Longevity and Occupational Choice*

Marius Guenzel[†] Ulrike Malmendier[‡] Denis Sosyura[§]

Abstract

The average adult spends nearly half of their waking hours at work. We document significant im-plications of occupational choice for individual health outcomes, which are not explained by other known determinants of socioeconomic status, such as income and wealth. Using administrative vital records for 15 percent of the U.S. population, we estimate substantial differences in lifespan associated with different occupations, after controlling for income, location, and demographic determinants. The estimated order of magnitude is large, comparable to the longevity gap between men and women. We also show that occupations associated with lower life expectancy are charac-terized by a high percentage of time working indoors (rather than outdoors), working in a sitting position, and lack of social interaction. Overall, occupational choice emerges as a key lifestyle determinant of longevity, with implications for occupation-based healthcare and retirement plans.

[†]University of Pennsylvania. mguenzel@wharton.upenn.edu.
[‡]UC Berkeley, NBER, and CEPR. ulrike@berkeley.edu.
[§]Arizona State University. denis.sosyura@asu.edu.

I. Introduction

Human life is one of the highest values in a democratic society. Thus, understanding the determinants of longevity is important for researchers, policy makers, and the general public. While many determinants of life expectancy are endowed at birth, others reflect lifetime experiences and individual choices. A large literature explores, for example, disparities in life expectancy by income, gender, race, and geographic location (see, e.g., Chetty et al. 2016, Couillard et al. 2021, Olshansky et al. 2012).

In this paper, we propose that an individual's occupation is a key predictor of their longevity, above and beyond the role of income and other known determinants. Occupational choice is a major lifestyle decision: the average person spends most of their adult life in employment and dedicates nearly half of their waking hours to work (Krueger and Mueller 2012). Job characteristics such as physical demands, social interaction or job-related stressors could thus plausibly affect long-term health and ultimately mortality. For example, Borgschulte et al. (2024) show that industry distress shocks shorten the lifespan of Fortune 500 CEOs by more than one year. Such an association between longevity and occupation would have significant implications for retirement planning, social security, and health insurance. Yet, our understanding of the empirical relationship remains limited. A major impediment is the lack of comprehensive and granular data linking individuals' career choices, other correlated observables, and later-life health outcomes.[1]

This paper uses administrative vital records for about 15% of the U.S. population across a wide

[1] Much of the existing work in economics and related fields is limited to small-scale comparisons based on survey data, limited samples comprising sub-populations, coarse occupation classifications, or single-city samples. For example, Brønnum-Hansen et al. (2020) use Danish survey data paired with register data ($N = 3,179$) to distinguish between blue- versus white-collar jobs, but do not provide an occupation-specific analysis of longevity. Paglione et al. (2020) study the relation between broadly-defined occupation categories and mortality in data from one city (Rome, Italy). In the U.S. context, Johnson et al. (1999) study mortality patterns across occupations in a national probability sample of households, but none of the sampled individuals are of retirement age. (All are younger than 65 at the time of the survey.)

range of socioeconomic groups to study the link between professional occupation and longevity. These data cover the universe of death records in several economically important states, including Florida, Connecticut, Massachusetts, and Ohio, and contain detailed personal data on millions of deceased individuals. Individual-level records include the person's dates of birth and death, their "usual" pre-retirement occupation, a granular classification of the primary and secondary death causes based on medical exams, and personal demographics such as sex, race, ethnicity, education, place of birth, and residential address. We supplement these data with information on occupational requirement and occupation-specific income profiles from the Occupational Requirements Survey (ORS), the American Time Use Survey (ATUS), and the Occupational Employment and Wage Statistics Survey (OEWS).

We organize our analysis into three parts. In the first part, we study the association between life expectancy and occupation. We map the occupations in vital records to the most granular, six-digit occupation categories of the Standard Occupational Classification (SOC) System used by U.S. government agencies including the Bureau of Labor Statistics (BLS) to classify and categorize different occupations (see, e.g., Katz et al. (2022) and Jaeger et al. (2023) for similar approaches).

We find large disparities in longevity across occupational categories, controlling for demographics and occupation-specific income profiles. In terms of economic magnitudes, the estimated occupation-based longevity differentials are comparable to the longevity gap between men and women, which is three years in our sample. These disparities and their magnitudes persist across individuals residing in the same ZIP code and subject to similar environmental factors.

We also find that occupational choice predicts variation in longevity as well as, or better than, other lifestyle characteristics. Relative to a baseline specification with solely demographic controls, adding indicators for broad occupation groups to the model increases the explained variation in longevity by 1 to 5 times more than controlling for the income profiles of broad occupation groups instead. Similarly, adding indicators for *granular* occupation groups to the baseline increases the explained variation by 1.2 times as much as when adding granular-occupation income profiles and granular place-of-residence controls (ZIP code fixed effects).

In the second part, we test whether we can identify job characteristics that help explain the occupation-dependent variation in mortality. To this end, we map the most granular SOC occupations in our data to various underlying job requirements included in the ORS and ATUS. Controlling for demographics and occupation-specific income, occupations associated with shorter lifespans are characterized by a high percentage of time spent working indoors as opposed to outdoors, long periods of working in a sitting position, a lack of social interaction, and being more stressful but less meaningful.

In the third part, we study the association between occupational choice and the primary and secondary causes of death, using the vital-records medical data based on official medical examinations. We focus on the most common death causes, such as heart disease and cancer. We find persistent variation in the causes of death across occupations, even when comparing individuals with the same lifespan. For example, outdoor occupations such as farming, fishing, and forestry predict a considerably reduced relative likelihood of mortality from heart disease. Conversely, construction labor jobs, associated with high stress in the ATUS, predict increased cancer mortality, possibly due to stress-induced behaviors like smoking and poor diet, and the resulting hormonal imbalances and inflammation conducive to cancer growth. These differential cause-of-mortality patterns suggest the occupational factors might enhance or deplete an individual's health capital.

In summary, our findings provide new evidence on the relationship between occupational choices, life expectancy, and mortality causes. Our results complement recent work on the association between U.S. life expectancy and income in the 21st century (Chetty et al. 2016), on rising geographic disparities in mortality in the U.S. (Couillard et al. 2021), on recession-induced mortality reductions (Finkelstein et al. 2024), and intergenerational transmission of lifespan in the U.S. (Black et al. 2024). Our results complement several studies on life expectancy with European data, for example, by income brackets and health conditions across managerial, intermediate, and manual occupational classes (Marmot et al. 2003), on mortality distinguishing manual and non-manual jobs (Mackenbach et al. 2008), and on mortality in working-age adults distinguishing finer occupational groups yet not accounting for correlates such as race, income, or location (Katikireddi et al. 2017). Using our rich administrative vital-records data, we uncover large life expectancy differentials across occupation groups and job requirements, which persists after controlling for income, location, and demographics.

Our findings also extend the work on inequality. Much of the inequality research has focused on absolute and relative income and wealth inequality (see, e.g., Saez and Zucman 2016, Mueller, Ouimet, and Simintzi 2017, Smith, Zidar, and Zwick 2021). Our results highlight the importance of job-related disparities not only for income but also for long-term health and lifespan. We document new dimensions of inequality in life expectancy that have implications for economic policy on retirement programs and senior healthcare.

The remainder of this article is structured as follows. Section II introduces the data. Section III describes the methodology. Section IV presents the results. Section V concludes.

II. Data

II.A. Vital Records

Our main data on individuals' lifespans, occupations, and demographics comes from the official vital records of four states: Connecticut, Florida, Massachusetts, and Ohio. Collectively, these states represent about 15% of the U.S. population. The records are collected by each state's Department of Health and cover the universe of death events in the respective state for the recorded period. They also contain dozens of variables for each death event, such as the decedent's occupation and industry, years of education, and close relatives. (For deceased individuals in retirement, the records list their "usual pre-retirement" occupation.) Vital records also provide the official cause of death, primary and secondary death factors, a distinction between natural and unnatural death events (such as accidents or homicides), and, for a subset of observations, the time interval elapsed between the primary death cause and the death event. The records also include demographic variables, such as gender, race, ethnicity, residential address, which allow us to account for local and environmental mortality factors.

Table 1 shows the sample sizes and sample periods for each state in our dataset. The sample period runs until 2020 for all states and is longest for Massachusetts, starting in 1990, and starts in 2005, 2007, and 2008 for Connecticut, Ohio, and Florida, respectively.[2] The largest fraction of death events come from Florida, followed by Massachusetts, then Ohio, and then Connecticut. In total, our dataset comprises more than four million deceased individuals in these states between 1990 and 2020.

