The Association Between Income and Adult Life Expectancy in the United States Emerged in the Mid 20th Century

Andrew Goodman-Bacon
Federal Reserve Bank of Minneapolis

Adriana Lleras-Muney
UCLA and NBER

Joseph Price
Brigham Young University and NBER

Dahai Yue
University of Maryland, College Park

Carson Zhang
UCLA

(Draft, 10/24/2025)

Abstract: We document for the first time that there was no association between household income and period life expectancy in either 1940 or 1950. Using occupation to proxy for income, we further document that the association between income and life expectancy was not positive in the 1900-1950 period. Since the previous literature shows positive gradients in 1960, our results imply that income gradients in lifespan began to emerge during the 1950s, a period during which income inequality was low but modern medical technologies advanced. Indeed, the relationship between income and *cohort* life expectancy turned from negative for those who were aged 40 in 1900 to positive for those who were 40 in 1950.

Between 2001 and 2014, the richest five percent of American 40-year-olds could expect to live 9.8 years longer than the poorest five percent (Chetty et al., 2016). This gap equals the *entire* gain in life expectancy for 40-year-olds from 1940 to 2019 (Arias & Xu, 2022; Greville, 1947), a period during which advances in medical science, expansions in health care capacity and insurance coverage, and improving health behaviors all contributed meaningfully to increased longevity (Cutler et al., 2006). Yet a basic question about these relationships remains open: Have the lifespans of lower- and higher-income Americans always been very different, or did contemporary inequality in lifespans arise recently?

Historical evidence on the association between income and lifespan is rare because the data necessary to estimate it is not systematically available in the US. Vital records alone are insufficient for this purpose: they do not contain income, income proxies like education and occupation are poorly recorded, and death registration in the United States was not complete until about 1933. Kitagawa and Hauser (1973)'s groundbreaking study of linked Census records from 1960 yielded the first nationally representative evidence that period mortality rates decline with household income, but did not examine life expectancy. Subsequent estimates that combine income data from Censuses, survey datasets, or tax records, with death information corroborate this finding but only as far back as the late 1970s (Chetty et al., 2016; Cristia, 2009; Duggan et al., 2008; Rogot et al., 1992; Schwandt et al., 2022; Singh & Siahpush, 2006), when more than half of the aforementioned life expectancy gains had already materialized. Although existing research documents that the association between income and life expectancy has increased since the 1960s, we do not know when these associations started growing and whether these associations have always existed at all (see Appendix 1 for a detailed description of the key papers in the literature). In fact, some historical evidence suggests that better off individuals lived shorter lives than their lower-income counterparts (Cutler et al., 2006).

This paper documents for the first time how income gradients in life expectancy evolved from 1900 to 1950. We create representative data for millions of individuals by combining census records containing information on income and its determinants, with family trees that provide birth and death dates. We first provide estimates of the association between household income and life expectancy among about 10 million respondents in the 1940 Census who are linked to a profile on the Family Tree at familysearch.org that also includes the date or year of death for the individual.

The 1940 Census is the first census that collected data on individual income. These data allow us to construct period and cohort measures of life expectancy conditional on surviving to 1941 and age 40, and to relate these measures to household income. We also conduct the same analysis in 1950 by linking about 5 million people who were given the long Census form in 1950 to death information. There is no income data prior to 1940 for nationally representative samples. But we use data on about 5, 8, 9, and 12 million individuals in the 1900, 1910, 1920, and 1930 censuses to investigate the evolution of the income gradient from 1900 to 1950 using proxies for household income that rely on occupations of household members.

We report several novel findings. Period life expectancy in 1941 is unrelated to household income in 1939 or to other proxies for individual income measures such as wage income and occupation income scores. This result holds within rural and urban areas and is robust to how we account for farming income. We also fail to find an income gradient in period life expectancy in 1951, and can rule out gradients of the size of those that prevailed in the 2000s (Chetty et al., 2016).

We then investigate whether the lack of an association between income and period life expectancy in the 1940s and 1950s is a historical anomaly. The post-war era was unique in the United States and in other Western countries: it was characterized by strong growth in incomes and historically low-income inequality. But income and wealth inequality were much larger in the 1900-1930 era (Piketty et al., 2017). Thus, it is possible that there was no gradient mid-century because income inequality was low, but that gradients were in fact large before and after.

We find no evidence that there was a significant income gradient in period life expectancy from 1900-1950. These results contrast with the findings in Kitagawa and Hauser (1973) for 1960 and other work showing positive income gradients in mortality. Our findings suggest then that income gradients are a modern phenomenon, that arose in the 1950s with the rise of modern medicine, despite low income inequality. These conclusions contrast with widely held hypotheses such as Link and Phelan (1995)'s fundamental cause of disease which state that those with the highest SES in society will always use their resources to live longer lives. According to this view, there are always gradients in life expectancy, regardless of the time period, society or causes of death. Our evidence suggests instead that historically there have been long periods during which period gradients in life expectancy did not exist. Our results are consistent with work for other

countries (Dribe & Eriksson, 2025) which also find no income gradients in lifespan prior to 1940 for men, though in contrast to that work, we do not find income gradients for women either.

Unlike prior US studies, such as Chetty et al. (2016), focusing on income gradients in period life expectancy, our data also allow us to investigate the relationship between income and *cohort* life expectancy which captures the actual average duration of life for a given cohort, instead of the predicted duration of life based on cross-sectional mortality rates at a point in time (period life expectancy). Given that the period gradient is rising in the second part of the century, we expect the cohort gradient to be more positive than the period gradient and to rise across birth cohorts. Indeed, we document that cohort life expectancy was negatively related to income for adult men and not related to income for adult women in 1900 (where income is measured by household occupation scores) and remained negative or insignificant until 1940. In 1940 and 1950, there is a positive income gradient using household income measures for men and women, but for women the cohort gradient is small in 1950. This evidence is consistent with rich adults, and in particular men, increasingly benefiting from technologies that became available in the 1940s and 1950s such as antibiotics and blood transfusions, and later on, from progress in cardiovascular disease.

II. Data

a. Data sources and data construction

We construct our datasets by matching the individual census data from 1900, 1910, 1920, 1930, 1940 and 1950 to the family trees from FamilySearch.

The 1940 and 1950 Census contains information on income and other socio-economic characteristics including educational attainment, occupation, residential location and family structure. Earlier censuses did not collect data on education or income, but they did ask all adult respondents to report their occupation from which rough measures of income can be constructed.

The database of family trees available from FamilySearch includes more than 1.7 billion people – the largest collection of its kind. Family trees contain death dates. The main advantage of the data is that it includes death information from various sources including national death databases such as the Social Security, but they also include data from other sources such as death

certificates from state registries, from tombs, obituaries and family members. This information is critical because national databases of deaths in the US are not complete until roughly 1970 and they are not made available after 2010. The data from FamilySearch allows us to track individuals from 1940 until 2023. The main drawback of the data derived from genealogies is that it underrepresents certain populations, in particular blacks, immigrants, individuals from small families and from certain regions of the US. To address this we follow the previous literature (Bailey et al., 2020; Newson & Falcaro, 2023; Seaman & White, 2013) and construct weights.

Ro make our results directly comparable to those in Chetty et al. (2016) and focus on adults aged 40 to 63 in the year after each census for whom we can observe their income two years ago, a better permanent income proxy.[1] We then link the individuals from these census records with their profiles on the Family Tree at familysearch.org. The Family Tree is a wiki-style platform that allows individuals doing family history research to collaborate in gathering information about their ancestors. The Family Tree includes 1.6 billion profiles of deceased individuals, and these profiles include information about the date and place of key life events (birth, marriage, and death) as well as the ability to attach historical records to the profile that corroborate this information. These attached historical records can be any one of the 20.5 billion records that FamilySearch has on their website, including the US census records. FamilySearch also has a machine-learning algorithm that predicts which records are a likely match for each profile. We use the FamilySearch API to identify each profile on the Family Tree that is attached or has likely match to each of the census records in our sample. We then use the birth and death date information from the profiles that we link to.

We are able to link about 50% of the target population in 1900-1930 censuses to a profile on the Family That includes both a birth and death date (sometimes just a year). Our link rate for the 1940 census is 32% and the link rate for the 1950 census is 22% (**Figure A1**). The link rates for the 1950 and 1940 census are lower since those records have had less time to be examined and connected to profiles on the Family Tree. US census records are made public after 72 years, so the 1950 census only recently became available on FamilySearch.

[1] For the 1941 census, we focus on adults aged 40 to 62 due to a lack of older cohorts in the tree.

The cohorts we study were born between 1838 and 1861 when we observed them in 1900 and between 1888 and 1911 when we observe them in 1950. Nearly all of these cohorts are now extinct. There are two main reasons why we don't have lifespan information for everyone. First, not everyone in our sample has a profile on the Family Tree. There are also likely cases where the individual has a profile on the Family Tree but it has not yet been successfully linked to the census records in our sample. Second, there are people in our sample for which we are able to link their census record to a profile on the Family Tree but that profile does not include any death information. This could stem from the fact that the death date is not available in any publicly available historical records and has not been recorded on the profile by family members with personal knowledge of the death date. Despite these two limitations, our link rates are high by historical standards, which typically link between 10 and 30% of the target populations (e.g. Abramitzky et al., 2014; Bailey et al., 2023). Previous work has documented that the data constructed by linking US census records to the Family Tree accurately describes population trends by sex (e.g. Black et al., 2025; Lleras-Muney et al., 2022). We verify that our data produces weighted mortality rates that are very similar to those provided by the Social Security Administration (SSA) in each census year for the nation as a whole (e.g. **Figure A2**).

b. Measurement of income and of lifespan

To make results comparable over time, we follow the approach by Chetty et al. (2016) and investigate the relationship between household income and life expectancy at age 40. Details on how our data construction compares to Chetty et al.'s can be found in **Table A1.** We excluded farmers for comparison across years as their income was not well captured even in the 1940 and 1950 censuses. We show that the inclusion of farmers does not substantively affect our results.

Income measures. We measure income in the 1940 census, which asked all respondents over the age of 14 about the "amount of money wages or salary received" in the previous year (1939). We restrict attention to individuals with positive income to match with Chetty et al. (2016) initially, but for most analyses we include all 0s. We construct our main measure of household income by summing over the income of all the household members. To match what the Chetty team does, we restrict attention to individuals who survive to 1941 and age 40-63 in 1941, so that the income is measured 2 years prior. Prior work has determined that annual income measured during these ages (40-62) is a good approximation of lifetime income (e.g. Haider & Solon,

2006). We classify individuals with positive household income into income percentiles, stratified by age and sex. In addition, we define a separate group for individuals with zero household income.

We also use the 1950 census questions on income. Although these were only asked of sample line respondents over the age of 14 (six per sheet) the questions on income are more comprehensive and include amounts earned separately from wages and salaries, from self-employment (own business or farm) and from other sources including rents and transfers (interest, dividends, veteran's allowances, pensions, rents, or other income). Among our targeted cohorts, about 32.5% of total household income comes from sources other than household wage and salary income.

Earlier censuses prior to 1940 do not have income measures. We assign individuals to income using their occupation and then calculate total household income summing over all individuals in the household. First, we use occupation scores, which are available for all persons who provide their occupation in the census. These assign each individual the 1950 income in their occupation and can be computed even among the unemployed. The 0 value (=NA) in the census code is set to missing. Alternatively, among those with an occupational response, we use the LIDO score (Lasso-adjusted industry, demographic, and occupation score), which was developed by Saavedra and Twinam (2020) to better predict income using a machine learning model that constructs predicted income in 1950 based on industry, occupation and demographics. The original LIDO score is on a log-income scale, and we take its exponential form throughout the analysis. By summing over the occupation score or LIDO score of all the household members, we get another proxy for household income. These proxies for income are available for all censuses from 1900 to 1950, and contain only positive values. Thus for 1940 and 1950 we can compare our results using both actual incomes and these proxies. But for earlier periods 1900-1930 we must rely on these alone.

Mortality rates and life expectancy. We construct one year mortality rates for individuals ages 40-62 by income percentiles. To assess the quality of the data we constructed, we compare our mortality rates to those published by the Social Security Administration for the nation in each census year (**Figure A2**). As expected, the raw mortality rates are lower in our data than they are for the nation for both men and women, but they closely follow the same trends. The same is true

for mortality rates in 1950 (though these are noisier) and for earlier census years. To make these more representative we use weights (see below).

To compare our estimates with Chetty et al. (2016), we follow their approach to compute period life expectancy. Mortality rates for 40-63 are observed in our data. Beyond the age of 63 years, mortality rates are estimated using the Gompertz model, which assumes that the log of mortality is linear in age after a certain age. The Gompertz model is fitted for each sex, income percentile, and census year. The data is strongly consistent with the Gompertz model overall and within income groups. For survivors at the age of 90 years, we assign sex-year-specific but income-independent mortality rates from the Social Security period tables. With these mortality rates, we construct period life expectancy by sex, income percentiles, and census year at age 40. The details for calculating the period life expectancy can be found in Chetty et al. (2016).

In addition, we also examine how cohort life expectancy relates to income. Because we observe the age at death of all individuals, this can be easily computed for the cohorts we study without making any assumptions, which is a unique advantage of our data. Recall that period and cohort life expectancies can differ a lot. Period life expectancies are a cross-sectional prediction of what the average age at death will be if the current population experiences the mortality rates observed in the current year. Cohort life expectancy is the actual average age at death that the cohort experiences. When mortality rates are falling, as was the case during the second half of the 20th century, cohort life expectancy greatly exceeds period life expectancy (**Table A2**).

A key issue for this study is whether the match rates in our data vary by income. To test this, we look at how matching rates vary by income percentile (**Figure A1**). Match rates are relatively stable across the income distribution particularly for women, and for individuals in households up to the 95th percentile of income, the differences in match rate by income are small. However, the match rates are 8.0 and 5.5 percentage points lower for men and women respectively (P<0.01) for individuals coming from households with incomes above the 95th percentile of income (vs. below the 95th percentile). We test whether our results are sensitive to how we measure income at the very top. We also conducted a small study to find the age at death of the missing individuals at the very top. Research assistants with experience in family history work were able to track 347 individuals out of 600 random individuals with missing data and incomes

in the top 5%. The mean age at death in this sample is 76.37 compared to 76.80 in the original data. This exercise suggests that the age at death of the missing rich is not very different.

c. Sample characteristics

To make the data representative we construct weights. We estimate the probability, $p(X_i)$, of a successful match between our sample and the 1940 full-count Census using a robit model, that estimates the likelihood of a match as a function of the covariates available in the 1940 Census (Newson & Falcaro, 2023; Seaman & White, 2013).[2] We then reweight matched observations by $p(X_i)^{-1}$ (IPW, rescaled to sum to one within the matched sample) so their covariate distributions match the 1940 Census. These IPW weights are used in all analysis.

Our final sample for 1940 includes about 10 million people ages 40-62 who are alive in 1941 and had non-missing household incomes in 1939 (about 8 million with positive income, and 2 million with zero income). We observe the ages at death of all these individuals, 72.3 for men and 79.2 for women conditional on being alive at the age 40 (**Table A2**). On average the mean age in this sample is 49.0 and the mean household income in 1940 is $1,636. Descriptive statistics show that the 1940 sample we obtain, even after reweighting, is not fully representative of the target population in 1940. For example, whites constitute 97.0% of the sample, instead of 91.8% in the 1940 Census and are 58.4% urban, compared to 64.9% in the 1940 Census. These patterns are typical of the genealogical data (e.g. Black et al., 2025).

Our final sample for 1950 includes 5 million individuals ages 40-63 alive in 1951 who had non-missing household incomes in 1949 (about 3 million with positive income and 2 million with positive income). This sample is smaller because as noted earlier the Census only collected income from the sample line respondents. In addition, we restrict attention to individuals who can be linked to 1940 so that we can obtain two measures of income for a subsample of the population. We also construct weights to make this sample representative.

[2] In addition, we further truncated IPW at the 99th percentile to stabilize the weights (Newson & Falcaro, 2023). We include the following covariates (X_i) available in the 1940 full-count census: sex, race, ethnicity, marital status, years of schooling, household total income, family size, labor force participation, occupation income scores, year of birth, state of residence, and place of birth. We estimated the model separately by sex.

The final sample for 1900 to 1930 includes 5, 8, 9 and 12 million individuals aged 40-63 alive in 1901, 1911, 1921 and 1931, who have a household occupation score measure. We also construct weights to make this sample representative.

For both men and women, the SSA period life expectancy at age 40 increased from 1900 to 1920 and rose from 1930 to 1950 (**Table A2**). This occurred despite most health innovations primarily saved children from dying (rather than adults) and likely saved less healthy children from dying. Additionally, we observed that women lived longer than men and that the sex gap grew substantially over the century as has been noted before (Goldin & Lleras-Muney, 2019).

III. The relationship between income and period life expectancy mid-century

a. Main results for 1940

Figure 1 plots the period one-year log mortality rates by age in 1941 for individuals in the 5th and 95th percentiles of income, corresponding to annual household incomes of $242 and $4,768 for men, and $170 and $4,605 for women. For both men (Panel A) and women (Panel B), we see that these mortality rates are noisy and although they are slightly higher for those in the 95th percentiles, the differences are small and statistically insignificant. This is quite different from the results reported by Chetty et al. (2016) which are shown in the same figure. For those results, here is a large and statistically significant difference between those in the 5th and 95th percentiles today. For both rich and poor, mortality rates are lower in recent times, but the decline is much larger for the top than for the bottom, resulting in very substantial gaps in the 2000s.

