Longevity and Occupational Choice*

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

March 15, 2024

Please do not circulate without permission.

Abstract

Using administrative vital records for 15 percent of the U.S. population, we find large variation in longevity across occupational groups. For some occupation pairs, the differences in lifespan are comparable to the longevity gap between men and women. This variation persists after controlling for demographic, environmental, and pecuniary drivers of longevity. A key mechanism linking the variation in longevity to occupations is the intensity of manual tasks. A higher manual intensity predicts a shorter lifespan. Consistent with the importance of on-the-job tasks for terminal health, occupational choice predicts the cause of death, even holding constant the age at death. Overall, we provide evidence on a key lifestyle determinant of longevity with implications for occupation-based healthcare and retirement plans.

* We thank Nathaniel Hendren and Andrei Shleifer as well as participants at the 2023 Fall NBER Economics of Health Meeting for helpful comments. For this paper, we have used the O*NET-SOC AutoCoder software developed by R. M. Wilson Consulting for the Department of Labor; we are grateful to Bob Wilson and Gary Sincick of R. M. Wilson Consulting for providing us with the AutoCoder results. We thank Sasmira Matti and Tim Zhang for outstanding research assistance.

[†]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 drivers of life expectancy are endowed at birth, others reflect personal lifestyle decisions. One of the crucial lifestyle decisions is an individual's choice of occupation, since 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).

Despite the importance of life expectancy for economic policy and retirement programs, our understanding of the association between longevity and occupational choice remains limited. A major impediment is the lack of comprehensive data linking individuals' career choices and health outcomes later in life. Most existing work relies on small-scale comparisons based on survey data or single-city samples.[1]

This paper uses administrative vital records for about 15% of the U.S. population across a variety 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, such as Florida, Connecticut, Massachusetts, and Ohio, and include detailed personal data on millions of deceased individuals. Examples of individual-level records include the person's dates of birth and death, a granular classification of the primary and secondary death causes based on medical exams, the usual occupation before retirement, and personal demographics, such as sex, race, ethnicity, education, place of birth, and residential address.

[1] Brønnum-Hansen, Foverskov, and Andersen (2020) use Danish survey data paired with register data ($N = 3,179$), and only distinguish blue versus white-collar jobs, instead of a more granular, occupation-specific analysis of longevity. Paglione, Angelici, Davoli, Agabiti, and Cesaroni (2020) study the relation between broadly-defined occupation categories and mortality using data from Rome, Italy.

We organize our analyses 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, such as 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).[2]

We find large disparities in longevity across occupational categories, controlling for demographics and occupation-level income. For many occupation pairs, life expectancy differentials are nearly as large as the longevity gap between men and women. For example, healthcare practitioners, on average, have a lifespan two years shorter than sales employees. These disparities persist with the same magnitudes across individuals residing in the same zip code and subject to similar environmental factors.

In the second part, we focus on occupational tasks and the intensity of routine and manual labor. We map the granular SOC occupations to the job task intensity measures developed in Autor and Dorn (2013).[3] Controlling for demographics and occupation-level income, life expectancy is roughly monotonically decreasing in the occupation's abstract task intensity. In contrast, we find an approximate inverse-U shaped association of life expectancy with manual task intensity and an approximate U-shaped association with routine task intensity.

In the third part, we study the association between occupational choice and the primary and secondary causes of death, using data from official medical exams. We focus on the most common

[2] As described in more detail in Section II, for this step, we have used the O*NET-SOC AutoCoder software developed by R. M. Wilson Consulting for the Department of Labor.

[3] For this part, we use the linking tables provided by David Dorn at `https://www.ddorn.net/data.htm`. Abstract task intensity measures an occupation's demand for "direction control and planning" and "General Educational Development (GED) Math." Manual task intensity measures the demand for "eye-hand-foot coordination." Routine task intensity measures the demand for "set limits, tolerances and standards" and "finger dexterity," and See Autor and Dorn (2013) for further detail.

death causes, such as heart disease and cancer.[4] We find persistent variation in the causes of death across occupations, even when comparing individuals with the same lifespan and the same age at death. For example, farming, fishing, and forestry occupations show a considerably reduced relative likelihood of mortality from heart disease. Conversely, cancer-related deaths are particularly unlikely among healthcare practitioners and technical occupations. These differential cause-of-mortality patterns suggest the existence of occupational factors that 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) and on rising geographic disparities in mortality in the U.S. (Couillard et al. 2021). We uncover large life expectancy differentials across occupation groups and job task requirements. This variation persists for individuals with similar demographics, comparable incomes, and nearby residences.

More broadly, our findings also extend the work on inequality. Much of this 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, and The New York Times 2023). Our results highlight the importance of job-related disparities not only for income but also for terminal health and lifespan. We uncover 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 described the methodology. Section IV presents the results. Section V concludes.

[4] See https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm.

II. Data

II.A. Vital Records

Our main data on individuals' lifespans, occupations, and demographics is based on official state vital records, collected by each state's department of health. These records cover the universe of death events in a state for the recorded period. Vital records contain dozens of variables for each death event, such as the decedent's occupation and industry, years of education, and close relatives. The coverage of family structures permits within-family analyses that account for familial and congenital mortality drivers. Vital records also provide the official medical conclusion regarding the death cause, a detailed classification of 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. These records also include a variety of demographic variables, such as gender, race, ethnicity, residential address, which account for local and environmental mortality factors.