II.B. SOC Occupation Categories

Vital records report occupations (dn sometimes industries and employer names) via free-form responses filled in by the deceased individual's surviving family members, most frequently the surviving spouse. Similar to Katz et al. (2022) and Jaeger et al. (2023), we map reported occupations to the most granular, six-digit SOC categories. We also aggregate these granular occupation groups into major occupation groups.

[2] Partial vital records data are sometimes available for additional years but without necessary information for our purposes on, e.g., decedents' occupation.

Table 1. Distribution of States

State	*N*	%	Sample Period
Connecticut	232,366	6	2005–2020
Florida	1,686,889	42	2008–2020
Massachusetts	1,196,289	30	1990–2020
Ohio	911,467	23	2007–2020

Notes: The table shows the distribution of deceased individuals in our final dataset for each of the four sample states, as well as associated sample periods.

For the mapping to SOC occupations, we use the O*NET-SOC AutoCoder software developed by R. M. Wilson Consulting for the Department of Labor. The software produces highly reliable mapping, even in the presence of abbreviations, alternative job titles, and typographical errors in the free-form vital-records responses. For example, “Elementary School Te” and “Elmntry Schl Teacher” are correctly mapped to “Elementary School Teachers, Except Special Education.” Similarly, “Office Mgr,” “Ret Clerk Typist,” “Hairdreser,” and “Babysitter” are correctly mapped to “First-Line Supervisors of Office and Administrative Support Worker,” “Word Processors and Typists,” “Hairdressers,” and “Childcare Workers,” respectively. The software also successfully handles other subtle complexities, such as mapping “Executive Chef” to “Chefs and Head Cooks” rather than “Executives.” (See Table OA.2 in the Online Appendix for these and other examples of reported and mapped occupations.)

Table 2 shows the resulting distribution of the 22 major SOC occupation groups in our data, along with sample free-form descriptions associated with them as reported in the vital records. Office and Administration Support workers constitute the largest group in our sample (14%), frequently corresponding to Secretary, Clerk, or Bookkeeper in the vital records. Other common occupations are Construction (e.g., Laborers, Carpenters, Electricians) and Production (e. g., Supervisors, Machinists, Seamstresses), each making up 11$ of the sample. Infrequent occupation classes include Farming, Fishing, and Forestry as well as Legal occupations.

Table 2. Descriptive Statistics on Occupations

Occupation Category	Sample Occupations	*N*	%
Architecture & Engineering	Engineer, Electrical Engineer, Draftsman	148,175	4
Arts, Entertain, & Media	Artist, Designer, Photographer	70,865	2
Building Cleaning & Maintenance	Custodian, Housekeeper, Landscaper	112,281	3
Business & Fin. Operations	Accountant, Inspector, Banker	166,631	4
Community & Social Service	Social Worker, Minister, Counselor	62,358	2
Computer & Math	Administrator, Computer Programmer, Systems Analyst	64,745	2
Construction & Extraction	Laborer, Carpenter, Electrician	425,564	11
Educational Instruction & Library	Teacher, Professor, Librarian	212,062	5
Farming, Fishing, & Forestry	Logger, Farming, Citrus Worker	16,552	0
Food Preparation & Serving	Cook, Waitress, Chef	153,844	4
Healthcare Pract. & Technical	Registered Nurse, Technician, LPN	208,097	5
Healthcare Support	Nurses Aide, Certified Nursing Assistant, Medical Assistant	65,402	2
Installation, Maintenance, & Repair	Mechanic, Maintenance, Handyman	204,764	5
Life, Physical, & Social Science	Chemist, Quality Control, Psychologist	34,700	1
Legal	Attorney, Lawyer, Paralegal	32,043	1
Management	Manager, Vice President, Executive	349,239	9
Office & Admin. Support	Secretary, Clerk, Bookkeeper	555,498	14
Personal Care & Service	Hairdresser, Beautician, Barber	63,863	2
Production	Supervisor, Machinist, Seamstress	446,904	11
Protective Services	Police Officer, Security Guard, Firefighter	116,470	3
Sales	Sales, Cashier, Realtor	279,859	7
Transportation & Material Moving	Truck Driver, Packer, Pilot	237,095	6
	Total:	4,027,011	100

Notes: The figure shows the distribution of occupations in our final dataset. Occupation categories are based on the 2018 Standard Occupational Classification (SOC) System. Occupations reported in the vital records are mapped to SOC categories using the O*NET-SOC AutoCoder software developed by R. M. Wilson Consulting for the Department of Labor.

II.C. Occupational Requirements

We map the SOC occupations to various associated occupational requirements using data from the 2022 ORS and the 2012 and 2013 ATUS, which included separate modules on the subjective, job-related well-being. Specifically, we map occupations to active versus sedentary requirements (ORS), outdoor versus indoor requirements (ORS), physical and mental job burdens (ATUS), job meaningfulness (ATUS), and the degree of social interactions involved in performing the job (ORS).

II.D. Occupation-Specific Income Profiles

We use the May 2021 OEWS for information on income profiles at the most granular, six-digit SOC occupation level. For each of these SOC occupation group, we retain the average annual income, as well as the 10th, 25th, and 50th percentile incomes, which are the three available percentiles with the least amount of missing data. We also use information from the May 2022 OEWS to obtain Metropolitan Statistical Area (MSA)-specific income profiles for each granular SOC occupation group.

II.E. Underlying Cause of Death

We also use the detailed vital-records information to categorize causes of death. The vital records data provides the official, medically assessed cause of death in the form of International Statistical Classification of Diseases and Related Health Problems (ICD) codes. We map the detailed ICD codes into broader categories using the linking tables provided by the Centers for Disease Control and Prevention (CDC).[3] The most common causes of death are heart disease and cancer, mirroring representative US data.[4]

III. Methods

We estimate the empirical association between longevity and occupation using the following model:

$$AgeAtDeath_i = \alpha + \beta' \mathbf{Occ_i} + \gamma' \mathbf{X_i} + \varepsilon_i, \qquad (1)$$

[3] See https://ftp.cdc.gov/pub/health_statistics/nchs/datalinkage/linked_mortality/.
[4] See https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm.

where $AgeAtDeath_i$ is computed as the difference between the exact dates of birth and death in the vital records, $\mathbf{Occ_i}$ is a vector of major SOC occupation classes, and $\mathbf{X_i}$ is a vector of demographics. In the baseline estimation, we include sex, race, ethnicity, and six-digit SOC group-level income profiles (mean, $p10$, $p25$, and $p50$). In additional specifications, we also include residence ZIP-code fixed effects as well as area-specific or more flexible income controls, as discussed in more detail in Section IV.A. We restrict the estimation to individuals passing at or after age 40, to focus on those who have had the opportunity to accumulate sufficient education to be qualified for various professions.

Our analysis also accounts for time trends in the relative frequencies of occupations. For example, the share of jobs in the construction sector is increasing over our sample period, whereas the share of jobs in the production sector is decreasing. Ignoring these trends in the estimation would conflate the results and underestimate longevity for occupation groups that have grown in importance over time (as we will observe disproportionately many young deceased individuals in the sample), and vice versa for occupation groups that have shrunk in importance over time.

We use weighted least squares to correct for these distortions. Specifically, our methodology assigns a larger weight to observations when the occupation class is "underrepresented" for a given birth year relative to the same birth cohort in the full-sample distribution, and vice versa when it is "overrepresented." We present a detailed discussion of this procedure in Section B.2 of the Online Appendix. Online Appendix Figure OA.1 shows the unweighted and weighted frequency distributions over time for selected occupation classes. The unweighted distributions in Panels b, c, and d (in orange) reveal time trends in occupation frequencies. Weighting observations removes these time trends, resulting in identical weighted frequency distributions (blue) that correspond to the full-sample distribution from Figure OA.1a.

We proceed analogously to estimate the association between longevity and occupational requirements (*OccReq*), modifying the estimating equation as follows:

$$AgeAtDeath_i = \alpha + \beta OccReq_i^{type} + \gamma' \mathbf{X_i} + \varepsilon_i, \tag{2}$$

where $type \in \{sedentary, indoor, stress, meaning, social\}$. As in estimating model (1) we include individuals aged 40 and older at their passing, and weight observations as described in Online-Appendix Section B.2.

