Association Between Income and Life Expectancy in the United States, 2001-2014

February 22, 2017

These notes are based off a presentation by Michael Stepner (MIT, Department of Economics) for the section on Economic Inequality in the Mechanism Design for Social Good Reading Group. The noted are written by members of the reading group. Figures and some of the text are taken from the accompanying project page and paper: `https://healthinequality.org/`.

1 Introduction

- The association between higher income and longer life expectancy is well known, but there are still many unresolved questions:

- **Question 1:** What is the "shape" of the relationship between income and life expectancy?

- **Question 2:** How is this relationship changing over time?

- **Question 3:** How does it vary by region?

- **Question 4:** Where does the longevity gap come from?

- In this talk, we take steps to answer all four questions above.

- This study was done using de-identified U.S. tax records from 1999-2014. It includes 1.4 billion observations, significantly more than in prior work. Not only does this allow for more precise national estimates, it also makes estimating statistics for smaller local areas possible. The data only includes individuals aged 40 and above.

- Significant contributions:

  - Broadly, this study finds that at a national level, gaps in longevity across income groups are growing (Questions 1 and 2).
  - This is not necessarily true at local levels – the change in life expectancy over time varies greatly by region, especially for those with low income (Question 3).
  - They also provide a characterization of correlates (but no claims of causality) with this spatial variation in longevity (Question 4).
  - Their data is publicly available, enabling further studies of this type and easier systematic measurement of progress.

- Limitations of data: Because it relies on tax records, this study does not include individuals with no income or individuals who participate in the informal economy, or individuals who do not pay taxes, unless their employers file W-2's on their behalf. So, there might be many missing individuals.
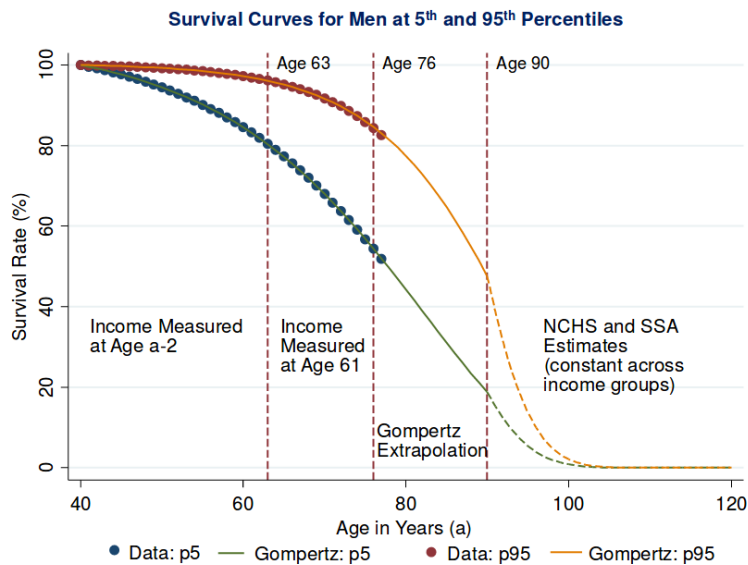
2 Data and Methodology

2.1 Data

- Income data comes from U.S. tax returns, and mortality data comes from Social Security Administration data, which is closely aligned with CDC statistics.

- The basic unit is *household income*, meaning we're not looking at each individual's income.

- Individuals with zero household income (8% people at age 40) are excluded because their mortality rates are not fully tracked in SSA data.

- We use *percentile ranks* to compare people – rank in income distribution within birth cohort, gender, and tax year.

- **Goal:** Estimate expected age of death conditioned on income at age 40, controlling for race and ethnicity differences. Since life expectancy changes over time, standard practice is to use *period life expectancy*, which is the life expectancy of a hypothetical individual who experiences today's mortality rates every year for the rest of his/her life. This is straightforward to compute if one could observe mortality rates at all agents for all racial groups conditioned on income at age 40. But, there are two problems:

- **Problem 1:** Since the data only spans 15 years, mortality rates aren't fully observed. If we look at the cohort for whom we have income at age 40, we don't observe their mortality rates past age 55.

- **Solution:** To extrapolate over missing data, we calculate mortality rates with whatever data we have and use this to fit a Gompertz model, which predicts that mortality rates grow exponentially with age. Then, we adjust for differences in ethnicity/race to standardize between areas with different demographics. In addition, the income used at time of death has to have a "lag," since income may vary dramatically very close to the time a person dies. Thus, mortality rates at age a are calculated as a function of income at age $a - 2$.