We have vital records from four states: Connecticut, Florida, Massachusetts, and Ohio. Collectively, these states represent about 15% of the U.S. population. Table 1 shows the distribution of individuals in our final dataset by state, as well as corresponding sample periods for death events.[5] The largest fraction of death events come from Florida, followed by Massachusetts, Ohio, and Connecticut. In total, our dataset comprises more than four million deceased individuals in these states between 1990 and 2020.

[5] 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: Figure shows the distribution of individuals in our final dataset by state, as well as associated sample periods (where sample period refers to the years for which we observe all individuals passing in a given state).

II.B. SOC Occupation Categories

Vital records report occupations, industries, and sometimes employers via free-form responses. Similar to Katz et al. (2022) and Jaeger et al. (2023), we map reported occupations to six-digit SOC categories, which we then aggregate further into major occupation groups for some analyses. 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.

Table OA.2 in the Online Appendix contains examples of reported and mapped occupations. The examples show that the mapping is highly reliable 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.[6]

Table 2 shows the distribution of major SOC occupation groups in our dataset, as well as sample

[6] The software also successfully handles other subtle complexities, such as mapping "Executive Chef" to "Chefs and Head Cooks" rather than "Executives."

free-form occupations reported in the vital records associated with them. Office and Administration Support workers constitute the largest group, oftentimes corresponding to Secretary, Clerk, or Bookkeeper in the vital records. Other frequent occupations are those related to Construction (Laborers, Carpenters, Electricians, etc.) and Production (Supervisors, Machinists, Seamstresses). Infrequent occupation classes include Farming, Fishing, and Forestry as well as Legal occupations.

II.C. Abstract, Manual, and Routine Task Intensities

We furthermore map the SOC occupations with the abstract, manual, and routine task intensity scores from Autor and Dorn (2013). Figures OA.2 to OA.4 in the Online Appendix plot the distribution of the task intensity scores by major SOC occupation class. As these figures reveal, most of the variation in jobs' abstract, manual, and routine task intensity is across, rather than within, major SOC occupation classes.

II.D. Underlying Cause of Death

To categorize causes of death into those related to heart diseases and cancer, we rely on the detailed vital-records information. This 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).[7]

III. Methods

To describe the baseline association between longevity and occupation, we estimate:

$$AgeAtDeath_i = \alpha + \beta' \mathbf{Occ_i} + \gamma' \mathbf{X_i} + \varepsilon_i \tag{1}$$

[7] See https://ftp.cdc.gov/pub/health_statistics/nchs/datalinkage/linked_mortality/.

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

Notes: 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.

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. 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 methodology 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).

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 relative to the full-sample distribution. We present a more 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 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 in Figure OA.1a.

We proceed analogously, in terms of the estimating equation, including individuals aged 40 and older at their passing, and weighting of observations, to describe the association between longevity and job task intensities, as well as that between underlying cause of death and occupation. With respect to the former, we estimate:

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

where $type \in \{abstract, manual, routine\}$. With respect to the latter, we estimate:

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

where $c \in \{cancer, heart\ attack\}$. We focus on the relationship with cancer and heart attack as they are the leading causes of death in the U.S. When estimating Eq. (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. Table OA.3 in the Online presents the regression estimates corresponding to Figure 1. 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 for men and women in that group, conditional on passing at or after age 40, is shown at the bottom left of the figure.[8] Average life expectancy in other SOC classes thus corresponds to these numbers plus the relevant coefficient estimate shown in the figure.

The figure uncovers large disparities in life expectancy across occupations. Frequently, the

[8] These statistics are derived from estimating Eq. (1) where the vector of control variables solely includes an indicator for decedent sex.