Finally, we also use the same methodology to relate occupational choice to an individual's cause of death:

$$\mathbf{1}_{cause=c} = \alpha + \beta' \mathbf{Occ_i} + \gamma' \mathbf{X_i} + \varepsilon_i. \tag{3}$$

We focus on the relationship with cancer and heart attack, $c \in \{cancer, heart\ attack\}$, as they are the leading causes of death in the U.S. When estimating model (3), we also include age-at-death fixed effects (rounded down to integer values), to analyze the relation between underlying cause of death and occupation among individuals of the same age.

IV. Results

IV.A. Baseline Association Between Longevity and Occupation

Figure 1 shows the baseline association between longevity and SOC occupational class, controlling for the decedent's demographics (sex, race, and ethnicity) and the six-digit SOC group-level income profiles. (Online-Appendix Table OA.3 presents the corresponding regression estimates.) The figure plots longevity differentials across occupations in years, with the lowest-longevity group omitted from the estimation and included at the top. Average life expectancy in that group, conditional on not passing before age 40, is indicated at the bottom left of the figure (mean and median).[5] Average life expectancy in other SOC classes thus corresponds to the lowest-longevity number plus the relevant coefficient estimate, as shown in the figure.

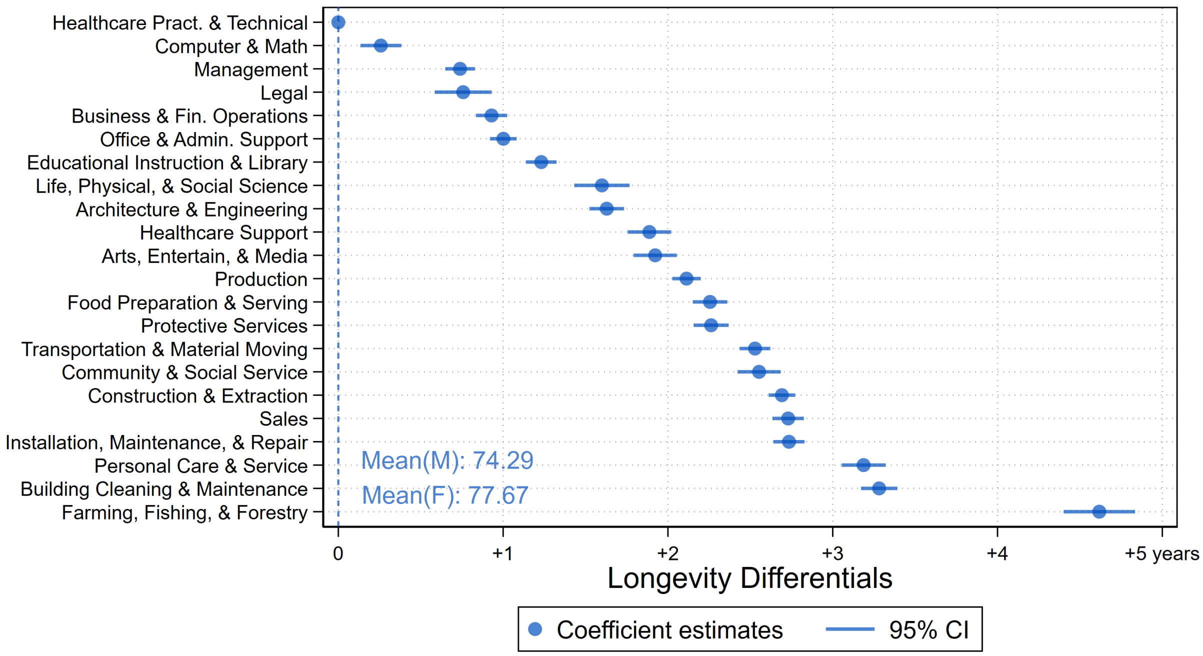


Figure 1. Baseline Estimates of Longevity Differentials by Occupation

Notes: The figure shows differences in longevity across occupation classes, defined as the major occupation groups based on the 2018 Standard Occupational Classification (SOC) system. Estimates based on administrative vital records data from CT, FL, MA, and OH. $N = 4,027,011$. Estimation controls for decedents' sex, race, ethnicity, and six-digit SOC group-level income. See Appendix A for variable definitions.

The figure reveals large occupation-based disparities in life expectancy, even after correlates are accounted for. The estimated order of magnitude associated with occupation-based longevity differentials is comparable to the roughly three-year gender longevity gap in the data. The occupation-based longevity gaps are as large 4.5 years in the case of highly outdoorsy farming,

[5] We estimate the longevity statistics for men and women using a simplified model (1) with an indicator for decedent sex as the sole control variable. We then average over all men and women in the respective group.

fishing, and forestry occupations, which show a multi-year residual longevity advantage against all other broad occupation groups.

Figure 2 presents additional estimates of the association between longevity and occupation, using different variants of the set of included income-related control variables. In varying shades of blue, the figure compares the coefficient estimates from the baseline specification in Figure 1 with specifications that include ZIP code fixed effects (accounting for environmental factors at the place of residence), area-specific income controls (allowing the occupation-specific income distributions to vary across MSAs), and piecewise linear income controls (allowing the slope of the occupation-specific income controls to vary across terciles of each income profile component included (mean, $p10$, $p25$, $p50$)), respectively. Additionally, the figure shows, in green, the estimated longevity differentials in the absence of any income controls.

The various alternative specifications in Figure 2 yield two insights. First, as indicated by the largely overlapping blue-colored coefficient plots, the precise method of controlling for income is of secondary importance, both in terms of the economic magnitudes of the estimated longevity differentials across occupation groups and in terms of the relative longevity ranking of occupations. Second, as indicated by the differing patterns in the green coefficient plot, excluding income controls altogether does make a significant difference. In the absence of any income controls, the longevity ranking of occupations changes considerably.

Zooming in on specific occupation groups with respect to the above two insights can also help guide our tests of possible underlying mechanisms in Section IV.C. For example, management occupations are positioned high in the pre-income-controls ranking. However, once the high pay associated with managerial positions is accounted for, management occupations fare significantly worse in the ranking, possibly reflecting that, despite being well-paid, these positions also come

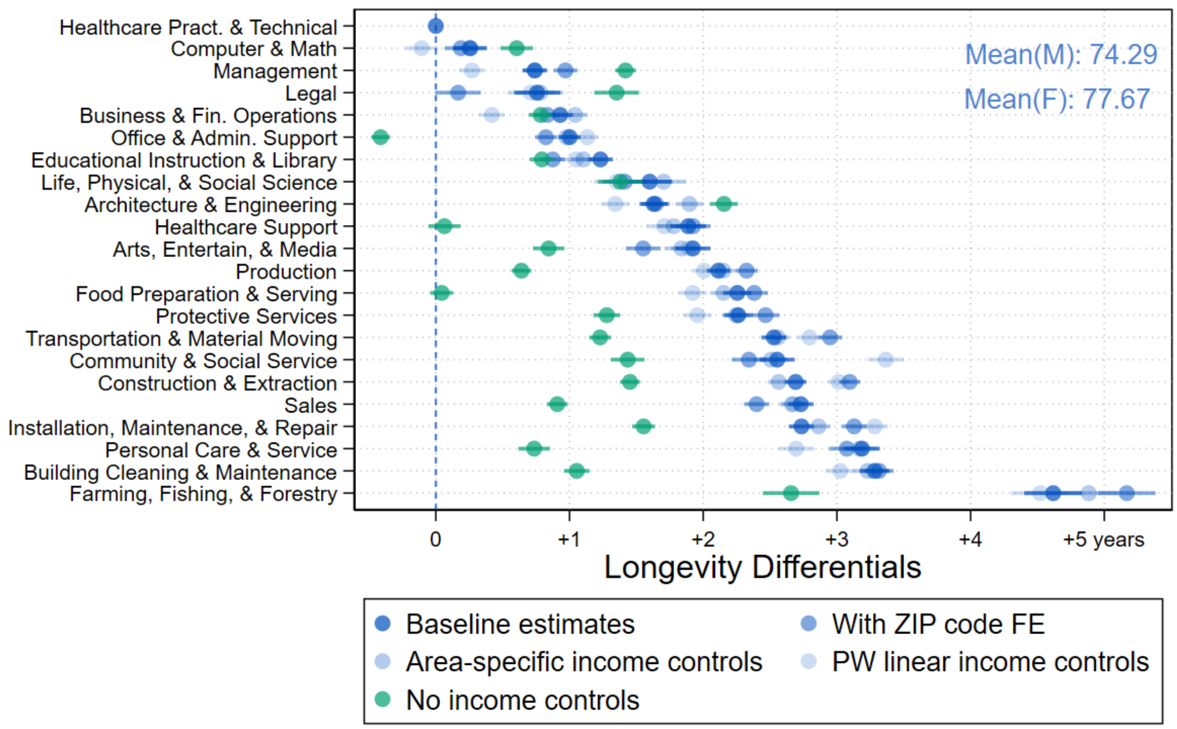


Figure 2. Alternative Estimates of Longevity Differentials by Occupation

Notes: The figure shows differences in longevity across occupation classes, defined as the major occupation groups based on the 2018 Standard Occupational Classification (SOC) system. Estimates based on administrative vital records data from CT, FL, MA, and OH. $N = 4,027,011$. Estimation controls for decedents' sex, race, ethnicity, and includes different variants of income controls as indicated in the figure legend. See Appendix A for variable definitions.

with heightened job stress and responsibilities (Borgschulte et al. 2024). Conversely, community and social service occupations benefit from the income controls, possibly reflecting that while these occupations are less highly paid, they are associated with other desirable characteristics, such as job meaningfulness.