For both men and women period life expectancy at all levels of income grew substantially between 1940 and 2000s but much more at the top than at the bottom. This can be seen in **Figure 2** which plots the period life expectancy at age 40 by income percentile for 1940 and the 2000s. In 1940, period life expectancy was not related to income at any point in the distribution. The non-parametric line does not show any slope at any point. Indeed, a regression of period life expectancy against income percentiles has a small and statistically insignificant slope of 0.003 (s.e. 0.004) for men and -0.007 (s.e. 0.006) for women in 1940s. By contrast in the 2000s the slope is 0.114 (s.e 0.002) for men and 0.072 (s.e. 0.001) for women. These slopes are statistically significantly different across periods for both sexes (**Table A3**).

Overall, the gaps in period life expectancy by income in 1940 are very small, compared to those in the 2000s. In 1940 the period life expectancy at age 40 for men is 69.4 for the bottom 10% and 69.6 for the top 10%, a difference of only 0.2 years. In the 2000s the period life expectancy for men at the bottom 10% is 75.6 compared to 86.4 for the top 10%, a gap of 10.8 years. For women, the gap is of 0.3 years in the 1940 but of 7.1 years in the 2000s.

b. Robustness: using alternative measures of income

Given that we find no relationship between income and period life expectancy using the 1940 census, a primary concern is that this null result is driven by measurement error in income. Unlike Chetty et al. (2016) who use tax records, our data comes from self-reports in the census which may have measurement error and cause attenuation. One advantage we have over the data used by Chetty et al. (2016) is that we observe individuals with no income and we include those individuals in all the results in our study, except for Figure 1.

To assess the role of measurement error, we replicate our results using four alternative income measures: 1-individual wages and salary including 0s, 2-occupation scores, 3-LIDO scores, and 4-our own predicted income measure, which we obtain by regressing household income on age, age squared, sex, marital status, race, Hispanic origin, homeownership, years of schooling, and occupation score. The regression is estimated among individuals aged 30-45 in 1940.

We do not find any evidence that period life expectancy at age 40 increased with income using any of these measures (**Figure A3**). For men (Panel A), some of these measures in fact show a negative association between income percentiles and period life expectancy. For women (Panel B), we compute additional measures by looking at whether period life expectancy varies as a function of the husband's wages or occupation score (only among married women). The non-parametric lines show a lot more variation for women but still no apparent trend across income percentiles, except perhaps at the top. We more formally assess if there are significant differences in the slopes by regressing period life expectancy at age 40 on income percentiles (**Table A3** Panel A). We do not estimate any positive and statistically significant slopes regardless of the measure we use, and in fact for men, we estimate small and statistically negative slopes for most alternative measures.

A second concern is that in 1940 a large share of the population (about 42%) was still living in rural areas and engaging primarily in agriculture. It is very difficult to impute incomes among rural agricultural households and it would be difficult for agricultural household to report it, since they may be engaged in a substantial amount of self-production and barter. To assess if this is affecting our findings, we re-estimate our models separately for urban and rural populations (**Figure A4)**. At all income percentiles, men and women living in rural areas have *higher* life expectancies than those living in urban ones, though this is more consistently true for men than for women. This suggests that the urban penalty was still quite large in the US in the 1940s, consistent with recent results from Feigenbaum et al. (2020). However, we find no strong relationship between income and life expectancy within urban or rural locations, although for men we do estimate a small positive association within urban locations (slope of 0.017, s.e. 0.005), driven primarily by increases at the bottom of the income distribution, though even this difference is still an order of magnitude smaller than the association in the 2000s. But there is no slope in rural areas. These results suggest that the urban penalty partly explains the lack of an association between income and period life expectancy, since within urban areas a small gradient is visible particularly for men.

A third concern is that we are investigating the relationship between income *percentiles* and period life expectancy, which may obscure the relationship between income *levels* and life expectancy. This is particularly concerning because many papers find that the relationship between income and life expectancy is concave, with a steeper slope among the poor than the rich (Lleras-Muney et al., 2025). However, when we re-estimate our models by income levels (instead of income percentiles) we do not find positive associations either regardless of the measure of income we use (**Table A3** Panel B).

A fourth issue is that our match rates are lower among the rich, particularly above the top 95th percentile of income. This could bias the mortality rates of the rich and potentially result in attenuated relationships. However, our results are very similar if we exclude the top 5% of the income distribution (**Table A3** Panel C).

A final concern is that the 1940 income data, even if reported accurately, does not properly describe lifetime incomes because the economic conditions during the Great Depression were unusual, even at the tail end in 1939 during which the unemployment rate was estimated to be

around 15%. We investigate this by looking at the relationship between income and period life expectancy in 1950 now.

c. Main Results for 1950

The IPUMS Multigenerational Longitudinal Panel matches individuals in the 1940 and 1950 censuses and has provided a public crosswalk (Helgertz et al., 2022; Ruggles et al., 2025). The 1950 Census also asked individuals about their income, although it did so for sample line respondents, about 1 in six respondents. Thus, we have a sample of 5 million individuals aged 40-63 in 1951 for whom we observe both income and lifespan.

The 1950 data, though smaller, has two main advantages. First the economic conditions in 1949 were markedly better than those in 1939: the unemployment rate hovered between 4.3% and 6.6% that year (U.S. Bureau of Labor Statistics), two to three times smaller than the rate in 1939. Therefore, the measure of income might be closer to permanent income. Second, as discussed earlier, the census asked more detailed questions regarding income and thus collected a more complete measure of income that includes income from various sources.

This data then allows to investigate two issues. First, is there an income gradient in period life expectancy in 1950? Knowing this allows us to understand when gradients emerged, since we know they were visible by 1960. Second, for the sample of individuals that survives from 1940 to 1950, we can combine this more complete measure of income with the 1940 measure of income to create an average measure of income that may more accurately represent lifetime incomes and contain less measurement error.

We do not find there is an income gradient in 1950 (**Figure 3** Panel A). The estimated slope of the relationship is 0.011 (s.e. 0.005) for men and 0.006 (s.e. 0.010) for women. These are small and statistically insignificant slopes that do not meaningfully differ from those we estimate for 1940 for either sex. The non-parametric lines do not show any positive slope at any point in the income distribution.

For those alive in both 1941 and 1951, we compute the average of the 1940 and 1950 incomes to create a measure of lifetime earnings, and we recompute period life expectancy by income percentile using this alternative measure of income. We do not find an income gradient using this average measure either for either sex (**Figure A5**). In fact, the relationship, measured

by the slope of the regression of period life expectancy on income percentile is almost identical in 1940 and 1950. None of these slopes are large or statistically significant.

Thus, we conclude that the lack of an association between income and period life expectancy is not due to measurement error in income, and furthermore that this association was not significant in the 1940s but emerged sometime in the 1950s after which it started to grow.

d. Results for 1900-1950

One interesting question that arises from these results is whether the lack of a gradient in the 1940 and 1950 is the exception to an otherwise rather universal pattern of advantage among the rich. The war and post war years were unique period in many ways but most interestingly it was a period of wage compression and historically low income and wealth inequality. Thus, a possible hypothesis is that if we go back further in time, to periods in which income inequality was larger we will observe a large gradient again. Income inequality in the years prior to the great depression (before 1929) was quite high, similar to or higher than it was in the 2000s (Piketty et al., 2017).

To investigate this, we follow the previous literature and use occupations to impute income in the 1900 to 1930 censuses. We then report period life expectancy gradients in household occupation score for each census year and we plot them alongside our estimates of the gradients in income for 1940, 1950 and the 2000s (Chetty et al., 2016).

We actually find *negative* income gradients in 1900 and 1930 for men (**Figure 4** Panel A**)**. While the gradients are small in magnitude, they are statistically significant when we use household occupation scores. In other words, in this 30-year period individuals in more highly paid occupations lived *shorter* lives. The figure also shows the gradient estimated by Chetty et al. (2016) in the 2000s. Relative to the small and unchanging associations we measure from 1910 to 1950, it is clear that there was a very large increase that occurred between 1950 and 2000.

For women (**Figure 4** Panel B), there are no significant gradients regardless of the measure we use, except for 1910 and 1940 when using the household LIDO score. Again, there was a very large increase in the gradient when comparing with the gradient estimated by Chetty et al. (2016). These conclusions hold even if we include farmers (**Figure A6**).

These results showing a negative gradient for men early in 1900 are similar to those obtained by work in Sweden, showing that men in highly paid occupations did not live longer than those in low paid occupations until after WWII (e.g. Dribe & Eriksson, 2025), and with work looking at famous individuals, such as politicians, which also finds that in many countries they did not live longer lives until mid-century (e.g. Clarke et al., 2022). However, our results stand in contrast to theirs for women: they document positive gradients for women going back to 1850 but we find no such association here.

IV. The relationship between income and cohort life expectancy over time

Previous work documents that over time the cross-sectional relationship between income and mortality or life expectancy has increased. Thus, we predict that cohort life expectancy will show a more positive association with income than the period life expectancy and that this association will grow across cohorts, as richer cohorts will benefit more from recent declines in mortality. We test this now.

a. Results for 1940

We first document this for 1940 (**Figure A7**)**.** For both men (Panel A) and women (Panel B) cohort gradients exceed period ones. While the period measure does not increase with income percentiles, the cohort measure does. Indeed, the slope of the relationship is now positive and significant for both men (0.019, s.e. 0.002) and women (0.026, s.e. 0.001) though it is still modest compared to the ones estimated by Chetty et al. (2016) for period life expectancy in the 2000s, which is about 5 times larger. Interestingly and contrary to what period measures show (see literature review in Appendix 1), the cohort slope is higher for women than men.

We re-estimate these slopes using our alternative measures of income (**Table A3** Panel A) to see if the slope is attenuated. There is a suggestion that it might be because the slope estimate is larger when we use our predicted income measure or when we use wage income. But even these slopes (the largest is 0.034 for men when we use wage income, and 0.038 for women when we use predicted household income) are much smaller than contemporary slopes of 0.11 for men and 0.07 for women in Chetty et al. (2016).

We also re-estimate the relationship in levels, dropping the top 5% of the income distribution who have low matching rates (**Figure A7**). We observe that in levels there is also a positive and

statistically significant association between income and lifespan for both men (slope 0.554, s.e. 0.048) and women (slope 0.651, s.e. 0.047). The cohort life expectancy is 71.5 for the bottom 10% and 73.0 for the top 10% for men, a gap of 1.5 years of life. For women the gap is larger (78.4 for the bottom 10% vs. 80.4 for the top 10%) and about 2.0 years

b. Tracing the emergence of the income gradient.

We now estimate cohort gradients in adult life expectancy for all census years 1900-1950 (**Figure 5**). As before, we use occupation-based household incomes for 1900-1930 and income for 1940 and 1950. The results show that for men (Panel A), cohort gradients in life expectancy were negative and statistically significant from 1900-1950 when using household occupation score, and from 1900 to 1930 when using household LIDO score. For women (Panel B) the pattern is similar when we use occupation scores, but the LIDO score results in zero or slightly positive gradients in 1900-1930, and moreover the gradient in 1950 is very small for all income measures in contrast to the gradients for men which are larger in 1950. The results are similar when we include farmers in the analysis (**Figure A8**) and they are similar when we investigate these relationships non-parametrically (**Figure A9**).

An important question here is whether we can conclude from these results that the income gradient was indeed negative. Here we can conduct a couple of exercises to assess how income gradients and occupation gradients compare. The 1940 and 1950 census results are informative about this. When we use household income instead of occupation-based household incomes, the gradients are positive in 1940 and 1950, and statistically bigger than the occupation-based gradients.

This evidence suggests that occupation-based income gradients do not fully account for income gradients. This could occur because there is significant variation in incomes within occupations that the occupation scores cannot account for. We can verify this using the 1940 individual data: we can regress individual lifespan at 40 on income and occupation scores. If incomes within occupation affect lifespan, then the coefficient on income will still be significant even when we control for occupation scores or occupation dummies. Indeed, this is what we find (**Table A4**).

If the occupation scores underestimate the relationship in the 1900-1930 period to the same extent as they do in 1940 and 1950 then the gradients may have been closer to zero throughout 1900-1930, or even positive. To assess this, we attempt to "correct" the earlier gradients by modeling household income as a linear function of household occupation score and an unobserved residual component. Using the covariance structure between household income, occupation score, and the residual term in 1940, we impute the implied income level gradients for earlier years. As shown in Panels C and D, the predicted gradients for 1940 and 1950 align closely with the actual data, and the predicted income level gradients were near zero prior to 1930 but turned positive in 1930 for both men and women. Note however that we can only do this exercise with income levels (not percentiles), which explains why the results are somewhat different in Panels a and b compared to panels c and d. Despite these small differences, the conclusions are similar: back in 1900 the cohort income gradient was small and/or insignificant and it emerges only later in the century.

Why is there a gradient in the cohort measure but not the period measure for 1940 and 1950? Recall that we are estimating cohort gradients in life expectancy at age 40, so that the gradients in 1930 may reflect events occurring in 1970 or later, in addition to selection/survival effects. This points primarily then to the benefit of cardiovascular disease innovation which according to Cutler et al. (2006) was responsible for at least 1/3 of the life expectancy gains since 1950 and which primarily benefitted men, explaining why cohort gradients for women are smaller than those for men.

V. Discussion

In this paper we use a large data set constructed by matching the 1900-1950 censuses with profiles on the Family Tree at familysearch.org to study the association between household income and lifespan. We document that, contrary to what has been documented for the US today, there is no association between household income and period life expectancy in any year from 1900 to 1950 for either men or women. The income gradient in lifespan thus emerges in the 1950s.

What distinguishes the second half of the century from the first half of the century primarily is the increasing availability of modern medicine which individuals must access through physicians and hospitals and pay for out of pocket and or through their insurance. The gains in

life expectancy in the earlier part of the century were primally obtained through public health efforts to contain infectious diseases. These efforts consisted in cleaning the water and separating it from sewage, disseminating of information and practices on how to avoid disease (boiling water, cleaning hands) and applying various techniques to prevent infectious diseases (such as the eradication of hookworm). Government and private non-profit institutions paid for these efforts which benefited large swaths of the population, who paid for these (if at all) only indirectly through taxes. In the second part of the century instead innovations were primarily made in the form of individual treatments that were sold in the market and from which innovators typically profited enormously, as was the case with the development of antibiotics (mostly developed from 1942 to 1960, the so called golden age of antibiotics) and of medicines to treat heart disease, such as statins to lower cholesterol and high blood pressure medications, developed later in the 1960-1980 period. Together with the existing evidence from the literature, the results suggest that the emergence of highly effective but expensive modern medicine is greatly responsible for the rises in lifespan inequality we observe today.

The rising period gradient results in a larger and also rising cohort gradient. We find that cohort gradients went from negative or insignificant in 1900 to positive by 1950. The gradient turns positive for cohorts that were 40 and older after 1970 which suggests that in fact innovation in cardiovascular disease is primarily responsible for the rising gradient.

We uncover other interesting patterns worth investigating further. First, in the process of investigating the income gradients, we found that in 1940 there was still a significant urban penalty and this was true for both our period and cohort life expectancy measures. Second, we find some important sex differences in the cohort gradients which show larger gradients for men than women consistent with the greater effect of cardiovascular disease technologies on men.

In conclusion, our study suggests the association between income and life expectancy likely began to emerge in the 1950s. Future research with better data could directly show the income gradients in LE from 1950-2000, further tracing their emergence and exploring the potential drivers of the substantial income gradients we observe today.

References

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2014). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy*, *122*(3), 467-506. https://doi.org/10.1086/675805

Arias, E., & Xu, J. (2022). US State life tables, 2019.

Backlund, E., Sorlie, P. D., & Johnson, N. J. (1996). The shape of the relationship between income and mortality in the United States: Evidence from the National Longitudinal Mortality Study. *Annals of Epidemiology*, *6*(1), 12-20. https://doi.org/https://doi.org/10.1016/1047-2797(95)00090-9

Bailey, M., Lin, P. Z., Mohammed, A. R. S., Mohnen, P., Murray, J., Zhang, M., & Prettyman, A. (2023). The creation of LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database project. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *56*(3), 138-159. https://doi.org/10.1080/01615440.2023.2239699

Bailey, M. J., Cole, C., Henderson, M., & Massey, C. (2020). How Well Do Automated Linking Methods Perform? Lessons from US Historical Data. *Journal of Economic Literature*, *58*(4), 997–1044. https://doi.org/10.1257/jel.20191526

Black, S. E., Duzett, N., Lleras-Muney, A., Pope, N. G., & Price, J. (2025). Intergenerational correlations in longevity.

Case, A., & Deaton, A. (2021). Life expectancy in adulthood is falling for those without a BA degree, but as educational gaps have widened, racial gaps have narrowed. *Proc Natl Acad Sci U S A*, *118*(11). https://doi.org/10.1073/pnas.2024777118

Chapin, C. V. (1924). Deaths among Taxpayers and Non-Taxpayers, Income Tax, Providence, 1865. *Journal of Public Health Policy*, *20*(2), 227-234.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*, *315*(16), 1750-1766. https://doi.org/10.1001/jama.2016.4226

Clarke, P. M., Tran-Duy, A., Roope, L. S. J., Stiles, J. A., & Barnett, A. G. (2022). The comparative mortality of an elite group in the long run of history: an observational analysis of politicians from 11 countries. *Eur J Epidemiol*, *37*(9), 891-899. https://doi.org/10.1007/s10654-022-00885-2

Couillard, B. K., Foote, C. L., Gandhi, K., Meara, E., & Skinner, J. (2021). Rising Geographic Disparities in US Mortality. *Journal of Economic Perspectives*, *35*(4), 123–146. https://doi.org/10.1257/jep.35.4.123

Cristia, J. P. (2009). Rising mortality and life expectancy differentials by lifetime earnings in the United States. *J Health Econ*, *28*(5), 984-995. https://doi.org/10.1016/j.jhealeco.2009.06.003

Cutler, D., Deaton, A., & Lleras-Muney, A. (2006). The Determinants of Mortality. *Journal of Economic Perspectives*, *20*(3), 97–120. https://doi.org/10.1257/jep.20.3.97

Dribe, M., & Eriksson, B. (2025). The Long-Term Development of the Mortality Gradient: Socioeconomic Differences in Adult Life Span of Swedish Cohorts 1841–1920. *Population and Development Review*. https://doi.org/10.1111/padr.70015

Duggan, J. E., Gillingham, R., & Greenlees, J. S. (2008). Mortality and Lifetime Income: Evidence from U.S. Social Security Records. *IMF Staff Papers*, *55*(4), 566-594. https://doi.org/10.1057/imfsp.2008.21

Duleep, H. O. (1989). Measuring socioeconomic mortality differentials over time. *Demography*, *26*(2), 345-351.