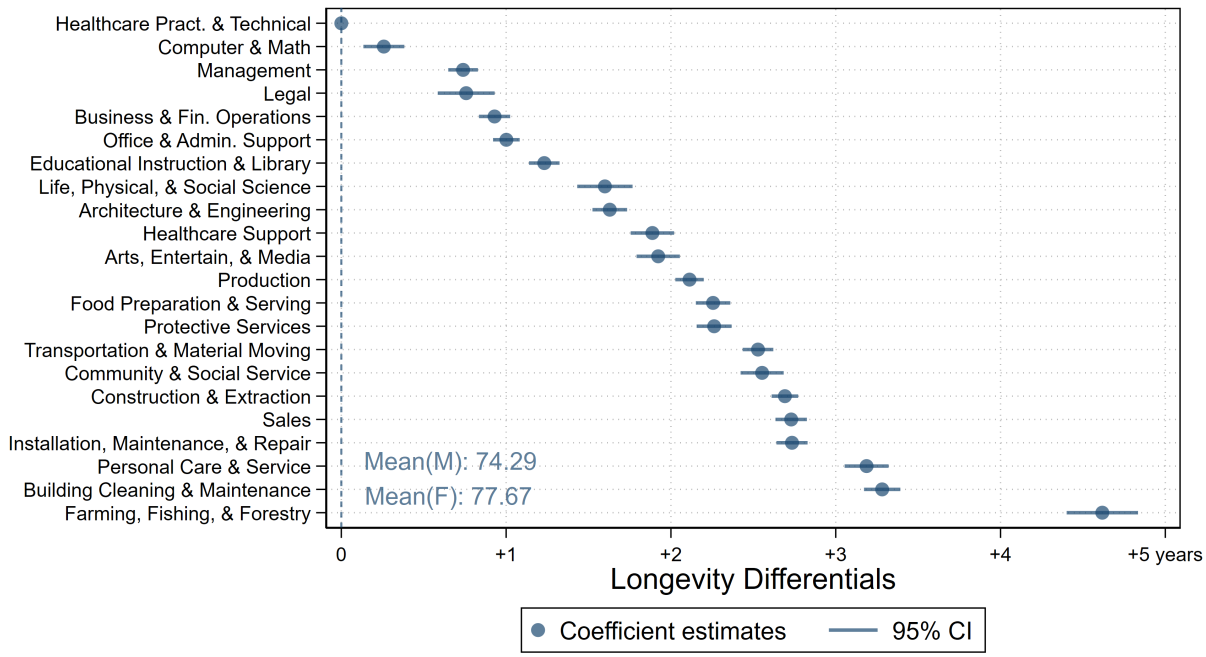

Figure 1. Baseline Estimates of Longevity Differentials by Occupation

Notes: 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.

cross-occupation differences are nearly as large as the longevity gap between men and women. For example, healthcare practitioners and office administration workers, on average, have a life expectancy of over two years less than employees in sales and related occupations. Management and legal occupations are associated with relatively shorter life expectancy as well, whereas farming, fishing, and forestry occupations are associated with the longest life expectancy (controlling, as noted above, for the decedent's demographics and six-digit SOC group-level income profiles).

Figure 2 presents within-geography estimates of the association between longevity and SOC occupational class (see again Table OA.3 for the corresponding regression results). Relative to

Figure 1, Figure 2 is estimated by including, in addition, decedents' residence ZIP code fixed effects in the estimation. As the figure reveals, the large longevity disparities across occupations persist when restricting to within-ZIP-code variation. Additionally, the ranking of occupations in terms of associated life expectancy is similar with and without ZIP code fixed effects.

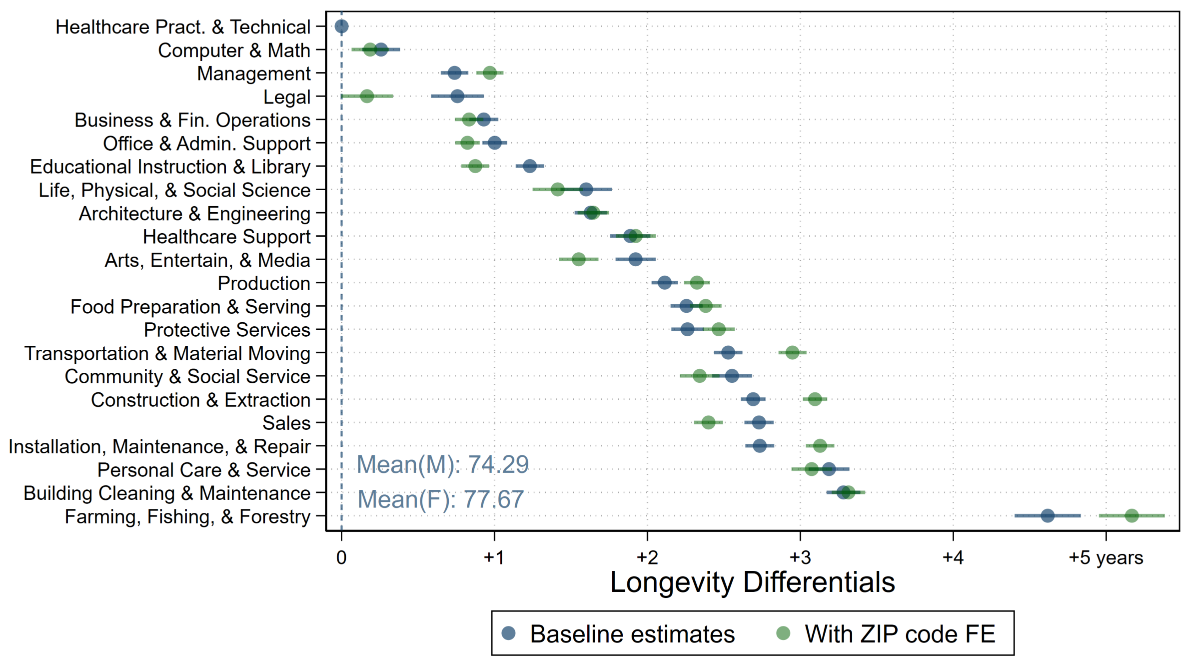

Figure 2. Within-Geography Longevity Differentials

Notes: 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, six-digit SOC group-level income, and includes ZIP code fixed effects. See Appendix A for variable definitions.