- **Problem 2:** Tax data does not contain race/ethnicity information.

- Some results on income vs mortality:

  - For age 50-54, the top 1% have a mortality rate that is an order of magnitude lower than the bottom 1%.

  - There is a significant gap in survival from age 40 to 76 between men in the 5th percentile (around 50%) and 95th percentile (over 80%) of income.

  - Medicare doesn't affect the exponential growth in mortality rate over time.

- – The Gompertz model holds for income subgroups, using the unique dataset. They also verify that the Gompertz model holds for race/ethnicity subgroups using public data.

Note: Data and statistics can be found at `www.heathinequality.org`.

2.2 Methodology

- As discussed above, there will be three steps to estimating life expectancy by income group:
  - – Calculate observed mortality rate: For working age sample (below age 63), start by calculating mortality rates as a function of income. Then, use this to construct a survival curve, which can change over time.
  - – Use income lagged by 2 years, which is a good approximation of income at age 40 due to the lag invariance property documented in the appendix of the paper.
  - – Use age profile of mortality rates to estimate Gompertz models: We want to get to life expectancy and in particular get past age 76, but data stops at 76. Gompertz documented a robust empirical regularity which is that mortality rates grow exponentially with age. We can use this to get up to age 90. We can then use NCHS and SSA estimates, which are constant across income groups.

Survival Curves for Men at 5th and 95th Percentiles

  - – Finally, we want to adjust for race and ethnicity. The CDC statistics shows that for males, life expectancy of whites is 3.8 years higher than blacks and 2.7 years lower than Hispanics. Race shares vary across income groups and especially across areas, which rules out raw comparisons. We would therefore like to perform race and ethnicity adjustment to answer the question: what would life expectancy be if each income group and area had the same black, Hispanic, and Asian shares as the U.S. population as a whole at age 40?
  - – We construct race-adjusted measures of life expectancy in four steps:

∗ Estimate differences in mortality by race controlling for income using data from the National Longitudinal Mortality Study

∗ Estimate racial demographics in each income group and area using Census data

∗ Recover mortality rates by race in each income group and area from aggregate rates in tax data and race differences from the NLMS

∗ Calculate life expectancy that would prevail if racial demographics were the same as the national demographics at age 40.

3   National-Level Results

- Summary of statistics:

  - Life for men expectancy ranges from 73 to 87 years for bottom 1% vs top 1%.
  - The bottom 1% of men have life expectancy comparable to men in Pakistan.
  - A 10-year gap in life expectancy exists between poorest and richest women.
  - Women tend to have higher life expectancies than men, but this gender gap decreases for higher incomes (both in terms of percentiles and in raw dollars). A similar trend holds when we consider individual instead of household income.

- **Time trends:** We want to know how these gaps change over time. This information is necessary to understand policy effects. Some past work has shown that the gap between life expectancy for the rich and the poor is increasing. Figure 1 shows that nationally, the life expectancy for the richest women is increasing more quickly than for the poorest women. A similar trend holds for men. The bottom 5% of both genders has experienced virtually no growth in life expectancy over the last 14 years. On the other hand, the highest-earning men and women have gained about 3 years of life expectancy during that time, which is equivalent to the estimated increase in life expectancy if cancer could be completely cured.

4   Local Areas

- We'll now focus on results on the level of a "commuting zone," which is essentially a metro area, though it's possible to get county-level results from the data as well.

- This requires disaggregating by income across area, which hasn't been done before due to lack of data.

- **Important result:** While life expectancy for high earners is consistent across area, life expectancy for the poor *varies widely by area*. Figure 2 demonstrates this for men, and similar results hold for women.

