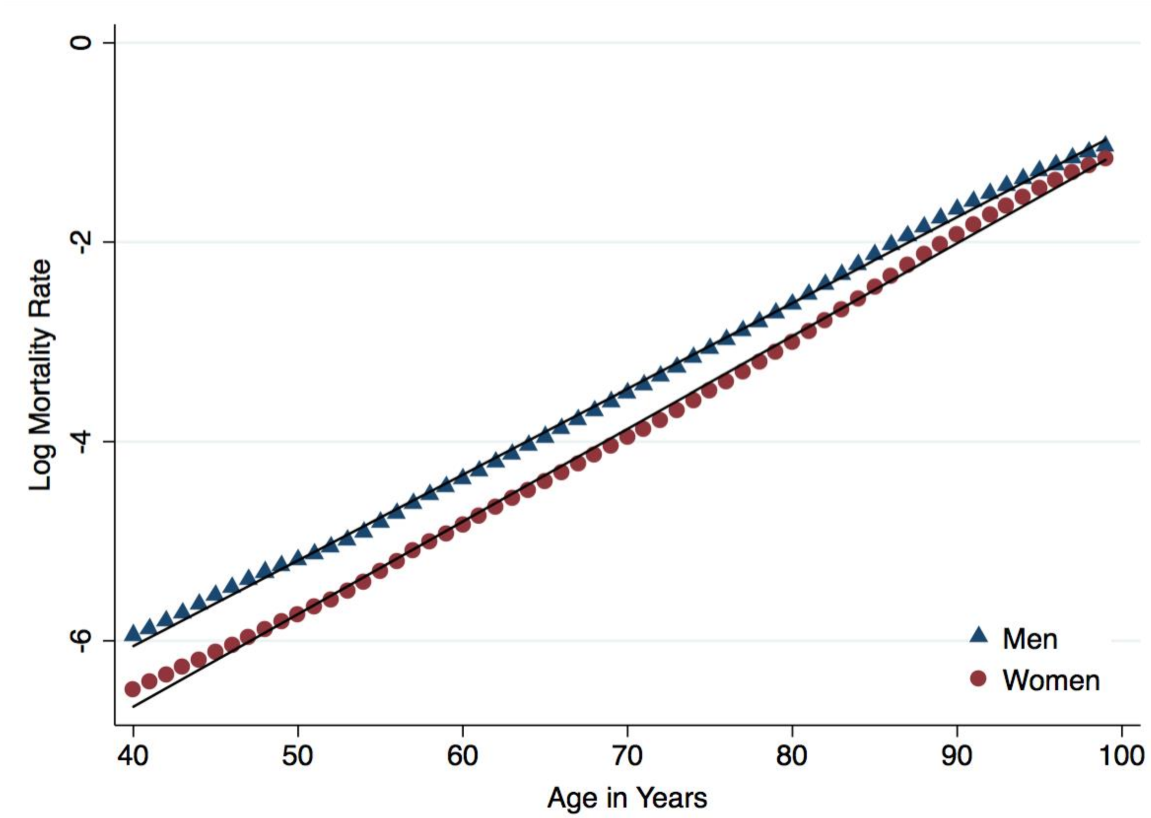
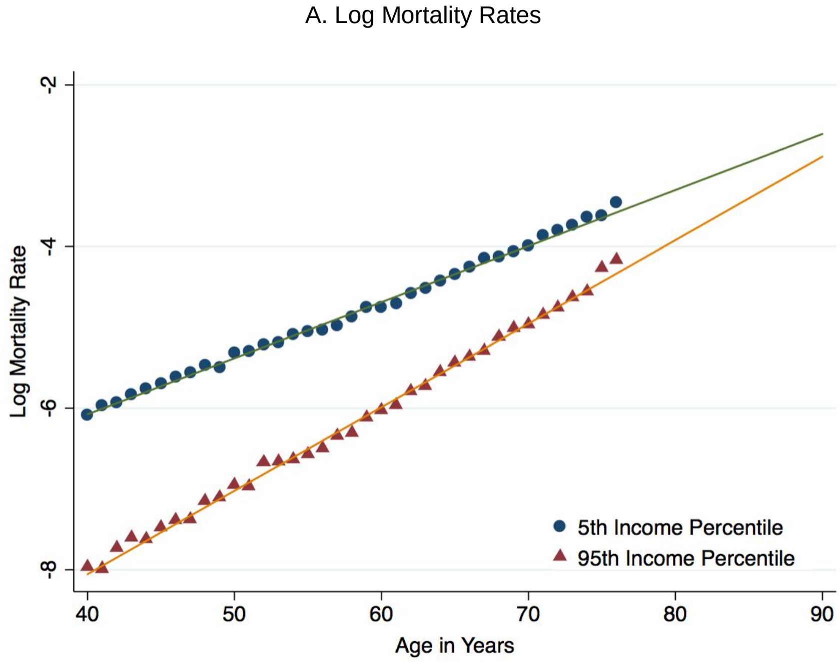


Figure shows log mortality rates from 2001 CDC NCHS life tables, and best fit lines by gender. The R^2 of the best fit line is 0.999 for men and 0.998 for women.

eFigure 4. Gompertz Approximations and Empirical Survival Curves for Women at 5th and 95th Income Percentiles, 2001-2014



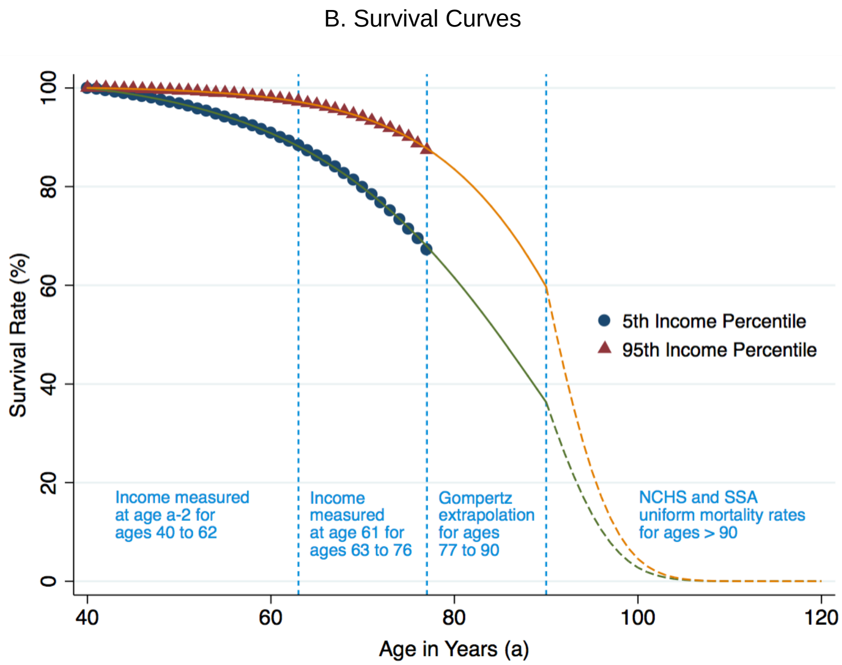
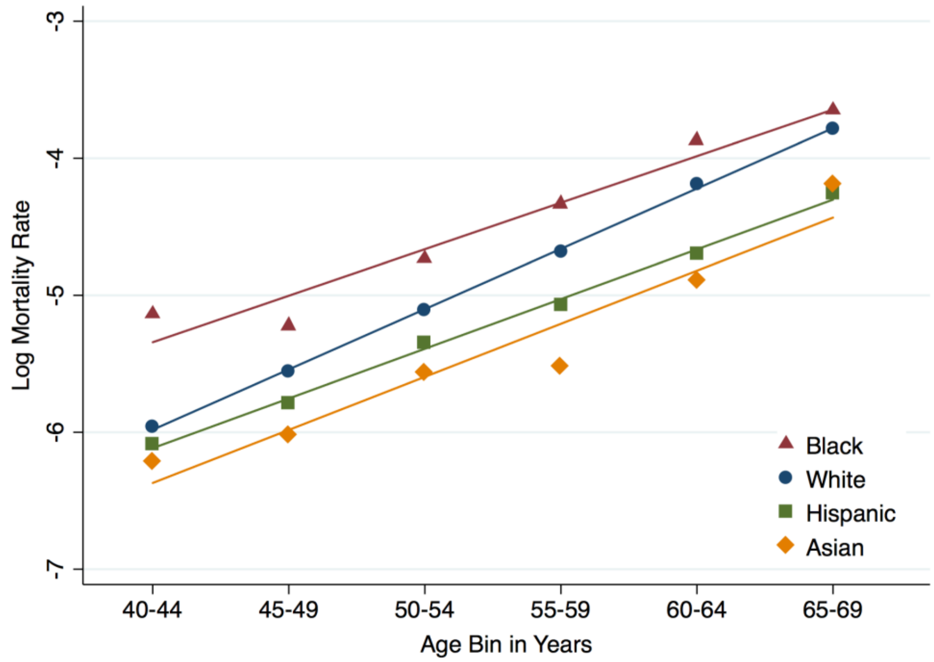


Figure replicates Figure 1 in the text for women instead of men. The mean household income for women in the 5th percentile in our sample is $7,089, and the mean income for women in the 95th percentile is $217,631.

eFigure 5. Log Mortality Rates vs. Age by Race and Ethnicity in NLMS Data, 1973-2011

A. Men



B. Women

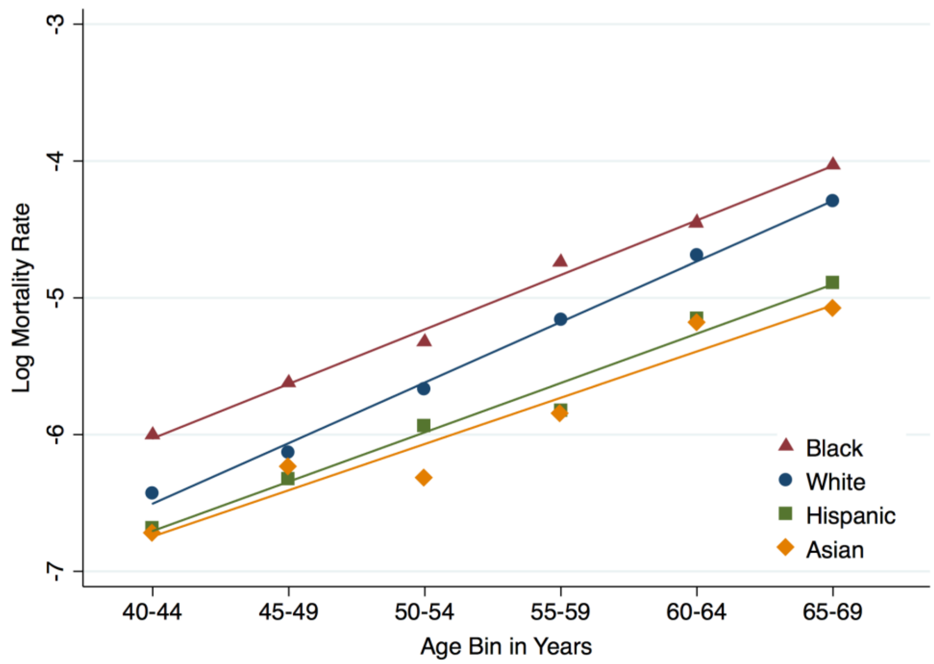
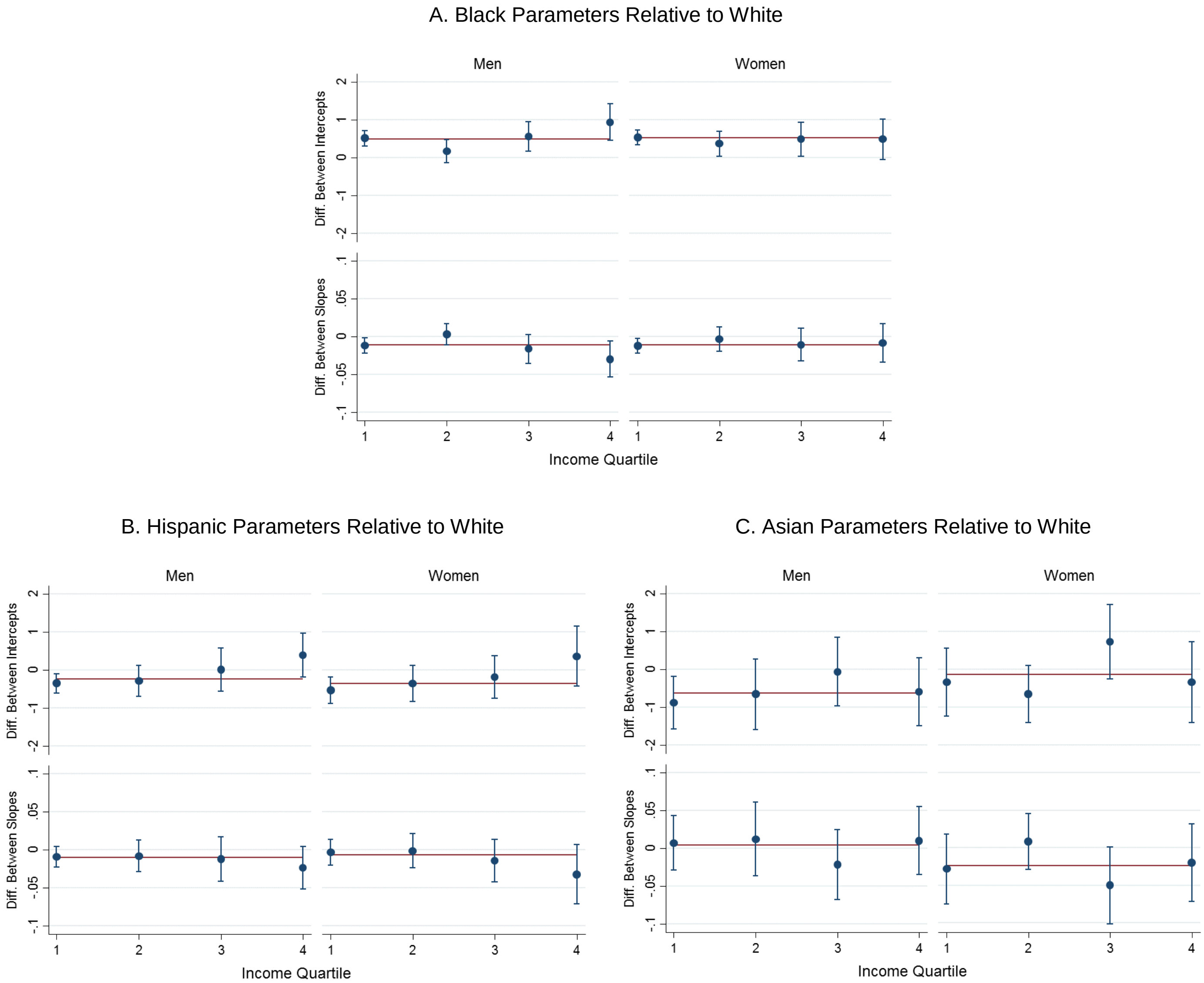