IV.B. Longevity and Job Task Intensities

Figure 3 plots the association between longevity and abstract (top figure), manual (middle figure), and routine (bottom figure) task intensities, derived as in Autor and Dorn (2013). The plots are constructed by residualizing longevity (age at death) and the task intensity measures with

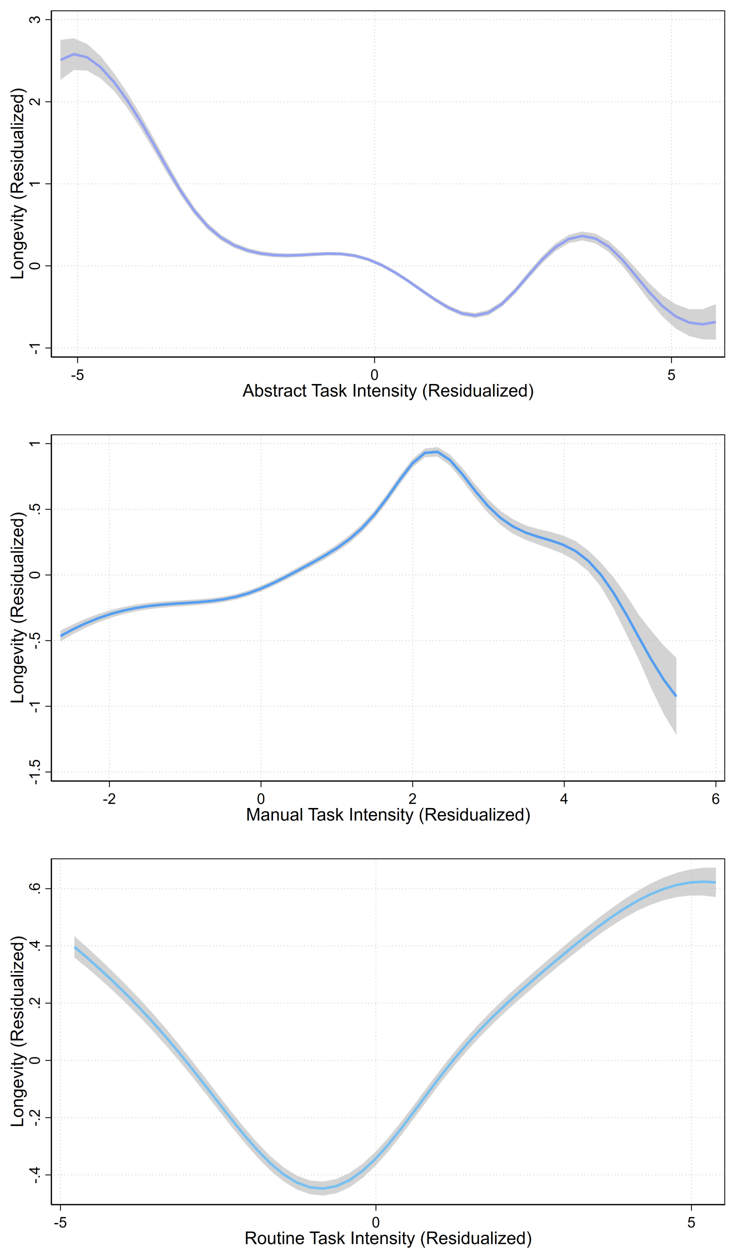

Figure 3. Longevity Differentials by Job Task Intensities

Notes: Figure shows association between longevity and jobs' abstract, manual, and routine task intensities (Autor and Dorn 2013). Longevity and task intensity 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. $N = 4,019,949$. 99% confidence intervals are shown in gray. See Appendix A for variable definitions.

respect to decedent demographics and SOC group-level income profiles (cf. Figure 1), and plotting the residuals against each other applying kernel-weighted local mean smoothing.

Life expectancy is roughly monotonically decreasing in the abstract task intensity (ATI) of an occupation. Occupations with the lowest ATI are associated with about two additional years of life compared to highest-ATI occupations. The longevity patterns for manual task intensity (MTI) and routine task intensity (RTI) are nonlinear. The longevity–MTI association is inverse-U shaped. Occupations with the lowest and highest MTI are, on average, associated with approximately one year less in life expectancy compared to medium-MTI occupations. By contrast, the longevity–RTI association is flipped and U-shaped. Lowest-RTI and highest-RTI occupations are associated with an approximately one-year higher life expectancy than medium-RTI occupations.

IV.C. Underlying Cause of Death and Occupation

Figures 4 and 5 show the relationship between occupation and the probability of a heart attack and cancer. The figures are constructed analogously to Figure 1 and additionally include age-at-death fixed effects. Consequently, the figures characterize occupation-related differences in the cause of death among individuals with the same lifespan. The figures are interrelated and best interpreted jointly because the outcome variables (death from a heart attack and death from cancer) are competing events.

Both figures reveal sizable differences in the underlying cause-of-death probabilities across occupations. These differences amount to up to four percentage points relative to a base probability of death from a heart attack or from cancer of about 20% each. Figure 4 focuses on heart failure as the primary cause of death. The figure reveals that occupations in farming, fishing, and forestry show a significantly lower probability of death from heart attacks compared to other occupation classes. Occupations with a medium probability of death from heart-related disease include job

