Income and Adult Life Expectancy in the US, 1900-1950

Authors: Carson Zhang, MA[1]; Andrew Goodman-Bacon, PhD[2]; Joseph Price, PhD[3]; Dahai Yue, PhD[4,5]; Adriana Lleras-Muney, PhD[1]

Affiliations:

[1]Department of Economics, University of California at Los Angeles; Los Angeles 90095, USA.

[2]Opportunity and Inclusive Growth Institute, Federal Reserve Bank of Minneapolis; Minneapolis, 55401, USA.

[3]Department of Economics, Brigham Young University; Provo, UT 84602, USA.

[4]Department of Health Policy and Management, University of Maryland; College Park, 20742, USA.

[5]Institute for Health Computing, University of Maryland; North Bethesda, 20852, USA.

Corresponding Author: Adriana Lleras-Muney, PhD, Department of Economics, University of California, Los Angeles, Bunche Hall, 315 Portola Plaza, Los Angeles, CA 90095 (alleras@econ.ucla.edu; telephone (310) 825-3925).

Key Points

Question Have the lifespans of rich and poor Americans always differed as much as they do today?

Findings Using records of 56 million Americans, we find no differences in adult period life expectancy by income in 1940 or 1950, and no positive association between income proxies and adult life expectancy for men or women between 1900 and 1950, despite high levels of income inequality during this period.

Meaning Income-related disparities in period life expectancy are not a fixed feature of unequal societies. In the US they emerged around the 1950s, a period of low income inequality but rapid medical innovation.

Abstract

Importance Today, at age 40, the rich can expect to live almost a decade longer than the poor.

Objective To test whether the rich have always outlived the poor in the U.S., and if not, when and why the current gradient emerged.

Design, Setting, and Participants We construct a representative dataset of about 56 million Americans, covering about 50% of the population in 1900-1940 and 35% in 1950, that links individual-level income from decennial censuses to death records derived from a genealogical website. We estimate associations between period (and cohort) life expectancy and income percentiles.

Exposures Household income, measured directly in the 1940 and 1950 censuses and, for 1900 through 1930, predicted from occupation and other individual and household characteristics.

Main Outcomes and Measures Period life expectancy at age 40 (the predicted age at death based on current mortality rates) across all income percentiles between 1900 and 1950; cohort life expectancy at age 40 (the realized age at death); estimated life expectancy gaps between the top and the bottom income deciles.

Results We find no differences in adult period life expectancy by income in 1940 or 1950. We can rule out gaps larger than a year in 1940, one-tenth as large as the gaps in the 2000s. The associations between income proxies and adult period life expectancy for men or women between 1900 and 1930 were even smaller or negative, despite high levels of income inequality during this period. Two factors help explain this: the rich lived in cities where life expectancy was lower and were more likely to consume goods associated with chronic diseases.

Conclusions and Relevance Income-related disparities in adult life expectancy are not a fixed feature of unequal societies. In the US these disparities emerged around the 1950s, as medicine

became more effective, expensive and distributed unequally across the market. How medical advances are made available to the population may therefore be a key determinant of lifespan inequality.

Introduction

Between 2001 and 2014, the richest 40-year-old Americans could expect to live a decade longer or more than the poorest.[1] High-quality Population Register data from other high-income countries show similar associations.[2-6] Income proxies are also associated with higher lifespans around the world.[7] That this gradient persists across time and space motivates theories in which socioeconomic factors are the central determinants of health[8-11] (eAppendix Part II.A in the Supplement). Because income inequality in the U.S. in the early 20th century was as large as it is today,[12] these theories predict that the rich should also have outlived the poor by a wide margin then. We test whether they did.

Estimating the income gradient in mortality over time is difficult because individual-level data on income and mortality are scarce. The earliest nationally representative U.S. estimates, based on the 1960 Census, find a moderate negative association between income and mortality among adults.[13] Later work suggests this association became stronger thereafter.[14-23] Pre-1960 U.S. evidence is neither nationally representative nor based on individual income: most studies compare mortality across selected geographic areas and are therefore subject to the ecological fallacy.[24] These studies find lower mortality among higher socioeconomic groups. In contrast, famous or very wealthy individuals in various countries did not outlive the average person until the mid-20th century.[25-27] Nationally representative data from Sweden similarly show that men in high-paying occupations only gained a longevity advantage after 1950.[28] Part II of the eAppendix in the Supplement reviews this literature and the methodological issues it raises.

In this study, we construct new representative data for the U.S. over the first half of the 20th century and investigate whether the rich have always outlived the poor in the U.S., and if not, when and why the current gradient emerged.

Methods

Data sources and measurement

We link individual census data from 1900-1950 to profiles on Family Tree from the FamilySearch genealogical website, which report lifespans (eTable 1 in the Supplement). Our target population is census respondents aged 40-63 in the year after the census was taken. The oldest cohorts studied were born between 1838 and 1861 and the youngest between 1888 and 1911.

We observe household income directly in the 1940 census for the entire population and in the 1950 census for about one in five respondents. Earlier censuses did not collect income but did collect occupation and other characteristics, from which we construct four income proxies: an occupation-based income score using 1950 average income by occupation; an alternative using 1890 average income estimates;[29] the LIDO score, which adds industry, region, state, sex, age, and race;[30] and a predicted household income constructed by machine learning from the 1940 census using additional household and personal characteristics. (See eAppendix Part I.A and eTable 2 in the Supplement for more details).

Our death information comes from individual profiles on the Family Tree, a wiki-style genealogy platform on familysearch.org where users doing family history research collaborate to gather information about their ancestors (eAppendix Part I.B in the Supplement). We link about 50% of the target population in 1900-1940 censuses to a profile on the Family Tree that includes both a

birth and death year. Our link rate is 34% for 1950. Match rates do not vary across the income distribution (eFigure 1 in the Supplement). We drop individuals living longer than 120 years.

Our main analysis sample for 1940 includes about 16 million people ages 40-63 alive in 1941 with non-missing household income in 1939 (13 million positive, 3 million zero). The 1950 sample includes 7 million individuals ages 40-63 alive in 1951 with non-missing household income in 1949 (5 million positive, 2 million zero) and who can be linked to 1940, to obtain two income measures for a subsample. Samples for 1900 to 1930 include 5, 7, 9 and 12 million individuals aged 40-63 alive in 1901, 1911, 1921 and 1931 with a household occupation score.

To confirm results for early decades when incomes are unobserved, we collected the lifespans and actual incomes for two additional samples: a sample from the Iowa Census of 1915 that includes income for 1,590 men and 2,261 women; and a sample of top-income New York State taxpayers in 1923, which includes 1,996 men and 1,530 women whose names and incomes were disclosed by newspapers in 1924 and 1925 (eAppendix Part I.C in the Supplement).

Mortality rates and life expectancy computation

We follow Chetty et al.[1]'s approach to compute period life expectancy. First, we compute one-year mortality rates for ages 40-63 by income percentile. To close the life table, mortality rates after age 63 are estimated using the Gompertz model, which assumes that in old age, log mortality rates are linear in age. The Gompertz model is fitted for each sex, income percentile 2 years earlier, and census year. For survivors at age 90, we assign sex-year-specific but income-independent mortality rates from Social Security period tables. We then construct period life expectancy at age 40 (expected age at death, i.e., life expectancy plus 40 years) by sex, income percentile, and census

year. See Chetty et al.[1] for details. (To assess data quality, we compare aggregate mortality rates to national estimates in eFigure 2 in the Supplement).

Since all our study objects were deceased, we can also examine how cohort life expectancy relates to income. Unlike period life expectancy, computing cohort life expectancy requires no assumptions: it is the average age at death by income percentile. Period and cohort life expectancies differ: the former predicts the average age at death if the current population experienced current mortality rates, while the latter is the average age at death the cohort actually experiences.

Data analysis and availability

We collapsed raw census records linked to the Family Tree into mortality rates and life expectancy estimates by sex, age, income percentile, and census year using Stata version 19.5. Tests of statistical significance were 2-sided with a threshold of 0.05. All estimates were weighted using raking weights so that the weighted sample matches the census population on year of birth, sex, race, marital status, urban residence and other characteristics (Part I.B of the eAppendix and eTable 3 in the Supplement).[31]

We measure income gradients using ordinary least squares (OLS) coefficients from regressions of life expectancy on income percentile, multiplied by 90 to give the estimated gap between the top and bottom income deciles. Data sets underlying every figure in this article and the Supplement are available at the Dryad Digital Repository (doi:10.5061/dryad.m63xsj4jf).

Results

Income gradients in mortality and period life expectancy in 1940

Figure 1 plots one-year log mortality rates by age in 1940 at the 5th and 95th percentiles of (positive) household income. For both men (Figure 1A) and women (Figure 1B), mortality rates tend to be lower at the 95th percentile, but differences are small and statistically insignificant. In contrast, in the 2000s Chetty et al.[1] report large and statistically significant differences at every age (Figure 1).

Figure 2 plots period life expectancy at age 40 by income percentile in 1940, using our data, and in the 2000s, using Chetty et al.[1]'s data. In 1940, period life expectancy was not related to income. A regression of 1940 period life expectancy on income percentile yields a small, statistically insignificant slope of 0.0051 (s.e. 0.0028) for men and -0.0018 (s.e. 0.0049) for women. In the 2000s the relationship is linear, with a slope of 0.1137 (s.e. 0.0016) for men and 0.0715 (s.e. 0.0012) for women. The implied gap between the top and bottom income deciles (=slope*90) in 1940 is 0.46 years for men and -0.16 for women, versus 10.8 and 7.1 years in the 2000s.

Higher income individuals in 1940 were much more likely to live in urban areas (eFigure 3 in the Supplement) and work for their income, but it is harder to measure incomes among the farming and rural populations. We re-estimate the results separately by rural-urban status. There is no strong relationship between income and life expectancy within urban or rural locations (Figure 3). Consistent with early work in the U.S., we find a small positive linear association for men (but not women) in urban areas (slope of 0.0194, s.e. 0.0036), driven by low-income households, but the slope is still less than one-fifth as large as in the 2000s. The figure also documents a significant urban penalty in 1940: urban residents live shorter lives than rural residents at all income percentiles for men and above the 10th percentile for women.

This null result is unchanged if we use individual rather than household income (eFigure 4 and eTable 4A in the Supplement), measure income in levels rather than percentiles (eFigure 5 and

eTable 4B in the Supplement), or exclude the richest 5%, for whom data are missing at higher rates (eTable 4C in the Supplement).

Income gradients in period life expectancy, 1900-1950

Figure 4 plots the estimated gap in period life expectancy between the top and bottom income deciles from 1900 to 1950, and for the 2000s. In 1940 and 1950 the gaps are the same whether we use actual income or income proxies, suggesting we can rely on the proxies for earlier periods. We estimate *negative* and statistically significant gaps in 1900, 1910 and 1920 for men (Figure 4A), and in 1910 for women (Figure 4B). Conclusions are similar whether or not we include farmers (about 10% of individuals in our data), whose incomes are difficult to measure (eFigure 6 in the Supplement), and in 1950 if we average 1940 and 1950 income for a linked sample to improve the measurement of income[32,33] (eFigure 7 in the Supplement).

The estimated gap between the top and bottom 10% of predicted income goes from -2.71 years in 1900 to 0.59 in 1940 for men and from -1.29 to -1.11 for women. Trends in observed (rather than predicted) gaps yield similar conclusions (eFigure 8 in the Supplement). Thus before 1940, adults in more highly paid occupations or with higher predicted incomes could expect to live slightly shorter lives than poorer counterparts, a pattern that disappeared by mid-century. Therefore, the positive relationship between income and lifespan observed in 1960 appears to have emerged around the 1950s.

Income gradients in cohort life expectancy, 1900-1950

Figure 5 reports the estimated gap in cohort rather than period life expectancy. In 1940, cohort gaps are positive and larger than period ones for both women (2.85, s.e. 0.16) and men (1.82, s.e.

0.18). But these relationships were less positive earlier: cohort gradients were negative and statistically significant for men from 1900 to 1920 (Figure 5A), and negative or zero in 1930 depending on the income proxy. Women follow similar patterns though their cohort gradients tend to be more positive (Figure 5B). Regardless of measure, gradients increase across cohorts for both sexes, except in 1950 when the cohort gradient unexpectedly falls; we treat this as suggestive given limitations in the 1950 census data (Part I.A of the eAppendix in the Supplement). These results are robust to the ages included (eFigure 9), the inclusion of farmers (eFigure 10), and to focusing on actual gaps between the top and bottom deciles (eFigure 11 in the Supplement).

In 1940, however, gradients computed with proxies underestimate those using actual income, raising the question of whether cohort gradients were truly negative or zero before 1940. A back-of-the-envelope correction suggests the gaps were still modest in 1900. The 1940 gap for men was 1.82 using actual income and -0.49 using 1950 occupation scores – the largest discrepancy in the 1940 data, implying that proxies understate gaps by 2.30 years. Adding this to the 1900 occupation gap of -1.96 gives a "corrected gap" in 1900 of 0.34; other proxies give similar corrections. In all cases the corrected 1900 gaps for men are small, negative, or both. Plotted across cohorts, these corrected gaps rise from 1900 to 1940 (eFigure 12 in the Supplement) for both sexes.

Evidence from unique historical samples

To verify that our null results for early cohorts are not due to mismeasured income, we examine how lifespan varies with actual income in our two small samples. There is no association in the 1915 Iowa sample (eTable 5 in the Supplement) and a weak association in the 1923 records. Rich taxpaying men (women) in the top 4% of the income distribution lived only 0.6 (1.1) years longer than other men (women), and the gap is not statistically significant for men (eTable 6 in the

Supplement). The implied gap between top and bottom income deciles is 1.20 for men and 2.01 for women, close to our corrected 1920 gaps of 0.81 and 1.71 – still an order of magnitude smaller than in the 2000s. Thus, the cohort gradient at age 40 was small early in the century but moderate by 1950, consistent with the positive and growing period gradient observed after 1960 (Part II.B of the eAppendix in the Supplement), when 40-year-olds lived most of their remaining lives.

Why didn't the rich live longer than the poor from 1900 to 1950?

We provide evidence for two offsetting factors that benefited the poor. First, higher-income households were much more likely to live in cities (eFigure 3 in the Supplement), where period life expectancy was lower – by about 2.13 years for men and 1.62 years for women in 1940 (eFigure 13 in the Supplement). Had the poor lived in cities at the same rate as the rich, in 1940 we would have observed a positive rich-poor gap among men (women) of about 0.86 to 1.33 years (1.29 to 1.85).

Second, personal consumption differed by income. Using consumption surveys from 1917-1919 and 1935-1936, we examined how expenditures varied with household income. The rich spent more on tobacco, alcohol and red meat (eTable 7 in the Supplement), all of which have been linked to increased risk of cardiovascular and cancer disease. The rich also spent more on medical care, but medical care was less effective then.[34] These patterns also explain why cohort gradients are more negative for men: women from these cohorts smoked and drank less.[35,36]

We ruled out two other explanations. The first is that the negative gradient reflects well-paid but dangerous occupations such as manufacturing and mining. It does not: these occupations were in fact less well paid, particularly among men (eTable 8 in the Supplement). The second is that rich

women historically had more children than poor women[37,38] and childbearing raised the risk of dying and of subsequent health complications. But in the 1900, 1910, 1940 and 1950 censuses where women reported all births, fertility is negatively related to income or income proxies (eTable 9 in the Supplement).

Discussion

We find no positive period gradient from 1900 to 1940, but Kitagawa and Hauser[13] document one in 1960. The gradient thus likely emerged around the 1950s, when infectious diseases had become negligible and chronic diseases were the main cause of death. Earlier gains in life expectancy came primarily from public health efforts to contain infectious disease (such as cleaning the water or eradicating hookworm).[27] Paid for by government and non-profit institutions, these efforts benefited large swaths of the population regardless of income, raised life expectancy substantially, and helped eliminate the urban penalty that had disproportionately hurt the rich.

In the second half of the century, innovation instead took the form of individual treatments sold in the market: antibiotics, developed mostly during the "golden age of antibiotics" from 1942 to 1960, and medicines (e.g. statins or high blood pressure medications) or surgical interventions (e.g. bypass surgery) to treat heart disease developed later in the 1960-1980 period. Drug patents were low and constant at around 200 per year between 1930 and 1950 but exceeded 1,000 per year by 1970.[39] Consistent with more expensive and more effective care, real per capita medical expenditures rose more than eightfold between 1960 and 2000.[40]

Differential access to these expensive innovations may explain the rise of the period gradient, particularly in the U.S., which, unlike other rich countries, did not implement universal health

insurance in the 20th century. Cohort gradients went from negative or insignificant in 1900 to positive by 1940: individuals aged 40 in 1940 and 1950 lived their remaining 30-40 years during this modern medicine era, potentially contributing to the rise of gradients. Previous research shows that adoption of new health innovations (such as statins or HAART) was greater among the rich and educated, and resulted in increased socioeconomic disparities in health.[41-43] Our results therefore suggest that highly effective but expensive modern medicine is at least partly responsible for the rise in lifespan inequality observed today. Consistent with this hypothesis, Hagen et al.[4] show using identical methods that income lifespan associations in Sweden rose after 1960 despite decreasing inequality.

Limitations

This study has several limitations. First, we did not compute gradients in life expectancy at birth, which may differ from those in adulthood: selective survival by income may have occurred at younger ages even if it did not after age 40. Existing evidence is not supportive of this hypothesis – infant mortality in 1900 in the U.S. was higher for children of fathers in upper white collar jobs[44] – but we cannot fully evaluate it because genealogies severely underrepresent individuals who died before age 25. Nevertheless, we constructed novel databases that allow us to offer the first evidence on income gradients in life expectancy prior to 1960. Second, our income measures suffer from measurement error. However, our results are robust to a variety of income measures and are verified in small samples with income.

Conclusions

Rich adults in America did not live substantially longer than poor adults in the first half of the 20th century whereas in the early 2000s, the richest Americans could expect to live nine years longer than the poorest, although inequality was just as large in both periods. Thus income inequality need not lead to lifespan inequality.

ARTICLE INFORMATION

Author Contributions: Carson Zhang had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Lleras-Muney, Goodman-Bacon, Yue, Price, Zhang.

Acquisition, analysis, or interpretation of data: Lleras-Muney, Price, Zhang.

Drafting of the manuscript: Goodman-Bacon, Lleras-Muney, Yue, Price, Zhang.

Critical revision of the manuscript for important intellectual content: Goodman-Bacon, Lleras-Muney, Yue, Price, Zhang.

Statistical analysis: Lleras-Muney, Zhang.

Obtained funding: Lleras-Muney, Yue.

Administrative, technical, or material support: Price.

Supervision: Lleras-Muney.

Conflict of Interest Disclosures: None reported.

Funding/Support: We acknowledge grant support (R01AG088639) from the National Institute on Aging of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

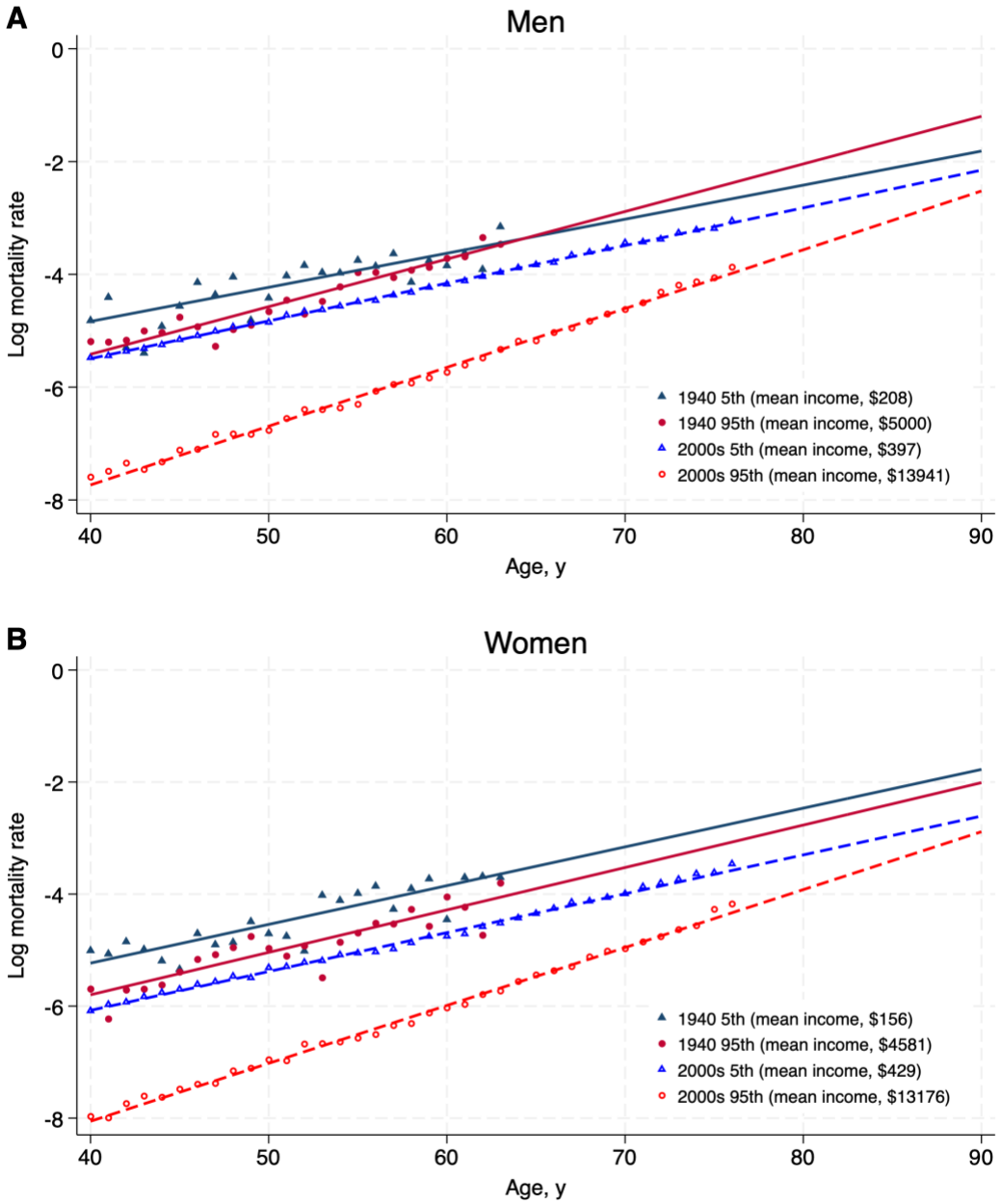
Additional Contributions: We are grateful to Angus Deaton, Marika Cabral, Patrick Heuveline, Evan Roberts, and Hiram Beltran-Sanchez and the seminar participants at All-UC Demography conference at UCSD and the CIREQ-McConnell Interdisciplinary Workshop for their valuable comments.

