Pathways to disability and death before disability in mid-ages: Estimates from the Health and Retirement Study Data

Lakshmi K. Raut[1]

2020-09-10

Table of Contents

Acknowledgement and disclaimer

This paper is dedicated in loving memory of my younger brother, Bishnu Pada Raut, who passed away in New Delhi on February 1, 2019, from lung cancer. He never smoked, never drank, and had normal BMI, CES-D and other standard biomarkers (personally observed)

[1] Social Security Administration,Office of Policy, Lakshmi.Raut@ssa.gov

throughout his life. Why are there incidence of diseases and death at premature ages? Scientific community is actively exploring the answers to these questions and the ways to improve life. This paper is an inquiry in this vein. An earlier draft was presented at the 2019 Annual Conference of the Society for Government Economists, April 5, 2019, Washington, DC. I got many useful comments from the discussant, Elizabeth Bass at Congressional Budget Office, and from the audience. I had many insightful comments from Han Altae-Tran at MIT, John Phillips at NIH, and Robert V. Gesumaria, Javier Meseguer, David Pattison, and Mark Sarney at SSA. David Pattison's detailed insightful comments on an earlier draft helped greatly the preparation of this draft. Thanks.

Disclaimer: The views, thoughts, and opinions expressed in the paper belong solely to the author, and do not necessarily represent the views of the Social Security Administration or the United States Government, other group or individual.

Abstract

This paper studies factors that affect enrollment onto a public insurance program such as the Social Security's DI (Disability Insurance) program and SSI (Supplemental Security Income) program, and the competing risk of death before disability enrollment by age 65. As individuals age, or misuse drugs, alcohols or intake less nutrient foods, the homeostatic regulatory mechanism that controls physiological body systems such as respiratory, cardiovascular, neuroendocrine, immune, and metabolic becomes more and more frail in its ability to face internal and external stressors. This leads to one-or-comorbid chronic diseases (such as diabetes, cancer, vascular, musculoskeletal, cognitive and mental disorders), DI-qualifying disabilities or death. The speed of progression through these health states depends on the rate of depletion of one's homeostatic health and frailty levels. Genetic and epigenetic factors comprised of internal and external environments, health care use, health related behavior and cognitive endowments modulate the depletion rate of homeostatic health over life cycle. These, in turn, determine the likelihood of various pathways through the health states. I use the Health and Retirement Study (HRS) data to estimate a multi-state time-to-event model of pathways through these health states. I use bio-markers such as BMI, CES-D, measures of cognitive health, and indicators of health related behaviors such as smoking, exercising and use of preventative care along the life-course as indicators of latent homeostatic health and frailty levels and how they affect the risks of following various pathways to disability or death before reaching age 65.

JEL Classifications: I12, C41, C51.

Keywords: Pathways to disability, OASDI, SSI, multistate time-to-event model, mortality, aging.

Introduction

Identification of factors that determine disability and mortality incident rates is important for disability programs such as Social Security Disability Insurance (DI) and the Supplemental Security Income (SSI) programs. According to the biology of living organisms, all individuals succumb to aging, and experience diseases and disabilities of various kinds as they age. Diseases and disabilities can also be caused by injuries, genetic abnormalities and epigenetic

reprogramming (epigenetic includes environmental factors and health-related individual behaviors). Some individuals stay in good health for a long period and then become disabled or die; some develop one or more chronic diseases such as diabetes, cancer, vascular, musculoskeletal, cognitive and mental disorders that expedite incidence of disability and death. I use the Health and Retirement Study (HRS) data to estimate a dynamic multi-state time-to-event econometric model of pathways to disability or to death before disability through various health states — specifically, normal health and one-or-more chronic diseases — before reaching age 65 for individuals in their early 50's. Genetic and environmental factors, health care use, health related behaviors and cognitive factors determine the progression of unobserved stock of internal health (also known as health-capital in economics, and frailty in gerontology). The state of internal health determines the risks of transitions to other health states and their transit times. I estimate the effects of these factors on the probabilities of transitions and the transit times along the pathways that individuals in their 50's follow before reaching age 65.

Before exploring pathways to disability, I must clarify the definition of disability that I study in this paper. The definition of disability depends on the purpose of its use. Disability is a multidimensional concept and is defined in the literature using simple descriptions, conceptual models, classification schemes, and measurement methods (see for details, (Hahn 1985; Marks 1997, 1997; Altman 2001, 2001; Albrecht and Verbrugge 2003; Snyder et al. 2008)). I use the following statutory definition of disability that the Social Security Administration use for the DI and SSI programs (specified in the Social Security Act, Title II, § 223(d), paragraph (1)A):

> "inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months"

and with a vocational grid addendum stated in paragraph (2)A,

> "An individual shall be determined to be under a disability only if his physical or mental impairment or impairments are of such severity that he is not only unable to do his previous work but cannot, considering his age, education, and work experience, engage in any other kind of substantial gainful work which exists in the national economy, regardless of whether such work exists in the immediate area in which he lives, or whether a specific job vacancy exists for him, or whether he would be hired if he applied for work. For purposes of the preceding sentence (with respect to any individual),"work which exists in the national economy" means work which exists in significant numbers either in the region where such individual lives or in several regions of the country."

The definition of disability for the SSI program is almost identical.[2]

While a lot has been said about the definition of disability, very few papers provide a biological or a behavioral mechanism of disablement process from which policy implications for clinical practice and health care policy can be derived. The first disablement model was introduced by the sociologist Nagi (1965), which he further refined in (Nagi 1976, 1991). This model was extended by (Verbrugge and Jette 1994; Verbrugge, Latham, and Clarke 2017) who

[2] For details, see https://www.ssa.gov/OP_Home/ssact/title02/0223.htm section 223(d)(2) for the OASDI program, and https://www.ssa.gov/OP_Home/ssact/title16b/1614.htm paragraph 3(A) for the SSI program.

added biological, environmental and behavioral risk factors affecting all four stages of the disablement process.[3] Disablement models are conceptual schemes that describe four distinct but related stages to arrive at a disability: starting from a pathology, leading to developments of impairments of body systems, then to functional limitations and finally to disability. I briefly describe these stages below.

Pathology is an interruption of the normal physiological process caused by developmental disorders (such as cerebral palsy, seizure disorders, mental retardation, hearing and vision impairments, autism, PKU, Huntington disease), infection, injury, trauma, metabolic imbalance (such as diabetes), degenerative disease processes (i.e., deterioration over time the functioning or the structure of tissues or organs leading to osteoarthritis, osteoporosis, cancer, Alzheimer or Parkinson's disease) or any other disease process. The impairments of body system involve loss or abnormality of an anatomical, physiological, mental, or emotional nature. Functional limitations include not being able to have one's ADL (activities of daily living) and IADL (instrumental activities of daily living), role activities (such as occupation, parenting, grand-parenting, and student roles), social activities (such as attending church and other group activities, and socializing with friends and relatives) and leisure activities (such as sports and physical recreations, reading, and distinct trips). The final stage is the disability, the definition of which depends on the purpose of the study and involves a combination of all the above models of disability.

The disablement models are useful for conceptualization, diagnosis and record keeping of disabilities but limited for the study of the causes of disability in epidemiological and policy research. The starting point of the above disablement models is an onset of a chronic disease or an injury causing a disability. For policy research on disability and mortality, it is important to study the biomedical processes modulated by genetic, epigenetic and behavioral factors in the manifestation and prognosis of disabling diseases and on the risk of disabling injuries. While an injury as the starting point of disablement process serves well for certain purpose such as for workers' compensation in sports, construction and factories, a large proportion of disabilities in the mid ages are caused by diseases–both physical and mental (see for instance, (Case and Deaton 2015; The US Burden of Disease Collaborators 2018). Mechanism for non-accidental death is similar. Diseases, leading to disabilities — both developmental disabilities and late age disabilities — and to mortality are the result of modulated biomedical processes, which at the microbiology level are the outcomes of cellular aging. While aging, an individual succumbs to diseases and injuries leading to disability or death. Not all individuals experience the same deterministic aging process — some experience faster aging and aging related diseases than others do. Why do people experience faster aging, diseases and mortality? At what stage of life, does it all begin — at mid-ages, at birth, or even earlier at conception? How do various genetic, epigenetic and behavioral factors modulate the aging process, culminating in diseases, disabilities and death? What biomarkers and epigenetic factors (including environmental factors

[3] Nagi model has been adapted by the World Health Organization in their classification scheme of disability, the latest one is World Health Organization (2001). See Bedirhan et al. (2010) for an application of the above disablement models in WHO's 2001 classification system, and see (Pope and Tarlov 1991; Institute of Medicine 2007) for more on this.

and individual health related behaviors) predict better the process of aging and incidence of disease, disability and death over the lifespan?

