

# UCSF

## UC San Francisco Previously Published Works

### Title

Predicting mortality from 57 economic, behavioral, social, and psychological factors

### Permalink

<https://escholarship.org/uc/item/7xm3h2k3>

### Journal

Proceedings of the National Academy of Sciences of the United States of America, 117(28)

### ISSN

0027-8424

### Authors

Puterman, Eli  
Weiss, Jordan  
Hives, Benjamin A  
[et al.](#)

### Publication Date

2020-07-14

### DOI

10.1073/pnas.1918455117

Peer reviewed

**Title: Predicting mortality from 57 economic, behavioral, social, and psychological factors**

**Authors:** Eli Puterman<sup>1\*</sup>, Jordan Weiss<sup>2</sup>, Ben Hives<sup>1</sup>, Alison Gemmill<sup>3</sup>, Deborah Karasek<sup>4</sup>, Wendy Berry Mendes<sup>5</sup>, David H. Rehkopf<sup>6\*</sup>

5 **Affiliations:**

<sup>1</sup> School of Kinesiology – The University of British Columbia, 210-6081 University Blvd, Vancouver BC, Canada, V6T1Z1

<sup>2</sup> Population Studies Center and the Leonard Davis Institute of Health Economics, University of Pennsylvania, 239 McNeil Building, 3718 Locust Walk, Philadelphia, PA, USA 19104.

<sup>3</sup> Department of Population, Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe Street, Baltimore, MD 21205

<sup>4</sup> Preterm Birth Initiative, Department of Obstetrics, Gynecology & Reproductive Sciences, University of California, San Francisco, CA 94158

15 <sup>5</sup> Department of Psychiatry, University of California San Francisco, 401 Parnassus Ave, San Francisco, CA 94143

<sup>6</sup> School of Medicine, Stanford University, 1701 Page Mill Road, Palo Alto, CA, USA, 94304

\* **Correspondence to:** [eli.puterman@ubc.ca](mailto:eli.puterman@ubc.ca), [drehkopf@stanford.edu](mailto:drehkopf@stanford.edu)

20 •

## **Abstract:**

Behavioral and social scientists have identified many non-biological predictors of mortality. An important limitation of much of this research, however, is that risk factors are not studied in comparison with one **another** 5 **or from across different fields of research**. It therefore remains unclear which factors should be prioritized for interventions and policy to reduce mortality risk. In the current investigation, we compare 57 factors within a multidisciplinary framework. These include: (1) adverse socioeconomic and psychosocial experiences during childhood, and (2) socioeconomic 10 conditions, (3) health behaviors, (4) social connections, (5) psychological characteristics, and (6) adverse experiences during adulthood. The current prospective cohort investigation with 13,611 adults from 50 to 104 years of age (mean age = 69.3) from the nationally representative Health and Retirement Study used weighted traditional (i.e. multivariate Cox 15 regressions) and machine learning (i.e. Lasso, random forest analysis) statistical approaches to identify the leading predictors of mortality over six years of follow-up time. We demonstrate that, in addition to the well-established behavioral risk factors of smoking, alcohol abuse and lack of physical activity, economic (e.g. recent financial difficulties, unemployment 20 history), social (e.g. childhood adversity, divorce history), and psychological (e.g. negative affectivity) factors were also among the strongest predictors of mortality among older American adults. The strength of these predictors should be used to guide future transdisciplinary investigations and

intervention studies across the fields of epidemiology, psychology, sociology, economics, and medicine to understand how changes in these factors alter individual mortality risk.

## Introduction

Population aging has given rise to an increase in non-communicable diseases that account for the majority of deaths in the United States (1). Although disease and mortality are caused by molecular, cellular, and physiological changes, non-biological processes play important roles in shaping mortality risk. The National Institutes of Health Office of Behavioral and Social Sciences designates the behavioral and social sciences as fundamental to our understanding of disease pathogenesis and mortality, with knowledge derived from these disciplines leading to scientific breakthroughs that can transform health policy (2). New efforts and approaches to the ways in which research guides policy are needed given the three decade stagnation of U.S. life expectancy relative to other industrialized countries (3).

Studies of the determinants of mortality have identified a wide range of behavioral risk factors across disciplines. McGinnis and Foege (4), followed by the work of Mokdad and colleagues (5), established the prevailing role that health behaviors—predominantly smoking, poor nutrition, and physical inactivity—play in mortality rates, accounting for nearly 35% of all deaths in the United States. While health behaviors have received the majority of attention in the scientific literature, other studies have identified specific economic, social, and psychological factors associated with higher risk of mortality, such as early childhood adversity (6), financial difficulties in adulthood (7), poor social relationships (8), lower levels of neuroticism and

conscientious (9, 10), and experiences of discrimination (11). An important limitation of much of this research, however, is that risk factors within and between these domains are often studied in isolation from each other with a priori hypotheses, so it is unclear which factors are the relatively strongest predictors of mortality risk. Additionally, a focus on single specific risk factors can result in publication of inflated effect sizes (12).

Several studies on mortality have sought to move beyond these siloed single-factor hypothesis testing approaches to incorporate independent predictors from across domains. For example, Ganna and Ingelsson (13) investigated 655 health, demographic, and lifestyle predictors of 5-year mortality in nearly 500,000 adults in the United Kingdom. These risk factors included circulating blood biomarkers, anthropometrics, health and medical histories, sociodemographics, early life health factors, family history, psychosocial factors, and health behaviors. Following an examination of the 655 factors in independent analyses, proxy measures of health itself, as measured through self-reported health, recent morbidities, disability, medication use, and walking pace were, not surprisingly, the strongest predictors of all-cause mortality as they capture the most proximal underlying pathophysiology preceding death. In contrast, when those with recent illnesses were excluded, smoking emerged as an important predictor of mortality.

Ganna and Ingelsson's investigation highlights the importance of an approach that incorporates factors from across different domains to

determine the strongest predictors of mortality. However, the inclusion of proximal health factors obstructs the discovery of behavioral and social factors from across the lifespan. A simultaneous investigation focused on predictors in the economic, behavioral, social, and psychological domains from across the life course will help advance our understanding of the importance of these factors as earlier, longer-term predictors of mortality (14). Such an investigation may potentially reveal important transdisciplinary life course contributors to mortality to inform future investigations, similar to Genome-Wide Association Studies (GWAS) (15, 16) and Environment-Wide Association Studies (EWAS) (17-19) that have revealed specific genetic factors or environmental toxins that are predictive of longevity, ill health, and mortality. Importantly, restricting our analyses to behavioral and social factors from across the life course, including childhood, establishes the importance of factors within these domains to include in future studies of the exposome (20) or in studies, such as Ganna and Ingelsson, which traverse the behavioral and social with the biological and medical sciences.

Comprehensively ranking independent predictors of mortality from across the behavioral and social sciences requires unique data that includes the most relevant potential risk factors. We used data from adults aged 50 years and older who participated in the U.S. Health and Retirement Study (HRS). The HRS is a nationally representative study of adults with biannual data collection since its inception in 1992. Adults were aged 52 to 104 at the time the exposures were measured between the years of 1992 and 2008,

with six years of follow-up for mortality. 13,611 adults are included in the current analysis.