REFERENCES

1. Chetty R, Stepner M, Abraham S, et al. The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*. 2016;315(16):1750-1766. doi:10.1001/jama.2016.4226
2. Kinge JM, Modalsli JH, Øverland S, et al. Association of Household Income With Life Expectancy and Cause-Specific Mortality in Norway, 2005-2015. *JAMA*. May 21 2019;321(19):1916-1925. doi:10.1001/jama.2019.4329
3. Milligan K, Schirle T. The evolution of longevity: Evidence from Canada. *Canadian Journal of Economics*. 2021;54(1):164-192. doi:https://doi.org/10.1111/caje.12497
4. Hagen J, Laun L, Lucke C, Palme M. The rising income gradient in life expectancy in Sweden over six decades. *Proceedings of the National Academy of Sciences*. Apr 8 2025;122(14):e2418145122. doi:10.1073/pnas.2418145122
5. Mortensen LH, Rehnberg J, Dahl E, et al. Shape of the association between income and mortality: a cohort study of Denmark, Finland, Norway and Sweden in 1995 and 2003. *BMJ Open*. Dec 23 2016;6(12):e010974. doi:10.1136/bmjopen-2015-010974
6. Ingleby FC, Woods LM, Atherton IM, Baker M, Elliss-Brookes L, Belot A. Describing socio-economic variation in life expectancy according to an individual's education, occupation and wage in England and Wales: An analysis of the ONS Longitudinal Study. *SSM - Population Health*. 2021/06/01/ 2021;14:100815. doi:https://doi.org/10.1016/j.ssmph.2021.100815
7. Mackenbach JP, Stirbu I, Roskam AJ, et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med*. Jun 5 2008;358(23):2468-81. doi:10.1056/NEJMsa0707519
8. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav*. 1995;Spec No:80-94.
9. Snyder-Mackler N, Burger JR, Gaydosh L, et al. Social determinants of health and survival in humans and other animals. *Science*. 2020;368(6493):eaax9553. doi:doi:10.1126/science.aax9553
10. Marmot M. The health gap: the challenge of an unequal world. *Lancet*. Dec 12 2015;386(10011):2442-4. doi:10.1016/s0140-6736(15)00150-6
11. Sapolsky RM. *Why zebras don't get ulcers: The acclaimed guide to stress, stress-related diseases, and coping*. Holt paperbacks; 2004.
12. Sherman A, Trisi D, Cureton J. *A guide to statistics on historical trends in income Inequality*. Center on Budget and Policy Priorities; 2024.
13. Kitagawa EM, Hauser PM. *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*. Harvard University Press; 1973.
14. Backlund E, Sorlie PD, Johnson NJ. The shape of the relationship between income and mortality in the United States: Evidence from the National Longitudinal Mortality Study. *Annals of Epidemiology*. 1996/01/01/ 1996;6(1):12-20. doi:https://doi.org/10.1016/1047-2797(95)00090-9
15. Duleep HO. Measuring socioeconomic mortality differentials over time. *Demography*. 1989;26(2):345-351.
16. Hadley J, Osei A. 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. *Medical Care*. Sep 1982;20(9):901-14. doi:10.1097/00005650-198209000-00004
17. Lantz PM, House JS, Lepkowski JM, Williams DR, Mero RP, Chen J. Socioeconomic Factors, Health Behaviors, and Mortality: Results From a Nationally Representative Prospective Study of US Adults. *JAMA*. 1998;279(21):1703-1708. doi:10.1001/jama.279.21.1703
18. Sorlie PD, Backlund E, Keller JB. US mortality by economic, demographic, and social characteristics: the National Longitudinal Mortality Study. *American Journal of Public Health*. Jul 1995;85(7):949-56. doi:10.2105/ajph.85.7.949
19. Cristia JP. Rising mortality and life expectancy differentials by lifetime earnings in the United States. *Journal of Health Economics*. Sep 2009;28(5):984-95. doi:10.1016/j.jhealeco.2009.06.003
20. Pappas G, Queen S, Hadden W, Fisher G. The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986. *New England Journal of Medicine*. 1993;329(2):103-109. doi:doi:10.1056/NEJM199307083290207
21. Rogot E, Sorlie PD, Johnson NJ. Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study. *Public Health Reports*. 1992;107(4):457.
22. Schalick LM, Hadden WC, Pamuk E, Navarro V, Pappas G. The widening gap in death rates among income groups in the United States from 1967 to 1986. *International Journal of Health Services*. 2000;30(1):13-26.

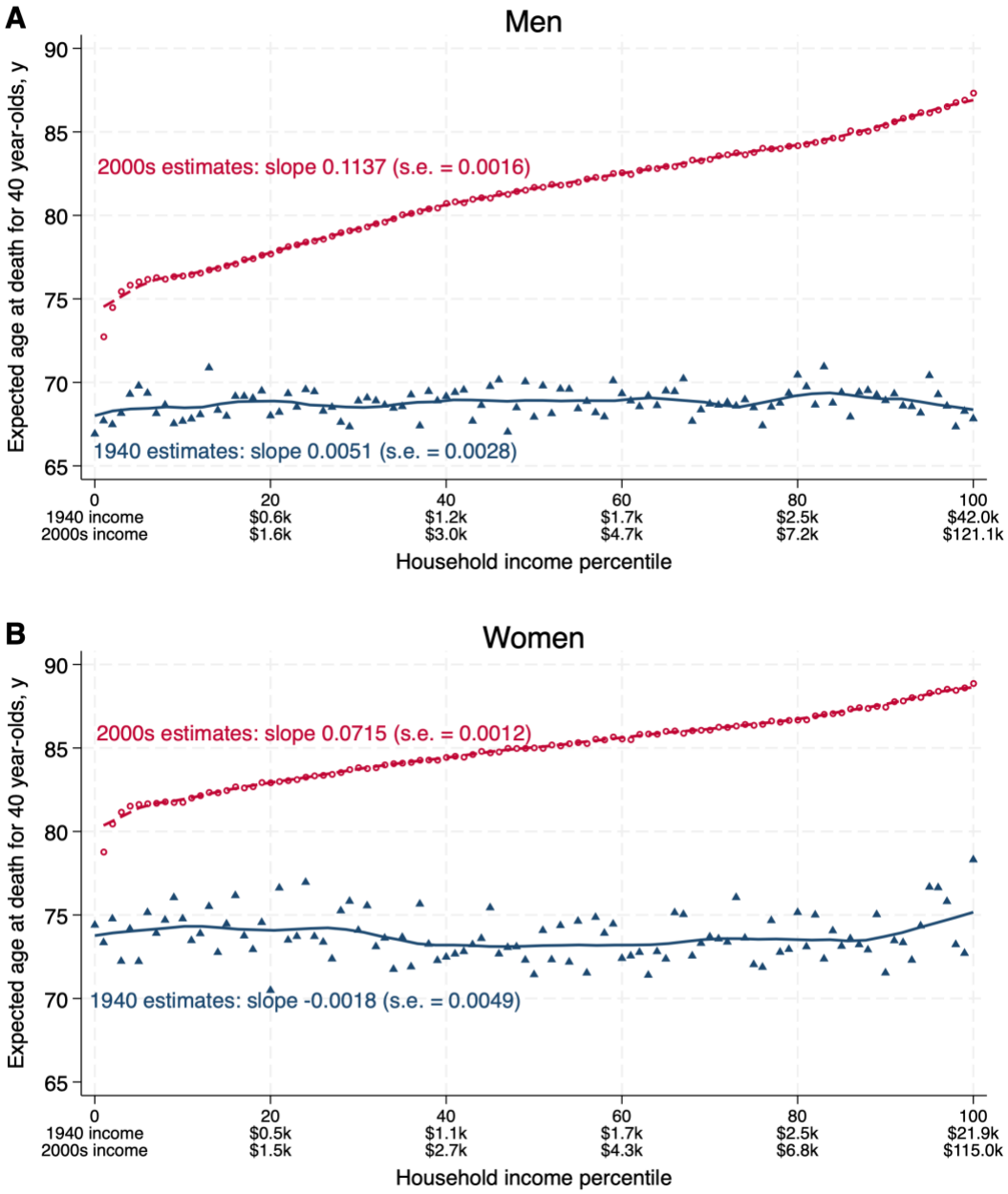
23. Waldron H. Trends in mortality differentials and life expectancy for male social security-covered workers, by socioeconomic status. *Soc Secur Bull*. 2007;67(3):1-28.
24. Preston SH. The changing relation between mortality and level of economic development. *Population studies*. 1975;29(2):231-248.
25. Bridgman B. A Century of Super-rich Longevity. *BEA Working Paper Series*. 2025;
26. Clarke PM, Tran-Duy A, Roope LSJ, Stiles JA, Barnett AG. The comparative mortality of an elite group in the long run of history: an observational analysis of politicians from 11 countries. *European Journal of Epidemiology*. 2022/09/01 2022;37(9):891-899. doi:10.1007/s10654-022-00885-2
27. Cutler D, Deaton A, Lleras-Muney A. The Determinants of Mortality. *Journal of Economic Perspectives*. 2006;20(3):97–120. doi:10.1257/jep.20.3.97
28. Dribe M, Eriksson B. The Long-Term Development of the Mortality Gradient: Socioeconomic Differences in Adult Life Span of Swedish Cohorts 1841–1920. *Population and Development Review*. 2025/6/4 2025;doi:10.1111/padr.70015
29. Sobek M. Work, status, and income: Men in the American occupational structure since the late nineteenth century. *Social Science History*. 1996;20(2):169-207.
30. Saavedra M, Twinam T. A machine learning approach to improving occupational income scores. *Explorations in Economic History*. 2020;75doi:10.1016/j.eeh.2019.101304
31. Kolenikov S. Calibrating Survey Data using Iterative Proportional Fitting (Raking). *The Stata Journal*. 2014;14(1):22-59. doi:10.1177/1536867x1401400104
32. Ruggles S, Drew J, Fitch CA, et al. The IPUMS multigenerational longitudinal panel: progress and prospects. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*. 2025:1-14. doi:10.1080/01615440.2025.2514110
33. Helgertz J, Price J, Wellington J, Thompson KJ, Ruggles S, Fitch CA. A New Strategy for Linking U.S. Historical Censuses: A Case Study for the IPUMS Multigenerational Longitudinal Panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*. 2022;55(1):12-29. doi:10.1080/01615440.2021.1985027
34. Beeson PB. Changes in medical therapy during the past half century. *Medicine*. Mar 1980;59(2):79-99. doi:10.1097/00005792-198003000-00001
35. Harris JE. Cigarette smoking among successive birth cohorts of men and women in the United States during 1900-80. *J Natl Cancer Inst*. Sep 1983;71(3):473-9.
36. Zhang Y, Guo X, Saitz R, et al. Secular trends in alcohol consumption over 50 years: the Framingham Study. *American Journal of Medicine*. Aug 2008;121(8):695-701. doi:10.1016/j.amjmed.2008.03.013
37. Vogl TS. Differential Fertility, Human Capital, and Development. *The Review of Economic Studies*. 2016;83(1):365-401. doi:10.1093/restud/rdv026
38. Malthus T. *An essay on the principle of population*. Electronic Scholarly Publishing Project; 1798.
39. Gross DP, Sampat BN. The Therapeutic Consequences of the War: World War II and the 20th-Century Expansion of Biomedicine. *National Bureau of Economic Research*. 2025;
40. Cutler DM, Rosen AB, Vijan S. The value of medical spending in the United States, 1960-2000. *New England Journal of Medicine*. Aug 31 2006;355(9):920-7. doi:10.1056/NEJMsa054744
41. Glied S, Lleras-Muney A. Technological innovation and inequality in health. *Demography*. Aug 2008;45(3):741-61. doi:10.1353/dem.0.0017
42. Chang VW, Lauderdale DS. Fundamental Cause Theory, Technological Innovation, and Health Disparities: The Case of Cholesterol in the Era of Statins∗. *Journal of Health and Social Behavior*. 2009;50(3):245-260. doi:10.1177/002214650905000301
43. Hamilton B, Hincapié A, Kalish EC, Papageorge NW. Medical innovation and health disparities. *National Bureau of Economic Research*. 2021;
44. Preston SH, Haines MR. *Fatal years: Child mortality in late nineteenth-century America*. Princeton University Press; 2014.

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



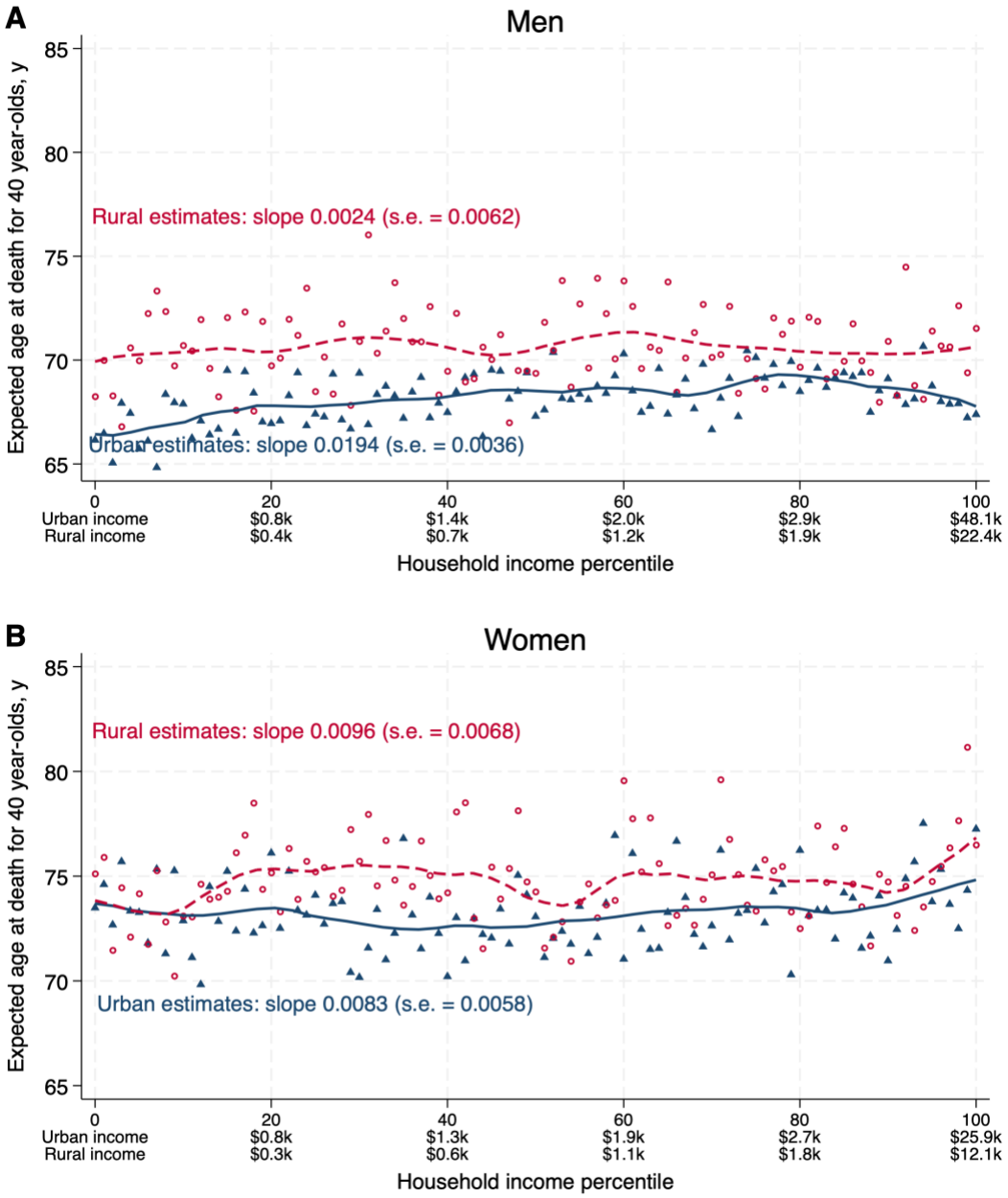
A, Gompertz approximations for men. B, Gompertz approximations for women. Log mortality rates are natural logarithms of observed mortality rates. The 2000s estimates are from Chetty et al.[1] Estimates for 1940 are based on mortality rates in 1941 and income in 1939. The solid (1940) and dashed (2000s) lines are fitted Gompertz curves using the points shown in the figure. Income values are adjusted to 1939 prices.

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



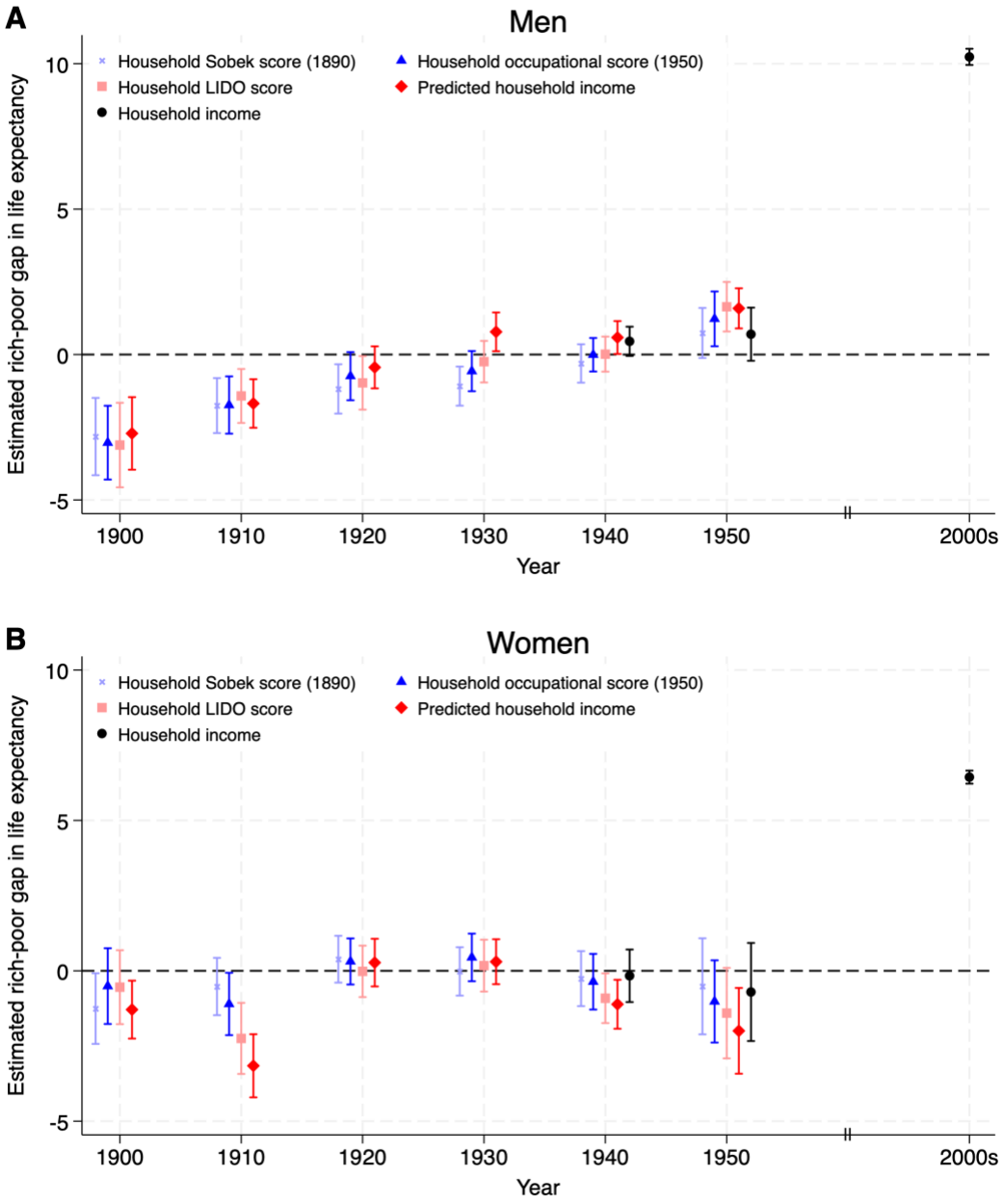
A, Period life expectancy for men. B, Period life expectancy for women. The 2000s estimates are from Chetty et al.[1] Estimates for 1940 are based on mortality rates in 1941 and income in 1939. Percentiles are defined within the positive income distribution. For 1940, an additional data point for zero income (x-axis value = 0) is included (about 18% of individuals come from households with zero income). The solid (1940) and dashed (2000s) lines are nonparametric fits, and the slopes are estimated using ordinary least squares (OLS) regressions based on the data points shown in the figure. Income values are adjusted to 1939 prices.