Feigenbaum, J., Hoehn-Velasco, L., & Wrigley-Field, E. (2020). *Did the Urban Mortality Penalty Disappear? Revisiting the Early Twentieth Century's Urban-Rural Mortality Convergence*.

Goldin, C., & Lleras-Muney, A. (2019). XX > XY?: The changing female advantage in life expectancy. *J Health Econ*, *67*, 102224. https://doi.org/10.1016/j.jhealeco.2019.102224

Greville, T. N. E. (1947). *United States life tables and actuarial tables, 1939-1941*. US Government Printing Office.

Hadley, J., & Osei, A. (1982). Does income affect mortality? An analysis of the effects of different types of income on age/sex/race-specific mortality rates in the United States. *Med Care*, *20*(9), 901-914. https://doi.org/10.1097/00005650-198209000-00004

Hagen, J., Laun, L., Lucke, C., & Palme, M. (2025). The rising income gradient in life expectancy in Sweden over six decades. *Proc Natl Acad Sci U S A*, *122*(14), e2418145122. https://doi.org/10.1073/pnas.2418145122

Haider, S., & Solon, G. (2006). Life-Cycle Variation in the Association between Current and Lifetime Earnings. *American Economic Review*, *96*(4), 1308–1320. https://doi.org/10.1257/aer.96.4.1308

Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S., & Fitch, C. A. (2022). A New Strategy for Linking U.S. Historical Censuses: A Case Study for the IPUMS Multigenerational Longitudinal Panel. *Hist Methods*, *55*(1), 12-29. https://doi.org/10.1080/01615440.2021.1985027

Kinge, J. M., Modalsli, J. H., Øverland, S., Gjessing, H. K., Tollånes, M. C., Knudsen, A. K., Skirbekk, V., Strand, B. H., Håberg, S. E., & Vollset, S. E. (2019). Association of Household Income With Life Expectancy and Cause-Specific Mortality in Norway, 2005-2015. *JAMA*, *321*(19), 1916-1925. https://doi.org/10.1001/jama.2019.4329

Kitagawa, E. M., & Hauser, P. M. (1973). *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*. Harvard University Press. https://doi.org/doi:10.4159/harvard.9780674188471

Lantz, P. M., House, J. S., Lepkowski, J. M., Williams, D. R., Mero, R. P., & Chen, J. (1998). Socioeconomic Factors, Health Behaviors, and MortalityResults From a Nationally Representative Prospective Study of US Adults. *JAMA*, *279*(21), 1703-1708. https://doi.org/10.1001/jama.279.21.1703

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *J Health Soc Behav*, *Spec No*, 80-94.

Lleras-Muney, A., Price, J., & Yue, D. (2022). The association between educational attainment and longevity using individual-level data from the 1940 census. *J Health Econ*, *84*, 102649. https://doi.org/10.1016/j.jhealeco.2022.102649

Lleras-Muney, A., Schwandt, H., & Wherry, L. R. (2025). Poverty and Health. *Annual Review of Economics*, *17*(Volume 17, 2025), 31-56. https://doi.org/https://doi.org/10.1146/annurev-economics-081624-083218

McDonough, P., Duncan, G. J., Williams, D., & House, J. (1997). Income dynamics and adult mortality in the United States, 1972 through 1989. *American Journal of Public Health*, *87*(9), 1476-1483.

Meara, E. R., Richards, S., & Cutler, D. M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981-2000. *Health Aff (Millwood)*, *27*(2), 350-360. https://doi.org/10.1377/hlthaff.27.2.350

Milligan, K., & Schirle, T. (2021). The evolution of longevity: Evidence from Canada. *Canadian Journal of Economics/Revue canadienne d'économique*, *54*(1), 164-192. https://doi.org/https://doi.org/10.1111/caje.12497

Newson, R. B., & Falcaro, M. (2023). Robit regression in Stata. *The Stata Journal*, *23*(3), 658-682.

Pappas, G., Queen, S., Hadden, W., & Fisher, G. (1993). The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986. *New England Journal of Medicine*, *329*(2), 103-109. https://doi.org/doi:10.1056/NEJM199307083290207

Piketty, T., Saez, E., & Zucman, G. (2017). Distributional National Accounts: Methods and Estimates for the United States*. *The Quarterly Journal of Economics*, *133*(2), 553-609. https://doi.org/10.1093/qje/qjx043

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies*, *29*(2), 231-248.

Rogot, E., Sorlie, P. D., & Johnson, N. J. (1992). Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study. *Public health reports*, *107*(4), 457.

Ruggles, S., Drew, J., Fitch, C. A., Hacker, J. D., Helgertz, J., Nelson, M. A., Ozder, N., Sobek, M., & Warren, J. R. (2025). The IPUMS multigenerational longitudinal panel: progress and prospects. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1-14. https://doi.org/10.1080/01615440.2025.2514110

Saavedra, M., & Twinam, T. (2020). A machine learning approach to improving occupational income scores. *Explorations in Economic History*, *75*. https://doi.org/10.1016/j.eeh.2019.101304

Schalick, L. M., Hadden, W. C., Pamuk, E., Navarro, V., & Pappas, G. (2000). The widening gap in death rates among income groups in the United States from 1967 to 1986. *International Journal of Health Services*, *30*(1), 13-26.

Schwandt, H., Currie, J., von Wachter, T., Kowarski, J., Chapman, D., & Woolf, S. H. (2022). Changes in the Relationship Between Income and Life Expectancy Before and During the COVID-19 Pandemic, California, 2015-2021. *JAMA*, *328*(4), 360-366. https://doi.org/10.1001/jama.2022.10952

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research*, *22*(3), 278-295.

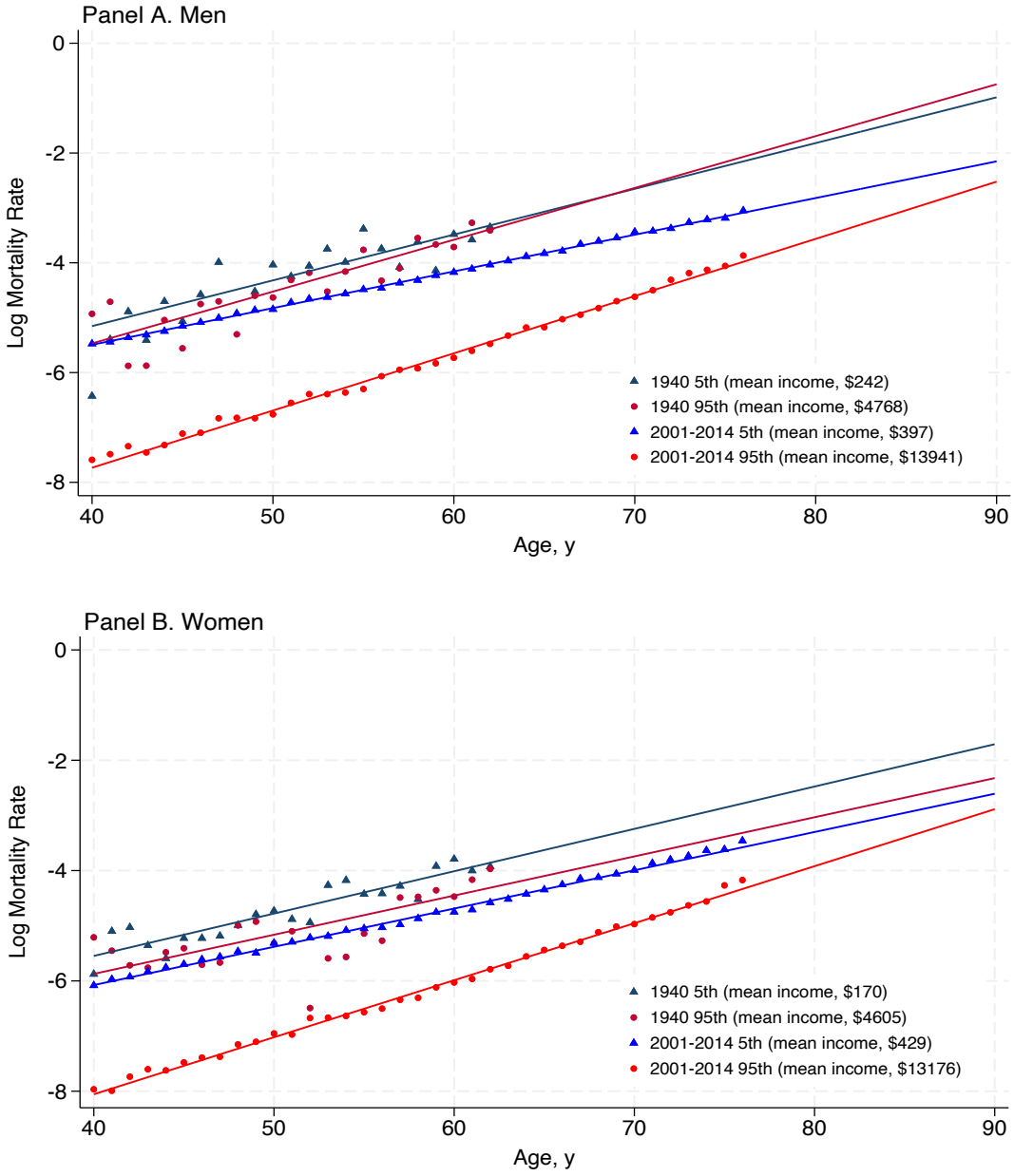
Singh, G. K., & Siahpush, M. (2006). Widening socioeconomic inequalities in US life expectancy, 1980-2000. *Int J Epidemiol*, *35*(4), 969-979. https://doi.org/10.1093/ije/dyl083

Sorlie, P. D., Backlund, E., & Keller, J. B. (1995). US mortality by economic, demographic, and social characteristics: the National Longitudinal Mortality Study. *Am J Public Health*, *85*(7), 949-956. https://doi.org/10.2105/ajph.85.7.949

U.S. Bureau of Labor Statistics. Unemployment Rate [UNRATE], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UNRATE, October 9, 2025.

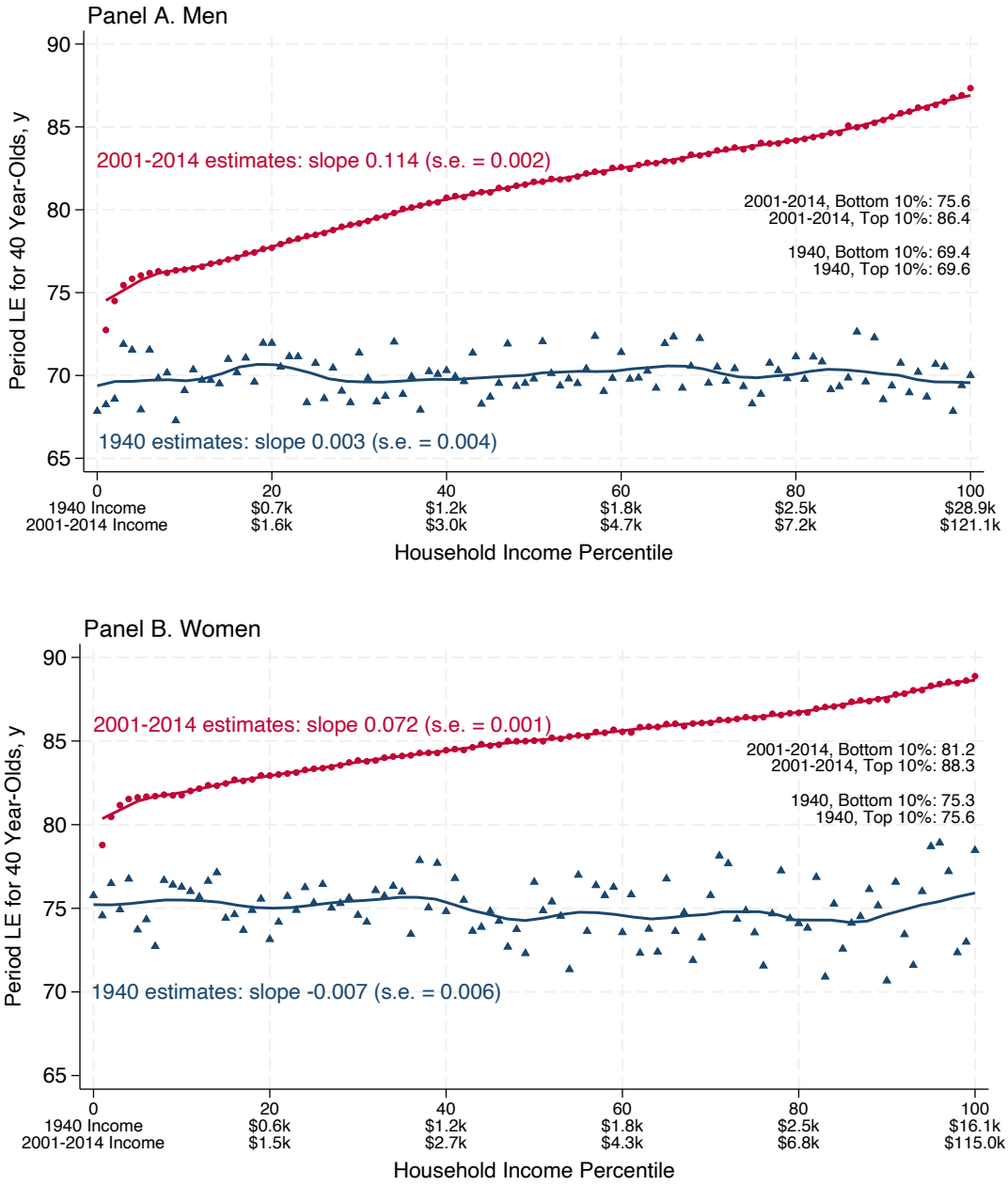
Figures and Tables

Figure 1. Gompertz Approximations in 1940 and 2000s in the 5th and 95th Income Percentiles



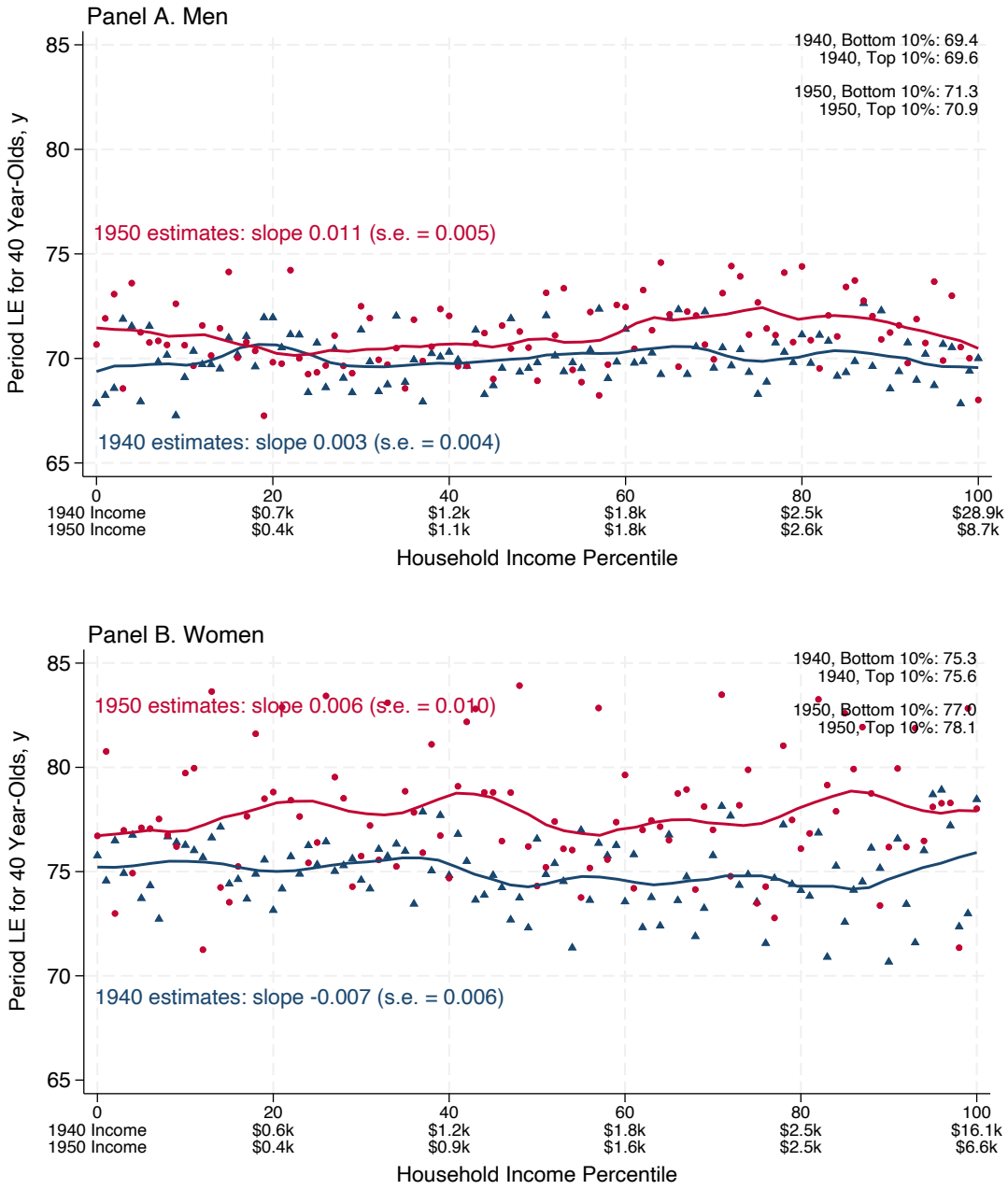
Notes: When regressing the log of mortality rates on age, a dummy for the 95th percentile, and the interaction between age*95th percentile dummy, the *P*-value of the interaction term is 0.911 for 1940 and <0.000 for 2001-2014 among men, and 0.301 for 1940 and <0.000 for 2001-2014 among women. Income values are adjusted to 1939 prices.