The HRS is not only unique in its comprehensive investigation of socioeconomic conditions, but also in its inclusion of several factors from the behavioral and social sciences, including early life factors, social relationships, and psychological characteristics. In the current investigation, we use three complementary analytical approaches to determine the relative contribution to prospective mortality of 57 independent predictors measured from years 1992 to 2008 selected from six commonly investigated domains in the behavioral and social sciences: (1) adverse socioeconomic and psychosocial experiences during childhood, and (2) socioeconomic conditions, (3) health behaviors, (4) social connections, (5) psychological characteristics, and (6) adverse experiences during adulthood. We further explore how each of these 6 identified domains, and the combination across all 57 factors, predict mortality.

First, we examine the contribution of each individual predictor, regardless of domain, to time to death using Cox regression models. Within the Cox regression framework, we also explore the proportion of the variance explained by each domain alone to mortality and the proportion explained by all 57 factors when included in the same analysis. Our analysis also capitalizes on advances in statistical science that have paved the way for data-driven approaches that simultaneously analyze a large number of predictors and their interactions (2, 21). Our analyses include these data-



driven approaches, namely the least absolute shrinkage and selection operator (Lasso) regression (22) and random forest survival analysis (RFSA) (23). Both of these approaches have contributed to understanding novel interdisciplinary and interactive pathways to health and mortality, while minimizing prediction error, such as over- or underestimating the predictive power of each factor, that can occur with traditional Cox regression models or survival decision trees when using data from individual datasets. Finally, we completed a replication of our primary Cox regression analyses in an independent dataset using another United States cohort of participants, the Midlife in the United States (MIDUS) Study. While not nationally representative, MIDUS offers the most comprehensive set of variables in the United States that best match the social and behavioral factors considered here in our primary study population (HRS).

## Results

**Participants.** The average age of participants was 69.3 years (SD = 9.7) with the majority women (58.6%), White (77.6%), and born in the United States (91.0%).

**Descriptive statistics.** Table S1 presents descriptive statistics for all 57 variables included in the study.

### **Main Results.**

#### ***Estimation of mortality risk for each independent factor***

We used Cox regression to estimate the excess mortality risk of having a particular level of exposure to a risk factor compared to no exposure in 57

independent models adjusted for the following demographic characteristics (corresponding hazard ratio): male gender (HR=1.28, 95%C.I. = 1.23, 1.33), race/ethnicity (HR<sub>Black</sub>=1.22, 95%C.I. = 1.13, 1.31); HR<sub>Other Race</sub>=1.08, 95%C.I. = 0.94, 1.24); HR<sub>Hispanic</sub>=1.02, 95%C.I. = 0.92, 1.12) and whether the individual was foreign-born (HR=0.87, 95%C.I. = 0.80, 0.95). Table 1 and Figure 1 shows the individual hazard ratios and confidence intervals for each predictor ranked from strongest to weakest association with mortality over the study period (2008 to 2014). Point estimates for the hazard ratios are represented by dots; the line widths for each predictor present the 95% confidence intervals adjusted for multiple comparisons. Confidence intervals in our study that do not include 1 are considered statistically significant at the 95% level.

The ten factors associated with the greatest risk of mortality over the study period were current or previous history as a smoker (HR = 1.91, 95% C.I. = 1.70, 2.14 and HR = 1.32, 95% C.I. = 1.22, 1.43, respectively), history of divorce (HR = 1.44, 95% C.I. = 1.31, 1.60), history of alcohol abuse (HR = 1.36, 95% C.I. = 1.14, 1.61), recent financial difficulties (HR = 1.32, 95% C.I. = 1.22, 1.43), history of unemployment (HR = 1.32, 95% C.I. = 1.10, 1.59), lower life satisfaction (HR = 1.31, 95% C.I. = 1.19, 1.45), never married (HR = 1.30, 95% C.I. = 1.03, 1.63), history of food stamps (HR = 1.28, 95% C.I. = 1.09, 1.49), and negative affectivity (HR = 1.23, 95% C.I. = 1.14, 1.33). Of the 57 predictors, 42 had confidence intervals that did not include 1, substantiating many previous *a priori* studies on these individual factors.

In order to estimate heterogeneity of associations by race, gender, education, or age group, we completed a series of follow-up analyses with interactions between each of these factors with each of the 57 factors considered here. We found little evidence that the strength of prediction differed by gender (Table S2), between White and non-White participants (Table S3), or those who completed high school and those who did not (Table S4). Many significant associations were apparent for individuals younger than 75 years old, whereas for those 75 years and older, significant associations were sparse (Table S5). Results were consistent when we censored participants who died within two years of 2008, where illness prior to death may be more likely to affect the level of the exposure (Table S6).

***Estimation of mortality risk for each domain and all domains combined***

Next, we calculated the proportion of variance explained from Cox regression models including (1) all demographic characteristics combined, (2) demographic characteristics and specific domains, and (3) one final model with all 57 variables. Estimates are conditional on age as the baseline hazard, but proportion of variance explained does not include age.

Demographic characteristics of gender, race/ethnicity, and place of birth explained 1.9% of the variation in mortality risk over the study period, and each domain independently provided an additional 1.1-4.7% increase in predictive power, with the smallest difference for childhood economic and psychosocial adverse experiences and largest difference for health

behaviors. Social connections, such as objective markers of marital status and subjective reports of positive and negative relationships with family and friends, ranked second in the increase in predictive power for mortality (3.1%). These were followed by psychological characteristics (2.6% increase), adulthood socioeconomic conditions (2.3% increase), and adulthood adverse experiences (1.3% increase). All variables together predicted an additional 9.5% of the variance in mortality in older adults above that of demographic characteristics, suggesting mostly unique variance in mortality accounted for by each domain.

As each domain included a different number of variables, we also ran principal component analyses within each domain and selected the top 3 principal components within each domain to predict mortality. In this approach, each domain predicted an additional 0.7% to 4.4% of variance, maintaining the same rank of lowest to greatest proportion of variance explained. **Specifically, childhood factors ranked lowest (0.7% increase), followed by adulthood adverse experiences (1.2% increase), adulthood socioeconomic conditions (1.3% increase), psychological characteristics (1.3% increase), social connections (1.9% increase), and finally behavioral factors (4.4% increase) which ranked highest.** Combined across all domains, all principal components explained an additional 7.2% of the variance in mortality.

### ***Estimation of mortality using Lasso and Random Forest Survival Analysis***

For a second complementary approach, we implemented a regression algorithm that allows a large number of predictors to be fit in a single model – lasso regression (22). As an alternative to prediction with several variables at a time, lasso allows an analysis of all 57 risk factors and demographic factors simultaneously. In order to produce robust estimates, the algorithm estimates hazard ratios that are shrunk closer to the null, with cross validation used to produce the best fitting parsimonious model. We present in Figure 2 hazard ratios estimated from lasso survival analysis (22). The 10 strongest behavioral and social predictors of mortality were largely similar to the results from the independent Cox survival models, even though all factors were included in the same model. The additional top 10 most predictive factors identified with the Lasso model were negative interactions with children, which was also a strong predictor in the Cox survival models, and childhood psychosocial adversity.