Figure 3. Period Life Expectancy for 40-Year-Olds in 1940 by Household Income Percentile and Urban/Rural Status



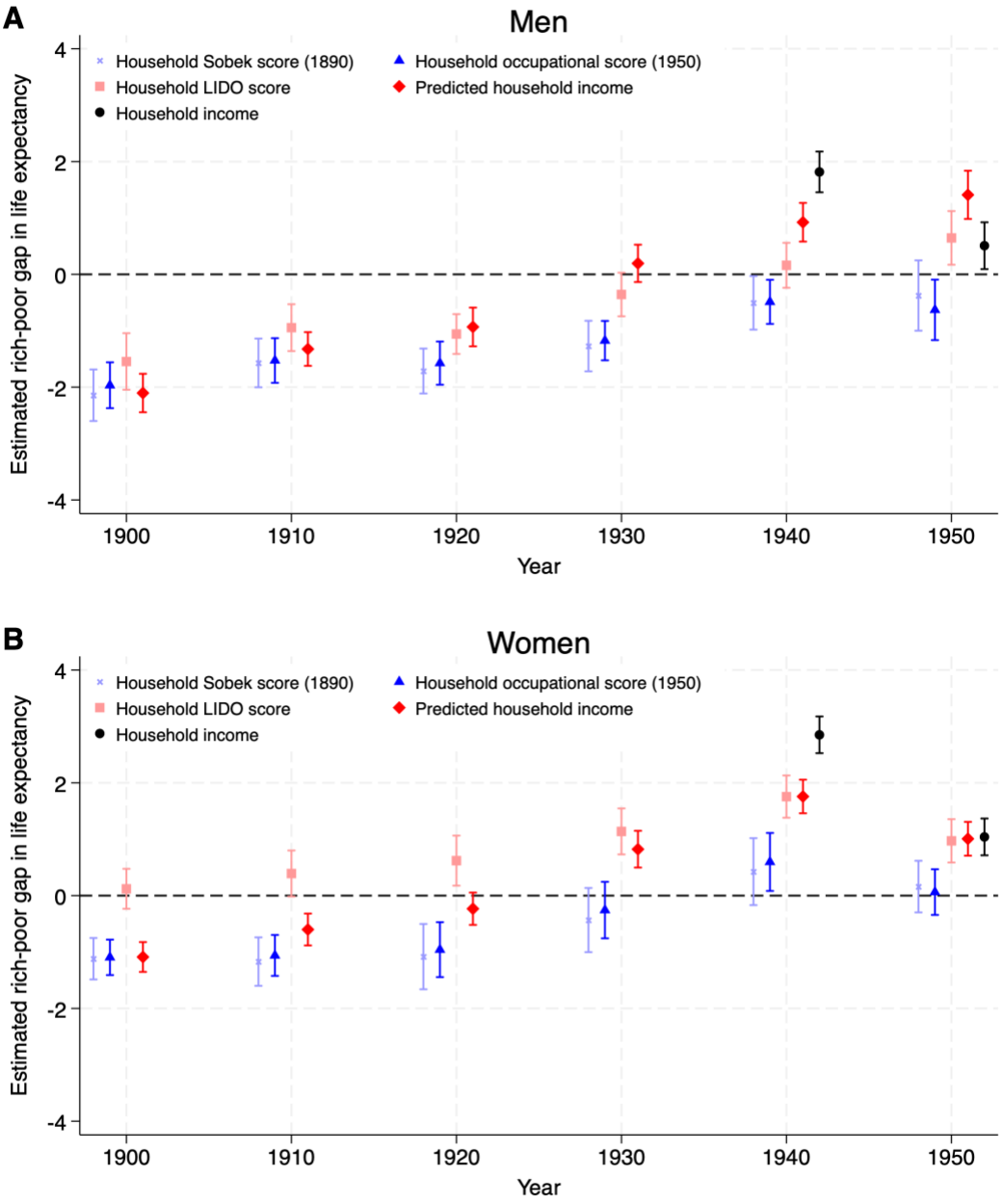
A, Period life expectancy for men. B, Period life expectancy for women. In our working sample, 65% of individuals lived in urban areas in 1940. Estimates are based on mortality rates in 1941 and income in 1939. Percentiles are defined within the positive income distribution. Additional data points for zero income (x-axis value = 0) are included. The solid lines are nonparametric fits, and the slopes are estimated with ordinary least squares (OLS) regressions based on the data points shown in the figure.

Figure 4. Estimated Period Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900



A, Estimated rich-poor gap for men. B, Estimated rich-poor gap for women. The 2000s estimates are based on data points from Chetty et al.[1] Estimates for each census year are based on mortality rates the year after the census, and income in the year before the census. Income is available in 1940 and 1950 only. For 1900, 1910, 1920 and 1930 we use various measures of predicted income. For each income measure and year, period life expectancy is estimated by percentiles of the corresponding income measure for each sex. Slopes are obtained from ordinary least squares (OLS) regressions of estimated period life expectancy on income percentiles and then multiplied by 90 to give the estimated gap between the top and bottom income deciles. Data markers indicate the estimated gap; error bars indicate 95% CIs.

Figure 5. Estimated Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900



A, Estimated rich-poor gap for men. B, Estimated rich-poor gap for women. Estimates for each census year are based on lifespan conditional on being alive the year after the census, and income the year before the census. Income is available in 1940 and 1950 only. For 1900, 1910, 1920 and 1930 we use various measures of predicted income. For each income measure and year, cohort life expectancy is estimated by percentiles of the corresponding income measure for each sex. Slopes are obtained from ordinary least squares (OLS) regressions of estimated cohort life expectancy on income percentiles and then multiplied by 90 to give the estimated gap between the top and bottom income deciles. Data markers indicate the estimated gap; error bars indicate 95% CIs.

Supplementary Online Content

The Rich Did Not Always Live Longer Than the Poor in the United States

Carson Zhang, Andrew Goodman-Bacon, Joseph Price, Dahai Yue, Adriana Lleras-Muney

eAppendix
eFigure 1. Proportion Matched With Lifespan by Household Income/Occupation Score Percentile in Each Census Year
eFigure 2. National Mortality Profile From Census Tree and SSA Life Table in Each Census Year
eFigure 3. Proportion Living in Urban Areas by Household Income Percentile and Sex in 1940
eFigure 4. Period Life Expectancy for 40-Year-Olds in 1940 by Percentile of Different Income Measures
eFigure 5. Period and Cohort Life Expectancy for 40-Year-Olds in 1940 by Household Income Levels
eFigure 6. Estimated Period Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900 (Including Farmers)
eFigure 7. Period Life Expectancy for 40-Year-Olds in 1950 by Household Income Percentile, 1950 Measure Versus Average 1940 and 1950 Measure
eFigure 8. Observed Period Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900
eFigure 9. Estimated Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 35-45-Year-Olds Since 1900
eFigure 10. Estimated Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900 (Including Farmers)
eFigure 11. Observed Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900
eFigure 12. Corrected Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900
eFigure 13. Period and Cohort Urban Penalty in Adult Life Expectancy for 40-Year-Olds Since 1900
eTable 1. Data construction compared with Chetty et al.
eTable 2. Predictive validity of household income proxies (OLS regression)
eTable 3. Weighted summary statistics of cohorts aged 40 to 63 in the year after each census
eTable 4. Slope estimates of life expectancy on income
eTable 5. Lifespan and income in the 1915 Iowa census
eTable 6. Lifespan for rich individuals in 1920 census for New York State
eTable 7. Household consumption and household income (OLS regression)
eTable 8. Hazardous occupation and income (OLS regression)
eTable 9. Number of children ever born and household income (OLS regression)
eTable 10. Summary statistics of cohorts aged 40 to 63 in 1941
eReferences

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix

Part I: Data and Measurement

A. Income measures

To make results comparable over time, we closely follow the approach in Chetty et al.[1] and investigate the relationship between household income in the year before each census and life expectancy at age 40 in the year after each census (eTable 1). Unlike Chetty et al.,[1] we exclude farmers for comparison across years as their income was not well captured even in the 1940 and 1950 censuses. However, we show that including farmers with imputed income does not substantively affect our results.

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 start by restricting our attention to individuals with positive income to be comparable with Chetty et al.,[1] but for most of our analyses, we also include individuals and households who report zero income. We also restrict attention to individuals who survive to 1941 and are 40-63 in 1941, so that the income is measured 2 years prior. Prior work has determined that annual earnings measured during these ages (40-63) is a reasonable approximation of lifetime earnings,[2] and that rank correlations between individual earnings in different age windows was relatively high in the 1950s.[3] We construct our main measure of household income by summing over the income of all the household members. 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

than in 1940 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 cohorts, about 32.5% of total household income comes from sources other than household wage and salary income. Since only a random subsample of individuals was asked about income, our household income measures, which sums over all individuals in the household, is measured with substantial error in 1950, much more so than in 1940. In addition, the underlying income data has not been accurately transcribed for an estimated 25-30% of the population in the currently released data (https://usa.ipums.org/usa/full_count.shtml#January2024).

Censuses prior to 1940 did not collect any direct measures of income. Because of this, we use four different proxies for income which we can compute in every census from 1900 to 1950. First, we use occupation scores, provided by IPUMS, which are available for each person who reported their occupation in the census. These assign individuals the 1950 income in their occupation and can be computed even among the unemployed. While this is a widely used measure of income in historical studies, it has two important downsides. First, it only assigns income based on the 1950 incomes in a given occupation. If the relative incomes across different occupations changed substantially from 1900 to 1950 then we will be misclassifying individual's income and income rank. For example, Starr[4] documents that in the early 1900s physicians earned about 40% more than average workers (p.85) but by the early 1930s physicians' earnings were four times larger than average (p.143). To address this, we also impute incomes to a given occupation using Sobek's estimated average incomes by occupation in 1890.[5]

The second issue is that these measures miss variation in income within occupations, which is significant: in 1940 about 57% of the variation in income is within rather than between

occupations. To address this, we compute two additional measures which use occupation but also use other information about recipients to predict income and generate variation in income within as well as across occupations. First, among those with an occupational response, we impute individual income using the LIDO score (Lasso-adjusted industry, demographic variables, and occupation score) developed by Saavedra and Twinam[6] to better predict historical incomes.[1] The original LIDO score is on a log-income scale, and we take its exponential form. Second, we also create our own predicted measure of household income using a machine learning model trained on the 1940 census. Unlike the LIDO score, we aim to predict household level income, and we include household level characteristics as well as individual characteristics as predictors.[2]

In all cases except for predicted household income, we first impute individual income and then we sum over all individuals in a household to obtain a measure of household income. For 1940 and 1950 we can compare our results using actual incomes and income proxies. But for 1900-1930 we must rely on proxies alone.

B. Mortality data

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 historical records to the profile that often corroborate this information. These attached historical records come from the 20.5 billion records that FamilySearch has on its website, including the

[1] Specifically, the LIDO score predicts individual income using age, sex, race, and state of residence in addition to occupation and industry.

[2] Household income was predicted using a gradient boosting model trained separately by sex on the 1940 census, using individual characteristics (race, marital status, region of birth, elite-occupation with occ1950=0-99, 200-290, and 400-410 corresponding to "professional and technical workers", "managers, officials, and proprietors", and "sales managers and higher-grade sales workers") and household characteristics (county, urban residence, farm status, number of servants, number of employees, and the summed occupational score of all household members). The model is trained among those aged 30-45 for a better approximation of lifetime income. The trained models were then applied to the 1900-1950 censuses to generate predicted household income for each individual.

U.S. 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 a likely match to each of the census records in our sample. We then use the birth and death date information from these profiles.

The main advantage of the data is that they include death information from various sources including national death databases such as the Social Security, death certificates from state registries, and dates from gravestones, obituaries and family members. The data from FamilySearch allow us to track individuals from 1940 until 2023, which is not possible using any other national database. The main drawback of the data derived from genealogies is that they underrepresent 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[7-9] and construct raking weights using the algorithm of Kolenikov.[103]

We link about 50% of the target population in 1900-1940 censuses to a profile on the Family Tree that includes both a birth and death date (sometimes just a year). Our link rate is 34% for 1950 census. The 1950 link rates are lower because those records have had less time to be examined and connected to profiles on the Family Tree. U.S. census records are made public after 72 years, so the 1950 census only recently became available on FamilySearch. Our link rates are high by historical standards, which typically link between 10% and 30% of the target populations.[11,12] Previous work has documented that the data constructed by linking U.S. census records to the Family Tree accurately describes population trends by sex.[13,14]

[3] The algorithm performs iterative proportional fitting to produce a set of calibrated survey weights such that the sample weighted totals of control variables match the known population totals. The variables include year of birth, sex, race, ethnicity, marital status, urban residence, years of schooling, family size, labor force participation, and occupation score. This weighting scheme performs better than the more commonly used inverse probability weighting, as shown in eTable 10.

Since we are not able to track the lifespan of the entire target population, a key issue for this study is whether the match rates in our data vary by income: in other words, do we observe the lifespan of rich and poor individuals at the same rates? Our match rates are relatively stable across the income distribution. Up to the 95th percentile of income, the differences in match rate by income are small (eFigure 1). However, the match rates are 10.0 and 5.6 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) in 1940. To see if missing data at the top results in biased estimates of life expectancy for the rich, we hand linked 347 out of 600 randomly selected individuals with high incomes and missing death dates. The mean age at death in this sample is 76.37 compared to 76.80 in the original data, suggesting that missing data at the top does not affect our conclusions. We test whether our results are sensitive to how we measure income at the very top.

C. Additional samples

To provide supporting evidence on the income gradients and the quality of our income proxies during the decades in which we do not observe income, we construct two additional samples in which income is measured directly rather than imputed. The first sample comes from the 1915 Iowa State Census. To our knowledge, the 1915 Iowa State Census is the only census conducted anywhere in the United States before 1940 to record individual income. This unique feature makes it especially valuable for our analysis, as it allows us to observe the relationship between income and longevity a quarter century before the federal census first collected information on earnings. We use the digitized sample of approximately 2.5% of the population prepared by Goldin and Katz,[15] and link its records to profiles on the Family Tree and recover valid birth and death dates for about 50% of them.

The second sample comes from high-income taxpayers in New York State in 1923. The Revenue Act of 1924 included a tax publicity provision under which the amount of tax paid by every filer became available for public inspection. Newspapers responded by publishing the names, addresses, and tax payments of tens of thousands of individuals and corporations, and *The New York Times* devoted several pages to New York tax returns in both 1924 and 1925. We use the published lists digitized by Marcin,[16] link them to the 1920 U.S. Census, and then match the resulting records to profiles on the Family Tree to recover individuals' lifespans. Among all individuals in the 1920 U.S. Census for New York State, we classify a person as rich if they or their spouse appears in these taxpayer lists. According to Marcin: "Newspapers were particularly interested in the highest tax payments. However, a tax payment of any size indicated high status, as only 6.5-7 percent of the population even filed tax returns, and only about 4 percent paid tax on a return (Statistics of Income)." The linked sample includes a number of well-known individuals such as John D. Rockefeller, Henry Ford, Andrew Mellon, Frederick Vanderbilt, Waldorf Astor and various members of the Guggenheim family.

For both samples, we focus on individuals aged 35 to 45, resulting in samples of 1,590 men and 2,261 women from the 1915 Iowa Census and 1,996 men and 1,530 women from the New York high-income taxpayer sample.

D. Income proxy validation

We used the 1915 Iowa census and the 1940 census to assess the validity of our income proxy measures (eTable 2). We regress actual household income percentiles on percentiles of household income proxies. Using occupation-based incomes result in a R2 of about 24-30%, the LIDO score has a slightly higher R2 of 32-34% (this cannot be computed in 1915 because the Iowa census

does not include industry which is a key input in the LIDO computation). Finally, our machine learning based prediction yields the highest R2 of 25-35% in 1915 and 42-44% in 1940.

Part II: Literature Review

A. Theories of how individual income affects lifespan

There are several theories linking resources broadly construed and health outcomes.

An important view of the relationship comes from animal studies linking social rank, stress and health outcomes, summarized by Sapolsky.[17] In small animal groups, individuals at the bottom of the social hierarchy are typically observed to have lower resources, to experience elevated stress (measured with biomarkers) and to be more susceptible to a range of diseases compared to individuals at the top. While this relationship is not always observed (for instance individuals at the very top have been observed to live shorter lives), and can vary substantially depending on the environment (and varies across species depending on its resources, social organization, hierarchy stability, and the availability of coping opportunities), it has been observed in many settings and has been shown to be causal in experimental settings.[18]

Work in epidemiology, most notably the Whitehall study,[19] has documented similar relationships between social rank and health and argued that the same mechanisms that are at play in animals hold for humans. This "social determinants of health" view, famously advocated by Sir Michael Marmot, argues that diseases have deep causes that are linked to social, environmental and economic conditions which together lead to the onset of disease, much before the health care system becomes involved. In this view a person's relative position in society matters and income inequality, in addition to income, determines health.[20]

A different perspective comes from the influential "fundamental causes of disease" theory of Link and Phelan[21] which argues that individuals with resources, including income, will always find

ways to increase their survival. Causes of death may change over time, as well as ways to avoid disease, but those at the top of the social hierarchy will always adapt and use their resources to their advantage. This theory emphasizes the ability of those at the top to use resources whereas work emphasizing the role of stress focus on the inability of those at the bottom to avoid or to compensate for it.

An important caveat to the theories above is that they generally link health with a broad notion of socio-economic status. They do not necessarily or uniquely describe the relationship between income *per se* and health. Education, knowledge, social connections, etc. may also contribute to the underlying notion of status.

Materialist views, such as the one advocated by the WHO, emphasize instead that income is needed to purchase basic goods and services such as food, housing and health care, which are essential in the determination of health. A key debate then among social scientists is whether absolute income levels matter or whether relative income plays an important role, once basic needs have been met. Altogether however these theories predict income and health will be positively associated but perhaps with some diminishing returns as individuals and societies get rich.

Theories from economics predict a more nuanced relationship between income and health because these theories suggest that individuals may choose to trade off health against consumption. Rosen's theory of compensating differentials[22] argues that individuals may accept higher mortality risks in exchange for higher wages. In Grossman's seminal model[23] of health production, the relationship between wages and health may be ambiguous due to offsetting income and substitution effects: individuals with higher wages may choose to work more, spend less time investing in their health, and thus have worse health outcomes. Basic consumer demand theory posits that individuals with higher incomes may spend more on consumer goods, including on goods that are known to be

health-harming or even addictive.[24] In contrast to the fundamental cause theory, these consumption patterns could lead to worse health if the perceived (or experienced) benefits of consumption outweigh its health risks.

Importantly the discussion so far focuses on the relationship between income in adulthood and health. Work from a variety of disciplines has demonstrated that health at any point in time is the result of the cumulative experiences and exposures of individuals. Conditions in utero, childhood, adolescence and young adulthood will leave a permanent imprint on a person's health. Thus income (and other resources) throughout the life course matter, and may result in a weak relationship between adult income and adult health. Finally, in addition to individual income, the aggregate level of income in a society may be influential since it will determine the provision of public goods and shape institutions and laws that together may impact health. Therefore, the relationship between adult income and health may be different in different societies and times.

As discussed below (Part II.E) there are other reasons why income and health may be associated that do not involve a causal relationship from income to health. For a recent review see Lleras-Muney et al.[25]

B. Representative estimates since 1960

Work in the U.S. since 1960 shows that income gradients in life expectancy have increased in recent times, though it is difficult to fully ascertain the magnitude of these changes because studies use different measures of life expectancy/mortality and income. Most papers in the literature report either gaps or ratios between the mortality rates of higher- and lower-income people by age. They rarely estimate life expectancy gaps, which often require modelling assumptions. Using Census data matched to death certificates, Kitagawa and Hauser[26] reported that in 1960 the standardized mortality ratio (the number of deaths over the number of expected deaths) declines with income,

and was 1.51 for individuals with incomes under $2k, 1.2 for those with incomes ranging from $2-4k, and 0.84 for those with incomes greater than $10k or more (among white men between the ages of 25 and 64). For women these ratios were 1.05, 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[27] reported that these ratios were 1.59, 1.79 and 0.71 for white men in 1973-1978. Using survey data from the National Health Interview Survey (NHIS) matched to death records, Pappas et al.[28] reported that these ratios for white men were 2.5 for under $11k, 0.6 for $25k or more in 1986. Schalick et al.[29] 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.[27,30-33]

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.[4] Using the NLMS, Rogot et al.[38] reported that between 1979-1985 the difference in period life expectancy for white men 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[39] used data from the Survey of Income and Program Participation (SIPP) to

[4] We ignore here papers studying the relationship between aggregate income and life expectancy, which started with the seminal study by Preston[34] documenting a positive and concave relationship between a country's GDP and its life expectancy. Papers in the U.S. looking at the relationship between aggregate county or state income and life expectancy include the early work of Hadley & Osei[31] or Singh & Siahpush.[35] Most recently Schwandt et al.[36] 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.[37] looked at the evolution of state income and state mortality since 1968 and show that state mortality was not strongly correlated with in state income in 1968 but by 2016 state life expectancy was strongly correlated with state income. As noted earlier these studies do not answer the same question we are interested in answering because of the ecological fallacy.

show that the gap in period life expectancy (between ages 35 and 76) between the top and bottom quantiles of income increased by 0.9 years between 1983-1997 and 1998-2003 for men, and 0.8 years for women. Note however that these later numbers are not quite comparable to the Rogot et al.[38] ones because Cristia[39] only computes partial life expectancy numbers.

The most prominent recent study on the question was conducted by Chetty et al.[1] 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 used more accurate income data and included information on more than 100 million adults in the U.S., a vastly larger sample compared to e.g. Kitagawa and Hauser's[26] paper (about 62,000 deaths) or the Rogot et al.[38] work (about 800,000 individuals). The sample size is larger despite Chetty et al.'s exclusion of individuals not reporting income to the IRS. They document significant gaps in life expectancy at age 40, with men in the top 1% expecting to live 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.

A study by Waldron[40] uses Social Security Administration data in which lifetime income (ages 45-55) is well measured from earnings records which are individually matched to deaths occurring from 1972 through 2001. It finds a 1.2 year gap in cohort life expectancy at age 60 between the top and bottom halves of the income distribution for the 1912 cohort of men who turned 40 in 1952, which widened to 5.8 years for the 1941 cohort, though the latter is computed used extrapolations rather than being observed.

Outside of the U.S. two studies have also found that income gradients have increased in recent times. Kinge et al.[41] essentially replicate the Chetty et al.[1] analysis using administrative data from Norway. Surprisingly they find that income gradients in life expectancy at age 40 are also large (though a bit smaller): there is a 13.8-year gap (men), 8.4-year gap (women) between top and bottom 1% of income. Similar to the U.S. case, these gaps increased from 2005 to 2015. Hagen et al.[42] look at the case of Sweden and further extend these results back to the 1960s, 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 2000s of 10.9 for men and 8.6 for women. A study using Canadian data however found smaller gradients that have not widened from 1970s to 2000s.[43]

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 representative 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 since the 1960s.[5] The earliest mortality gradients by education are reported by Lleras-Muney et al.[13] who find small gaps in annual mortality by education in 1950 that grow by 1970. That work also reports that education gradients in lifespan grew across successive cohorts.

C. Evidence from early studies (prior to 1960)

[5] The association between education and life expectancy among adults has grown substantially since 1960 in the US. For example, Kitagawa and Hauser[26] 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.[44] More recently Case and Deaton[45] 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.

This evidence has been discussed in great detail by Antonovsky.[46] We briefly summarize here the findings from these studies.

