

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

Raj Chetty, Stanford

Michael Stepner, MIT

Sarah Abraham, MIT

Shelby Lin, McKinsey

Benjamin Scuderi, Harvard

Nicholas Turner, Office of Tax Analysis

Augustin Bergeron, Harvard

David Cutler, Harvard

The opinions expressed in this paper are those of the authors alone and do not necessarily reflect the views of the Internal Revenue Service, the U.S. Treasury Department, or any other agency of the Federal Government.

Introduction

- Well known that higher income is associated with longer life

[e.g., Kitagawa and Hauser 1973, Pappas et al. 1993, Williams and Collins 1995, Meara et al., Olshansky et al. 2012, Waldron 2007, 2013]

- But several aspects of relationship between income and longevity remain unclear
 1. What is the shape of the income–life expectancy gradient?
 2. How are gaps in life expectancy changing over time?
 3. How do the gaps vary across local areas?
 4. What are the sources of the longevity gap?

This Paper

- We use de-identified data from tax records covering the U.S. population from 1999-2014 to characterize income-mortality gradients
 - 1.4 billion observations → more granular analysis of relationship between income and mortality than in prior work
- Characterize life expectancy by income, over time, and across areas
 - More precise estimates at *national* level than in prior work
 - Large and growing gaps in longevity across income groups
 - New *local area* estimates by income group
 - Substantial variation in level and change in life expectancy across areas, especially for the poor

This Paper

- We also characterize correlates of the spatial variation we document
 - But we do not identify causal mechanisms in this paper
- Focus primarily on constructing publicly available statistics
 - To facilitate future work on mechanisms and to measure progress systematically

Outline

1. Data and Methodology
2. National Statistics on Income and Life Expectancy
3. Local Area Estimates
4. Predictors of Local Area Variation

Part 1: Data and Methodology

Data and Sample Definition

- Income data from de-identified 1999-2014 tax returns
- Mortality data from SSA DM-1 file
 - DM-1 death counts are closely aligned with CDC NCHS counts by year and across age distribution (less than 2% difference)

Income Definition

- Baseline income concept: household earnings
 - For tax filers: Adjusted Gross Income minus Social Security and Disability benefits
 - For non-filers: W-2 earnings + UI benefits
- Exclude individuals with zero household income (8% of population at age 40)
 - Mortality rates for individuals with zero income measured imperfectly because deaths of non-residents are not tracked fully in SSA data
- Focus on percentile **ranks** in income distribution
 - Rank individuals in national income distribution within birth cohort, gender, and tax year

Methodology

- Goal: estimate expected age of death conditional on an individual's income at age 40, controlling for differences in race and ethnicity
 - *Period* life expectancy: life expectancy for a hypothetical individual who experiences mortality rates at each age observed in a cross-section
- Straightforward to compute if one could observe mortality rates at all ages for all racial groups conditional on income at age 40
- Two missing data problems:
 1. Mortality rates conditional on income at age 40 unobserved at age > 55
 2. Race and ethnicity not observed in tax data

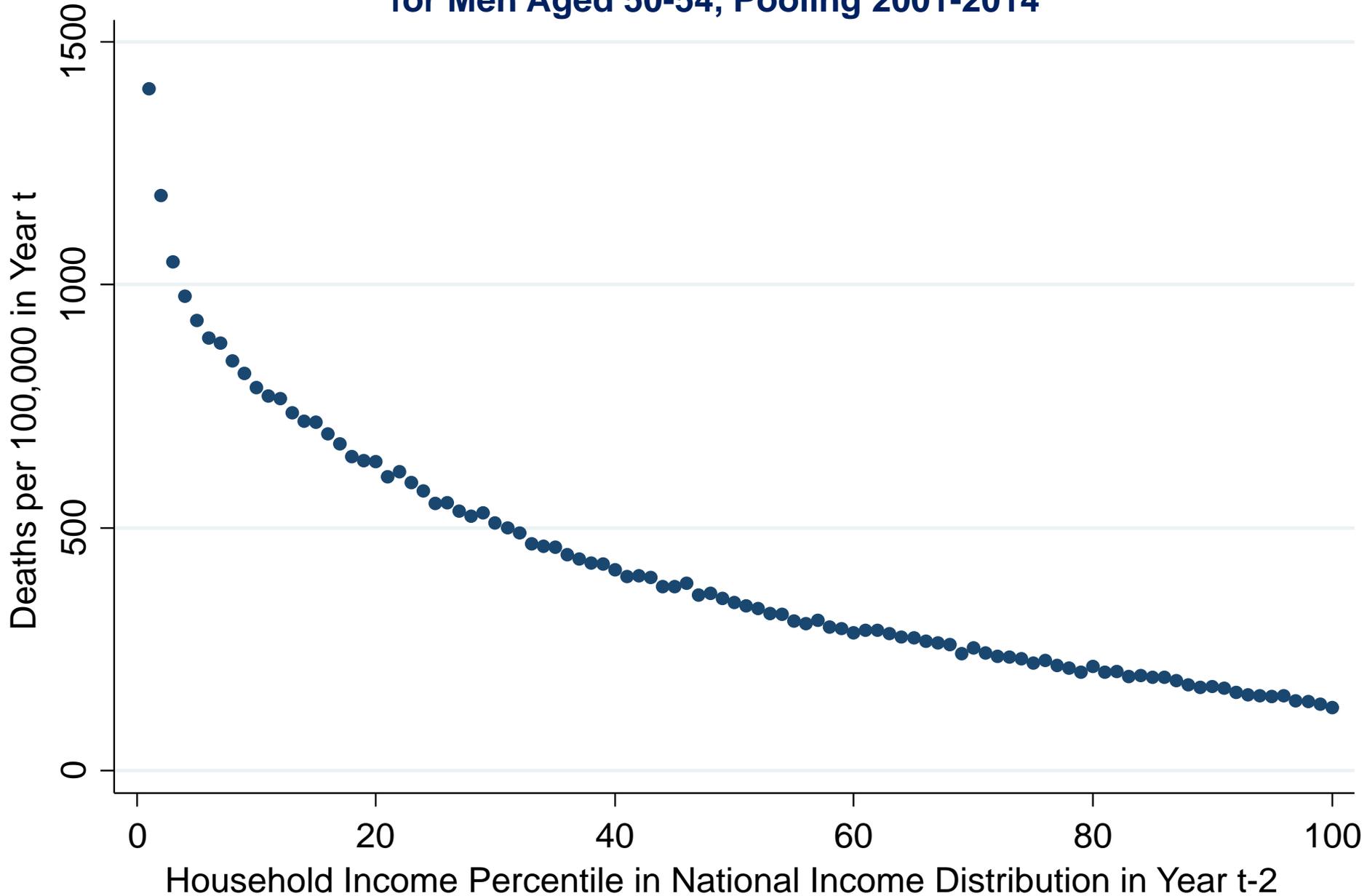
Methodology

- Three steps to estimate life expectancy by income group:
 1. Calculate mortality rates by income rank and age for available ages
 2. Use age profile of mortality rates to estimate Gompertz models
 3. Adjust for racial differences in mortality rates

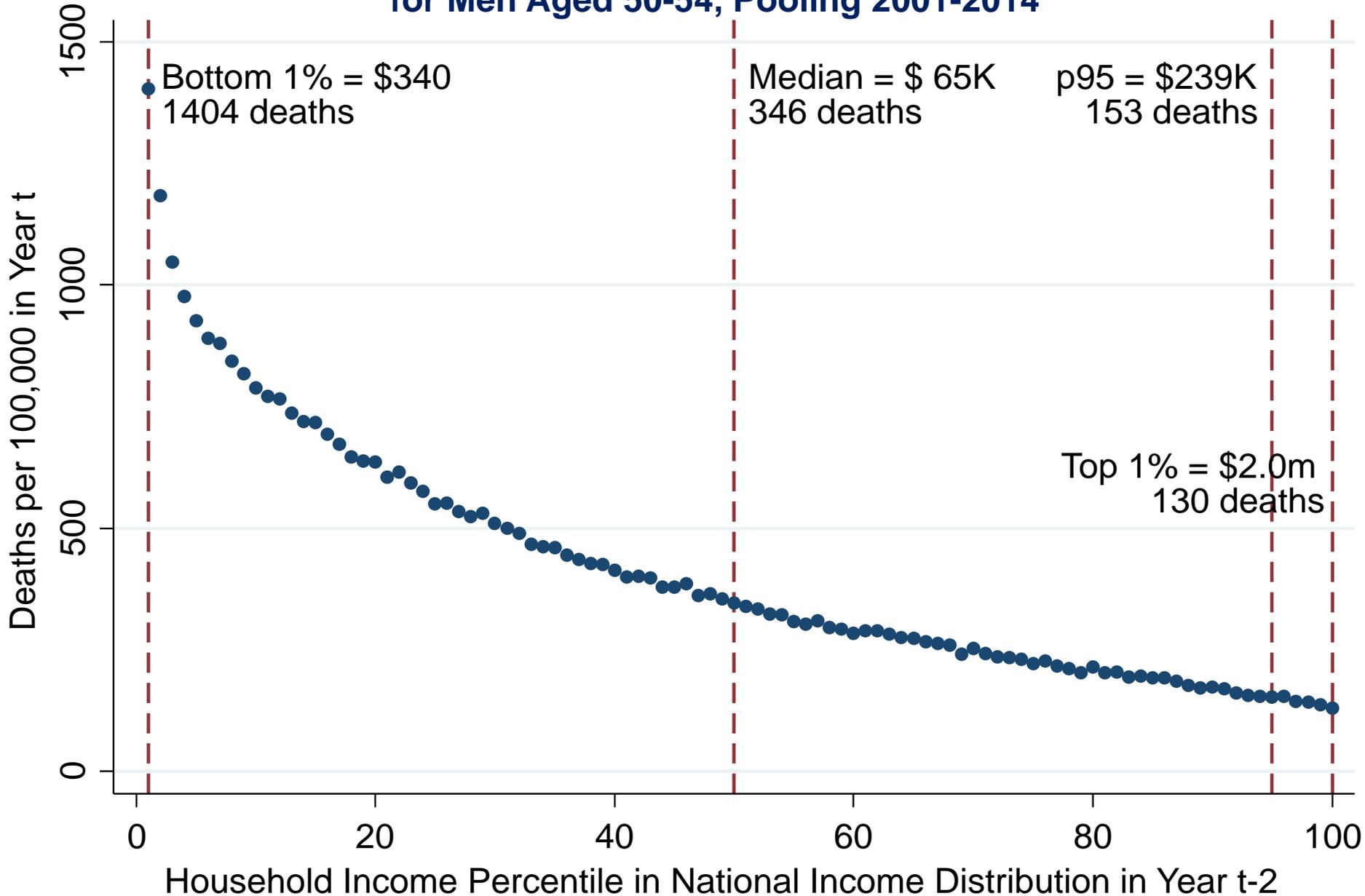
Step 1: Calculating Observed Mortality Rates

- For “working age” sample (below age 63), start by calculating mortality rates as a function of income percentile at age $a - 2$ (two year lag)
- Then return to original goal of estimating mortality rates as a function of income percentile at age 40

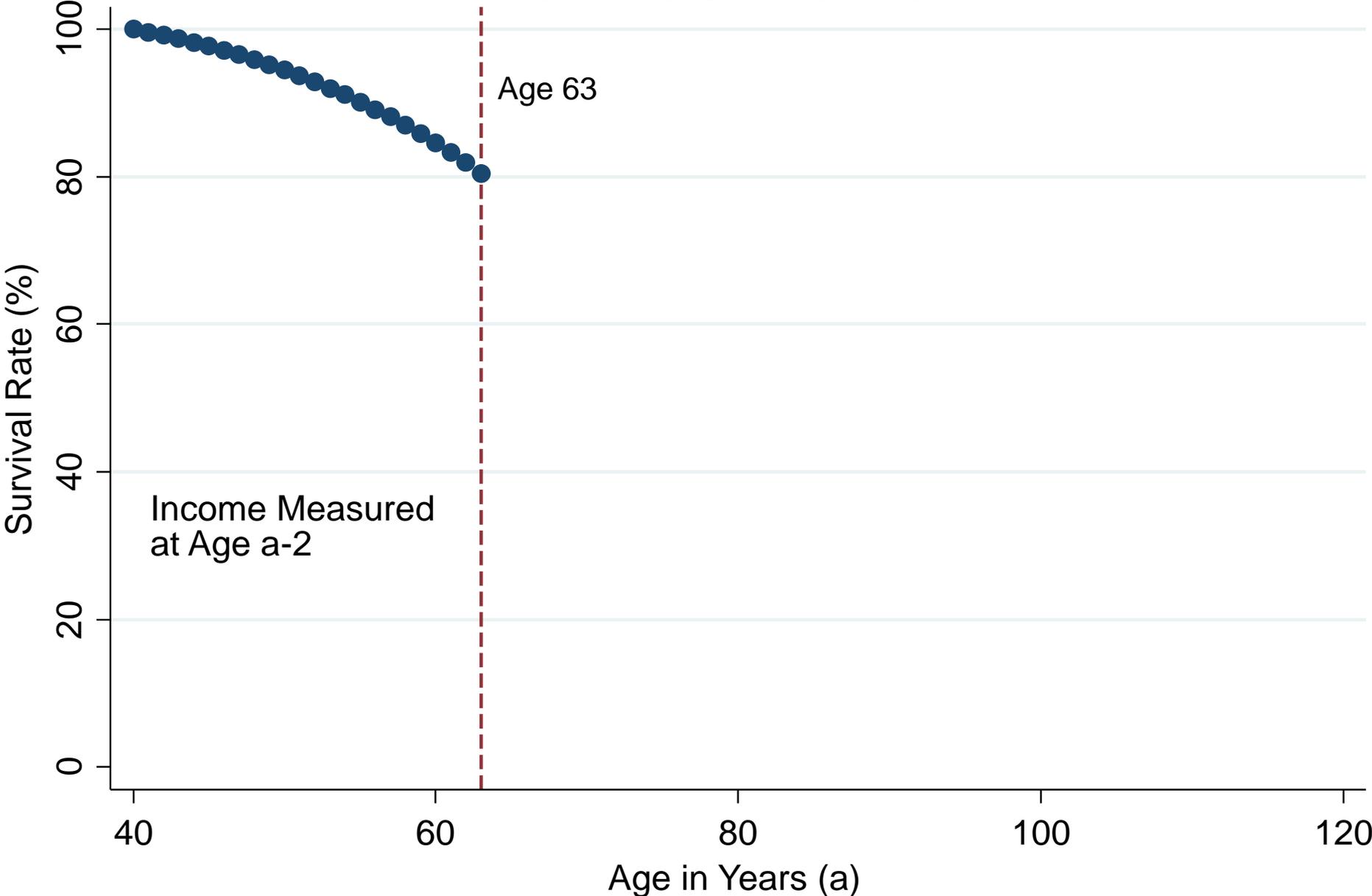
Annual Mortality Rates vs. Household Income Percentile for Men Aged 50-54, Pooling 2001-2014



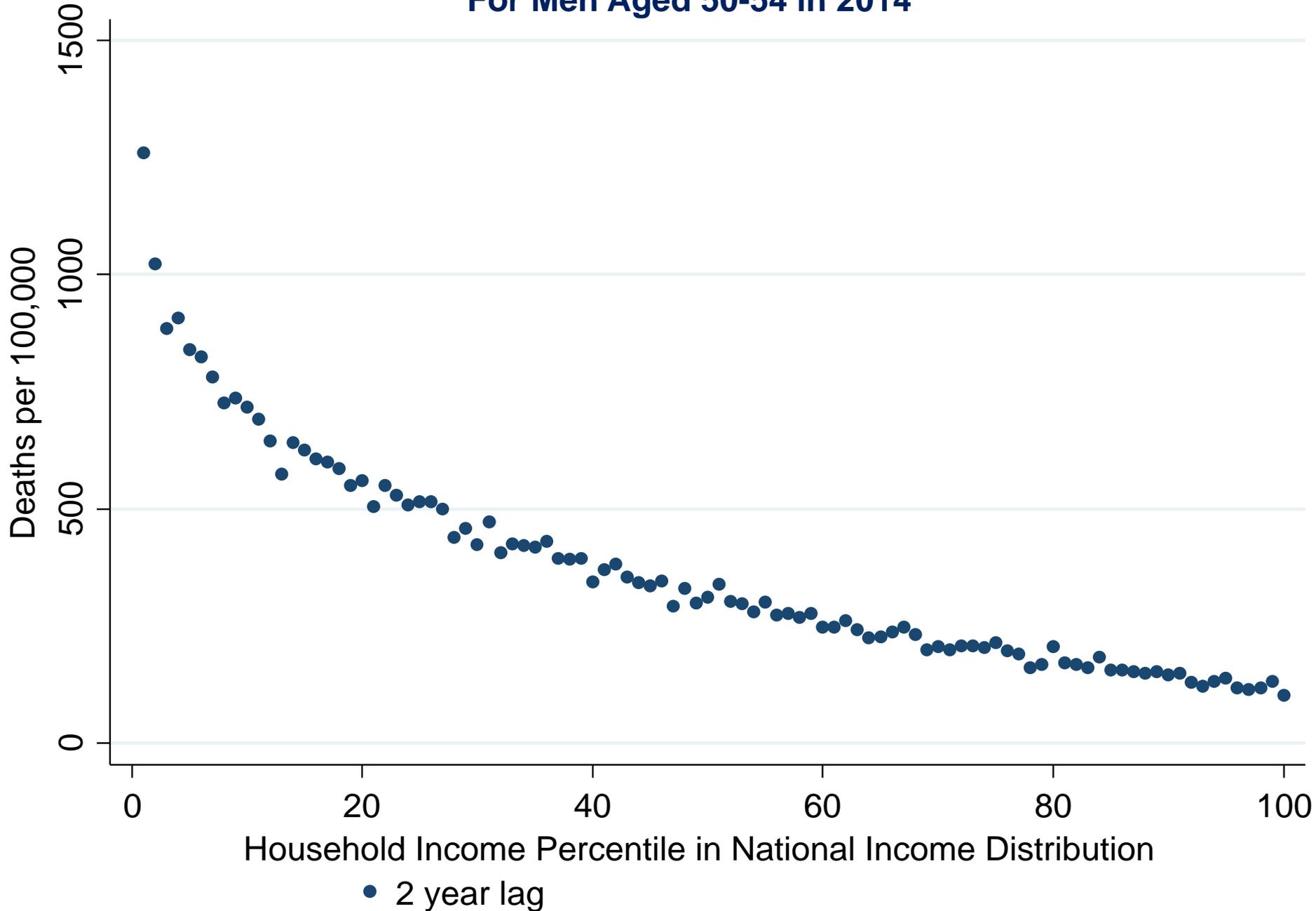
Annual Mortality Rates vs. Household Income Percentile for Men Aged 50-54, Pooling 2001-2014



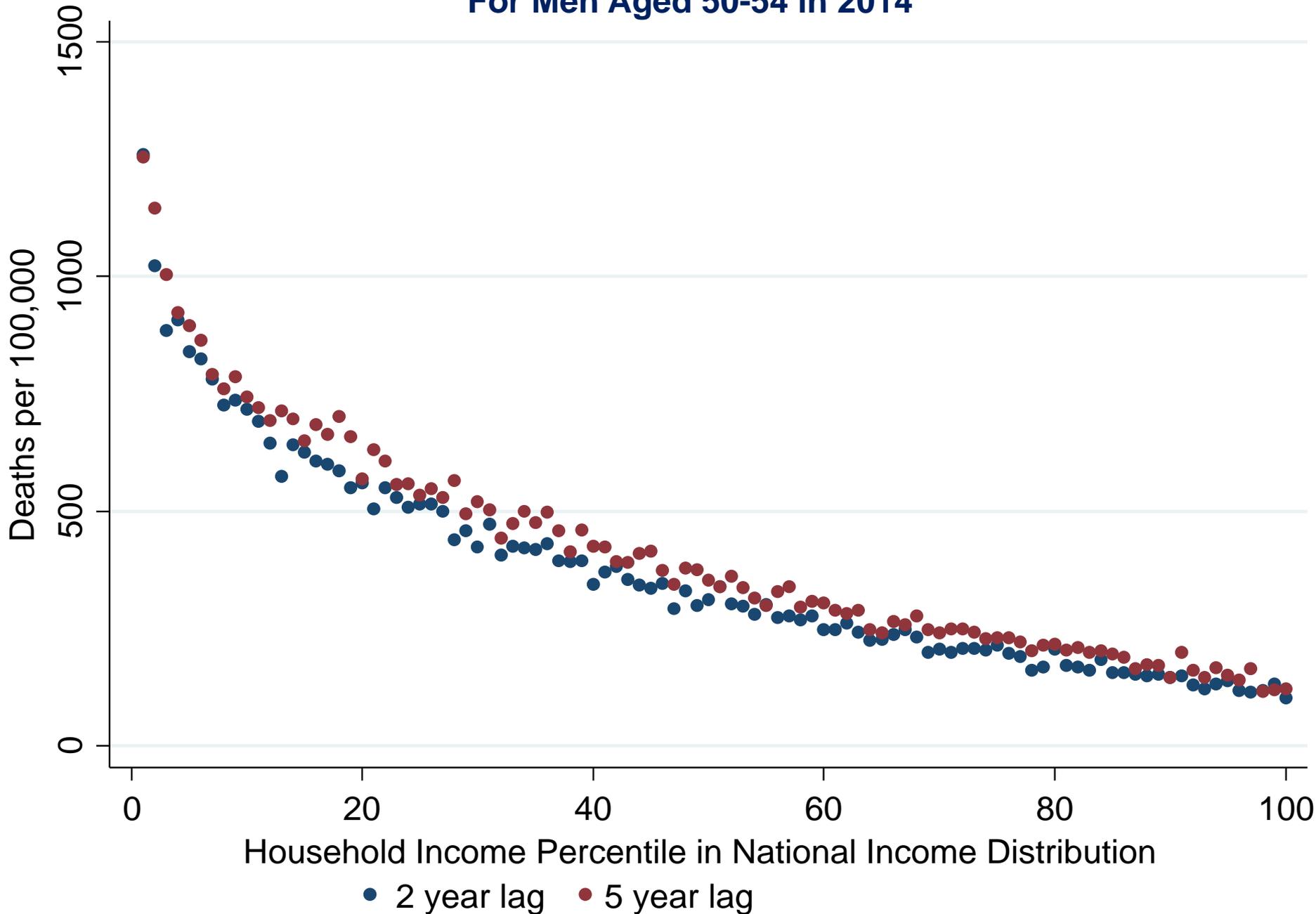
Survival Curve Using Period Life Table For Men at 5th Percentile