IV.B. Explanatory Power of Occupation Versus Income in Predicting Longevity

We next assess the explanatory power of occupational choice, relative to other known determinants of longevity. Figure 3 shows the baseline and incremental adjusted R^2 for various specifications. A baseline specification with solely demographic controls (sex, race, ethnicity) explains 2.47% of the variation in longevity. The bars in the figure visualize the incremental adjusted R^2 relative to this baseline when augmenting the specification with income or occupation-based characteristics, or both. The first two bars, in gray and light blue, show that adding indicators for the 22 broad SOC occupation classes increases the R^2 by as much as adding the more granular income profile controls for the 794 detailed SOC occupation groups. In that sense, occupation is as good of a predictor of longevity as income. The third bar, in dark blue, shows that the R^2 increases further when adding both occupation and income characteristics jointly. In other words, occupation and income are complements rather than substitutes in predicting longevity, indicating that the nature of a person's job encompasses more than just their earnings.

The final three bars of Figure 3 repeat the exercise, allowing both the occupation indicators and the income controls to vary by MSA. These specifications account for the possibility that even for the same job, both income and other job characteristics may significantly vary across regions—for example, . being realtor in rural Ohio may be quite different from being a realtor in Miami. Allowing for area-specific heterogeneity, occupational choice emerges as a much stronger predictor of longevity than income profile. Occupation-by-area indicators increase the R^2 by five times as much as area-specific controls for occupation-group income profiles. Additionally, as before, adding both area-specific occupation and area-specific income characteristics leads to the largest increase in R^2.

Building on the evidence on broad occupational indicators in Figure 3, Figure 4 compares the predictive power of occupational choice to that of other characteristics, now using indicators for granular occupation groups. Relative to the baseline, detailed occupation increases the R^2 by 1.2 more than a specification that augments the baseline with detailed occupation-group income

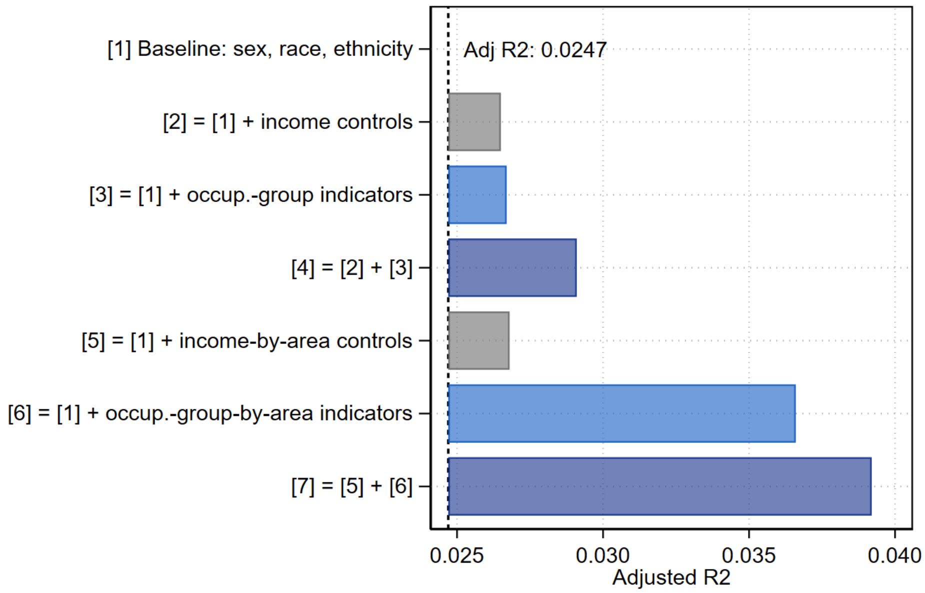


Figure 3. Explained Variation in Longevity

Notes: This figure shows the baseline and incremental adjusted R^2 for various specifications as indicated in the figure descriptions. Income controls are controls for the income profile (mean, $p10$, $p25$, $p50$) associated with the 794 detailed, six-digit SOC occupation groups. Occupation group indicators are indicators for the 22 broad SOC occupation classes. Area refers to the 47 MSA in our sample.

controls and 13,413 ZIP code fixed effects. These results further underscore the significant role of occupation in predicting longevity relative to other observable characteristics.

IV.C. Longevity and Occupational Requirements

This section tests whether we can identify specific job requirements that help explain the occupation-dependent variation in longevity. Reflecting the evidence and related discussion in Section IV.A, we examine a job's outdoor versus indoor requirements, job stress, and job meaningfulness as possible undelying mechanisms. We additionally examine sedentary requirements as well as on-the-job social requirements, given the frequent discussions in academia and the popular

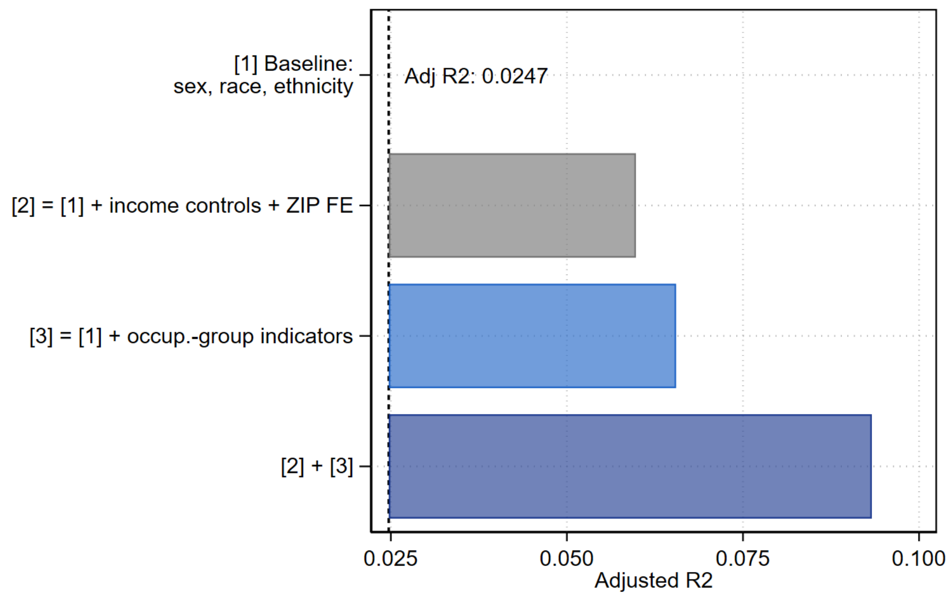


Figure 4. Additional Evidence on Explained Variation in Longevity

Notes: This figure shows the baseline and incremental adjusted R^2 for various specifications as indicated in the figure descriptions. Income controls are controls for the income profile (mean, $p10$, $p25$, $p50$) associated with the 794 detailed, six-digit SOC occupation groups. ZIP FE are ZIP code fixed effects for the 13,413 ZIP codes in the sample. Occupation group indicators are indicators for the 794 detailed SOC occupation groups.

press about their respective health risks and benefits.[6]