Figure 2. Period Life Expectancy for 40-Year-Olds in 1940 and 2000s by Household Income Percentile



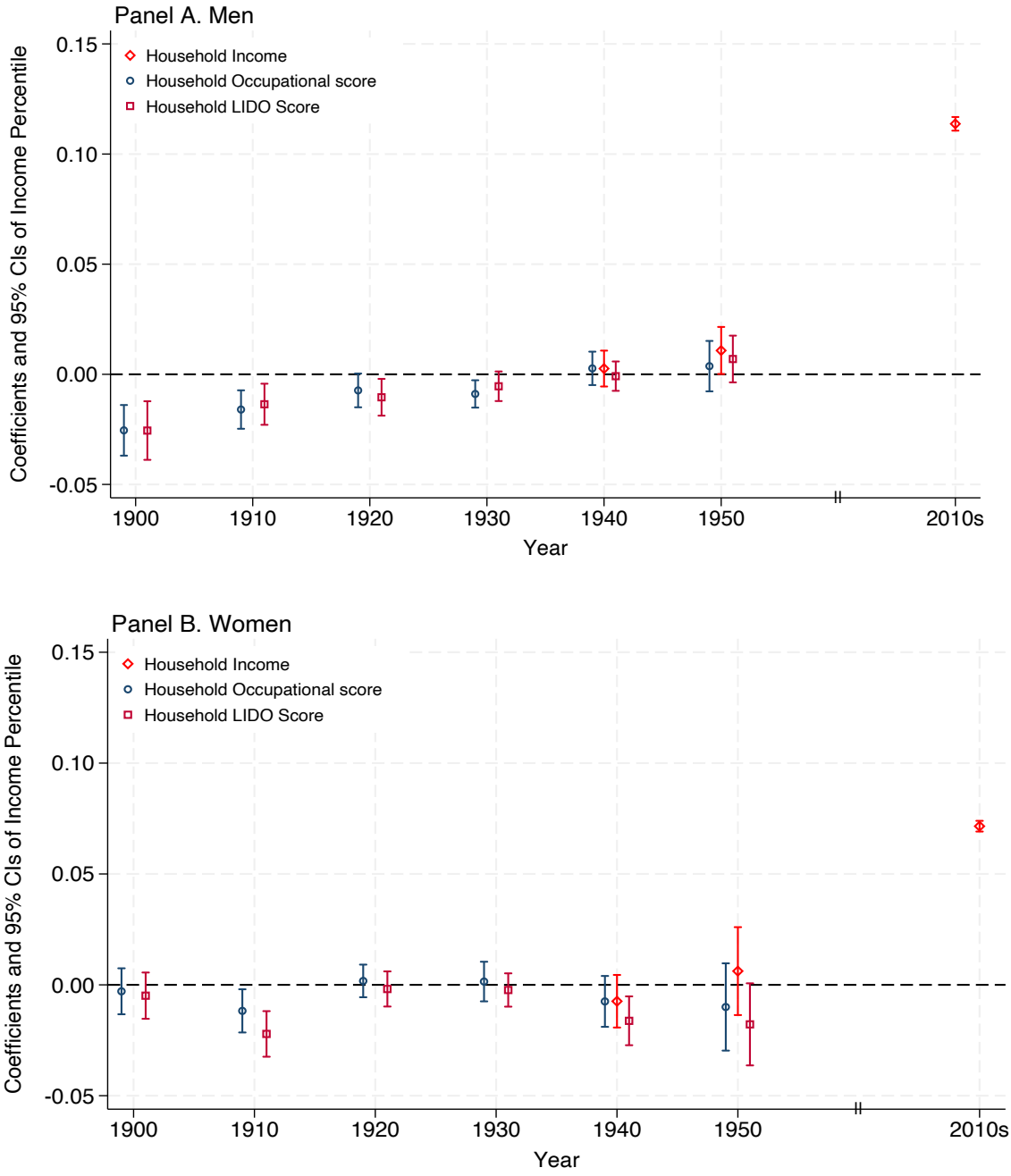
Notes: Percentiles are defined within the positive income distribution. For 1940, an additional data point for zero income (x-axis value = 0) is included. The nonparametric line and OLS slope estimates are based on regressions that incorporate this zero-income observation. The reported mean life expectancy in the text reflects both the bottom 10% and the zero-income group. Income values are adjusted to 1939 prices.

Figure 3. Period Life Expectancy for 40-Year-Olds in 1940 and 1950 by Household Income Percentile



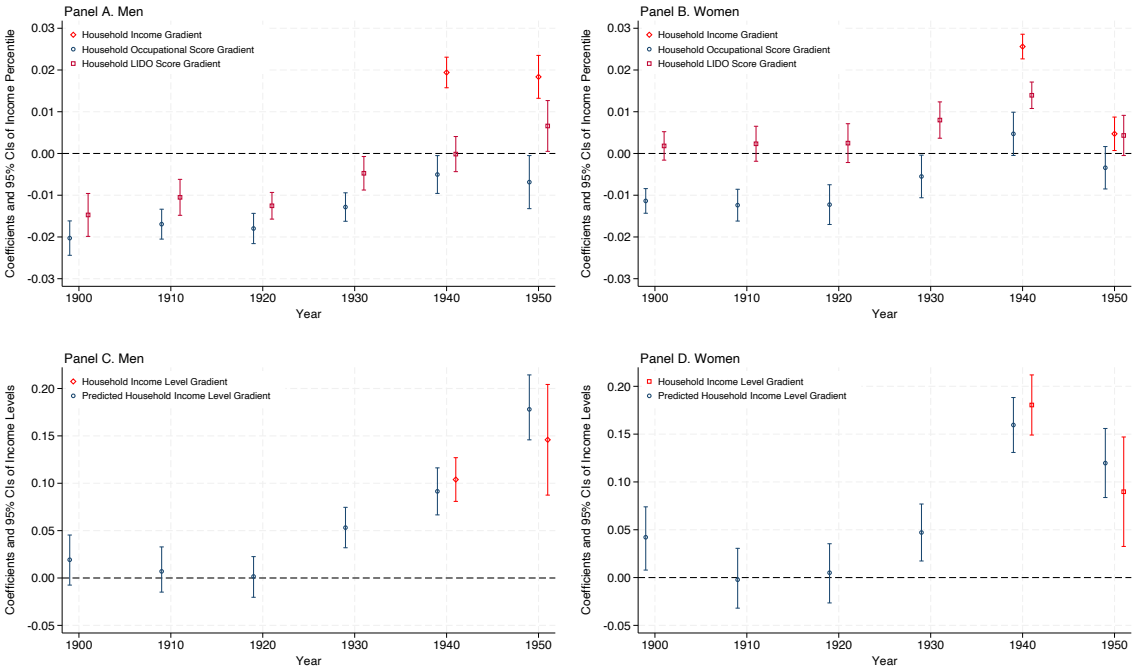
Notes: Percentiles are defined within the positive income distribution. Additional data points for zero income (x-axis value = 0) are included. The nonparametric line and OLS slope estimates are based on regressions that incorporate these zero-income observations. The reported mean life expectancy in the text reflects both the bottom 10% and the zero-income group. Income values are adjusted to 1939 prices.

Figure 4. Period Life Expectancy Income Gradient for 40-Year-Olds since 1900



Notes: For each income measure and year, period life expectancy is estimated by percentiles of the corresponding income measure for each sex. Coefficients and 95% confidence intervals are obtained from separate regressions of estimated period life expectancy on percentiles of the corresponding income measure. Weights are applied during the analysis.

Figure 5. Cohort Life Expectancy Income Gradient for 40-Year-Olds since 1900



Notes: For each income measure and year in Panel A and B, cohort life expectancy is estimated by percentiles of the corresponding income measure for each sex. Coefficients and 95% confidence intervals are obtained from separate regressions of estimated cohort life expectancy on percentiles of the corresponding income measure. Weights are applied during the analysis. In Panel C and D, the actual household income level gradient (red) is obtained by regressing cohort life expectancy on household income ($1000s). To predict household income gradient (blue), we model household income (y) as a linear function of household occupation score (x) and an unobserved residual component (z), such that $y = \beta x + z$. This implied income level gradient is given by $\frac{cov(LE,y)}{var(y)} = \frac{\beta cov(LE,x)+cov(LE,z)}{\beta^2 var(x)+var(z)+2\beta cov(x,z)}$. To estimate the income level gradient for each year for each sex, we compute $cov(LE,x)$ and $var(x)$ using year-specific data while the remaining right-hand side terms are estimated using 1940 data. The Confidence intervals are obtained via bootstrap resampling.

Appendix 1

I. Previous literature.

a. Income, mortality rates and life expectancy since 1960

Previous work studying income gradients in life expectancy in the US shows these have increased in recent times, though it is difficult to fully ascertain the magnitude of these changes because studies differ in the outcome measures and their income measures. Most papers in the literature report either gaps or ratios in mortality rates by age, and abstain from reporting estimates of life expectancy gaps, which often require modelling assumptions. Using census data matched to death certificates, Kitagawa and Hauser (1973) reported that in 1960 the standardized mortality ratio (the number of deaths over the number of expected deaths) declines with income going from 1.51 for 1.14 for individuals with incomes under $2k, 1.2 for those with incomes ranging from $2-4k, and was 0.84 for those with incomes greater than $10k or more for white men between the ages of 25 and 64. For women these ratios went from 1.05 to 1.02 and 0.92. Using the National Longitudinal Mortality Study NLMS (which matched data from various Current Population Surveys from 1979-1985 with death certificates from the National Death Index), Duleep (1989) reported that in 1973-1978 these ratios were 1.59, 1.79 to 0.71 for white men. Using survey data from the National Interview Survey (NIHS) matched to death records, Pappas et al. (1993) reported that these ratios for white men were 2.5 for under $11k, 0.6 for $25k or more in 1986. Schalick et al. (2000) report similar changes between 1967 and 1986 using similar data. A number of other studies have simply reported mortality rates by income at a point in time for various sub-populations and investigated other aspects of the relationship such as whether it is linear, different by race, or accounted for by education and other observables (Backlund et al., 1996; Duleep, 1989; Hadley & Osei, 1982; Lantz et al., 1998; Sorlie et al., 1995).

These changes in relative ratios suggest that income gradients widened considerably since 1960. But it is difficult to know how large these changes were in terms of life expectancy. A handful of studies do report estimates of differences in life expectancy.[3] Using the NLMS, Rogot

[3] We ignore here papers studying the relationship between aggregate income and life expectancy, which started with the famous study by Preston (1975) documenting a positive and concave relationship between a country's GDP and its life expectancy. Papers in the US looking at the relationship between aggregate county or state income and life expectancy include the early work of Hadley & Osei (1982) or Singh & Siahpush (2006). Most recently Schwandt et al. (2022) looked at the relationship between census tract (grouped by median household income) income and census tract period life expectancy at birth in California from 2015 to 2021 and Couillard et al. (2021) looked at the

et al. (1992) reported that around 1979-1985 for white men the difference in period life expectancy between the highest (50k+) and the lowest income groups (<5k) was about 10 years at age 25, 8 years at age 45, and 4 years at age 65. For white women the corresponding gaps were smaller (4.3 years, 4 years, and 1). Cristia (2009) uses data from the Survey of Income and Program Participation (SIPP) to show that the gap in period life expectancy (between ages 35 and 76) between the top and bottom quantiles of income increased by 0.8 years between 1983-1997 and 1998-2003 for men, and 0.7 years for women. Note however that these later numbers are not quite comparable to the Rogot et al. (1992) ones because Cristia (2009) only computes partial life expectancy numbers.

The most prominent recent study on the question was conducted by Chetty et al. (2016). It constructed mortality rates by income levels by matching tax records of individuals to death certificates for the 2001–2014 period. This study vastly improved upon previous work because it included information on more than 100 million adults in the US, a vastly larger sample compared to e.g. Kitagawa and Hauser (1973) paper which included about 62,000 individuals, or the Rogot et al. (1992) work which included about 800,000 individuals; though it excluded individuals not reporting income to the IRS. It also had access to significantly better income data. They document significant gaps in life expectancy at age 40, with men in top 99% living a full 14.6 years longer than men in the bottom 1% of the income distribution (the gap for women is of about 10 years). This study also investigated trends and found significant increases in the adult life expectancy gaps between the top and bottom ventiles of income between 2001 and 2014. These are difficult to compare to those in earlier studies because they focus on (life expectancy at age 40) and because they focus on percentiles of income. Nevertheless, within the study gradients did rise significantly since 2000.

Outside of the US two studies have also found that income gradients have increased in recent times. Kinge et al. (2019) use administrative data from Sweden and essentially replicate the Chetty et al. (2016) analysis in that country. Surprisingly they find that income gradients in life expectancy at age 40 are also large (though a bit smaller) in Sweden: there is 13.8-year gap (men), 8.4-year

evolution of state income and state life expectancy since 1968 and show that state income was not strongly correlated with state life expectancy in 1968 but it was strongly correlated with state income by 2019. Aggregate studies present additional challenges since they confound the effect of individual income and average income and may also misrepresent the shape of the relationship between income and health measures.

gap (women) between top and bottom 1% of income. Similar to the US case, these gaps increased from 2005 to 2015. Hagen et al. (2025) further extend these results back to 1960, when the gap between top and bottom income percentile was 3.5 years for men and 3.8 years for women, much smaller than the gap in the 2010s of 10.9 for men and 8.6 for women. A study using Canadian data however found smaller gradients that have not widened over time (Milligan & Schirle, 2021).

In sum the literature has documented that since 1960 there have been significant and widening differences in mortality by income in the US. In general, across studies these gaps by income appear to be smaller for women than men, and smaller at older ages. Gaps in adult period life expectancy by income have also risen but the exact magnitudes of these increases and particularly their timing is less clear in the US. All studies we reviewed found income gradients in either mortality or life expectancy, but all start in 1960. Consistent with these findings, research using education also finds large and growing education gradients in adult lifespan going since the 1960s.[4] The earliest mortality gradients by education are reported by Lleras-Muney et al. (2022) who find small gaps in 5-year mortality by education in 1950.[5]

b. Secular trends

Studies investigating how gradients have evolved over a longer period cannot observe income and thus typically rely on other measures to assess the gaps between the rich and the poor. That literature cast doubts on the assertion that the rich have always lived longer than the poor. There are two types of studies looking at secular trends. The first relies on measures of occupation to proxy for resources. The best studies in this area use data from Europe and combines census data, where occupation is reported with register data in the which the birth and death dates of individuals can be observed. The best evidence again comes from Sweden which has one of the oldest and most accurate vital registration systems in the world and can track mortality by occupation from 1840 onwards. Here recent work shows that men in highly paid occupations did not live longer than those in low paid occupations until after WWII (e.g., Dribe and Eriksson (2025)) suggesting

[4] The association between education and life expectancy among adults has grown substantially since 1960 in the US. For example, Kitagawa and Hauser reported a gap in life expectancy of 1.5 (women)-3 (men) years between those with college degree and individuals with only some high school (Table 2.4, chapter 2). The gap in life expectancy between adults with some college and high school dropouts was about 3 years in 1980 but it was 7 years in 2000 (Meara et al. 2008). More recently Case and Deaton (2021) calculated that the gap in life expectancy at age 25 between those with college and those with less than college rose from 2.6 years in 1992 to 6.3 years in 2018.
[5] See **Figure A6** of that paper.

that class gaps in lifespan are a relatively recent phenomenon, though it would appear that women from higher classes did, even though their disparities were smaller in the past.

Another strand of work looking at secular trends has investigated the lifespan of famous individuals and has compared it to that of the population. This work assumes that famous individuals have greater access to resources than the average person in the population. For example, Clarke et al. (2022) show that in 11 different now-rich countries, politicians did not live longer than average until the 1940s at which point a gap emerged and increased steadily thereafter. Both approaches suggest that in socio-economic gradients are primarily a post-WWII phenomenon.

However, these approaches have some important limitations. Occupations cannot be easily ranked over time: for example, famers are relatively poor today but there were many rich farmers in the past. Additionally, occupations such as farming may in fact confound the effects of income with the effects of living in rural areas: up until recently individuals living in cities died at much greater rates than those living in rural areas who were frequently engaged in farming related work. Similarly, it is difficult to assess the work looking famous individuals. The differences over time in their mortality might reflect changes in who become famous, and moreover individuals who live longer are more likely to become famous. Nevertheless, these studies do suggest that the 1940s mark the start of increasing gradients in mortality.