Prior to 1960, many studies compare mortality rates between neighborhoods with different median rents in selected U.S. cities. Mayer and Hauser[47] compare life expectancies by census tract in the city of Chicago in 1920-1940 and document greater life expectancy at birth in tracts with higher median rents. They compute a gap of about 10-11 years in life expectancy between the highest and lowest quantile in 1920 and a smaller gap of about 7.6 years in 1940. Yeracaris[48] conducts a similar study for the city of Buffalo in 1939-1941 and finds an identical gap of about 7.5 years in 1940. Tayback[49] follows the same approach for the city of Baltimore and finds a smaller gap (7 years for men and 4.7 for women) in 1949-1951. This evidence would seem to suggest that life expectancy gradients were as large in 1920 as they are today, and *declined* by mid-century. This would be consistent with the observed decline in income inequality.

Other similar studies - Coombs[50] studying Chicago in 1928-1932; Sheps and Watkins[51] studying New Haven in 1930-34; Patno[52] studying Pittsburgh in 1940-1950; Stockwell[53] studying Providence and Hartford in 1949-1951; and Ellis[54] studying the white population living in Houston in 1949-1951 - do not investigate life expectancy gaps but instead report greater mortality rates in low median-rent tracts (or tracts with worse overall characteristics like education, occupation or income) compared to the highest median-rent tracts.

A related set of studies compares mortality rates across larger geographic units. Altenderfer[55] compares mortality rates across 92 large U.S. cities in 1940 and finds larger age-adjusted mortality rates in cities with lower per capita income. Hamilton[56] investigates the relationship between age standardized mortality rates at the county level in North Carolina in 1950 and various measures of socio-economic status (home ownership levels, average education). He finds only weak

associations between county mortality and most economic measures, but strong relationships with measures of racial composition and hospital access.

Overall, the studies cited so far suffer from several limitations. First, none include representative samples of the U.S. population. Most look at single cities (ignoring the population living in rural areas which comprised 40% or more of the America population prior to 1960) or restrict attention to a given state. Second, these studies suffer from the aggregation or ecological fallacy: they confound the effect of individual income and average income. They can also misrepresent the shape of the individual relationship between income and health measures.[34] Lastly these studies do not observe incomes.

A different set of studies focuses on mortality rates by occupation. Britten[57] and Guralnick[58] look at the ten states in the death registration area in 1890-1900 and report higher mortality rates for men among the "laboring and servant class". Whitney[59] uses death certificate data from ten states and finds greater mortality rates for men working in unskilled occupations relative to those in professional ones. Breslow and Buell[60] focus on California in 1950 and report greater mortality rates among men in lower occupational classes relative to those at the top. These studies using occupation tend to report the largest gaps among "prime age" individuals (roughly around ages 40-50).

A major concern with these studies is that mortality rates are computed by counting deaths in a certain occupation using the information in the death certificate, and dividing by the number of people in that occupation counted in the census. Thus, occupation comes from two different sources, measured at different times, and by different people (families, coroners, or medical examines for deaths, Census respondents for population denominators). Additionally, occupation titles listed on death certificates are incomplete and poorly measured, potentially causing mortality

rates to be over-estimated for some groups and underestimated for others. Finally, these studies focus exclusively on men, and they only include a subset of states and thus are not representative of the population.

There are three pre-1960 studies based on individual data that are closest to our study. Bruno[61] uses data from 1.3 million wage earners with life insurance in 1915-1916 and shows lower mortality rates for individuals in professional or skilled occupations compared to semi-skilled or unskilled workers. Similarly, Dublin and Vane[62] used data from 3.25 million policyholders in 1922-4 and 1937-1939. They document that "industrial workers" have higher mortality rates than other workers for men though the difference falls between the 1920s and the 1930s. Finally Chapin[63] matched the 1865 Rhode Island tax records with death certificates and documented lower mortality rates for taxpayers relative to the population. The main limitation of these studies is that they are not representative of the population. Additionally, they do not measure income, though the tax payers group is presumably wealthier.

In sum early studies do not include nationally representative population, do not measure individual or household income and are often based on aggregate comparisons rather than individual data.

D. Secular trends

Studies investigating how gradients have evolved over a longer period of time 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 combine 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 1750 onwards. Here recent work shows that men in highly paid occupations did not live longer than those in low paid occupations until after WWII[64] suggesting that class gaps in lifespan are a relatively recent phenomenon, though it would appear that women from higher classes did live longer, even though their disparities were smaller in the past.

Another strand of work looking at secular trends has investigated the lifespan of famous or high ranking individuals and has compared it to that of the population. This work assumes that famous or high ranking individuals have greater access to resources than the average person in the population. For example, Clarke et al.[65] 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. Similar approaches based on very rich American adults[66] or focusing on child mortality in the 1900 Census[67] come to similar conclusions: that socio-economic gradients are primarily a post-WWII phenomenon. However the longest study on the question comparing the lifespans of aristocrats in England to those of the average population found that gradients did not exist prior to 1800 but started increasing since then.[68,69] This study also confirms that gradients have not always existed but provides conflicting evidence on the timing of their emergence.

However, these approaches have some important limitations. Occupations cannot be easily ranked over time: for example, farmers 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 at famous individuals. The differences over time in their mortality might reflect changes in who becomes famous, and moreover individuals who

live longer are more likely to become famous. The studies on aristocrats also suffer from potential biases as it is not clear that records survive of those that died young for example.

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. Of course it is possible that the emergence of the gradient varies across countries or that gradients come and go, even within countries.

E. 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 changes 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.[1] 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 can attenuate gaps or ratios across income groups.[6] As a result, some have argued that averaging income over multiple years provides a more accurate picture of the true differences across groups.[39,70] Indeed, others claim that the ideal measure of income would approximate lifetime or permanent income measures.[71] This is what more recent work such as Chetty et al.[1] attempts to do by measuring income during prime adult years. A related issue is whether comprehensive measures of income are included – most studies observe income from wages and salaries but often not income from other sources. There is also a substantive question of whether income from different sources has a different impact on health and longevity.

[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.[39]

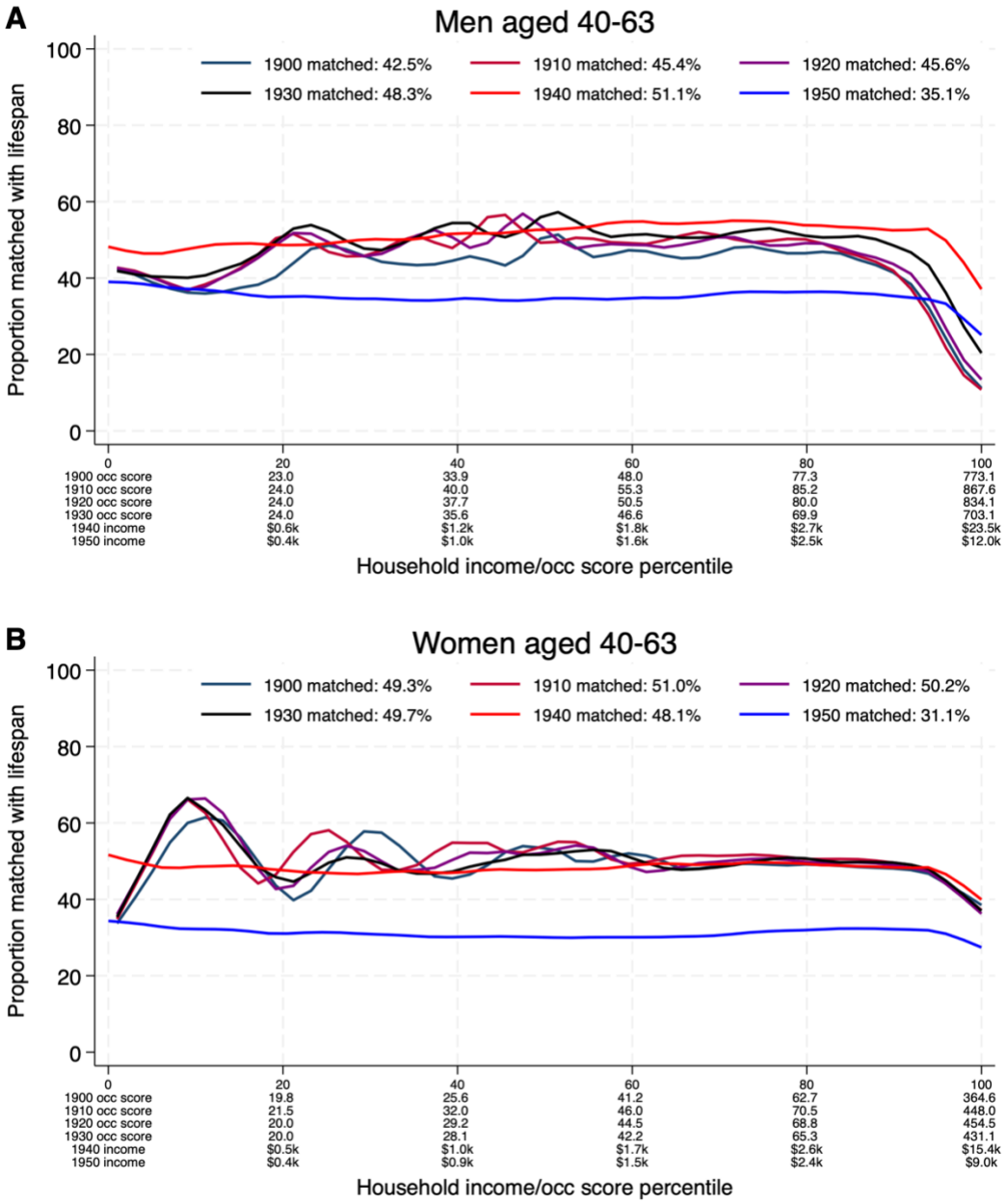
Third, the literature has found conflicting evidence on whether the effects of income are linear or whether the relationship is instead concave (or even flat after a certain threshold). But most studies investigating the relationship at the individual level tend to find some evidence that period life expectancy rises with income more sharply at the bottom[1,30] consistent with other research showing that other health measures are also more strongly associated with income at the bottom of the income distribution.[25] Importantly, however, granular studies using individual data do not find evidence of a poverty threshold effect beyond which income no longer matters: instead greater incomes are always associated with greater lifespans.[72] 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 age-specific 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 (realized) average lifespan of a cohort. When mortality rates are trending over time, cohort and period life expectancy can differ substantially. Period life expectancy has been found to better describe the experience of cohorts born 40-50 years before.[73] Moreover the two measures can give very different pictures of the gradient as noted for instance by Milligan and Schirle[43] who document income gradients in life expectancy in Canada.

Finally, a substantial number of papers have investigated why income and lifespan are related. This entails assessing 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

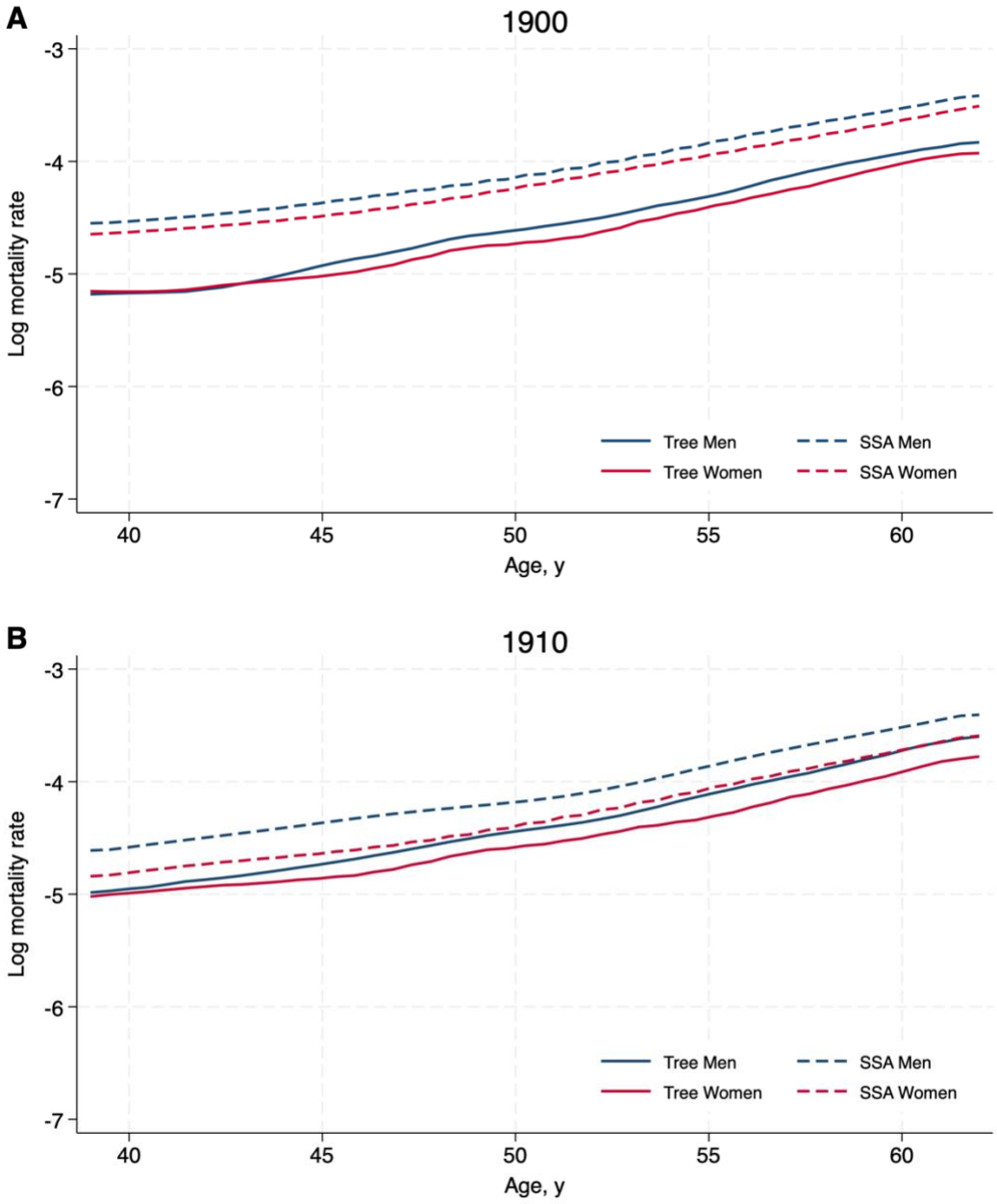
differences by other socio-economic markers (such as occupation) or health behaviors, typically finding that income remains an important determinant of mortality and lifespan even after accounting for other covariates.[32,33,70] But studies investigating causal effects have come to more mixed conclusions, 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 improving 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.[25] for a more extensive review.

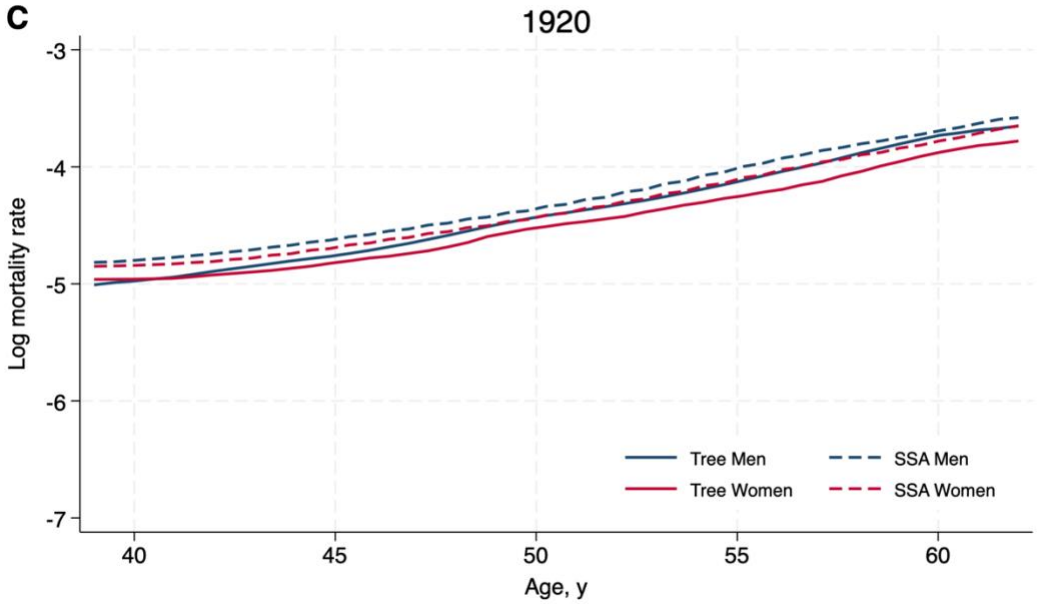
eFigure 1. Proportion Matched With Lifespan by Household Income/Occupation Score Percentile in Each Census Year

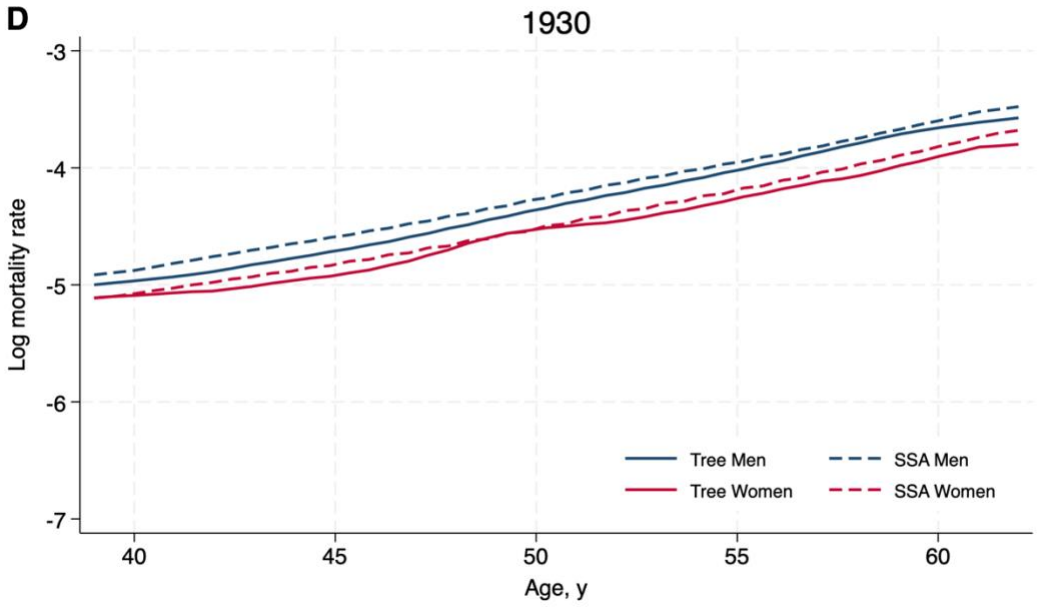


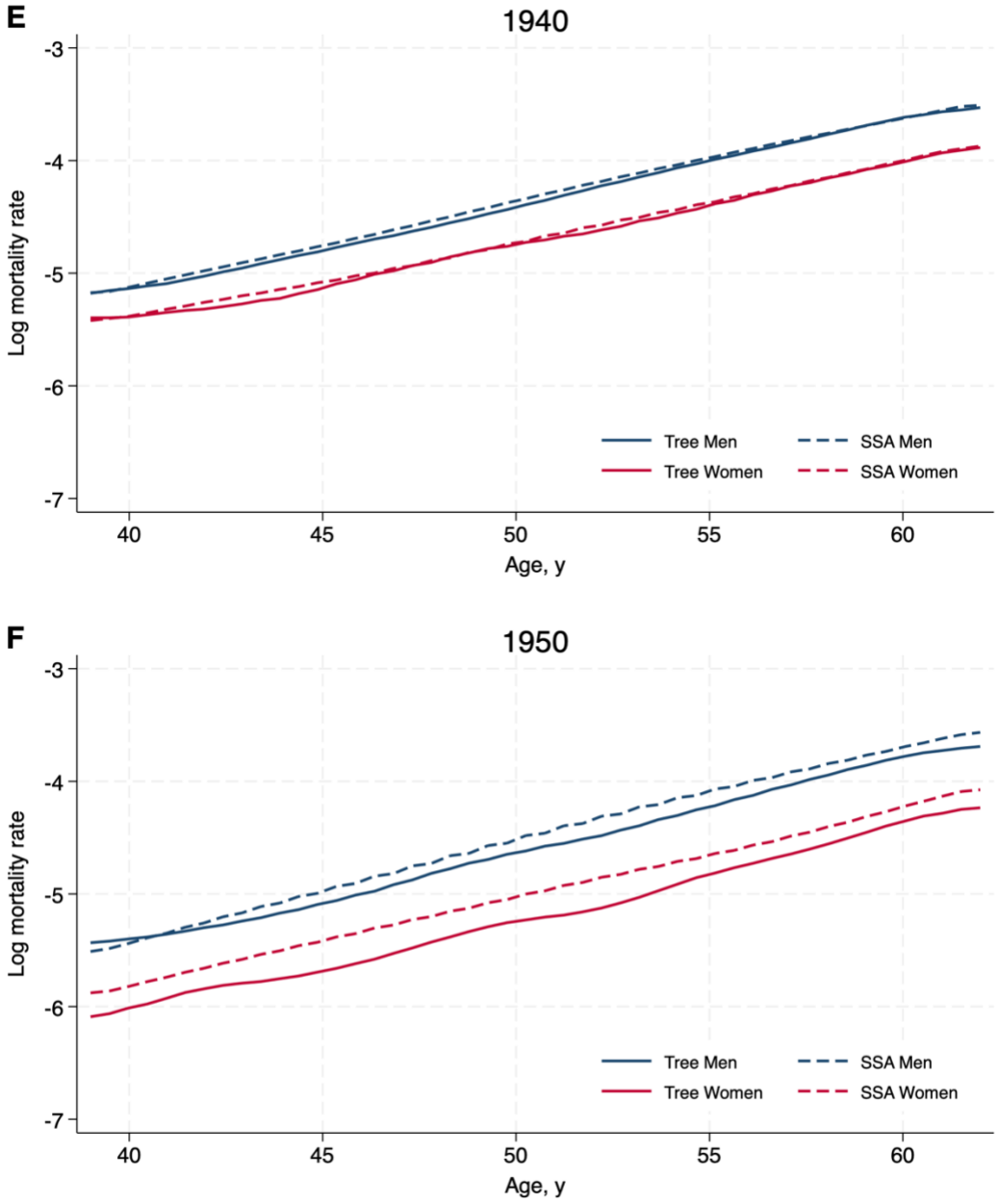
A, Match rates for men. B, Match rates for women. The sample includes individuals aged 40 to 63 in the year after each census. Percentiles are defined within the positive income distribution. For 1940 and 1950, additional data points for zero income (x-axis value = 0) are included. The solid lines are nonparametric fits. Income values are adjusted to 1939 prices.