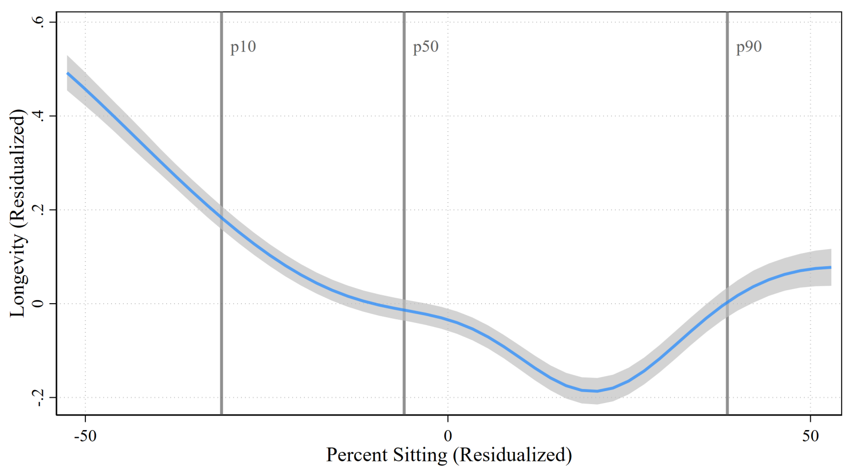
Figure 5 illustrates the association between longevity and occupational requirements. Each plot is constructed by residualizing longevity (age at death) and occupational requirement measure with respect to demographics and SOC group-level income profiles, and plotting the residuals against each other applying kernel-weighted local mean smoothing.

Life expectancy decreases as the percentage of indoor requirements in an occupation increases (Figure 5a). Occupations in the left tail of the percentage of indoor work (e.g., maintenance and

[6] See, e.g., Diaz et al. (2017) and Holt-Lunstad et al. (2010), as well as https://www.cnn.com/2017/09/11/health/sitting-increases-risk-of-death-study/index.html, and https://longevity.stanford.edu/lifestyle/2023/12/18/how-social-connection-supports-longevity/.

(a) Indoor Versus Outdoor Requirements

Notes: Representative occupations: Maintenance and Repair Workers ($p10$), General and Operations Managers ($p50$), Industrial Engineers ($p90$).

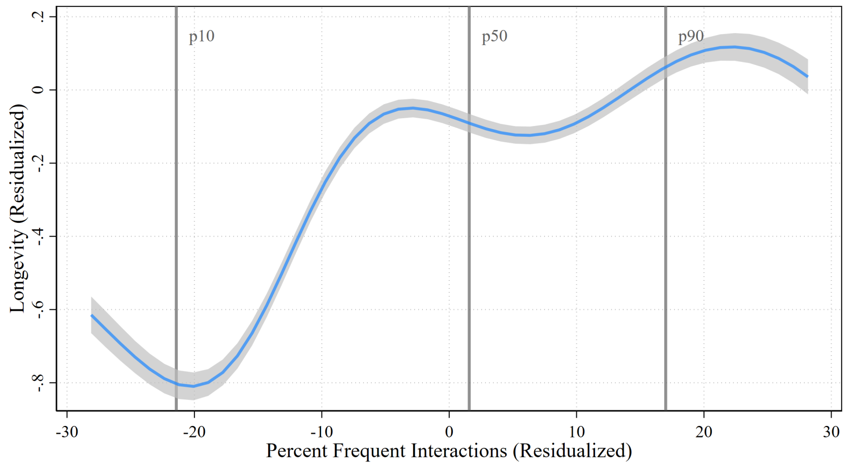


(b) Sedentary Requirements

Notes: Representative occupations: Carpenters ($p10$), Retail Salespersons ($p50$), Office Clerks ($p90$).

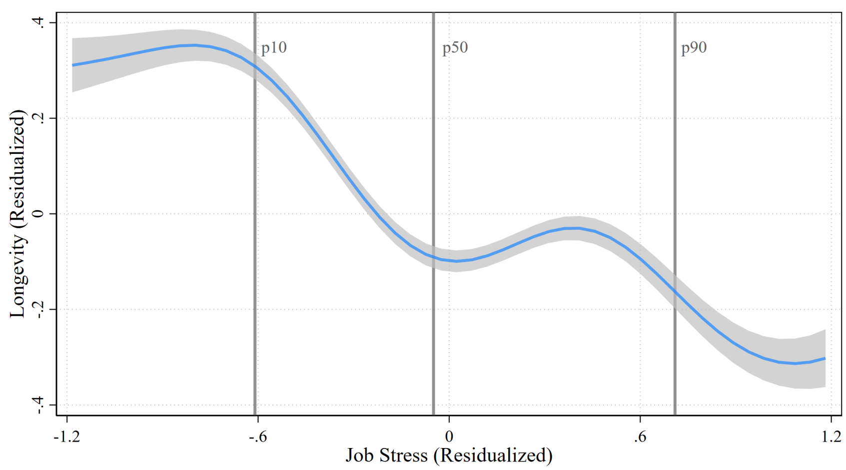
Figure 5. Longevity Differentials by Occupational Requirements

Notes: The figure shows the association between longevity and occupational requirements. Longevity and occupational requirement measures are residualized by decedents' sex, race, ethnicity, and six-digit SOC group-level income. Estimates based on administrative vital records data from CT, FL, MA, and OH. 99% confidence intervals are shown in gray. See Appendix A for variable definitions.



(c) On-the-Job Social Interactions

Notes: Representative occupations: Heavy and Tractor-Trailer Truck Drivers ($p10$), Construction Laborers ($p50$), Secretaries and Administrative Assistants ($p90$).

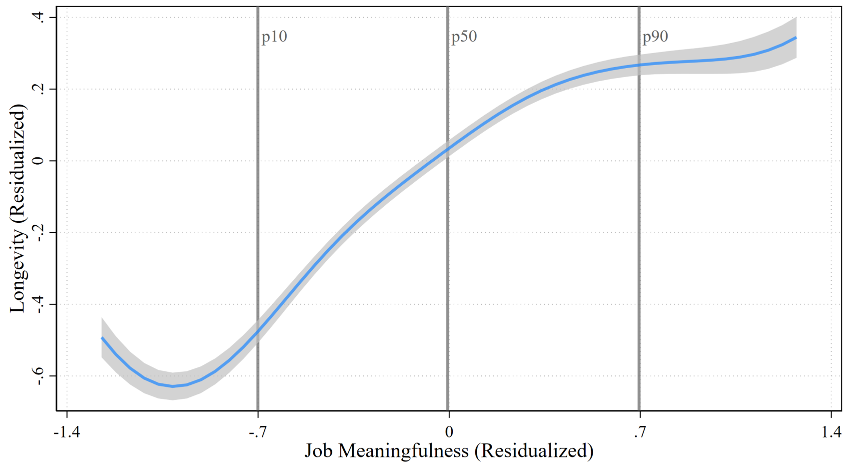


(d) Job Stress

Notes: Representative occupations: Automotive Service Technicians and Mechanics ($p10$), Retail Salespersons ($p50$), Construction Laborers ($p90$).

Figure 5. Longevity Differentials by Occupational Requirements

Notes: Continued from previous page.



(e) Job Meaningfulness

Notes: Representative occupations: Construction Laborers ($p10$), Cooks, Restaurant ($p50$), Clergy ($p90$).

Figure 5. Longevity Differentials by Occupational Requirements

Notes: Continued from previous page.

repair workers) are associated with about 0.5 additional years of life compared to occupations at the median (e.g., general and operations managers), and about 0.75-1.5 additional years of life compared to occupations in the right tail (e.g., industrial engineers).

Performing more of the job in a sitting position (e.g., retail salespersons) is also associated with lower longevity compared to having a very active job (e.g., carpenters), with the possible exception of those in the extreme right tail (e.g., office clerks) (Figure 5b). With respect to sedentary requirments, the maximum estimated longevity difference across percentiles is approximately 0.7 years.

Conversely, being exposed to more social interactions on the job (e.g., secretaries and administrative assistants versus construction laborers or heavy and truck drivers) is monotonically associated with higher longevity (Figure 5c). The maximum longevity difference associated with varying degrees of social interaction versus isolation is about 0.9 years.

Finally, higher job stress is monotonically associated with lower longevity (e.g., construction laborers versus retail salespersons and automotive services technicians) (Figure 5d), whereas higher job meaningfulness is monotonically associated with higher longevity (e.g., jobs in the clergy versus cooks and construction laborers) (Figure 5e). The maximum longevity gap associated with varying degrees of job stress and meaningfulness is approximately 0.7 and 0.9 years, respectively.