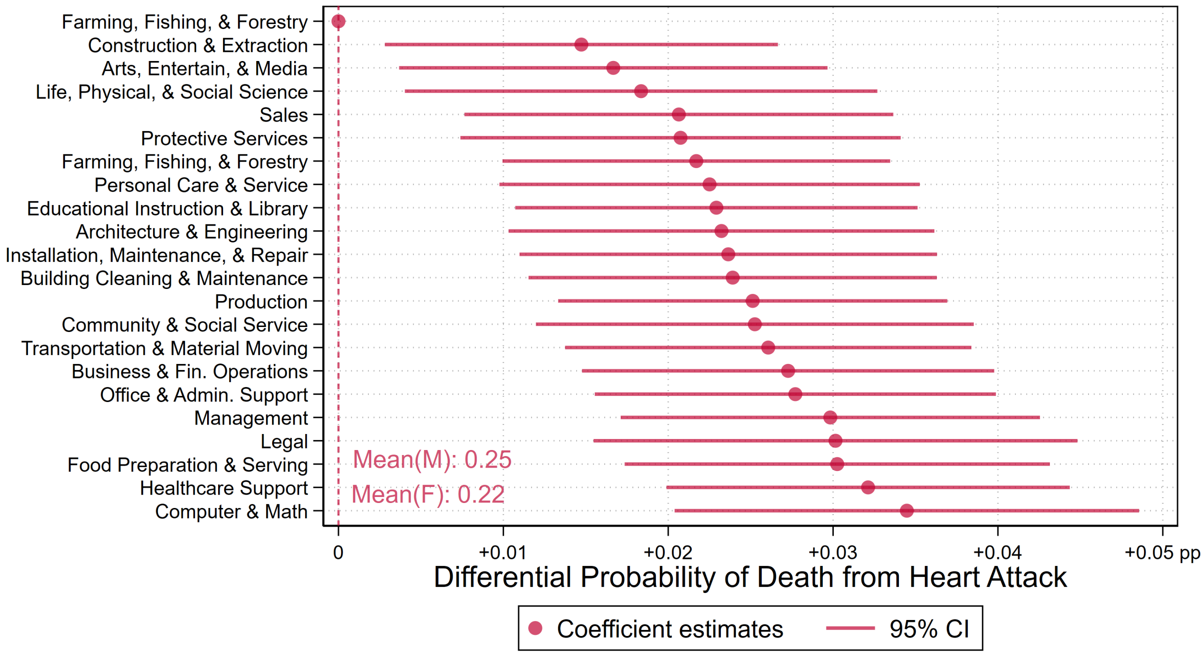

Figure 4. Underlying Cause of Death (Heart Attack) and Occupation

Notes: 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.

in installation and maintenance. Occupations in healthcare support and computers/mathematics are associated with the highest probably of heart failure as the underlying cause of death. When we focus on cancer as the primary cause of death, we find that healthcare practitioners and food preparation occupations are associated with the lowest probabilities of death from cancer. In contrast, sales and Arts/Entertainment occupations have the highest probability of death from cancer.

Overall, there are significant differences in the cause of death across occupations, and these differentials persist while holding fixed the age at death. This pattern suggests that some occupation-specific tasks contribute to the erosion of individuals' health capital.

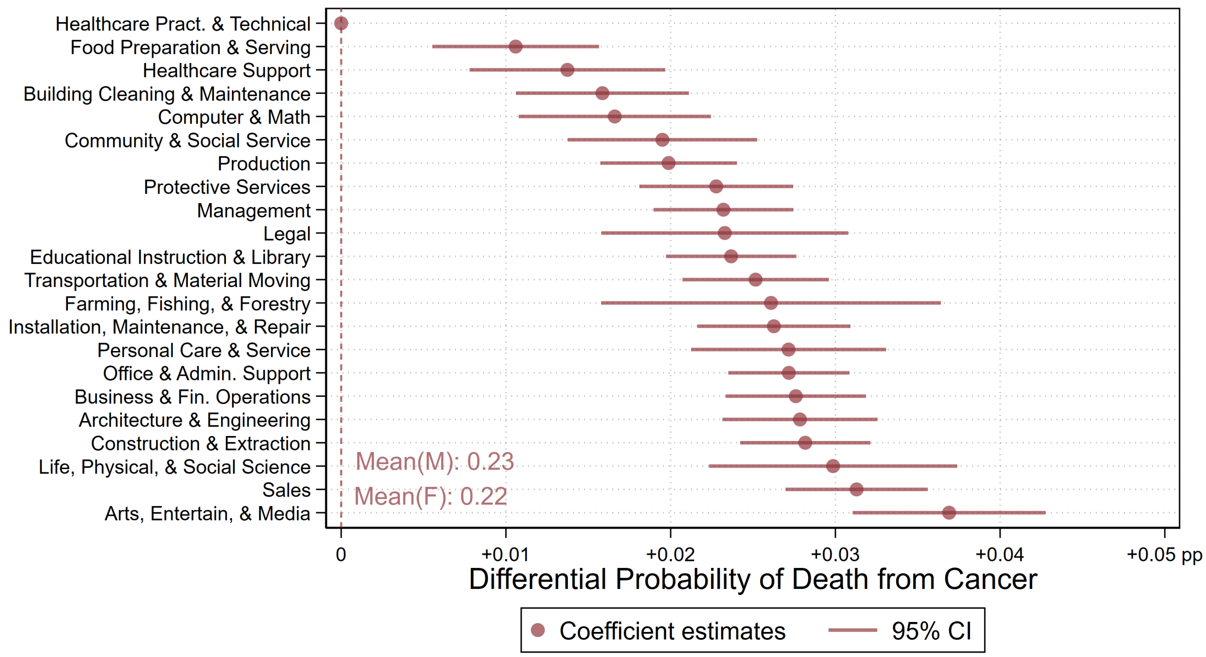

Figure 5. Underlying Cause of Death (Cancer) and Occupation

Notes: 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.