I will not get into the details of the biomedical literature on these issues. Similar to the literature of behavioral genetics of personality and intelligence, the nature-nurture controversy exists in the health literature: Is it all nature (i.e., all genetics or genome) or is it all nurture (i.e., all epigenetics or epigenome modulated by the environment and health related individual behaviors) that determines the progression of health over the life span of an individual? The consensus so far is that it is neither the nature nor the nurture; it is a combination of the two that determines health developments over one's life span. The research so far found that certain genetic make-ups (i.e., certain sequences of DNA) predispose one to certain diseases, (Barondes 1999; Khoury et al. 2009; Bookman et al. 2011), but the epigenetic inputs — especially at the very early stage of life, i.e. in the womb, but not the least at later stages of life — are also very important determinants of life expectancy and quality of life. The biomedical research so far has not found genes that are responsible for aging and age related diseases, leading to early disability and mortality. The twenty-first century biomedical research emphasizes more on the epigenetic factors than the genetic factors to explain the pattern of health developments over the life span.

At the cellular level, aging means cellular senescence — i.e., after a certain number of cell divisions, it stops dividing or have defective replications, causing tissues or organs to increasingly deteriorate over time. Senescence leads to incidence of degenerative diseases. It is generally observed that women live longer than men and those with better life styles in terms of smoking, exercising and diets delay the aging process (for evidence, Austad and Fischer 2016; Blair et al. 1989; Vaupel 2010; Zarulli et al. 2018). This line of biological inquiry led to explore the (cellular) molecular mechanism of aging process and to find biomarkers of aging that can be used to diagnose, monitor, and improve the age related physiological decline and disease. A good indicator of the aging process at the cellular level is the rate of decay in the telomere length. Telomeres are the caps at the end of chromosomes in a DNA sequence. They look like the plastic caps at the end of shoelaces. The main function of telomeres is to protect cells preserving the genetic content within each chromosome during cell divisions. Unfortunately, the telomere length shortens in the course of each cycle of chromosomal replication during cell division, reaching the Hayflick limit (about 40 to 60 cell divisions, Hayflick (1965)) with a critically short telomere length, after which the cells stop dividing or divide with chromosomal abnormalities. The rate of shortening of the telomere length is modulated by telomerase enzyme. Why the rate of decay in telomere length varies for individuals is an active area of biomedical research and the mechanism for it is not yet fully understood. Many studies find that higher stress of any kind — psychological, financial, social and chemical — is strongly associated with higher oxidative stress, lower level of telomerase enzyme, and shorter telomere length. Furthermore, shorter telomere length is associated with health related phenotypes of poorer health and higher risks for cardiovascular and immune diseases (see, DiLoreto and Murphy 2015; Epel et al. 2004; Shalev and Belsky 2016; Shalev et al. 2013; Simons et al. 2016).

More recently emerged second line of biomedical research on aging and aging related diseases explores the epigenetic (which literally means on top of genetic) mechanism for these life-cycle processes. (See for instance, Alisch et al. 2012; Barres and Zierath 2011; Boks et al. 2009; Esteller 2008; Hannum et al. 2013; Horvath 2013).

The above literature emphasizes that aging and age related diseases are associated with shortening of telomere length and changes in global methylation, and that stress, smoking, drinking, chemical misuse, and diet are important modulators for these changes. The question remains, what are the critical periods or the developmental milestones in life cycle that program the motions of health developments over the life span of an individual?

Research along this line began with the striking findings of Barker (Barker 1990, 1998) and later of Gluckman et al. (2008). They found strong associations between birth weight and many later life chronic diseases, including hypertension, coronary artery diseases, type 2 diabetes, and osteoporosis. Many other studies find that much of health developments in later life is determined very early in life — specifically during the prenatal period, right after conception, i.e. in the womb. Sometimes it is said in social sciences that inequality begins in the womb. The effect of an environmental stress in the womb on later life diseases and developmental outcomes is known as programming. Gluckman et al. (2008) observes that "like the long latency period between an environmental trigger and the onset of certain cancers, the etiology of many later life diseases such as cardiovascular disease, metabolic disease, or osteoporosis originate as early as in the intrauterine development and the influence of environments that created by the mother." For more empirical evidence on the developmental origin of later life diseases, see (Barker 2007; Thornburg et al. 2010). The papers by Barbara, Abdilla, and Calleja-Agius (2017) and Kanherkar, Bhatia-Dey, and Csoka (2014) provide detailed descriptions of the biological process of development of life and health, starting from the conception. They explain how the global DNA demethylation of the fertilized egg right after conception creates an epigenetic — clean slate — to start a new life, followed by rapid remythylation to reprogram the maternal and paternal genomes to create epigenetic configurations in the fetus which rapidly produce specialized cells of the body with cell divisions. The environment provided in mother's womb during those times has long-term effects on the child's later cognitive and other health developments. While inputs at early milestone ages are important for later age health, healthy living and good healthcare are still important for maintaining health in mid ages.

Studies in social sciences find that low socio-economic status (SES) are associated with inflammation, metabolic dysregulation, and various chronic and age-related diseases such as type 2 diabetes, coronary heart disease, stroke, and dementia, and that low SES create epigenetic changes in individuals that lead to faster biological aging even after controlling for health-related behaviors such as diet, exercise, smoking, alcohol consumption, or having health insurance, see for evidence, Simons et al. (2016). The study by Karakus and Patton (2011) uses the Health and Retirement Studies data and after controlling for education, race, income, health risk indicators like BMI and smoking, functional limitations like gross motor index, health limitations for work, and income, they find depression at baseline leads to significantly higher risk for developing diabetes, heart problems, and arthritis and no significant effect on developing cancer during the 12 years follow-up period. Renna (2008) uses National Longitudinal Survey of Youth data to find no significant effect of alcohol use on labor market outcomes such as on earnings or hours of work. Seib et al. (2014) collected data on a sample of older women in Australia and found that severe traumatic life events create strong stress levels that influence them to have unhealthy living and diet measured by BMI and develop stronger and earlier health problems. Conti, Berndt, and Frank (2009) utilize the CES-D data in the Health and Retirement Survey dataset to construct a measure of depression, and find that depression of men and women have significant negative effect on employment status, early retirement, and application for DI/SSI benefits. More

recently, Case and Deaton (2015) found a racial reversal in the mortality rates of the US mid-age population between 1993 and 2013. They found that all-cause mortality and morbidity of non-Hispanic white men and women of ages 45-55 have been increasing during the period, mainly due to increases in their incidence rates of drug and alcohol poisoning, suicide, chronic liver diseases and cirrhosis. Morbidity of the group culminate into serious disabilities and crowding into DI and SSI rolls and to lower labor force participation rates, especially among women. Such time reversals are confined to that age and racial group only, and the rates are higher for less educated than educated groups. They attribute such behavioral changes to increased (within and inter-generational) income inequality and rises in prescription of pain killer drugs and opioid, and falling price and easier availability of heroin.

I adopt all the above views and formulate a statistical model of disablement process. I postulate that as individuals age or misuse drugs, alcohols or intake less nutrient foods, the homeostatic regulatory mechanism that controls physiological systems respiratory, cardiovascular, neuroendocrine, immune, and metabolic becomes more and more fragile in its ability to face internal and external stressors, leading to early occurrence of disease, disability and the death. I draw from the microbiology literature that study the genetic and epigenetic mechanism for aging and timing and severity of aging related diseases, disability and death. I use available bio-markers (such as BMI, CES-D, cognition) and health related behaviors such as smoking, exercising and using preventative care along the life-course to explain how they affect the risk of chronic diseases, disabilities and premature death. I use a multi-state time to event statistical framework to estimate the effects of these factors on the probabilities of following various pathways through normal health, diseases, disability or death before reaching age 65. The multi-state framework is more useful to study the effect of various covariates — the covariates that are specific to intermediate health states — on the risk of becoming disabled or being dead.

The rest of the paper is organized as follows. In Section 2, I provide an extended disablement model of this paper. In Section 3, I describe the subset of the Health and Retirement Survey dataset that I use for this study. In Section 5, I describe the econometric specification, estimation issues, present the estimates, and discuss the results. Section 5concludes the paper.

The Model

With insights from the disablement modeling literature and the biomedical literature on aging process, I formulate and then estimate an econometric model of paths to enter disability rolls. An individual can be on the disability rolls if the individual has a qualifying disability and has not reached age 65 and has not died before applying for disability benefits. I assume that an individual's getting on the disability rolls is a terminal event, i.e., the individual does not move to the normal or the diseased health state from the disability health state.[4] After reaching this state, the individual is not followed any further. A competing risk for getting on the rolls is death

[4] The focus is on the first time entitlement onto a disability program. A few people, however, recover and move to normal or diseased health state, but more likely they come back later to the disability rolls, see Raut (2017) for details on some of these probabilities.

before age 65. This is a competing risk because an individual cannot be at risk for disability enrollment if the individual is already dead and thus not at risk to get on the disability rolls. In the technical terms defined below, I treat health states — disability and death — as absorbing states, i.e., once in that health state, an individual remains in that health state and is removed from the sample for later considerations.