- In general, the Midwest and South have low life expectancy for the poor (Figure 3). Furthermore, there's a huge difference between California and Nevada, though they are neighbors. Interestingly, although the Southeast generally has low life expectancy, when adjusting for race and looking at the bottom quartile, things aren't bad there. This is especially true for women (Figure 4). The worst is really a belt that stretches from Texas through Michigan.
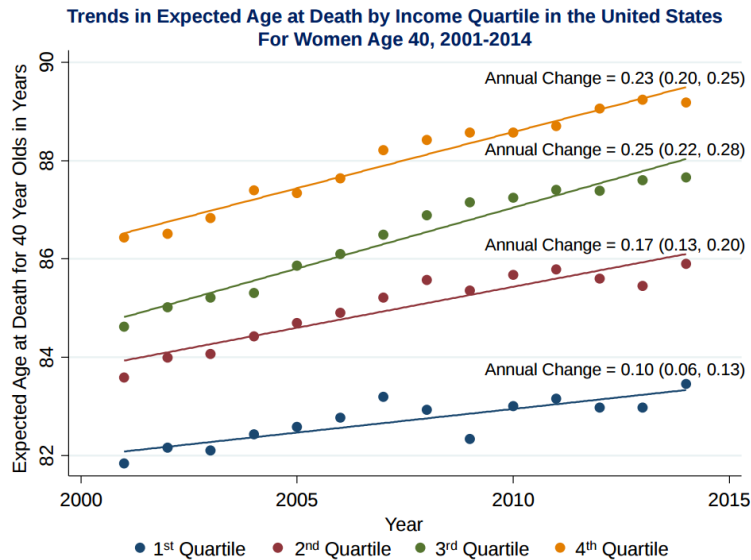
Figure 1: Life expectancy is increasing faster for the rich than for the poor.

- Given that this is the case, one might think that improving incomes in the Southeast would help life expectancy there; however, causality might work in the other direction (good health makes it easier to earn an income). Also, there are plenty of other variables that are unaccounted for here.

- We might want to separate by cause of death, but this is difficult given the data we have.

- Even within a commuting zone, there is large variation in life expectancy for the poor (Figure 5).

- **Local trends over time:** Different areas have very different rates of change for life expectancy for the poor. For example, Birmingham AL has had an annual change of .37 years gained per year for women in the bottom quartile over the last 15 years, while Tampa FL has lost .16 years of life expectancy per year. In general, the coasts (except Florida) have experienced positive change in life expectancy for the bottom quartile, while the Midwest has had smaller or even negative change. The natural question to ask is what causes these differences, or at least, what correlates can we find?

5   Correlates of Variation in Life Expectancy by Region

- The goal of this section is to test various hypotheses of the form, "Is this attribute correlated with higher life expectancy among the poor?"

- As one might expect, smoking and obesity rates are negatively correlated with life expectancy for the bottom quartile, while exercise rate is positively correlated.

- We consider four other classes of commonly proposed correlates:
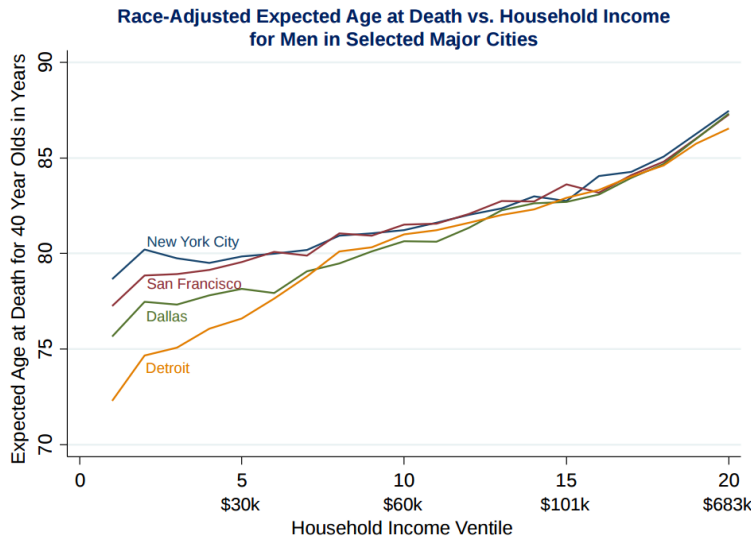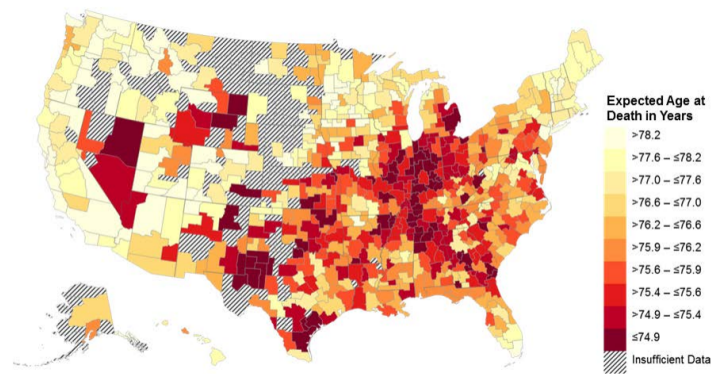
Figure 2: Life expectancy for the poor varies widely by area.