Overall, the evidence in Figure 5 shows that occupational requirements such as indoor work, sedentary tasks, social interaction, job stress, and job meaningfulness are significantly associated with variations in longevity, with specific job characteristics either enhancing or diminishing individuals' health capital and life expectancy. They are consistent with the view of occupational choice as a significant element of lifestyle that has strong ramifications of individual health outcomes.

IV.D. Underlying Cause of Death and Occupation

In the final part, we turn to the analysis of occupation-based influences on underlying causes of death. Figures 6 and 7 show the relationship between occupation and the likelihood of death from the two leading causes of death in the U.S., heart attack and cancer. The figures are constructed analogously to Figure 1 but substituting longevity with an indicator for the respective cause of death and additionally include age-at-death fixed effects. Thus, the figures characterize occupation-related differences in the cause of death among individuals with the same lifespan.

Figure 6 examines heart failure as the primary cause of death. Occupations in farming, fishing, and forestry, characterized by high outdoor activity and active lifestyles, are associated with the lowest probability of death from heart attacks, showing a significantly lower risk compared to all other occupational classes.

Figure 7 examines cancer-related deaths. One occupation group that stands out with one of

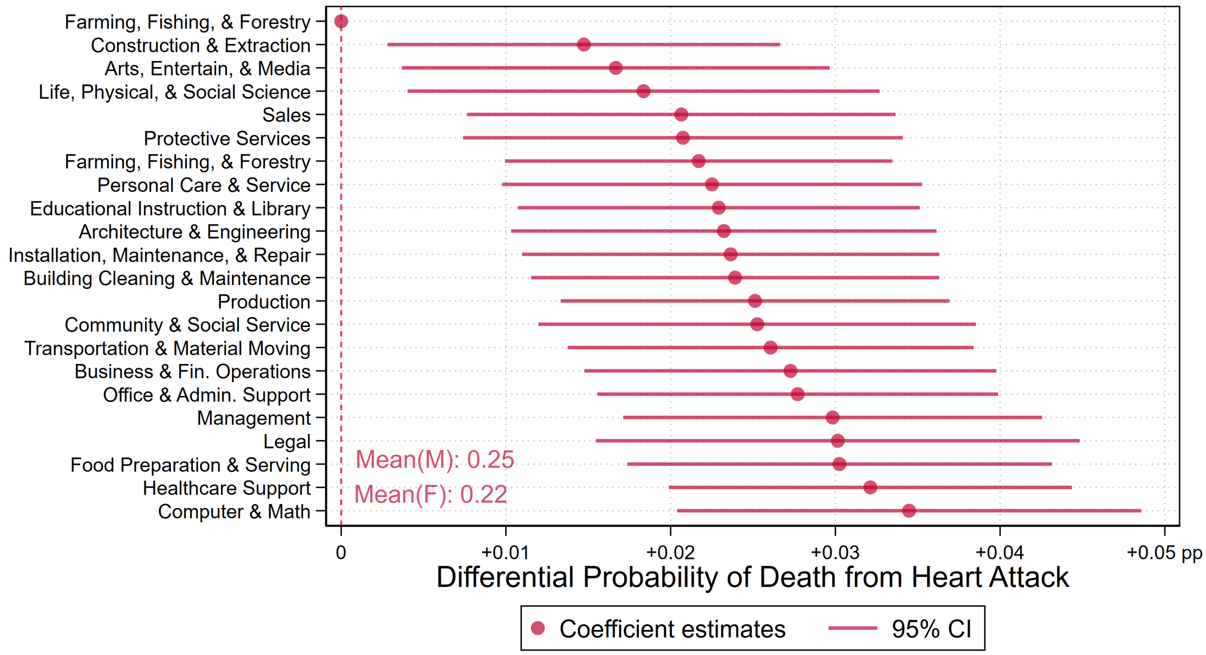


Figure 6. Underlying Cause of Death (Heart Attack) and Occupation

Notes: The figure shows differences in probability of death from heart attack across occupation classes, defined as the major occupation groups based on the 2018 Standard Occupational Classification (SOC) system. Estimates based on administrative vital records data from CT, FL, MA, and OH. $N = 4,027,011$. Estimation controls for decedents' sex, race, ethnicity, and six-digit SOC group-level income, and includes age-at-death fixed effects. See Appendix A for variable definitions.

the highest probabilities of cancer is construction labor jobs. Given that these jobs are associated with high stress levels (Section IV.C), the increased cancer risk may be due to stress-induced behaviors like smoking and poor diet, leading to hormonal imbalances and inflammation that could promote cancer growth. Conversely, healthcare-related occupations are associated with the lowest probabilities of death from cancer, possibly reflecting increased awareness of monitoring needs and better access to healthcare. This access and awareness could lead to early detection and treatment, ultimately reducing cancer mortality among these professionals, highlighting another aspect of how occupation may affect longevity.

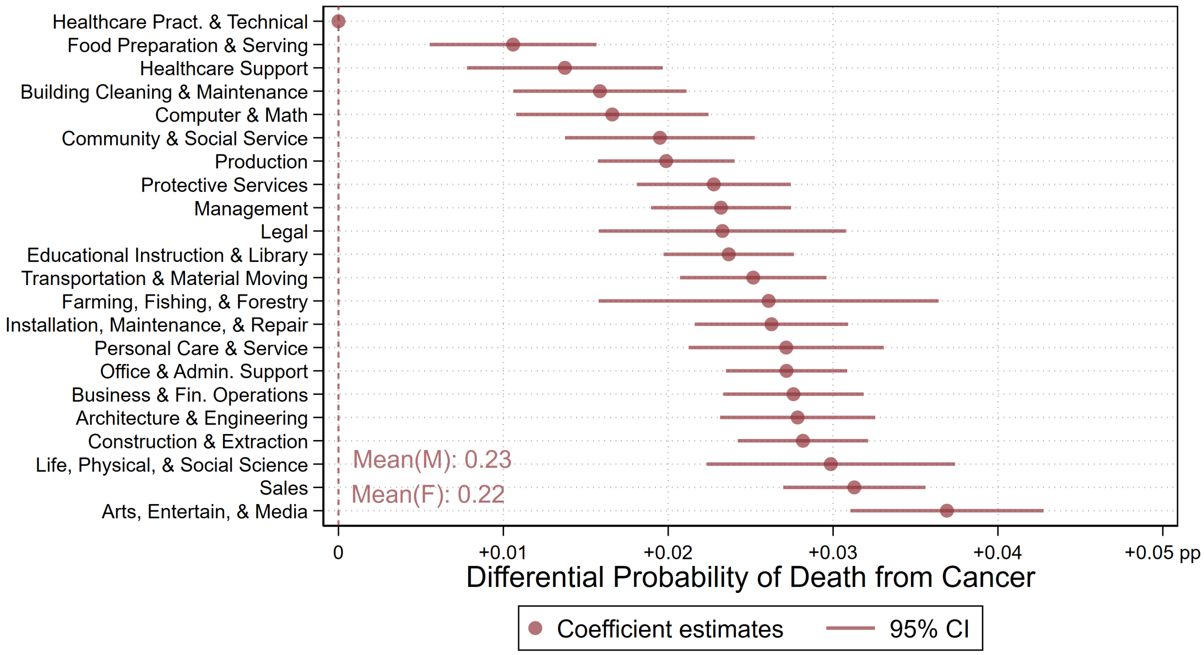


Figure 7. Underlying Cause of Death (Cancer) and Occupation

Notes: The figure shows differences in probability of death from cancer across occupation classes, defined as the major occupation groups based on the 2018 Standard Occupational Classification (SOC) system. Estimates based on administrative vital records data from CT, FL, MA, and OH. $N = 4,027,011$. Estimation controls for decedents' sex, race, ethnicity, and six-digit SOC group-level income, and includes age-at-death fixed effects. See Appendix A for variable definitions.

In sum, these differential, occupation-related cause-of-mortality patterns suggest the existence of occupational factors that could significantly influence an individual's health capital and longevity.

V. Conclusion

This article provides large-scale evidence on the relationship between longevity and occupational choice. Life expectancy varies substantially across occupations, and these differentials persist after controlling for demographics, occupation-group income profiles, and location. Occupations associated with shorter lifespans are characterized by a high fraction of indoor work, working in a sitting position, a lack of on-the-job social interactions, and being more stressful while less meaningful. Finally, there exist systematic, occupation-specific differences in the underlying cause of death among individuals with the same lifespan, suggesting that occupation-dependent tasks and lifestyles contribute to the erosion of individuals' health capital.