However, this is not completely clear. In the US the only study to directly measure income or wealth gradients in lifespan prior to 1960 is the study by Chapin (1924) who compared the mortality rates of taxpayers and non-taxpayers by matching the 1865 Rhode Island Census with death certificates and documented lower mortality rates for taxpayers. Thus, it is unclear whether the rich always lived longer and if not, when the gaps first emerged (or if instead they have fluctuated over time) in the US.

c. Methodological issues

This literature investigating the link between mortality and income has raised some important methodological issues. First, it is difficult to compare across identical income levels over time because of differences in the cost of living and in standards of living. For this reason, some of the more recent studies compare across income percentiles rather than across income groups (e.g.

Chetty et al., 2016). This approach makes comparisons across groups over time easier but restricts attention to relative changes and cannot measure absolute improvements.

Second, measurement error in income leads to lowers gaps or ratios across income groups.[6] As a result, some have argued that measures of income that average over multiple years provide a more accurate picture of the true differences across groups (e.g. Cristia, 2009; McDonough et al., 1997). Indeed, some argue that the ideal measure of income would approximate lifetime or permanent income measures (e.g. Duggan et al., 2008). This is what more recent work such as Chetty et al. (2016) attempts to do by measuring income during prime adult years (around ages 35-45). A relatedly issue is whether comprehensive measures of income are included – most studies observe income from wages and salaries but often not from other sources. There is also a substantive question of whether income from different sources has a different impact on health and longevity.

Third the literature has found conflicting evidence on whether the effects of income are linear or whether the relationship is instead concave (or flat after a certain threshold). But most studies investigating the relationship at the individual level tend to find some evidence that period life expectancy rises with income more sharply at the bottom (e.g. Backlund et al., 1996; Chetty et al., 2016) consistent with other research showing that other health measures are also more strongly associated with income at the bottom of the income distribution (Lleras-Muney et al., 2025). A key issue here is that respondents with low income are often not observed, as is the case with those that do not pay taxes, or work in subsistence agriculture settings.

Fourth, most studies report income gaps in *period* life expectancy, rather than *cohort* life expectancy. Period life expectancy is the average number of years a person would live if they experienced the prevailing mortality rates throughout their entire life. Cohort life expectancy is computed using the age-specific mortality for each cohort, and so it more accurately reflects the actual average lifespan in a cohort. When mortality rates are trending over time, cohort and period life expectancy can differ substantially. Moreover the two measures can give very different pictures of the gradient as noted for instance by Milligan and Schirle (2021)who document income gradients in life expectancy in Canada.

[6] In general, the quality of the income data in both surveys and administrative data sets has improved over time potentially explaining some of the rise in the gradient (e.g. Cristia 2009).

Finally, a significant number of papers has investigated why income and lifespan are related. This entails investigations trying to assess if there is a causal effect of income on health and mortality, or whether the association is due to reverse causality (the unhealthy becoming poor) or is driven by other factors. A number of studies have attempted to assess if income gradients can be accounted for by differences by other socio-economic markers (such as occupation) or health behaviors, typically finding that income remains an important determinant of mortality and lifespan even after accounting for other covariates (Lantz et al., 1998; McDonough et al., 1997; Sorlie et al., 1995). But studies investigating causal effects have come to more mixed conclusions, with many studies finding negative or zero causal effects of income on health. In trying to explain this surprising result the literature has documented important reasons why, in addition to increasing health, income can be detrimental to health. There is evidence for several mechanisms. Income may be used to purchase alcohol cigarettes illegal drugs and other goods and services (e.g. sex) that can damage health. Income may be obtained by working in risky occupations, so individuals may in fact trade off income and health optimally. Income may be larger in times and places that are unhealthy such as polluted cities or factories. See Lleras-Muney et al. (2025) for a more extensive review.

Table A1. Data Construction Compared with Chetty et al. (2016)

	This Paper	Chetty et al. (2016)
Time period	1900-1950	2001-2014
Sample restrictions	a. matched with lifespan; b. aged 40-63 and being alive one year after each census; c. with positive or zero income measures; d. excluding farmers	a. aged 40-76; b. with positive income
Sample size	**Sample with positive household income (1940-1950) or household occ score (1900-1930)** 5,240,781 (1900), 7,743,943 (1910), 9,272,209 (1920), 12,306,258 (1930), 8,133,097 (1940), 3,167,577 (1950) **Sample with zero household income (1940-1950)** 1,874,000 (1940), 1,674,203 (1950) **U.S. population** 75,824,712 (1900), 92,043,618 (1910), 105,730,985 (1920), 122,777,512 (1930), 131,902,398 (1940), 152,287,077 (1950)	1,408,287,218 person-year observation (100,591,944 per year) U.S. population: 4,255,544,576 person year (303,967,462 per year)
Household income measures	1940-1950: Household income measured at a-2 y for ages 40 y to 63 y; 1900-1950: Household occscore measured at a-2 y for ages 40 y to 63 y; 1900-1950: Household LIDO score measured at a-2 y for ages 40 y to 63 y	Household income measured at a-2 y for ages 40 y to 62 y; measured at age 61 y for ages 63 y to 76 y
Household income measures construction	**Household income** 1940: the sum of wage income for all household members; 1950: the sum of income from all potential sources across household members **Household occscore or LIDO score** 1900-1950: the sum of occscore or LIDO score (conditional on an occupational response)	
Alternative income measures in 1940	Predicted household income measured at a-2 y for ages 40 y to 63 y; Individual wage and salary income measured at a-2 y for ages 40 y to 63 y; Individual occscore measured at a-2 y for ages 40 y to 63 y; Individual LIDO score measured at a-2 y for ages 40 y to 63 y	
Income measures universe in the census	**Individual income** 1940 wage and salary income: Persons age 14+, not institutional inmates 1950 total personal income: Sample-line persons age 14+; all persons age 14+ in Alaska **Individual occupation (000-999)** 1910-1930: All persons. 1940-1950: Persons age 14+ and in the labor force; not institutional inmates; not new workers Codes 000-970 represents occupational response Codes 970+ includes Not-yet classified, Non-occupational response, and Occupation missing/unknown **Individual Occscore (00 = N/A)** 1900-1950: All persons with an occupational response (codes 000-970) recorded in OCC1950 **Individual LIDO score** 1900-1950: All persons with census region and division, state, sex, age, race, occupation, and industry	
Gompertz approximations	Based on 1-y mortality rate for ages 40 y to 63 y	Based on 1-y mortality rate for ages 40 y to 76 y
Mortality rates	Gompertz extrapolation for ages 63 y to 90 y SSA mortality rates in each year for ages > 90 y	Gompertz extrapolation for ages 77 y to 90 y NCHS and SSA mortality rates for ages > 90 y
Period life expectancy calculation	Based on Gompertz parameters and SSA Weighted LE	Based on Gompertz parameters, NCHS, and SSA Race- and ethnicity-adjusted LE

Notes: LIDO refers to the Lasso-adjusted industry, demographic, and occupation score, measured on a log-income scale; in our analysis, we use its exponential transformation. Predicted household income in 1940 is obtained by regressing household income for the 1895–1910 birth cohorts on a set of covariates, including age, age squared, sex, marital status, race, Hispanic origin, homeownership, schooling, and occupational score. For all analyses based on 1940 data, the sample is restricted to individuals aged 40–62 (rather than 40–63).

Table A2. Summary Statistics of Cohorts aged 40 to 63 in the Year after Each Census

	1900		1910		1920		1930		1940		1950	
	Census	Working sample	Census	Working sample	Census	Working sample	Census	Working sample	Census	Working sample	Census	Working sample
With data on lifespan	0.456		0.480		0.476		0.483		0.323		0.215	
SSA Period LE at 40 for men		67.520		67.640		69.200		68.820		69.420		70.870
SSA Period LE at 40 for women		68.690		69.570		70.060		71.010		72.540		75.180
Period LE at 40 for men		70.982		69.750		69.813		69.146		69.469		70.894
Period LE at 40 for women		74.311		73.449		72.484		73.251		75.059		77.368
Cohort LE at 40 for men		70.034		69.588		69.732		70.567		72.333		72.866
Cohort LE at 40 for women		72.333		72.227		73.356		75.799		79.245		80.149
Panel A. Sociodemographics												
Age	48.295	48.324	48.394	48.462	48.581	48.651	48.663	48.708	48.812	49.012	49.354	49.219
Female	0.561	0.583	0.540	0.558	0.533	0.548	0.530	0.539	0.528	0.535	0.526	0.530
White	0.910	0.950	0.916	0.943	0.919	0.946	0.920	0.946	0.918	0.970	0.909	0.964
Non-Hispanic	0.994	0.995	0.994	0.995	0.992	0.993	0.990	0.991	0.988	0.989	.	.
Marital status												
Single	0.113	0.092	0.116	0.099	0.119	0.105	0.110	0.101	0.105	0.085	0.088	0.069
Married	0.755	0.788	0.758	0.788	0.761	0.785	0.771	0.790	0.783	0.819	0.777	0.835
Separated/divorced/widowed	0.133	0.119	0.126	0.114	0.120	0.111	0.119	0.109	0.113	0.096	0.135	0.096
Urban	0.497	0.442	0.576	0.531	0.630	0.588	0.662	0.628	0.649	0.584	0.657	0.563
Years of schooling	.	.	.	.	.	.	.	.	8.152	8.487	8.197	8.461
Family size	4.451	4.694	4.248	4.456	4.048	4.218	3.902	4.030	3.692	3.894	.	.
Employment status												
Not in labor force	.	.	0.493	0.529	.	.	0.446	0.464	0.460	0.474	0.402	0.428
Unemployed	.	.	0.027	0.023	.	.	0.057	0.050	0.044	0.036	0.023	0.017
Employed	.	.	0.479	0.448	.	.	0.497	0.487	0.497	0.490	0.575	0.554
Panel B. Income measures												
Household income ($1,000s)	.	.	.	.	.	.	.	.	1.577	1.636	1.112	1.084
Household occ score	53.537	48.162	60.921	54.171	57.819	52.802	52.358	49.613	48.514	48.089	46.617	50.255
Household LIDO score	38.431	34.952	45.023	39.161	44.472	39.980	41.683	39.277	39.019	38.688	36.059	37.445
Predicted household income ($1,000s)	.	.	.	.	.	.	.	.	1.379	1.467	.	.
Wage income ($1,000s)	.	.	.	.	.	.	.	.	0.610	0.653	0.754	0.762
Occ score	24.396	26.420	24.417	26.356	25.278	26.871	25.787	27.088	25.864	27.247	26.533	27.366
LIDO score	21.492	23.333	22.054	23.693	23.089	24.431	23.655	24.837	23.279	24.635	23.908	24.752
Wage income of spouse ($1,000s)	.	.	.	.	.	.	.	.	0.551	0.574	.	.
Occ score of spouse	.	.	.	.	.	.	.	.	25.288	25.425	.	.
Number of observations	12,291,602	5,604,426	16,148,482	7,743,943	20,623,024	9,808,632	26,656,666	12,883,017	31,888,095	10,253,005	39,200,251	4,841,780

Notes: The summary statistics for the working sample are weighted. The working sample consists of linked individuals aged 40 to 63 (62) and being live in the year after each census (1940), excluding farmers.

Table A3. Slope Estimates of LE on Income

	Period LE for 40-year-olds				Cohort LE for 40-year-olds			
	Men		Women		Men		Women	
	Slope	SE	Slope	SE	Slope	SE	Slope	SE
Panel A. 1940 gradient by percentiles of income measures								
Household income	0.003	0.004	-0.007	0.006	0.019***	0.002	0.026***	0.001
Predicted household income	-0.017***	0.004	0.019	0.014	0.026***	0.002	0.038***	0.003
Wage income	-0.007	0.004	0.001	0.015	0.034***	0.001	0.027***	0.003
Occ score	-0.018***	0.004	0.001	0.014	0.016***	0.002	0.025***	0.003
LIDO score	-0.025***	0.004	0.015	0.016	0.017***	0.002	0.029***	0.004
Wage income of spouse	-0.017	0.011	0.001	0.008	0.008*	0.004	0.031***	0.002
Panel B. 1940 gradient by levels of income measures including top 5%								
Household income ($1,000s)	-0.003	0.040	0.072	0.092	0.010	0.026	0.292***	0.035
Predicted household income ($1,000s)	-0.654***	0.178	1.199	0.698	0.939***	0.102	1.844***	0.164
Wage income ($1,000s)	-0.310*	0.126	0.382	0.655	0.944***	0.047	0.953***	0.137
Occ score	-0.052***	0.011	-0.012	0.047	0.043***	0.006	0.080***	0.012
LIDO score	-0.087***	0.014	0.065	0.081	0.053***	0.006	0.153***	0.019
Wage income of spouse ($1,000s)	-0.481	0.518	0.266	0.205	0.567***	0.166	0.770***	0.056
Panel C. 1940 gradient by levels of income measures excluding top 5%								
Household income ($1,000s)	0.066	0.121	-0.260	0.163	0.554***	0.048	0.651***	0.047
Predicted household income ($1,000s)	-0.847**	0.314	1.262	0.909	1.523***	0.143	2.339***	0.199
Wage income ($1,000s)	-0.156	0.173	-0.944	0.877	1.192***	0.050	1.576***	0.161
Occ score	-0.052***	0.013	-0.005	0.058	0.043***	0.007	0.126***	0.013
LIDO score	-0.087***	0.017	0.045	0.094	0.053***	0.007	0.191***	0.021
Wage income of spouse ($1,000s)	-1.247	0.723	-0.143	0.274	0.496*	0.233	1.015***	0.068
Panel D. Income gradient in different years by percentiles of income measures								
Household occ score								
1900 gradient	-0.025***	0.006	-0.003	0.005	-0.020***	0.002	-0.011***	0.001
1910 gradient	-0.016***	0.004	-0.012*	0.005	-0.017***	0.002	-0.012***	0.002
1920 gradient	-0.007	0.004	0.002	0.004	-0.018***	0.002	-0.012***	0.002
1930 gradient	-0.009**	0.003	0.001	0.005	-0.013***	0.002	-0.006*	0.003
1940 gradient	0.003	0.004	-0.007	0.006	-0.005*	0.002	0.005	0.003
1950 gradient	0.004	0.006	-0.010	0.010	-0.007*	0.003	-0.003	0.003
Household LIDO score								
1900 gradient	-0.026***	0.007	-0.005	0.005	-0.015***	0.003	0.002	0.002
1910 gradient	-0.014**	0.005	-0.022***	0.005	-0.011***	0.002	0.002	0.002
1920 gradient	-0.010*	0.004	-0.002	0.004	-0.013***	0.002	0.002	0.002
1930 gradient	-0.005	0.003	-0.002	0.004	-0.005*	0.002	0.008***	0.002
1940 gradient	-0.001	0.003	-0.016**	0.006	-0.000	0.002	0.014***	0.002
1950 gradient	0.007	0.005	-0.018	0.009	0.007*	0.003	0.004	0.002
Household income								
1940 gradient	0.003	0.004	-0.007	0.006	0.019***	0.002	0.026***	0.001
1950 gradient	0.011*	0.005	0.006	0.010	0.018***	0.003	0.005*	0.002
2010s gradient	0.114***	0.002	0.072***	0.001				
Panel E. Income gradient in different years by levels of income measures including top 5%								
Household occ score								
1900 gradient	-0.010***	0.003	-0.005	0.004	-0.007***	0.001	-0.008***	0.001
1910 gradient	-0.003	0.002	-0.000	0.004	-0.005***	0.001	-0.006***	0.001
1920 gradient	-0.004**	0.001	0.001	0.003	-0.005***	0.001	-0.005**	0.002
1930 gradient	-0.005***	0.001	0.003	0.003	-0.005***	0.001	-0.001	0.002
1940 gradient	-0.000	0.002	0.005	0.004	-0.005***	0.001	0.004*	0.002
1950 gradient	-0.008*	0.004	0.000	0.011	-0.009***	0.002	-0.005	0.003
Household LIDO score								
1900 gradient	-0.012**	0.004	-0.007	0.006	-0.007***	0.001	-0.001	0.002
1910 gradient	-0.002	0.002	-0.005	0.005	-0.006***	0.001	-0.001	0.002
1920 gradient	-0.005**	0.002	-0.002	0.004	-0.004***	0.001	0.001	0.002
1930 gradient	-0.006***	0.002	0.002	0.004	-0.005***	0.001	0.005*	0.002
1940 gradient	-0.002	0.002	-0.006	0.006	-0.005***	0.001	0.010***	0.002
1950 gradient	-0.009	0.005	-0.007	0.016	-0.008**	0.003	-0.001	0.004
Household income ($1,000s)								
1940 gradient	-0.003	0.040	0.072	0.092	0.010	0.026	0.292***	0.035
1950 gradient	0.079	0.108	0.121	0.209	0.269***	0.056	0.096*	0.042

Panel F. Income gradient in different years by levels of income measures excluding top 5%								
Household occ score								
1900 gradient	-0.028***	0.008	-0.008	0.008	-0.024***	0.002	-0.016***	0.002
1910 gradient	-0.022***	0.005	-0.013*	0.007	-0.019***	0.002	-0.016***	0.003
1920 gradient	-0.007	0.005	0.002	0.005	-0.024***	0.002	-0.017***	0.004
1930 gradient	-0.011*	0.005	0.002	0.007	-0.019***	0.002	-0.007	0.004
1940 gradient	0.004	0.007	-0.010	0.010	-0.006	0.004	0.006	0.005
1950 gradient	0.008	0.009	-0.029	0.015	-0.009	0.005	-0.005	0.004
Household LIDO score								
1900 gradient	-0.025*	0.012	-0.010	0.012	-0.022***	0.004	0.008*	0.004
1910 gradient	-0.025**	0.008	-0.044***	0.011	-0.014***	0.003	0.006	0.005
1920 gradient	-0.010	0.007	-0.003	0.009	-0.021***	0.003	0.006	0.005
1930 gradient	-0.006	0.007	-0.007	0.009	-0.008*	0.004	0.018***	0.005
1940 gradient	0.004	0.008	-0.031*	0.013	0.003	0.004	0.032***	0.004
1950 gradient	0.022*	0.011	-0.051*	0.022	0.016*	0.006	0.014*	0.006
Household income ($1,000s)								
1940 gradient	0.066	0.121	-0.260	0.163	0.554***	0.048	0.651***	0.047
1950 gradient	0.394**	0.144	0.195	0.272	0.522***	0.065	0.157**	0.056