Finally, we used a random forest algorithm to determine if there were any additional factors identified through this approach, which additionally demonstrated that household wealth was a leading predictor of mortality (see Figure S1).

### ***Replication in the MIDUS Study***

Thirty-nine of the 57 variables across the 6 domains were able to be matched between the MIDUS Study and the HRS (see Table S7). The replication of the unweighted Cox regression models in the MIDUS study sample demonstrated generally similar unweighted hazard ratios and

overlapping confidence intervals across 85% of factors from the HRS (**see Table S8**), without accounting for any corrections for multiple comparisons. **Additionally, for the 10 highest ranked predictors in the HRS in this refined set of 39 variables, 5 directly matched the 10 highest ranked in MIDUS, including past or current smoking status, alcohol abuse, history of employment, and negative affectivity. Two additional factors in the HRS's highest ranked, life satisfaction and recent financial difficulties, are closely linked conceptually to hopelessness and wealth which were among MIDUS's highest ranked.** There were some exceptions **to the overlapping confidence intervals**, however, with a lack of overlap in confidence intervals in 15% (n = 6) of the 39 factors, including history of divorce, history of renting, smoking status, lower positive interactions with friends, hopelessness, and lower neuroticism (see Figure 3 for HRs and confidence intervals from both studies). Of these, the hazard ratios for smoking, divorce, and neuroticism were higher in the HRS whereas the other factors (i.e. maternal education, history of renting, positive interactions with friends and hopelessness) were higher in MIDUS, although the direction of association only differed meaningfully for neuroticism. Methods for the replication are included in the Supplementary Materials.

## **Discussion**

In the present study, factors from across the behavioral and social sciences were explored as independent predictors of mortality, and many were discovered as especially important regardless of the applied statistical

approach. These included, behaviors (e.g. smoking, alcohol abuse, physical inactivity), financial wellbeing (e.g. reported financial difficulties or lower wealth), social experiences (e.g. divorce, negative interactions with children, discrimination), and psychological characteristics (e.g. trait negative affectivity, lower life satisfaction). In addition, we found that each overall domain contributed substantially to explaining the variance in time to death. Because of different strengths of the complementary analytic approaches, some predictors were better identified with a particular approach. While adverse psychosocial experiences in childhood were associated with mortality in independent Cox regression analysis, but its importance ranked higher in the Lasso model. Conversely, differences were apparent for discrimination, lower life satisfaction, and sleep problems, although important when examined independently, they reduced in strength when considered in a model simultaneously with recent financial difficulties, health behaviors, negative social interactions with children, and early psychosocial adversity. These findings provide evidence to support the suggestion that other prospective and national cohort studies should widen the net that is cast when testing behavioral and social factors by including these and related measures of social experiences and psychological characteristics from across the life course.

There were also some individual variables that were not predictive of mortality across domains in the individual Cox regression analyses and machine learning algorithms. For example, childhood economic difficulties

were unrelated with any approach. Adverse psychosocial experiences in adulthood, including for example, death of a child, being in combat, or experiencing a natural disaster, were also unrelated to mortality in the three statistical approaches. Finally, in the current study of older adults, religiosity was unrelated to mortality in any of the analyses.

There were some findings that ran counter to those in the literature. In the Lasso regression analysis, older adults whose fathers were unemployed when they were children or whose mothers had lower education levels were more likely to live longer. These results may result from a cohort effect for adults born in the 1930s, during the Great Depression, and whose fathers were sent to Europe to fight in World War II. As adults in the HRS continue to age, we can potentially examine the effects of childhood socioeconomic conditions as a function of the different generations of children included in the study. Another finding that seems counterintuitive in the Lasso regression is that increased adversity in adulthood was related to reduced risk of mortality. Some research suggests that moderate levels of psychosocial adversity may predict lower distress across the lifespan compared to no adversity or high adversity (24). In the current investigation, we did not develop our aims to examine the curvilinear relationship between our selected factors, including adulthood adversity, with mortality. Our current findings provide fruitful grounds for future explorations. These counterintuitive results may also be due to mortality selection.



There are several factors not included in the current study that were either not measured in the HRS or purposefully excluded from our analyses. The HRS does not have data on dietary consumption, nor measures of environmental contaminants, which have been the focus of prior prediction models of mortality (4, 18). We excluded genetic and health markers because our primary focus was to examine predictors typically examined in the behavioral and social sciences (12). A recent investigation by Liu and colleagues (25) using data from the HRS determined that genetic factors, health behaviors, and life course circumstances accounted for one-third of the variation of a novel summary marker of health calculated from 9 biomarkers. Importantly, there was also a significant gene-environment interaction by which cardiovascular disease polygenic risk scores and socioeconomic disadvantage compounded accelerated aging. While the study did not use similar measures to ours - we included social connection with family and friends and personality characteristics - Liu's study demonstrates the added information that can be gleaned when including genetic markers in studies of health and mortality. Liu and colleagues' study also highlights that each investigated domain accounts for significant variability in the selected outcome, similar to our investigation.

Future studies should develop lifespan models that include genetic, biological, and health markers, in addition to the individual level factors considered here (e.g. economic, behavioral, social, and psychological) and more macrolevel markers (e.g. built, natural, and social environments,

including structural racism, neighborhood cohesion, policies at state/provincial and national levels) that have previously been demonstrated as determinants of health and mortality. Social trajectory theories (26–28) highlight the pathways through which **social and economic** experiences early in life and through adulthood foster health habit formation, economic attainment, and psychological characteristics that ultimately shape the healthspan and life expectancy. **While our investigation did not demonstrate that early economic experiences impact mortality later in life, social trajectory research is clear on the importance of early economic experiences in laying the ground for economic attainment, health habits, and personality in midlife and later in life. Importantly, those adults who emerged from poor economic conditions earlier in life may have died at earlier ages and, resultingly, may not have been included in the HRS or our analyses with the HRS data (i.e. selective mortality).** Lifespan modeling using longitudinal prospective studies that **started collecting data with younger participants than the HRS and that includes genetic, built, and political environment measures would allow for testing the role of early social and economic conditions in laying the foundation for social and economic attainment in adulthood, health habits, and personality formation.**

Our study has several limitations. It is important to note that our study design does not allow for causal interpretations. Although we can rank which

factors best predict mortality from across disciplines, we cannot assert that modifying levels of these factors would change an individual's mortality risk. The utility of our approach is the ranking of multiple factors from across the disciplines which can expand future considerations of what types of

5 predictors should be more thoroughly investigated, and is intended as a means of hypothesis generation for future observational and clinical studies. Second, our measures of adulthood psychosocial adversities were limited in scope and number, and do not include the full breadth and number of events that are often measured by life event scales (31). As noted elsewhere, other

10 relevant adversities in adulthood (e.g. food insecurity, domestic abuse) were not included in the HRS (32). Third, there may be some concern that the reporting of childhood events is prone to reporting and recall error (33, 34), as seen in previous studies where early abuse experiences are

15 underreported (35, 36). Hardt and colleagues (36, 37) have noted, however, that when adverse experiences are clearly defined, as in the case in our report, reporting bias should have little effect in studies. Nationally

20 representative studies have supported Hardt's conclusion, demonstrating that rates and health impacts of childhood health, economic, and traumatic experiences reported as adults are similar to prospective cohort studies that commenced in childhood (38-40). Our results are specific to six years of follow-up time for mortality, and the factors we identify could differ with longer follow-up time.