V. Conclusion

This article provides one of the first large-scale investigations of the relationship between longevity and occupational choice. Life expectancy varies substantially across occupations, and these differentials persist after controlling for demographics and occupation group income profiles. Life expectancy also varies systematically with jobs' abstract, manual, and routine task intensities (Autor and Dorn 2013). Finally, there exist systematic, occupation-specific differences in the underlying cause of death among individuals with the same lifespan, suggesting that occupation-specific tasks contribute to the erosion of individuals' health capital.

Given the aging population in the U.S. and many other countries, as well as the evolving composition of professional occupations in the economy, these findings provide insights for policy makers. Our results also inform questions around optimal retirement savings that account for occupation-driven differences in life expectancy. 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. Even 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

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review 103*(5), 1553–1597.

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.

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.

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.

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

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.

The New York Times (2023). The Greatest Wealth Transfer in History Is Here, With Familiar (Rich) Winners. https://www.nytimes.com/2023/05/14/business/economy/wealth-generations.html.

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.
Hispanic/Latinx	Hispanic or Latinx, Not Hispanic or Latinx
Cause of Death	Primary cause of death (based on ICD-9 or ICD-10 codes) specified in state vital records.
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).
Task Intensities	Abstract, manual, and routine task intensities of jobs from Autor and Dorn (2013).

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: 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 detailed in the main text 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 in the full sample corresponding to deceased individuals born in year t, $p(t)$:

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

Figure OA.1a plots this distribution.

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

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

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

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

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, Production, and Construction. As alluded to above, the unweighted show that over time, Production jobs have become fewer, whereas Construction jobs have increased. 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