eFigure 2. National Mortality Profile From Census Tree and SSA Life Table in Each Census Year



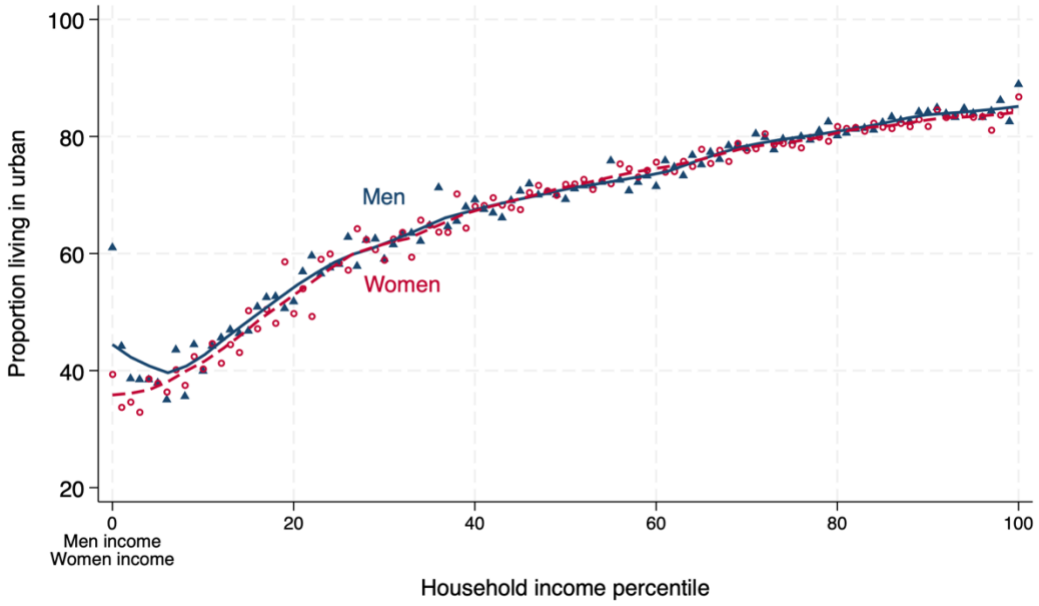

C
1920
-3
-4
-5
-6
-7
Log mortality rate
40
45
50
55
60
Age, y
Tree Men
Tree Women
SSA Men
SSA Women


D
1930
-3
-4
-5
-6
-7
Log mortality rate
40
45
50
55
60
Age, y
Tree Men
Tree Women
SSA Men
SSA Women



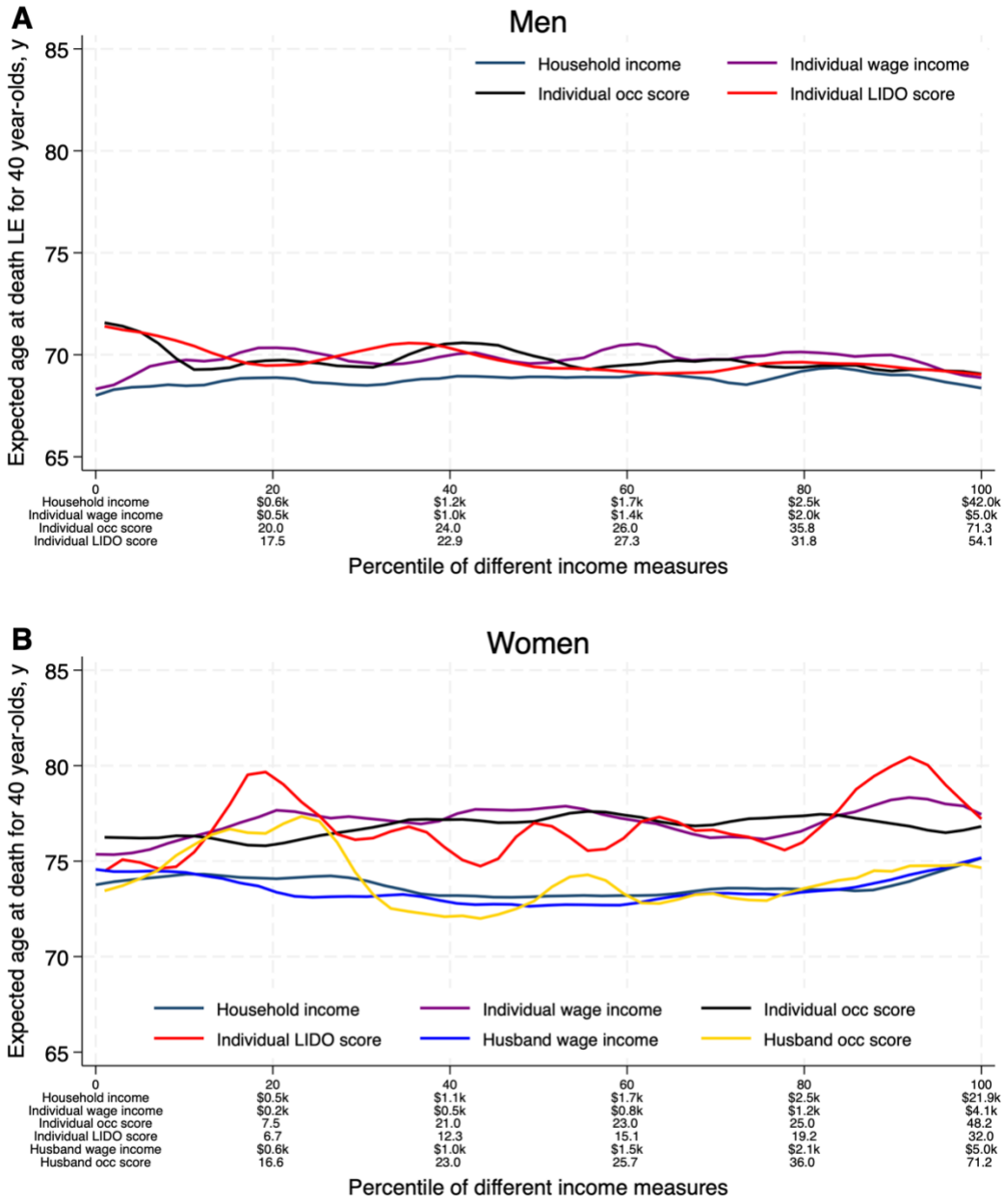
Log mortality rates are natural logarithms of observed mortality rates. Mortality profiles are calculated from census matched to family tree (solid) and SSA data (dashed) for (A) 1900, (B) 1910, (C) 1920, (D) 1930, (E) 1940, and (F) 1950. The sample used for calculating mortality from the Census Tree is restricted to linked individuals who were alive in each census year.

eFigure 3. Proportion Living in Urban Areas by Household Income Percentile and Sex in 1940



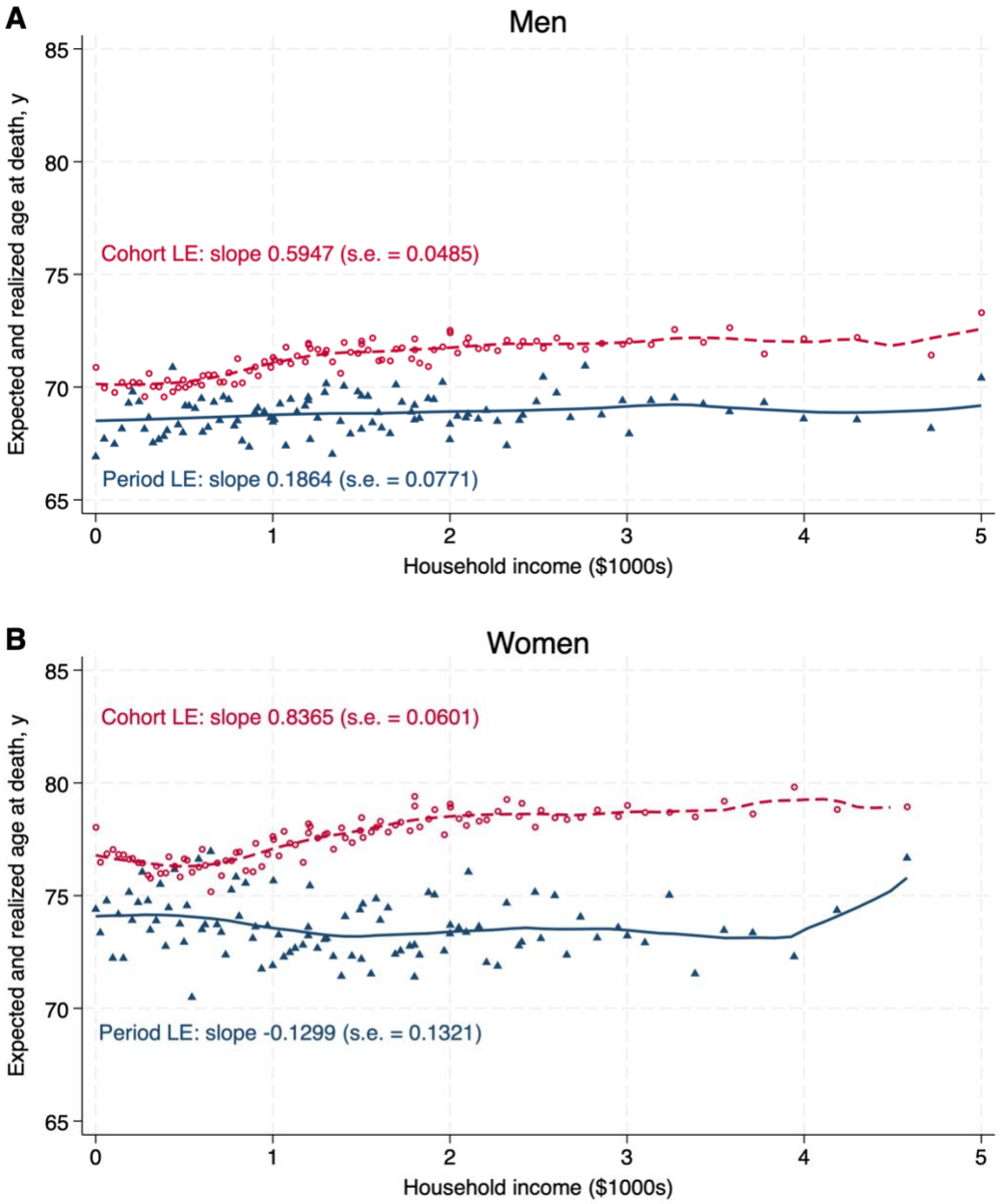
The sample includes individuals aged 40 in 1941. Percentiles are defined within the positive income distribution. Additional data points for zero income (x-axis value = 0) are included. The solid lines are nonparametric fits.

eFigure 4. Period Life Expectancy for 40-Year-Olds in 1940 by Percentile of Different Income Measures



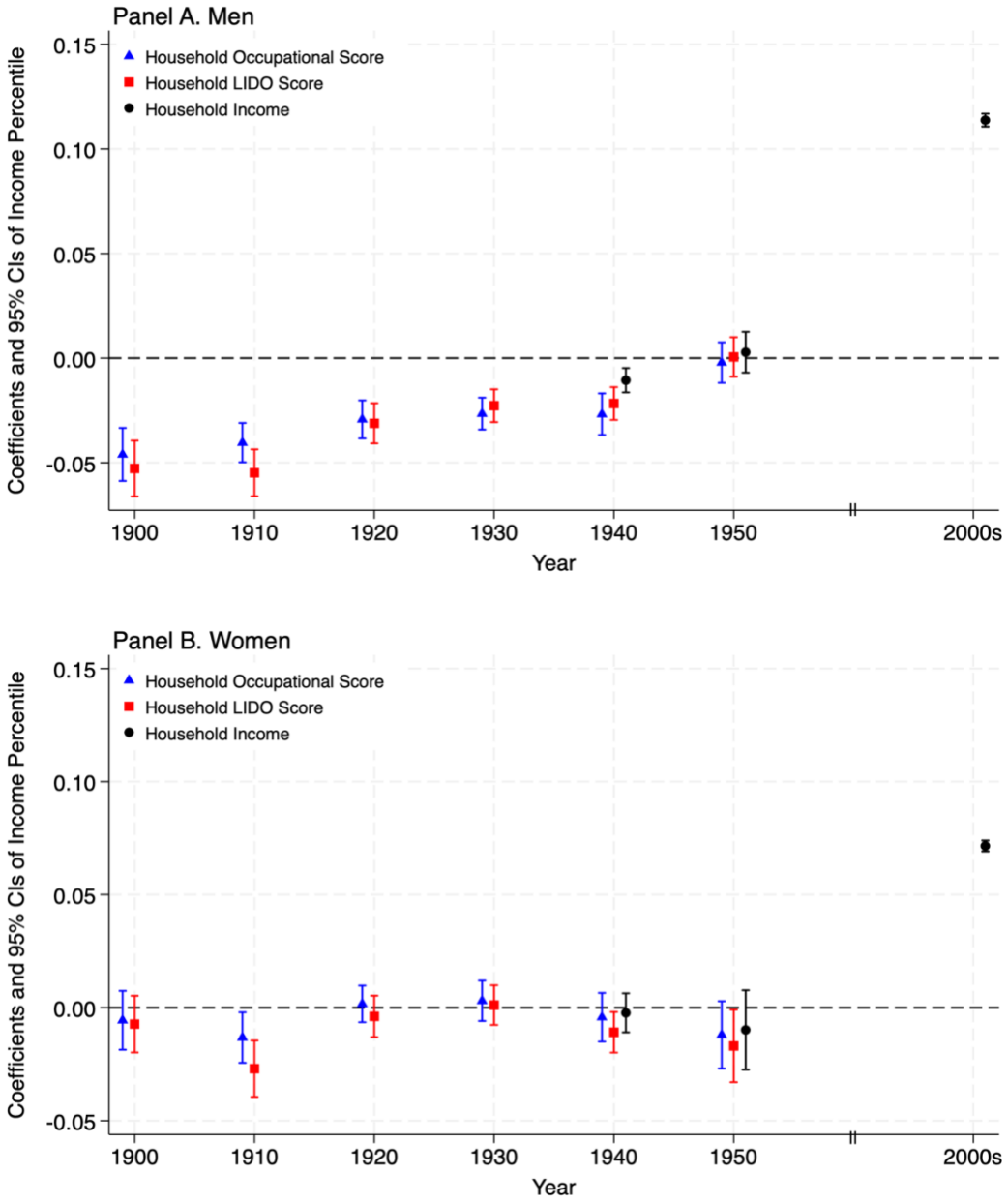
A, Period life expectancy in 1940 for men. B, Period life expectancy in 1940 for women. Estimates are based on mortality rates in 1941 and income in 1939. For household income, individual wage income, and husband's wage income, percentiles are defined within the positive income distribution, and additional data points for zero income (x-axis value = 0) are included. The solid lines are nonparametric fits.

eFigure 5. Period and Cohort Life Expectancy for 40-Year-Olds in 1940 by Household Income Levels



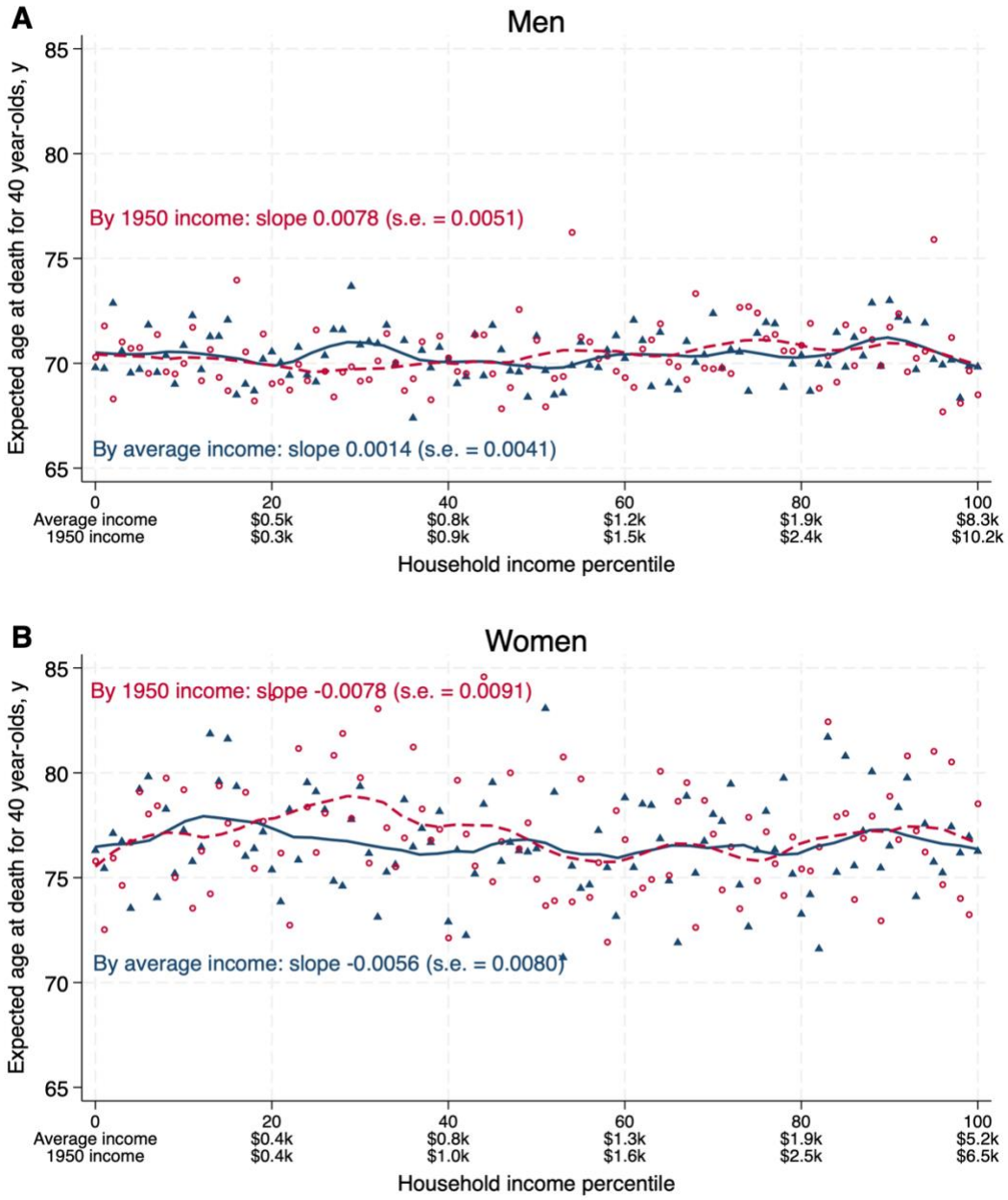
A, Period and cohort life expectancy in 1940 by income levels for men. B, Period and cohort life expectancy in 1940 by income levels for women. Estimates are based on mortality rates in 1941 (period) and lifespan conditional on being alive in 1941 (cohort), and income in 1939. Each data point represents household income at a given percentile, and the top 5% are excluded due to extreme values. The solid lines are nonparametric fits, and the slopes are estimated with ordinary least squares (OLS) regressions based on the data points shown in the figure.

eFigure 6. Estimated Period Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900 (Including Farmers)



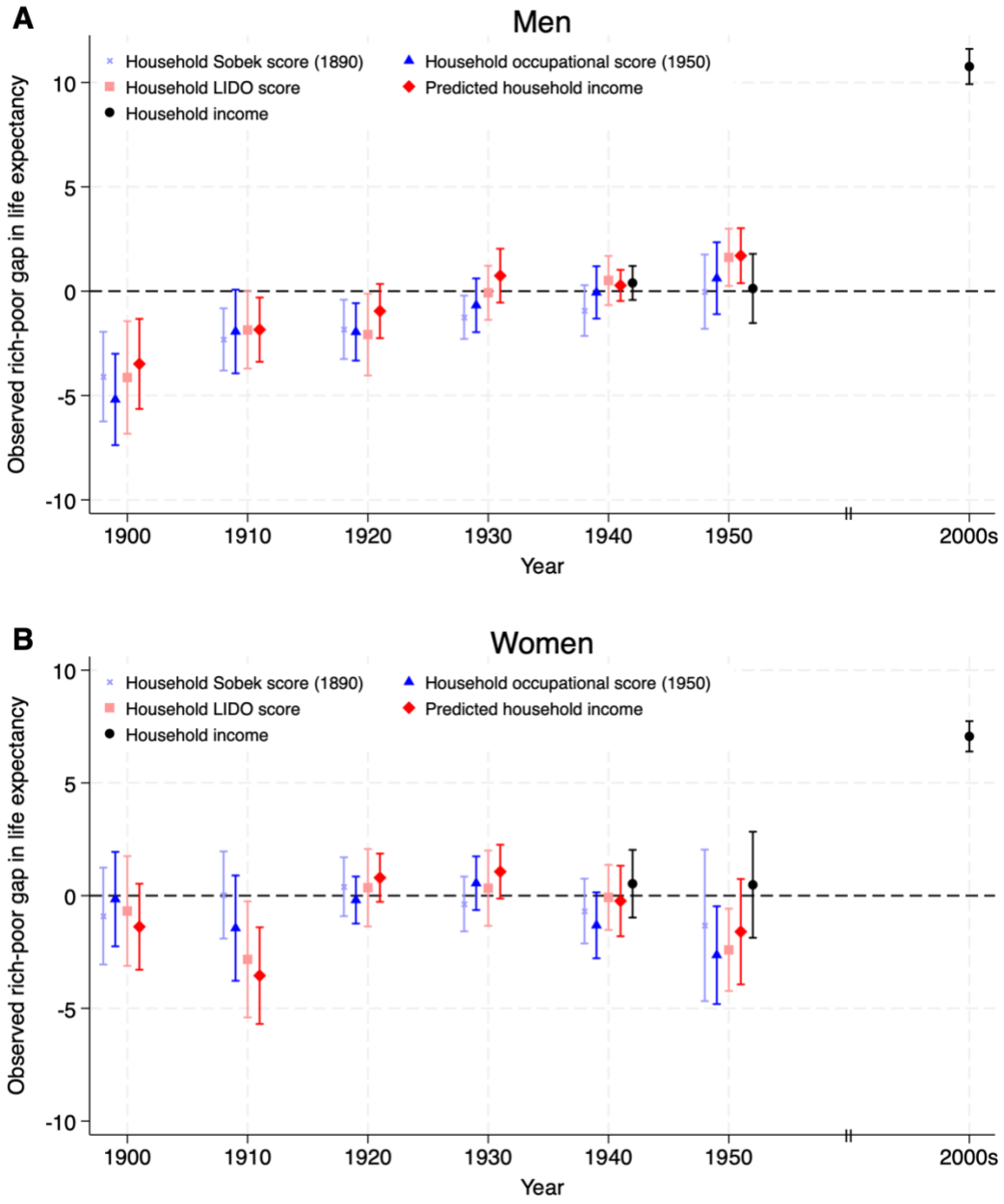
A, Estimated rich-poor gap for men. B, Estimated rich-poor gap for women. The 2000s estimates are based on data points from Chetty et al.[1] Estimates for each census year are based on mortality rates the year after the census, and income in the year before the census. For each income measure and year, period life expectancy is estimated by percentiles of the corresponding income measure for each sex. Slopes are obtained from ordinary least squares (OLS) regressions of estimated period life expectancy on income percentiles and then multiplied by 90 to give the estimated gap between the top and bottom income deciles. Data markers indicate the estimated gap; error bars indicate 95% CIs.