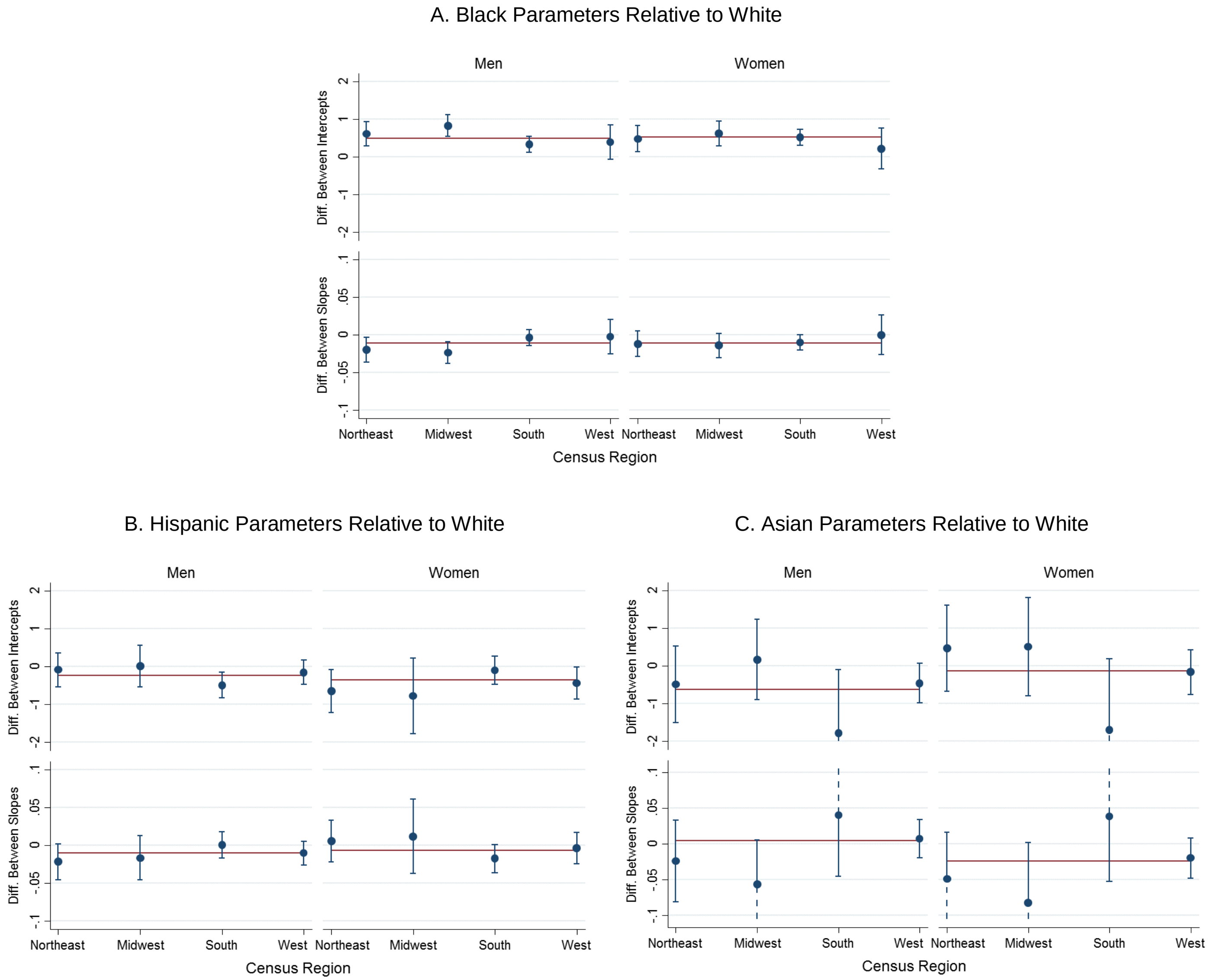
Figure shows log mortality rates by age bin in NLMS data for black, Hispanic, and Asian individuals, as well as members of all other races/ethnicities, who are labeled "white" (the modal group). To control for differences in income across the groups, we first compute mortality rates by income quartile, racial group, and age bin, and then plot an unweighted average across the four income quartiles within each age bin and racial group. The NLMS sample is constructed from 39 cohorts over 1973-2011, re-weighted to reflect the US non-institutionalized population on April 1, 1990.

eFigure 6. Gompertz Parameter Estimates by Income Quartile in NLMS Data, 1973-2011



Points show estimated differences in Gompertz intercepts (at age 40) and slopes for black, Hispanic, and Asian individuals relative to members of all other races/ethnicities, who are labeled "white" (the modal group). Parameter differences were estimated separately by income quartile using NLMS data. Vertical bars show 95% confidence intervals. Horizontal lines show estimates pooling all income quartiles. The NLMS sample is constructed from 39 cohorts over 1973-2011, re-weighted to reflect the US non-institutionalized population on April 1, 1990.

eFigure 7. Gompertz Parameter Estimates by Census Region in NLMS Data, 1973-2011



Points show estimated differences in Gompertz intercepts (at age 40) and slopes for black, Hispanic, and Asian individuals relative to members of all other races/ethnicities, who are labeled “white” (the modal group). Parameter differences were estimated separately by census region using NLMS data. Vertical bars show 95% confidence intervals. Horizontal lines show estimates pooling all income quartiles. Dashed vertical bars are used when the 95% confidence interval extends beyond the vertical axes, in order to preserve scaling across panels; all confidence intervals depicted are symmetric around the point estimate. The NLMS sample is constructed from 39 cohorts over 1973-2011, re-weighted to reflect the US non-institutionalized population on April 1, 1990.

eFigure 8. Race- and Ethnicity-Adjusted Life Expectancy vs. Income in Dollars, 2001-2014

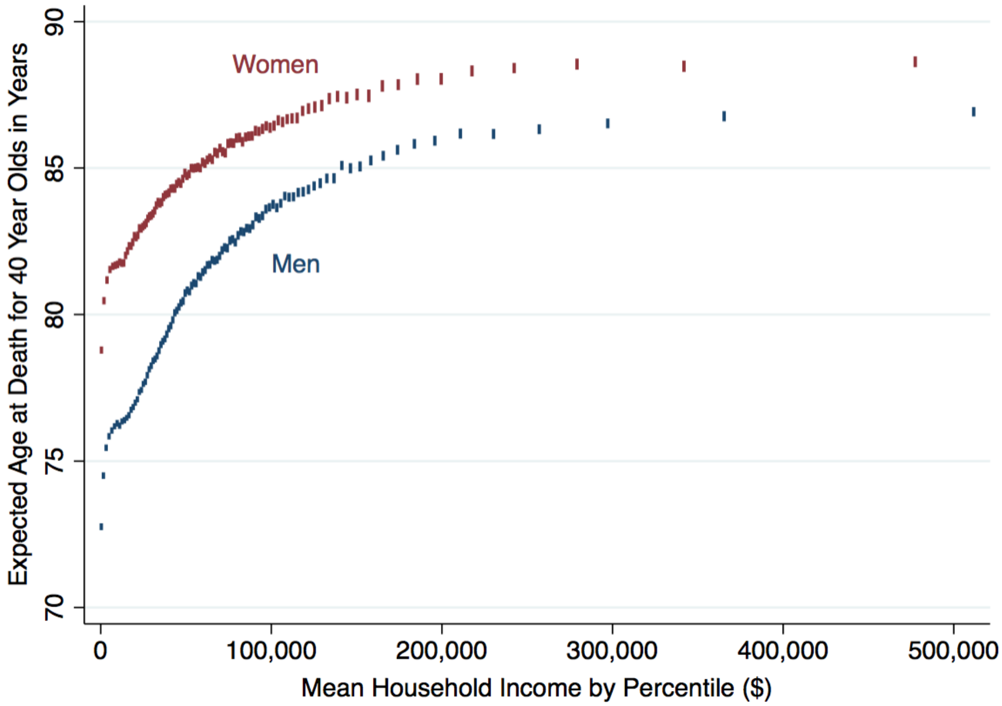
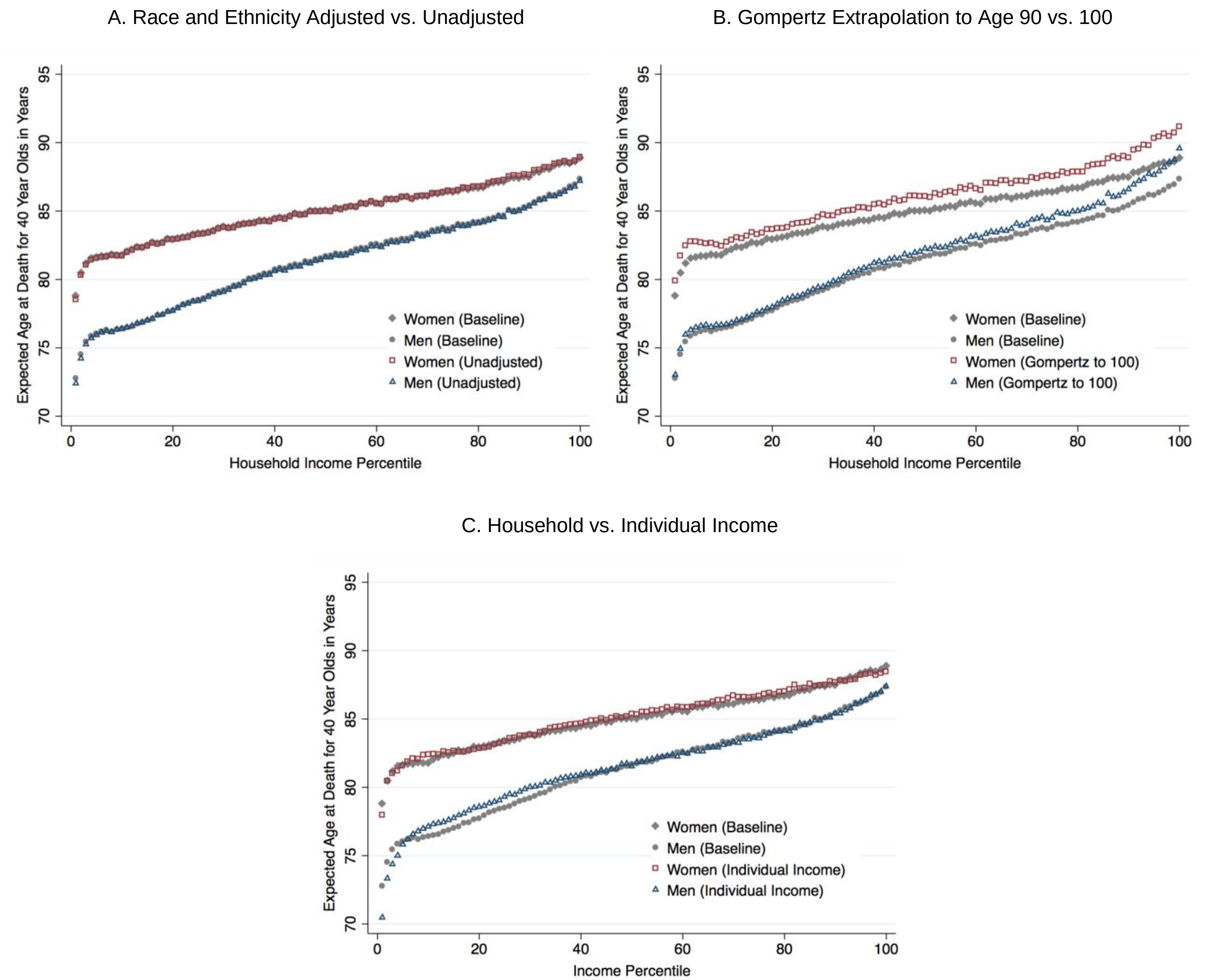