We replicated our findings in the national Midlife in the United States Study, which was selected for its breadth of psychological factors, its location (i.e. the U.S.), the capacity to compare similar age groups, and access to linked mortality data. There may also be utility in comparing how factors across the behavioral and social sciences predict mortality compared to morbidity or physical functioning assessed with objective data collected in electronic health records. Replication should be further considered with nationally representative datasets from across the globe, including the U.K. Biobank and the International Sister Studies of which the HRS is just one.

These nationally representative sister studies have been conducted in Brazil, China, Costa Rica, England, Europe, India, Indonesia, Ireland, Japan, Korea, Malaysia, Mexico, Northern Ireland, South Africa, Scotland, and Thailand and collected data from these studies have been harmonized through the effort to create the Gateway to Global Aging Data platform. Establishing global and country-specific behavioral and social predictors of mortality across these developed and developing countries may expand our understanding of the challenges faced globally by an aging population as well as clarify unique challenges facing specific countries.

The purpose of our investigation was to draw attention to specific early life and adulthood socioeconomic, behavioral, social, and psychological factors that may impact mortality in a nationally representative and prospective study in the United States. Our goal was to expand the types of factors that current and future investigations could consider in their

methodologies in order to obtain a more comprehensive understanding of the impact of behavioral and social factors on health and mortality. Similar investigations with other national studies may help consolidate around a critical set of behavioral, economic, social, and psychological factors to include in future studies.

## **Methods**

### Predictors.

Fifty-seven predictors measured from years 1992 to 2008 across the following domains were measured: (1) adverse socioeconomic and psychosocial experiences during childhood, (2) socioeconomic conditions, (3) health behaviors, (4) social connections, (5) psychological characteristics, and (6) adverse experiences during adulthood. For a full list of the 57 variables in the current analysis, see Table 1. See Supplementary Materials and Methods for further details of the measures used, and how data were processed for the current analysis.

### Outcome.

Mortality records in the HRS were linked to National Death Index (NDI) through 2011. Beyond 2011, mortality information was ascertained through exit interviews obtained from a family member of the respondent. For more information see (50).

### **Data Source.**

All data is publicly available and can be retrieved upon request from <https://hrs.isr.umich.edu/data-products/access-to-public-data>.

## Statistical Analysis

Means and standard deviations (SDs) were calculated for each continuous variable, and totals and percent for each categorical variable. Correlations among all risk factors are presented in Supplementary Figure S2. The analyses were weighted to account for the survey design of the Health and Retirement Study. The outcome of interest across our analyses was time to death. Multivariate Cox regression analyses (51) were used to examine the contribution of each predictor to the hazard of death, with age as the time unit. Continuous variables were standardized. Binary variables were coded -1 and 1, and categorical variables with three categories -1, 0, and 1. These were implemented using the 'survival' package in R (52). For Figure 1, predictors were examined independently in multivariate cox models adjusted for gender, race, ethnicity, and migrant status. We estimated cluster-robust standard errors by household due to the sampling of respondents and their spouses in the HRS. We obtained standardized hazard ratios (HRs) and 95% confidence intervals for each predictor which were corrected for multiple comparisons using Bonferroni correction for 57 comparisons (53). We also completed follow-up, sensitivity multivariate Cox regression analyses excluding those who died within two years from 2008 and with an interaction between each predictor and gender (male versus females, racial identification (White versus non-White), educational status (completed high school or not), and age (<75 versus 75+).

Next, to estimate the proportion of variance explained by each predictor domain, we estimated a multivariate Cox regression model for each domain independently, adjusted for sex, race, ethnicity, and migrant status, and obtained the pseudo R-squared for each domain. **We used exploratory**

5 **Principal Components analyses to reduce the dimensionality of our data. This process was completed in two steps. First, the imputePCA function from the *missMDA* package was used to impute any missing values. Then, for each category of variable, the PCA function from *FactoMineR* was used to identify the dimensions within that**

10 **category. The three dimensions which explained the greatest amount of variance were used to in the multivariate Cox regression models for each domain independently and the domains combines.**

Next, we performed Lasso regression (22) in order to fit a prediction model which selects the best fitting parsimonious model given a large

15 number of potential predictors of mortality. Given the time to event nature of the data, we fit our models using the `glmnet()` function in R with `family="cox"`. Lasso is a generalized linear model that is fit with penalized maximum likelihood and is appropriate when correlations are not high enough to cause any issues with multicollinearity, as is the case in the

20 current study. It was developed to provide a variable selection approach which has the advantage, as compared to backward or forward selection models, of penalizing coefficients rather than completely dropping or adding variables to the model, even though variables can be set to zero if they do

not add sufficiently to the model (54). We used cross-validation with the `cv.glmnet` function to select a model with the minimum mean cross-validation error.

5 Finally, we used random forest survival analysis to predict survival over the survey period, using the domain specific predictors and age as the baseline hazard. The random forest algorithm is a non-parametric, ensemble machine learning tool first introduced by Breiman (55) as an extension of classification and regression trees (CART) and bagging.

10 The random forest algorithm works by repeatedly drawing bootstrap samples from the original sample and a random selection of predictors to grow a predetermined number of decision trees across which results are pooled. A training data set consisting of  $n$  of  $N$  cases (two thirds of the original sample) is generated for each of  $k$  decision trees and the remaining cases (one third of the original sample) are used as test data to estimate the out of bag (OOB) classification error. A random sample  $m$  of  $M$  predictors is  
15 selected at each node and the one predictor that best discriminates discrepancies in survival is chosen for that particular split. As a result, the root node of each decision tree represents the strongest predictor and the splits that follow are based on the successively strongest predictors. A final  
20 classification is made using a majority of votes across all trees.

We implemented random forest survival analysis with all predictors included in a single model. Classification was performed using the `'rfsrc'` function from the R package `'randomForestSRC v2.8.0'` (56, 57) with a



maximum of five iterations—each with 500 classification trees. We obtained variable importance plots for each predictor. Missing values were imputed using missForest (58).

We evaluated the extent to which each variable contributes to predicting survival using the metric of variable importance, using the ‘vimp’ function from the R package ‘randomForestSRC v2.8.0’ (56, 57). We calculate variable importance using random permutation of the variable approach. This is done by comparing the accuracy of prediction of each tree as estimated from the model with the tree with each particular variable permuted. A summary of these differences across all of the trees is used to calculate variable importance (59). The figure shows positive variable importance in blue, which indicates this factor increases the predictive nature of the model. Negative variable importance is shown in red, indicating these variables decrease the accuracy of prediction. This is likely due to random error. Note that the variable importance scale is quite low, meaning that based on the random forest model with baseline hazard of age, these variables are not highly predictive of survival relative to age in this model.