An individual can follow many possible health paths. For instance, beginning with a normal health state, an individual can become disabled or die before becoming disabled after some passage of time. Or the individual may first become diseased with one or more diseases, or start from the beginning at this health state and after some passage of time become disabled or die before becoming disabled. There are many possible paths that an individual can follow. Even when the health states they pass through are the same, the duration of stay in each health state (also known as the *waiting time* in stochastic process literature) could vary. From the diagram below one can see various health paths that an individual may traverse. The focus of the paper is to study the probabilities of various transitions and the duration of stay in each health state.

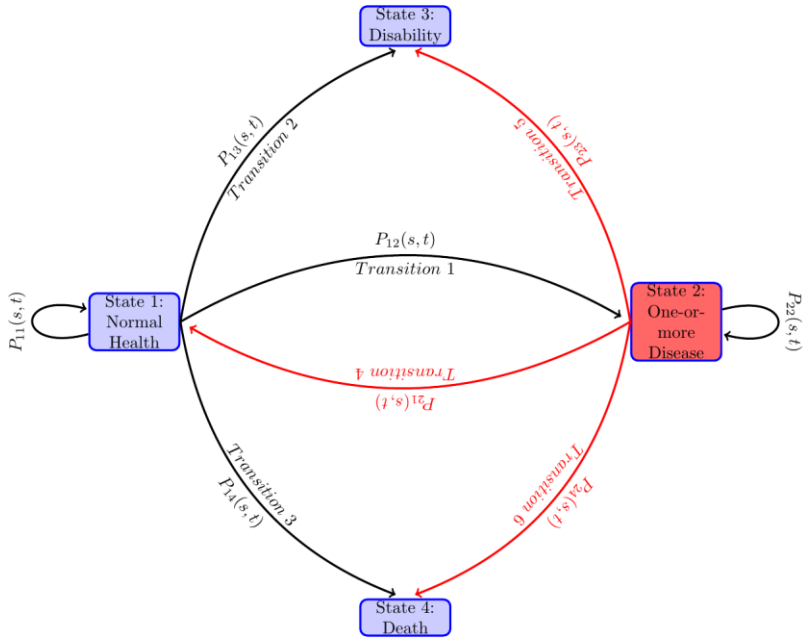

Figure 1: Path diagram of health trajectors.

Each configuration of visited states and waiting times in those visited states constitute one path. When time is continuous, the number of paths that one can follow is infinite. For an individual, one path maybe more likely than another. The likelihood or risk of following a particular path may depend on the individual's genetic make-up and prior health conditions and health related behaviors. Various factors affect individual risks of various transitions to different health states and the time they stay in each health state along their life-spans. Both, in turn, determine the timing of getting on to the disability rolls.

The goal is to estimate the probabilities of getting onto the disability rolls or die before disability sequentially over time as an individual progresses over life experiencing health shocks or changing one's health related behaviors. For instance, it will be important to get estimate of the probability of individuals in normal health at age 51 to get on to the disability rolls or die

before disability by age 55, 60, by age 65 (the last age before one can be in disability program) or by any other age? How these probabilities change if the individual from normal health at age 51 becomes diseased with one or more chronic diseases at age 54? Or if the individual starts his life-course with diseased health health state at age 51, how do these probabilities change? How do these probabilities depend on an individual's childhood factors such as socioeconomic status, health and education and how do they depend on race and sex? To get such estimates is important for policy point of view to see what kind of social policies can reduce disability enrollments or death before disability, and also to get quantitative effect of such public policies. To that end, one needs to formulate an appropriate statistical model of the health trajectories, incorporating the effects of various time-varying covariates and estimate it using a nationally representative sample.

I model the paths through various health states that individuals follow along their life-spans as a continuous-time finite-state Markov process $X(t), t \in T$, where at each time point t during the study period T, the random variable $X(t)$ takes a value from a finite set S of health states. In the present study, I take $T = [51,65]$. The state space S of the stochastic process contains health states $1 =$ "healthy or normal health", $2 =$ "diseased with one or more chronic diseases", $3 =$ "disabled with DI-or SSI-qualifying disability" and $4 =$ "death before disability". Sometimes I will use $S = \{h, i, d, D\}$ in place of $\{1,2,3,4\}$.

Let the transition probabilities of our stochastic process $X(t)$ be given by

$$P_{hj}(s,t) = Prob(X(t) = j | X(s) = h),$$

for all $h, j \in S$, $s, t \in T$, $t \geq s$. Denote the matrix of transition probabilities by

$$P(s,t) \equiv \left(P_{hj}(s,t)\right)_{h,j=1\ldots4}.$$

An individual at time t may be in any of the health states in S, the probability of which is known as the *occupation probability*. The occupation probability at time t can be viewed as the proportion of population of age t who are in health state j. Let $\pi_j(t)$ be the occupation probability of an individual in health state j at time t. Denote all the occupation probabilities as a column vector $\pi(t) \equiv (\pi_j)(t), j \in S$. Then the occupation probabilities move over time recursively as follows,

$$\pi(t) = \pi'(s)P(s,t), 0 \leq s < t.$$

Note that given initial distribution $\pi(0)$ and the transition probabilities $\left(P_{hj}(s,t), 0 \leq s < t, s, t \in T\right)$, one can calculate the occupation probabilities in all time periods in T from the above recursive equation.

It is known that the transition probabilities of a stochastic process satisfies the following *Chapman-Kolmogorov* equation

$$P(s,t) = P(s,u) \cdot P(u,t), \text{ for all } s, u, t \in T \text{ with } s < u < t \qquad (1)$$

I assume that the transition probabilities in $P(s,t)$ are absolutely continuous in s and t. A *transition intensity* — also known as the *hazard rate* in the survival analysis literature when exits occur because of one event, and as the *cause-specific hazard rate* in the competing risk analysis[5] when exits occur because of many events — of the health process X_t from health state h to health state j at time t is given by the derivative

$$\begin{aligned}
\lambda_{hj}(t) &= \lim_{\Delta t \to 0} \frac{P_{hj}(t, t+\Delta t) - P_{hj}(t,t)}{\Delta t}, \text{ for } j \in S \\
&= \lim_{\Delta t \to 0} \frac{P_{hj}(t, t+\Delta t)}{\Delta t}, \text{ for } j \neq h \\
\lambda_{hh}(t) &= \lim_{\Delta t \to 0} \frac{P_{hh}(t, t+\Delta t) - 1}{\Delta t}, \text{ for } j = h, \\
&= -\lim_{\Delta t \to 0} \frac{\sum_{j \neq h} P_{hj}\,(t, t+\Delta t)}{\Delta t} \\
&= -\sum_{j \neq h} \lambda_{hj}\,(t)
\end{aligned}$$

For absorbing states $h = 3,4$, the transition intensities $\lambda_{hj}(t) = 0$, for all $j \in S$. Denote the matrix of transition intensities by

$$\Gamma(t) = \begin{pmatrix} -\big(\lambda_{12}(t) + \lambda_{13}(t) + \lambda_{14}(t)\big) & \lambda_{12}(t) & \lambda_{13}(t) & \lambda_{14}(t) \\ \lambda_{21}(t) & -\big(\lambda_{21} + \lambda_{23}\big(t\big) + \lambda_{24}(t)\big) & \lambda_{23}(t) & \lambda_{24}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (2)$$

Notice that while $P_{hj}(s,t)$ is an unconditional probability, the transition intensity or the hazard rate $\lambda_{hj}(t)\Delta t$ is the conditional (instantaneous) probability of an individual experiencing the event j in a small time interval $[t, t+\Delta t)$ given that he has been in state h at time t. This conditional probability may depend on time t and other characteristics of the individual and the path through various health states that is followed by the individual to be in the health state h at time t. I am assuming that the process is Markovian, i.e., it depends only on the health state h that he is in at time s, and on the times s and t, but it does not depend on the path that he followed to come to state h at time s. Or in other words, the future health condition depends on the current health condition but not how one came to the current health state.

It can be shown that the Chapman-Kolmogorov equation leads to the following Kolmogorov forward equation[6]

---

[5] See for instance, Raut (2017) for a competing risk analysis in a similar context using the SSA Administrative data and compare that with the present framework.

[6] There is also a Chapman-Kolmogorov backward equation, which goes backward in time to trace probabilities of various paths leading to a particular state of interest at some time. For our purpose, the forward equation is of interest.

$$\frac{\partial P(s,t)}{\partial t} = P(s,t) \cdot \Gamma(t) \qquad (3)$$

and the initial condition $P(s,s) = I$ for all $s \in T$. Thus, given intensities $\Gamma(t)$, the solution $P(s,t)$ of the above differential equation defines a continuous time Markov chain and conversely, given a set of absolutely continuous transition probabilities $P(s,t)$ for a continuous time Markov chain, one has the Kolmogorov forward Equation (3).

Given an intensity matrix function $\Gamma(t)$, from the Fundamental Theorem of ordinary differential equations, we know that there exists a unique solution $P(s,t)$ to the system of ordinary differential equations in Equation (3). It is, however, not possible to find analytic solution of the Kolmogorov forward equation without further restrictions on the nature of transitions. One can use a numerical method to solve the Kolmogorov Forward Equation with an estimate of the intensity matrix function $\hat{\Gamma}(t)$.[7] No statistical distribution theory is readily available to compute standard errors of the estimated transition probabilities. Generalizing the classic product limit procedure of Kaplan and Meier (1958) for the estimation of survival function in survival analysis, the statistical literature introduced the notion of *Product Limit* to solve and transition probabilities and study the statistical distribution theory of the estimates using Martingale theory. I follow this approach in this paper and describe the procedure briefly in the section 3.