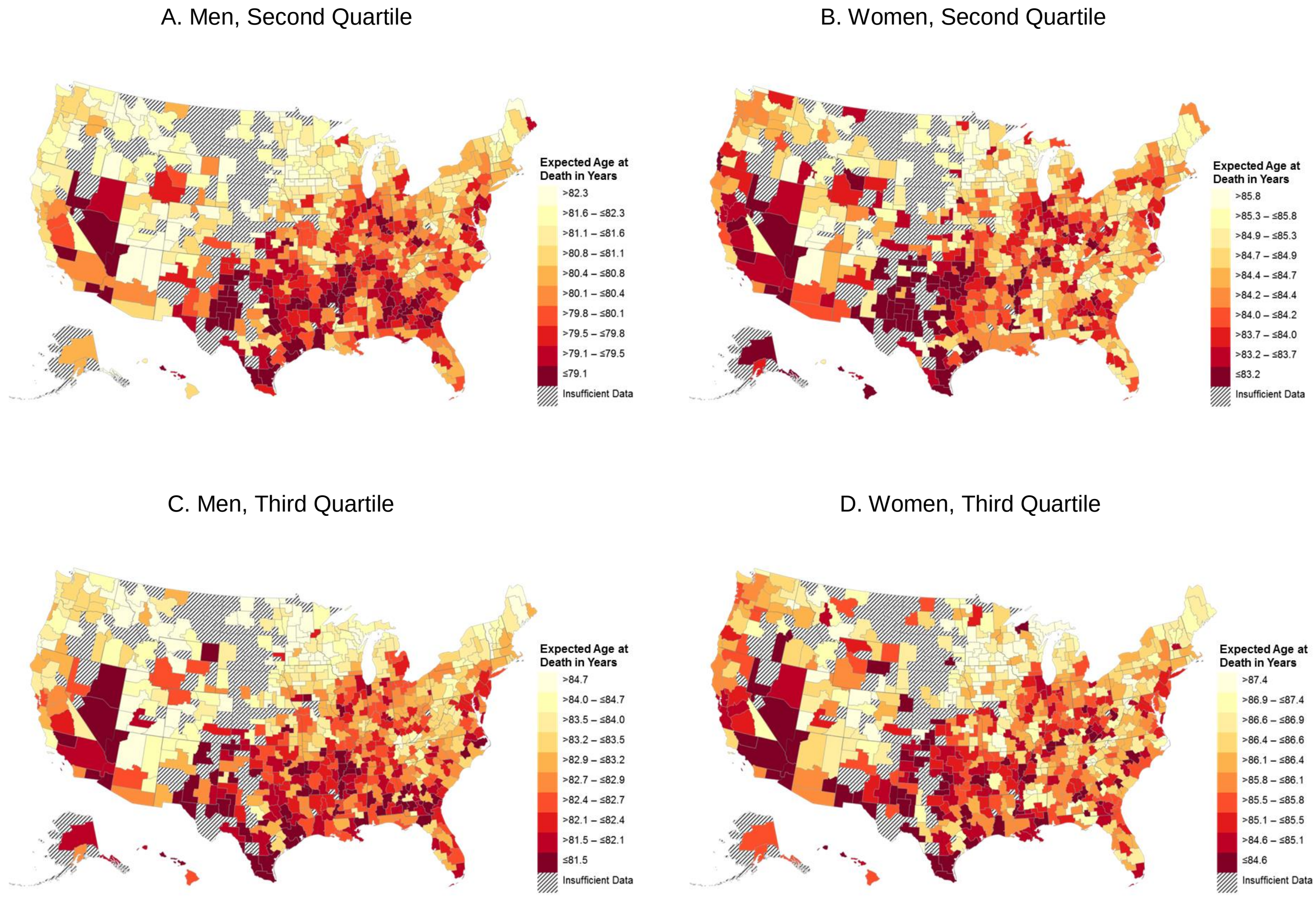
Figure plots race- and ethnicity-adjusted life expectancies by percentile as in Figure 2 vs. the mean level of household income (measured at ages 38-61) in each percentile bin. The vertical height of each bar depicts the 95% confidence interval. The top income percentile is omitted for scaling purposes. The mean household income for men in the top percentile is $1.98 million and their expected age at death is 87.3 years. The mean household income for women in the top percentile is $1.92 million and their expected age at death is 88.9 years. Men and women with household incomes of $25,000 are in the 19th and 21st income percentiles, respectively. Men and women are in percentiles 40 and 45 at $50,000 in household income, percentiles 59 and 62 with $75,000 in household income, percentiles 73 and 75 with $100,000 in household income, and percentiles 93 and 94 with $200,000 in household income, respectively.

eFigure 9. National Life Expectancy Estimates: Sensitivity Analysis, 2001-2014



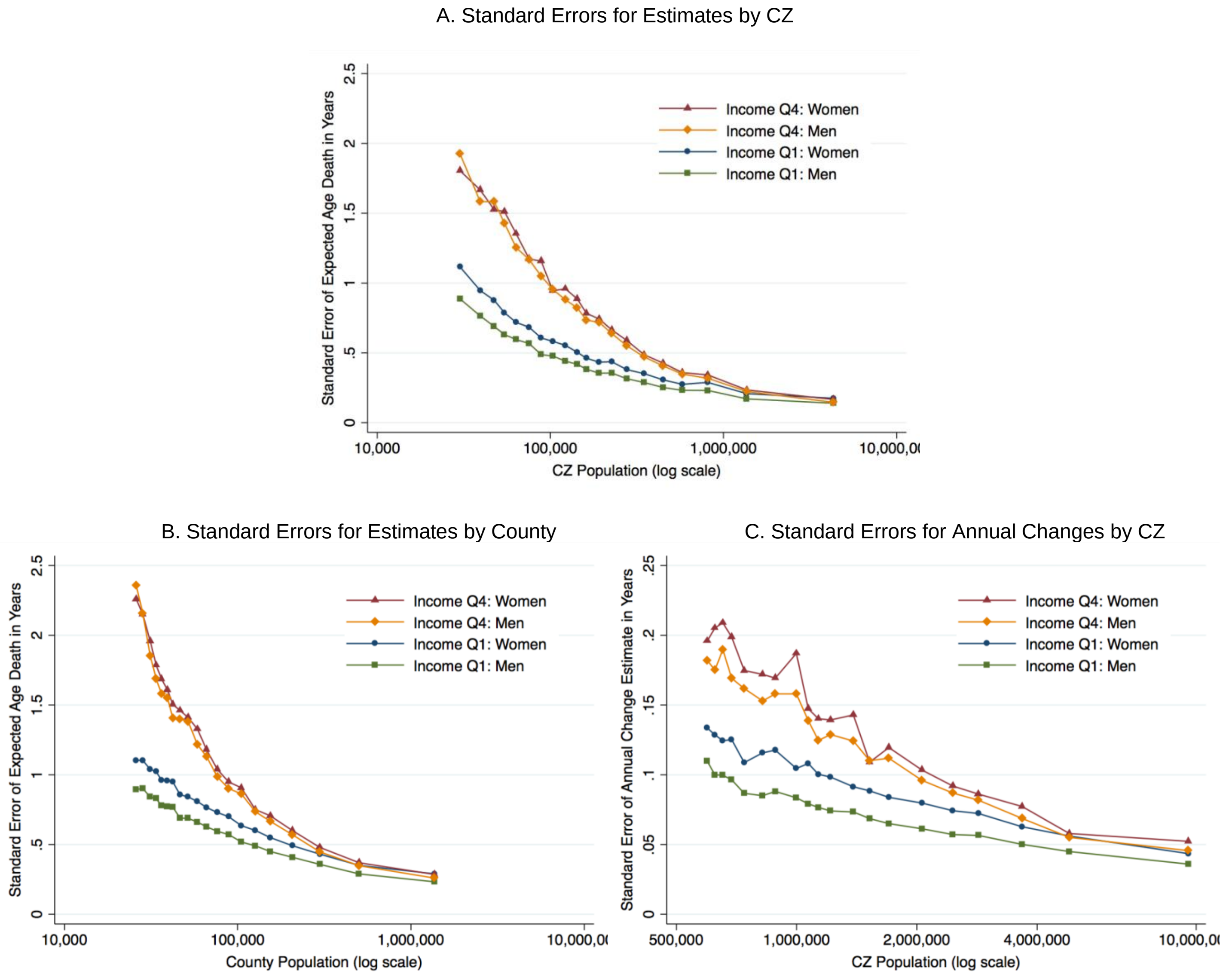
Alternative estimates of expected age at death for 40 year olds by income percentile and gender. Estimates using baseline specification (shown in Figure 2) are plotted in solid gray in all panels. The first panel compares unadjusted estimates with the baseline race- and ethnicity-adjusted estimates. The second panel compares estimates that use Gompertz extrapolations to age 100 before switching to uniform mortality rates, instead of Gompertz extrapolations to age 90, when estimating expected age at death. The third panel compares estimates based on classifying individuals into percentiles based on individual (own) earnings rather than household earnings.

eFigure 10. Race- and Ethnicity-Adjusted Life Expectancy by Commuting Zone for Second and Third Income Quartiles, 2001-2014