Code for data analysis is provided in the Supplement Dataset.

## References

1. M. Mather, L. Jacobsen, K. Pollard, "Aging in the United States," *Population Bulletin* 70, no. 2 (2015).
- 5 2. P. L. Mabry, D. H. Olster, G. D. Morgan, D. B. Abrams, Interdisciplinarity and systems science to improve population health: a view from the NIH Office of Behavioral and Social Sciences Research. *Am. J. Prev. Med.* **35**, S211-24 (2008).
3. V. Kontis, *et al.*, Future life expectancy in 35 industrialised countries:  
10 projections with a Bayesian model ensemble. *Lancet* **389**, 1323–1335 (2017).
4. J. M. McGinnis, W. H. Foege, Actual Causes of Death in the United States. *JAMA J. Am. Med. Assoc.* **270**, 2207 (1993).
5. A. H. Mokdad, J. S. Marks, D. F. Stroup, J. L. Gerberding, Actual Causes  
15 of Death in the United States, 2000. *JAMA* **291**, 1238 (2004).
6. D. W. Brown, *et al.*, Adverse Childhood Experiences and the Risk of Premature Mortality. *Am. J. Prev. Med.* **37**, 389–396 (2009).
7. A. C. Carlsson, *et al.*, Financial stress in late adulthood and diverse risks  
of incident cardiovascular disease and all-cause mortality in women and  
20 men. *BMC Public Health* **14**, 17 (2014).
8. J. Holt-Lunstad, T. B. Smith, J. B. Layton, Social Relationships and Mortality Risk: A Meta-analytic Review. *PLoS Med.* **7**, e1000316 (2010).
9. M. Jokela, *et al.*, Personality and All-Cause Mortality: Individual-Participant Meta-Analysis of 3,947 Deaths in 76,150 Adults. *Am. J.*

*Epidemiol.* **178**, 667–675 (2013).

10. C. R. Gale, *et al.*, When Is Higher Neuroticism Protective Against Death? Findings From UK Biobank. *Psychol. Sci.* **28**, 1345–1357 (2017).

11. L. L. Barnes, *et al.*, Perceived discrimination and mortality in a  
5 population-based study of older adults. *Am. J. Public Health* **98**, 1241–7  
(2008).

12. D. Fanelli, J. P. A. Ioannidis, US studies may overestimate effect sizes in softer research. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15031–6 (2013).

13. A. Ganna, E. Ingelsson, 5 year mortality predictors in 498 103 UK  
10 Biobank participants: a prospective population-based study. *Lancet*  
**386**, 533–540 (2015).

14. C. J. Patel, J. P. A. Ioannidis, Studying the elusive environment in large scale. *JAMA* **311**, 2173–4 (2014).

15. L. C. Pilling, *et al.*, Human longevity: 25 genetic loci associated in  
15 389,166 UK biobank participants. *Aging (Albany. NY)*. **9**, 2504–2520  
(2017).

16. A. Ganna, *et al.*, Genetic determinants of mortality. Can findings from genome-wide association studies explain variation in human mortality? *Hum. Genet.* **132**, 553–561 (2013).

20 17. C. J. Patel, J. Bhattacharya, A. J. Butte, An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One* **5** (2010).

18. C. J. Patel, *et al.*, Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United

States national health and nutrition examination survey. *Int. J. Epidemiol.* **42**, 1795–810 (2013).

19. X. Zhuang, *et al.*, Environment-wide association study to identify novel factors associated with peripheral arterial disease: Evidence from the National Health and Nutrition Examination Survey (1999–2004).

5

*Atherosclerosis* **269**, 172–177 (2018).

20. G. M. Buck Louis, R. Sundaram, Exposome: time for transformative research. *Stat. Med.* **31**, 2569–2575 (2012).

21. M. M. Glymour, T. L. Osypuk, D. H. Rehkopf, Invited Commentary: Off-Roading With Social Epidemiology--Exploration, Causation, Translation.

10

*Am. J. Epidemiol.* **178**, 858–863 (2013).

22. R. Tibshirani, Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).

23. S. Aichele, P. Rabbitt, P. Ghisletta, Think Fast, Feel Fine, Live Long.

15

*Psychol. Sci.* **27**, 518–529 (2016).

24. M. D. Seery, E. A. Holman, R. C. Silver, Whatever Does Not Kill Us: Cumulative Lifetime Adversity, Vulnerability, and Resilience. *J. Pers. Soc. Psychol.* **99**, 1025–1041 (2010).

25. Z. Liu, *et al.*, Associations of genetics, behaviors, and life course circumstances with a novel aging and healthspan measure: Evidence from the Health and Retirement Study. *PLOS Med.* **16**, e1002827 (2019).

20

26. D. Umberson, R. Crosnoe, C. Reczek, Social Relationships and Health

- Behavior Across Life Course. *Annu. Rev. Sociol.* **36**, 139–157 (2010).
27. S. Liu, R. N. Jones, M. M. Glymour, Implications of Lifecourse Epidemiology for Research on Determinants of Adult Disease. *Public Health Rev.* **32**, 489–511 (2010).
- 5 28. L. F. Berkman, Social Epidemiology: Social Determinants of Health in the United States: Are We Losing Ground? *Annu. Rev. Public Health* **30**, 27–41 (2009).
29. S. Oishi, J. Graham, Social Ecology. *Perspect. Psychol. Sci.* **5**, 356–377 (2010).
- 10 30. O. Solar, A. Irwin, World Health Organization, “A conceptual framework for action on the social determinants of health Paper 2 (Policy and practice).” (2010).
31. G. Brown, T. Harris, “Life events and measurement.” in *Life Events and Illness.*, G. Brown, T. Harris, Eds. (The Guilford Press, 1989), pp. 3–45.
- 15 32. E. Puterman, *et al.*, Lifespan adversity and later adulthood telomere length in the nationally representative US Health and Retirement Study. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6335–E6342 (2016).
33. B. Maughan, M. Rutter, Retrospective reporting of childhood adversity: Issues in assessing long-term recall. *J. Pers. Disord.* **11**, 19–33 (1997).
- 20 34. S. R. Dube, D. F. Williamson, T. Thompson, V. J. Felitti, R. F. Anda, Assessing the reliability of retrospective reports of adverse childhood experiences among adult HMO members attending a primary care clinic. *Child Abuse Negl.* **28**, 729–37 (2004).