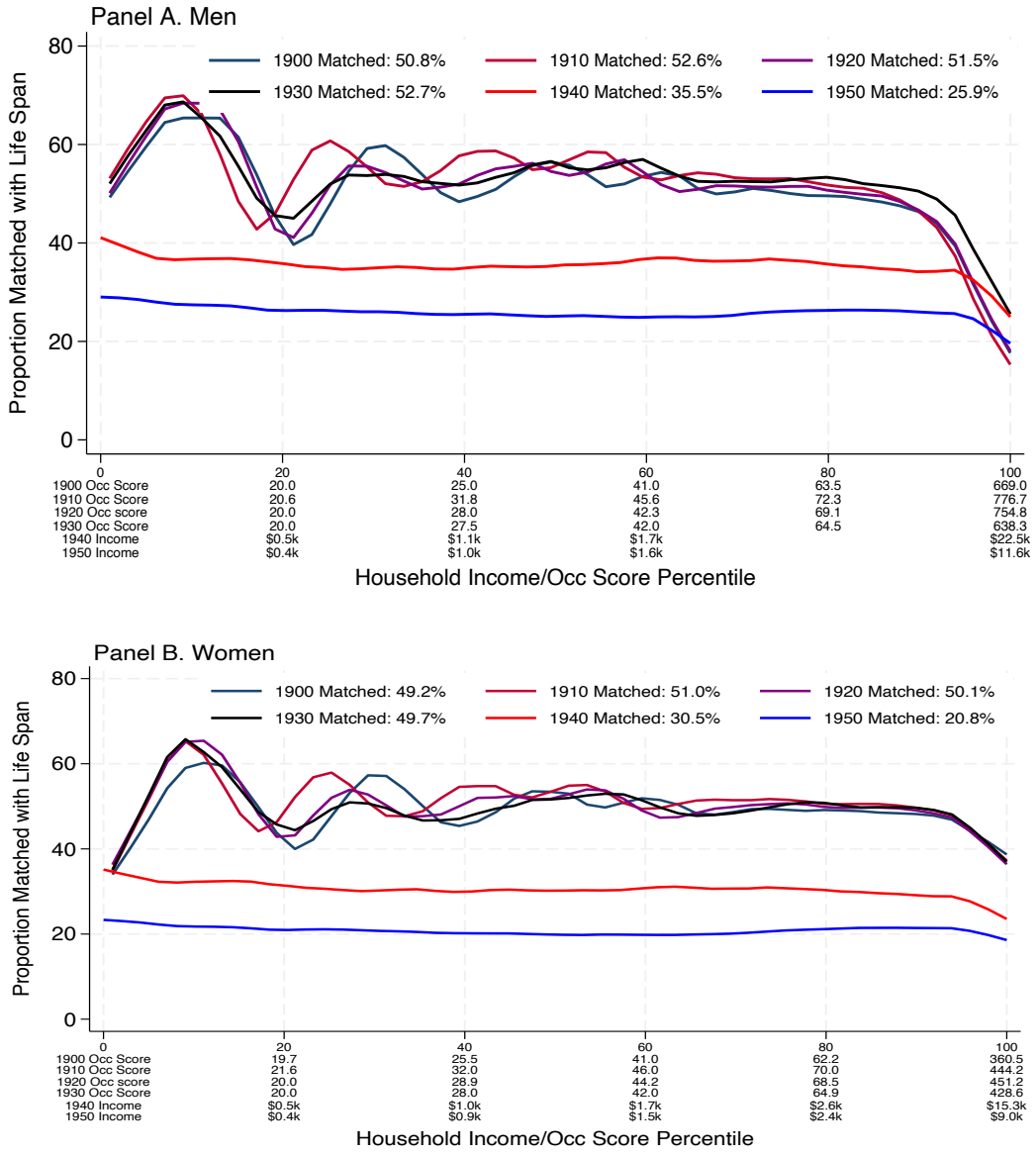
Notes: * p<0.05, ** p<0.01, *** p<0.001

Table A4. OLS Estimates of Cohort LE for 40-Year-Olds

	1901 Cohort			
	(1)	(2)	(3)	(4)
Panel A. Men				
Household income percentiles	0.0123***	0.0173***	0.0164***	0.0076***
	(0.0010)	(0.0011)	(0.0012)	(0.0012)
Household occ score percentiles		-0.0134***		
		(0.0012)		
Household LIDO score percentiles			-0.0092***	
			(0.0013)	
Occupation FE	No	No	No	Yes
Number of observations	240742	240742	240742	240742
Panel B. Women				
Household income percentiles	0.0119***	0.0129***	0.0070***	0.0086***
	(0.0010)	(0.0011)	(0.0012)	(0.0010)
Household occ score percentiles		-0.0022		
		(0.0013)		
Household LIDO score percentiles			0.0099***	
			(0.0014)	
Occupation FE	No	No	No	Yes
Number of observations	253104	253104	253104	253076

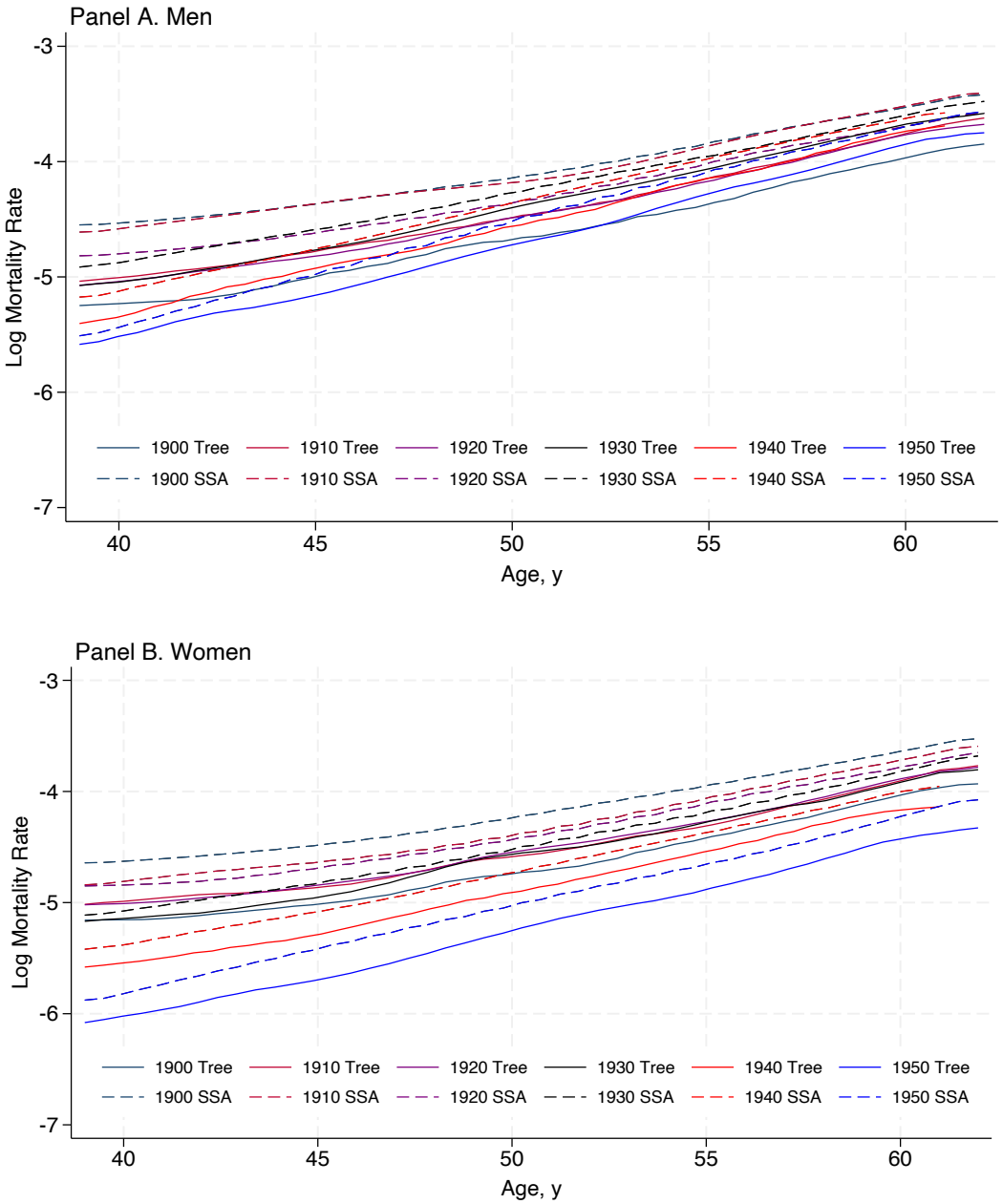
Notes: The sample consists of linked individuals aged 40 and being live in 1941, excluding farmers.
* p<0.05, ** p<0.01, *** p<0.001

Figure A1. Proportion Matched with Lifespan by Household Income/Occupation Score Percentile in Each Census Year



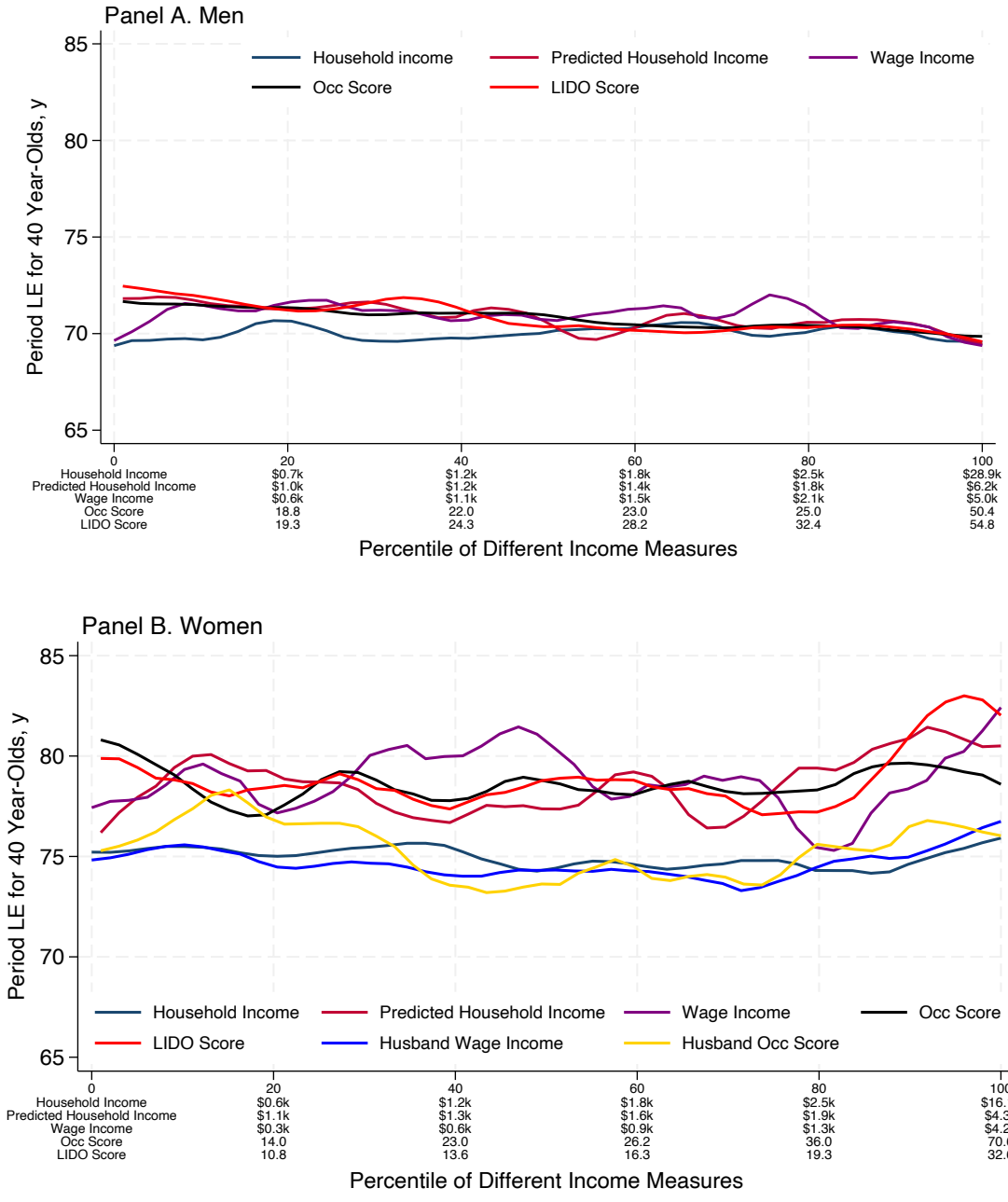
Notes: The sample consists of individuals aged 40 to 63 (62) in the year after each census (1940). Percentiles are defined within the positive income distribution. For 1940 and 1950, an additional data point for zero income (x-axis value = 0) is included. The nonparametric line estimates are based on regressions that incorporate this zero-income observation.

Figure A2. National Mortality Profile from Census Tree and SSA Life Table in Each Census Year



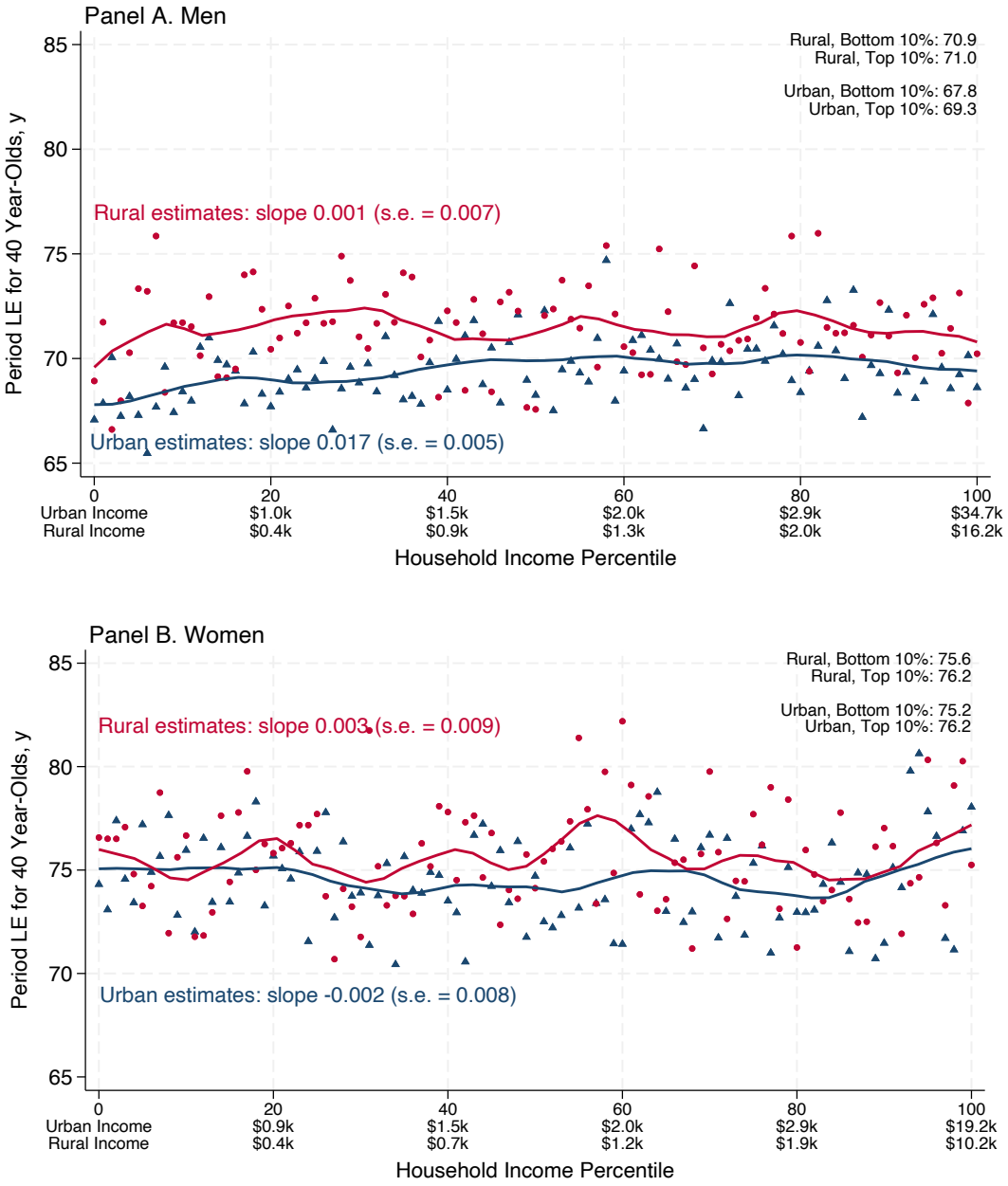
Notes: The sample used for calculating mortality from the census tree is restricted to linked individuals aged 39-62 (61) and being alive each census year (1940).