eFigure 7. Period Life Expectancy for 40-Year-Olds in 1950 by Household Income Percentile, 1950 Measure Versus Average 1940 and 1950 Measure



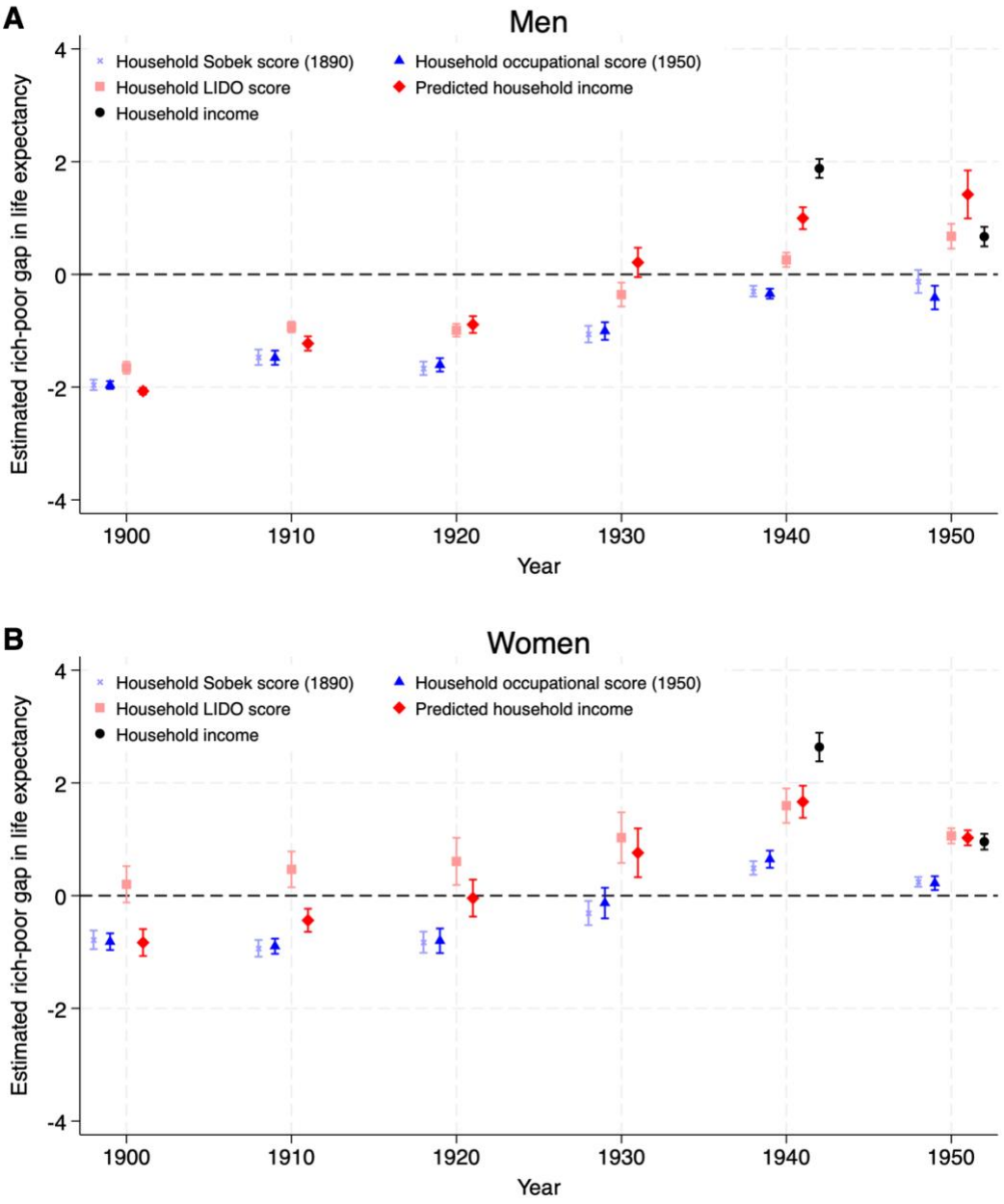
A, Period life expectancy in 1950 for men. B, Period life expectancy in 1950 for women. Estimates are based on mortality rates in 1951, and income in 1949 and the average income in 1939 and 1949 recorded in the census. Percentiles are defined within the positive income distribution. Additional data points for zero income (x-axis value = 0) are included. The solid lines are nonparametric fits, and the slopes are estimated with ordinary least squares (OLS) regressions based on the data points shown in the figure. Income values are adjusted to 1939 prices.

eFigure 8. Observed Period Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900



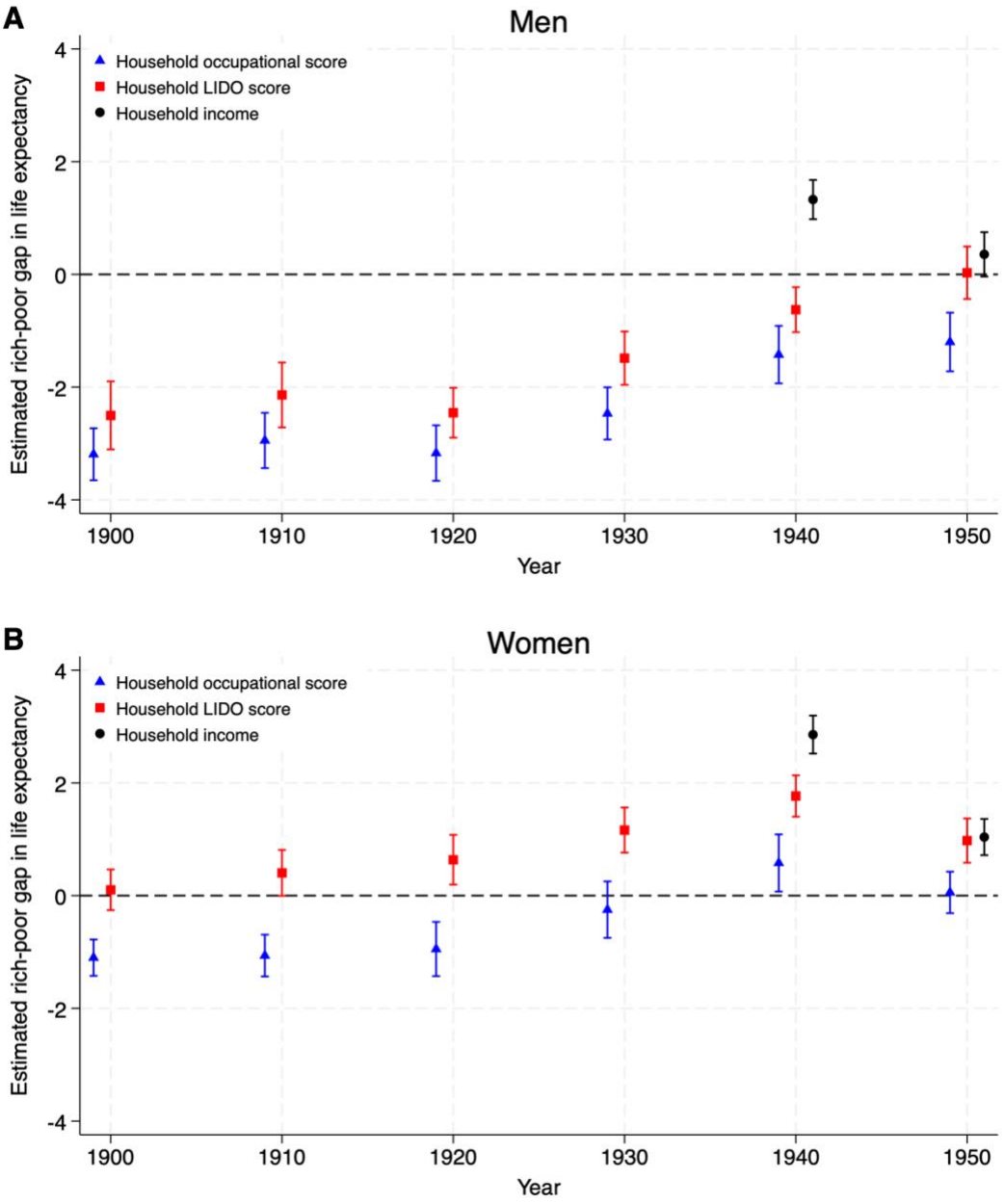
A, Observed rich-poor gap for men. B, Observed rich-poor gap for women. The 2000s estimates are based on data points from Chetty et al.[1] Estimates for each census year are based on mortality rates the year after the census, and income in the year before the census. For each income measure and year, period life expectancy is estimated by percentiles of the corresponding income measure for each sex. The observed gap is obtained by comparing the top and bottom income deciles directly.

eFigure 9. Estimated Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 35-45-Year-Olds Since 1900



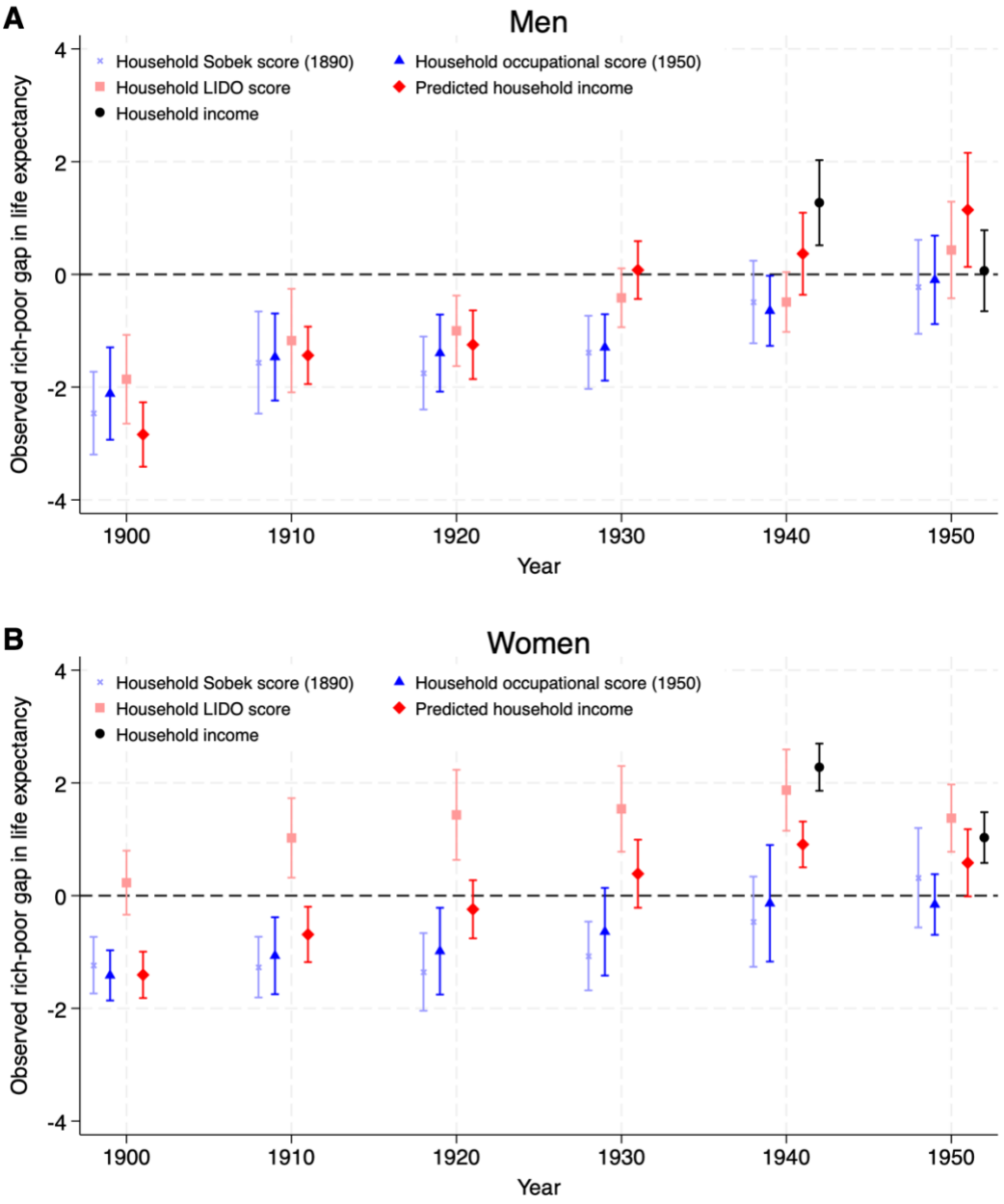
A, Estimated rich-poor gap for men. B, Estimated rich-poor gap for women. Estimates for each census year are based on lifespan conditional on being alive the year after the census, and income the year before the census. For each income measure and year, cohort life expectancy is estimated by percentiles of the corresponding income measure for each age (35-45) and sex. Slopes are obtained from ordinary least squares (OLS) regressions of estimated cohort life expectancy on income percentiles controlling for age fixed effects and then multiplied by 90 to give the estimated gap between the top and bottom income deciles. Data markers indicate the estimated gap; error bars indicate 95% CIs.

eFigure 10. Estimated Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900 (Including Farmers)



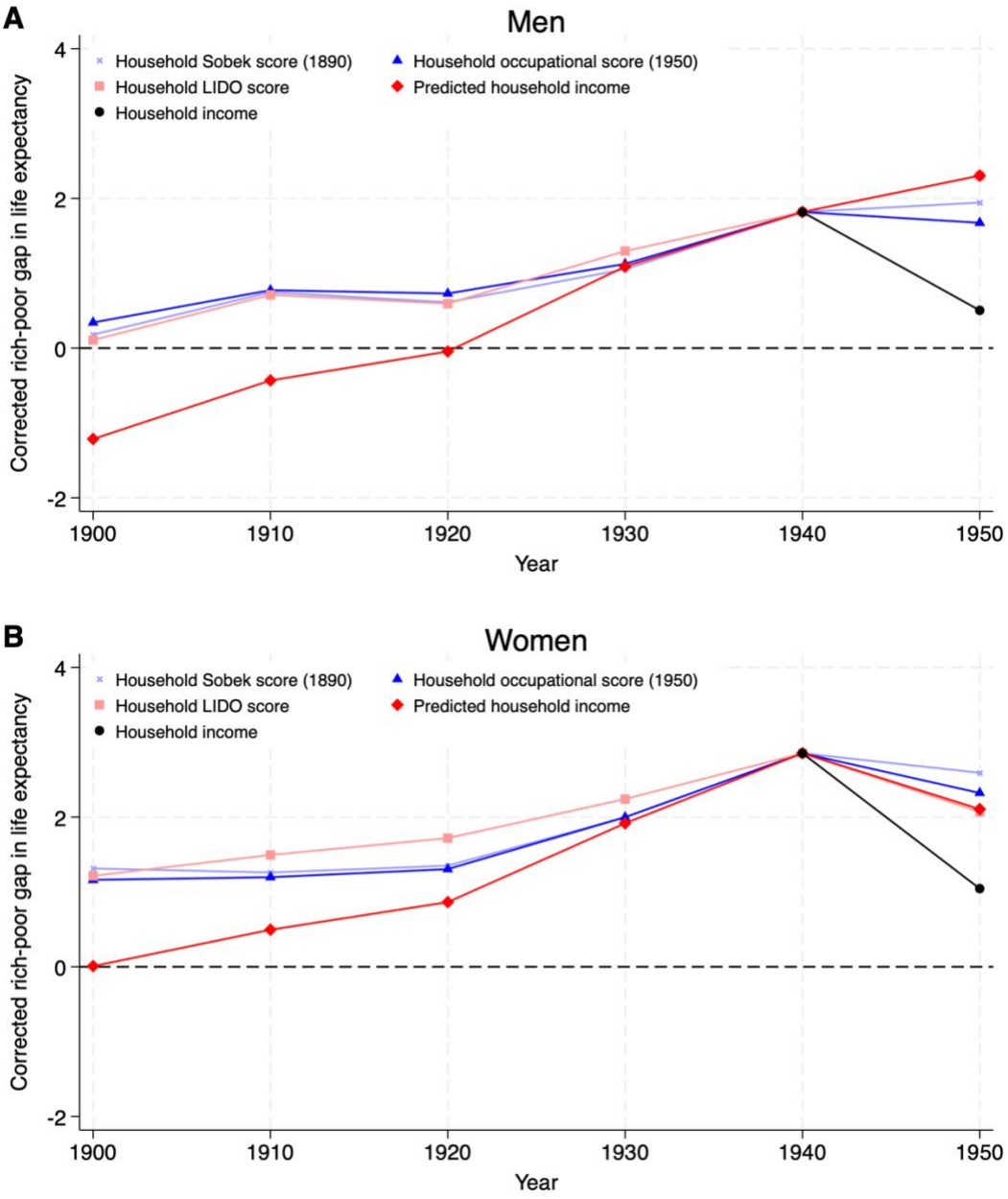
A, Estimated rich-poor gap for men. B, Estimated rich-poor gap for women. Estimates for each census year are based on lifespan conditional on being alive the year after the census, and income the year before the census. For each income measure and year, cohort life expectancy is estimated by percentiles of the corresponding income measure for each sex. Slopes are obtained from ordinary least squares (OLS) regressions of estimated cohort life expectancy on income percentiles and then multiplied by 90 to give the estimated gap between the top and bottom income deciles. Data markers indicate the estimated gap; error bars indicate 95% CIs.

eFigure 11. Observed Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900



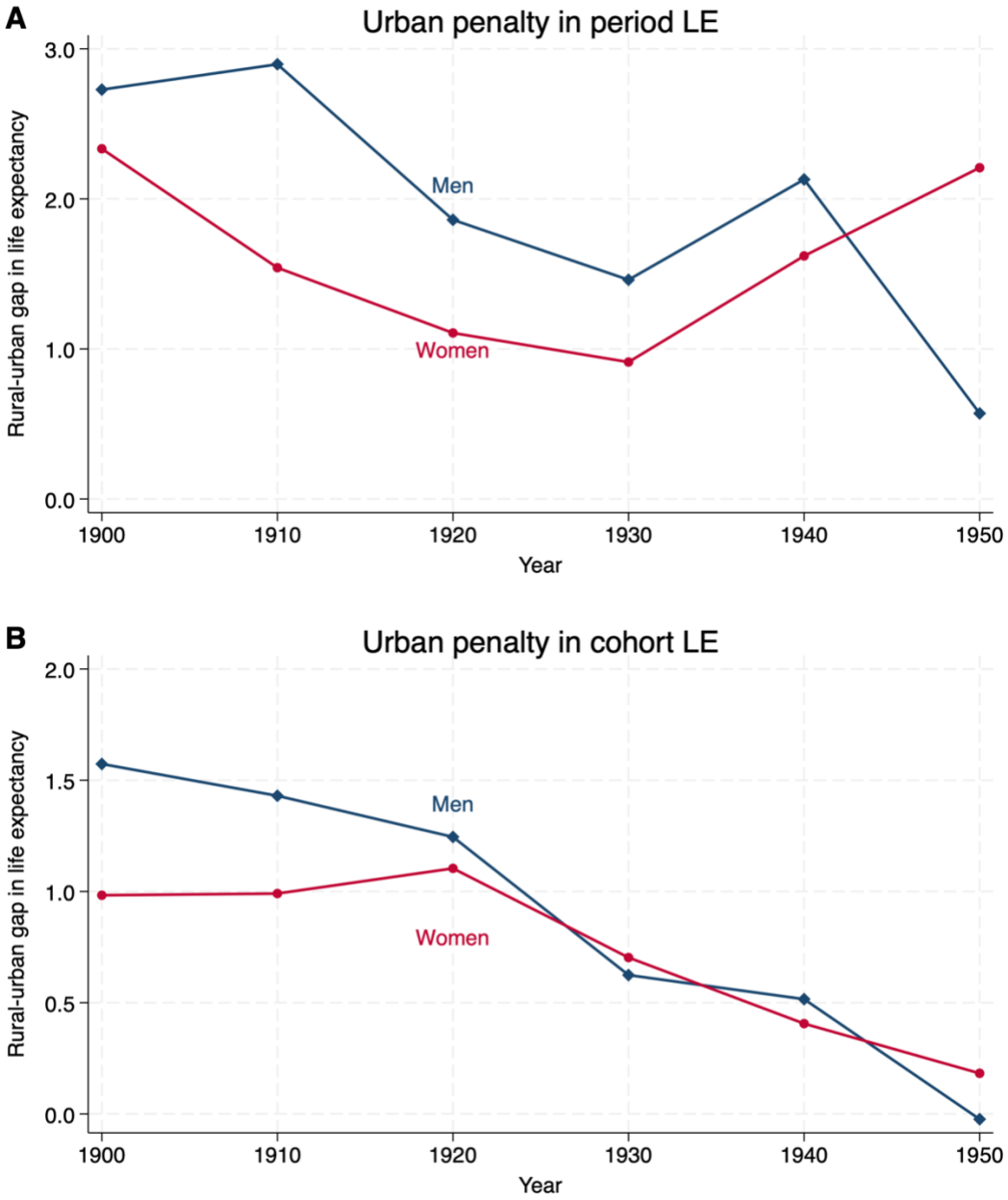
A, Observed rich-poor gap for men. B, Observed rich-poor gap for women. Estimates for each census year are based on lifespan conditional on being alive the year after the census, and income the year before the census. For each income measure and year, cohort life expectancy is estimated by percentiles of the corresponding income measure for each sex. The observed gap is obtained by comparing the top and bottom income deciles directly.

eFigure 12. Corrected Cohort Life Expectancy Gap Between Top and Bottom Income Deciles for 40-Year-Olds Since 1900



A, Corrected rich-poor gap for men. B, Corrected rich-poor gap for women. For each household income proxy and year, the corrected gap equals that proxy's unadjusted gap in the given year plus the 1940 difference between the gap based on actual household income and the gap based on the same proxy.

eFigure 13. Period and Cohort Urban Penalty in Adult Life Expectancy for 40-Year-Olds Since 1900



A, Rural-urban gap in period life expectancy. B, Rural-urban gap in cohort life expectancy. Period and cohort life expectancy are computed by urban residence and sex, and the urban penalty is defined by the gap between rural and urban estimates for each year and sex.

eTable 1. Data construction compared with Chetty et al.[1]

	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 occupation score (1900-1930) 5,240,781 (1900), 7,326,580 (1910), 9,272,209 (1920), 12,306,258 (1930), 12,649,070 (1940), 4,657,630 (1950) Sample with zero household income (1940-1950) 2,847,951 (1940), 2,450,303 (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 occupation score/Sobek score measured at a-2 y for ages 40 y to 63 y; 1900-1950: Household LIDO score/predicted household income 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 occupation score, Sobek score, or LIDO score 1900-1950: the sum of occupation score, Sobek score, or LIDO score (conditional on an occupational response) Predicted household income Gradient boosting mode: trained in 1940 on personal and household characteristics.	
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 Occupation score (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

eTable 2. Predictive validity of household income proxies (OLS regression).