35. K. MacDonald, *et al.*, Minimization of Childhood Maltreatment Is Common and Consequential: Results from a Large, Multinational Sample Using the Childhood Trauma Questionnaire. *PLoS One* **11**, e0146058 (2016).
- 5 36. J. Hardt, M. Rutter, Validity of adult retrospective reports of adverse childhood experiences: review of the evidence. *J. Child Psychol. Psychiatry* **45**, 260–273 (2004).
37. J. Hardt, A. Sidor, M. Bracko, U. T. Egle, Reliability of retrospective assessments of childhood experiences in Germany. *J. Nerv. Ment. Dis.* **194**, 676–83 (2006).
- 10 38. J. P. Smith, Reconstructing childhood health histories. *Demography* **46**, 387–403 (2009).
39. E. Havari, F. Mazzonna, Can We Trust Older People’s Statements on Their Childhood Circumstances? Evidence from SHARELIFE. *Eur. J. Popul.* **31**, 233–257 (2015).
- 15 40. J. Hardt, P. Vellaisamy, I. Schoon, Sequelae of prospective versus retrospective reports of adverse childhood experiences. *Psychol. Rep.* **107**, 425–40 (2010).
41. E. C. Donny, *et al.*, Randomized Trial of Reduced-Nicotine Standards for Cigarettes. *N. Engl. J. Med.* **373**, 1340–1349 (2015).
- 20 42. M. C. Fiore, Tobacco Control in the Obama Era — Substantial Progress, Remaining Challenges. *N. Engl. J. Med.* **375**, 1410–1412 (2016).
43. H. W. Kohl, *et al.*, The pandemic of physical inactivity: global action for

public health. *Lancet* **380**, 294–305 (2012).

44. K. W. Davidson, T. McGinn, Screening for Social Determinants of Health. *JAMA* **322**, 1037 (2019).

45. N. Lurie, T. Dubowitz, Health Disparities and Access to Health. *JAMA* **297**, 1118 (2007).

46. J. Stiel, S. Menendian, S. Gambhir, “Responding to Rising Inequality: Policy Brief” (2014).

47. X. Ramos Salas, *et al.*, Addressing weight bias and discrimination: moving beyond raising awareness to creating change. *Obes. Rev.* **18**, 1323–1335 (2017).

48. D. R. Williams, J. A. Lawrence, B. A. Davis, Racism and Health: Evidence and Needed Research. *Annu. Rev. Public Health* **40**, annurev-publhealth-040218-043750 (2019).

49. M. L. Hatzenbuehler, Structural stigma: Research evidence and implications for psychological science. *Am. Psychol.* **71**, 742–751 (2016).

50. D. Weir, “Validating mortality ascertainment in the Health and Retirement Study” (2016).

51. D. R. Cox, “Regression Models and Life-Tables” in (Springer, New York, NY, 1992), pp. 527–541.

52. T. Therneau, T. L.-S. analysis P. on CRAN, undefined 2014, Package “survival.” *tbrieder.org* (August 5, 2019).

53. R. J. SIMES, An improved Bonferroni procedure for multiple tests of

significance. *Biometrika* **73**, 751-754 (1986).

54. T. Hastie, R. Tibshirani, M. Wainwright, *Statistical learning with Sparsity: the lasso and generalizations* (August 5, 2019).

55. L. Breiman, Random Forests. *Mach. Learn.* **45**, 5-32 (2001).

5 56. C. Strobl, T. Hothorn, Zeileis A, Party on! A New, Conditional Variable Importance Measure for Random Forests Available in the party Package (2009) (April 24, 2017).

57. H. Ishwaran, U. Kogalur, M. Kogalur, "Package 'randomForestSRC'" (2019).

10 58. D. J. Stekhoven, P. Buhlmann, MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112-118 (2012).

59. H. Ishwaran, Variable importance in binary regression trees and forests. *Electron. J. Stat.* **1**, 519-537 (2007).

60. J. P. Shaffer, Multiple Hypothesis Testing. *Annu. Rev. Psychol.* **46**, 561-15 584 (1995).



**Acknowledgments:** The HRS is supported by the National Institute on Aging (NIA; U01 AG009740) and the Social Security Administration. The Midlife in the United States (MIDUS) investigation was supported by National Institute of Aging (NIA) Grants P01-AG020166 and R01-AG019239. The original study was supported by the John D. and Catherine T. MacArthur Foundation Research Network on Successful Midlife Development. The funding sources had no involvement in the study design; data collection, analysis, or interpretation; nor the writing and submission of this article.

This research was supported by the Canada Research Chairs program, (EP), and the Population Research Training Grant (NIH T32 HD007242) awarded to the Population Studies Center at the University of Pennsylvania by the National Institutes of Health's (NIH)'s Eunice Kennedy Shriver National Institute of Child Health and Human Development (JW). DHR was supported by the National Institute on Aging at the National Institutes of Health (K01AG047280). This work was supported by the National Institute on Aging at the National Institutes of Health with a grant to the Stress Measurement Network (R24AG048024).

**Author contributions:** EP conceptualized the investigation, received funds to support data management, developed the list of included variables, prepared the original draft of the manuscript and supported data visualization. JW curated and prepared the data for analysis, contributed to the R code used for analysis, completed the analysis under the supervision of

EP and DHR, prepared all tables and figures, reviewed and edited the original draft, and wrote the original draft of the supplementary materials. AG and DK helped to curate the data for analysis and reviewed and edited the original draft. WBM received funding from the NIH to support this work,

5 conceptualized the investigation and reviewed and edited the original draft.

DHR conceptualized the investigation, developed the appropriate data management and analytic procedures to complete the statistical approach including the R code used for analysis, conceptualized the figures, and helped write the original draft.

10 **Competing interests:** Authors declare no competing interests.

**Fig. 1.** Independent Cox regression hazard ratios of each predictor for mortality. Confidence intervals that include 1 indicate that a predictor is not statistically significant at the 5% level, corrected for multiple tests using the Bonferroni Method (60). Age is used as the baseline hazard. Predictor coefficients on the right of the line increase risk of mortality, and predictor coefficients on the left decrease risk of mortality.

**Fig. 2.** Hazard ratios from time to event Lasso regression. Age is used as the baseline hazard. All 57 factors were used to fit the model, along with demographic factors of male gender, black race, other race and Hispanic ethnicity. Only hazard ratios that differ from 0 in the final model are presented in the figure.

**Fig. 3.** Independent Cox regression hazard ratios and confidence intervals (unweighted) of each predictor for mortality in the HRS (red) and MIDUS (blue) studies with 39 harmonized variables from the 6 domains. Results are Age is used as the baseline hazard and are not corrected for multiple tests. Predictor coefficients on the right of the line increase risk of mortality, and predictor coefficients on the left decrease risk of mortality.

Figure 1.