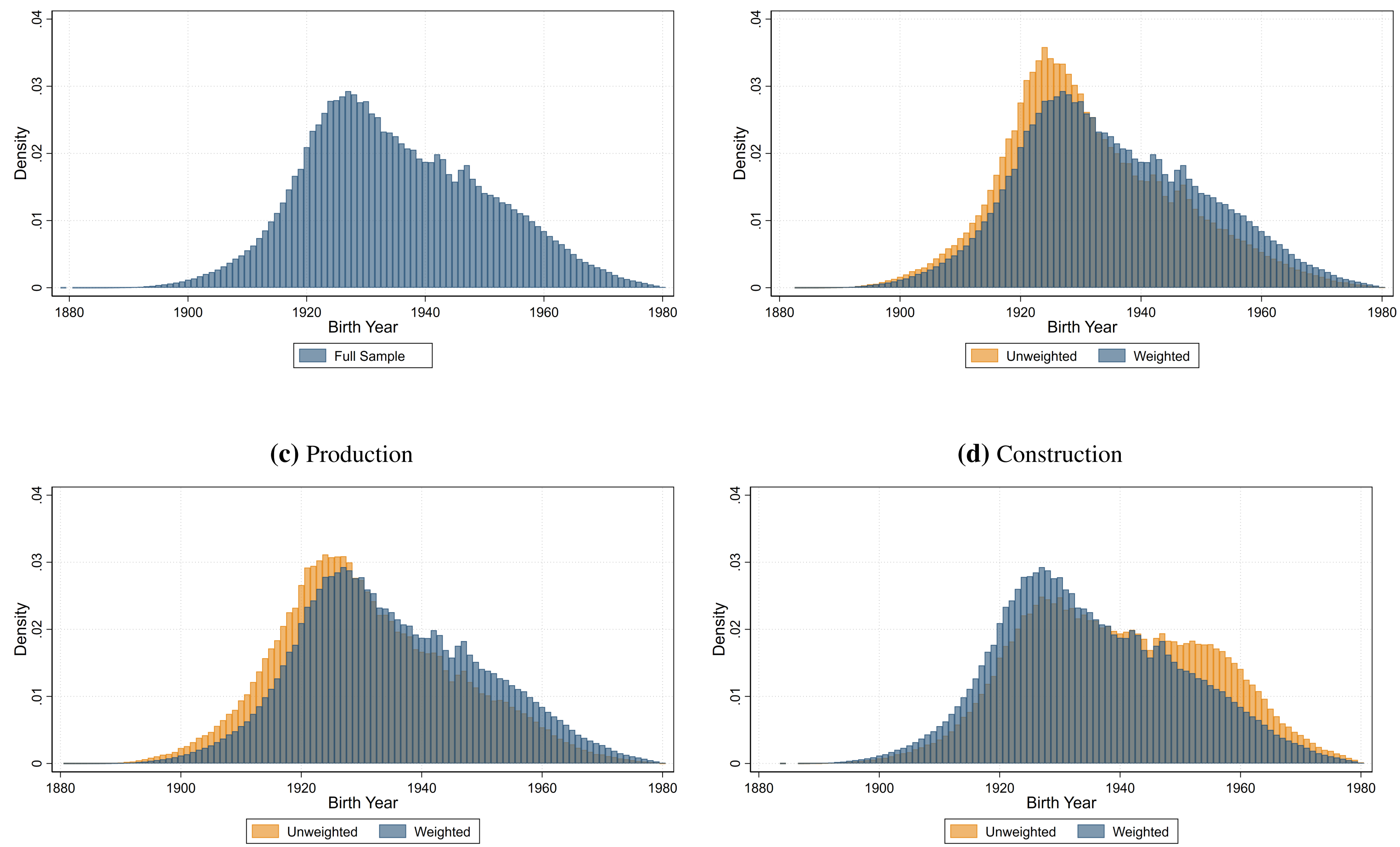

Figure OA.1. Occupation Frequencies Over Time for Selected Occupations

Notes: 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

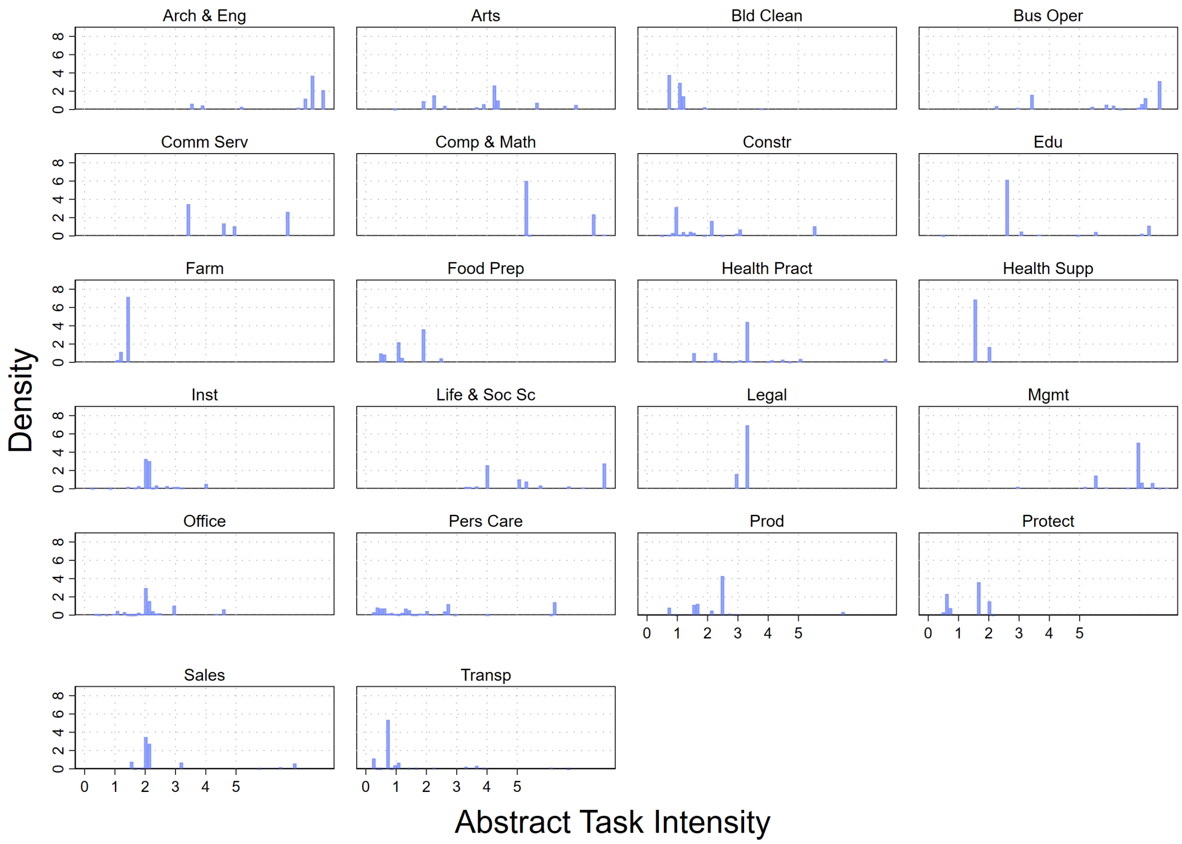

Figure OA.2. Distribution of Abstract Task Intensity by Occupation

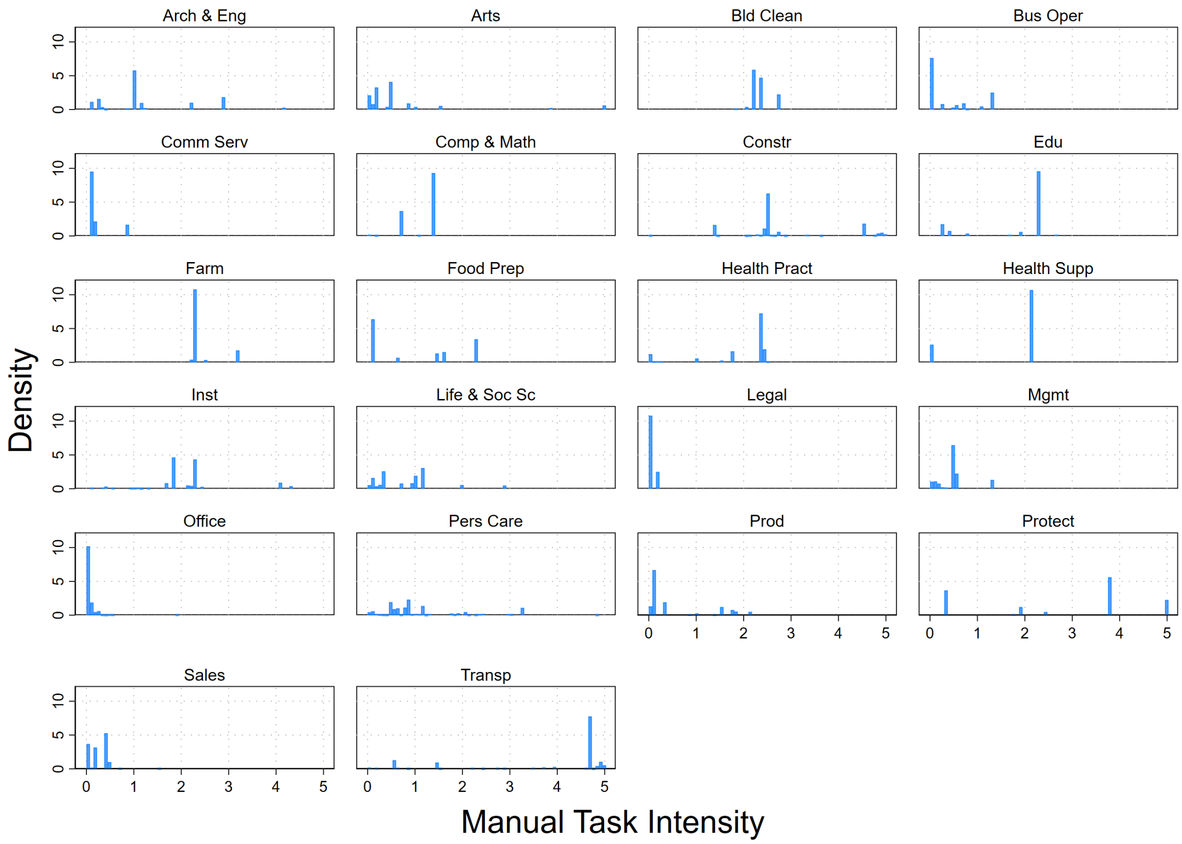

Figure OA.3. Distribution of Manual Task Intensity by Occupation