- **Health care:** The only health care metric that was found to be strongly correlated with life expectancy was a 30-day hospital mortality rate index, which is measured over several dangerous conditions and is used as a proxy for hospital quality.
- **Environmental factors:** Income segregation, which measures how segregated an area is by income, is used to test the hypothesis that when the poor live in lower-quality areas than the rich, they may be exposed to environmental factors that decrease life expectancy. However, no strong correlation was found here.
- **Inequality and social cohesion:** Using measures such as Gini index, as well as other metrics for social cohesion, no strong correlations were found.
- **Economic decline:** Again, no correlations were found when considering unemployment rates, change in population, and change in labor force.

• Since none of these proposed factors were strongly correlated with life expectancy, what is?

- Immigrant percentage
- Median house value
- Local government expenditures
- Population density
- Percent of college graduates in population

• The pattern here is that people with low income living in affluent cities with well-educated inhabitants tend to live longer. Some potential explanations of why this is:

- Shared government revenue can cause transfers from rich to poor.
- Culturally, healthier behaviors may be the norm in these areas.
- Self-selection of people who choose to live in expensive cities.

More work is needed to understand the causal effects at work here.
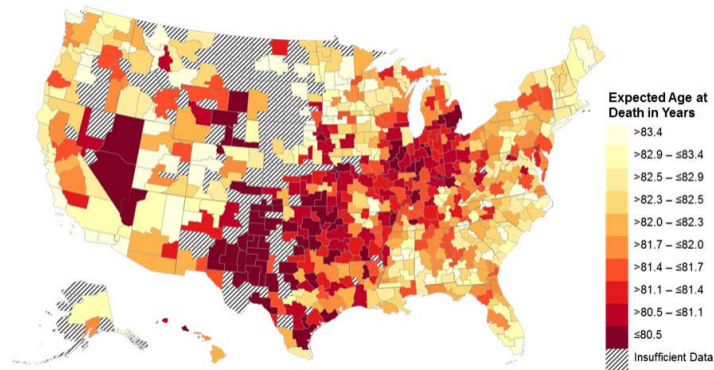
Note: Lighter Colors Represent Areas with Higher Life Expectancy

Figure 3: Life expectancy by region for men

6   Conclusion/Discussion

- While inequality in life expectancy is growing, many areas show the opposite trend. Identifying the causes behind these could help us improve life expectancy on a national scale. Understanding what's going right in Birmingham vs. what's going wrong in Tampa is key to deciding what approaches to take.

- Proposed policy changes, such as raising the Social Security retirement age or Medicare eligibility age as life expectancies grow, need to be carefully considered for how they differentially impact the rich and the poor, namely since most growth in life expectancy has been accruing to the rich.

- Since there is so much local variation, it is likely that many of these policy changes must occur at a local level so that they can be tailored to the specific factors in play. There has been some interesting work in cities including Las Vegas, Tulsa, and Oklahoma City.

- Why does Medicaid not affect life expectancy significantly? Because Medicaid doesn't usually change whether or not people get life-saving treatment, only what they pay for it – Medicaid just reduces financial exposure to poor health. In Oregon, there were more people who needed Medicaid than they had money to provide for, so they randomly assigned coverage to some people. Those with Medicaid had better financial outcomes, but not significantly better heath.
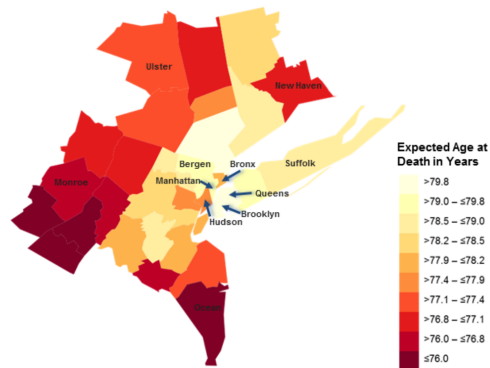
Race-Adjusted Expected Age at Death for 40 Year Old Women
Bottom Quartile of U.S. Income Distribution

Note: Lighter Colors Represent Areas with Higher Life Expectancy

Figure 4: Life expectancy by region for women

Race-Adjusted Expected Age at Death for 40 Year Old Men in Bottom Quartile
By County in the New York Area

Note: Lighter Colors Represent Areas with Higher Life Expectancy

Figure 5: Variation in life expectancy for men in the New York area