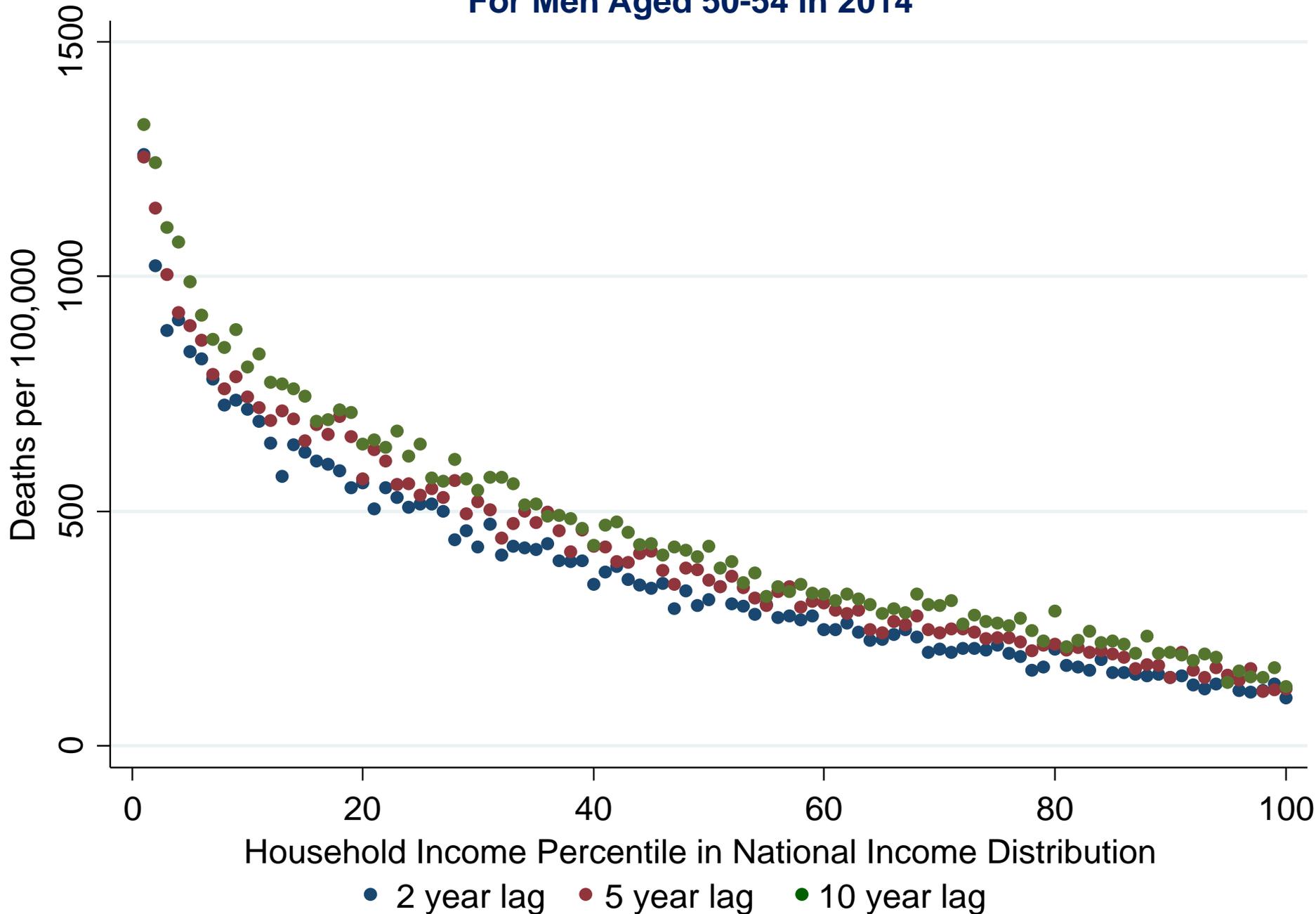
Annual Mortality Rates vs. Household Income Percentile For Men Aged 50-54 in 2014



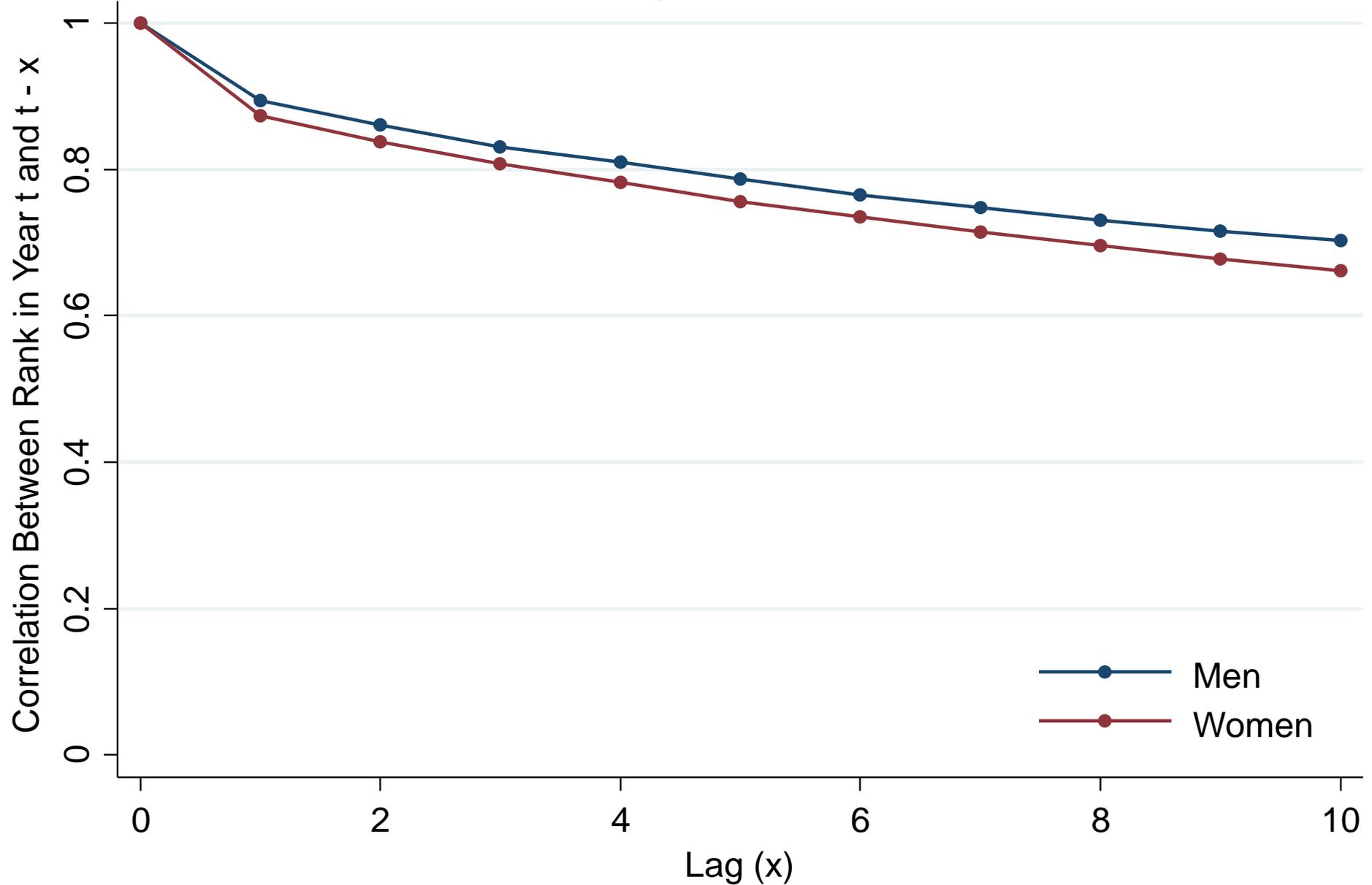
Annual Mortality Rates vs. Household Income Percentile For Men Aged 50-54 in 2014



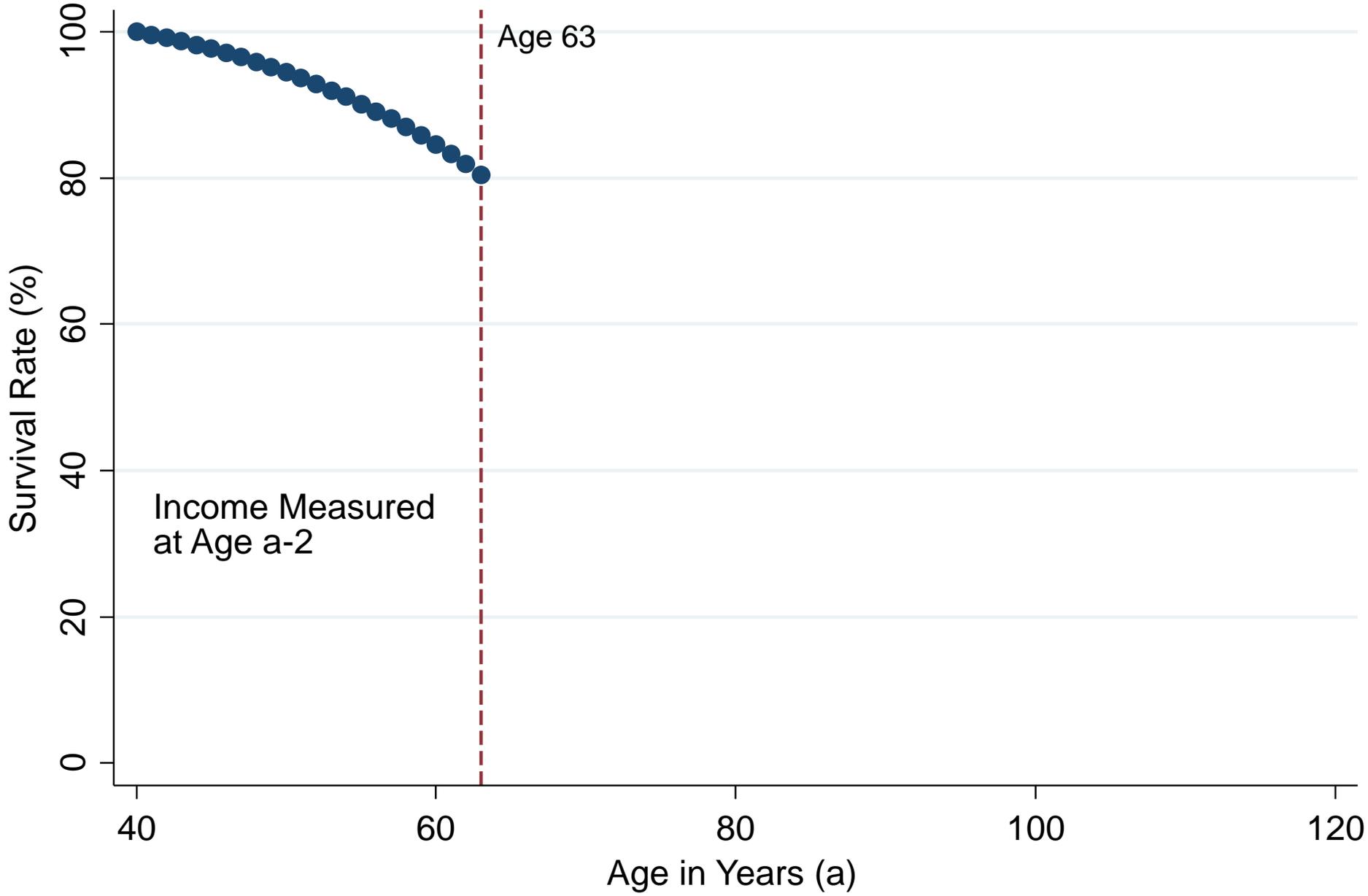
Annual Mortality Rates vs. Household Income Percentile For Men Aged 50-54 in 2014



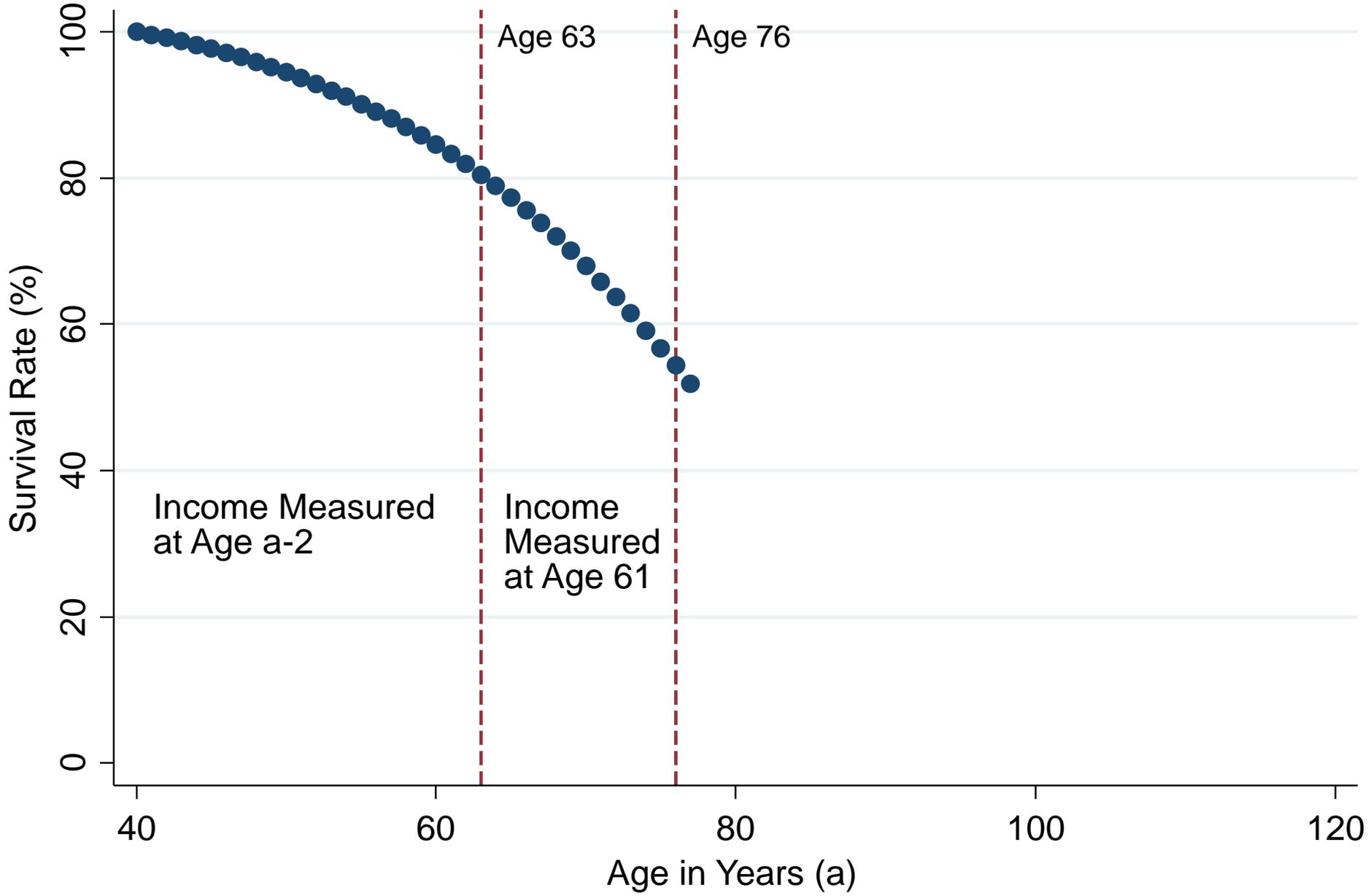
Correlation of Current Income Percentile with Lagged Percentiles by Gender



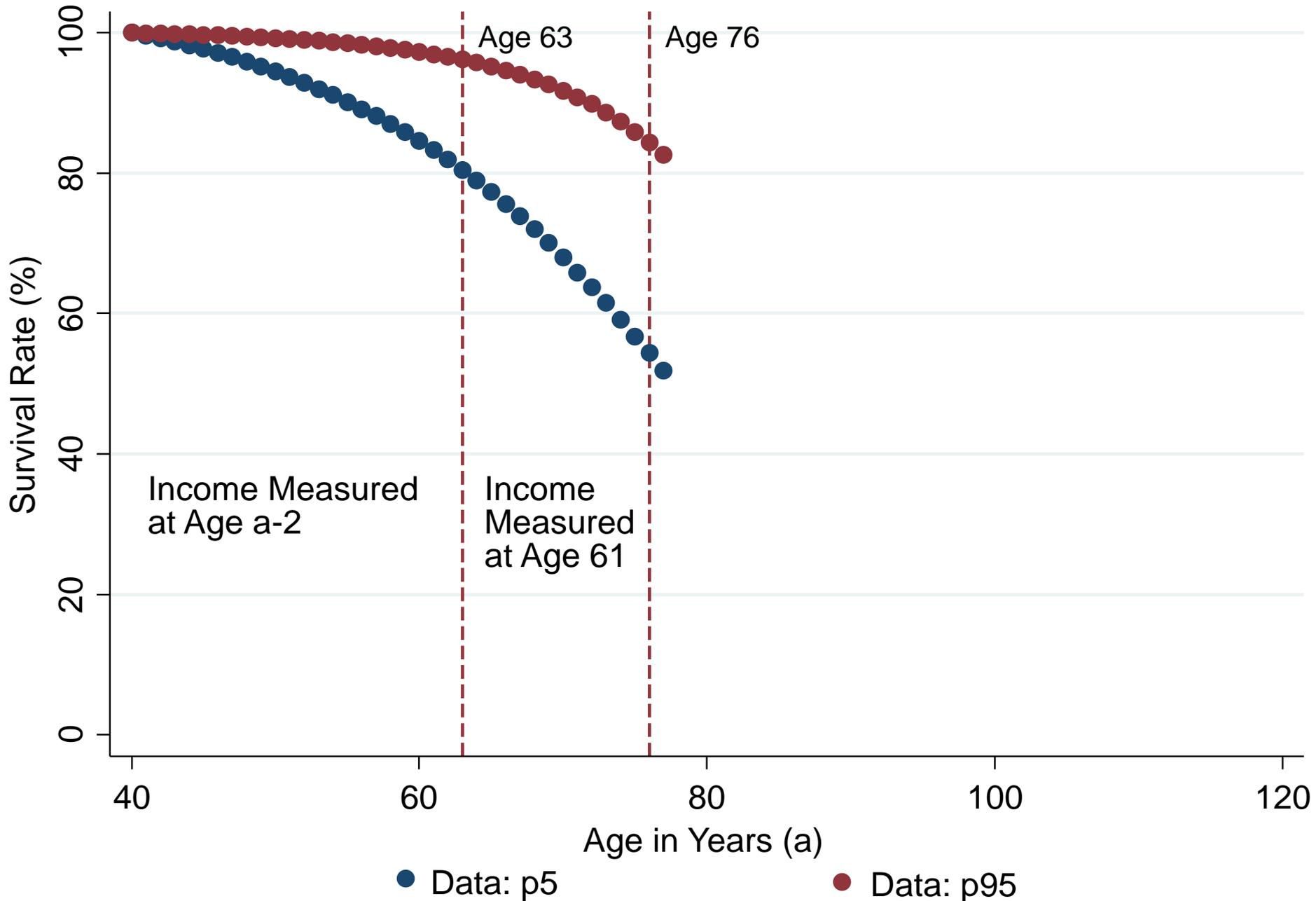
Survival Curve for Men at 5th Percentile



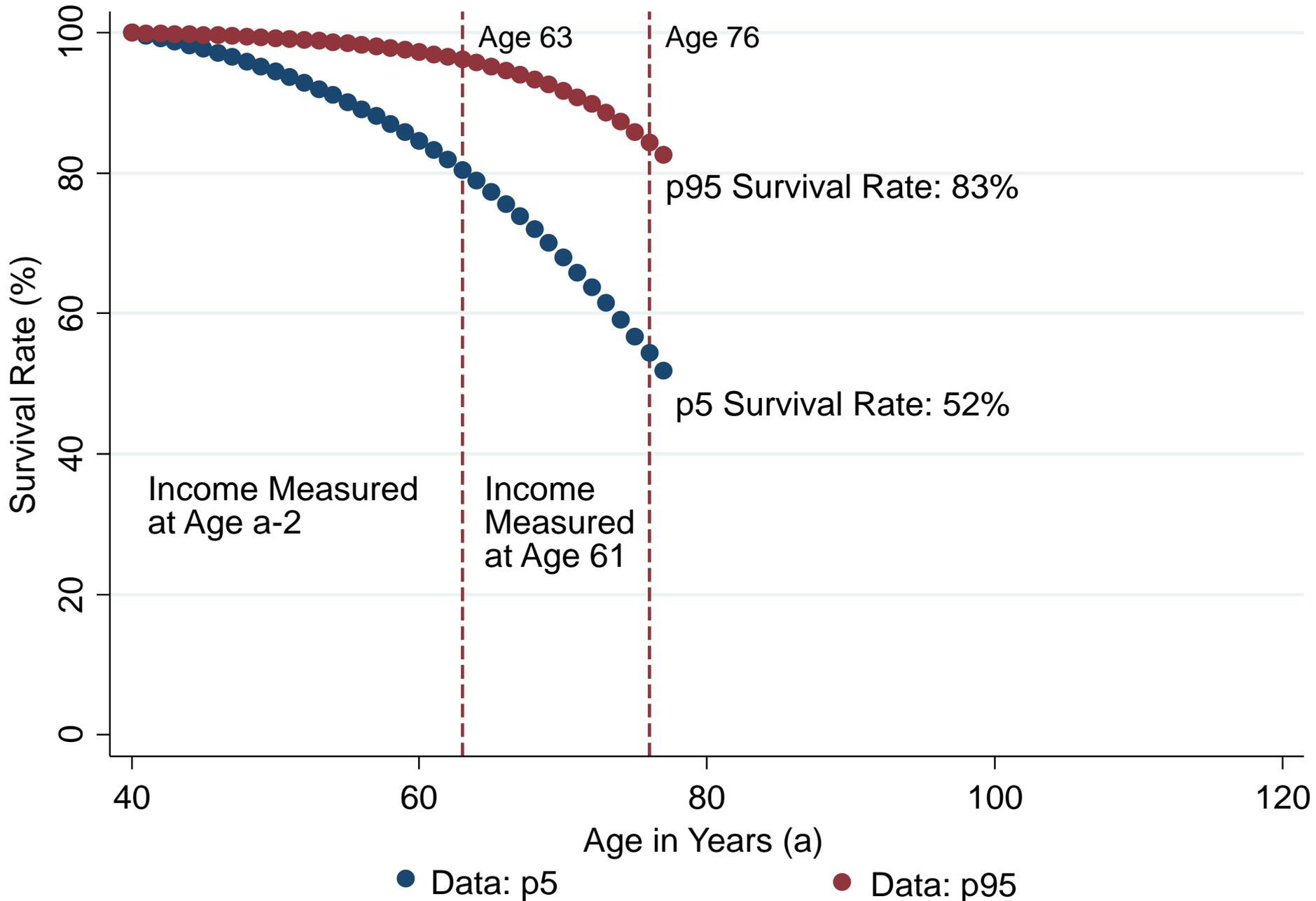
Survival Curve for Men at 5th Percentile