Our findings have implications for several core features of labor markets, ranging from job choice and job design to policy design. Regarding job choice, an important question is to what extent people account for job-related health risks in their career decisions, especially given that these decisions are typically made at a young age, while health conditions often unfold decades later. Regarding job design, a key consideration is identifying which changes to work practices may help reduce health strains, such as reducing sedentary requirements, increasing opportunities for social interaction, and minimizing stress to create healthier work environments.

Finally, regarding policy design, our results inform a range of questions, including those related to optimal retirement savings plans that account for occupation-driven differences in life expectancy. The latter is becoming increasingly important given the aging population in the U.S. and many other countries, as well as the evolving composition of professional occupations in the economy. In fact, in many countries, including France and Germany, efforts to reform the retirement system and adjust it to the rapidly aging populations face significant obstacles due to the physical and psychological demands of work and their adverse effects on health and aging. To date, this debate is largely happening with scarce empirical evidence. In the U.S., a critical policy issue in retirement planning pertains to the underfunding of pension plans, both at the state and federal levels. Given the occupation-centric structure of many pension plans, such as the California Teachers Retirement Fund (CalSTRS), understanding the relationship between occupation and life expectancy is crucial for ensuring adequate plan funding.

References

Black, S. E., N. Duzett, A. Lleras-Muney, N. G. Pope, and J. Price (2024). Intergenerational Transmission of Lifespan in the US. Working Paper, National Bureau of Economic Research.

Borgschulte, M., M. Guenzel, C. Liu, and U. Malmendier (2024). CEO Stress, Aging, and Death. *The Journal of Finance, forthcoming*.

Brønnum-Hansen, H., E. Foverskov, and I. Andersen (2020). Occupational inequality in health expectancy in Denmark. *Scandinavian Journal of Public Health 48*(3), 338–345.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016). The association between income and life expectancy in the United States, 2001-2014. *Jama 315*(16), 1750–1766.

Couillard, B. K., C. L. Foote, K. Gandhi, E. Meara, and J. Skinner (2021). Rising geographic disparities in US mortality. *Journal of Economic Perspectives 35*(4), 123–46.

Diaz, K. M., V. J. Howard, B. Hutto, N. Colabianchi, J. E. Vena, M. M. Safford, S. N. Blair, and S. P. Hooker (2017). Patterns of sedentary behavior and mortality in US middle-aged and older adults: a national cohort study. *Annals of Internal Medicine 167*(7), 465–475.

Finkelstein, A., M. J. Notowidigdo, F. Schilbach, and J. Zhang (2024). Lives vs. livelihoods: The impact of the great recession on mortality and welfare. Working Paper, National Bureau of Economic Research.

Holt-Lunstad, J., T. B. Smith, and J. B. Layton (2010). Social relationships and mortality risk: a meta-analytic review. *PLoS Medicine 7*(7), e1000316.

Jaeger, D. A., J. M. Nunley, R. A. Seals Jr, C. Shandra, and E. J. Wilbrandt (2023). The demand for interns. *Journal of Economic Behavior & Organization 209*, 372–390.

Johnson, N. J., P. D. Sorlie, and E. Backlund (1999). The impact of specific occupation on mortality in the US National Longitudinal Mortality Study. *Demography 36*(3), 355–367.

Katikireddi, S. V., A. H. Leyland, M. McKee, K. Ralston, and D. Stuckler (2017). Patterns of mortality by occupation in the UK, 1991–2011: a comparative analysis of linked census and mortality records. *The Lancet Public Health 2*(11), e501–e512.

Katz, L. F., J. Roth, R. Hendra, and K. Schaberg (2022). Why do sectoral employment programs work? Lessons from WorkAdvance. *Journal of Labor Economics 40*(S1), S249–S291.

Krueger, A. B. and A. I. Mueller (2012). Time use, emotional well-being, and unemployment: Evidence from longitudinal data. *American Economic Review 102*(3), 594–599.

Mackenbach, J. P., I. Stirbu, A.-J. R. Roskam, M. M. Schaap, G. Menvielle, M. Leinsalu, and A. E. Kunst (2008). Socioeconomic inequalities in health in 22 European countries. *New England Journal of Medicine 358*(23), 2468–2481.

Marmot, M., J. Banks, R. Blundell, C. Lessof, and J. Nazroo (2003). Health, wealth and lifestyles of the older population in England. *London: Institute of Fiscal Studies*.

Mueller, H. M., P. P. Ouimet, and E. Simintzi (2017). Within-firm pay inequality. *The Review of Financial Studies 30*(10), 3605–3635.

Olshansky, S. J., T. Antonucci, L. Berkman, R. H. Binstock, A. Boersch-Supan, J. T. Cacioppo, B. A. Carnes, L. L. Carstensen, L. P. Fried, D. P. Goldman, et al. (2012). Differences in life expectancy due to race and educational differences are widening, and many may not catch up. *Health Affairs 31*(8), 1803–1813.

Paglione, L., L. Angelici, M. Davoli, N. Agabiti, and G. Cesaroni (2020). Mortality inequalities by occupational status and type of job in men and women: results from the Rome Longitudinal Study. *BMJ open 10*(6), e033776.

Saez, E. and G. Zucman (2016). Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *The Quarterly Journal of Economics 131*(2), 519–578.

Smith, M., O. M. Zidar, and E. Zwick (2021). Top wealth in America: New estimates and implications for taxing the rich. Working Paper, National Bureau of Economic Research.

Online Appendix

Not for Publication

A. Variable Definitions

Table OA.1. Variable Definitions

Variable Name	Description
Age at Death	Calculated from exact dates of birth and death included in official state vital records.
Cause of Death	Primary cause of death (based on ICD-9 or ICD-10 codes) specified in state vital records.
Ethnicity	Hispanic or Latinx, Not Hispanic or Latinx
Income	SOC occupation group-level (six-digit) 2021 income profiles (mean, $p10, p25, p50$) from Bureau of Labor Statistics.
Occupation	Occupation categories are based on the 2018 Standard Occupational Classification (SOC) System. Occupations reported in the vital records are mapped to SOC categories using the O*NET-SOC AutoCoder software developed by R. M. Wilson Consulting for the Department of Labor.
Race	White, Black or African American, Native American or Alaska Native, Asian, Native Hawaiian or Other Pacific Island, Multiracial or Other Race, Unknown.
Sex	Female, male (other values are dropped).

B. Data and Methodology Details

B.1. Data Details

Table OA.2. Examples of SOC Category Mapping

Reported Occupation	Mapped SOC Category
Elementary School Te	Elementary School Teachers, Except Special Education
Elmntry Schl Teacher	Elementary School Teachers, Except Special Education
University Professor of Accounting	Business Teachers, Postsecondary
Office Supv	First-Line Supervisors of Office and Administrative Support Workers
Office Mgr	First-Line Supervisors of Office and Administrative Support Workers
Ret Clerk Typist	Word Processors and Typists
Hairdreser	Hairdressers, Hairstylists, and Cosmetologists
Babysitter	Childcare Workers
Parks Supervisor	First-Line Supervisors of Entertainment and Recreation Workers, Except Gambling Services
CNA	Nursing Assistants
Hospital Instrument Sterilizer	Medical Equipment Preparers
Orthodontic Asst.	Dental Assistants
Executive Chef	Chefs and Head Cooks
Sandwich Maker	Food Preparation Workers
Mixologist	Bartenders
Seamtress	Sewing Machine Operators
Silversmith	Jewelers and Precious Stone and Metal Workers
Turret Lathe Opr	Lathe and Turning Machine Tool Setters, Operators, and Tenders, Metal and Plastic

Notes: The table shows examples of reported and mapped occupations. Reported occupations in the left column are exactly as included in the vital records, including abbreviations and typographical errors.

B.2. Details on Methodology

As stated in Section III, we correct for occupation time trends by weighting observations (i.e., using weighted least squares), assigning a larger weight to observations when the occupation class is "underrepresented" in a given birth year.

We first calculate the fraction of observations (deaths) in the full sample corresponding to deceased individuals born in year t, $p(t)$:

$$p(t) = \frac{N_t}{N}.$$

Figure OA.1a (in the upper left) plots this distribution.