	1915 IOWA Census Men	1915 IOWA Census Women	1940 Census Men	1940 Census Women
Panel A. Household Sobek score				
Household Sobek score percentiles	0.5193***	0.5753***	0.5600***	0.5398***
	(0.0117)	(0.0126)	(0.0002)	(0.0003)
R-squared	0.2342	0.2751	0.2433	0.2270
Number of observations	6,440	5,478	16,056,046	14,878,497
Panel B. Household Occupation score				
Household occupation score percentiles	0.5305***	0.6047***	0.5952***	0.5939***
	(0.0115)	(0.0122)	(0.0002)	(0.0002)
R-squared	0.2450	0.3022	0.2749	0.2755
Number of observations	6,580	5,690	16,678,523	15,578,117
Panel C. Household LIDO score				
Household LIDO score percentiles			0.6588***	0.6476***
			(0.0002)	(0.0002)
R-squared			0.3361	0.3274
Number of observations			15,768,779	14,572,256
Panel D. Predicted household income				
Predicted household income percentiles	0.5402***	0.6684***	0.7475***	0.7601***
	(0.0114)	(0.0116)	(0.0002)	(0.0002)
R-squared	0.2499	0.3503	0.4255	0.4362
Number of observations	6,711	6,181	17,071,565	16,515,884

The table regresses observed household income percentiles on percentiles based on each of four proxy measures for individuals aged 40 to 63 in the year after each census. Household LIDO score cannot be calculated for the 1915 Iowa census because industry data are not available.

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

eTable 3. Weighted 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	0.463		0.485		0.480		0.491		0.499		0.340	
Period LE at 40 - Men		70.819		69.694		69.601		68.885		68.448		70.235
Period LE at 40 - Women		74.225		73.656		72.438		73.192		73.901		76.663
Cohort LE at 40 - Men		69.848		69.413		69.520		70.377		71.182		71.891
Cohort LE at 40 - Women		72.204		72.030		73.094		75.567		77.728		79.202
Panel A. Sociodemographic												
Age	48.220	48.241	48.266	48.291	48.453	48.465	48.527	48.545	48.974	49.008	49.205	49.328
Female	0.560	0.559	0.537	0.537	0.529	0.529	0.525	0.526	0.528	0.528	0.526	0.526
White	0.908	0.911	0.914	0.915	0.917	0.917	0.918	0.918	0.918	0.918	0.904	0.913
Non-Hispanic	0.994	0.994	0.994	0.994	0.992	0.992	0.990	0.989	0.987	0.988	.	.
Marital status												
Single	0.109	0.104	0.111	0.106	0.114	0.109	0.104	0.101	0.099	0.104	0.087	0.088
Married	0.766	0.772	0.768	0.774	0.771	0.777	0.783	0.786	0.786	0.780	0.790	0.778
Separated/divorced/widowed	0.126	0.124	0.121	0.120	0.115	0.114	0.112	0.113	0.115	0.116	0.123	0.135
Urban	0.498	0.497	0.581	0.580	0.634	0.633	0.665	0.664	0.659	0.650	0.657	0.658
Years of schooling	.	.	.	.	.	.	.	.	8.143	8.132	8.173	8.191
Family size	4.558	4.598	4.332	4.378	4.139	4.171	3.995	4.010	3.715	3.671	.	.
Employment status												
Not in labor force	.	.	0.459	0.459	.	.	0.426	0.426	0.448	0.462	0.407	0.402
Unemployed	.	.	0.029	0.029	.	.	0.057	0.057	0.045	0.044	0.024	0.023
Employed	.	.	0.512	0.512	.	.	0.518	0.517	0.507	0.494	0.570	0.575
Panel B. Income measures												
Household income ($1,000s)	.	.	.	.	.	.	.	.	1.574	1.703	1.112	1.064
Household occupation score	53.537	53.013	60.921	60.570	57.819	57.897	52.358	52.582	48.943	50.950	53.883	53.246
Household Sobek score	1.208	1.206	1.378	1.380	1.292	1.298	1.193	1.198	1.105	1.146	1.197	1.179
Household LIDO score	38.431	38.308	45.023	43.969	44.472	43.933	41.683	41.513	39.372	41.044	40.368	39.032
Predicted household income ($1,000s)	1.498	1.480	1.703	1.690	1.689	1.688	1.593	1.593	1.483	1.547	1.561	1.514
Individual age income ($1,000s)	.	.	.	.	.	.	.	.	0.605	0.606	0.753	0.764
Individual occupation score	24.396	24.387	24.417	24.411	25.278	25.279	25.787	25.793	25.828	25.855	26.319	26.545
Individual LIDO score	21.492	21.619	22.054	22.184	23.089	23.228	23.655	23.769	23.288	23.433	23.759	24.001
Wage income of spouse ($1,000s)	.	.	.	.	.	.	.	.	0.546	0.568	.	.
Occupation score of spouse	.	.	.	.	.	.	.	.	25.353	25.366	.	.
Number of observations	11,312,483	5,240,781	15,121,422	7,326,580	19,309,731	9,272,209	25,086,459	12,306,258	31,068,435	15,497,021	20,919,237	7,107,933

The sample includes individuals aged 40 to 63 in the year after census with non-missing household occupation score (1900, 1910, 1920, and 1930) or household income (1940 and 1950), and excludes farmers. The working sample further restricts to those linked with lifespan. The summary statistics for the working sample are weighted using raked weights.

eTable 4. Slope estimates of life expectancy 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
A. 1940 gradient by percentiles of income measures								
Household income	0.0051	0.0028	-0.0018	0.0049	0.0202***	0.0020	0.0317***	0.0018
Household occupation score	-0.0001	0.0032	-0.0040	0.0052	-0.0054*	0.0022	0.0066*	0.0029
Household *Sobek* score	-0.0034	0.0037	-0.0029	0.0051	-0.0056*	0.0027	0.0047	0.0033
Household LIDO score	0.0001	0.0034	-0.0101*	0.0046	0.0018	0.0022	0.0195***	0.0021
Predicted household income	0.0065*	0.0032	-0.0124**	0.0046	0.0103***	0.0019	0.0195***	0.0017
Individual wage income	0.0026	0.0036	0.0120	0.0130	0.0358***	0.0012	0.0383***	0.0027
Individual occupation score	-0.0104**	0.0039	0.0105	0.0124	0.0164***	0.0020	0.0364***	0.0038
Individual LIDO score	-0.0156***	0.0040	0.0224	0.0127	0.0203***	0.0017	0.0479***	0.0038
Wage income of husband			0.0004	0.0064			0.0402***	0.0019
Occupation score of husband			-0.0145*	0.0065			0.0187***	0.0032
B. 1940 gradient by levels of income measures including top 5%								
Household income ($1,000s)	-0.0168	0.0190	0.1582**	0.0576	0.0073	0.0190	0.2223***	0.0389
Household occupation score	-0.0010	0.0010	0.0080**	0.0029	-0.0029***	0.0006	0.0024	0.0017
Household *Sobek* score	-0.0000	0.0000	0.0003*	0.0001	-0.0001***	0.0000	0.0001	0.0001
Household LIDO score	-0.0017	0.0012	0.0044	0.0038	-0.0033***	0.0007	0.0068**	0.0022
Predicted household income ($1,000s)	0.0017	0.0293	0.1638	0.0959	-0.0464*	0.0192	0.2273***	0.0480
Individual wage income ($1,000s)	-0.0335	0.1041	0.3710	0.5556	0.9698***	0.0526	1.3372***	0.1484
Individual occupation score	-0.0359**	0.0106	0.0370	0.0375	0.0437***	0.0058	0.1060***	0.0119
Individual LIDO score	-0.0551***	0.0129	0.0888	0.0597	0.0622***	0.0058	0.2206***	0.0184
Wage income of husband ($1,000s)	-0.9069*	0.4554	0.2308	0.1700	0.9259***	0.1547	0.9748***	0.0687
Occupation score of husband	-0.0011	0.0258	-0.0204	0.0178	0.0665***	0.0104	0.0511***	0.0086
C. 1940 gradient by levels of income measures excluding top 5%								
Household income ($1,000s)	0.1528	0.0814	-0.2486	0.1341	0.5941***	0.0517	0.8796***	0.0619
Household occupation score	0.0042	0.0055	-0.0107	0.0089	-0.0039	0.0033	0.0098	0.0054
Household *Sobek* score	-0.0002	0.0003	-0.0006	0.0004	-0.0002	0.0002	0.0003	0.0003
Household LIDO score	0.0080	0.0073	-0.0215	0.0111	0.0081	0.0042	0.0459***	0.0053
Predicted household income ($1,000s)	0.3762*	0.1623	-0.9808***	0.1905	0.7124***	0.0676	0.9165***	0.0840
Individual wage income ($1,000s)	0.2515	0.1460	0.6075	0.7758	1.2971***	0.0596	2.2913***	0.1632
Individual occupation score	-0.0418**	0.0134	0.0565	0.0478	0.0420***	0.0072	0.1595***	0.0120
Individual LIDO score	-0.0569***	0.0157	0.1020	0.0682	0.0661***	0.0071	0.2677***	0.0197
Wage income of husband ($1,000s)	-1.7827*	0.7070	-0.1831	0.2252	0.9865***	0.2343	1.3827***	0.0783
Occupation score of husband	0.0235	0.0324	-0.0501*	0.0218	0.0940***	0.0127	0.0514***	0.0109
D. Income gradient in different years by percentiles of income measures								
Household income								
1940 gradient	0.0051	0.0028	-0.0018	0.0049	0.0202***	0.0020	0.0317***	0.0018
1950 gradient	0.0078	0.0051	-0.0078	0.0091	0.0056*	0.0023	0.0116***	0.0018
2000s gradient	0.1137***	0.0016	0.0715***	0.0012				
Household occupation score								
1900 gradient	-0.0337***	0.0071	-0.0056	0.0071	-0.0218***	0.0023	-0.0122***	0.0018
1910 gradient	-0.0193***	0.0055	-0.0122*	0.0058	-0.0170***	0.0022	-0.0118***	0.0020
1920 gradient	-0.0083	0.0046	0.0035	0.0043	-0.0175***	0.0021	-0.0106***	0.0027
1930 gradient	-0.0063	0.0039	0.0050	0.0044	-0.0131***	0.0020	-0.0029	0.0028
1940 gradient	-0.0001	0.0032	-0.0040	0.0052	-0.0054*	0.0022	0.0066*	0.0029
1950 gradient	0.0136*	0.0053	-0.0113	0.0077	-0.0070*	0.0030	0.0007	0.0023
Household Sobek score								
1900 gradient	-0.0313***	0.0074	-0.0139*	0.0066	-0.0238***	0.0026	-0.0124***	0.0021
1910 gradient	-0.0195***	0.0053	-0.0058	0.0053	-0.0175***	0.0024	-0.0130***	0.0024
1920 gradient	-0.0131**	0.0047	0.0043	0.0044	-0.0190***	0.0022	-0.0120***	0.0032
1930 gradient	-0.0121**	0.0038	-0.0002	0.0045	-0.0141***	0.0025	-0.0048	0.0032
1940 gradient	-0.0034	0.0037	-0.0029	0.0051	-0.0056*	0.0027	0.0047	0.0033
1950 gradient	0.0082	0.0048	-0.0057	0.0089	-0.0042	0.0035	0.0018	0.0026
Household LIDO score								
1900 gradient	-0.0346***	0.0081	-0.0060	0.0069	-0.0172***	0.0028	0.0013	0.0020
1910 gradient	-0.0158**	0.0052	-0.0250***	0.0066	-0.0105***	0.0023	0.0044	0.0023
1920 gradient	-0.0108*	0.0051	-0.0002	0.0048	-0.0118***	0.0020	0.0069**	0.0025
1930 gradient	-0.0027	0.0040	0.0019	0.0048	-0.0040	0.0022	0.0127***	0.0023
1940 gradient	0.0001	0.0034	-0.0101*	0.0046	0.0018	0.0022	0.0195***	0.0021
1950 gradient	0.0183***	0.0048	-0.0156	0.0084	0.0072**	0.0027	0.0108***	0.0022
Predicted household income								
1900 gradient	-0.0301***	0.0070	-0.0143**	0.0054	-0.0234***	0.0019	-0.0121***	0.0015
1910 gradient	-0.0187***	0.0047	-0.0351***	0.0059	-0.0147***	0.0017	-0.0067***	0.0016

1920 gradient	-0.0049	0.0040	0.0030	0.0044	-0.0104***	0.0019	-0.0026	0.0016
1930 gradient	0.0087*	0.0037	0.0034	0.0042	0.0022	0.0019	0.0091***	0.0018
1940 gradient	0.0065*	0.0032	-0.0124**	0.0046	0.0103***	0.0019	0.0195***	0.0017
1950 gradient	0.0176***	0.0039	-0.0222**	0.0080	0.0157***	0.0024	0.0112***	0.0017
E. Income gradient in different years by levels of income measures including top 5%								
Household income ($1,000s)								
1940 gradient	-0.0168	0.0190	0.1582**	0.0576	0.0073	0.0190	0.2223***	0.0389
1950 gradient	0.0449	0.0957	-0.1316	0.1942	0.0273	0.0442	0.2325***	0.0396
Household occupation score								
1900 gradient	-0.0080***	0.0023	-0.0096*	0.0044	-0.0047***	0.0008	-0.0054***	0.0013
1910 gradient	-0.0030	0.0016	0.0055	0.0031	-0.0039***	0.0007	-0.0039**	0.0012
1920 gradient	-0.0041**	0.0012	0.0002	0.0026	-0.0029***	0.0007	-0.0018	0.0017
1930 gradient	-0.0044**	0.0013	0.0046	0.0027	-0.0042***	0.0007	0.0006	0.0018
1940 gradient	-0.0010	0.0010	0.0080**	0.0029	-0.0029***	0.0006	0.0024	0.0017
1950 gradient	-0.0039	0.0029	-0.0103	0.0080	-0.0072***	0.0015	0.0019	0.0023
Household Sobek score								
1900 gradient	-0.0004***	0.0001	-0.0005**	0.0002	-0.0002***	0.0000	-0.0002***	0.0001
1910 gradient	-0.0001	0.0001	0.0002	0.0001	-0.0002***	0.0000	-0.0002*	0.0001
1920 gradient	-0.0002***	0.0001	0.0000	0.0001	-0.0001***	0.0000	-0.0001	0.0001
1930 gradient	-0.0002***	0.0001	0.0001	0.0001	-0.0002***	0.0000	0.0000	0.0001
1940 gradient	-0.0000	0.0000	0.0003*	0.0001	-0.0001***	0.0000	0.0001	0.0001
1950 gradient	-0.0002	0.0001	-0.0002	0.0004	-0.0003***	0.0001	0.0001	0.0001
Household LIDO score								
1900 gradient	-0.0116***	0.0033	-0.0138*	0.0059	-0.0055***	0.0012	-0.0002	0.0017
1910 gradient	-0.0029	0.0018	0.0058	0.0051	-0.0041***	0.0008	-0.0003	0.0017
1920 gradient	-0.0047**	0.0016	-0.0030	0.0039	-0.0031***	0.0007	0.0041*	0.0021
1930 gradient	-0.0052**	0.0016	0.0054	0.0041	-0.0040***	0.0008	0.0059**	0.0021
1940 gradient	-0.0017	0.0012	0.0044	0.0038	-0.0033***	0.0007	0.0068**	0.0022
1950 gradient	-0.0043	0.0036	-0.0252*	0.0124	-0.0061**	0.0019	0.0124***	0.0033
Predicted household income ($1,000s)								
1900 gradient	-0.3345***	0.0758	-0.4016**	0.1274	-0.1596***	0.0289	-0.1885***	0.0418
1910 gradient	-0.1134*	0.0472	-0.0896	0.1416	-0.1129***	0.0183	-0.1118**	0.0337
1920 gradient	-0.1227***	0.0340	-0.0108	0.0967	-0.0666***	0.0182	0.0354	0.0352
1930 gradient	-0.0810	0.0434	0.1461	0.0961	-0.0829***	0.0199	0.1767***	0.0438
1940 gradient	0.0017	0.0293	0.1638	0.0959	-0.0464*	0.0192	0.2273***	0.0480
1950 gradient	-0.0386	0.0795	-0.7287*	0.2780	-0.0529	0.0534	0.3852***	0.0584
F. Income gradient in different years by levels of income measures excluding top 5%								
Household income ($1,000s)								
1940 gradient	0.1528	0.0814	-0.2486	0.1341	0.5941***	0.0517	0.8796***	0.0619
1950 gradient	0.2741*	0.1354	-0.2157	0.2651	0.2497***	0.0609	0.3474***	0.0531
Household occupation score								
1900 gradient	-0.0286***	0.0081	-0.0102	0.0103	-0.0200***	0.0024	-0.0159***	0.0026
1910 gradient	-0.0196***	0.0055	-0.0158*	0.0069	-0.0136***	0.0022	-0.0153***	0.0027
1920 gradient	-0.0064	0.0053	0.0041	0.0062	-0.0214***	0.0022	-0.0161***	0.0039
1930 gradient	-0.0052	0.0056	0.0047	0.0072	-0.0168***	0.0027	-0.0036	0.0046
1940 gradient	0.0042	0.0055	-0.0107	0.0089	-0.0039	0.0033	0.0098	0.0054
1950 gradient	0.0250**	0.0076	-0.0250*	0.0114	-0.0056	0.0043	-0.0013	0.0035
Household Sobek score								
1900 gradient	-0.0009*	0.0004	-0.0008*	0.0004	-0.0009***	0.0001	-0.0006***	0.0001
1910 gradient	-0.0007**	0.0002	-0.0003	0.0003	-0.0006***	0.0001	-0.0007***	0.0001
1920 gradient	-0.0003	0.0002	0.0002	0.0003	-0.0009***	0.0001	-0.0007***	0.0002
1930 gradient	-0.0004	0.0002	0.0001	0.0003	-0.0007***	0.0001	-0.0003	0.0002
1940 gradient	-0.0002	0.0003	-0.0006	0.0004	-0.0002	0.0002	0.0003	0.0003
1950 gradient	0.0009**	0.0003	-0.0001	0.0006	-0.0001	0.0002	-0.0000	0.0002
Household LIDO score								
1900 gradient	-0.0262*	0.0123	-0.0090	0.0147	-0.0192***	0.0039	0.0076	0.0043
1910 gradient	-0.0233***	0.0065	-0.0527***	0.0122	-0.0121***	0.0028	0.0095*	0.0047
1920 gradient	-0.0019	0.0072	0.0017	0.0102	-0.0177***	0.0029	0.0142*	0.0055
1930 gradient	0.0020	0.0075	0.0014	0.0109	-0.0050	0.0039	0.0285***	0.0052
1940 gradient	0.0080	0.0073	-0.0215	0.0111	0.0081	0.0042	0.0459***	0.0053
1950 gradient	0.0420***	0.0093	-0.0175	0.0189	0.0163**	0.0049	0.0260***	0.0046
Predicted household income ($1,000s)								
1900 gradient	-1.0400**	0.3067	-0.6180*	0.2604	-0.8312***	0.0727	-0.4973***	0.0773
1910 gradient	-0.6800***	0.1787	-1.5598***	0.2433	-0.5210***	0.0603	-0.2768***	0.0699
1920 gradient	-0.0026	0.1703	0.0537	0.2006	-0.4100***	0.0745	-0.2028**	0.0689
1930 gradient	0.6210***	0.1719	-0.0383	0.1947	0.2649***	0.0768	0.3631***	0.0838
1940 gradient	0.3762*	0.1623	-0.9808***	0.1905	0.7124***	0.0676	0.9165***	0.0840
1950 gradient	1.1606***	0.1672	-1.0033**	0.3681	0.9834***	0.0770	0.4663***	0.0761

The slopes are from ordinary least squares (OLS) regressions of estimated period/cohort life expectancy on the corresponding income percentiles/levels. The period life expectancy income gradient in the 2000s is calculated using data from Chetty et al.[1] The 2000s income level gradient used only five data points for 20^{th}, 40^{th}, 60^{th}, 80^{th}, and 100^{th} percentiles as income levels are not available for other percentiles.
Notes: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

eTable 5. Lifespan and income in the 1915 Iowa census.