Survival Curves for Men at 5th and 95th Percentiles



Survival Curves for Men at 5th and 95th Percentiles

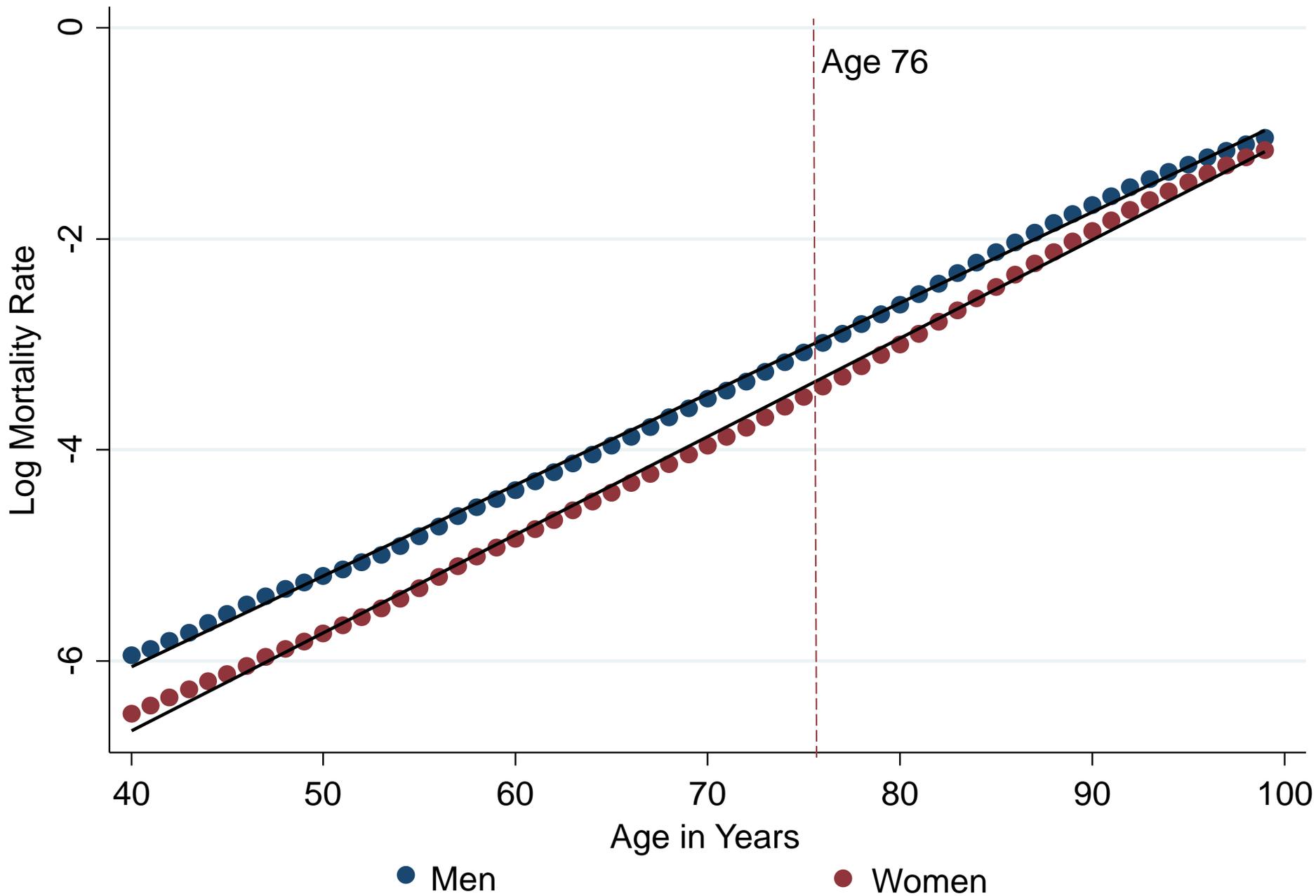


Step 2: Predicting Mortality Rates at Older Ages

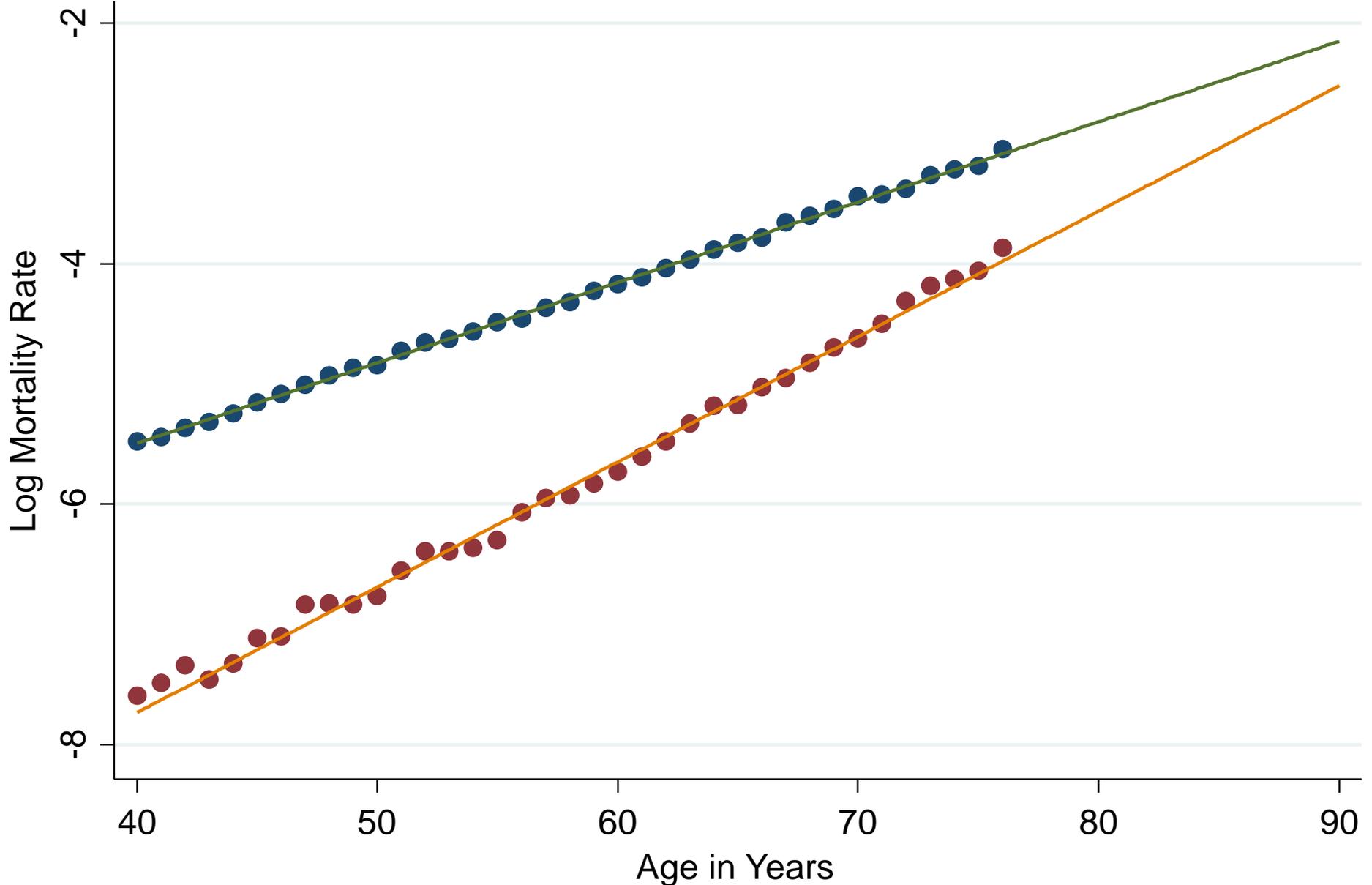
- To calculate life expectancy, need estimates of mortality rates beyond age 76
- Gompertz (1825) documented a robust empirical regularity: mortality rates grow exponentially with age

$$m(a) = k e^{\beta a}$$
$$\Rightarrow \log m(a) = \kappa + \beta a$$

CDC NCHS Mortality Rates by Gender in the United States in 2001

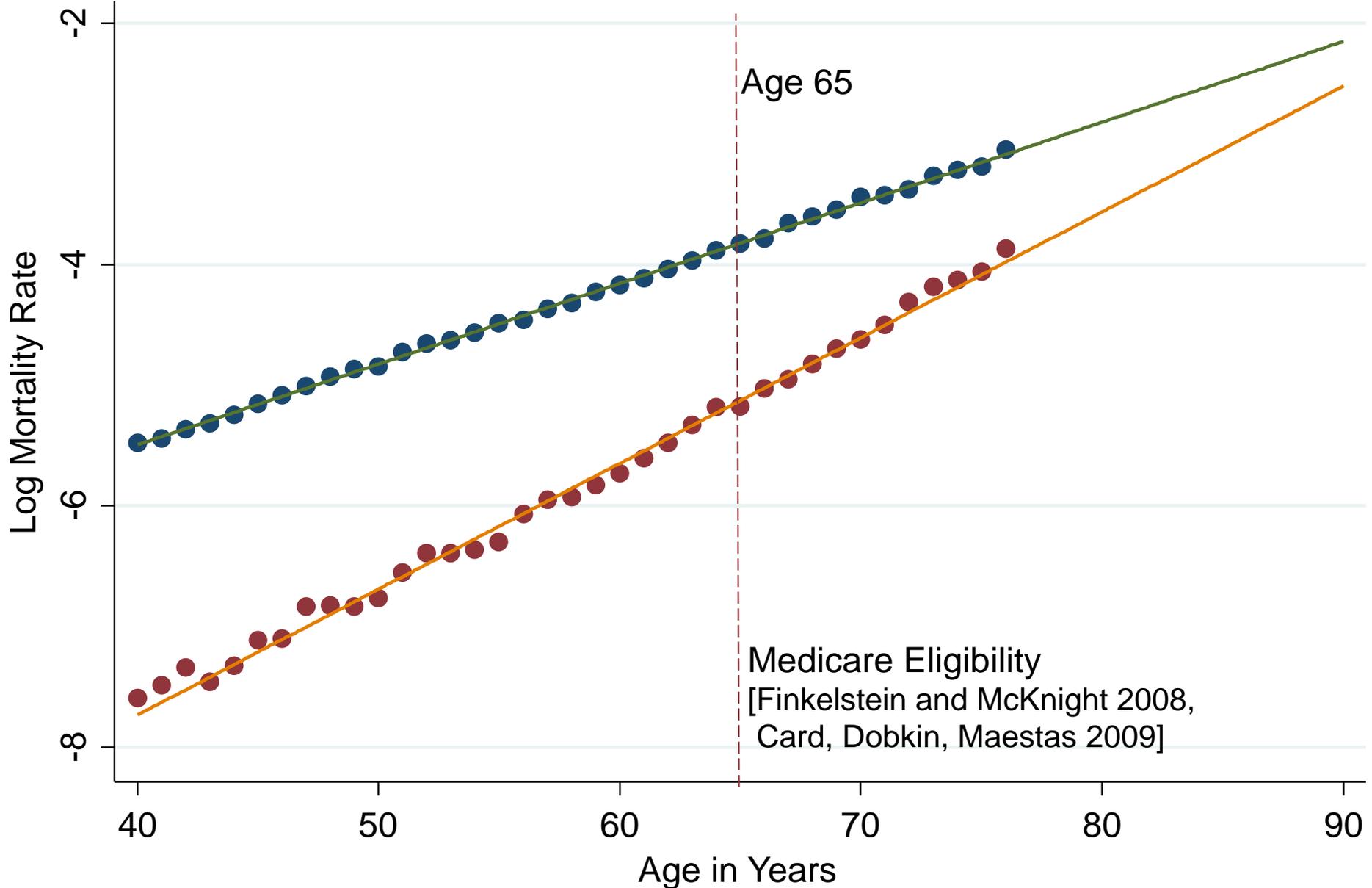


Log Mortality Rates For Men at 5th and 95th Percentiles



● Data: p5 — Gompertz: p5 ● Data: p95 — Gompertz: p95

Log Mortality Rates For Men at 5th and 95th Percentiles

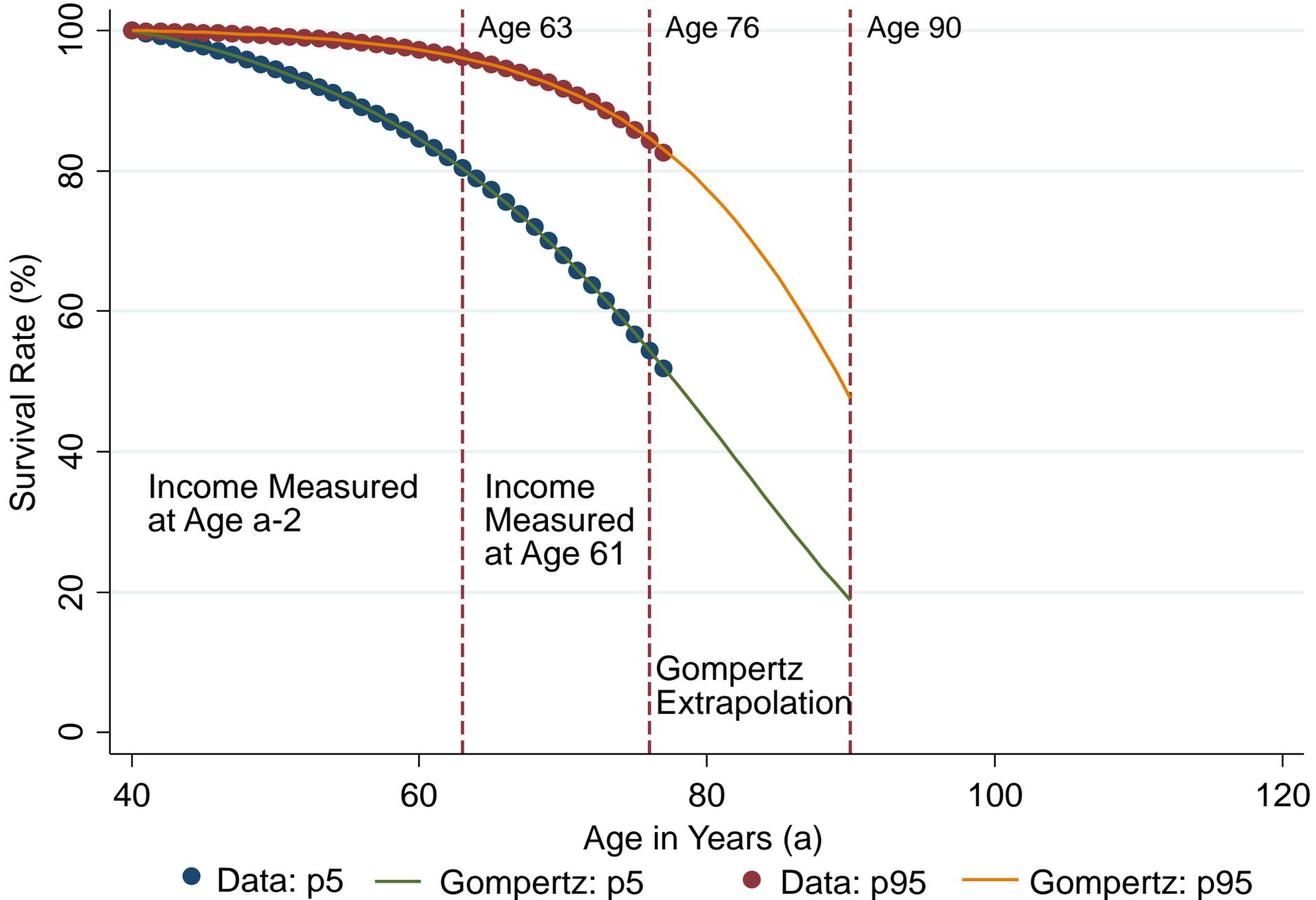


Age 65

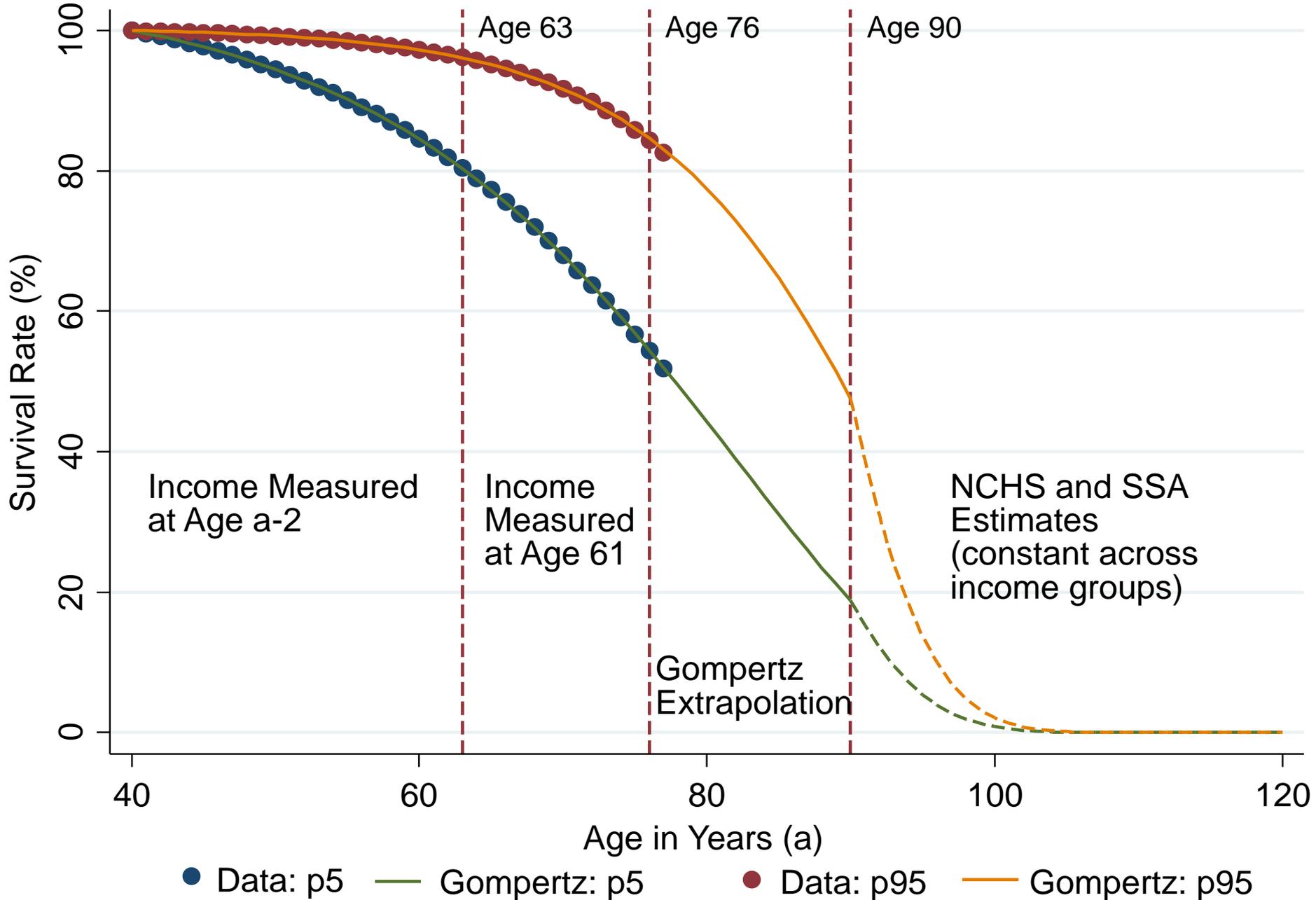
Medicare Eligibility
[Finkelstein and McKnight 2008,
Card, Dobkin, Maestas 2009]

● Data: p5 — Gompertz: p5 ● Data: p95 — Gompertz: p95

Survival Curves for Men at 5th and 95th Percentiles



Survival Curves for Men at 5th and 95th Percentiles



Step 3: Race and Ethnicity Adjustment

- Final step: adjust for racial and ethnic differences in life expectancy
 - CDC statistics show that for males, life exp. 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, potentially biasing raw comparisons
- Perform race (and ethnicity) adjustment to answer the question:

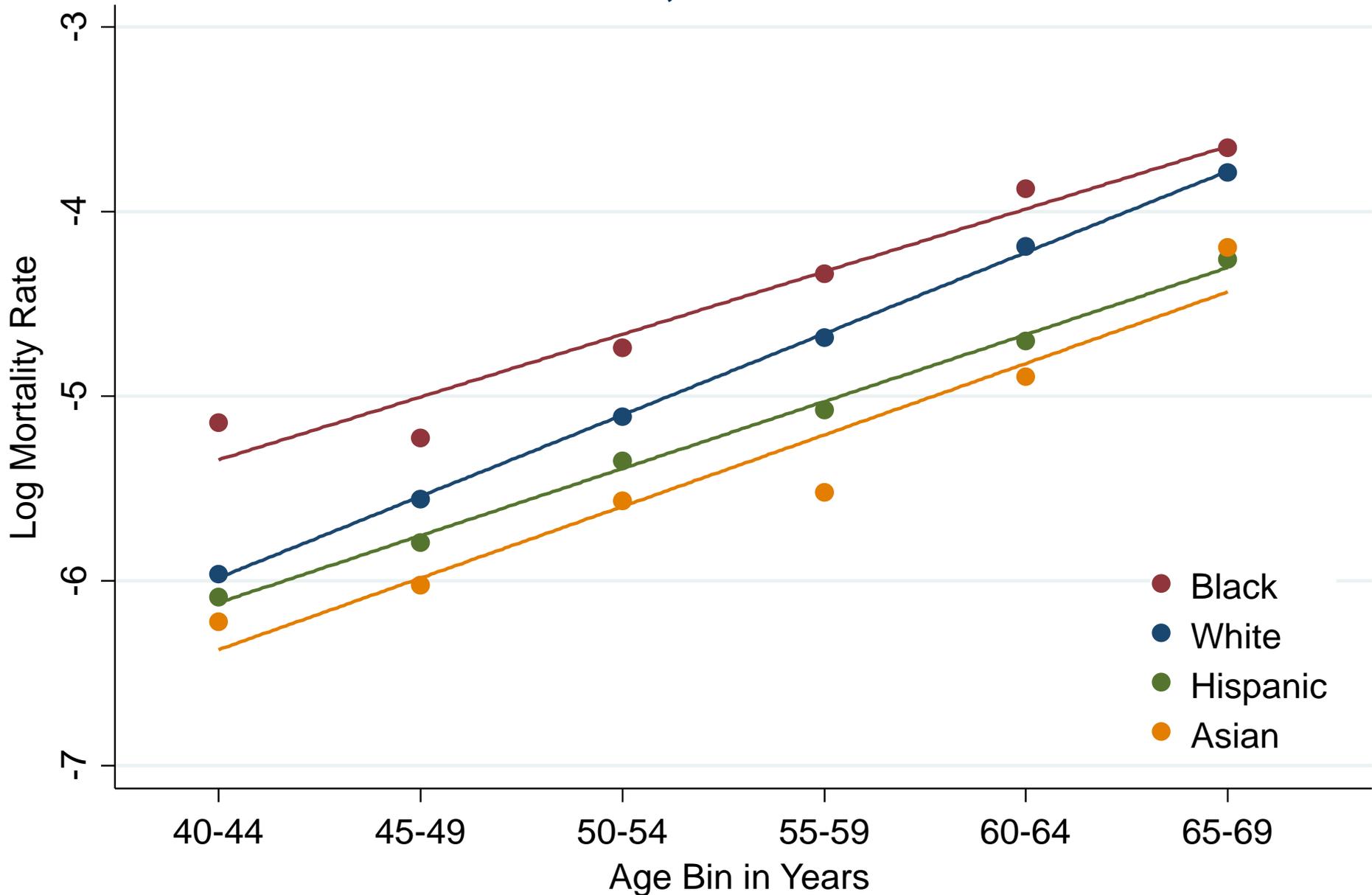
*“What would life expectancy be if each **income group** and **area** had the same black, Hispanic and Asian shares as the U.S. population as a whole at age 40?”*

Race and Ethnicity Adjustment

*“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?”*

- Construct race-adjusted measures of life expectancy in four steps:
 1. Estimate differences in mortality by race controlling for income using data from National Longitudinal Mortality Study
 - Assume racial differences do not vary across areas