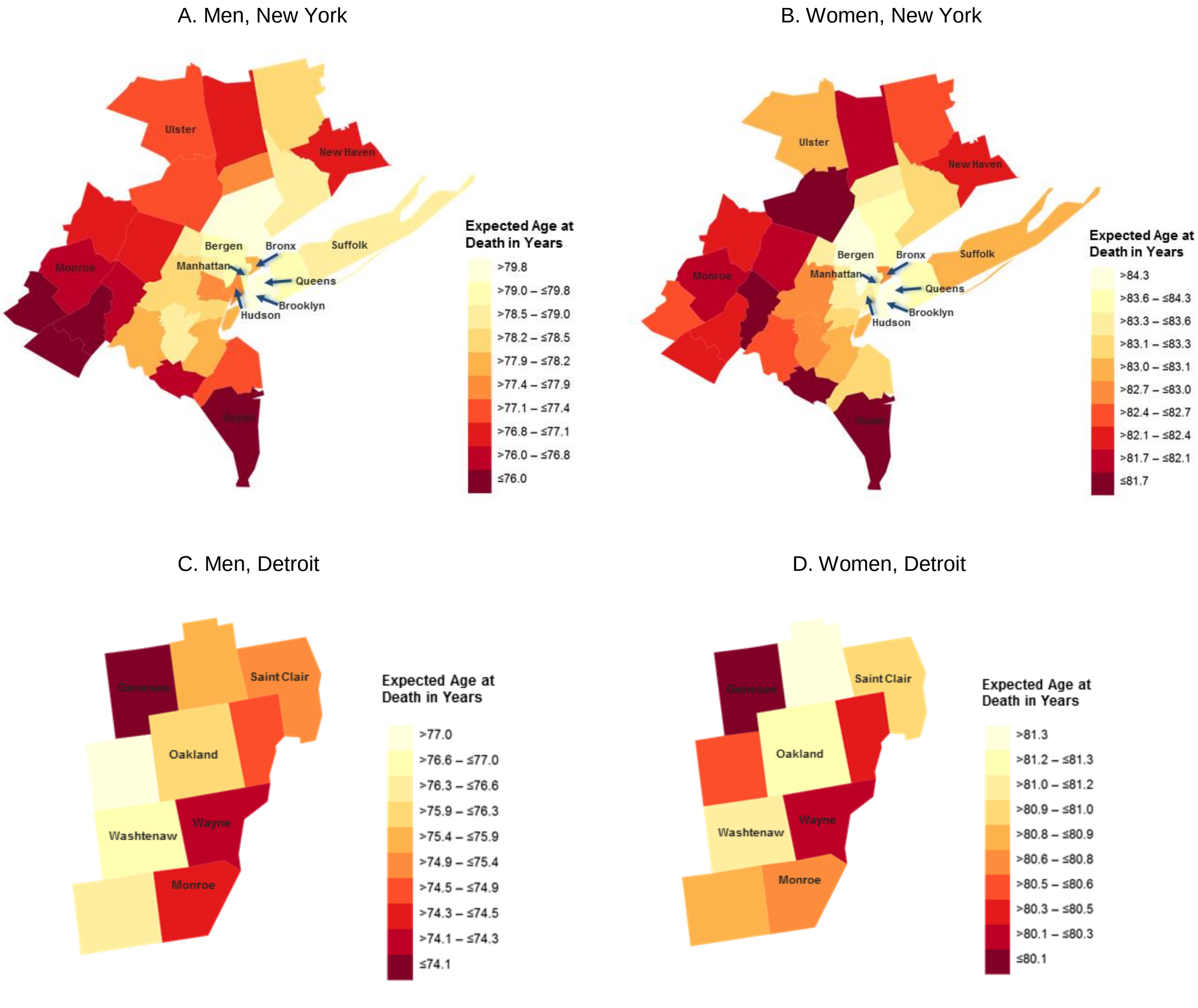
Estimates of race/ethnicity adjusted expected age at death for 40 year olds, computed at the commuting zone level. Commuting zones are grouped into deciles and colored from dark to light as expected age at death increases. Left column shows estimates for men, right column for women. Second and third quartiles are displayed, and top and bottom quartiles are shown in Figure 5. 595 CZs with populations above 25,000 are depicted.

eFigure 11. Standard Errors of Life Expectancy Estimates, 2001-2014



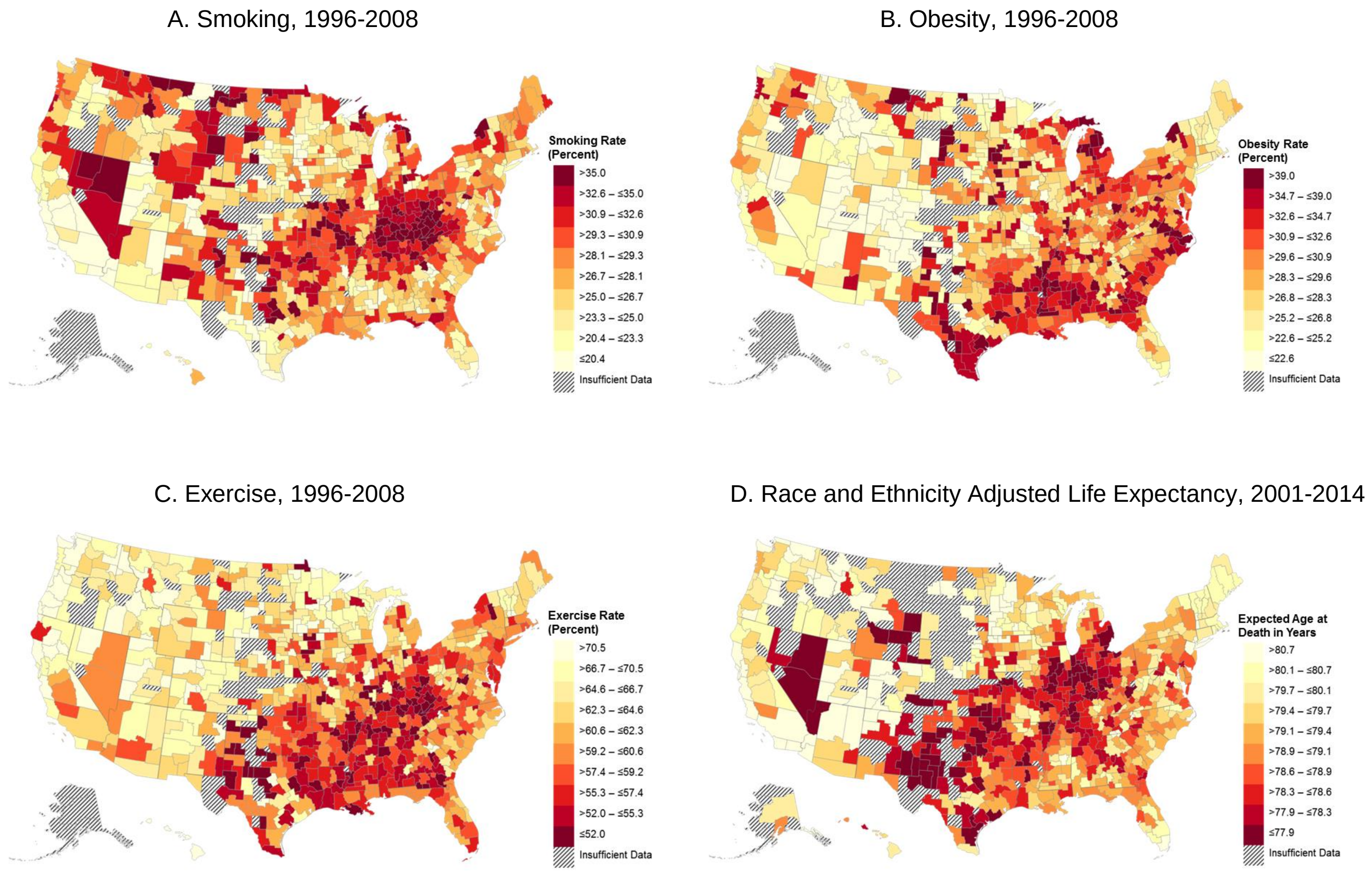
Panel A shows a binned scatter plot of the average standard error of race- and ethnicity-adjusted expected age at death estimates at the CZ level vs. CZ population size. This figure is constructed by dividing CZs into twenty population ventiles (5% bins) and plotting the mean standard error vs. the mean population size in each bin, using a log scale for the x axis. Panel B replicates Panel A at the county level. Panel C shows standard errors for estimates of annual trends in life expectancy by CZ, for the 100 most populated CZs, with populations above 590,000. Standard errors are calculated based on 1,000 bootstrap samples of tax data and NLMS data as described in part II.E of the Supplemental Appendix.

eFigure 12. Race- and Ethnicity-Adjusted Life Expectancy by County in the New York and Detroit Combined Statistical Areas, 2001-2014



Estimates of race-adjusted expected age at death for 40 year olds in the bottom income quartile, computed at the county level. Maps show estimates for counties in the New York and Detroit Combined Statistical Areas. Counties are grouped into deciles and colored from dark to light as expected age at death increases. Left column shows estimates for men, right column for women.

eFigure 13. Rates of Smoking, Obesity, Exercise, and Life Expectancy by CZ for Individuals in Bottom Quartile



Panels A-C show the percentage of adults in the bottom income quartile who currently smoke, have a BMI above 30, or exercised in the last 30 days by CZ. The last panel shows an average of the life expectancy estimates for men and women in the bottom quartile, plotted separately by gender in Figure 5. Commuting zones are grouped into deciles, and lighter colors represent areas with lower smoking rates, lower obesity rates, higher rates of exercise, and higher life expectancy. Data for 652 CZs in the first three panels are obtained from the BRFSS. Data for 595 CZs with population above 25,000 are shown in Panel D.

eFigure 14. Correlations between Life Expectancy and Inequality by Income Quartile, 2001-2014

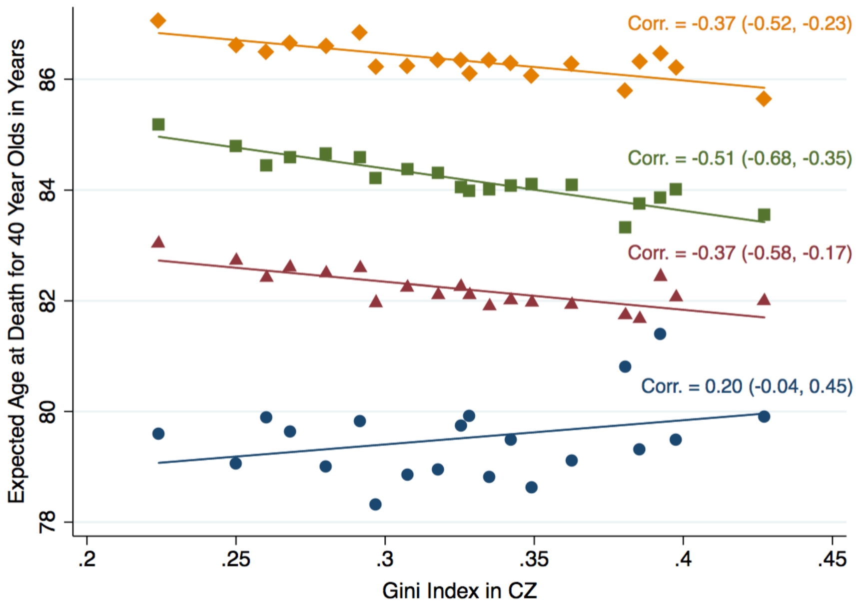
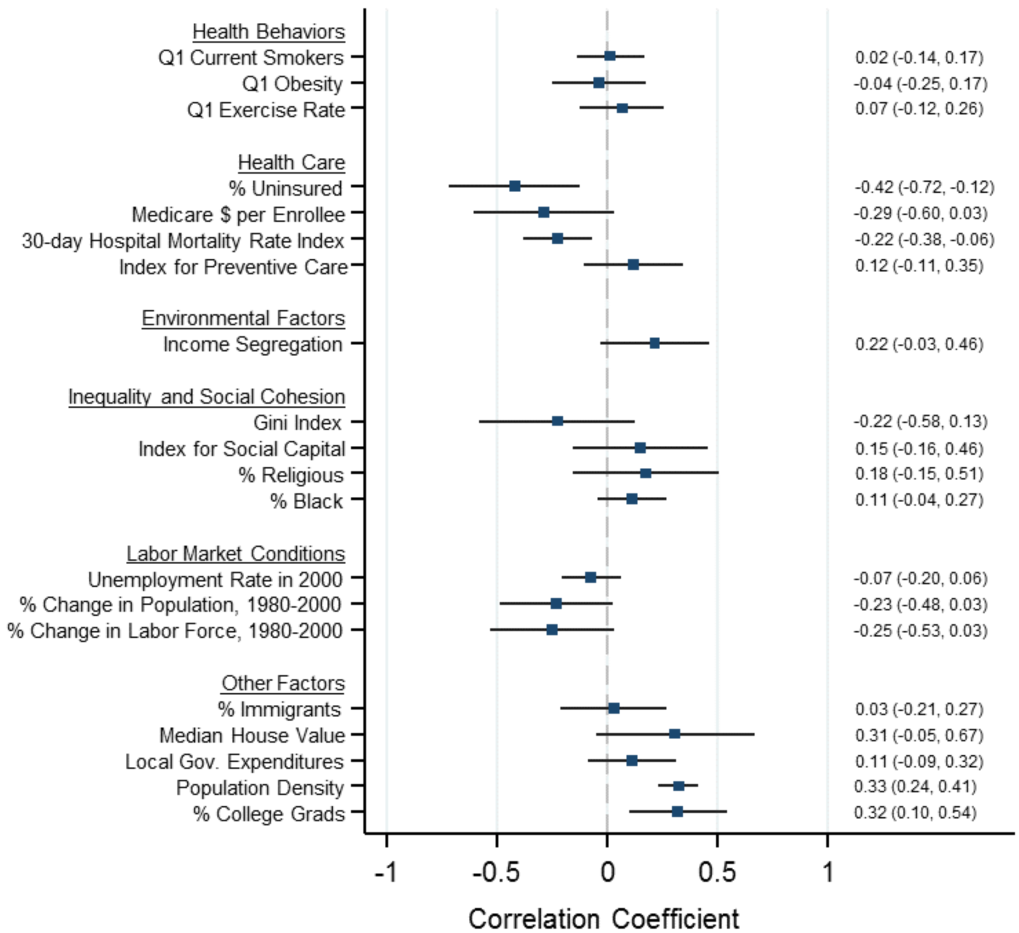