Next, we calculate the cumulative fraction of observations (deaths) of people born prior to or in year t relative to the full sample, $P(t)$:

$$P(t) = \sum_{s \leq t} p(t).$$

We then repeat these calculations separately for each occupation class Occ, calculating:

$$p^{Occ}(t) = \frac{N_t^{Occ}}{N^{Occ}}$$

and

$$P^{Occ}(t) = \sum_{s \leq t} p^{Occ}(t).$$

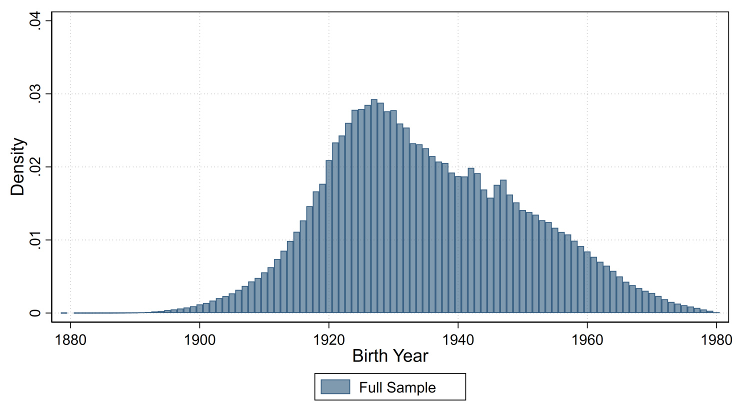
Finally, we compute the weight assigned to observation i in occupation class Occ born in year t, $w_i^{Occ}(t)$, as:

$$w_i^{Occ}(t) = \frac{P(t) - P(t-1)}{P^{Occ}(t) - P^{Occ}(t-1)}.$$

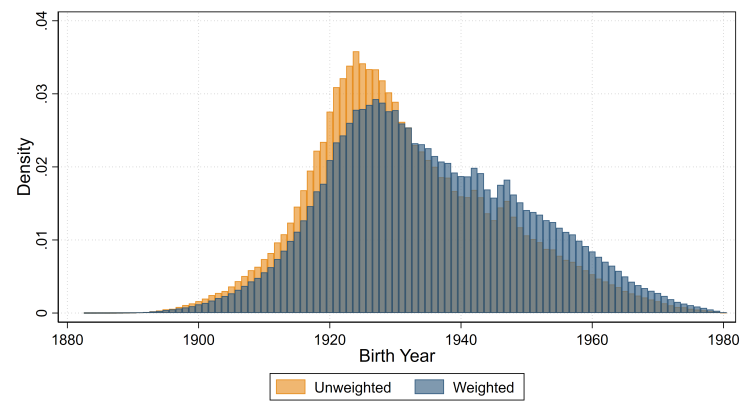
That is, we weight the observation of the death of person i from birth cohort t and occupation Occ by the fraction of all deaths of people born in t (relative to the full sample) over the fraction of all deaths of people born in t who also had occupation Occ (relative to the Occ sample). For example, if birth cohort t has been overrepresented in an occuption early since that occupation is declining in workers of later cohorts, the denominator would be higher than the numerator and hence the weight low to undo the overrepresentation. Vice versa, this fraction is high if, for birth cohort t, deaths of workers in a given occupation are underrepresented.

Figures OA.1b, OA.1c, and OA.1d plot the unweighted (in orange) and weighted (in blue) frequency distributions for three occupation classes in our sample: Office and Administrative Support (Panel b), Production (Panel c), and Construction (Panel d). The orange mass to the left of the blue distribution in panels (b) and (c) indicates that Office and Production jobs have decreased over time, leading to the overrepresentation of earlier birth cohorts from 1900 to 1925 in the occupation-specific subsamples (relative to their representation in the full sample of deaths). Vice versa, the additional orange mass to the right of the blue distribution in panel (d) indicates that the opposite is the case for Construction jobs – Construction jobs have increased over time. Weighting observations removes these time trends, resulting in identical weighted frequency distributions that correspond to the full-sample distribution in Figure OA.1a.

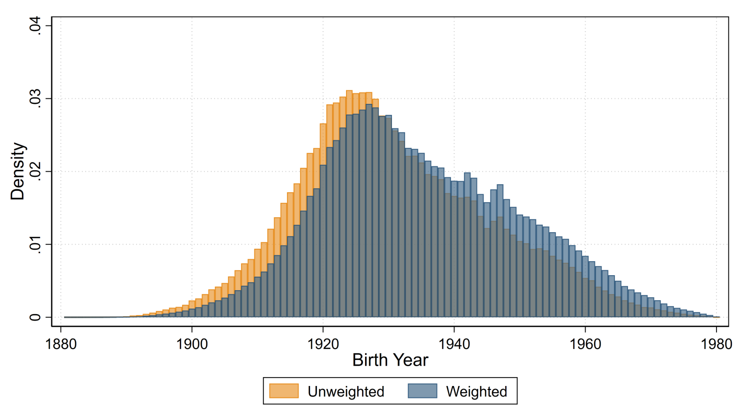
(a) Full Sample



(b) Office & Admin. Support



(c) Production



(d) Construction

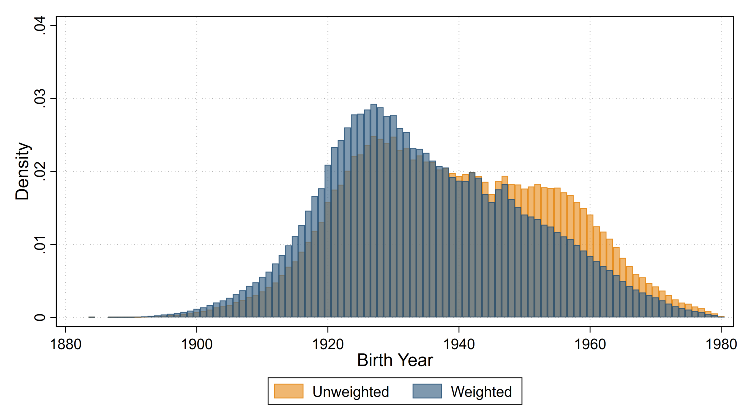


Figure OA.1. Occupation Frequencies Over Time for Selected Occupations

Notes: The figure shows unweighted (orange) and weighted frequency distributions for selected occupation categories. See Sections III and B.2 for details on the weighting procedure.

C. Supplementary Figures and Tables

Table OA.3. Estimates of Longevity Differentials by Occupation

	Baseline		With Zip Code FE	
	(1)		(2)	
Computer & Math	0.258***	(0.063)	0.187***	(0.062)
Management	0.739***	(0.046)	0.970***	(0.045)
Legal	0.758***	(0.088)	0.167*	(0.086)
Business & Fin. Operations	0.930***	(0.048)	0.834***	(0.048)
Office & Admin. Support	1.001***	(0.041)	0.823***	(0.041)
Educational Instruction & Library	1.231***	(0.047)	0.874***	(0.047)
Life, Physical, & Social Science	1.599***	(0.086)	1.414***	(0.084)
Architecture & Engineering	1.629***	(0.054)	1.646***	(0.053)
Healthcare Support	1.888***	(0.067)	1.924***	(0.067)
Arts, Entertain, & Media	1.923***	(0.067)	1.551***	(0.066)
Production	2.113***	(0.044)	2.324***	(0.043)
Food Preparation & Serving	2.256***	(0.053)	2.381***	(0.053)
Protective Services	2.262***	(0.054)	2.466***	(0.053)
Transportation & Material Moving	2.528***	(0.047)	2.949***	(0.047)
Community & Social Service	2.553***	(0.067)	2.342***	(0.066)
Construction & Extraction	2.692***	(0.041)	3.096***	(0.041)
Sales	2.730***	(0.049)	2.399***	(0.048)
Installation, Maintenance, & Repair	2.735***	(0.048)	3.129***	(0.048)
Personal Care & Service	3.187***	(0.068)	3.074***	(0.067)
Building Cleaning & Maintenance	3.282***	(0.056)	3.315***	(0.056)
Farming, Fishing, & Forestry	4.618***	(0.111)	5.168***	(0.110)
Sex Control	Yes		Yes	
Race and Ethnicity FE	Yes		Yes	
Income Controls	Yes		Yes	
ZIP Code FE	No		Yes	
N	4,027,011		4,027,011	

Notes: The table shows estimates of longevity differentials by occupation corresponding to Figure 1 (Column (1)) and the plot with ZIP code fixed effects in Figure 2 (Column (2)), respectively. See these figures for additional details.