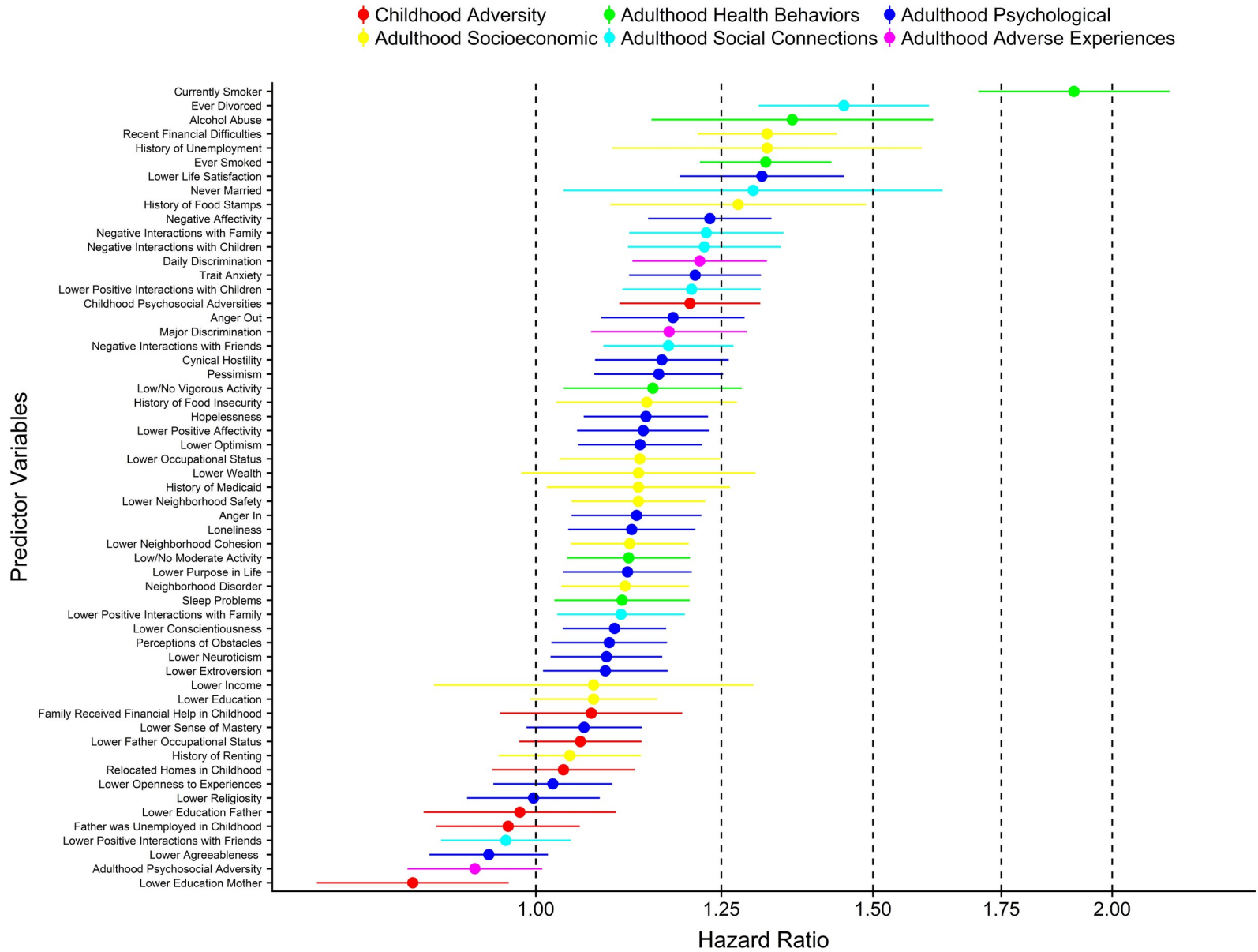


Figure 2.

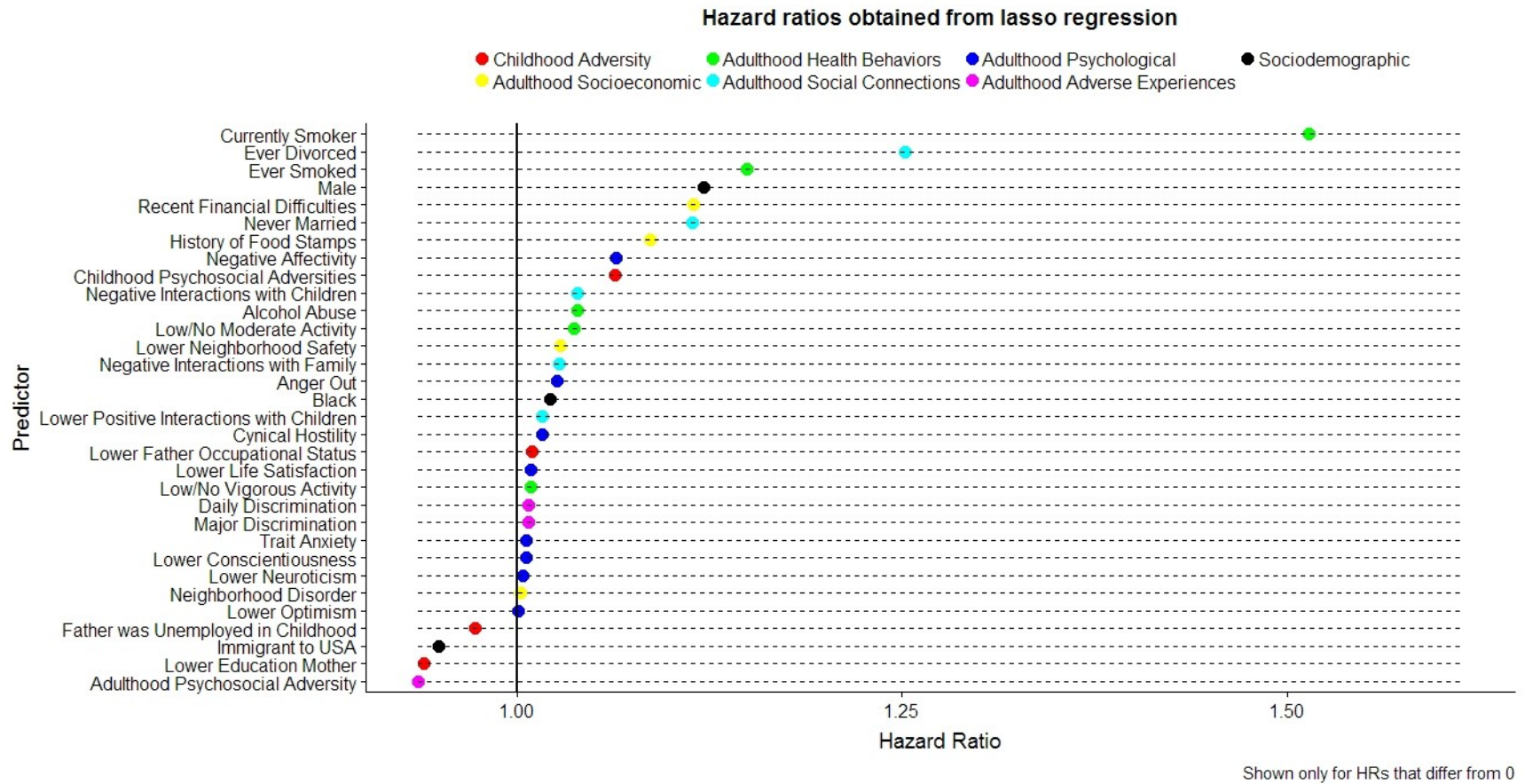
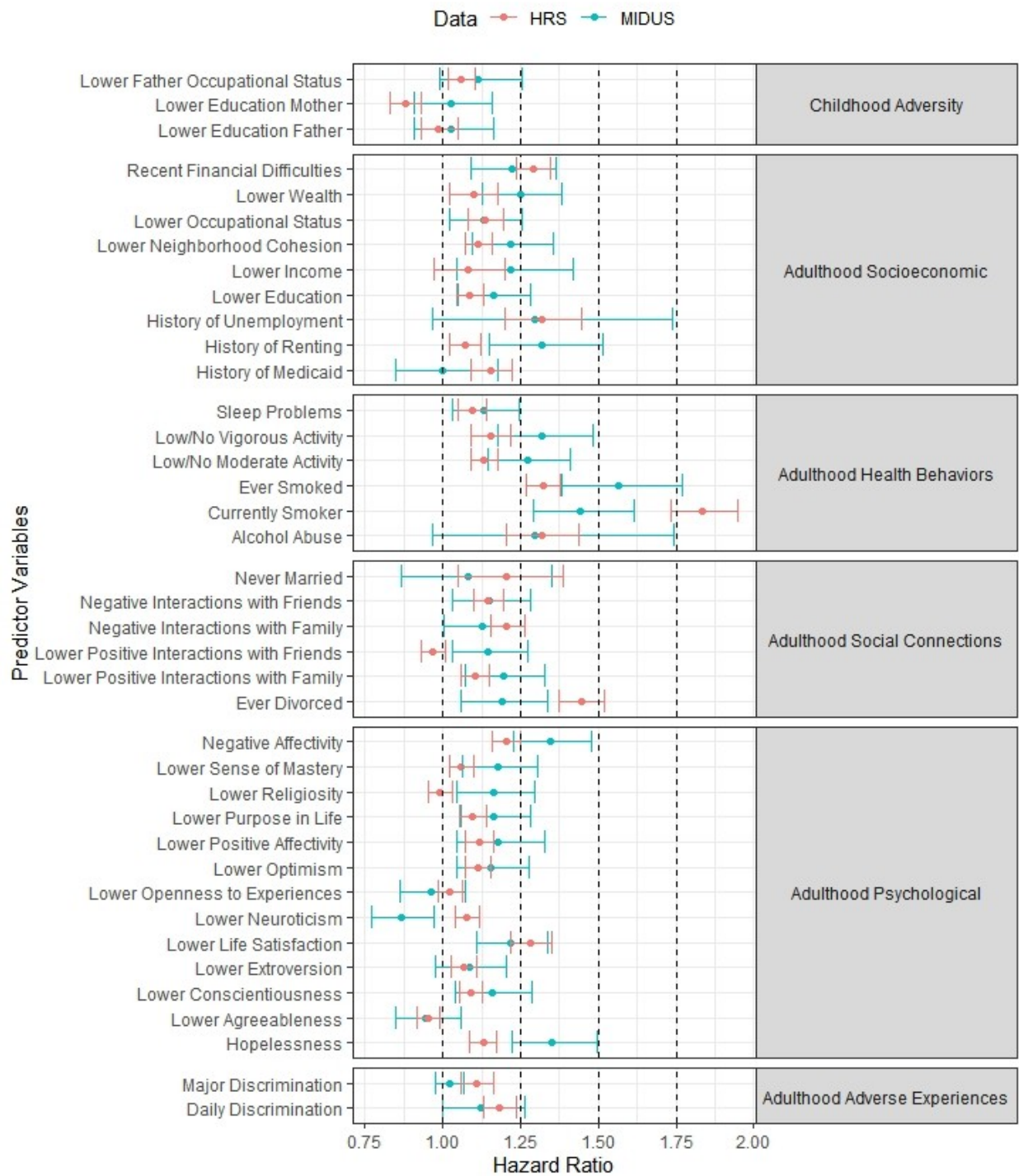