Figure shows a binned scatter plot of race- and ethnicity-adjusted expected age at death for 40 year olds by income quartile vs. the Gini index, a measure of income inequality in the commuting zone. A higher Gini index indicates greater income inequality. CZs are divided into 20 equal-sized bins (ventiles) based on their Gini index; each point shows the mean Gini index and expected age at death (mean across men and women) in each bin. Best-fit lines are estimated using ordinary least-squares regressions. CZ-level Pearson correlations between expected age at death and Gini index are reported for each quartile, along with 95% confidence intervals (in parentheses).

eFigure 15. Correlations between Changes in Race and Ethnicity Adjusted Life Expectancy in Bottom Quartile and Local Area Characteristics, 2001-2014



Population-weighted univariate Pearson correlations of local area characteristics with the annual change (from 2001-14) in race- and ethnicity-adjusted expected age at death for 40 year olds in the bottom income quartile. Computed at the commuting zone level after averaging annual change in life expectancy across genders. Estimates are based on data for the 100 most populated commuting zones, with populations above 590,000. Dots show point estimates and error bars show 95% confidence intervals, with errors clustered by state. 95% confidence intervals are also listed in parentheses. Definitions and sources of all variables appear in eTable 3.

eFigure 16. U.S. Life Expectancies by Percentile in Comparison to Mean Life Expectancies Across Countries

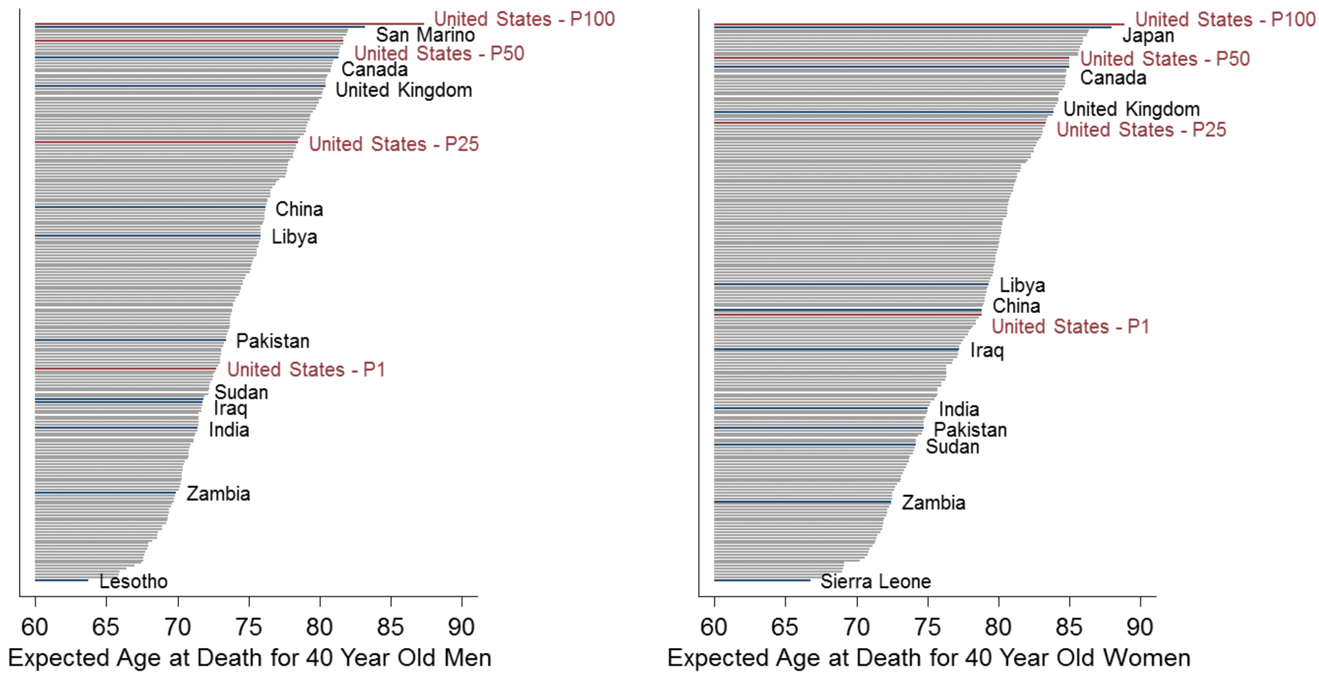
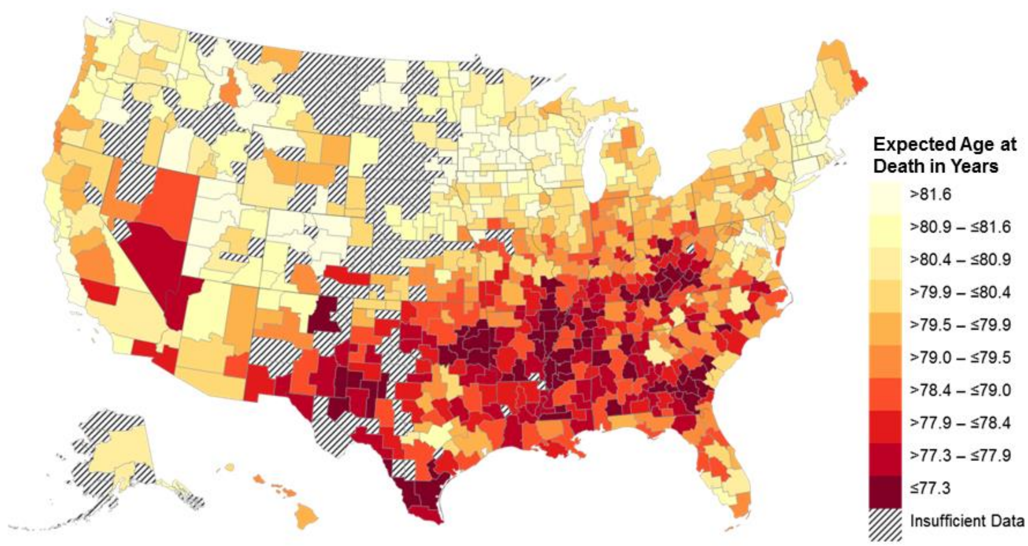


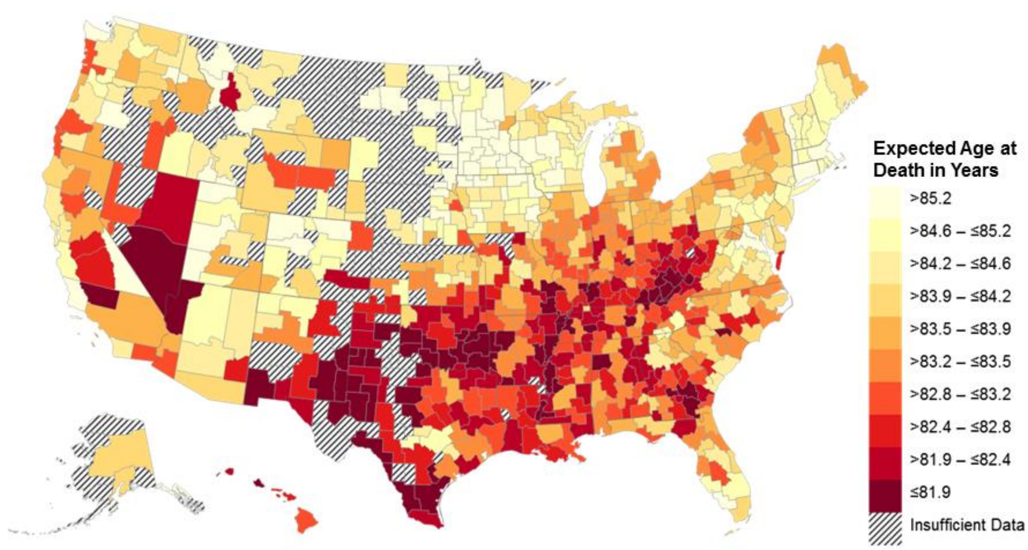
Figure shows mean life expectancy by country in 2013 based on World Health Organization statistics as well as U.S. life expectancies at the 1st, 25th, 50th, and 100th percentiles (among individuals with positive income) based on data in Figure 2.

eFigure 17. Race- and Ethnicity-Adjusted Life Expectancy by Commuting Zone, Pooling All Income Groups and Including Individuals with Zero Income, 2001-2014

A. Men



B. Women



Estimates of race- and ethnicity-adjusted expected age at death for 40 year olds, computed at the commuting zone level, pooling all income groups and including individuals with zero income. Commuting zones are grouped into deciles and colored from dark to light as expected age at death increases. Top panel shows estimates for men, bottom panel shows estimates for women. 595 CZs with populations above 25,000 are depicted.

eTable 1. Comparison of NCHS and SSA Mortality Rates and Counts, 2001-2014

A. Population Counts	NCHS	SSA	Tax Data	Tax Data, Pos. Earnings	Tax Data, Zero Earnings
Count of Men	46,750,217	57,144,071	51,460,457	42,604,804	8,402,743
Percentage of NCHS Count		122%	110%	91%	18%
Count of Women	48,677,810	56,463,456	51,323,370	44,149,965	6,779,436
Percentage of NCHS Count		116%	105%	91%	14%

B. Death Counts	NCHS	SSA	Tax Data	Tax Data, Pos. Earnings	Tax Data, Zero Earnings
Count of Men	298,465	292,460	285,462	187,757	98,675
Percentage of NCHS Count		98%	96%	62%	33%
Count of Women	189,761	181,343	177,067	118,233	59,151
Percentage of NCHS Count		96%	93%	62%	31%

C. Mortality Rates per 100,000	NCHS	SSA	Tax Data	Tax Data, Pos. Earnings	Tax Data, Zero Earnings
Rate for Men	638.4	511.8	554.7	433.1	1,174.3
Percentage of NCHS Rate		80%	87%	68%	184%
Rate for Women	389.8	321.2	345.0	264.5	872.5
Percentage of NCHS Rate		82%	89%	68%	224%

Table presents a comparison of population counts, death counts, and mortality rates across various samples. All statistics are for individuals aged 40-63 averaged across 2001-2014. Column 1 shows statistics from the CDC NCHS mortality files. Column 2 shows statistics for all individuals with a valid SSN recorded by the SSA. Column 3 shows statistics for the subsample of individuals with a valid SSN who file a tax return or have an information return (such as a W-2 form or social security income form) filed on their behalf. Column 4 further subsets to individuals with positive household earnings two years earlier, our primary analysis sample. Column 5 presents statistics for individuals with zero household earnings in the tax records. Approximately 1% of individuals have negative household earnings, which corresponds to the difference between Column 3 and the sum of Columns 4 and 5.