Variable	Men	Women
Panel A. Income percentiles		
Household income percentiles	-0.0019	0.0075
	(0.0118)	(0.0081)
Number of observations	1,590	2,261
Panel B. Income percentiles with basic controls		
Household income percentiles	0.0015	0.0086
	(0.0130)	(0.0082)
Number of observations	1,588	2,261
Panel C. Income percentiles with extended controls		
Household income percentiles	0.0089	0.0042
	(0.0152)	(0.0096)
Number of observations	1,578	2,253
Panel D. Income levels		
Household income ($1,000s)	-0.0087	0.0279
	(0.3078)	(0.2069)
Number of observations	1,590	2,261
Panel E. Income levels with basic controls		
Household income ($1,000s)	0.0299	0.0363
	(0.2931)	(0.2014)
Number of observations	1,588	2,261
Panel F. Income levels with extended controls		
Household income ($1,000s)	0.1585	-0.0864
	(0.2964)	(0.2343)
Number of observations	1,578	2,253

The table reports results from regressing lifespan on household income percentiles. The sample includes individuals aged 35 to 45 and being alive in 1916 Iowa with linked lifespan. Basic controls are birth cohorts, white, and birth places (out of U.S. or specific U.S. states). Extended controls further include married, urban residence, years of schooling, home ownership, and family size. Income is adjusted to 1939 prices.
Notes: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

eTable 6. Lifespan for rich individuals in 1920 census for New York State.

	Men	Women
Panel A. Rich sample		
Share of population	0.0052	0.0052
Share of rich matched to lifespan	0.5123	0.4237
Panel B. Rich lifespan		
Lifespan of the rich	69.6881	73.2255
Number observations	1,996	1,530
Panel C. Rich – others in lifespan		
Gap in lifespan between rich and the others	0.6639	1.1193*
	(0.4541)	(0.4493)
Panel D. Estimated gap between top and bottom income deciles		
Estimated gap between top and bottom income deciles	1.1950	2.0147*
	(0.8174)	(0.8087)
Number of observations	215,445	187,284

The table reports lifespan for rich individuals aged 35 to 45 and being alive in 1923, together with the lifespan gaps relative to the rest of the population in the 1920 census for New York State. Individuals are classified as rich if they or their spouses were taxpayers in 1923. Lifespan gaps are from ordinary least squares (OLS) regressions of lifespan on a rich indicator adjusted for age fixed effects. To impute the slope of lifespan on income percentiles, lifespan is assumed to rise linearly with income percentile. We assign the rich in our sample to the 98.5^{th} percentile and assign all other sample members to the 48.5^{th} percentile (their average percentile by definition of not being in the top 4%), and calculate the slopes by dividing the gap in lifespan from panel C by 98.5-48.5 = 50. Slopes are then multiplied by 90 to give the estimated gap between the top and bottom income deciles.
Notes: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

eTable 7. Household consumption and household income (OLS regression).

Dependent variable: Expenditures on	Medical Care	Tobacco	Liquor (1917)/ Beverages including beer (1936)	Red meat
Panel A. Cost of Living Survey, 1917-1919				
Household income	0.0356***	0.0075***	0.0008***	0.0432***
	(0.0019)	(0.0004)	(0.0001)	(0.0026)
Mean dependent variable	54.2117	15.1268	1.2706	107.7511
Number of observations	12,805	12,805	12,805	12,805
Panel B. Study of Consumer Purchases, 1935-1936				
Household income	0.0247***	0.0071***	0.0002***	
	(0.0032)	(0.0010)	(0.0001)	
Mean dependent variable	79.9838	29.3649	0.3594	
Number of observations	5,321	5,321	5,321	

The table reports the results from regressing expenditures on household level income. Husband-cohort fixed effects are included in each model. Standard errors are clustered at the husband-cohort level. Income and expenditure values are adjusted to 1939 prices. The Cost of Living 1917-1919 sample has a mean household income of $1,382, and mean husband age of 37. The Study of Consumer Purchases 1935-1936 sample has a mean household income of $1,751, and mean husband age of 44. In the Cost of Living Survey alcohol consumption was asked about separately, but in the 1935-36 data the question asked about expenditures on "soft drink, beers, etc." per month.

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

eTable 8. Hazardous occupation and income (OLS regression).

	Occupational Score – 1900 Census	Occupational Score – 1910 Census	Wage and Salary Income – 1915 IOWA Census	Occupational Score – 1920 Census	Occupational Score – 1930 Census	Wage and Salary Income – 1940 Census	Wage and Salary Income – 1950 Census
Panel A. Men (reference group: low risk occupation)							
High risk	-11.5323***	-11.8101***	0.0654**	-11.5987***	-11.1490***	-0.4922***	-0.4087***
	(0.0076)	(0.0069)	(0.0245)	(0.0064)	(0.0059)	(0.0006)	(0.0015)
Extreme risk	-1.4586***	-2.9135***	-0.1154*	-2.1359***	-1.0138***	0.1553***	0.1310***
	(0.0140)	(0.0121)	(0.0479)	(0.0106)	(0.0095)	(0.0009)	(0.0020)
Share of high risk	0.5064	0.4447	0.3844	0.4018	0.3516	0.3002	0.2555
Share of extreme risk	0.0692	0.0767	0.0649	0.0833	0.0848	0.0985	0.1030
Number of observations	6,995,400	9,348,481	5,767	11,525,318	14,301,155	15,276,008	3,201,642
Panel B. Women (reference group: low risk occupation)							
High risk	-0.8875***	-0.0496*	0.2847***	-1.2867***	-3.8279***	-0.1452***	-0.4144***
	(0.0227)	(0.0207)	(0.0495)	(0.0223)	(0.0226)	(0.0014)	(0.0032)
Extreme risk	10.6613***	9.5654***	0.1912	8.5665***	8.7079***	0.1393***	-0.0518***
	(0.1142)	(0.1225)	(0.1864)	(0.1068)	(0.1168)	(0.0045)	(0.0091)
Share of high risk	0.2401	0.2137	0.1083	0.1405	0.0964	0.0526	0.0660
Share of extreme risk	0.0055	0.0033	0.0059	0.0039	0.0026	0.0041	0.0077
Number of observations	1,078,947	1,719,176	925	2,056,814	2,797,579	4,347,676	1,254,973

The sample for each census year regression includes individuals aged 40 to 63 in the year after each census. Occupations are classified as high or extreme risk based on occupation codes (occ1950). High-risk occupations include metalworking, construction trades, farming, and transportation-related jobs. Extreme-risk occupations include mining, logging, fishing, railroad work, and public safety and military roles, reflecting very high fatality risk. Each regression includes cohort fixed effects and county fixed effects. Years of schooling are additionally controlled in the 1940 and 1950 regression. Wage and salary income ($1,000s) are adjusted to 1939 prices. In 1940, about 7.8% of men have no occupation, and 71.5% of women have no occupation.

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

eTable 9. Number of children ever born and household income (OLS regression).

	1900 Census	1910 Census	1940 Census	1950 Census
Panel A. Household Sobek score				
Household Sobek score	-0.1027*	0.0203	-0.0351	-0.0426*
	(0.0336)	(0.0251)	(0.0318)	(0.0154)
Panel B. Household Occupation score				
Household occupation score	0.0002	0.0029**	0.0008	0.0001
	(0.0009)	(0.0007)	(0.0007)	(0.0004)
Panel C. Household LIDO score				
Household LIDO score	-0.0111***	-0.0077***	-0.0050**	-0.0074***
	(0.0010)	(0.0007)	(0.0011)	(0.0006)
Panel D. Predicted household income				
Predicted household income ($1,000s)	-0.1626***	-0.0516***	-0.2894***	-0.6310***
	(0.0053)	(0.0013)	(0.0192)	(0.0118)
Panel E. Household income				
Household income ($1,000s)			-0.1549***	-0.1096***
			(0.0131)	(0.0056)
Mean dependent variable	4.4848	4.5084	2.9656	2.8381
Number of observations	3,745,495	4,393,433	354,784	216,708

The universe in 1940 is limited to ever-married sample-line women. In 1950, the universe includes only ever-married 1-in-6 sample-line women. For 1900, 1910, 1940, and 1950 census year regressions, the sample includes ever-married women aged 40 to 50. Each regression is at individual level with cohort fixed effects, and the dependent variable is the number of children ever born. The 1920 and 1930 censuses did not ask about ever born children. Standard errors are clustered at the cohort level. Household income is adjusted to 1939 prices.
Notes: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

eTable 10. Summary statistics of cohorts aged 40 to 63 in 1941.

	1940			
	Census	Working sample Unweighted	Working sample IPW weights	Working sample Raked weights
	(1)	(2)	(3)	(4)
Panel A. Sociodemographic				
Age	48.974	48.759	48.977	49.008
Female	0.528	0.517	0.532	0.528
White	0.918	0.978	0.939	0.918
Non-Hispanic	0.987	0.993	0.988	0.988
Marital status				
Single	0.099	0.051	0.091	0.104
Married	0.786	0.879	0.799	0.780
Separated/divorced/widowed	0.115	0.069	0.110	0.116
Urban	0.659	0.577	0.625	0.650
Years of schooling	8.143	8.657	8.299	8.132
Family size	3.715	4.204	3.824	3.671
Employment status				
Not in labor force	0.448	0.474	0.458	0.462
Unemployed	0.045	0.035	0.040	0.044
Employed	0.507	0.491	0.502	0.494
Panel B. Income measures				
Household income ($1,000s)	1.574	1.489	1.623	1.703
Household occupation score	48.943	45.809	48.554	50.950
Household Sobek score	1.105	1.022	1.088	1.146
Household LIDO score	39.372	36.603	38.939	41.044
Predicted household income ($1,000s)	1.483	1.388	1.477	1.547
Individual wage income ($1,000s)	0.605	0.641	0.632	0.606
Individual occupation score	25.828	27.751	26.661	25.855
Individual LIDO score	23.288	25.262	24.072	23.433
Wage income of spouse ($1,000s)	0.546	0.524	0.556	0.568
Occupation score of spouse	25.353	25.450	25.471	25.366
Number of observations	31,068,435	15,497,021	15,497,021	15,497,021

The sample includes individuals aged 40 to 63 with non-missing household income, and excludes farmers. The working sample further restricts to those linked with family tree. The summary statistics for the working sample are unweighted (column 2), weighted using IPW weights (column 3), and weighted using raked weights (column 4).

eReferences

1. Chetty R, Stepner M, Abraham S, et al. The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*. 2016;315(16):1750-1766. doi:10.1001/jama.2016.4226
2. Haider S, Solon G. Life-Cycle Variation in the Association between Current and Lifetime Earnings. *American Economic Review*. 2006;96(4):1308–1320. doi:10.1257/aer.96.4.1308
3. Kopczuk W, Saez E, Song J. Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937. *The Quarterly Journal of Economics*. 2010;125(1):91-128. doi:10.1162/qjec.2010.125.1.91
4. Starr P. *The social transformation of American medicine: The rise of a sovereign profession and the making of a vast industry*. 1982.
5. Sobek M. Work, status, and income: Men in the American occupational structure since the late nineteenth century. *Social Science History*. 1996;20(2):169-207.
6. Saavedra M, Twinam T. A machine learning approach to improving occupational income scores. *Explorations in Economic History*. 2020;75doi:10.1016/j.eeh.2019.101304
7. Newson RB, Falcaro M. Robit regression in Stata. *The Stata Journal*. 2023;23(3):658-682.
8. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*. 2013;22(3):278-295.
9. Bailey MJ, Cole C, Henderson M, Massey C. How Well Do Automated Linking Methods Perform? Lessons from US Historical Data. *Journal of Economic Literature*. 2020;58(4):997–1044. doi:10.1257/jel.20191526
10. Kolenikov S. Calibrating Survey Data using Iterative Proportional Fitting (Raking). *The Stata Journal*. 2014;14(1):22-59. doi:10.1177/1536867x1401400104
11. Abramitzky R, Boustan LP, Eriksson K. A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy*. 2014;122(3):467-506. doi:10.1086/675805
12. Bailey M, Lin PZ, Mohammed ARS, et al. The creation of LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database project. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*. 2023/07/03 2023;56(3):138-159. doi:10.1080/01615440.2023.2239699
13. Lleras-Muney A, Price J, Yue D. The association between educational attainment and longevity using individual-level data from the 1940 census. *J Health Econ*. Jul 2022;84:102649. doi:10.1016/j.jhealeco.2022.102649
14. Black SE, Duzett N, Lleras-Muney A, Pope NG, Price J. Intergenerational correlations in longevity. *Conditionally accepted at the American Economic Review*. 2026;
15. Goldin C, Katz L. The 1915 Iowa State Census Project. 2010;doi:10.3886/ICPSR28501.v1
16. Marcin D. United States Individual Income Tax Payments, 1923-1924. 2017;doi:10.3886/ICPSR36640.v1
17. Sapolsky RM. *Why zebras don't get ulcers: The acclaimed guide to stress, stress-related diseases, and coping*. Holt paperbacks; 2004.
18. Snyder-Mackler N, Sanz J, Kohn JN, et al. Social status alters immune regulation and response to infection in macaques. *Science*. Nov 25 2016;354(6315):1041-1045. doi:10.1126/science.aah3580

19. Marmot M. The health gap: the challenge of an unequal world. *Lancet*. Dec 12 2015;386(10011):2442-4. doi:10.1016/s0140-6736(15)00150-6
20. Wilkinson RG. *Unhealthy societies: the afflictions of inequality*. Routledge; 2002.
21. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav*. 1995;Spec No:80-94.
22. Rosen S. *The theory of equalizing differences*. vol 1. Handbook of Labor Economics. Elsevier; 1986:641-692.
23. Grossman M. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*. 1972;80(2):223-255.
24. Becker GS, Murphy KM. A Theory of Rational Addiction. *Journal of Political Economy*. 1988;96(4):675-700. doi:10.1086/261558
25. Lleras-Muney A, Schwandt H, Wherry LR. Poverty and Health. *Annual Review of Economics*. 2025;17(Volume 17, 2025):31-56. doi:https://doi.org/10.1146/annurev-economics-081624-083218
26. Kitagawa EM, Hauser PM. *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*. Harvard University Press; 1973.
27. Duleep HO. Measuring socioeconomic mortality differentials over time. *Demography*. 1989;26(2):345-351.
28. Pappas G, Queen S, Hadden W, Fisher G. The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986. *New England Journal of Medicine*. 1993;329(2):103-109. doi:doi:10.1056/NEJM199307083290207
29. Schalick LM, Hadden WC, Pamuk E, Navarro V, Pappas G. The widening gap in death rates among income groups in the United States from 1967 to 1986. *International Journal of Health Services*. 2000;30(1):13-26.
30. Backlund E, Sorlie PD, Johnson NJ. The shape of the relationship between income and mortality in the United States: Evidence from the National Longitudinal Mortality Study. *Annals of Epidemiology*. 1996/01/01/ 1996;6(1):12-20. doi:https://doi.org/10.1016/1047-2797(95)00090-9
31. Hadley J, Osei A. 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. *Medical Care*. Sep 1982;20(9):901-14. doi:10.1097/00005650-198209000-00004
32. Lantz PM, House JS, Lepkowski JM, Williams DR, Mero RP, Chen J. Socioeconomic Factors, Health Behaviors, and Mortality: Results From a Nationally Representative Prospective Study of US Adults. *JAMA*. 1998;279(21):1703-1708. doi:10.1001/jama.279.21.1703
33. Sorlie PD, Backlund E, Keller JB. US mortality by economic, demographic, and social characteristics: the National Longitudinal Mortality Study. *American Journal of Public Health*. Jul 1995;85(7):949-56. doi:10.2105/ajph.85.7.949
34. Preston SH. The changing relation between mortality and level of economic development. *Population studies*. 1975;29(2):231-248.
35. Singh GK, Siahpush M. Widening socioeconomic inequalities in US life expectancy, 1980-2000. *International Journal of Epidemiology*. Aug 2006;35(4):969-79. doi:10.1093/ije/dyl083
36. Schwandt H, Currie J, von Wachter T, Kowarski J, Chapman D, Woolf SH. Changes in the Relationship Between Income and Life Expectancy Before and During the COVID-19 Pandemic, California, 2015-2021. *JAMA*. Jul 26 2022;328(4):360-366. doi:10.1001/jama.2022.10952

37. Couillard BK, Foote CL, Gandhi K, Meara E, Skinner J. Rising Geographic Disparities in US Mortality. *Journal of Economic Perspectives*. 2021;35(4):123–46. doi:10.1257/jep.35.4.123
38. Rogot E, Sorlie PD, Johnson NJ. Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study. *Public Health Reports*. 1992;107(4):457.
39. Cristia JP. Rising mortality and life expectancy differentials by lifetime earnings in the United States. *Journal of Health Economics*. Sep 2009;28(5):984-95. doi:10.1016/j.jhealeco.2009.06.003
40. Waldron H. Trends in mortality differentials and life expectancy for male social security-covered workers, by socioeconomic status. *Soc Secur Bull*. 2007;67(3):1-28.
41. Kinge JM, Modalsli JH, Øverland S, et al. Association of Household Income With Life Expectancy and Cause-Specific Mortality in Norway, 2005-2015. *JAMA*. May 21 2019;321(19):1916-1925. doi:10.1001/jama.2019.4329
42. Hagen J, Laun L, Lucke C, Palme M. The rising income gradient in life expectancy in Sweden over six decades. *Proceedings of the National Academy of Sciences*. Apr 8 2025;122(14):e2418145122. doi:10.1073/pnas.2418145122
43. Milligan K, Schirle T. The evolution of longevity: Evidence from Canada. *Canadian Journal of Economics*. 2021;54(1):164-192. doi:https://doi.org/10.1111/caje.12497
44. Meara ER, Richards S, Cutler DM. The gap gets bigger: changes in mortality and life expectancy, by education, 1981-2000. *Health Aff (Millwood)*. Mar-Apr 2008;27(2):350-60. doi:10.1377/hlthaff.27.2.350
45. Case A, Deaton A. Life expectancy in adulthood is falling for those without a BA degree, but as educational gaps have widened, racial gaps have narrowed. *Proceedings of the National Academy of Sciences*. Mar 16 2021;118(11)doi:10.1073/pnas.2024777118
46. Antonovsky A. Social Class, Life Expectancy and Overall Mortality. *The Milbank Memorial Fund Quarterly*. 1967;45(2):31-73. doi:10.2307/3348839
47. Mayer AJ, Hauser PM. Class Differentials in Expectation of Life at Birth. *Review of the International Statistical Institute*. 1950;18(3/4):197-200. doi:10.2307/1401037
48. Yeracaris CA. Differential Mortality, General and Cause-Specific in Buffalo, 1939–41. *Journal of the American Statistical Association*. 1955/12/01 1955;50(272):1235-1247. doi:10.1080/01621459.1955.10501302
49. Tayback M. The Relationship of Socioeconomic Status and Expectation of Life. *Baltimore Health News*. 1957;
50. Coombs LC. Economic differentials in causes of death. *Medical Care*. 1941;1(3):246-255.
51. Sheps C, Watkins JH. Mortality in the socioeconomic districts of New Haven. *Yale J Biol Med*. Oct 1947;20(1):51-80.
52. Patno ME. Mortality and economic level in an urban area. *Public Health Rep (1896)*. Sep 1960;75(9):841-51.
53. Stockwell EG. A critical examination of the relationship between socioeconomic status and mortality. *American Journal of Public Health and the Nations Health*. 1963;53(6):956-964.
54. Ellis JM. Socio-Economic Differentials in Mortality from Chronic Diseases. *Social Problems*. 1957;5(1):30-36. doi:10.2307/798946
55. Altenderfer ME. Relationship between per Capita Income and Mortality, in the Cities of 100,000 or More Population. *Public Health Reports (1896-1970)*. 1947;62(48):1681-1691. doi:10.2307/4586354

56. Hamilton CH. Ecological and social factors in mortality variation. *Eugenics Quarterly*. 1955;2(4):212-223.
57. Britten RH. Mortality Rates by Occupational Class in the United States. *Public Health Reports (1896-1970)*. 1934;49(38):1101-1111. doi:10.2307/4581310
58. Guralnick L. *Mortality by occupation level and cause of death among men 20 to 64 years of age: United States, 1950*. vol 53. US Public Health Service, National Vital Statistics Division; 1963.
59. Whitney JS. *Death rates by occupation, based on data of the US Census Bureau, 1930*. National Tuberculosis Assn.; 1934.
60. Breslow L, Buell P. Mortality from coronary heart disease and physical activity of work in California. *J Chronic Dis*. Apr 1960;11:421-44. doi:10.1016/0021-9681(60)90047-3
61. Bruno FJ. *Illness and Dependency*. vol No. 9. Miscellaneous Contributions on the Costs of Medical Care. Committee on the Cost of Medical Care; 1933.
62. Dublin LI, Vane RJ. Occupational Mortality Experience of Insured Wage Earners. *Monthly Labor Review*. 1947;64(6):1003-1018.
63. Chapin CV. Deaths among Taxpayers and Non-Taxpayers, Income Tax, Providence, 1865. *Journal of Public Health Policy*. 1924;20(2):227-234.
64. Dribe M, Eriksson B. The Long-Term Development of the Mortality Gradient: Socioeconomic Differences in Adult Life Span of Swedish Cohorts 1841–1920. *Population and Development Review*. 2025/6/4 2025;doi:10.1111/padr.70015
65. Clarke PM, Tran-Duy A, Roope LSJ, Stiles JA, Barnett AG. The comparative mortality of an elite group in the long run of history: an observational analysis of politicians from 11 countries. *European Journal of Epidemiology*. 2022/09/01 2022;37(9):891-899. doi:10.1007/s10654-022-00885-2
66. Bridgman B. A Century of Super-rich Longevity. *BEA Working Paper Series*. 2025;
67. Preston SH, Haines MR. *Fatal years: Child mortality in late nineteenth-century America*. Princeton University Press; 2014.
68. Harris B. Public Health, Nutrition, and the Decline of Mortality: The McKeown Thesis Revisited. *Social History of Medicine*. 2004;17(3):379-407. doi:10.1093/shm/17.3.379
69. Deaton A. *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton University Press; 2013.
70. McDonough P, Duncan GJ, Williams D, House J. Income dynamics and adult mortality in the United States, 1972 through 1989. *American Journal of Public Health*. 1997;87(9):1476-1483.
71. Duggan JE, Gillingham R, Greenlees JS. Mortality and Lifetime Income: Evidence from U.S. Social Security Records. *IMF Staff Papers*. 2008;55(4):566-594. doi:10.1057/imfsp.2008.21
72. Waldron H. Mortality differentials by lifetime earnings decile: implications for evaluations of proposed Social Security law changes. *Soc Secur Bull*. 2013;73(1):1-37.
73. Goldstein JR, Wachter KW. Relationships between period and cohort life expectancy: gaps and lags. *Population Studies*. Nov 2006;60(3):257-69. doi:10.1080/00324720600895876