Figure 3.



**Table 1.** Hazard ratios and 95% confidence intervals for each predictor ranked from strongest to weakest association with mortality over the study period

Variable	HR	95% CI
Current Smoker	1.91	(1.70, 2.14)
History of Divorce	1.45	(1.31, 1.60)
Alcohol Abuse	1.36	(1.15, 1.61)
Recent Financial Difficulties	1.32	(1.22, 1.44)
History of Unemployment	1.32	(1.10, 1.59)
History of Smoking	1.32	(1.22, 1.43)
Lower Life Satisfaction	1.31	(1.19, 1.45)
Never Married	1.30	(1.03, 1.63)
History of Food Stamps	1.28	(1.09, 1.49)
Negative Affectivity	1.23	(1.14, 1.33)
Negative Interactions with Family	1.23	(1.12, 1.35)
Negative Interactions with Children	1.22	(1.12, 1.34)
Daily Discrimination	1.22	(1.12, 1.32)
Trait Anxiety	1.21	(1.12, 1.31)
Lower Positive Interactions with	1.21	(1.11, 1.31)
Children		
Childhood Psychosocial Adversities	1.20	(1.11, 1.31)
Anger Out	1.18	(1.08, 1.28)
Major Discrimination	1.17	(1.07, 1.29)
Negative Interactions with Friends	1.17	(1.08, 1.27)
Cynical Hostility	1.16	(1.07, 1.26)
Pessimism	1.16	(1.07, 1.25)
Low/No Vigorous Activity	1.15	(1.03, 1.28)
History of Food Insecurity	1.14	(1.02, 1.27)
Hopelessness	1.14	(1.06, 1.23)
Lower Positive Affectivity	1.14	(1.05, 1.23)
Lower Optimism	1.13	(1.05, 1.22)
Lower Occupational Status	1.13	(1.03, 1.25)
Lower Wealth	1.13	(0.98, 1.30)
History of Medicaid	1.13	(1.01, 1.26)
Lower Neighborhood Safety	1.13	(1.04, 1.23)
Anger In	1.13	(1.04, 1.22)
Loneliness	1.12	(1.04, 1.21)
Lower Neighborhood Cohesion	1.12	(1.04, 1.20)
Low/No Moderate Activity	1.12	(1.04, 1.20)
Lower Purpose in Life	1.12	(1.03, 1.21)
Neighborhood Disorder	1.11	(1.03, 1.20)
Sleep Problems	1.11	(1.02, 1.20)
Lower Positive Interactions with	1.11	(1.03, 1.20)
Family		



Lower Conscientiousness	1.10	(1.03, 1.17)
Perceptions of Obstacles	1.09	(1.02, 1.17)
Lower Neuroticism	1.09	(1.02, 1.16)
Lower Extroversion	1.09	(1.01, 1.17)
Lower Income	1.07	(0.88, 1.30)
Lower Education	1.07	(0.99, 1.16)
Family Received Financial Help in	1.07	(0.96, 1.19)
Childhood		
Lower Sense of Mastery	1.06	(0.99, 1.14)
Lower Father Occupational Status	1.06	(0.98, 1.14)
History of Renting	1.04	(0.96, 1.14)
Relocated Homes in Childhood	1.03	(0.95, 1.13)
Lower Openness to Experiences	1.02	(0.95, 1.10)
Lower Religiosity	1.00	(0.92, 1.08)
Lower Education Father	0.98	(0.87, 1.10)
Father was Unemployed in Childhood	0.97	(0.89, 1.05)
Lower Positive Interactions with	0.96	(0.89, 1.04)
Friends		
Lower Agreeableness	0.94	(0.88, 1.02)
Adulthood Psychosocial Adversity	0.93	(0.86, 1.01)
Lower Education Mother	0.86	(0.77, 0.97)