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



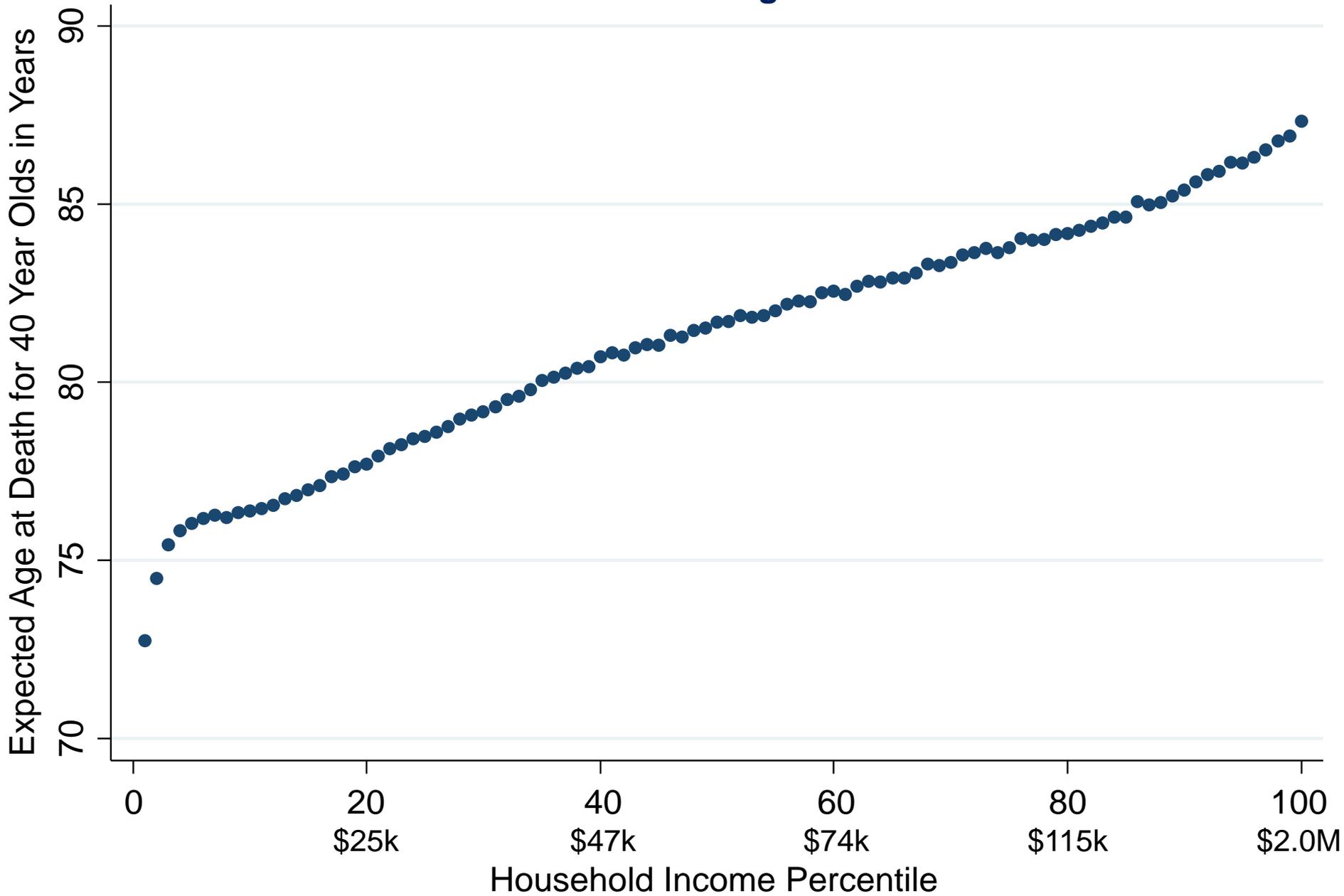
Race and Ethnicity Adjustment

*“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?”*

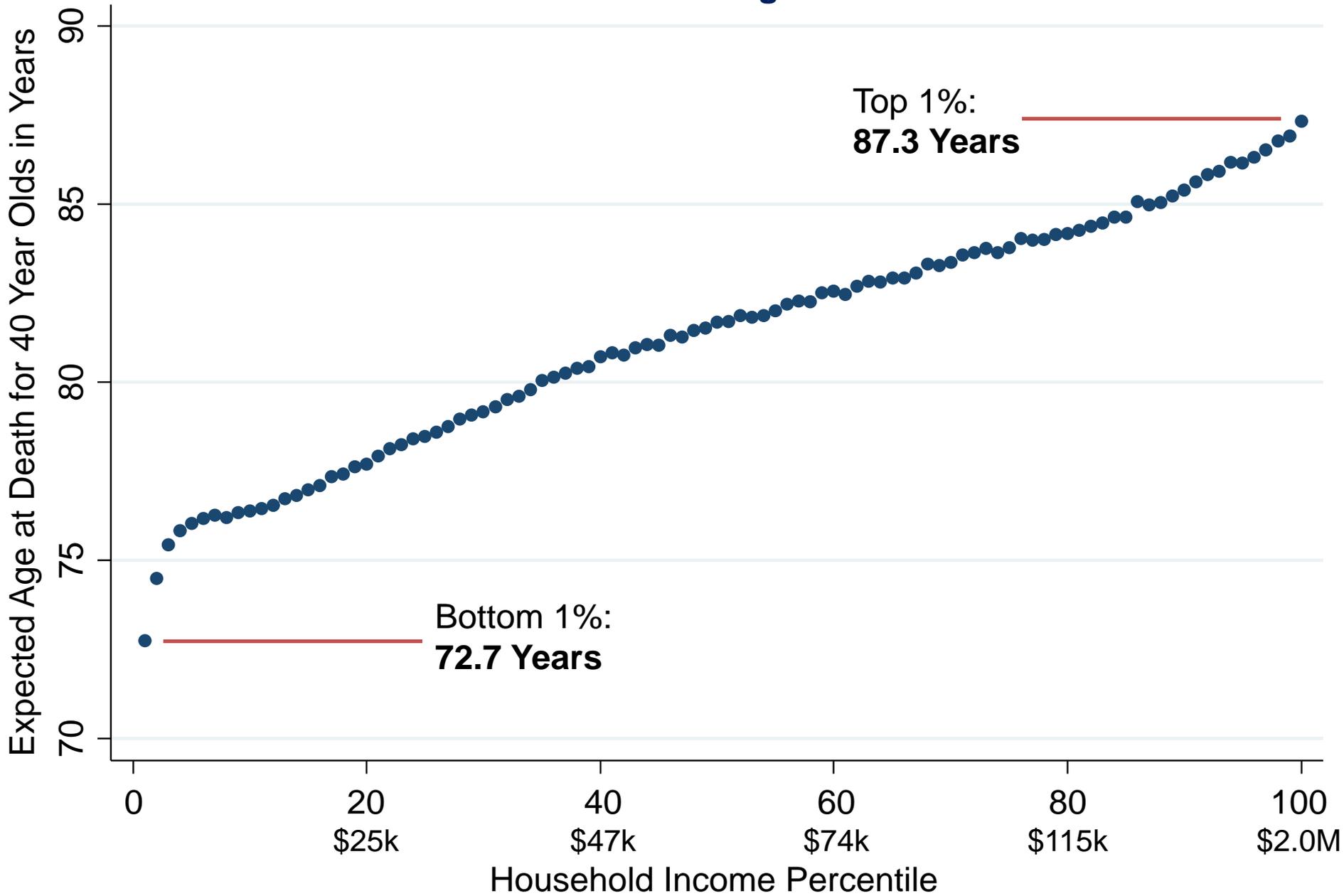
- Construct race-adjusted measures of life expectancy in four steps:
 1. Estimate differences in mortality by race controlling for income using data from National Longitudinal Mortality Study
 2. Estimate racial demographics in each income group and area using Census data
 3. Recover mortality rates by race in each income group and area from aggregate rates in tax data and race differences from NLMS
 4. Calculate life expectancy that would prevail if racial demographics were the same as the national demographics at age 40 (for men, 12% black, 12% Hispanic, 4% Asian)

Part 2: National Statistics on Income and Life Expectancy

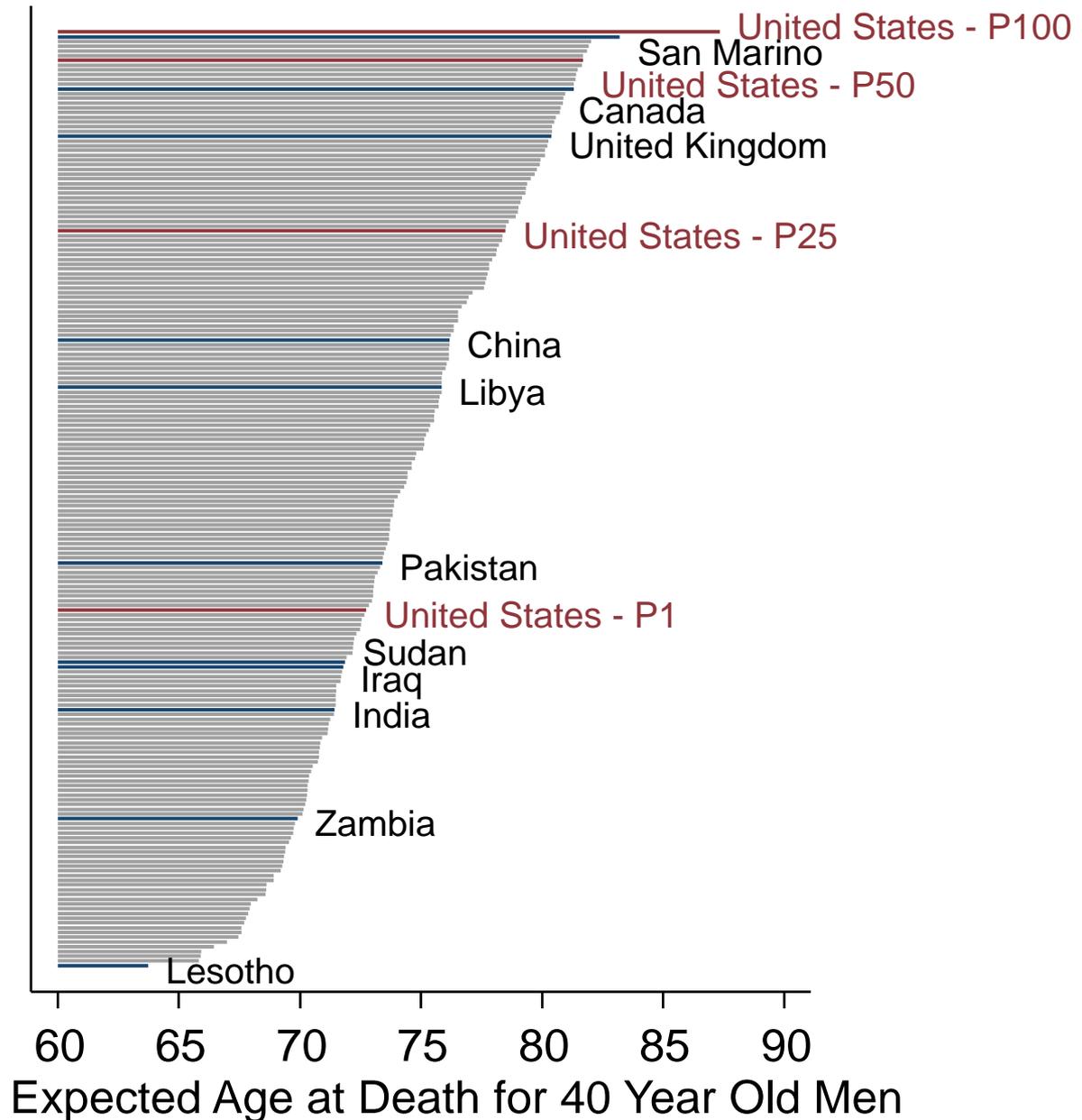
Expected Age at Death vs. Household Income Percentile For Men at Age 40



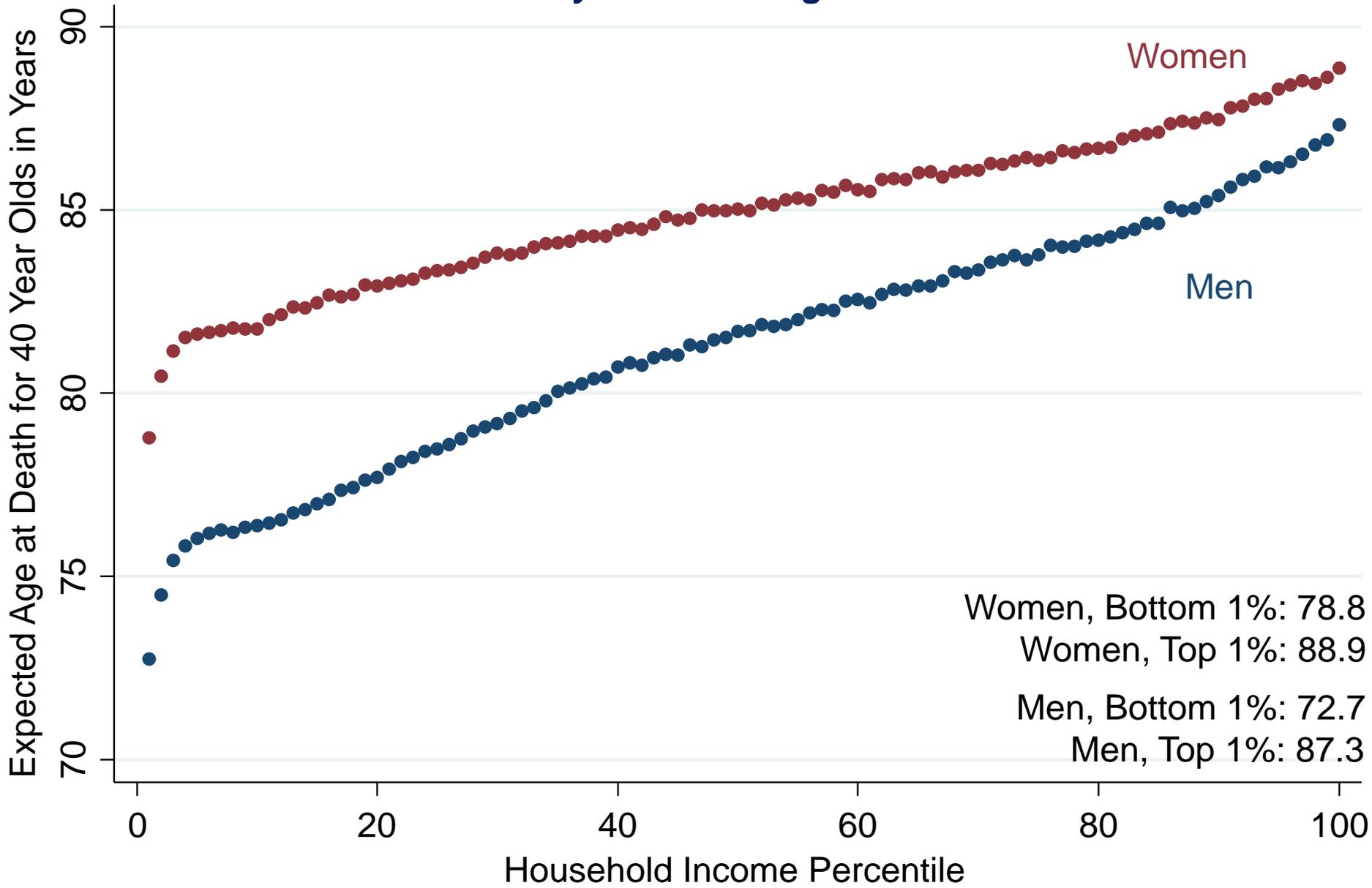
Expected Age at Death vs. Household Income Percentile For Men at Age 40



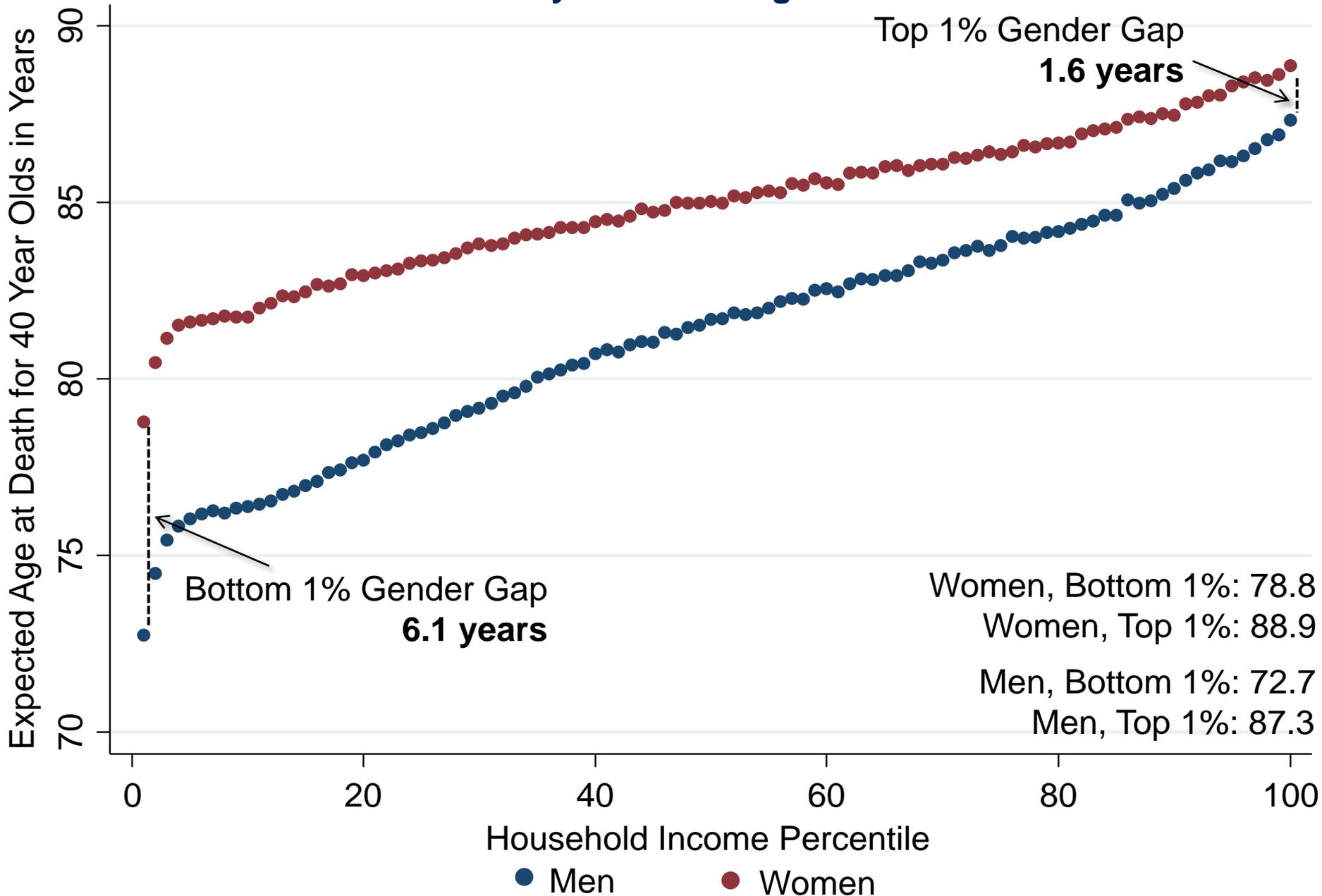
U.S. Life Expectancies by Percentile in Comparison to Mean Life Expectancies Across Countries



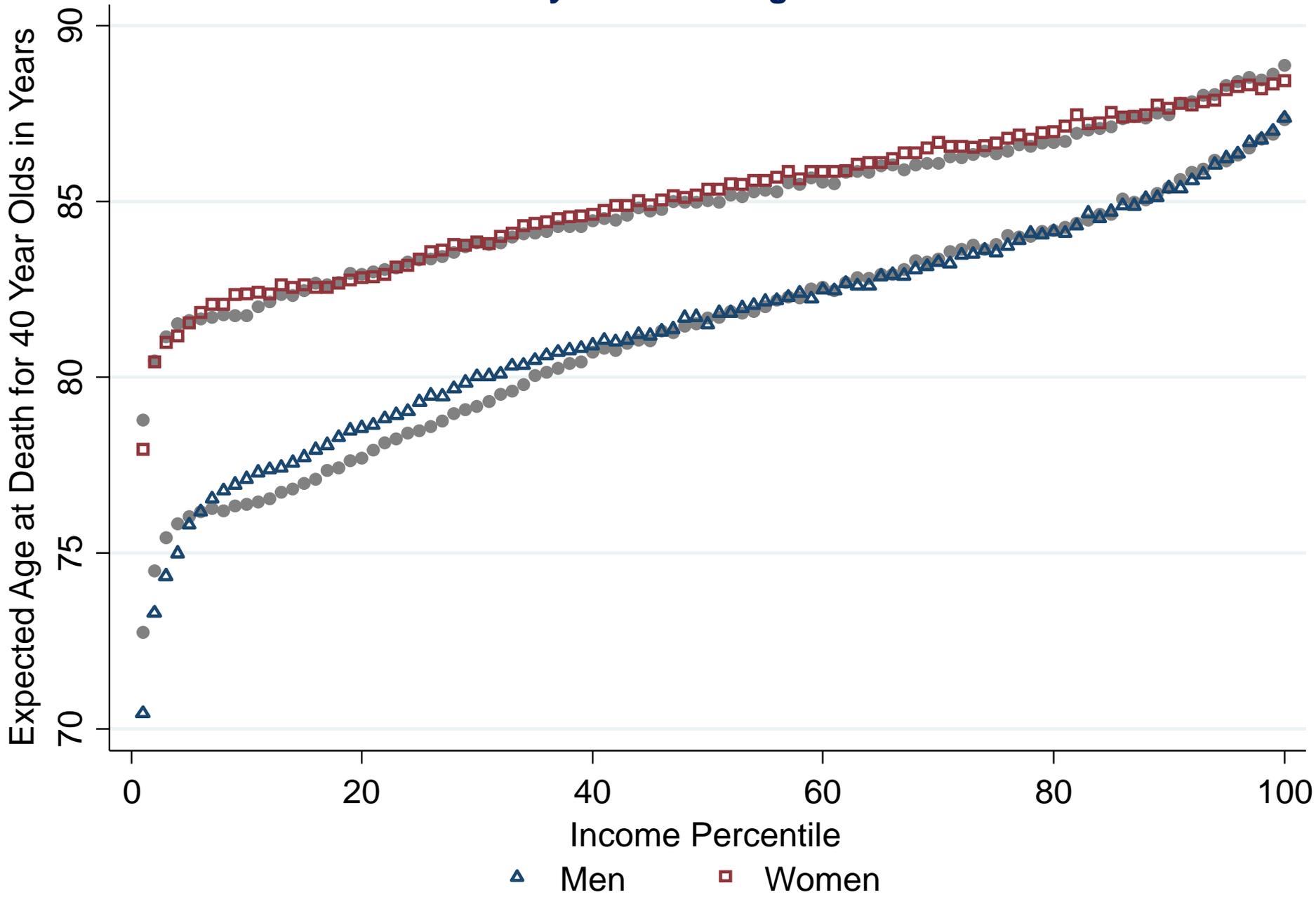
Expected Age at Death vs. Household Income Percentile By Gender at Age 40



Expected Age at Death vs. Household Income Percentile By Gender at Age 40



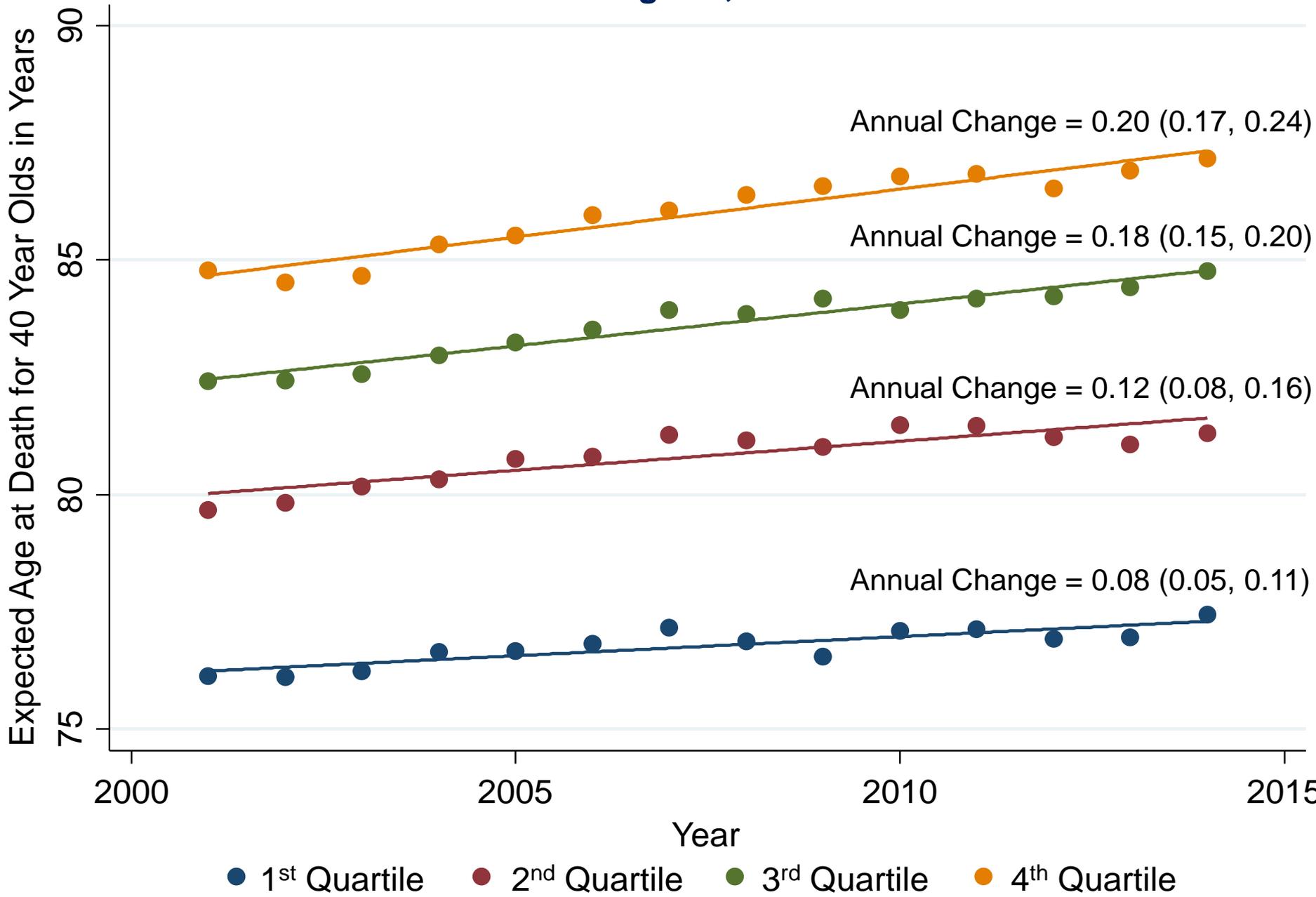
Expected Age at Death vs. Individual Income Percentile By Gender at Age 40