Econometric specifications and Estimation Methods

First, I introduce a few concepts briefly. An *integrated transition intensity function* $\Lambda_{hj}(t)$ for a transition $h \to j$ is defined by $\Lambda_{hj}(t) = \int_0^t \lambda_{hj}(u)du$. Let the time interval $[s,t]$ be subdivided into a partition of m sub-intervals with cut-off points $s = t_0 < t_1 < \ldots < t_m = t$. Applying repeatedly the Chapman-Kolmogorov Equation (1) on the sub-intervals of the partition, we have

$$P(s,t) = P(t_0,t_1) \cdot P(t_1,t_2) \cdot \ldots \cdot P(t_{m-1},t_m) = \prod_{i=1}^{m} P(t_{i-1},t_i) \qquad (4)$$

Note that as $|t_{i-1} - t_i| \to 0$, the transition probability matrix $P(t_{i-1},t_i) \to P(t_{i-1},(t_{i-1} + dt) = I + \Gamma(t)dt$.[8] With finer subdivisions of the interval $[s,t]$ such that the maximum length of sub intervals tends to 0, the right hand side of Equation (4) converges to a matrix called the *the*

[7] I have given analytic solution in Raut (2019) under the assumption that individuals once become chronically ill never return to normal health, and also I have computed the transition probabilities based on the estimated hazard matrix.

[8] From the definition of transition intensity above and writing it in the matrix form, we have $\Gamma(t) = \lim_{\Delta t \downarrow 0} \frac{P(t,t+\Delta t) - P(t,t)}{\Delta t} = \lim_{\Delta t \downarrow 0} \frac{P(t,t+\Delta t) - I}{\Delta t}$. From this it follows that for small Δt, we have $P(t,t+\Delta t) = I + \Gamma(t)\Delta t$.

product integral[9] of the integrated hazard functions $\Lambda(s,t)$, which is denoted as $\widetilde{\prod}_s^t\big(I + d\Lambda(u)\big)$. Or in other words, the transition probabilities of a stochastic process parameterized via an intensity process is given by the product integral of integrated hazard functions.

$$P(s,t) = \widetilde{\prod}_s^t(I + d\Lambda(u)). \qquad (5)$$

The above product-integral solution is a generalization of the Kaplan-Meier (Kaplan and Meier 1958) product-limit formula for the survival function in survival analysis. The product integral formula unifies both discrete time and continuous time Markov processes. It is an extremely useful apparatus for statistical analysis of Markov processes. I now describe the statistical methods followed in this paper.

The effects of covariates are incorporated by conditioning transition intensity functions for each trnsition on the covariates process $X(t)$. Denote this as $\Gamma(t; X(t))$. The most widely used statistical procedure is to estimate the transition probabilities $P(s,t)$, $s,t \in T, s < t$ with or without covariates is to plug in an estimate of $\Lambda(u)$ in Equation (5) and then compute the matrix products.

There are broadly two ways to get the estimates — parametric and semi-parametric methods. I will follow the widely used non-parametric Aalen-Johnson-Fleming method via the Nelson-Aalen estimate of each-component transition intensity function in $\Lambda(u; X)$ with Cox proportional hazard model to incorporate the time-varying covariate effects in the next sub-section 4.1.

Aalen-Johansen-Fleming Estimator for Transition Probabilities

Most widely used statistical procedure incorporates the time-varying covariates for the transition probabilities by specifying a semi-parametric functional forms for the intensity hazard functions

$$\lambda_{hj}\big(t; X(t)\big) = \lambda_{hj}^0(t)e^{\beta_{hj}' X(t)}. \qquad (6)$$

In the above specification, $\lambda_{hj}^0(t)$ is known as the *baseline hazard function*. The specification of transition intensity in Equation (6) is known as the *proportional hazard model*. It aggregates the effects of the regressors linearly as a measure of some kind of latent factor, and that latent factor shifts the baseline hazard proportionately, i.e., the effect on hazard is uniform over time. Two papers (Fleming 1978; and Aalen and Johansen 1978) independently extended the Kaplan-Meier nonparametric product limit estimator from survival analysis to the multi-state time to event models. While Fleming gave the estimator for complete data, Aalen and Johansen gave the estimator for censored data. To describe the Aalen-Johansen estimator, I introduce some concepts and notation. For each individual $i, i = 1,2,\ldots,n$ and corresponding to each transient health state, $h, h = 1,2$, define two types of stochastic processes: (1) the counting processes $N_{hj,i}(t)$ denoting the **observed** number of transitions from health state h to health state j that the individual i has made by time t — which in our case is either 0 or 1, since by assumption when

[9] For a more formal treatment of product integral see Gill and Johansen (1990) and for a lucid exposition with some applications, see Gill (2005).

an individual exits a health state, the individual does not return to it in future ; and (2) $Y_{h,i}(t)$, taking value 1 if individual i is at risk at time t for transition to another possible health state, and taking value 0 otherwise.

Let us focus on one transition $h \to j$. Denote by $\overline{N}_{hj}(t) = \sum_i^n N_{hj,i}(t)$, a counting process measuring the number of transitions of the type $h \to j$ in the sample at time t, $\overline{Y}_h(t) = \sum_i^n Y_{h,i}(t)$, a counting process measuring the number of individuals in the sample at risk for a transition at time t, and $\overline{M}_{hj}(t) = \sum_i^n M_{hj,i}(t)$. In any empirical study the data will be at the discrete times, say in ordered times $0 = t_0 < t_1 < \ldots < t_m$. At each time t_i, we calculate

$$\hat{\lambda}_{hj}(t_i) = \frac{\triangle \overline{N}_{hj}(t_i)}{\overline{Y}_h(t_i)}, j \neq h, \qquad (7)$$

Without covariates, the *Nelson-Aalen non-parametric estimate* of the integrated intensity functions is given by, for each $h = 1,2$

$$\begin{aligned} \hat{\Lambda}_{hj}(t) &= \sum_{i:t_i \leq t} \hat{\lambda}_{hj}(t_i), j \neq h, \\ \hat{\Lambda}_{hh}(t) &= -\sum \hat{\Lambda}_{hj}(t) \text{ and} \\ \hat{\Lambda}_{hj}(t) &= 0, \text{ for } h = 3,4; \quad j = 1,2,3,4 \qquad (8) \end{aligned}$$

The *Aalen-Johansen-Fleming estimator* $\hat{P}(s,t), s,t, \in T, s < t$ for the transition probabilities is obtained by substituting for each component hj the Nelson-Aalen estimates $\hat{\Lambda}_{hj}(t)$ and then applying the product integral formula Equation (5) as follows

$$\hat{P}(s,t) = \widetilde{\prod_{s<u<t}} (I + d\hat{\Lambda}(u)) = \widetilde{\prod_{i:t_i \leq t}} (I + [\hat{\Lambda}(t_i) - \hat{\Lambda}(t_{i-1})]). \qquad (9)$$

For this counting process framework, Andersen et al. (1993) derive the following generalized Cox partial likelihood to get parameter estimates of the Cox regression models

$$CL(\beta) = \prod_i \prod_{h=1,2,j=2,3,4,h\neq j} \prod_i \left(\frac{Y_{h,i}(t)\exp(\beta'_{hj} X_{h,i})}{\sum_r \exp(\beta'_{hj} X_{h,r})} \right)^{\triangle N_{hj,i}(t)} \qquad (10)$$

With covariates one obtains the Cox partial likelihood estimate for $\hat{\beta}_{hj}$ for each transition $h \to j$ separately and then computes an weighted risk set defined by

$$\overline{Y}^*_{hj}(t) = \sum_{i=1}^n Y_{hj,i}(t)\exp(\hat{\beta}'_{hj} X^0_{h,i}). \qquad (11)$$

The estimates of cumulative intensities with covariates are obtained from Equation (7) by replacing, $\overline{Y}_h(t)$ with $\overline{Y}^*_{hj}(t)$.

Nelson-Aalen estimator has nice statistical property. For instance, using Martingale calculus, it can be shown that the estimator is asymptotically unbiased. Using the results from Martingale

theory, one can derive the formula for variance-covariance estimates of parameter estimates and the normalized estimate is normally distributed (central limit theorem holds for normalized parameter estimates), see for details, (Aalen, Borgan, and Gjessing 2008; Andersen et al. 1993; Fleming and Harrington 2005).