eTable 2. Association between Mortality Rates and Income Averages, 2001-2014

A. Men

Dependent Variable: Mortality (deaths per 100,000)			
	(1)	(2)	(3)
Inc. Pctile 2 Years Ago	-15.51 (0.326)		
Inc. Pctile 5-Year Average		-15.15 (0.350)	
Inc. Pctile 10-Year Average			-14.53 (0.362)
Obs.	949273	949273	949273

B. Women

Dependent Variable: Mortality (deaths per 100,000)			
	(1)	(2)	(3)
Inc. Pctile 2 Years Ago	-8.06 (0.250)		
Inc. Pctile 5-Year Average		-7.96 (0.268)	
Inc. Pctile 10-Year Average			-7.852 (0.277)
Obs.	1017086	1017086	1017086

Table presents estimates from an OLS regression of an indicator variable for death in a given year on the individual's income percentile at various lags. Estimates are rescaled to represent the change in deaths per 100,000 people for a 1 unit increase in income percentile. Standard errors are reported in parentheses. Column 1 reports regression estimates using income with a 2 year lag, our baseline specification. Columns 2-3 report regression estimates using income averaged over the preceding 5 (years *t-2* to *t-6*) and 10 years (years *t-2* to *t-11*). All columns use a constant sample of individuals for whom all income lags are present. Estimates are based on a random sample of individuals whose income is observed between the ages of 40 and 61.

eTable 3. Local Area Characteristics: Definitions and Data Sources

	Variable	Definition	Source
Health Behaviors			
	Smoking Rate in Corresponding Quartile	Fraction of respondents who report currently smoking in each income quartile of the pooled BRFSS sample over years 1996-2008	CDC Behavioral Risk Factor Surveillance System (BRFSS) 1996-2008[1]
	Obesity Rate in Corresponding Quartile	Fraction of respondents who are obese (BMI≥30) in each income quartile of the pooled BRFSS sample over years 1996-2008	CDC Behavioral Risk Factor Surveillance System (BRFSS) 1996-2008
	Exercise Rate in Corresponding Quartile	Fraction of respondents who have exercised in the past 30 days in each income quartile of the pooled BRFSS sample over years 1996-2008	CDC Behavioral Risk Factor Surveillance System (BRFSS) 1996-2008
Health Care			
	% Uninsured	County-level estimates of the fraction of the population without health insurance in 2010	2010 Small Area Health Insurance Estimates[2]
	Medicare $ per Enrollee	Age, sex, race, and price-adjusted Medicare reimbursements per enrollee in 2010 ($)	Dartmouth Atlas 2010[3]
Acute Care	30-day Mortality for Heart Attacks	30-day mortality rates for heart attack (acute myocardial infarction) patients	Joynt et al. (2014)[4]
	30-day Mortality for Pneumonia	30-day mortality rates for pneumonia patients	Joynt et al. (2014)
	30-day Mortality for Heart Failure	30-day mortality rates for heart failure patients	Joynt et al. (2014)
	30-day Hospital Mortality Rate Index	Standardized index combining 30-day mortality rates for heart attack, heart failure, and pneumonia patients	
Preventive Care	% with at least 1 Primary Care Visit	Average annual percentage of Medicare enrollees with at least one ambulatory visit to a primary care physician	Dartmouth Atlas 2010[5]
	% Diabetic with Hemoglobin Test	Average annual percentage of diabetic Medicare enrollees ages 65-75 with a hemoglobin A1c test	Dartmouth Atlas 2010
	% Diabetic with Eye Exam	Average annual percentage of diabetic Medicare enrollees ages 65-75 with an eye examination	Dartmouth Atlas 2010
	% Diabetic with Lipids Test	Average annual percentage of diabetic Medicare enrollees ages 65-75 with a blood lipids (LDL-C) test	Dartmouth Atlas 2010
	% Female Ages 67-69 with Mammogram	Average annual percentage of female Medicare enrollees ages 67-69 with at least one mammogram over a two-year period	Dartmouth Atlas 2010
	Discharges for Ambulatory Care Sensitive Conditions	Discharges for ambulatory care sensitive conditions per 1,000 Medicare enrollees	Dartmouth Atlas 2010
	Index for Preventive Care	Standardized index combining measures of primary care visits, hemoglobin tests, eye exams, and lipids tests among diabetics, mammograms among females ages 67-69, and discharges for ambulatory care	
Environmental Factors			
Segregation	Income Segregation	Rank-Order index estimated at the census-tract level using equation (13) in Reardon (2011)[6]; the δ vector is given in Appendix A4 of Reardon's paper. H(pk) is computed for each of the income brackets given in the 2000 census. See Appendix D of Chetty et al. (2014) for further details.	2000 Census SF3 Sample Data Table P052
	Segregation of Poverty (<p25)	H(p25) estimated following Reardon (2011); we compute H(p) for 16 income groups defined by the 2000 census. We estimate H(p25) using a fourth-order polynomial of the weighted linear regression in equation (12) of Reardon (2011).	2000 Census SF3 Sample Data Table P052

eTable 3 (continued)

	Variable	Definition	Source
	Segregation of Affluence (>p75)	Same definition as segregation of poverty, but using p75 instead of p25	2000 Census SF3 Sample Data Table P052
	Racial Segregation	Multi-group Theil Index calculated at the census-tract level over four groups: White alone, Black alone, Hispanic, and Other	2000 Census SF1 100% Data Table P008
<u>Inequality and Social Cohesion</u>			
Income Distribution	Gini Index (Within Bottom 99%)	Gini coefficient minus top 1% income share	Tax Records, Core Sample, Chetty et al. (2014)[7]
	Poverty Rate	Fraction of population below the poverty line	2000 Census SF3 Sample Data Table P087
	Top 1% Income Share	The percentage of income within a CZ going to the top 1% defined within the CZ, computed using parents of children in the Chetty et al. (2014) core sample	Tax Records, Core Sample, Chetty et al. (2014)
	Fraction Middle Class (between p25 and p75)	Fraction of parents in the Chetty et al. (2014) core sample whose income falls between the 25th and 75th percentile of the national parent income distribution	Tax Records, Core Sample, Chetty et al. (2014)
Social Cohesion	Index for Social Capital	Standardized index combining measures of voter turnout rates, the fraction of people who return their census forms, and measures of participation in community organizations	Rupasingha and Goetz (2008)[8]
	% Religious	Fraction of religious adherents	2000 Association of Religion Data Archives[9]
Race & Ethnicity	% Black	Number of individuals who are black alone divided by total population	2000 Census SF1 100% Data Table P008
	% Hispanic	Number of individuals who are Hispanic divided by total population	
<u>Labor Market Conditions</u>			
	Unemployment Rate	Unemployed civilian population 16 years and over divided by civilian labor force population 16 years and older	2000 Census SF3 Sample Data Table DP-3[10]
	% Change in Population, 1980-2000	Fraction change in CZ non-institutional civilian population from 1980 to 2000	1980, 2000 Census
	% Change in Labor Force, 1980-2000	Fraction change in CZ civilian labor force population from 1980 to 2000	1980, 2000 Census
	Labor Force Participation	Fraction of people at least 16 years old that are in the labor force	2000 Census SF3 Sample Data Table P043
	Share Working in Manufacturing	Fraction of employed persons 16 and older working in manufacturing	2000 Census SF3 Sample Data Table P049
<u>Other Factors</u>			
Migration	% Immigrants	Fraction of CZ residents born outside the United States	2000 Census SF3 Sample Data Table P021
	Migration Inflow Rate	Migration into the CZ from other CZs (divided by CZ population from 2000 Census)	IRS Statistics of Income 2004-2005
	Migration Outflow Rate	Migration out of the CZ from other CZs (divided by CZ population from 2000 Census)	IRS Statistics of Income 2004-2005
Local Geography	Population Density	Population divided by the land area in square miles	2000 Census Gazetteer Files
	Fraction with Commute < 15 Mins	Number of workers that commute less than 15 minutes to work divided by total number of workers. Sample restricted to workers that are 16 or older and not working at home.	2000 Census SF3 Sample Data Table P031

eTable 3 (continued)

	Variable	Definition	Source
Affluence	Mean Household Income	Aggregate household income in the 2000 Census divided by the number of people aged 16-64 ($)	2000 Census SF3 Sample Data Table P054
	Median House Value	Median value of housing units at the county level (population-weighted to CZ level for CZ covariate) ($)	2000 Census SF3a
K-12 Education	School Expenditure per Student	Average expenditures per student in public schools	NCES CCD 1996-1997 Financial Survey
	Student Teacher Ratio	Average student-teacher ratio in public schools	NCES CCD 1996-1997 Universe Survey
	Test Score Percentile (income adjusted)	Residual from a regression of mean math and English standardized test scores on household income per capita in 2000	George Bush Global Report Card
	High School Dropout Rate (income adjusted)	Residual from a regression of high school dropout rates on household income per capita in 2000. Coded as missing for CZs in which dropout rates are missing for more than 25% of school districts.	NCES CCD 2000-2001
College Education	% College Grads	Percentage of people at least 25 years old that have a bachelors degree	2000 Census SF3 Sample Data Table DP-2
	College Tuition	Mean in-state tuition and fees for first-time, full-time undergraduates ($)	IPEDS 2000
	College Graduation Rate (income adjusted)	Residual from a regression of graduation rate (the share of undergraduate students that complete their degree in 150% of normal time) on household income per capita in 2000	IPEDS 2009
Socioeconomics	Absolute Upward Mobility	Expected income rank of child born to parents at 25th Percentile	Tax Records, Core Sample, Chetty et al. (2014)
	Fraction of Children with Single Mothers	Number of single female households with children divided by total number of households with children	2000 Census SF3 Sample Data Table P015
	Total Crime Rate	Per capita crime rate	2000 Uniform Crime Reports
Local Taxation	Local Gov. Expenditures	Total local government expenditures per capita ($)	1992 Census of Government county-level summaries
	Local Tax Rate	Total tax revenue per capita divided by mean household income per capita for working age adults (in 1990)	1992 Census of Government county-level summaries
	Tax Progressivity	The difference between the top state income tax rate and the state income tax rate for individuals with taxable income of $20,000 in 2008	2008 state income tax rates from the Tax Foundation

Definitions and sources of all local area characteristics used for correlational analysis in Figures 8 and 9, eFigure 15, and eTables 8 and 9. All of these variables are available for download from the project website (www.healthinequality.org).

eTable 4. Trends in Race-Adjusted Life Expectancy by Quartile Controlling for Income Levels, 2001-2014

	Income Q1	Income Q2	Income Q3	Income Q4
Men	0.10	0.17	0.19	0.21
	(0.08,0.13)	(0.13,0.21)	(0.17,0.21)	(0.17,0.24)
Women	0.11	0.19	0.26	0.23
	(0.09,0.14)	(0.16,0.22)	(0.23,0.28)	(0.21,0.25)

Table reports annual changes in life expectancy by gender and income quartile (in years), controlling for changes in income levels across years. 95% confidence intervals shown in parentheses. Estimates are obtained from linear regressions of race-adjusted life expectancy on year and level of income at age 40, with one observation per gender-year-percentile bin and standard errors clustered by year. See part IV.B of Supplementary Appendix for details.

eTable 5. Standard Deviation of Life Expectancy Levels and Trends by Income Quartile, 2001-2014

	Levels (Years)			
	Q1	Q2	Q3	Q4
Men	1.39	0.79	0.81	0.70
Women	0.97	0.67	0.74	0.60
N (CZs)	595	595	595	595

	Trends (Annual Change)			
	Q1	Q2	Q3	Q4
Men	0.071	0.078	0.040	0.033
Women	0.072	0.051	0.041	0.055
N (States)	51	51	51	51

Upper panel of table shows the signal standard deviations of expected age at death across commuting zones, by income quartile and gender. Lower panel shows the signal standard deviations of the annual trends in life expectancy across states, by income quartile and gender. Standard deviations are population weighted and exclude variation due to sampling error in life expectancy estimates using the method described in part V.C of the Supplementary Appendix.

eTable 6. Local Area Estimates of Life Expectancy: Sensitivity Analysis

Men

Income Quartile	Not Race and Ethnicity Adjusted	Life Years up to 77	Gompertz to Age 100	Ordinary Least Squares	Income Controlled	Cost of Living Adjusted
1	0.938	0.971	0.998	0.993	0.990	0.989
2	0.711	0.899	0.994	0.983	0.960	0.832
3	0.753	0.941	0.995	0.984	0.965	0.788
4	0.871	0.912	0.988	0.957	0.894	0.889
N (CZs)	595	595	595	595	207	595