Figure A3. Period Life Expectancy for 40-Year-Olds in 1940 by Percentile of Different Income Measures



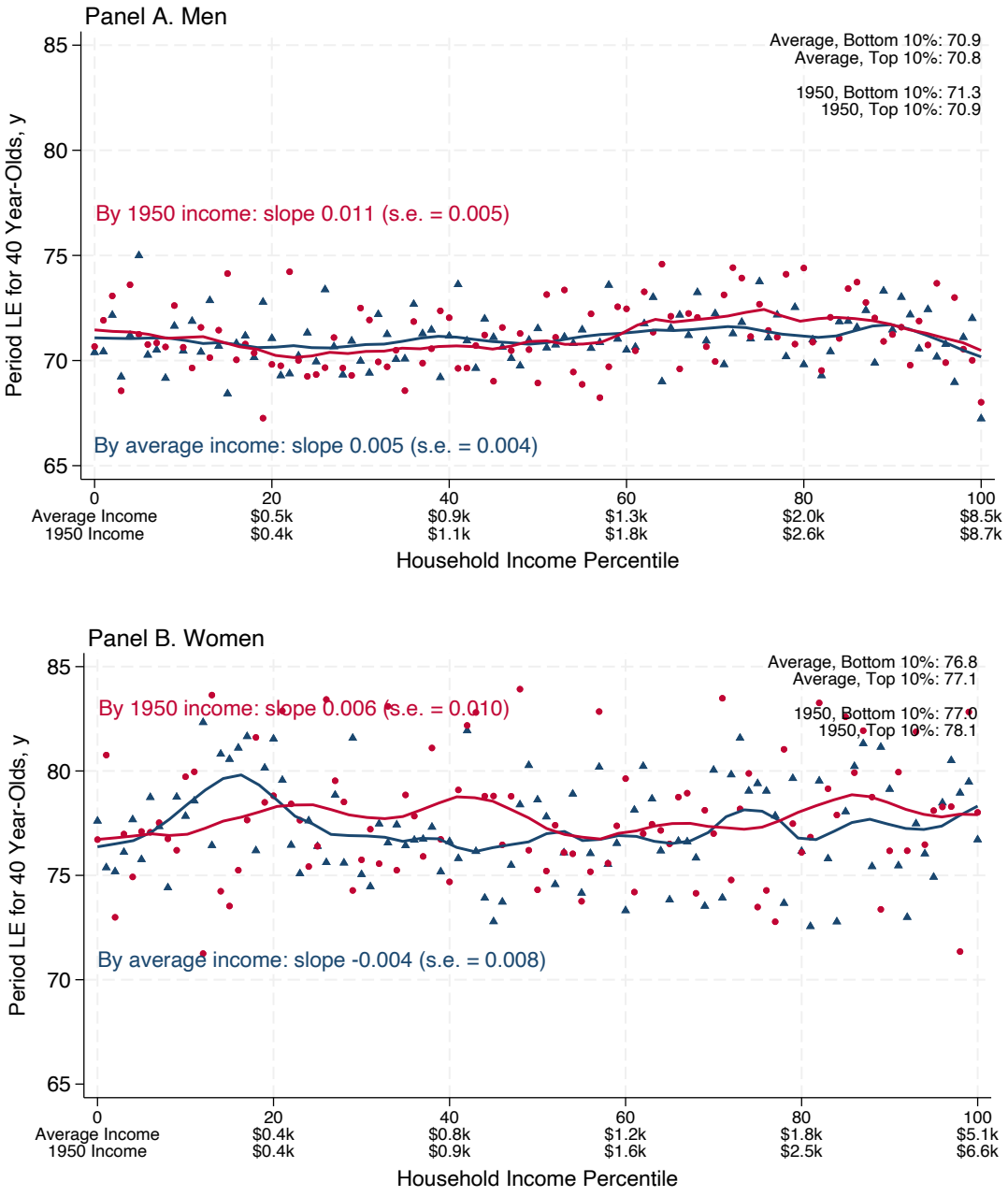
Notes: Percentiles of household income, wage income, and husband's wage income are defined within the positive income distribution, and additional data points for zero income (x-axis value = 0) are included. For these measures, the nonparametric line estimates are based on regressions that incorporate these zero-income observations.

Figure A4. Period Life Expectancy for 40-Year-Olds in 1940 by Household Income Percentile and Residence



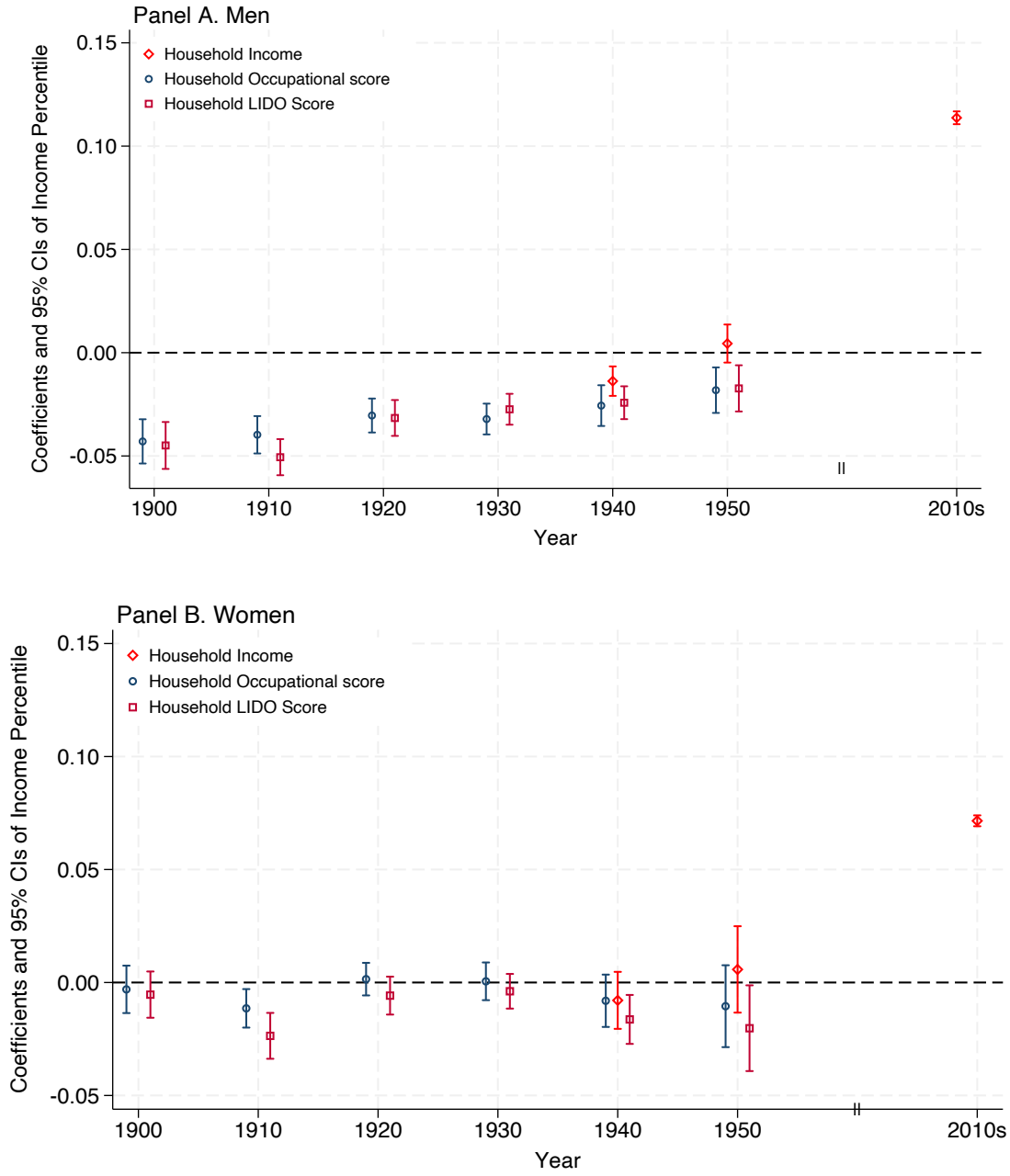
Notes: In our working sample, 58.4% lived in urban area in 1940, which is made up for the most part of households in cities and incorporated places with 2,500+ inhabitants. Percentiles are defined within the positive income distribution. Additional data points for zero income (x-axis value = 0) are included. The slope estimates are based on regressions that incorporate these zero-income observations. The reported mean life expectancy in the text reflects both the bottom 10% and the zero-income group.

Figure A5. Period Life Expectancy for 40-Year-Olds in 1950 by Household Income Percentile



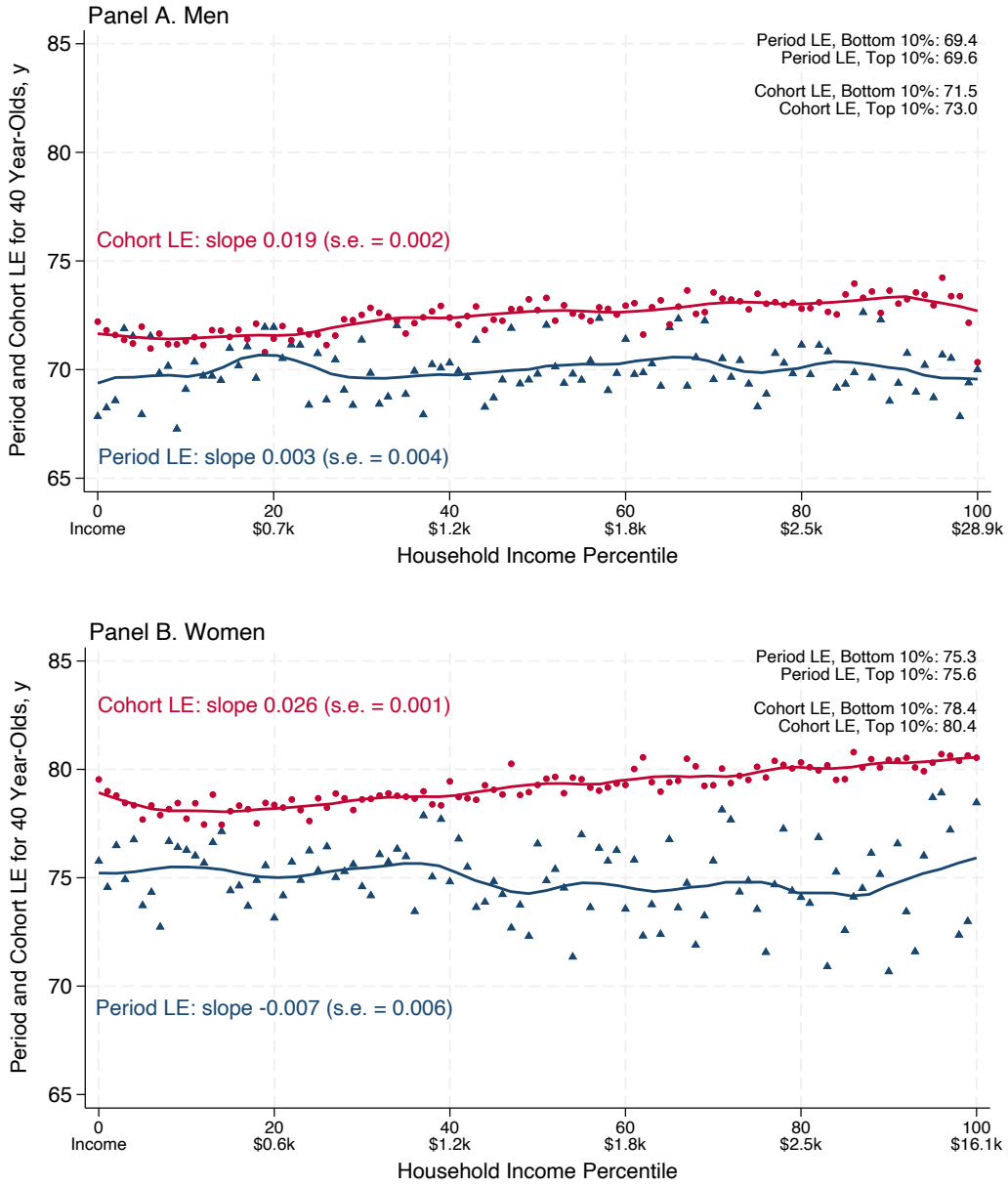
Notes: Percentiles are defined within the positive income distribution. Additional data points for zero income (x-axis value = 0) are included. The slope estimates are based on regressions that incorporate these zero-income observations. The reported mean life expectancy in the text reflects both the bottom 10% and the zero-income group. The average income refers to the average income in 1939 and 1949 recorded in the census. Income values are adjusted to 1939 prices.

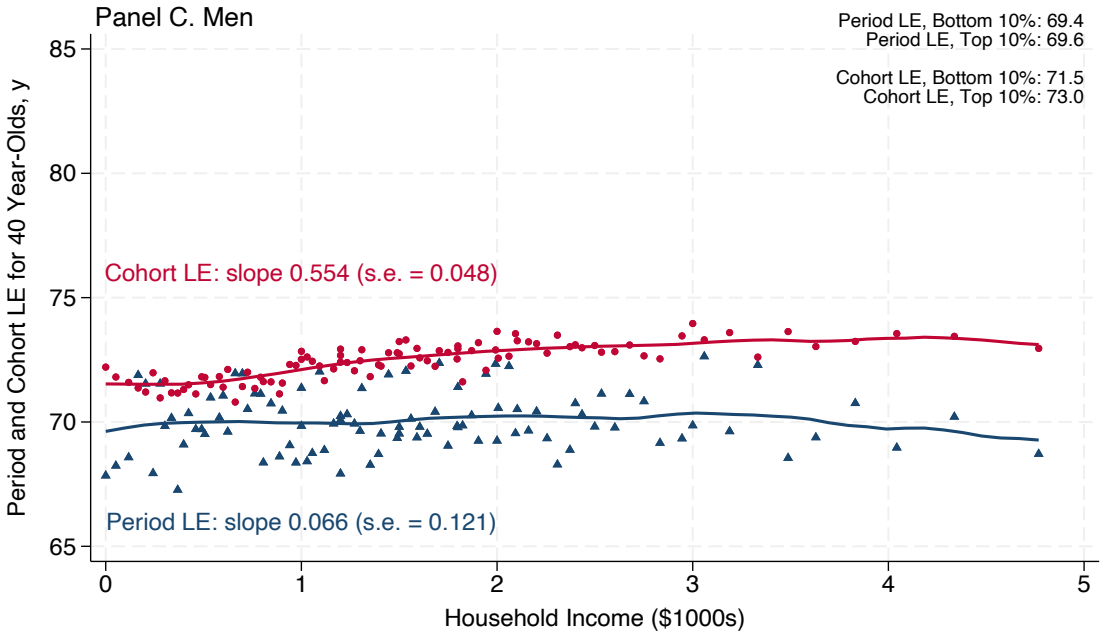
Figure A6. Period Life Expectancy Income Gradient for 40-Year-Olds since 1900 (Including Farmers)

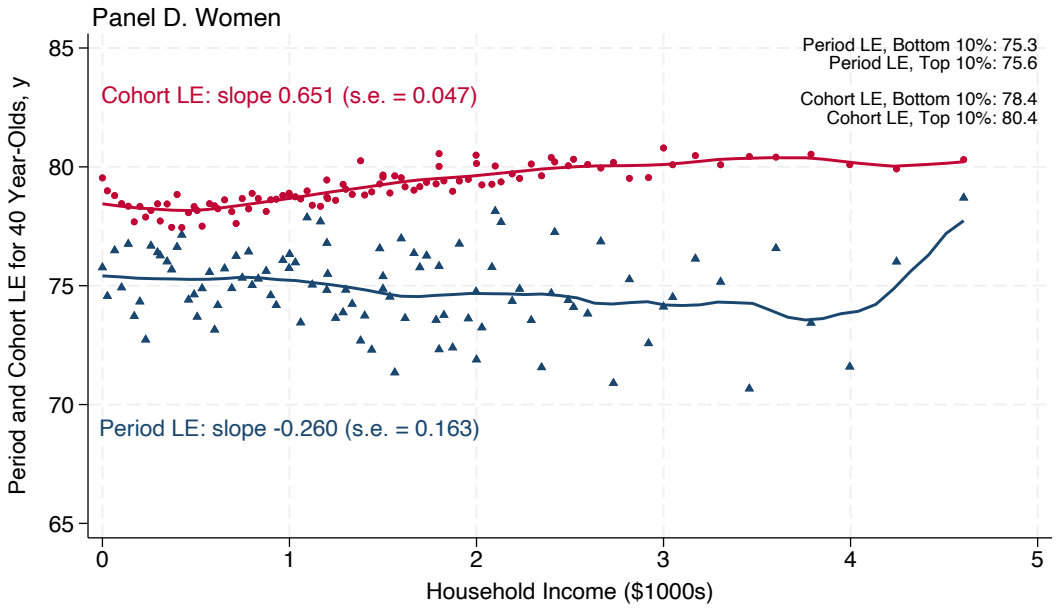


Notes: For each income measure and year, period life expectancy is estimated by percentiles of the corresponding income measure for each sex. Coefficients and 95% confidence intervals are obtained from separate regressions of estimated period life expectancy on percentiles of the corresponding income measure. Weights are applied during the analysis.