The Dataset and the Variables

The dataset

I use the Health and Retirement Study (HRS) dataset for empirical analysis. A lot has been reported on the family of HRS datasets — about its structure, purpose, and various modules collecting data on genetics, biomarkers, cognitive functioning, and more, see for instance (Juster and Suzman 1995; Sonnega et al. 2014; Fisher and Ryan 2017). The first survey was conducted in 1992 on a representative sample of individuals living in households i.e., in non-institutionalized, community dwelling, in the United States from the population of cohort born during 1931 to 1941 and their spouses of any age. "The sample was drawn at the household financial unit level using a multistage, national area-clustered probability sample frame. An oversample of Blacks, Hispanics (primarily Mexican Americans), and Florida residents was drawn to increase the sample size of Blacks and Hispanics as well as those who reside in the state of Florida", (Fisher and Ryan 2017).

The number of respondents were 13,593. Since 1992, the survey were repeated every two years, each is referred to as a wave of survey. New cohorts were added in 1993, 1998, 2004 and 2010, ending the survey up with the sample size of 37,495 from around 23,000 households in wave 12 in 2014. The RAND created many variables from the original HRS data for ease of use. I create all the variables (with a few exceptions noted below) from the RAND HRS dataset version P. The details of the Rand HRS version P can be found in Bugliari et al. (2016). I use the original cohort first interviewed in 1992 so that we have a homogeneous group of individuals with data for many years to avoid cohort effects in our analysis. This sample has the largest sample size.

As mentioned in the introduction, I define the disability health state to be the one that qualifies one to be on the disability programs OASDI or SSI. The data on disability is self-reported. Later I plan to use the Social Security Administration's matched administrative data on this variable and earnings variables not included here. The matched data will, however, reduce the sample size to half, as only 50 percent of the respondents are used for matching HRS with SSA Administrative data. The HRS data collected information on if and when the doctor diagnosed that the respondent has any of the severe diseases such as high blood pressure, diabetes, cancer, lung disease, heart attack, stroke, psychiatric disorder and severe arthritis.

I drop respondents who were enrolled on to disability programs before the first survey year 1992 and I also drop the spouses in the sample who were not born between 1931 to 1941, so that the respondents in our sample are between ages 51 to 61 and are not disabled or dead by the first survey year 1992. I ended up with the final sample size of 9601 for this analysis. Table 1 provides summary health statistics of the cohorts in our sample over the survey years.

Table 1: Summary of the health statuses of the individuals in the sample over the survey years.

Survey Year	alive with normal health	alive with disease	became disabled	died before disability	65+, censored	Total
1992	3026	6483	92	0	0	9601
1994	2591	6608	166	144	0	9509
1996	2291	6623	139	146	0	9199
1998	1726	5478	134	123	727	8188
2000	1279	4313	86	113	741	6532
2002	848	3148	54	58	759	4867
2004	468	1893	34	48	781	3224
2006	132	636	4	14	795	1581

The table reports these statistics up to the survey year 2006, as the individuals exited the study because of disability or death before disability or censored because they are over age 65 after this survey year. The table shows that the first period of this study in 1992 has 3026 individuals, which is 32 percent of the sample, in good health, 6483 individuals (i.e., 68 percent) in diseased health state with one-or-more chronic diseases and 92 individuals (i.e., 1 percent) left the study as they become disabled. No individuals died or were censored because of ages higher than 65 — this is the result of sample selection criterion mentioned above. In the next survey year 1994, out of 9509 non-exited individuals, 144 died without any disability. In the survey year 1998 for the first time, 727 individuals in the sample left our study because they reached ages above 65. The total number of individuals during the last survey round of 2006 before they all become older than 65 is 1581, i.e. about 16 percent of the original sample.

Variables

I have noted earlier the importance of the early childhood factors such as childhood socioeconomic status, childhood health status, cognitive and non-cognitive skills in determining the health developments in the mid-ages. Other important factors are biomarkers measuring the initial physical and mental health status in the mid-ages and health related behaviors. Furthermore, the health development may vary by race and sex. I describe the construction of these variables in this subsection.

I use the Item Response Theory (IRT) from the latent variable analysis literature to construct an aggregate measure of childhood socioeconomic status, **cSES**, and two health related behavioral traits, one capturing the propensity for using preventive care, the variable **behav_prev**, and the other one is to measure penchant for drug and alcohol use, the variable **behave_drink**.

IRT techniques are not commonly used in Economics. Originally the IRT techniques were used in the psychometry literature to measure latent traits such as cognitive ability and personality of individuals. More recently this technique has been used in health care fields to measure health status of individuals in clinical trials and treatments. In this procedure, the latent trait, known as *score*, is assumed to be a continuous variable and individuals differ in the levels of its possession. The procedure uses responses on a number of test items usually with true/false or with multiple choices to estimate the level of the latent trait that an individual possesses. The probability of a particular response to an item depends on the individual's trait level and on item

characteristics such as difficulty level to answer objectively a question or the imperfection of the item question to measure the trait, or an individual might be guessing a response. The IRT procedure specifies a probability model of the responses to each item as a function of the level of the latent trait and item characteristics. The procedure uses various statistical methods to estimate the latent trait level and the charactersitcs of the item. Mainly three statistical estimation procedures are used in the literature — the maximum likelihood (ML) procedure, Bayesian maximum a posteriori (MAP) procedure and expected a posteriori (EAP) procedure. I have used a two parameter model (which includes the well known Rasch model as special case) of the probabilities of item responses and the MAP procedure to estimate the individual scores and the set of item parameters. I did this in SAS. See Embretson and Reise (2000) for a lucid exposition of the basic one-dimensional IRT models and the above three estimation procedures, see Cai et al. (2016) for a survey of IRT models of multi dimensional traits and extensions to dynamic scoring, and see An and Yung (2014) for details on the SAS IRT procedure and general introduction to various IRT procedures that SAS can perform.

The demographic variables **White** and **Female** have the standard definition. The variable **College** is a binary variable taking value 1 if the respondent has education level of completed college and above (does not include some college), i.e., has a college degree and more and taking value 0 otherwise.

cesd: I used the score on the Center for Epidemiologic Studies Depression (CESD) measure in various waves that is created by RAND release of the HRS data. RAND creates the score as the sum of five negative indicators minus two positive indicators. "The negative indicators measure whether the Respondent experienced the following sentiments all or most of the time: depression, everything is an effort, sleep is restless, felt alone, felt sad, and could not get going. The positive indicators measure whether the Respondent felt happy and enjoyed life, all or most of the time." I standardize this score by subtracting 4 and dividing 8 to the RAND measure. The wave 1 had different set of questions so it was not reported in RAND HRS. I imputed it to be the first non-missing future CESD score. In the paper, I refer the variable as cesd. Steffick (2000) discusses its validity as a measure of stress and depression.

cogtot: This variable is a measure of cognitive functioning. RAND combined the original HRS scores on cognitive function measure which includes "immediate and delayed word recall, the serial 7s test, counting backwards, naming tasks (e.g., date-naming), and vocabulary questions''. Three of the original HRS cognition summary indices—two indices of scores on 20 and 40 words recall and third is score on the mental status index which is sum of scores ``from counting, naming, and vocabulary tasks"—are added together to create this variable. Again due to non-compatibility with the rest of the waves, the score in the first wave was not reported in the RAND HRS. I have imputed it by taking the first future non-missing value of this variable.

bmi: The variable body-mass-index (BMI) is the standard measure used in the medical field and HRS collected data on this for all individuals. If it is missing in 1992, I impute it with the first future non-missing value for the variable. Following the criterion in the literature, I create the variable bmi taking value 1 if BMI> 25 and value 0 otherwise.

Now I describe the construction of the behavioral variables.

behav_prev: The original HRS surveys starting in 1998 contain responses to a set of questions to capture the respondent's behavior towards preventive care. I used the IRT procedure

on these responses and get the estimated score of each individual, and define the variable behav_prev to take value 1 if the score is above one mean plus one standard deviation of the score, and 0 otherwise.

behav_smoke: This variable is constructed to be a binary variable taking value 1 if the respondent has reported yes to ever smoked question during any of the waves as reported in the RAND HRS data and then repeated the value for all the years.

behav_drink: This variable created using the dynamic IRT on the categorical variables in the RAND HRS reporting the number of days per week the respondent drinks. The data is available from wave 3 (i.e., 1996) onward. Using the same methodology as for the behav_prev described above, I create this binary variable.

behav_vigex: The RAND HRS has data on whether the respondent did vigorous exercise three or more days per week. I created in each time period to be 1 if the respondent did vigorous exercise three or more days per week in any of the waves and then that value is assigned to all the years.

cSES: This variable is a binary variable measuring childhood SES. I constructed it using the IRT procedure as follows. From the HRS data I created four binary variables using the original categorical data on family moved for financial reason, family usually got financial help during childhood, father unemployed during childhood, father's usual occupation during childhood (0 = disadvantaged and 1 = advantaged), and three tertiary variables two on each parent's educational levels (0 = High School dropout, 1 = some college, 2 = completed college and higher) and third on family financial situation (0 = poor, 1 = average, 2 = well-off). I used these seven variables as items in the IRT procedure to first compute a continuous score estimate and then I define **cSES** = 1 if the score is above mean plus one standard deviation of the scores and 0 otherwise. Vable et al. (2017) used Mplus software and other information from the HRS data to create their cSES measure. I have not followed this complex procedure to check robustness of the results this paper. Other studies such as Luo and Waite (2005) used father's and mother's education and family financial well-being directly without aggregating them into a single measure. The qualitative results did not change much when I used these individual variables as measures of childhood SES.

cHLTH is a binary measure of childhood health constructed from the self-reported qualitative childhood health variable in HRS. I define **cHLTH** = 1 if the respondent reported very good or excellent, and zero otherwise.