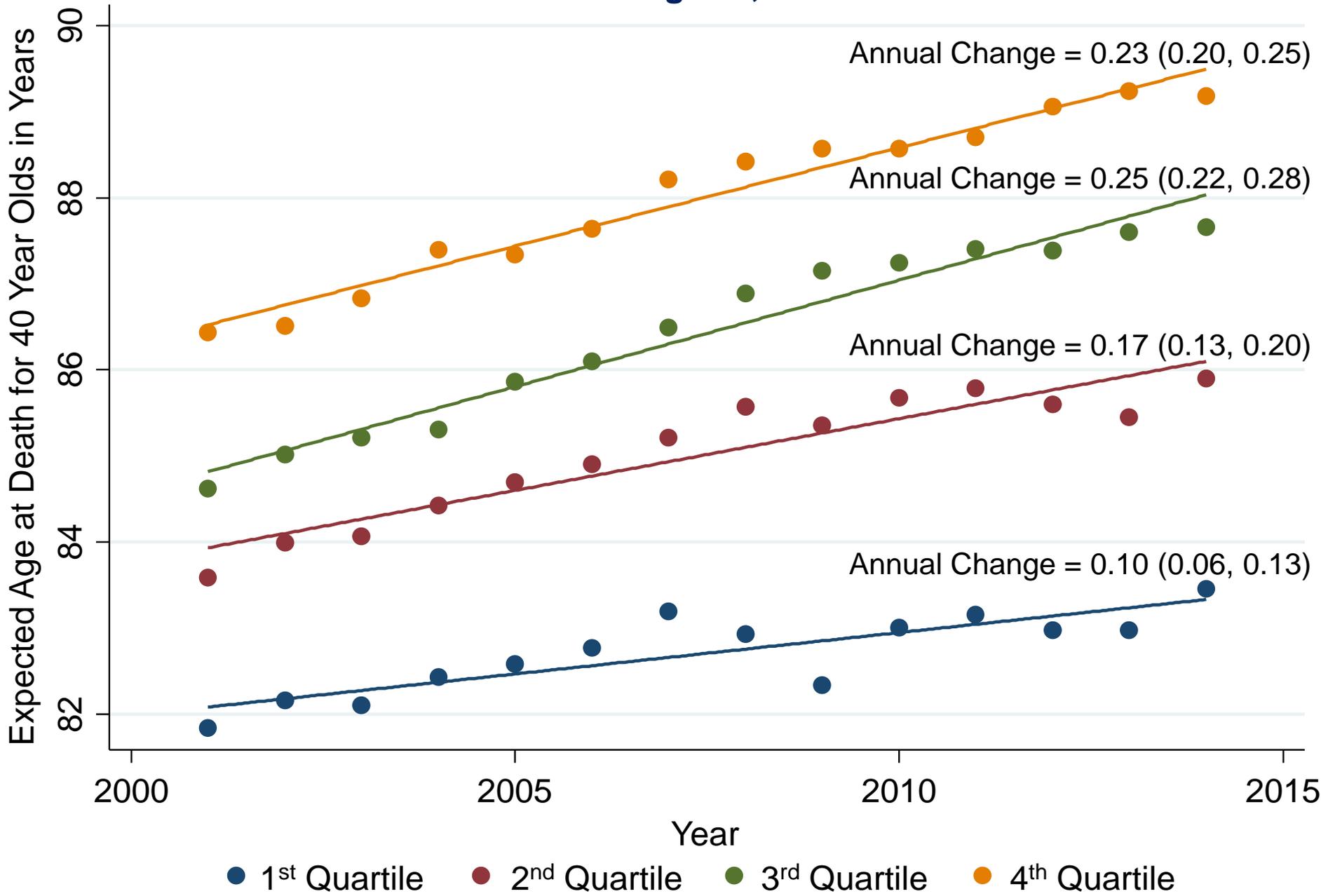
Time Trends

- How are gaps in life expectancy changing over time?
 - Relevant for understanding distributional consequences of various policies, e.g. increasing age of eligibility for social security
- Some studies have found that gap between low- and high-SES groups has grown [Waldron 2007, Meara et al. 2008, Goldring et al. 2015]
 - Some evidence of *declining* life expectancy for low-SES subgroups, but results debated [Olshansky et al. 2012, Bound et al 2015]

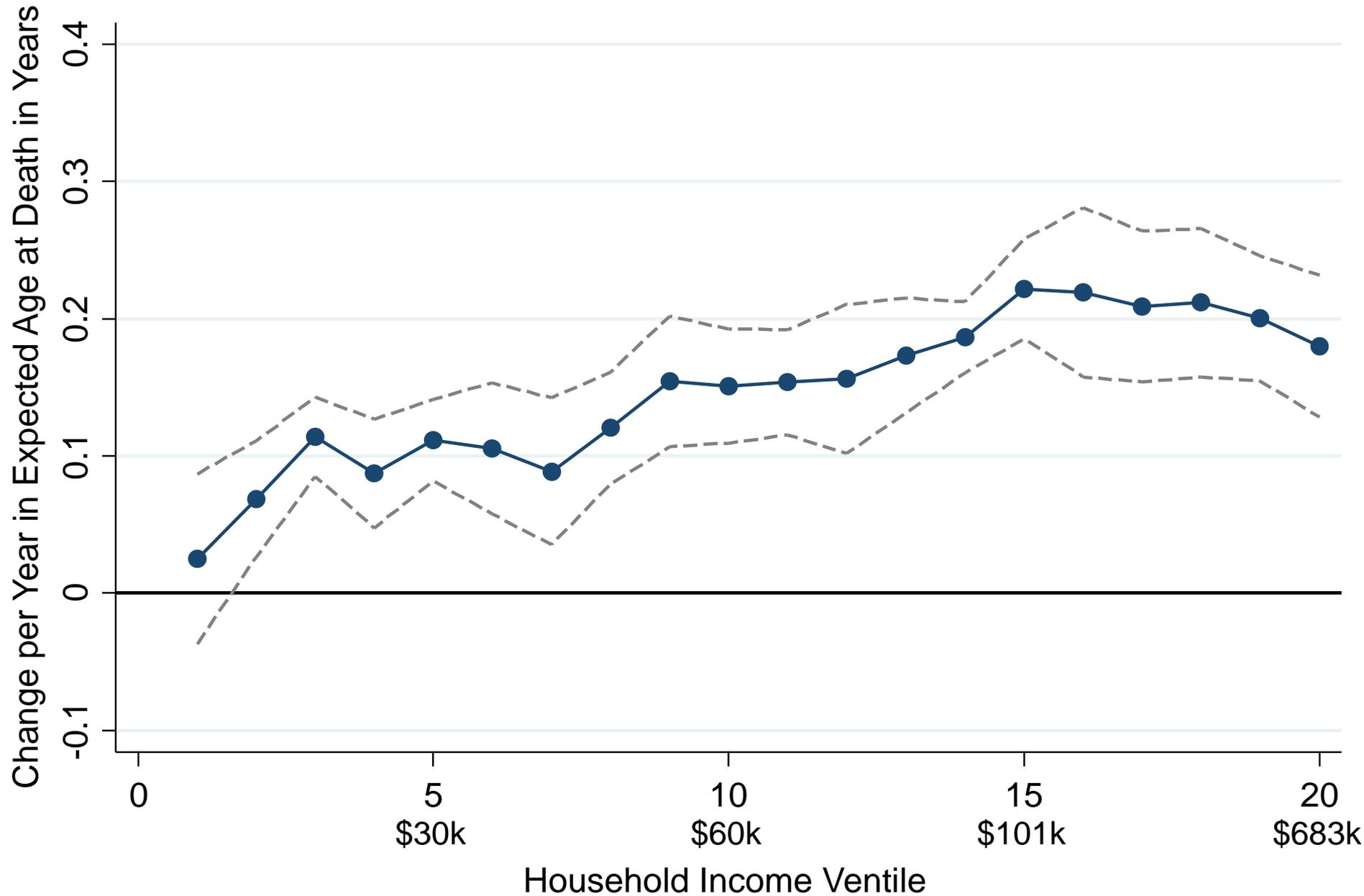
Trends in Expected Age at Death by Income Quartile in the United States For Men Age 40, 2001-2014



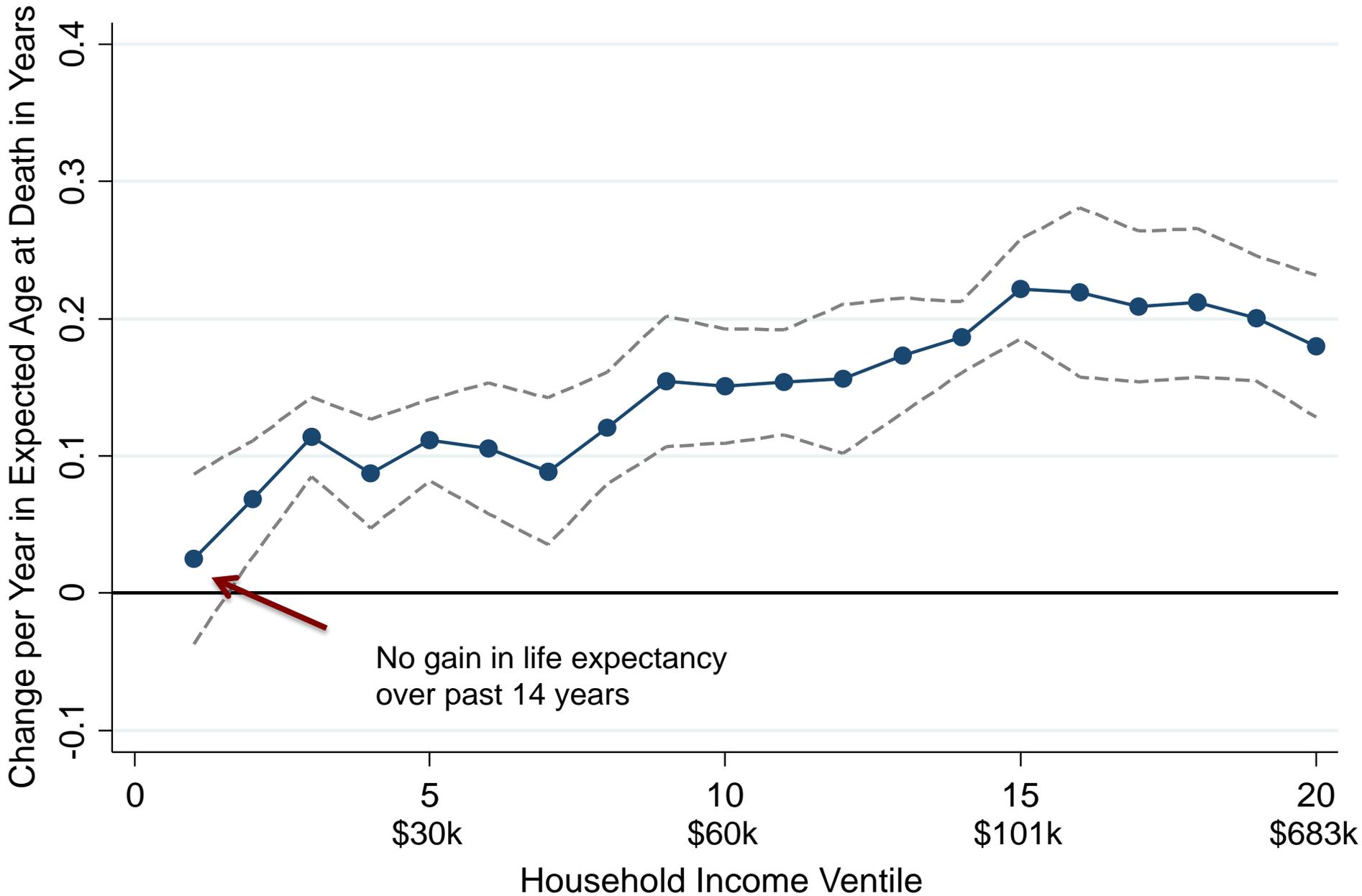
Trends in Expected Age at Death by Income Quartile in the United States For Women Age 40, 2001-2014



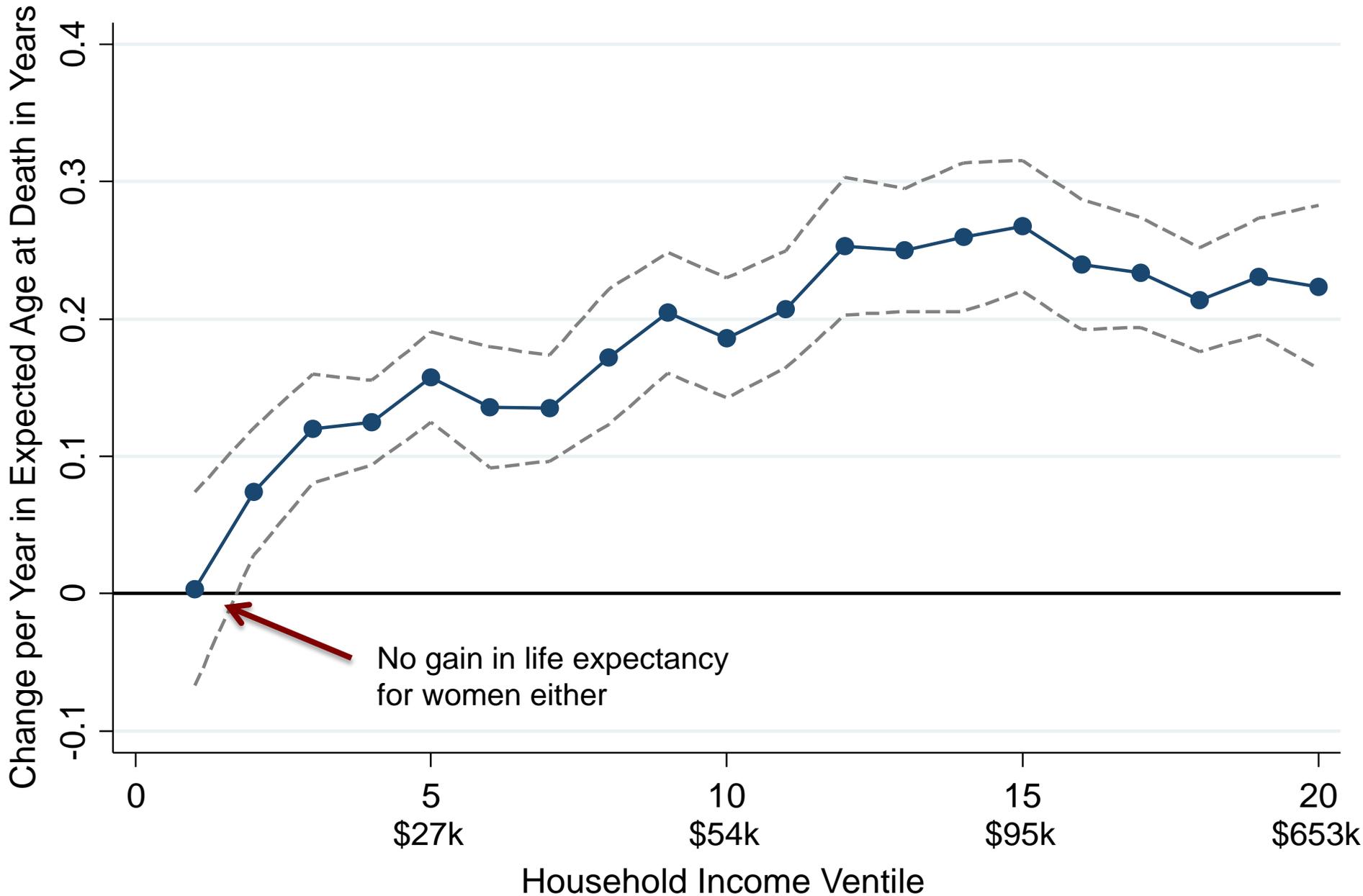
Change in Life Expectancy Per Year by Income Ventile, Men



Change in Life Expectancy Per Year by Income Ventile, Men



Change in Life Expectancy Per Year by Income Ventile, Women

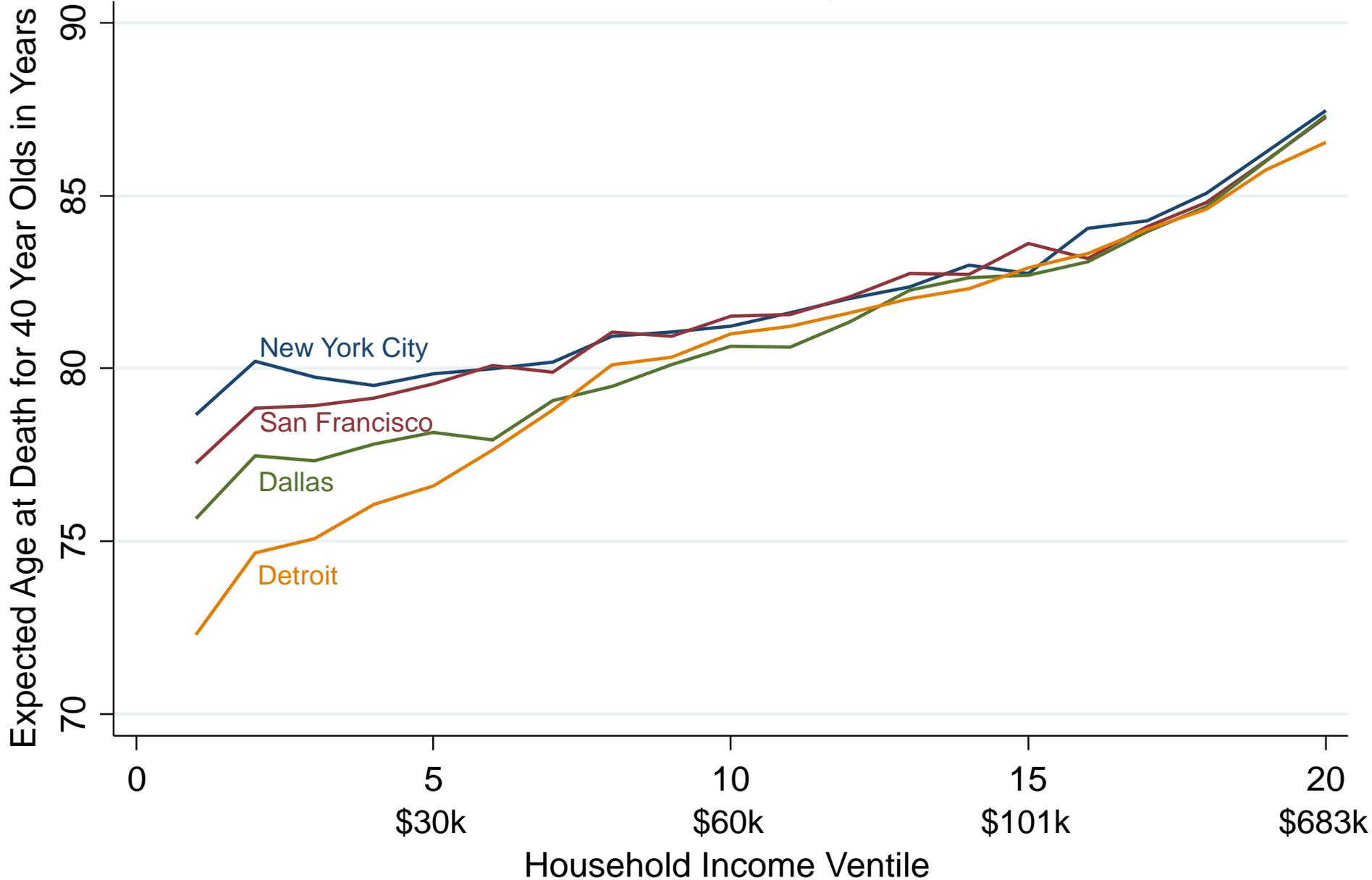


Part 3: Local Area Variation

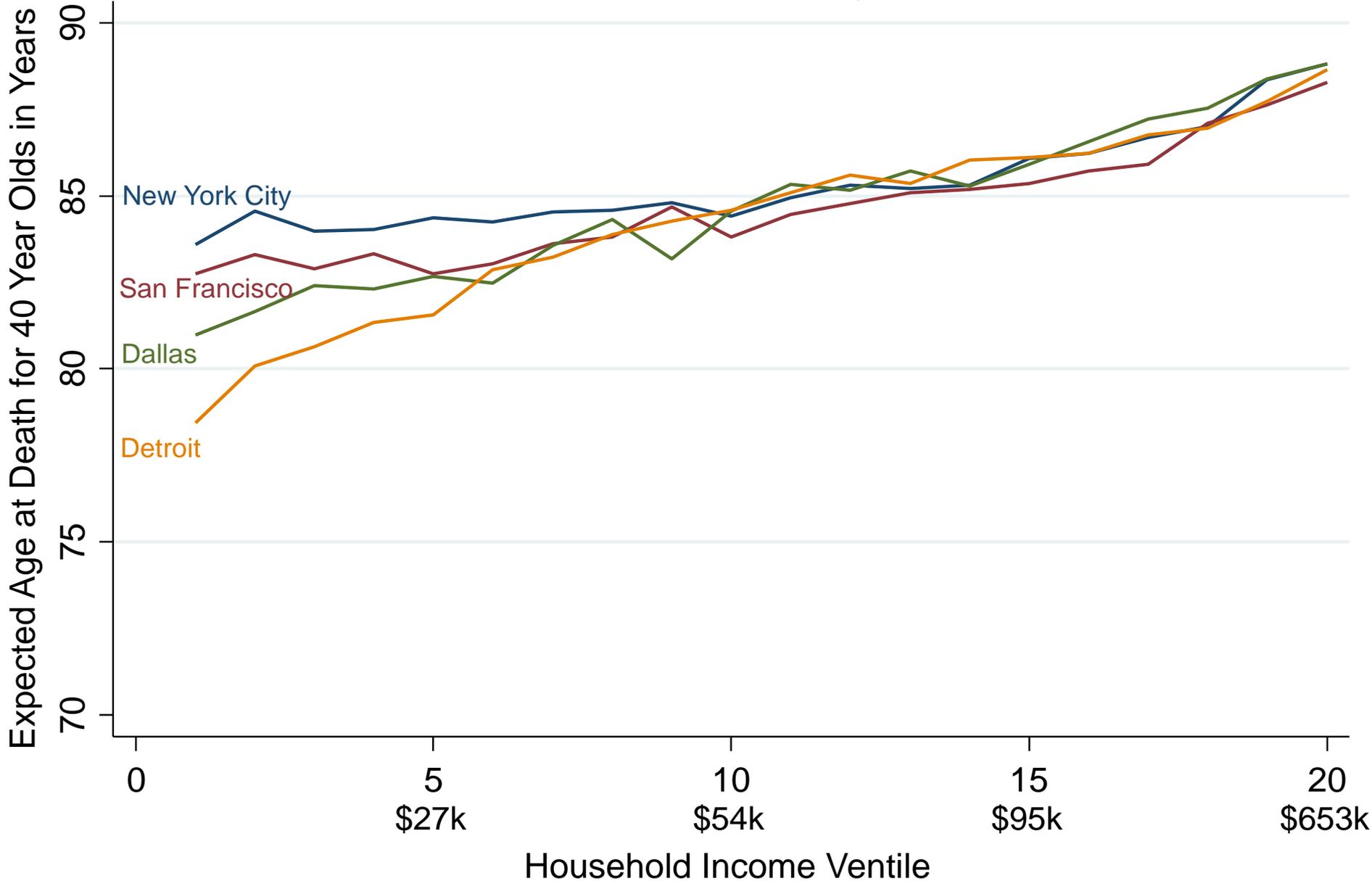
Local Area Variation

- Long literature analyzing geographical differences in mortality [e.g., Fuchs (1974), Murray et al. 2006, Berkman et al 2014]
- We analyze geographic variation at the level of commuting zones
 - Commuting zones are aggregations of counties (analogous to metro areas)
 - Also report county-level results
- Prior work has not disaggregated geographical variation in mortality by income
 - This turns out to be quite important...