Women

Income Quartile	Not Race and Ethnicity Adjusted	Life Years up to 77	Gompertz to Age 100	Ordinary Least Squares	Income Controlled	Cost of Living Adjusted
1	0.843	0.953	0.994	0.990	0.971	0.988
2	0.399	0.845	0.987	0.975	0.946	0.882
3	0.530	0.933	0.993	0.977	0.951	0.875
4	0.655	0.895	0.987	0.964	0.926	0.905
N (CZs)	595	595	595	595	207	595

Table reports population-weighted correlations across commuting zones of baseline estimates of life expectancy with alternative estimates, separately by income quartile and gender. The first column correlates baseline race/ethnicity adjusted estimate with race/ethnicity unadjusted estimates. The second column correlates the baseline life expectancy estimates, which use extrapolated mortality rates after age 77, with the expected number of life years up to age 77. The third column correlates the baseline estimates, which use a Gompertz extrapolation to age 90, with estimates that use Gompertz extrapolations to age 100. The fourth column correlates the baseline estimates, which use Gompertz parameters estimated using maximum likelihood, with estimates where the Gompertz parameters are estimated using ordinary least squares. The fifth column correlates the baseline estimates with estimates that reweight the within-quartile income distribution in each CZ to match the national distribution. The sixth column correlates the baseline estimates with estimates that classify individuals into income quartiles based on real income, adjusting for differences in the cost of living across areas.

eTable 7. Race and Ethnicity Adjusted Expected Age at Death by County and Income Quartile, 2001-2014

Race and Ethnicity Adjusted Expected Age at Death in Years, Bottom Quartile

Rank	County	Bottom Income Quartile, Mean	Bottom Income Quartile, Men	Bottom Income Quartile, Women	Diff. Between Top and Bottom Income Quartiles, Mean
1	Queens, NY	82.6 (82.3, 83.0)	80.2 (79.8, 80.7)	85.0 (84.5, 85.5)	2.9 (2.5, 3.3)
2	Kings, NY	82.6 (82.3, 82.8)	80.3 (80.0, 80.7)	84.8 (84.5, 85.2)	3.3 (2.9, 3.8)
3	Montgomery, MD	82.2 (81.7, 82.7)	79.9 (79.2, 80.6)	84.5 (83.8, 85.2)	5.1 (4.6, 5.7)
4	Nassau, NY	82.0 (81.7, 82.3)	79.8 (79.3, 80.2)	84.2 (83.7, 84.7)	4.7 (4.3, 5.1)
5	Bergen, NJ	81.9 (81.5, 82.3)	79.5 (79.0, 80.0)	84.3 (83.7, 84.9)	4.7 (4.2, 5.2)
6	New York, NY	81.8 (81.5, 82.3)	79.8 (79.3, 80.3)	83.9 (83.3, 84.6)	5.6 (5.1, 6.1)
7	Westchester, NY	81.8 (81.5, 82.2)	79.8 (79.3, 80.3)	83.9 (83.3, 84.4)	5.5 (5.1, 6.0)
8	Ventura, CA	81.8 (81.4, 82.2)	80.1 (79.6, 80.7)	83.5 (82.8, 84.2)	4.6 (4.1, 5.2)
9	Orange, CA	81.6 (81.3, 82.0)	79.6 (79.2, 80.1)	83.6 (83.0, 84.2)	4.7 (4.3, 4.9)
10	San Francisco, CA	81.6 (81.0, 82.3)	79.3 (78.6, 80.1)	83.9 (83.0, 85.1)	3.9 (3.3, 4.6)
...	*Entire United States*	*79.4 (79.4, 79.5)*	*76.7 (76.7, 76.8)*	*82.1 (82.1, 82.2)*	*7.0 (6.9, 7.1)*
91	Franklin, OH	77.9 (77.6, 78.2)	75.3 (74.9, 75.7)	80.5 (80.0, 81.0)	8.7 (8.2, 9.2)
92	Tulsa, OK	77.8 (77.5, 78.2)	75.3 (74.9, 75.8)	80.3 (79.8, 80.8)	8.1 (7.6, 8.7)
93	Montgomery, OH	77.8 (77.4, 78.1)	74.8 (74.3, 75.2)	80.8 (80.3, 81.4)	9.0 (8.4, 9.6)
94	Hamilton, OH	77.8 (77.5, 78.1)	75.4 (74.9, 75.8)	80.2 (79.7, 80.7)	8.6 (8.1, 9.2)
95	Bexar, TX	77.8 (77.3, 78.2)	75.0 (74.5, 75.6)	80.5 (79.7, 81.3)	8.1 (7.5, 8.6)
96	Clark, NV	77.6 (77.4, 77.9)	75.2 (75.0, 75.5)	80.0 (79.6, 80.4)	6.4 (6.1, 6.8)
97	Oklahoma, OK	77.5 (77.2, 77.8)	74.9 (74.5, 75.3)	80.0 (79.6, 80.5)	8.5 (8.0, 9.1)
98	Macomb, MI	77.5 (77.2, 77.8)	74.6 (74.2, 75.1)	80.3 (79.9, 80.9)	7.6 (7.1, 8.0)
99	Marion, IN	77.1 (76.9, 77.4)	74.4 (74.0, 74.7)	79.9 (79.5, 80.4)	8.9 (8.4, 9.4)
100	Wayne, MI	77.1 (76.9, 77.4)	74.1 (73.8, 74.4)	80.1 (79.7, 80.5)	7.9 (7.6, 8.3)

Race and Ethnicity Adjusted Expected Age at Death in Years, Top Quartile

Rank	County	Top Income Quartile, Mean	Top Income Quartile, Men	Top Income Quartile, Women	Diff. Between Top and Bottom Income Quartiles, Mean
1	Salt Lake, UT	87.7 (87.3, 88.1)	86.6 (86.1, 87.1)	88.8 (88.3, 89.3)	8.1 (7.6, 8.6)
2	Washington, DC	87.6 (87.2, 88.1)	86.9 (86.4, 87.6)	88.3 (87.7, 89.0)	8.4 (7.8, 8.9)
3	Fulton, GA	87.6 (87.2, 87.9)	86.4 (85.9, 86.9)	88.8 (88.3, 89.3)	7.8 (7.2, 8.3)
4	Fairfield, CT	87.5 (87.3, 87.8)	86.4 (86.1, 86.8)	88.6 (88.3, 89.0)	6.6 (6.1, 7.0)
5	Hennepin, MN	87.5 (87.2, 87.8)	86.7 (86.3, 87.1)	88.2 (87.8, 88.7)	7.5 (7.1, 8.0)
6	New York, NY	87.5 (87.2, 87.7)	86.8 (86.5, 87.2)	88.1 (87.8, 88.5)	5.6 (5.1, 6.1)
7	Monroe, NY	87.4 (87.1, 87.9)	86.6 (86.0, 87.1)	88.3 (87.8, 88.9)	7.1 (6.6, 7.7)
8	Westchester, NY	87.3 (87.1, 87.6)	86.1 (85.8, 86.5)	88.6 (88.2, 88.9)	5.5 (5.1, 6.0)
9	Kent, MI	87.3 (86.9, 87.8)	86.2 (85.6, 86.9)	88.4 (87.8, 89.1)	8.0 (7.4, 8.7)
10	Montgomery, MD	87.3 (87.0, 87.6)	86.4 (86.0, 86.8)	88.3 (87.9, 88.7)	5.1 (4.6, 5.7)
...	*Entire United States*	*86.4 (86.3, 86.5)*	*85.3 (85.2, 85.4)*	*87.5 (87.4, 87.6)*	*7.0 (6.9, 7.1)*
91	Dade, FL	85.1 (84.6, 85.7)	83.9 (83.2, 84.5)	86.4 (85.6, 87.3)	3.8 (3.4, 4.2)
92	Wayne, MI	85.0 (84.7, 85.3)	83.7 (83.3, 84.1)	86.3 (86.0, 86.8)	7.9 (7.6, 8.3)
93	Kern, CA	85.0 (84.5, 85.6)	84.1 (83.4, 84.8)	86.0 (85.2, 86.8)	6.1 (5.5, 6.8)
94	Macomb, MI	85.0 (84.7, 85.4)	83.5 (83.1, 84.0)	86.5 (86.0, 87.1)	7.6 (7.1, 8.0)
95	Middlesex, NJ	84.9 (84.5, 85.4)	83.7 (83.1, 84.3)	86.1 (85.4, 86.9)	4.3 (3.7, 4.9)
96	Honolulu, HI	84.8 (83.8, 86.0)	84.2 (83.0, 85.6)	85.3 (83.8, 87.3)	6.6 (6.0, 7.2)
97	San Bernardino, CA	84.7 (84.2, 85.0)	83.3 (82.8, 83.8)	86.0 (85.5, 86.6)	5.5 (5.0, 5.9)
98	Clark, NV	84.0 (83.7, 84.3)	82.8 (82.4, 83.2)	85.3 (84.8, 85.8)	6.4 (6.1, 6.8)
99	El Paso, TX	83.9 (83.1, 84.8)	81.8 (80.7, 82.8)	86.0 (84.9, 87.3)	4.6 (3.8, 5.5)
100	Hudson, NJ	83.6 (82.9, 84.3)	81.5 (80.6, 82.3)	85.8 (84.8, 86.8)	3.2 (2.4, 4.0)

Table shows estimates of race-adjusted expected age at death for top 10 and bottom 10 counties among the 100 largest counties by population. Column 1 reports means across genders; columns 2 and 3 report estimates by gender; column 4 reports longevity gap (top income quartile minus bottom income quartile), pooling genders. Estimates for top and bottom income quartiles are shown.

eTable 8. Correlations between Life Expectancy and Local Area Characteristics across Commuting Zones by Income Quartile, 2001-2014

	Variable	Mean	Men Q1	Women Q1	Men Q4	Women Q4
Health Behaviors						
	Smoking Rate in Corresponding Quantile	0.27; 0.12	-0.70	-0.60	-0.34	-0.24
	Obesity Rate in Corresponding Quantile	0.28; 0.20	-0.53	-0.34	-0.37	-0.12
	Exercise Rate in Corresponding Quantile	0.61; 0.87	0.39	0.18	0.47	0.34
Health Care						
	% Uninsured	17.47	0.09	0.10	-0.40	-0.37
	Medicare $ per Enrollee	9,617.46	-0.13	-0.03	-0.47	-0.40
Acute Care	30-day Hospital Mortality Rate Index	0.00	-0.32	-0.26	-0.25	-0.08
	30-day Mortality for Heart Attacks	0.15	-0.36	-0.30	-0.33	-0.18
	30-day Mortality for Pneumonia	0.10	-0.24	-0.24	0.03	0.17
	30-day Mortality for Heart Failure	0.12	-0.20	-0.11	-0.33	-0.20
Preventive Care	Index for Preventive Care	0.01	0.02	0.09	0.47	0.51
	% with at least 1 Primary Care Visit	77.81	-0.64	-0.48	-0.05	0.22
	% Diabetic with Hemoglobin Test	83.76	-0.09	0.01	0.35	0.42
	% Diabetic with Eye Exam	67.21	0.21	0.26	0.43	0.45
	% Diabetic with Lipids Test	80.86	0.21	0.27	0.16	0.10
	% Female Ages 67-69 with Mammogram	65.02	-0.03	0.06	0.44	0.51
	Discharges for Amb. Care Sensitive Conds.	65.55	-0.42	-0.24	-0.49	-0.25
Environmental Factors						
Segregation	Income Segregation	0.09	0.32	0.16	0.16	-0.11
	Segregation of Poverty (<p25)	0.08	0.25	0.10	0.17	-0.07
	Segregation of Affluence (>p75)	0.10	0.35	0.18	0.15	-0.12
	Racial Segregation	0.24	0.08	0.08	-0.09	-0.13
Inequality and Social Cohesion						
Income Distribution	Gini Index (Within Bottom 99%)	0.33	0.15	0.25	-0.33	-0.33
	Poverty Rate	0.12	0.07	0.13	-0.41	-0.37
	Top 1% Income Share	0.15	0.61	0.51	0.13	-0.10
	Fraction Middle Class (between p25 and p75)	0.50	-0.34	-0.30	0.13	0.28
Social Cohesion	Index for Social Capital	-0.42	-0.24	-0.28	0.37	0.45
	% Religious	50.15	0.10	0.13	-0.04	0.07
Race & Ethnicity	% Black	12.13	-0.14	0.07	-0.19	-0.11
	% Hispanic	12.56	0.48	0.30	-0.20	-0.39
Labor Market Conditions						
	Unemployment Rate in 2000	0.05	0.11	0.10	-0.36	-0.30
	% Change in Population, 1980-2000	0.30	0.20	0.10	0.02	-0.09
	% Change in Labor Force, 1980-2000	0.36	0.10	0.05	0.09	0.04
	Labor Force Participation	0.64	-0.02	-0.15	0.42	0.34
	Share Working in Manufacturing	0.14	-0.34	-0.22	-0.13	0.08
Other Factors						
Migration	% Immigrants	11.10	0.76	0.59	-0.04	-0.33

eTable 8 (continued)