EstimatedTransition probabilities for various populations

First I report the transition probabilities for the overall population. Transition probabilities are calculated using the *mstate* package.

For the overall population

I have reported the probabilities of each transition by age 60 and 65, starting from age 51 in the path diagram Figure 2.

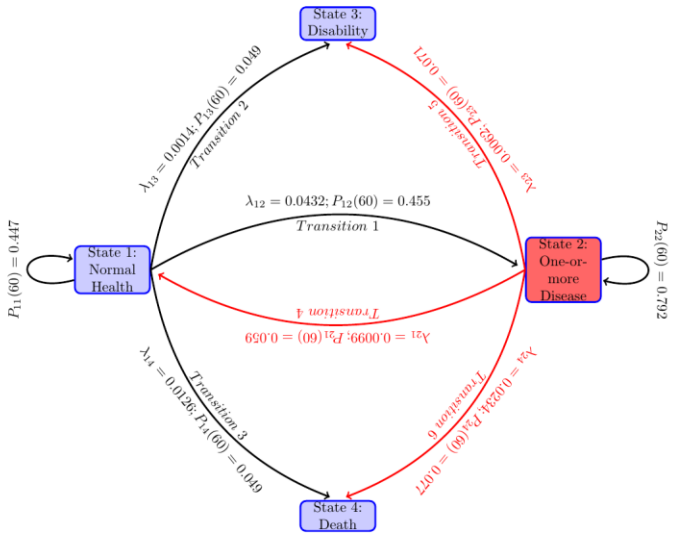

Figure 2: Path diagram of the estimated model.

Table 2 show the transition probabilities. Transition from state i to to j is denoted $i \rightarrow j$, for i=1, 2 and j=1,2,3,4.

Table 2: Estimated transition Probabilities by duration of stay.

Age	$1 \rightarrow 1$	$2 \rightarrow 2$	$1 \rightarrow 2$	$2 \rightarrow 1$	$1 \rightarrow 3$	$2 \rightarrow 3$	$1 \rightarrow 4$	$2 \rightarrow 4$
51	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
52	0.995	0.972	0.000	0.000	0.000	0.017	0.005	0.010
53	0.898	0.953	0.086	0.012	0.005	0.021	0.011	0.014
54	0.837	0.930	0.135	0.016	0.014	0.033	0.013	0.021
55	0.762	0.900	0.202	0.032	0.018	0.041	0.018	0.027
56	0.703	0.877	0.251	0.041	0.025	0.048	0.022	0.034
57	0.635	0.854	0.308	0.050	0.029	0.056	0.028	0.041
58	0.582	0.830	0.347	0.054	0.037	0.068	0.033	0.049
59	0.529	0.805	0.389	0.062	0.044	0.078	0.038	0.055
60	0.489	0.782	0.416	0.067	0.051	0.087	0.044	0.064
61	0.442	0.763	0.448	0.067	0.058	0.096	0.052	0.074
62	0.400	0.744	0.473	0.067	0.068	0.106	0.060	0.084
63	0.362	0.730	0.500	0.068	0.072	0.111	0.066	0.091
64	0.360	0.719	0.493	0.067	0.074	0.114	0.073	0.100
65	0.299	0.711	0.548	0.069	0.075	0.115	0.078	0.105

To get an idea about these transition probabilities over time, I plot transition probabilities out of state 1 in Figure 3, panel (a) and out of state 2 in panel (b).

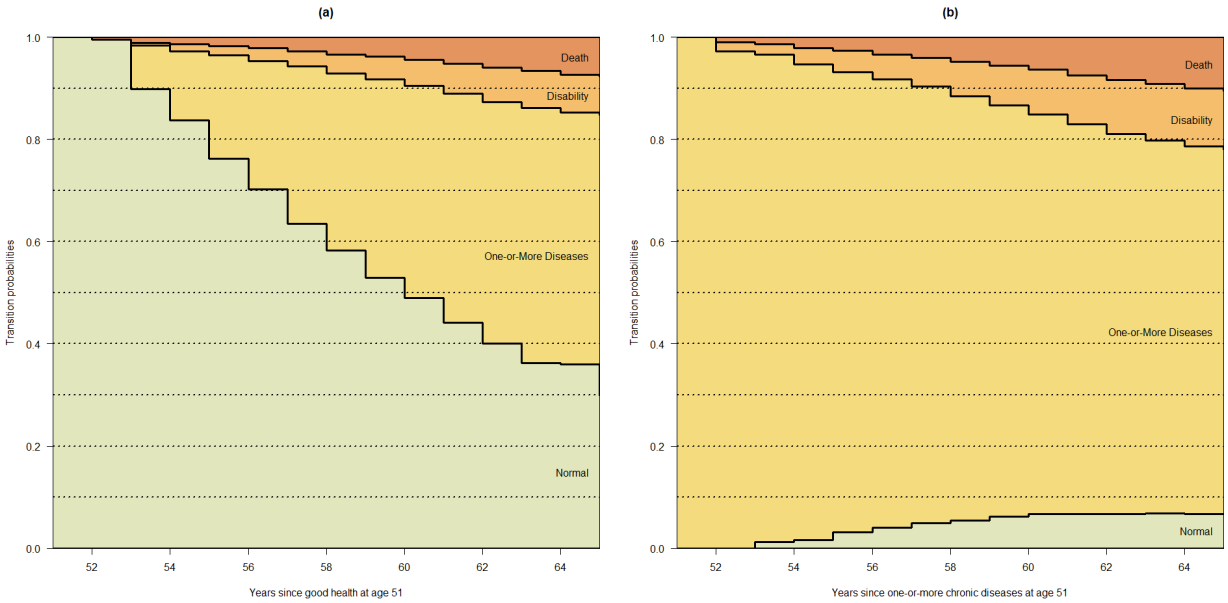

Figure 3: Transition probabilities (a) from normal health state, (b) from one-or-more diseased health state

For various sub populations

Table 3 shows for various groups the parameter estimates and transition probabilities at 10 years, i.e., at age 60 starting, the probability that an individual will exit a health because of disability or mortality before disability.

Table 3: Transition Probabilities by age 60 and 65 for various groups.

group	$P_{13}(60)$	$P_{23}(60)$	$P_{14}(60)$	$P_{24}(60)$	$P_{13}(65)$	$P_{23}(65)$	$P_{14}(65)$	$P_{24}(65)$
Overall	0.051	0.087	0.044	0.064	0.075	0.115	0.078	0.105
White	0.048	0.080	0.039	0.064	0.071	0.105	0.068	0.101
Non-White	0.062	0.113	0.068	0.061	0.093	0.153	0.120	0.120
Female	0.051	0.101	0.039	0.048	0.072	0.126	0.064	0.079
Male	0.049	0.070	0.049	0.083	0.078	0.102	0.093	0.136
College+	0.023	0.041	0.026	0.047	0.032	0.052	0.044	0.073
Non-college	0.058	0.096	0.049	0.067	0.086	0.128	0.086	0.111
cSES:advantaged	0.034	0.048	0.015	0.013	0.054	0.072	0.037	0.041
cSES:disadvantaged	0.055	0.096	0.053	0.075	0.081	0.125	0.089	0.119
cHLTH: good	0.048	0.078	0.006	0.009	0.074	0.108	0.023	0.032
cHLTH: poor	0.054	0.104	0.128	0.162	0.076	0.128	0.189	0.232
BMI: high	0.053	0.088	0.046	0.060	0.081	0.119	0.078	0.099
BMI: low	0.047	0.085	0.042	0.071	0.067	0.107	0.078	0.118

Effects of Childhood SES, Health Status, Educational Attainment on later life health trajectories

We want to examine policies improving social mobility, that is improving child SES helping the parents. We compute polychoric correlations of cSES_R with the component categorical variables to see which variables are most important.

First note that our measure of childhood cSES has polychoric correlations with the component variables (all are ordered categorical variables, with higher values mean better condition) as follows: family financial situation = 0.60, family moved for financial reason = 0.44, family usually got financial help during childhood = 0.35, father unemployed during childhood = 0.49, father's usual occupation during childhood = 0.73, father's education = 0.93, mother's education = 0.85.

From the estimates we see that the most important factors to improve cSES are policies that help a father to have steady jobs during a child's childhood and parents' to have higher education. In the appendix, I have shown for some parents with poor SES, how they can provide better SES for their children through education, steady employment etc.

Logit models of childhood health, College education and initial mid-age health

(#tab:table7) Effects of childhood factors, race and sex on childhood health, college education and initial health in early 50's.

Table 4: Effects of childhood factors, race and sex on childhood health, college education and initial health in early 50's.