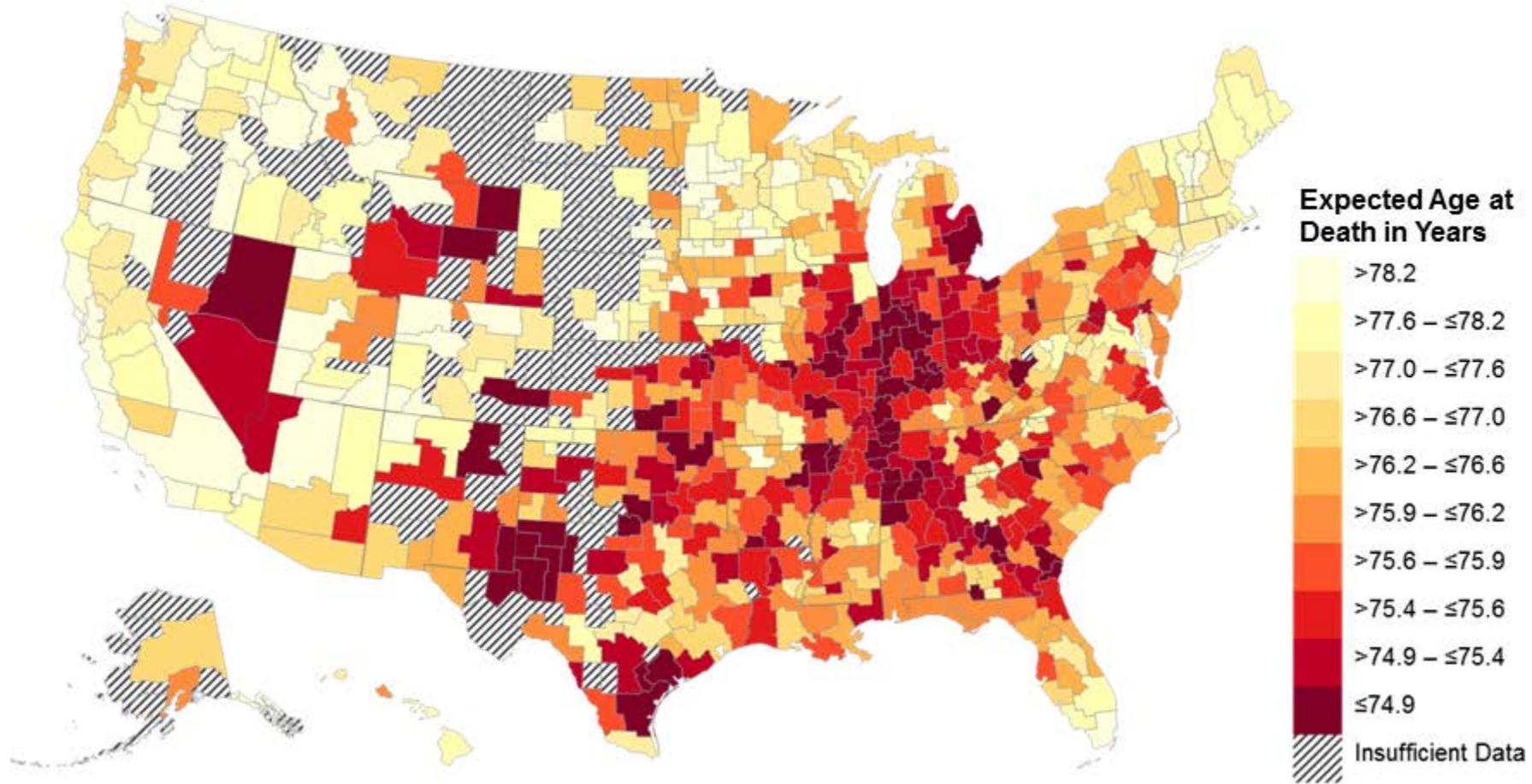
Race-Adjusted Expected Age at Death vs. Household Income for Men in Selected Major Cities



Race-Adjusted Expected Age at Death vs. Household Income for Women in Selected Major Cities

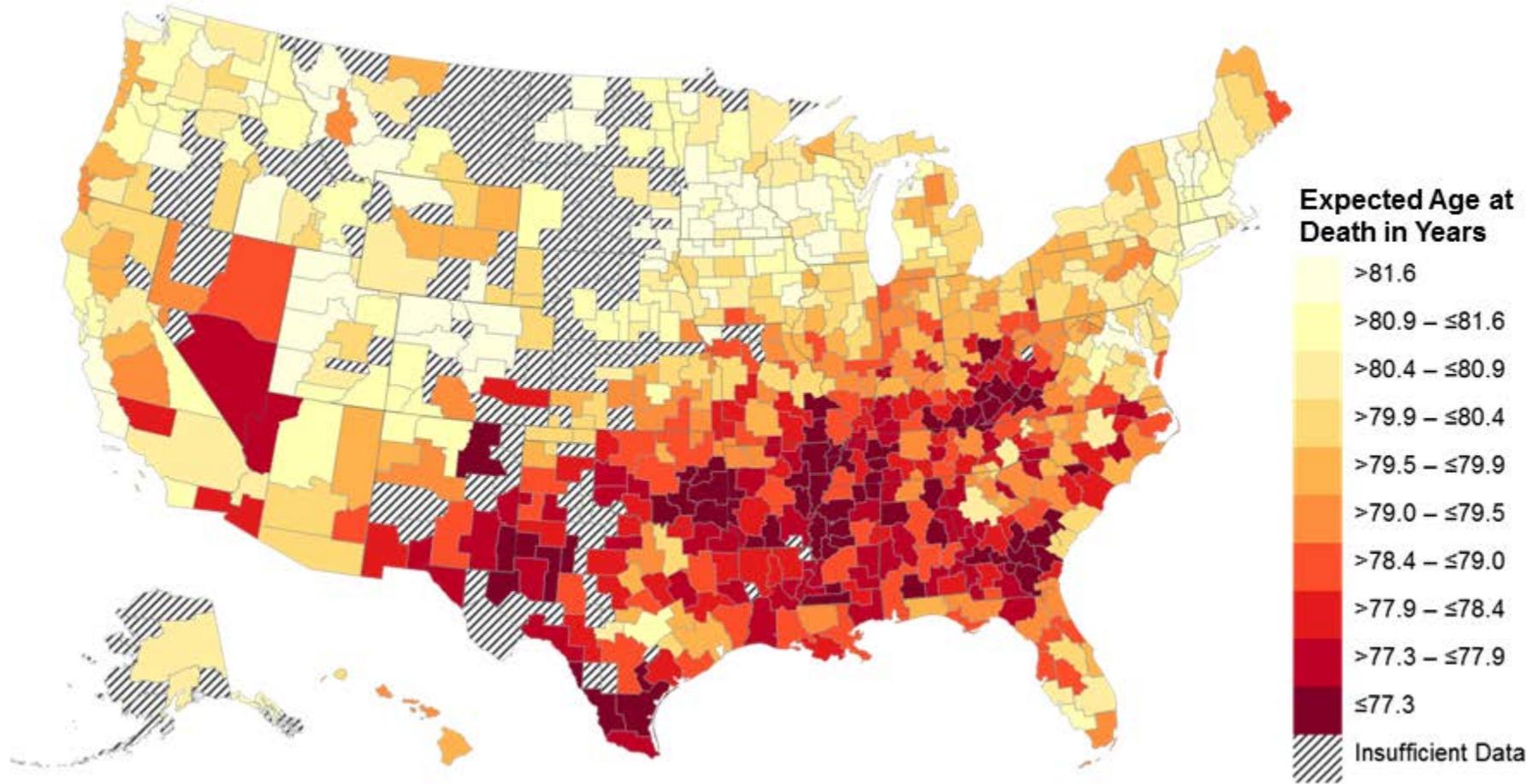


Race-Adjusted Expected Age at Death for 40 Year Old Men Bottom Quartile of U.S. Income Distribution



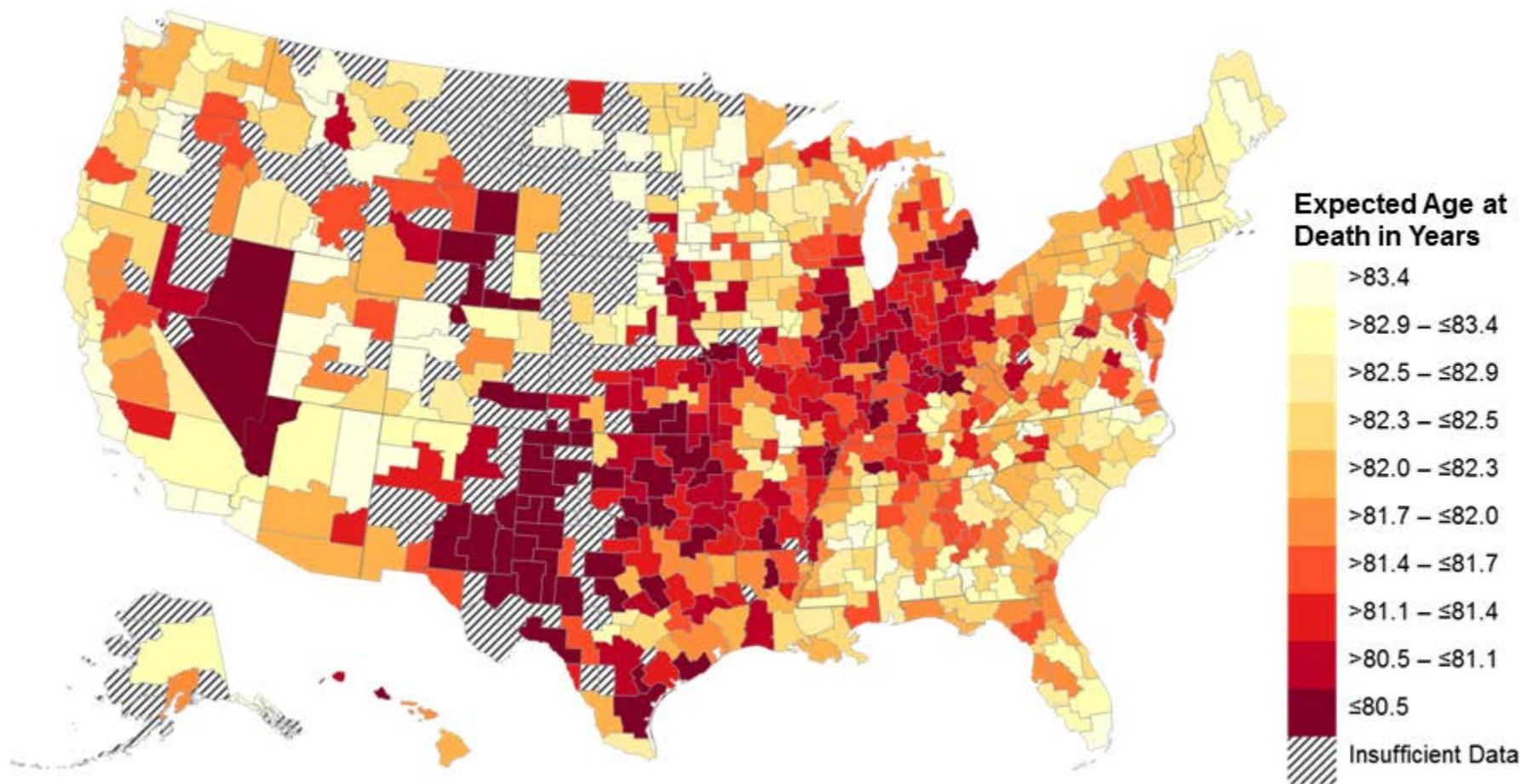
Note: Lighter Colors Represent Areas with Higher Life Expectancy

Race-Adjusted Expected Age at Death for 40 Year Old Men Pooling All Income Groups



Note: Lighter Colors Represent Areas with Higher Life Expectancy

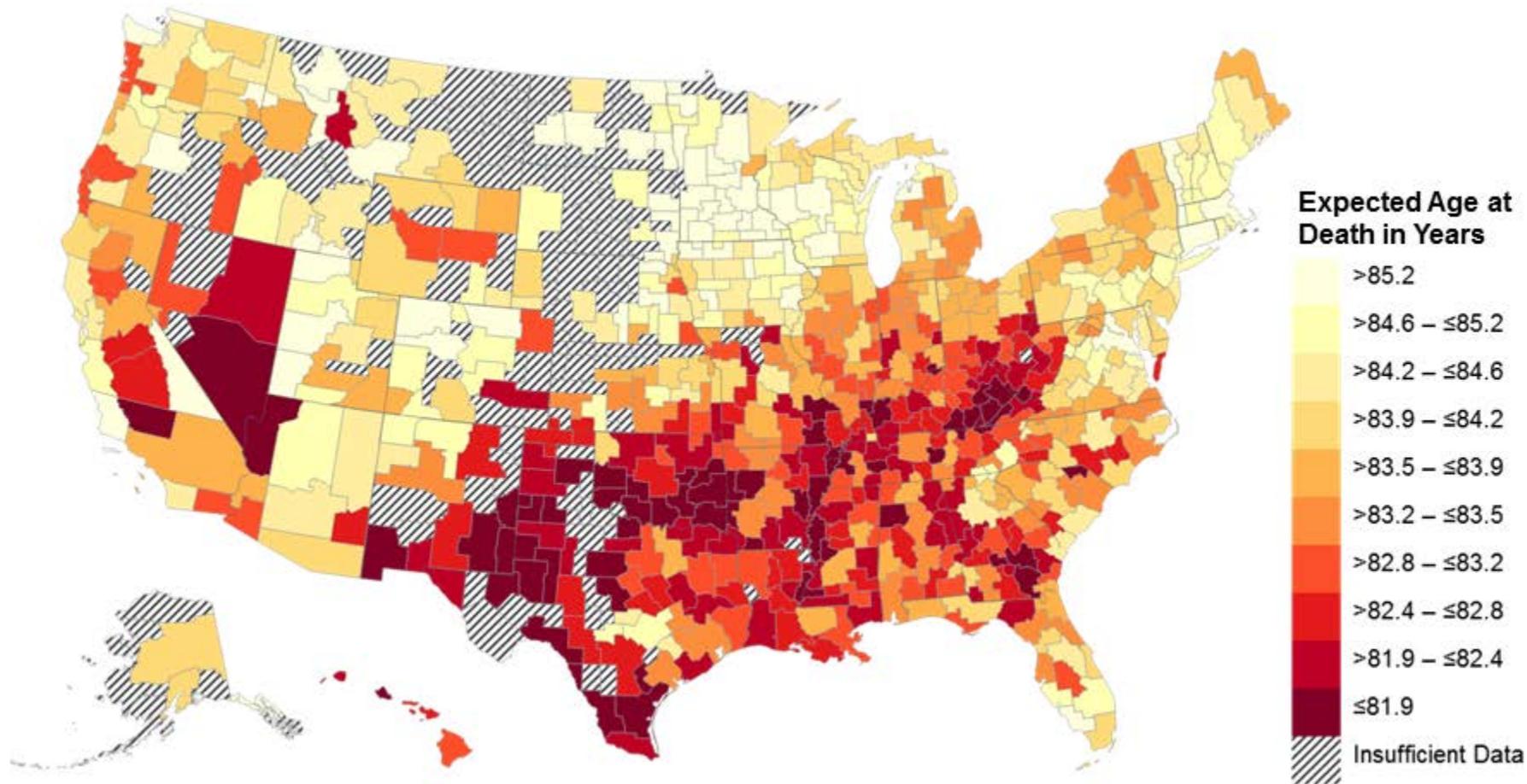
Race-Adjusted Expected Age at Death for 40 Year Old Women Bottom Quartile of U.S. Income Distribution



Note: Lighter Colors Represent Areas with Higher Life Expectancy

Race-Adjusted Expected Age at Death for 40 Year Old Women

Pooling All Income Groups



Note: Lighter Colors Represent Areas with Higher Life Expectancy

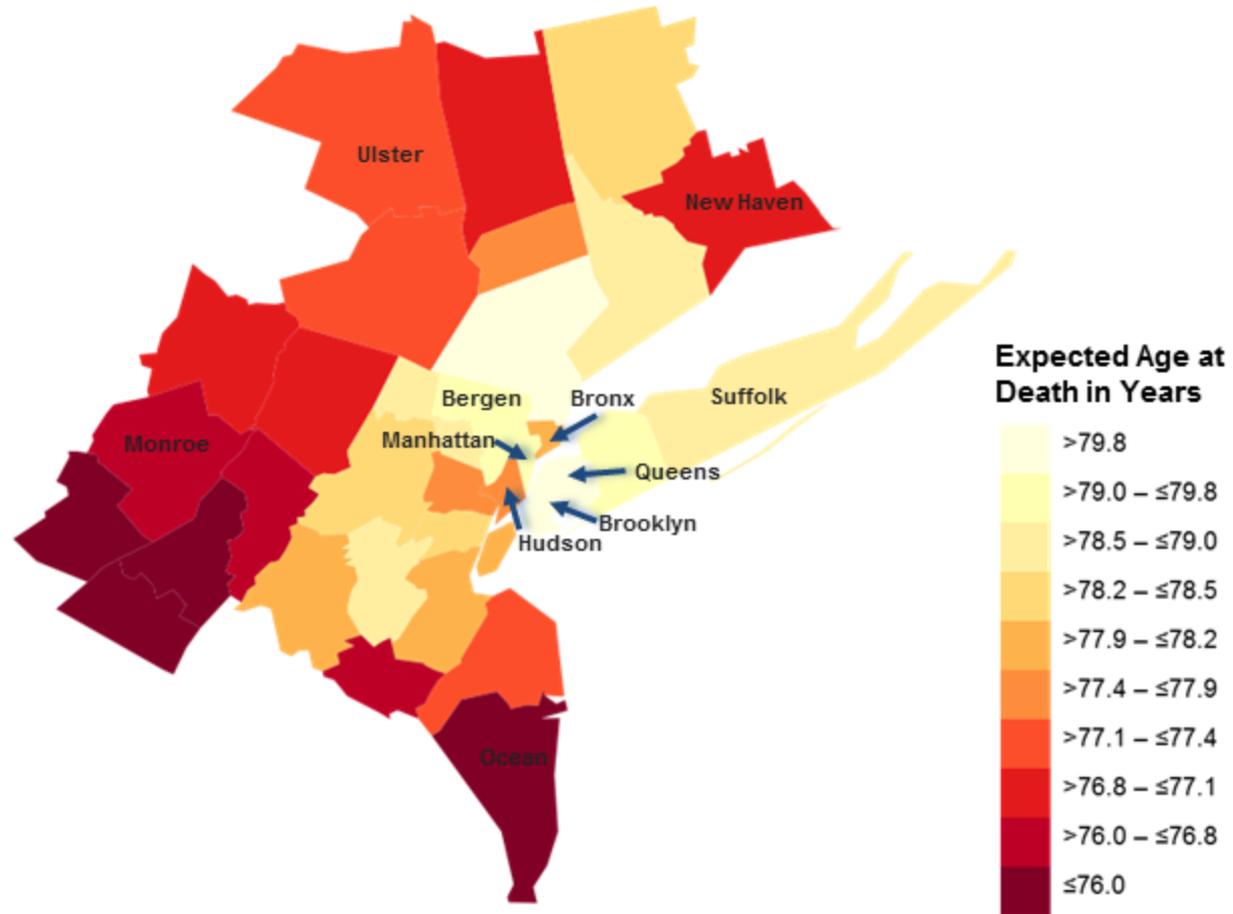
Race-Adjusted Expected Age at Death for 40 Year Olds in Bottom Quartile

Top 10 and Bottom 10 CZs Among 100 Largest CZs

Top 10 CZs			Bottom 10 CZs		
Rank	CZ	Expected Age at Death	Rank	CZ	Expected Age at Death
1	New York, NY	81.8 (81.6, 82.0)	91	San Antonio, TX	78.0 (77.6, 78.4)
2	Santa Barbara, CA	81.7 (81.3, 82.1)	92	Louisville, KY	77.9 (77.7, 78.2)
3	San Jose, CA	81.6 (81.2, 82.0)	93	Toledo, OH	77.9 (77.6, 78.2)
4	Miami, FL	81.2 (80.9, 81.6)	94	Cincinnati, OH	77.9 (77.7, 78.1)
5	Los Angeles, CA	81.1 (80.9, 81.4)	95	Detroit, MI	77.7 (77.5, 77.8)
6	San Diego, CA	81.1 (80.8, 81.4)	96	Tulsa, OK	77.6 (77.4, 77.9)
7	San Francisco, CA	80.9 (80.6, 81.3)	97	Indianapolis, IN	77.6 (77.4, 77.8)
8	Santa Rosa, CA	80.8 (80.5, 81.2)	98	Oklahoma City, OK	77.6 (77.3, 77.8)
9	Newark, NJ	80.7 (80.5, 80.9)	99	Las Vegas, NV	77.6 (77.4, 77.8)
10	Port St. Lucie, FL	80.7 (80.5, 80.9)	100	Gary, IN	77.4 (77.1, 77.8)

Note: 95% confidence intervals shown in parentheses

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



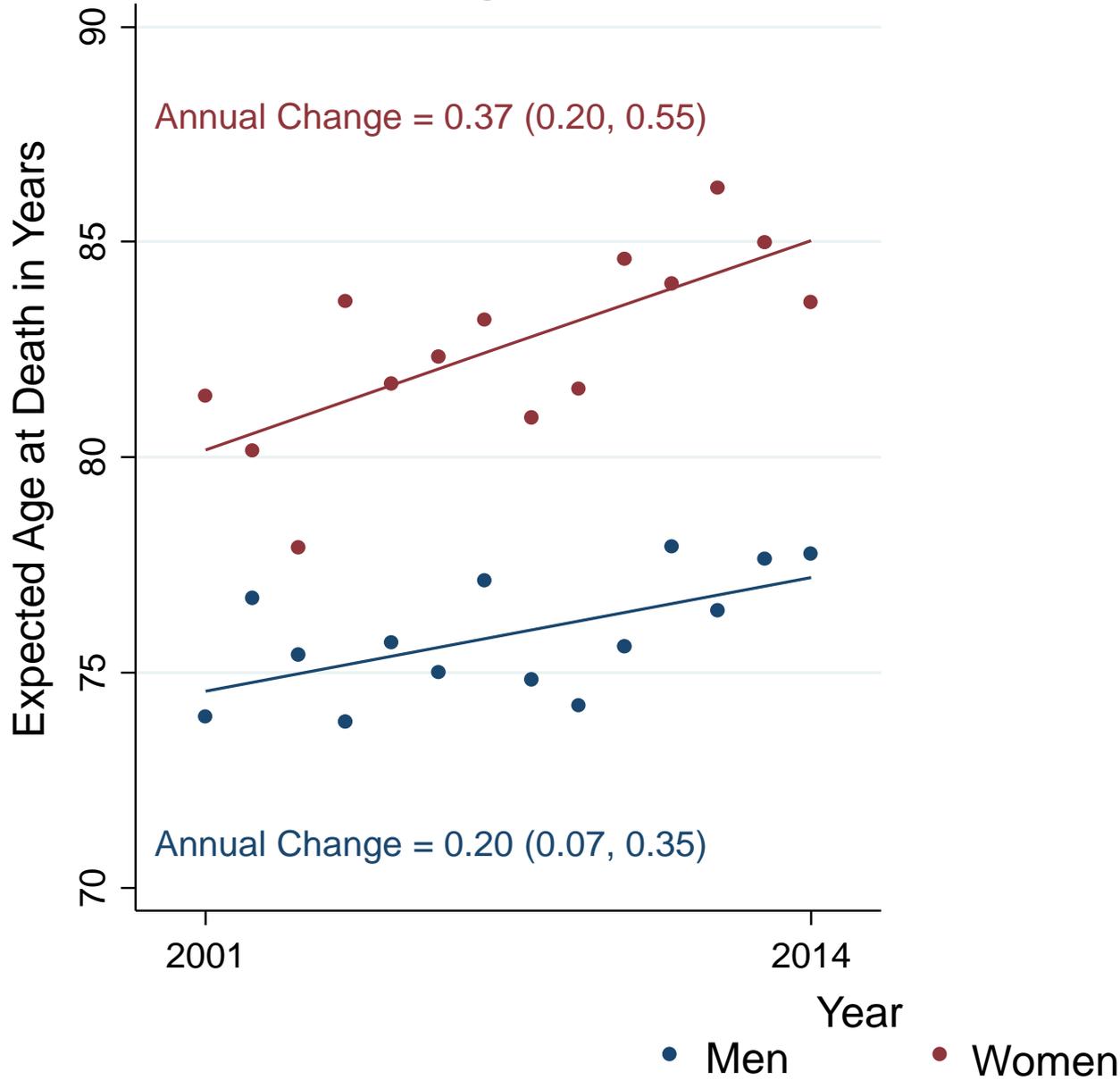
Note: Lighter Colors Represent Areas with Higher Life Expectancy