	Variable	Mean	Men Q1	Women Q1	Men Q4	Women Q4
	Migration Inflow Rate	0.02	-0.01	-0.04	0.07	0.04
	Migration Outflow Rate	0.02	0.35	0.24	0.13	-0.07
Local Geography	Population Density	649.55	0.48	0.44	0.07	-0.02
	Fraction with Commute < 15 Mins	0.29	-0.33	-0.28	-0.01	0.19
Affluence	Mean Household Income	39,483.21	0.40	0.27	0.34	0.13
	Median House Value	179,657.31	0.71	0.53	0.26	-0.10
K-12 Education	School Expenditure per Student	6.39	0.29	0.25	0.30	0.22
	Student Teacher Ratio	18.62	0.46	0.27	0.02	-0.26
	Test Score Percentile (income adjusted)	-3.55	-0.33	-0.21	0.20	0.43
	High School Dropout Rate (income adjusted)	0.01	0.06	0.08	-0.10	-0.15
College Education	% College Grads	24.40	0.48	0.31	0.50	0.21
	College Tuition	5,986.82	0.14	0.17	0.21	0.19
	College Graduation Rate (income adjusted)	-0.02	-0.05	0.05	0.22	0.26
Socioeconomics	Absolute Upward Mobility	41.60	0.42	0.26	0.20	0.09
	Fraction of Children with Single Mothers	0.22	-0.15	0.02	-0.20	-0.14
	Total Crime Rate	0.01	0.03	-0.02	-0.10	-0.14
Local Taxation	Local Government Expenditures	2,586.10	0.63	0.42	0.18	-0.10
	Local Tax Rate	0.02	0.33	0.23	0.19	0.08
	Tax Progressivity	1.41	0.45	0.30	0.00	-0.21

First column shows the population-weighted mean of each variable. Remaining columns show population-weighted univariate Pearson correlations of race-adjusted expected age at death for 40 year old men and women in the bottom and top income quartiles with local area characteristics, computed at the commuting zone level. CZs with populations below 25,000 are omitted from all columns. See eTable 3 for definitions and sources of all variables.

eTable 9. Correlations between Life Expectancy and Local Area Characteristics across Counties by Income Quartile, 2001-2014

	Variable	Mean	Men Q1	Women Q1	Men Q4	Women Q4
Health Behaviors						
	Smoking Rate in Corresponding Quantile	0.26; 0.13	-0.59	-0.49	-0.27	-0.18
	Obesity Rate in Corresponding Quantile	0.28; 0.20	-0.44	-0.31	-0.31	-0.15
	Exercise Rate in Corresponding Quantile	0.62; 0.87	0.33	0.19	0.37	0.25
Health Care						
	% Uninsured	17.46	0.07	0.08	-0.31	-0.29
	Medicare $ per Enrollee	9,650.66	-0.10	0.00	-0.37	-0.32
Acute Care	30-day Hospital Mortality Rate Index	-0.05	-0.28	-0.24	-0.19	-0.08
	30-day Mortality for Heart Attacks	0.15	-0.30	-0.25	-0.22	-0.14
	30-day Mortality for Pneumonia	0.10	-0.20	-0.20	0.02	0.11
	30-day Mortality for Heart Failure	0.12	-0.20	-0.15	-0.26	-0.16
Preventive Care	Index for Preventive Care	0.02	0.07	0.12	0.41	0.44
	% with at least 1 Primary Care Visit	77.39	-0.51	-0.37	0.01	0.21
	% Diabetic with Hemoglobin Test	83.66	0.00	0.06	0.27	0.31
	% Diabetic with Eye Exam	67.24	0.22	0.27	0.36	0.36
	% Diabetic with Lipids Test	80.90	0.22	0.26	0.11	0.10
	% Female Ages 67-69 with Mammogram	65.04	-0.01	0.05	0.40	0.46
	Discharges for Amb. Care Sensitive Conds.	64.26	-0.38	-0.22	-0.45	-0.27
Environmental Factors						
Segregation	Income Segregation	0.08	0.18	0.06	0.20	0.02
	Segregation of Poverty (<p25)	0.07	0.07	-0.02	0.19	0.02
	Segregation of Affluence (>p75)	0.09	0.24	0.10	0.20	0.02
	Racial Segregation	0.19	0.10	0.10	-0.06	-0.14
Inequality and Social Cohesion						
Income Distribution	Gini Index (Within Bottom 99%)	0.46	0.28	0.28	0.03	-0.10
	Poverty Rate	0.12	-0.01	0.04	-0.29	-0.27
	Top 1% Income Share	0.14	0.43	0.36	0.23	0.04
	Fraction Middle Class (between p25 and p75)	0.49	-0.38	-0.30	-0.08	0.04
Social Cohesion	Index for Social Capital	-0.45	-0.14	-0.18	0.38	0.38
	% Religious	49.93	0.09	0.13	0.06	0.07
Race & Ethnicity	% Black	12.24	-0.14	-0.01	-0.15	-0.11
	% Hispanic	13.07	0.40	0.25	-0.19	-0.30
Labor Market Conditions						
	Unemployment Rate in 2000	0.05	0.03	0.03	-0.35	-0.28
	% Change in Population, 1980-2000	0.34	0.19	0.11	0.01	0.00
	% Change in Labor Force, 1980-2000	0.41	0.13	0.09	0.05	0.07
	Labor Force Participation	0.64	0.03	-0.08	0.36	0.29
	Share Working in Manufacturing	0.14	-0.28	-0.17	-0.11	0.02
Other Factors						
Migration	% Immigrants	11.67	0.69	0.54	-0.06	-0.26

eTable 9 (continued)

	Variable	Mean	Men Q1	Women Q1	Men Q4	Women Q4
	Migration Inflow Rate	0.04	0.04	-0.02	0.10	0.09
	Migration Outflow Rate	0.04	0.25	0.14	0.17	0.07
Local Geography	Population Density	2,271.20	0.32	0.28	0.03	-0.05
	Fraction with Commute < 15 Mins	0.29	-0.33	-0.28	0.06	0.18
Affluence	Mean Household Income	40,020.01	0.40	0.28	0.38	0.23
	Median House Value	189,790.58	0.58	0.44	0.25	0.03
K-12 Education	School Expenditure per Student	6.44	0.28	0.24	0.24	0.18
	Student Teacher Ratio	18.90	0.42	0.24	0.00	-0.21
	Test Score Percentile (income adjusted)	-6.00	-0.26	-0.16	0.05	0.22
	High School Dropout Rate (income adjusted)	0.01	0.02	0.01	-0.04	-0.11
College Education	% College Grads	25.11	0.47	0.34	0.51	0.30
	College Tuition	6,176.33	0.07	0.10	0.16	0.14
	College Graduation Rate (income adjusted)	0.01	0.07	0.09	0.26	0.17
Socioeconomics	Absolute Upward Mobility	44.33	0.30	0.20	0.15	0.12
	Fraction of Children with Single Mothers	0.22	-0.19	-0.11	-0.13	-0.12
	Total Crime Rate	0.01	-0.06	-0.09	-0.07	-0.10
Local Taxation	Local Government Expenditures	2,548.62	0.24	0.10	0.17	0.01
	Local Tax Rate	0.02	0.05	-0.02	0.15	0.10
	Tax Progressivity	1.45	0.35	0.21	-0.03	-0.17

First column shows the population-weighted mean of each variable. Remaining columns show population-weighted univariate Pearson correlations of race-adjusted expected age at death for 40 year old men and women in the bottom and top income quartiles with local area characteristics, computed at the county level. Counties with populations below 25,000 are omitted from all columns. See eTable 3 for definitions and sources of all variables.

eTable 10. Decomposition of Mortality Rates by Cause of Death

	No College		*Some College*	
	(1)	(2)	(3)	(4)
	External	Medical	External	Medical
Total Mortality	0.127 (0.009)	0.873 (0.009)	0.147 (0.014)	0.853 (0.014)
N (CZs)	237	237	239	239

Table shows population-weighted regressions of external and medical mortality rates on total mortality rates by CZ. Standard errors are reported in parentheses. The first two columns include individuals with a high school degree or less, columns 3 and 4 include those with some education beyond high school. Data are from the 2004 NCHS and 2000 Census. Mortality rates are for 40-64 year olds, and are reweighted to match the national distribution by age and gender. Mortality rates are expressed at the CZ level as the number of deaths from external or medical causes divided by CZ population, aggregated from county data. External causes include accidents, self-harm, and assault. Medical causes include diseases and deaths related to medical problems. The NCHS does not report data for counties with populations below 100,000; these counties are excluded from CZ averages, and CZs that have no counties with unsuppressed data are excluded entirely.

References

1. Arias E. United States life tables, 1999-2010. In: National Center for Health Statistics, ed. National Vital Statistics Reports 1999-2014.
2. Olshansky SJ, Carnes BA. Ever since gompertz. Demography 1997;34.1:1-15.
3. Bell FC, Miller ML. Life Tables for the United States Social Security Area 1900-2100. In: Actuary OotC, ed. Actuarial Study No 120. SSA Pub. No. 11-11536: Social Security Administration; 2005.
4. Arias E. United States life tables, 2010. National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System 2014;63:1-63.
5. Case A, Deaton A. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. Proceedings of the National Academy of Sciences 2015.
6. Current Estimates Data. U.S. Census Bureau; 2000-2014.
7. Intercensal Estimates. U.S. Census Bureau; 2000-2010.
8. IPUMS USA. (Accessed September 1, 2015, at https://usa.ipums.org/usa/.)
9. Census 2000 SF3, Tables P55 and PCT72. American FactFinder: U.S. Census Bureau; 2000.
10. Joynt KE, Orav E, Jha AK. Association between hospital conversions to for-profit status and clinical and economic outcomes. JAMA 2014;312:1644-52.
11. Kling JR, Liebman JB, Katz LF. Experimental Analysis of Neighborhood Effects. Econometrica 2007;75:83-119.
12. Chetty R, Hendren N, Kline P, Saez E. Where is the Land of Opportunity: The Geography of Intergenerational Mobility in the United States. Quarterly Journal of Economics 2014;129:1553-623.
13. Underlying Cause of Death 1999-2013 on CDC WONDER Online Database. Centers for Disease Control and Prevention; released 2015.
14. Behavioral Risk Factor Surveillance System. Centers for Disease Control and Prevention; 1996-2008.
15. Model-based Small Area Health Insurance Estimates (SAHIE) for Counties and States. United States Census Bureau; 2010.
16. Medicare Spending: County level. The Dartmouth Atlas of Health Care; 2010.
17. Selected measures of primary care access and quality. The Dartmouth Atlas of Health Care; 2010.
18. Reardon S. Measures of Income Segregation. In: Stanford Center on Poverty and Inequality, ed. Stanford University,2011.
19. Rupashinga A, Goetz SJ. US County-Level Social Capital Data, 1990-2005. University Park, PA: The Northeast Regional Center for Rural Development, Penn State University; 2008.
20. Religious Congregations and Membership Study. In: Bodies AoSoAR, ed.: Association of Religion Data Archives; 2000.
21. National Historical Geographic Information System: Version 2.0. . Minneapolis, MN: University of Minnesota Minnesota Population Center; 2011.