Childhood Health

College and above

Initial mid-age health
Intercept
0.7419***
-1.9915***
-0.9649***

(0.0580)
(0.0978)
(0.0702)
White
0.2981***
0.1430
0.2137***

(0.0579)
(0.0831)
(0.0599)
Female
-0.0775
-0.5452***
-0.2140***

(0.0482)
(0.0601)
(0.0461)
Childhood SES
0.3201***
1.0714***
0.1646***

(0.0321)
(0.0404)
(0.0312)
Childhood Health

0.4533***
0.2128***

(0.0738)
(0.0526)
College

0.1157

(0.0621)
AIC
10239.7368
7190.0004
10951.3864
BIC
10268.0358
7225.3741
10993.8349
Log Likelihood
-5115.8684
-3590.0002
-5469.6932
Deviance
10231.7368
7180.0004
10939.3864
Num. obs.
8732
8732
8732
p < 0.001; *p < 0.01;* p < 0.05

Early Childhood Factors and health trajectories in mid-ages

Table 5: Effects of childhood factors, race and sex on health transitions in mid ages.

1 → 2
1 → 3
1 → 4
2 → 1
2 → 3
2 → 4
White
0.0216
-0.0855
-0.6303**
0.3198**
-0.3231***
-0.2330*

(0.0613)
(0.2497)
(0.2235)
(0.1132)
(0.0982)
(0.0939)
Female
0.0321

-0.3773
-0.4106
-0.4071***
-0.0422
-0.6275***

(0.0440)
(0.1997)
(0.2130)
(0.0793)
(0.0883)
(0.0861)
Childhood SES
-0.0959*
-0.1130
-0.5008
0.0049
-0.2014
-0.6573***

(0.0480)
(0.2232)
(0.2988)
(0.0903)
(0.1141)
(0.1303)
Childhood Health
0.2929***
-0.1070
-2.4043***
0.3508***
-0.0554
-1.5831***

(0.0541)
(0.2081)
(0.2780)
(0.0909)
(0.0931)
(0.0927)
College
-0.1031
-0.9697**
-0.4653
0.0722
-0.9722***
-0.2439

(0.0561)
(0.3356)
(0.3536)
(0.1029)
(0.1834)
(0.1452)
AIC
28661.7668
1555.6432
1267.7251
11265.7729
8599.9834
9197.5035
R2
0.0101
0.0044
0.0387
0.0074
0.0080
0.0556
Max. R2
0.9997
0.3535
0.3235
0.7550
0.6590
0.6986
Num. events
1918
105
94
667
512
564
Num. obs.
3580
3580
3580
8045
8045
8045
Missings
1018
1018
1018
134
134
134

PH test
0.0044
0.8809
0.0112
0.3365
0.0028
0.0000
p* < *0.001; *p < 0.01;* p < 0.05. The number in parenthesis below a parameter estimate is its standard error.

For the Cox regression parameter estimates, I have used both the R package **mstate** (see, Wreede, Fiocco, and Putter 2010 for details) and also used the SAS procedure **phreg** (both produced the same estimates) and to estimate all the transition probabilities, I have used the R mstate package (SAS does not have readily available procedure for this purpose). The parameter estimates are shown in Table 8 and Table.

With only demographic covariates (most that can be done with the Administrative data) the parameter estimates in Table 8 show that significantly lower risks of transitions $1 \rightarrow 4$; $2 \rightarrow 3$; $2 \rightarrow 4$ for whites and $1 \rightarrow 4$ and $2 \rightarrow 4$ for women. This may entail that the genetic make-up of being white or female sex yield favorable genetic predisposition to have better health outcomes and longer life. This is a misleading inference as we will see next that when we control for epigenetic factors that the biomedical literature pointed out to have significant effects on aging process and health outcomes, the above effects disappear.

Effects of childhood factors, mid age health and health related behavioral factors on mid age health trajectories

(#tab:table9) Cox regression estimates of the effects of childhood factors, middle age health status and health behaviors.
Table 6: Cox regression estimates of the effects of childhood factors, middle age health status and health behaviors.

$1 \rightarrow 2$
$1 \rightarrow 3$
$1 \rightarrow 4$
$2 \rightarrow 1$
$2 \rightarrow 3$
$2 \rightarrow 4$
White
0.0259
0.2498
-0.5956
0.2610*
-0.1460
-0.1507

(0.0647)
(0.3153)
(0.4154)
(0.1185)

(0.1054)
(0.1426)
Female
0.0611
-0.4111
-0.3106
-0.4576***
-0.0844
-0.3984**

(0.0473)
(0.2327)
(0.4190)
(0.0844)
(0.0984)
(0.1300)
Childhood SES
-0.1273**
0.1401
0.3017
-0.0950
-0.1289
-0.0206

(0.0488)
(0.2428)
(0.4549)
(0.0906)
(0.1177)
(0.1534)
Childhood Health
-0.0027
-0.0297
-1.4293***
0.1852*
-0.0848
-0.5879***

(0.0548)
(0.2493)
(0.3673)
(0.0932)
(0.0985)
(0.1227)
College
-0.0551
-0.6318
-1.1051

0.0517
-0.6942***
-0.4091

(0.0592)
(0.3958)
(0.7715)
(0.1083)
(0.1938)
(0.2236)
CES-D
0.6611***
2.2509***
-0.9044
-0.6775***
1.2337***
0.4866*

(0.1154)
(0.3414)
(1.1577)
(0.2036)
(0.1556)
(0.2259)
Total cognitive scores
0.0022
-0.0808**
-0.0018
0.0096
-0.0291**
0.0104

(0.0054)
(0.0270)
(0.0369)
(0.0099)
(0.0090)
(0.0137)
BMI
0.0378***
0.0263
0.0326
-0.0630***
0.0208*
-0.0016

(0.0053)
(0.0281)

(0.0347)
(0.0095)
(0.0086)
(0.0133)
Behavior: Preventive care
0.3051***
0.7102**
-2.8812**
-0.4377***
0.2483**
-1.4560***

(0.0455)
(0.2224)
(1.0151)
(0.0809)
(0.0939)
(0.1522)
Behavior: Smoking
0.0838
0.2443
2.1072**
0.0240
0.3252**
0.8934***

(0.0471)
(0.2233)
(0.7333)
(0.0870)
(0.1033)
(0.1631)
Behavior: Drinking
0.1141*
-0.2398
0.7305
0.0666
-0.1651
-0.3584**

(0.0568)
(0.2508)
(0.6231)
(0.1006)
(0.1005)
(0.1325)
Behavior: Exercising
-0.0715

-0.9220***
-0.6686
0.6836***
-0.5666***
-0.8770***

(0.0685)
(0.2507)
(0.4244)
(0.1407)
(0.1009)
(0.1273)
AIC
26713.0814
1293.3868
407.9690
10560.9478
7909.5672
4299.1891
R2
0.0440
0.0278
0.0237
0.0258
0.0329
0.0412
Max. R2
0.9998
0.3459
0.1340
0.7771
0.6792
0.4729
Num. events
1824
93
31
639
486
266
Num. obs.
3203
3203
3203
7145
7145
7145
Missings

1395
1395
1395
1034
1034
1034
PH test
0.0463
0.2693
0.0168
0.3266
0.0040
0.0238

p* < *0.001; *p < 0.01;* p < 0.05. The number in parenthesis below a parameter estimate is its standard error.

Table 6 shows the Cox regression coefficient estimates of the effects of various factors on the risk of having transitions $h \rightarrow j$ from health status $h = 1,2$ to health status $j, j = 2,3,4$.

These estimates show that the parameter estimates for the demographic covariates in Table 5 are biased as they are capturing the effects of excluded epigenetic and behavioral factors in that model. After controlling for these epigenetic and behavioral factors, the significance of those effects disappear. Furthermore, women show significantly lower probability of transition from diseased health states onto the disability health state.

Most important factors in Table 6 are cesd, measuring depression and stress and college graduation or higher level of education, with positive effect on all transitions with the exception of no effect on transition from normal health to death. Other important factors are smoking, with significant adverse effect on transitions, and exercising three or more times regularly has significant favorable effect on most transitions. The alcohol use has no significant detrimental effect. Instead it reduces the risk of disability and death for people with diseases.

To understand the quantitative significance of these estimates, say the parameter estimate of -1.0042 in Table 6 for the parameter of $behav\_vigex$ in the model for transition $1 \rightarrow 3$ is that given all other factors equal, an individual of a given age with normal health will have a $0.37 \equiv exp(-1.0042)$ times lower risk of becoming disabled before age 65 if the person does moderately vigorous exercise compared to the probability of one who does not. The corresponding probability of becoming disabled before age 65 from the diseased health state is $0.56 \equiv exp(-0.5811))$ times lower compared to the probability of one who does not. The effect of other parameters can be easily read from the table.