Local Area Variation in Trends

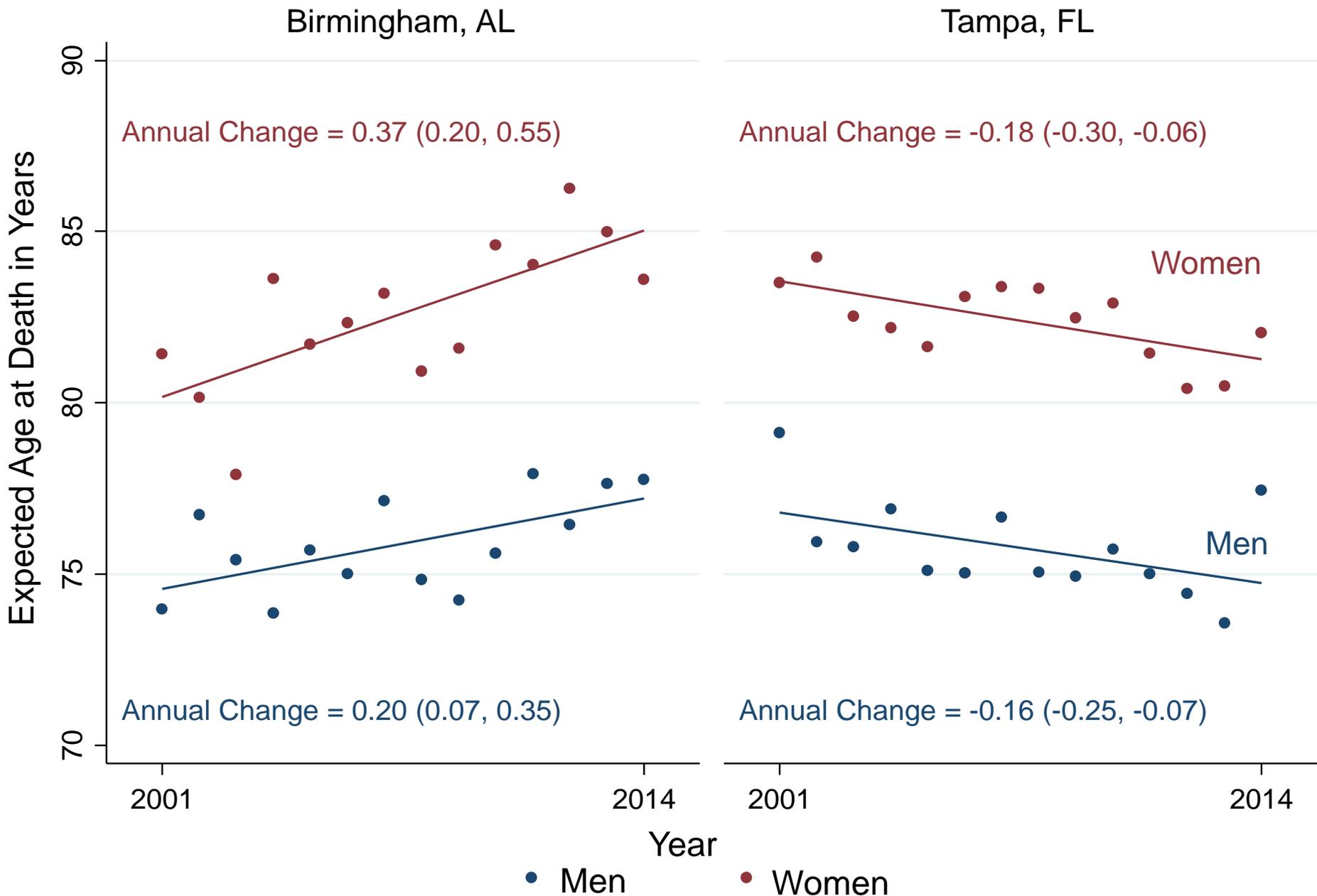
- Next, analyze how *trends* in life expectancy vary across areas

Change in Race-Adjusted Expected Age at Death in Bottom Quartile

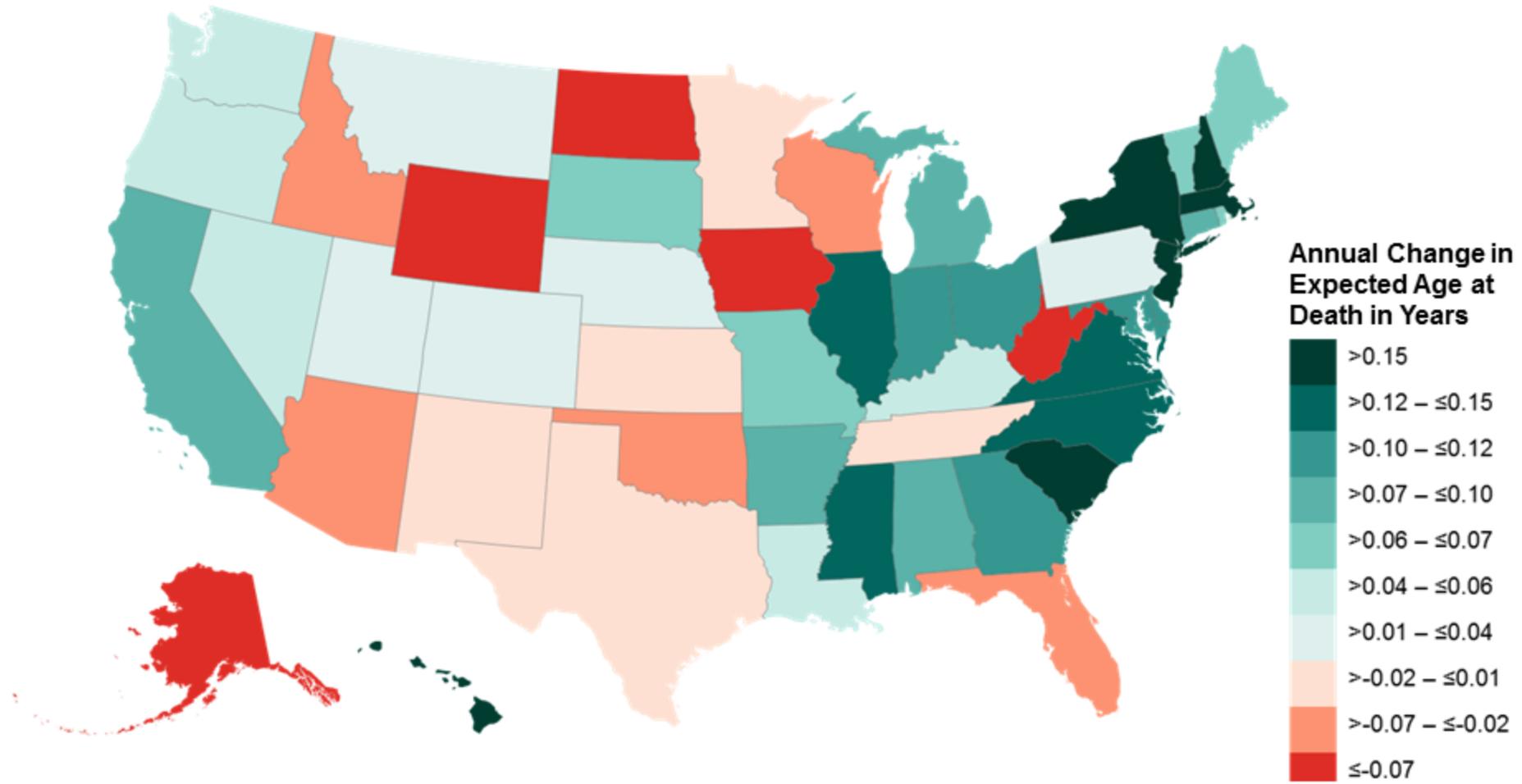
Birmingham, AL



Change in Race-Adjusted Expected Age at Death in Bottom Quartile

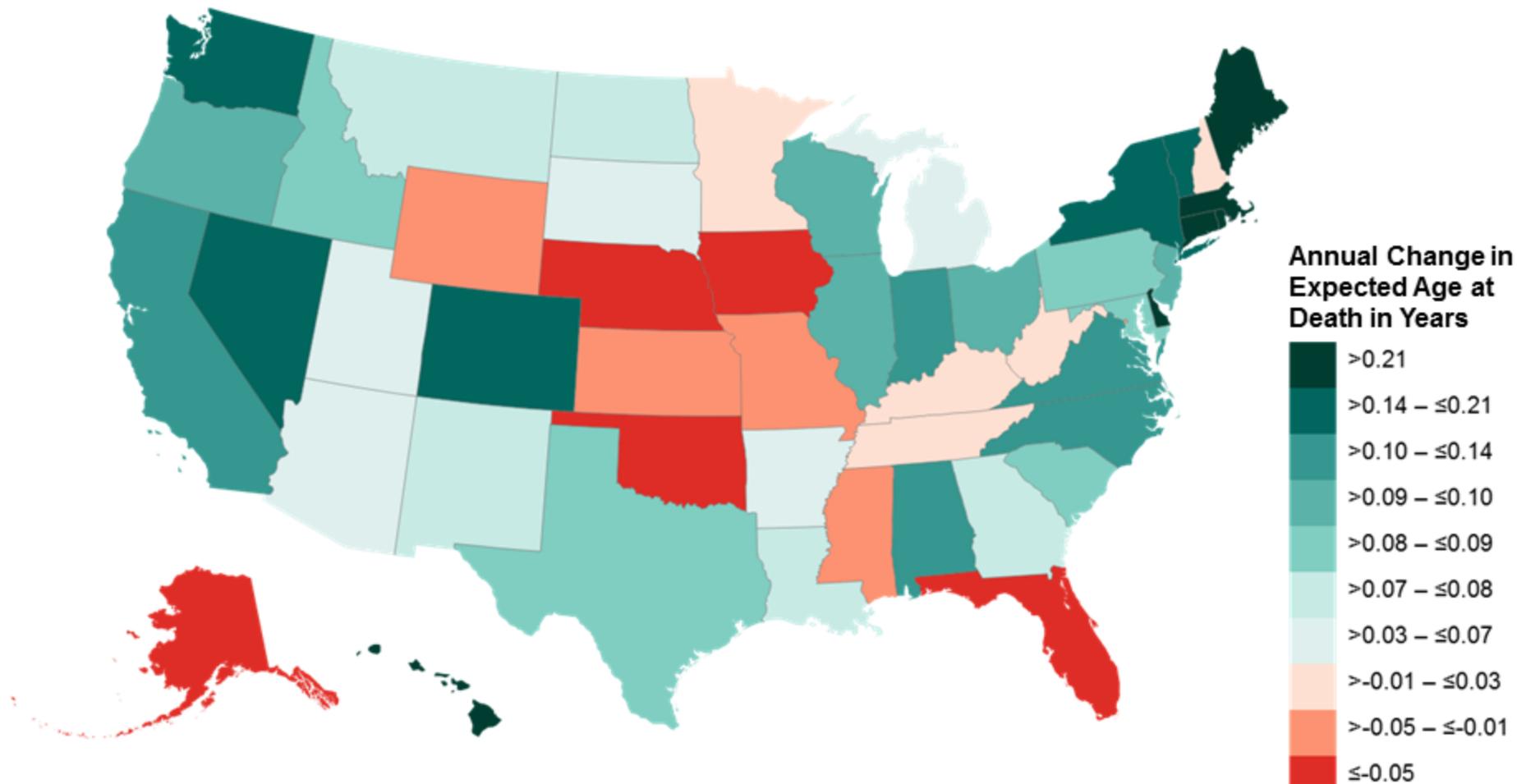


Annual Change in Race-Adjusted Expected Age at Death for Men in Bottom Quartile by State



Note: Turquoise represents rising life expectancy; red represents falling life expectancy

Annual Change in Race-Adjusted Expected Age at Death for Women in Bottom Quartile by State



Note: Turquoise represents rising life expectancy; red represents falling life expectancy

Change in Race-Adjusted Expected Age at Death in Bottom Quartile

Top 10 and Bottom 10 CZs Among 100 Largest CZs

Top 10 CZs			Bottom 10 CZs		
Rank	CZ	Change over Decade	Rank	CZ	Change over Decade
1	Toms River, NJ	3.8 (2.4, 5.2)	91	Cape Coral, FL	-0.7 (-2.1, 0.6)
2	Birmingham, AL	2.9 (1.8, 4.1)	92	Miami, FL	-0.7 (-1.4, -0.1)
3	Richmond, VA	2.6 (1.3, 3.9)	93	Tucson, AZ	-0.7 (-2.0, 0.5)
4	Syracuse, NY	2.5 (1.1, 4.0)	94	Albuquerque, NM	-0.8 (-2.2, 0.6)
5	Cincinnati, OH	2.4 (1.5, 3.4)	95	Sarasota, FL	-0.8 (-2.0, 0.3)
6	Fayetteville, NC	2.4 (1.0, 3.8)	96	Des Moines, IA	-1.0 (-3.0, 0.8)
7	Springfield, MA	2.3 (0.6, 4.1)	97	Bakersfield, CA	-1.2 (-2.8, 0.3)
8	Gary, IN	2.2 (0.8, 3.8)	98	Knoxville, TN	-1.2 (-2.6, 0.1)
9	Scranton, PA	2.1 (0.8, 3.4)	99	Pensacola, FL	-1.5 (-3.0, -0.2)
10	Honolulu, HI	2.1 (0.5, 3.8)	100	Tampa, FL	-1.7 (-2.5, -0.9)

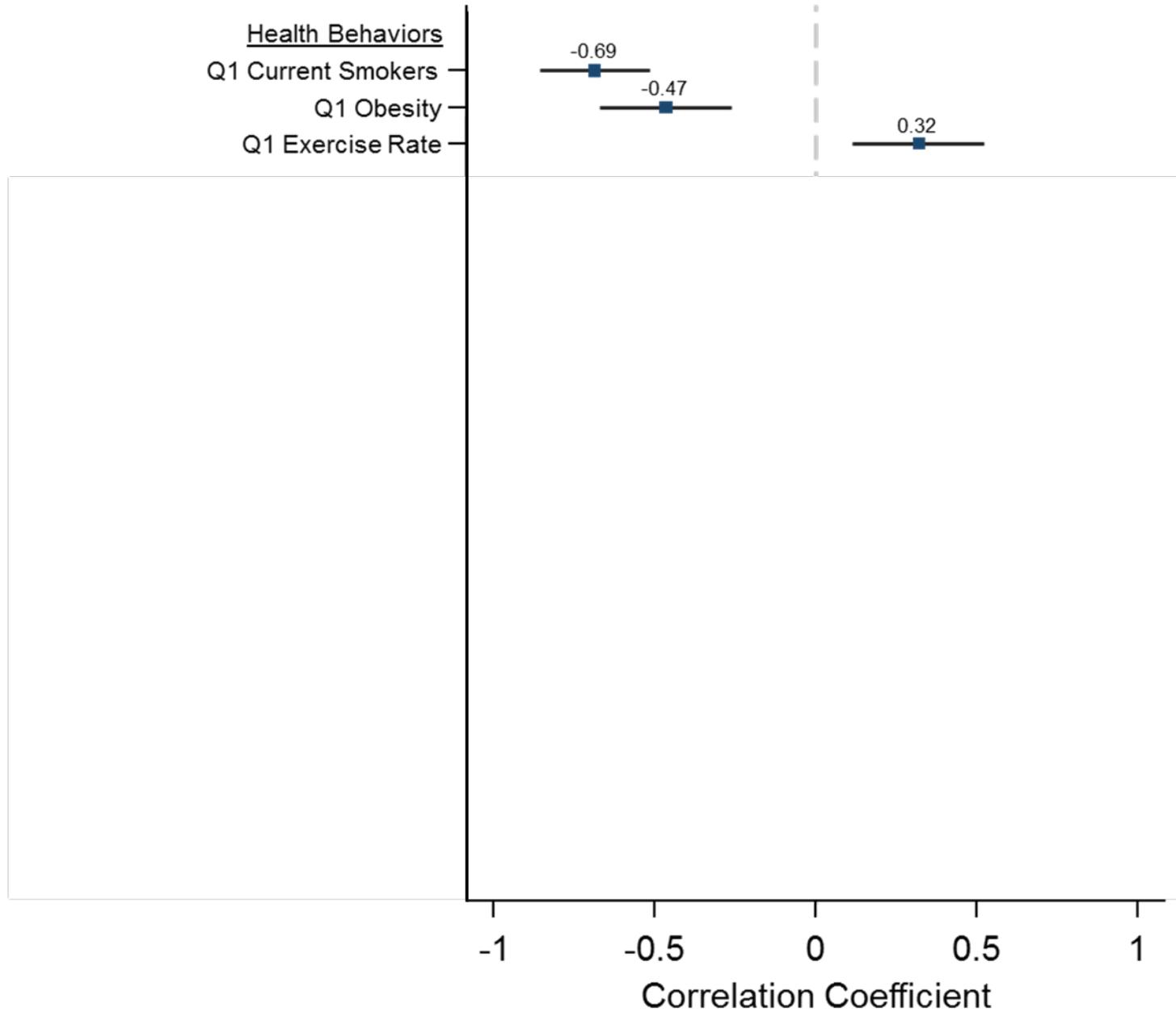
Note: 95% confidence intervals shown in parentheses

Part 4: Correlates of Spatial Variation in Mortality

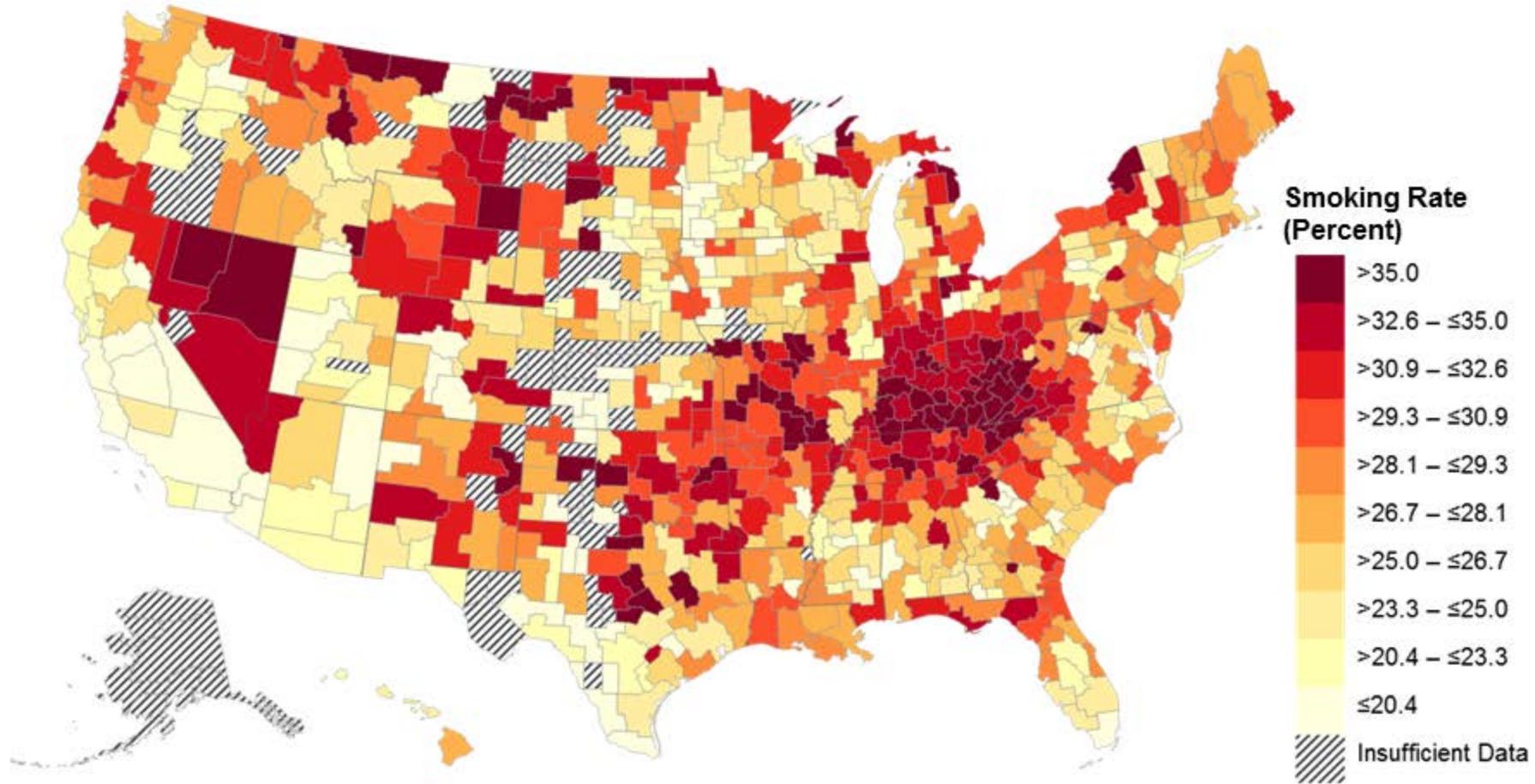
Why Does Life Expectancy Vary Across Areas?