Figure A7. Period and Cohort Life Expectancy for 40-Year-Olds in 1940 by Household Income Percentile

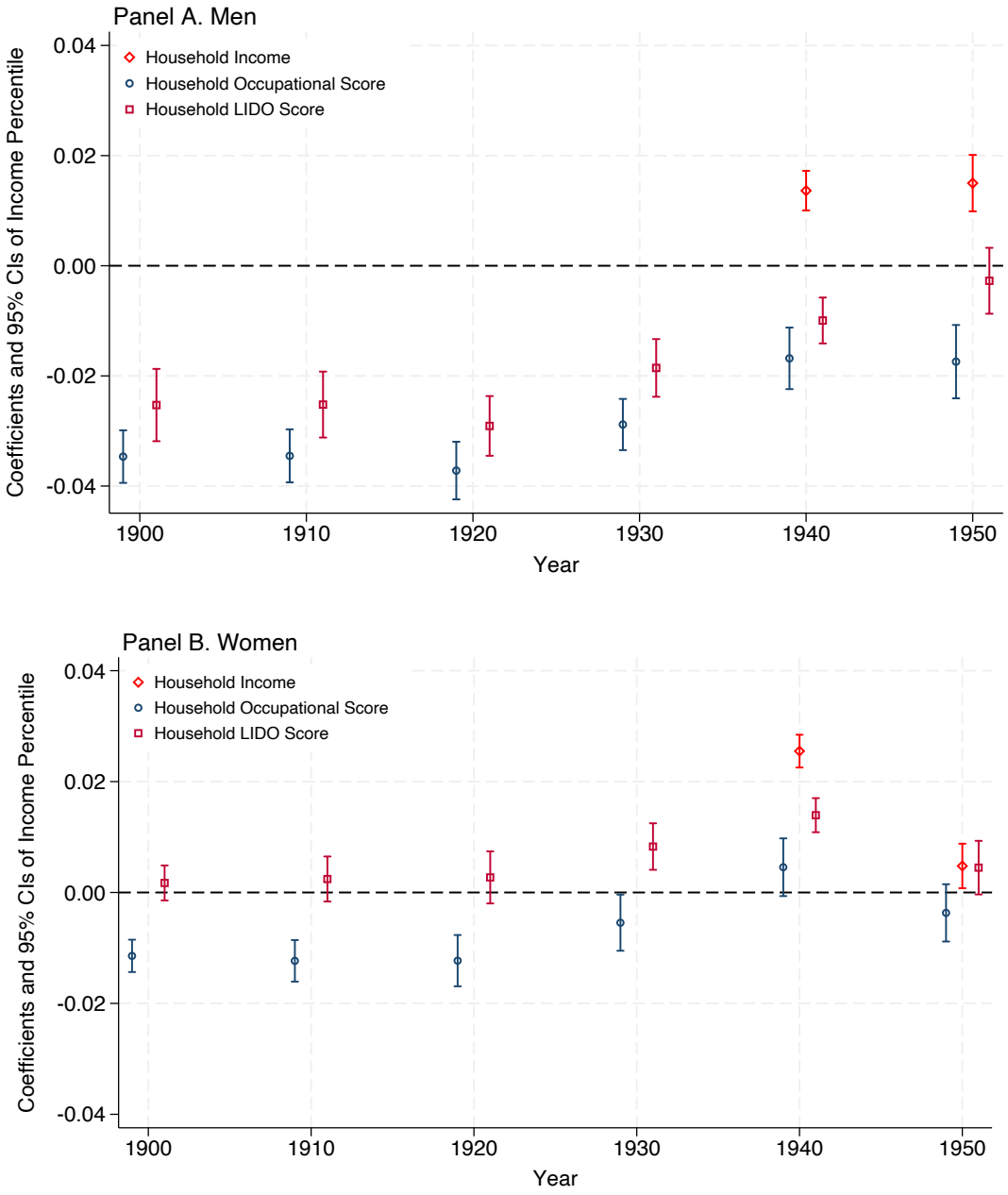






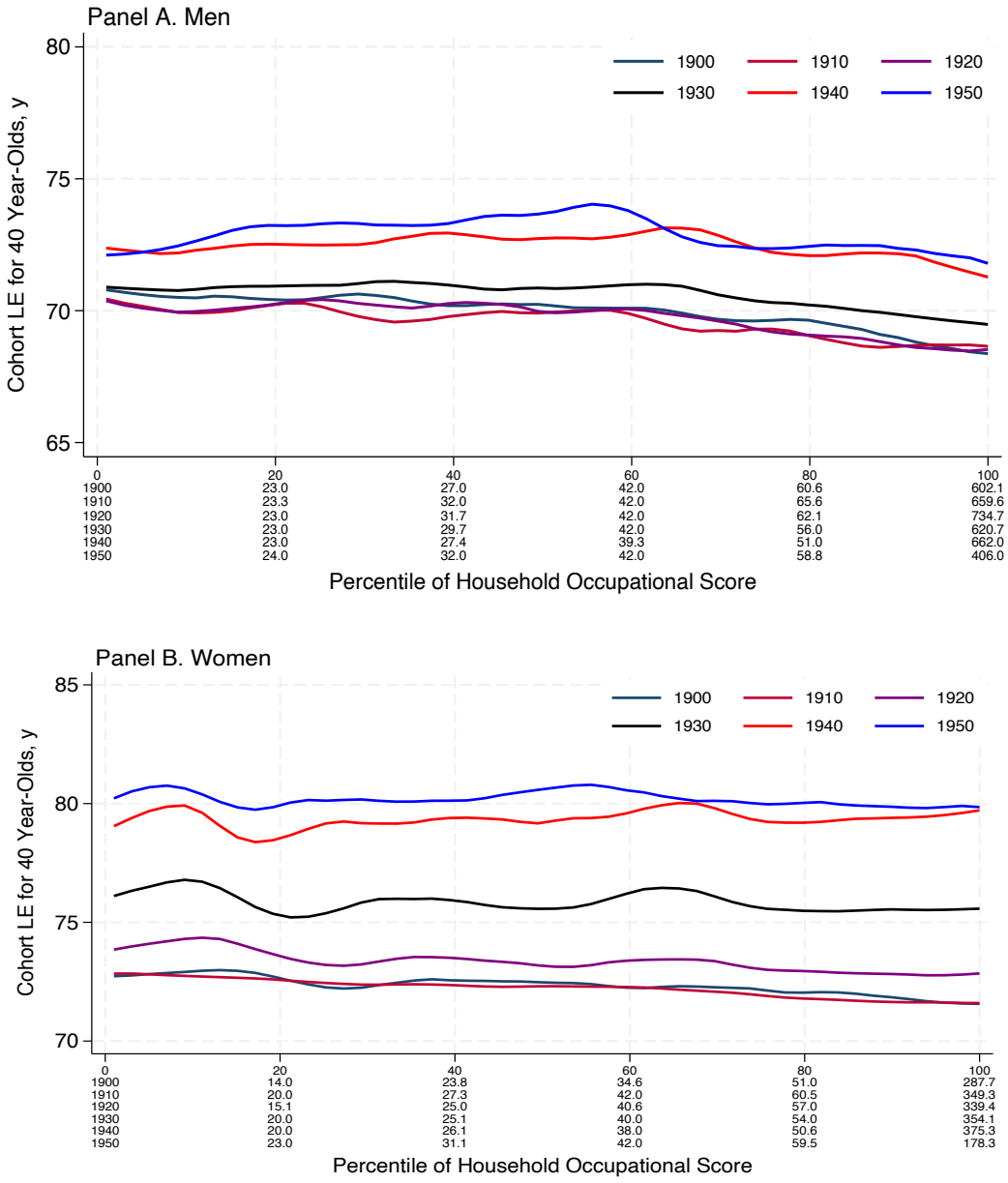
Notes: Percentiles are defined within the positive income distribution. Additional data points for zero income (x-axis value = 0) are included. The nonparametric line and OLS slope estimates are based on regressions that incorporate these zero-income observations. The reported mean life expectancy in the text reflects both the bottom 10% and the zero-income group. For Panel C and D, each data point represents household income at a given percentile of the positive income distribution.

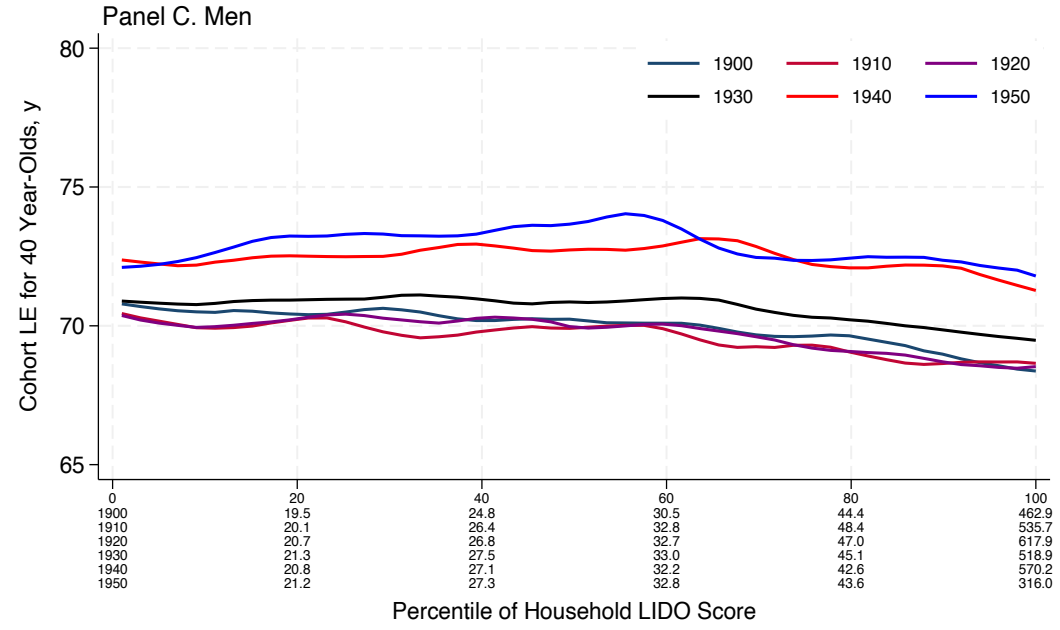
Figure A8. Cohort Life Expectancy Income Gradient for 40-Year-Olds since 1900 (Including Farmers)

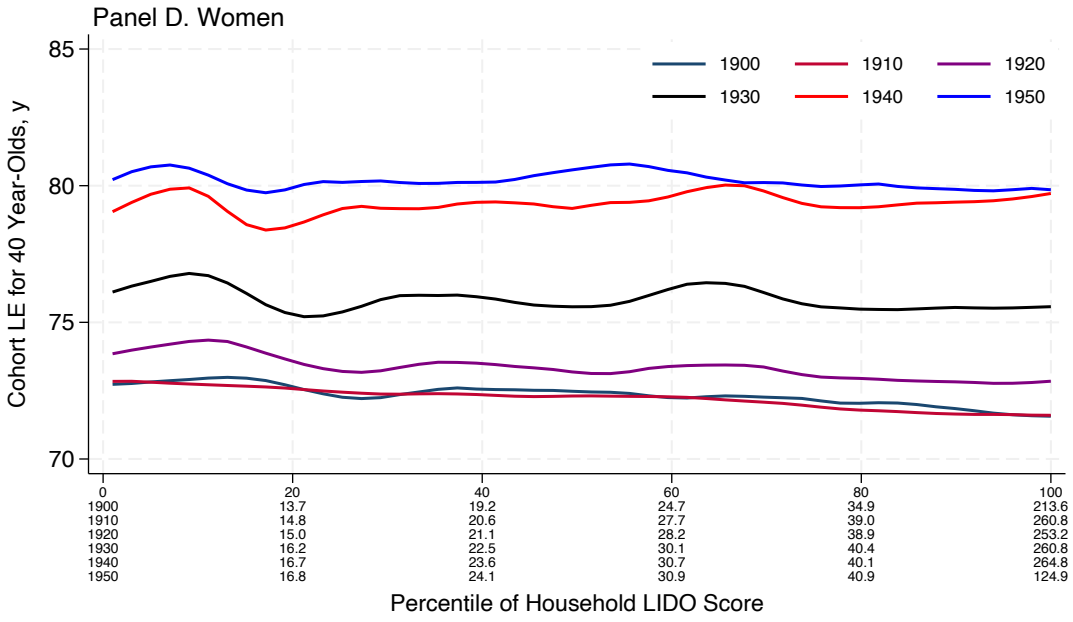


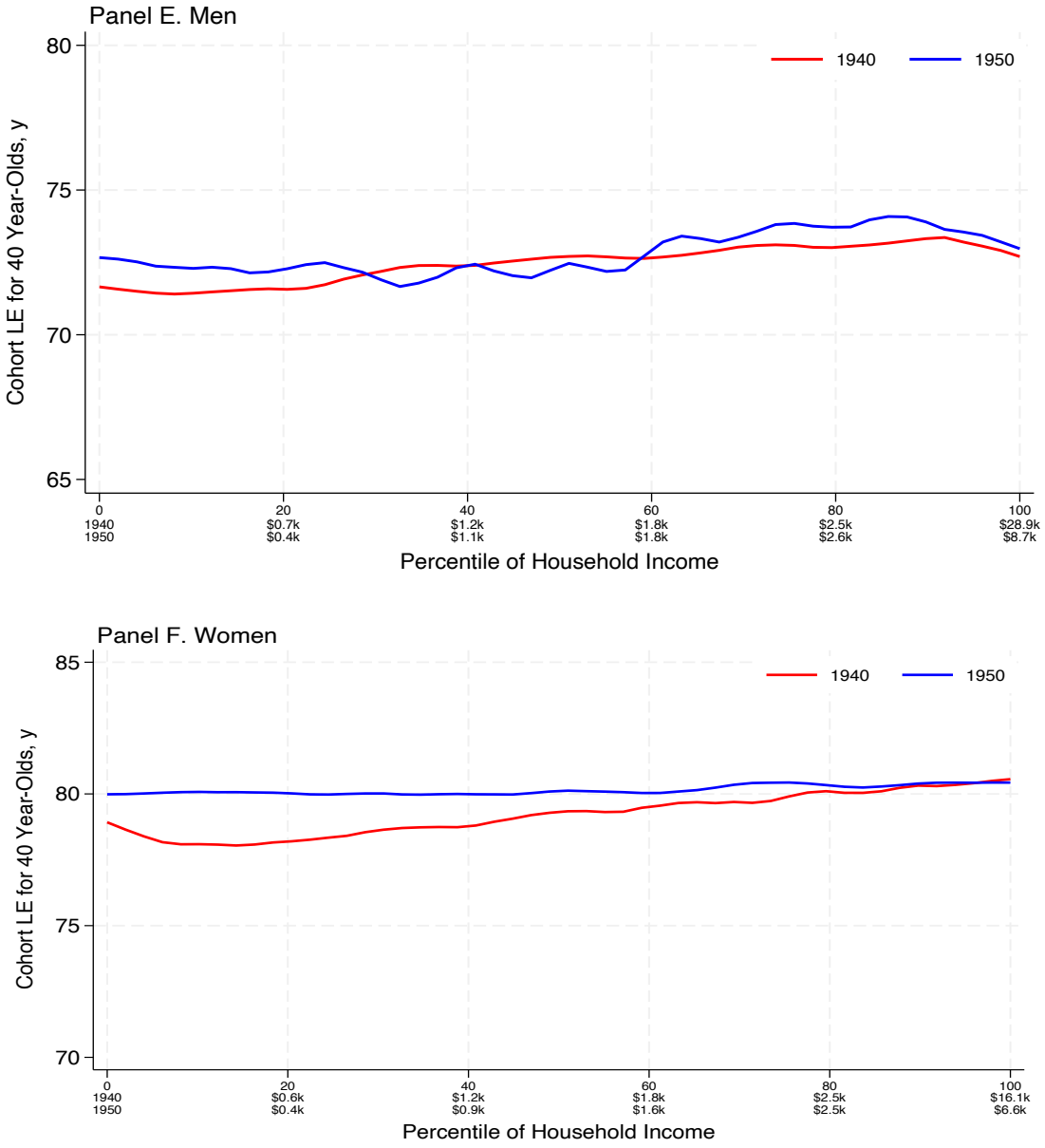
Notes: For each income measure and year in Panel A and B, cohort life expectancy are estimated by percentiles of the corresponding income measure for each sex. Coefficients and 95% confidence intervals are obtained from separate regressions of estimated cohort life expectancy on percentiles of the corresponding income measure. Weights are applied during the analysis.

Figure A9. Cohort Life Expectancy by Household Income/Occscore Percentile for Men (Left) and Women (Right) by Year




Panel C. Men
Cohort LE for 40 Year-Olds, y
80
75
70
65
1900
1910
1920
1930
1940
1950
0 20 40 60 80 100
1900 19.5 24.8 30.5 44.4 462.9
1910 20.1 26.4 32.8 48.4 535.7
1920 20.7 26.8 32.7 47.0 617.9
1930 21.3 27.5 33.0 45.1 518.9
1940 20.8 27.1 32.2 42.6 570.2
1950 21.2 27.3 32.8 43.6 316.0
Percentile of Household LIDO Score


Panel D. Women
Cohort LE for 40 Year-Olds, y
85
80
75
70
1900
1910
1920
1930
1940
1950
0 20 40 60 80 100
1900 13.7 19.2 24.7 34.9 213.6
1910 14.8 20.6 27.7 39.0 260.8
1920 15.0 21.1 28.2 38.9 253.2
1930 16.2 22.5 30.1 40.4 260.8
1940 16.7 23.6 30.7 40.1 264.8
1950 16.8 24.1 30.9 40.9 124.9
Percentile of Household LIDO Score



Notes: For Panel E and F, percentiles are defined within the positive income distribution. Additional data points for zero income (x-axis value = 0) are included. The nonparametric line and OLS slope estimates are based on regressions that incorporate these zero-income observations.