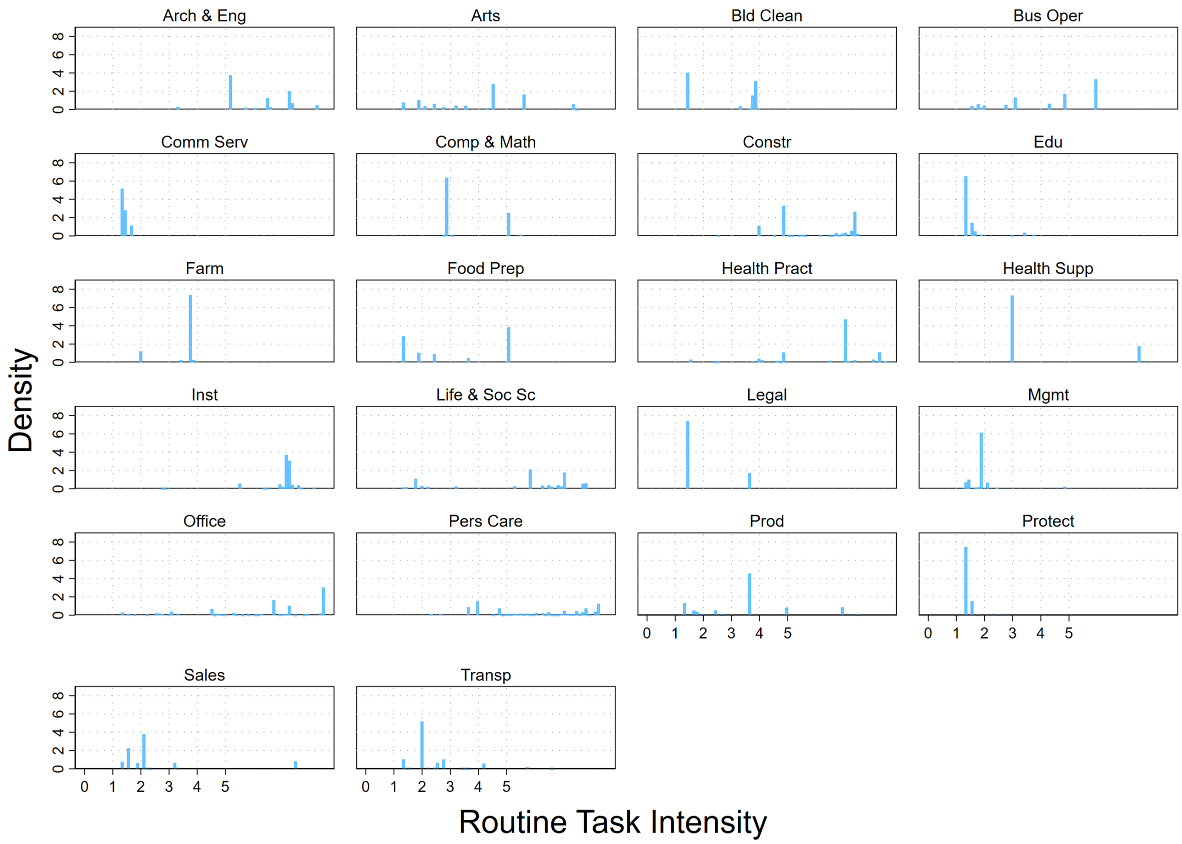

Figure OA.4. Distribution of Routine Task Intensity by Occupation

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: Table shows estimates of longevity differentials by occupation corresponding to Figure 1 (Column (1)) and Figure 2 (Column (2)), respectively. See these figures for additional details.