- Finally, we characterize the features of areas with high vs. low life expectancy conditional on income
 - Analysis is purely correlational: does not directly identify causal pathways that can be manipulated to change mortality
- Begin by assessing measures of health behavior using data from the BRFSS [Fuchs 1974]

Correlations of Expected Age at Death with Health and Social Factors For Individuals in Bottom Quartile of Income Distribution



Smoking Rates by Commuting Zone in Bottom Quartile

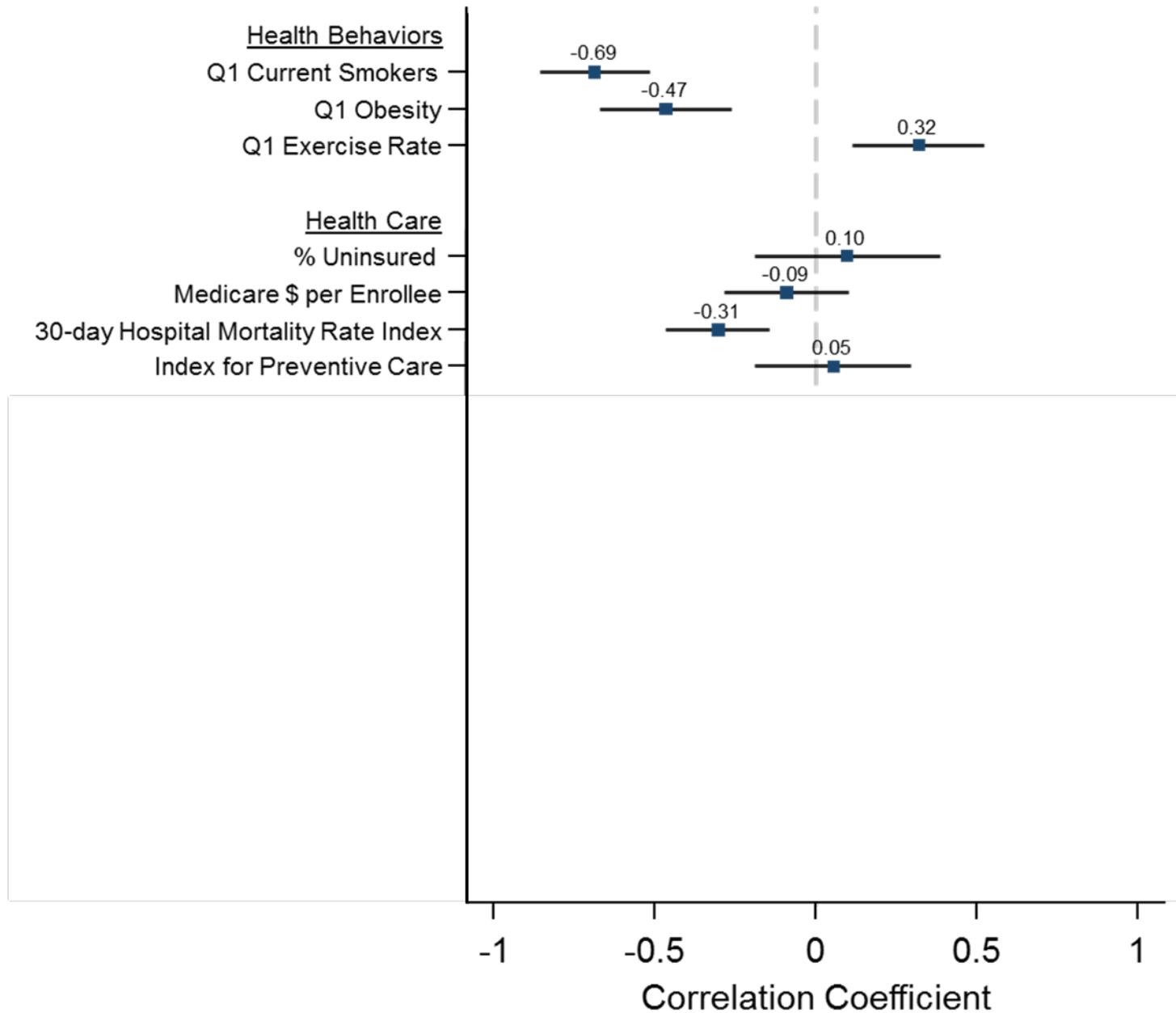


Note: Lighter Colors Represent Areas Lower Smoking Rates

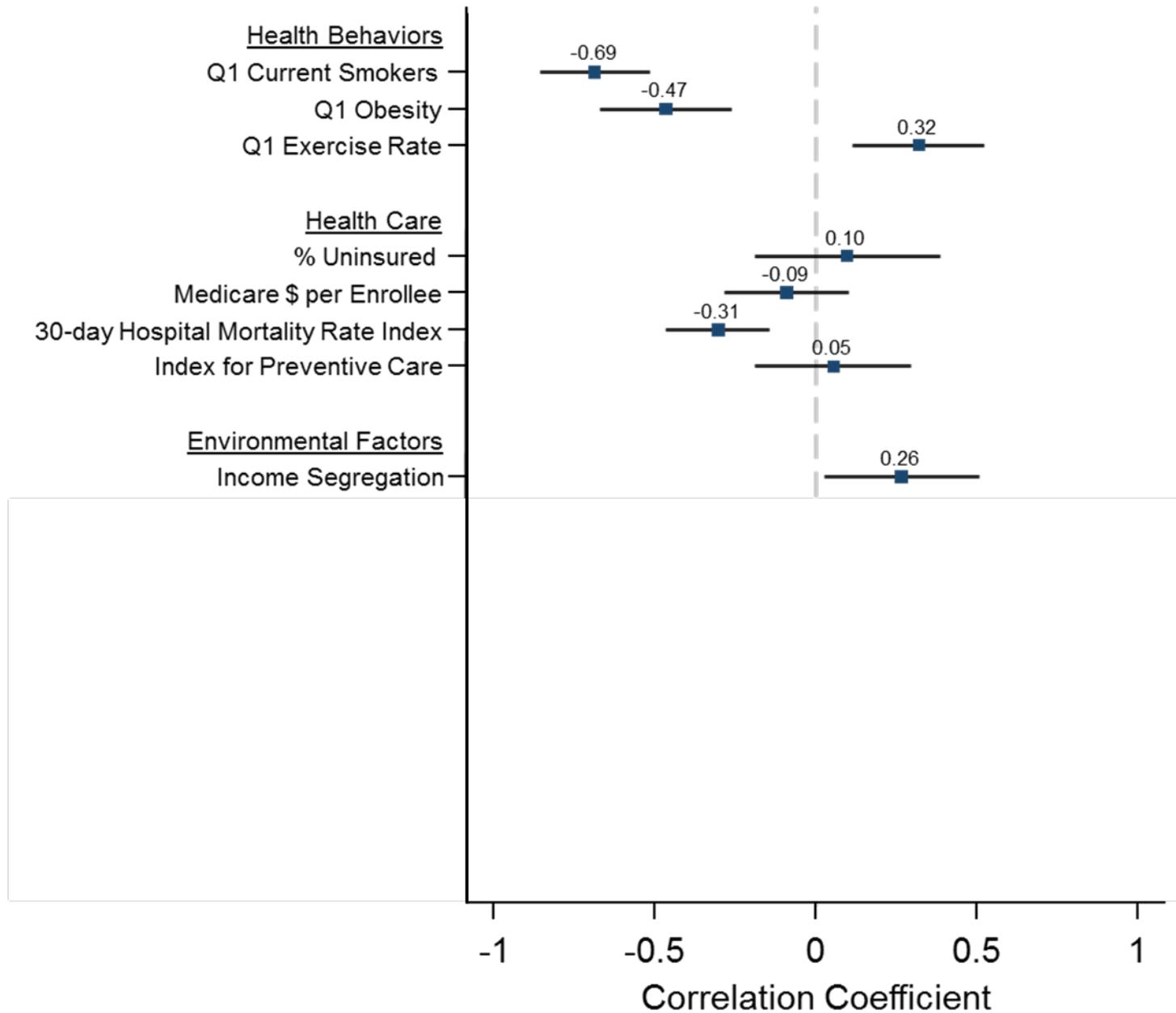
Why Does Life Expectancy Vary Across Areas?

- Variation in life expectancy among low income individuals is strongly related to variation in health behaviors
- What generates spatial variation in health behaviors and outcomes?
- We focus here on four theories discussed widely in literature:
 1. Health care [Fisher et al. 1993, Almond et al. 2010, Doyle et al. 2015]
 2. Environmental factors [Dockery et al. 1993, Currie and Neidell 2005]
 3. Income inequality [Lynch et al. 1998, Deaton and Lubotsky 2001, Wilkinson 2005]
 4. Economic decline [Ruhm 2000, Sullivan and von Wachter 2009]

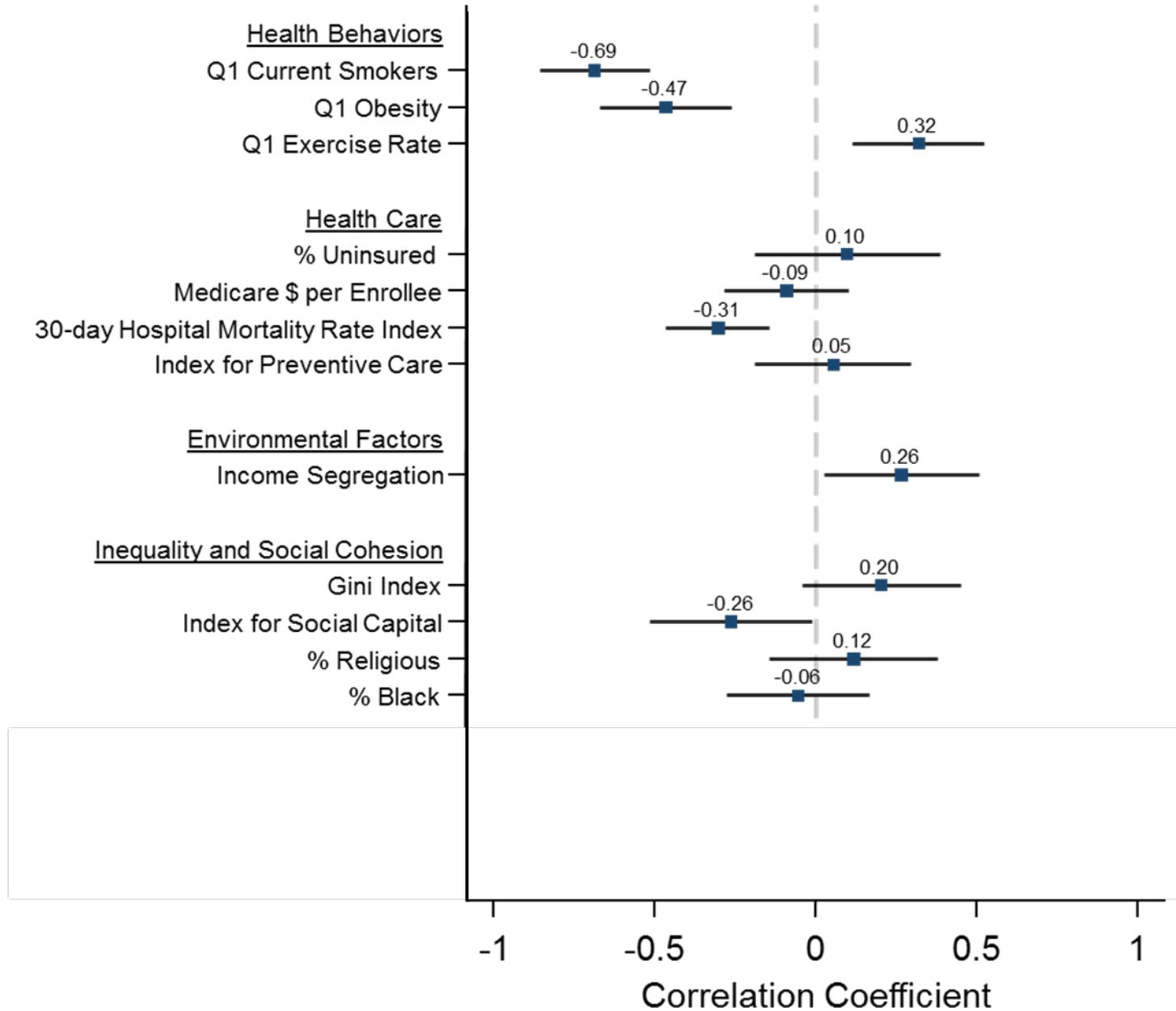
Correlations of Expected Age at Death with Health and Social Factors For Individuals in Bottom Quartile of Income Distribution



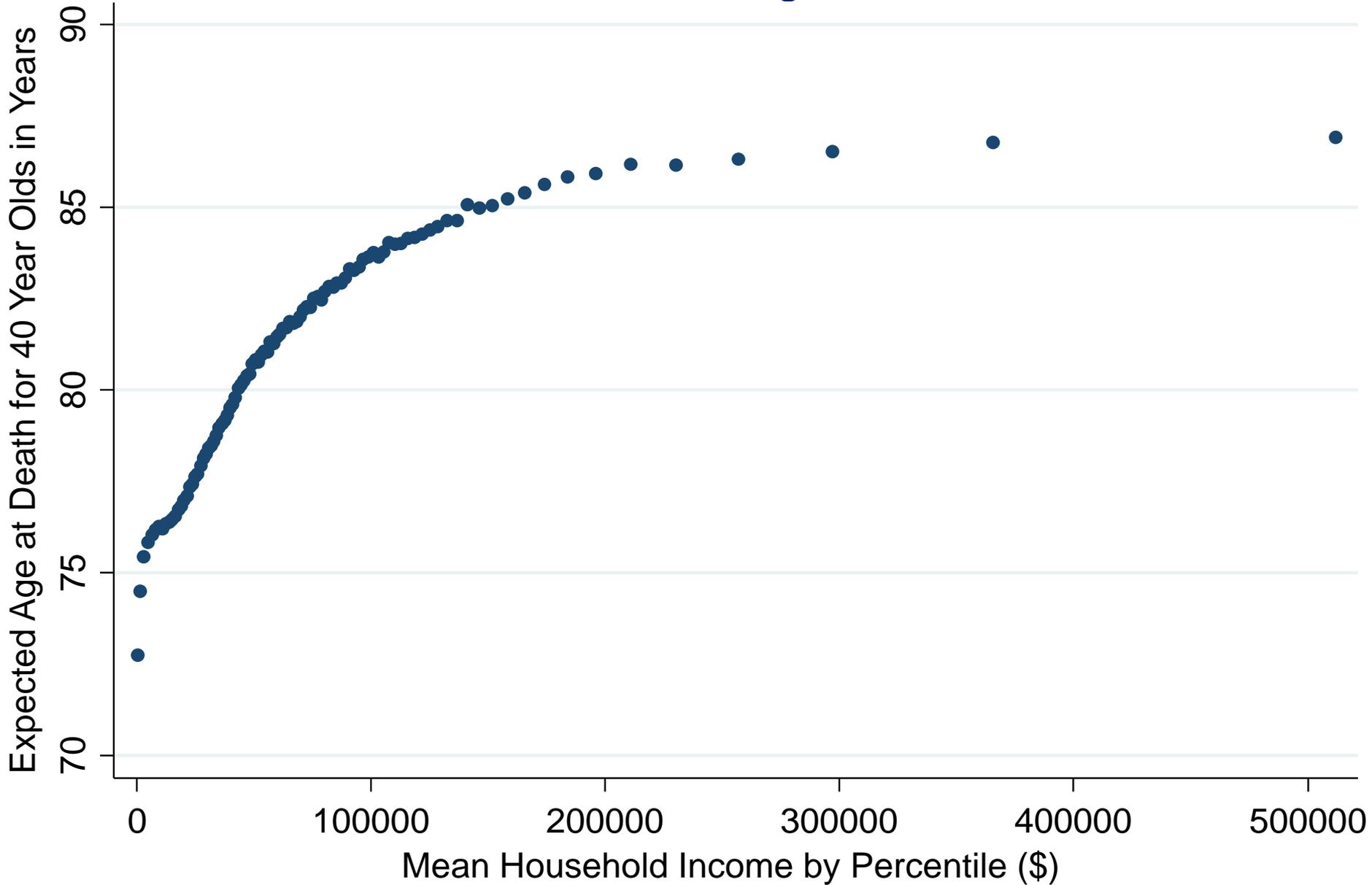
Correlations of Expected Age at Death with Health and Social Factors For Individuals in Bottom Quartile of Income Distribution



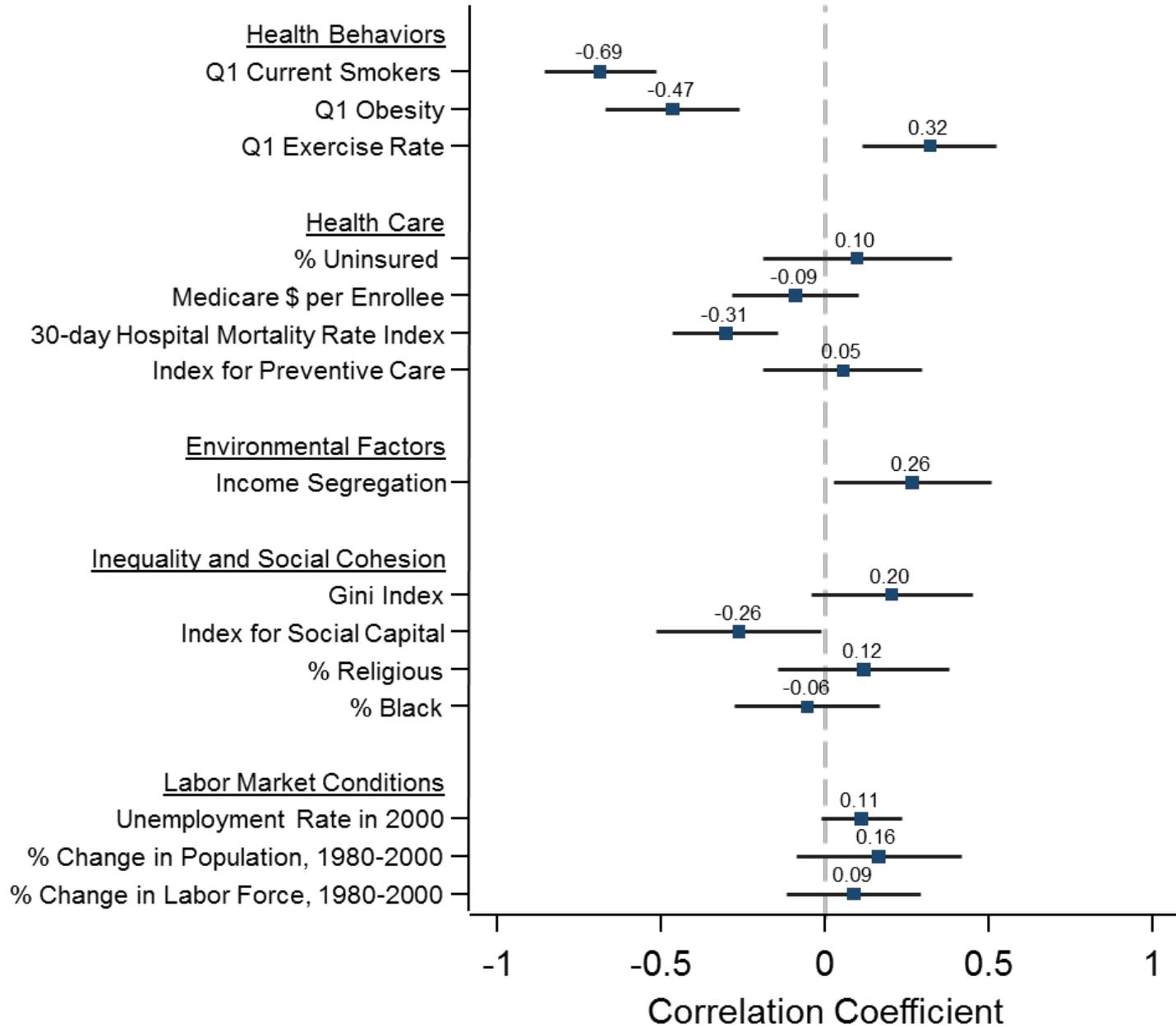
Correlations of Expected Age at Death with Health and Social Factors For Individuals in Bottom Quartile of Income Distribution



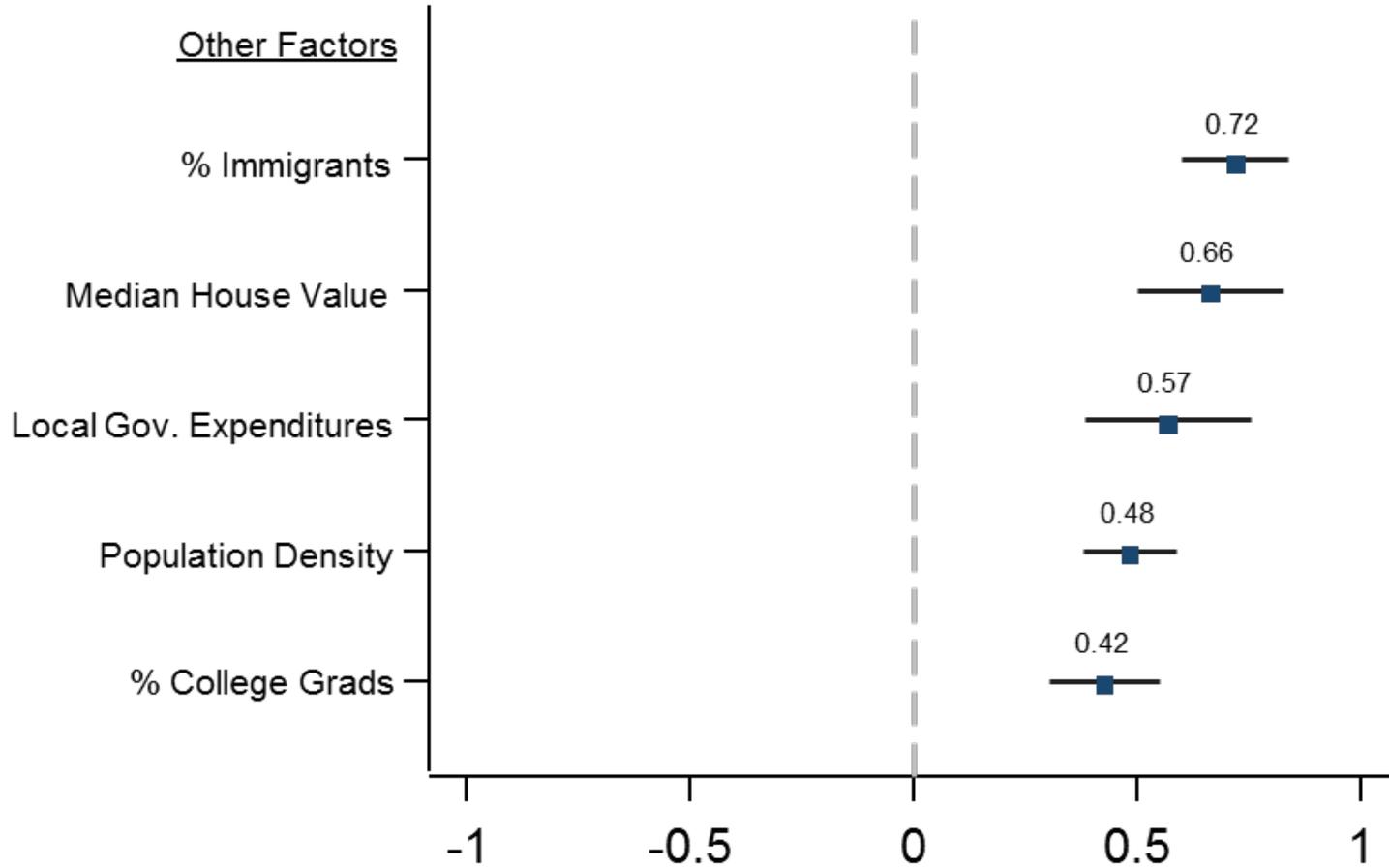
Expected Age at Death vs. Household Income For Men at Age 40



Correlations of Expected Age at Death with Health and Social Factors For Individuals in Bottom Quartile of Income Distribution



Correlations of Expected Age at Death with Other Factors For Individuals in Bottom Quartile of Income Distribution



Correlations: Summary

- General pattern: Low-income people in affluent, educated cities live longer (and have healthier behaviors)
- Why is this the case?
 - Spillovers from rich to poor: regulation, public revenues/transfers
 - Exposure to people with healthier behaviors
 - Sorting: low-income people who live in expensive cities are a selected group with different characteristics
 - Ongoing work by other researchers will shed light on these alternative mechanisms

Conclusion

- Inequality in life expectancy is large and growing, but not immutable: some areas in the U.S. have relatively small and shrinking gaps
- Differential trends imply that indexing eligibility for Social Security and Medicare to mean life expectancy will affect progressivity
- Reducing health disparities likely to require local policy interventions
 - Ex: targeted efforts to improve health among low-income individuals in Las Vegas, Tulsa, and Oklahoma City
 - Changing health behaviors at local level likely to be important
- Statistics constructed here (available at www.healthinequality.org) provide a tool to monitor local progress and identify solutions