As I have shown above that it is not the genetic make-up of race that is important for the pathways to diseases, disability and to death. The other behavioral and biomedical factors reported in Table 9 are important. To see if the non-college educated non-white group had on the average higher values of the behavioral and biomedical indicators that lead to higher probabilities of diseases, disability and death, I computed these averages for the two groups shown in Table 10. We saw from the estimates in Table 9 that a level of cesd, higher level of of cogtot , lower level of bmi, no smoking and moderate amount of vigorous exercising lower incidence of diseases, disability and death. From Table 10, however, we see that the white has all the good attributes — lower cesd, higher cogtot , lower bmi, higher percent of population doing

moderately vigorous exercising — all leading to lower probabilities of diseases, disability and death. Although they have slightly higher percent smoke and drink, which are detrimental, but these did not offset the other conducive effects mentioned above.

Conclusion

In this paper, I study the determinants of falling chronically ill with one-or-more diseases, the likelihood of getting on the Social Security's DI (Disability Insurance) program or the SSI (Supplemental Security Income) program and the likelihood of dying before becoming disabled by age 65. I use the Health and Retirement Studies (HRS) dataset. I surveyed the biomedical literature to gain insights into the genetic and epigenetic mechanisms at the molecular (i.e., cellular) level for the process of aging and developing age related chronic diseases, disability and death. I view aging as depletion in one's homeostatic regulatory health level that controls physiological body systems such as respiratory, cardiovascular, neuroendocrine, immune, and metabolic. The higher rate of depletion of one's homeostatic health level makes the individual become more and more frail in his/her ability to face internal and external stressors. The depleted level of homeostatic health leads to one-or-more chronic diseases (such as diabetes, cancer, vascular, musculoskeletal, cognitive and mental disorders), and to disability or death. The consensus in the biomedical literature is that much of an individual's later life health outcomes is programmed at an early stage of life as early as the prenatal stage, most importantly right after the conception stage. The programming is strongly modulated by the epigentic inputs created by the environment in mother's womb. The genetic predisposition also matters. But epigentic factors modulate quite strongly the programming for later life developments in cognitive and non-cognitive health. The most important epigenetic factor is stress of any kind psychological, financial, social and chemical and other significant factors are diet, smoking, substance use, and exercising. These modulating factors are important throughout life, with stronger effects imparted in early stages of life.

I used a multi-state time-to-event model to estimate the effects of the above epigenetic factors (that include health related behaviors), demographic factors, education level (taken as college graduated or not). I use biomarkers like BMI, CES-D and cognition scores as noisy measurements of internal homeostatic health level, depression and stress level. I then study their effects on the probabilities of following various transition paths through the health states of normal health, illness with one-or-more chronic diseases, disability and death before reaching age 65. Disability and death are both treated as an absorbing state (i.e., final state) for this analysis. Death is always an absorbing state. I have treated disability also as an absorbing state because the focus of this study is the determinants of DI/SSI enrollments. Individuals are entitled to these programs only up to age 65 with severe disabilities.

I first estimate the model with only demographic variables as covariates, and found that whites have significantly lower risk of transiting from good health to death and from diseased health state to disability or death. The women found to have significantly lower risk of death than men, both from normal health and diseased health, and have no difference with men in becoming disabled. These are commonly observed phenomena found in other studies. These are the effects one can study using only the Administrative data.

After controlling for other variables from our list of variables, I find that the above effects for women and for whites disappeared. I find that college graduates have significantly lower probability of all transitions. The variable CES-D measuring the level of depression and stress has significant positive effects on transiting from normal health to acquiring one-or-more diseases, from normal health to becoming disabled and from diseased health state to becoming disabled or dying. The other most significant behavioral variables are smoking and sufficiently vigorous level of exercising regularly. The smoking has significantly adverse effects and exercising has favorable effects on most transitions.

For individuals in their early fifties in normal health, I have computed the risks of their acquiring diseases, or becoming disabled or dying before certain age. Similarly, for individuals who are ill with one-or-more diseases in their early fifties, I have calculated their risks of becoming disabled or dying before certain age. These probabilities can also be calculated for individuals with given values of the covariates of the Cox regression model of the paper.

Unlike the findings of (Case and Deaton 2015), the non-Hispanic whites in their mid-ages, 50-60 in 1992, do not show higher morbidity, higher incidence of disability nor higher mortality. This could be due to the fact that the cohort in this study is older and the studied period is little earlier, 1992-2006, as compared to the age-group, 40-45, and the period, 1999-2013, for which their findings hold. Future work using the younger cohorts in the HRS data can shed better light on this issue, and on the effects of other variables studied in this paper.

References

[1.] Aalen, Odd O., Ørnulf Borgan, and Håkon K. Gjessing. 2008. *Survival and Event History Analysis*. Survival Analysis, Stochastic Process. Springer New York. https://doi.org/10.1007/978-0-387-68560-1.

[2.] Aalen, Odd O., and Søren Johansen. 1978. "An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations." Duration, mult-state, counting process. *Scandinavian Journal of Statistics* 5 (3): 141–50. http://www.jstor.org/stable/4615704.

[3.] Albrecht, Gary L., and Lois M. Verbrugge. 2003. "The Global Emergence of Disability." In *Handbook of Social Studies in Health and Medicine*, edited by Gary L Albrecht, Ray Fitzpatrick, and Susan C Scrimshaw, 293–307. London: Sage.

[4.] Alisch, R. S., B. G. Barwick, P. Chopra, L. K. Myrick, G. A. Satten, K. N. Conneely, and S. T. Warren. 2012. "Age-Associated DNA Methylation in Pediatric Populations." *Genome Research* 22 (4): 623–32. https://doi.org/10.1101/gr.125187.111.

[5.] Altman, Barbara M. 2001. "Definitions, Models, Classifications, Schemes, and Applications." In *Handbook of Disability Studies*, 97–122. SAGE Publications, Inc. https://doi.org/10.4135/9781412976251.n4.

[6.] An, Xinming, and Yiu-Fai Yung. 2014. "Item Response Theory: What It Is and How You Can Use the Irt Procedure to Apply It." DFA, IRT, SAS. *SAS Institute Inc. SAS364-2014*. https://pdfs.semanticscholar.org/d85a/7025441f5685b287b53234ce6456dcd40192.pdf.

[7.] Andersen, Per Kragh, Ornulf Borgan, Richard D. Gill, and Niels Keiding. 1993. *Statistical Models Based on Counting Processes*. Survival Analysis, Multi-state. Springer-Verlag, New York.

[8.] Austad, Steven N., and Kathleen E. Fischer. 2016. "Sex Differences in Lifespan." *Cell Metabolism* 23 (6): 1022–33. https://doi.org/10.1016/j.cmet.2016.05.019.

[9.] Barbara, Maria Andria, Ylenia Abdilla, and Jean Calleja-Agius. 2017. "An Introduction to Epigenetics." *Neonatal Network* 36 (3): 124–28. https://doi.org/10.1891/0730-0832.36.3.124.

[10.] Barker, David J. P. 1990. "The Fetal and Infant Origins of Adult Disease." *BMJ: British Medical Journal* 301 (6761): 1111.

[11.] ———. 1998. "In Utero Programming of Chronic Disease." *Clinical Science* 95 (2): 115–28. https://doi.org/10.1042/cs0950115.

[12.] ———. 2007. "The Origins of the Developmental Origins Theory." Developmental origin of health. *Journal of Internal Medicine* 261 (5): 412–17. https://doi.org/10.1111/j.1365-2796.2007.01809.x.

[13.] Barondes, Samuel. 1999. *Molecules and Mental Illness*. Scientific American Library.

[14.] Barres, Romain, and Juleen R Zierath. 2011. "DNA methylation in metabolic disorders." *The American Journal of Clinical Nutrition* 93 (4): 897S–900S. https://doi.org/10.3945/ajcn.110.001933.

[15.] Bedirhan, Otsun T, Somnath Chatterji, Nenad Kostanjsek, Jorgen Rehm, Cille Kennedy, Joanne Epping-Jordan, Shekhar Saxena, Michael von Korff, and Charles Pull. 2010. "Developing the World Health Organization Disability Assessment Schedule 2.0." *Bulletin of the World Health Organization* 88 (11): 815–23. https://doi.org/10.2471/blt.09.067231.

[16.] Blair, Steven N., III Kohl Harold W., Jr Paffenbarger Ralph S., Debra G. Clark, Kenneth H. Cooper, and Larry W. Gibbons. 1989. "Physical Fitness and All-Cause Mortality: A Prospective Study of Healthy Men and Women." *JAMA* 262 (17): 2395–2401. https://doi.org/10.1001/jama.1989.03430170057028.

[17.] Boks, Marco P., Eske M. Derks, Daniel J. Weisenberger, Erik Strengman, Esther Janson, Iris E. Sommer, René S. Kahn, and Roel A. Ophoff. 2009. "The Relationship of DNA Methylation with Age, Gender and Genotype in Twins and Healthy Controls." Edited by Joseph Najbauer. *PLoS ONE* 4 (8): e6767. https://doi.org/10.1371/journal.pone.0006767.

[18.] Bookman, Ebony B., Kimberly McAllister, Elizabeth Gillanders, Kay Wanke, David Balshaw, Joni Rutter, Jill Reedy, et al. 2011. "Gene-Environment Interplay in Common Complex Diseases: Forging an Integrative Model-Recommendations from an Nih Workshop." *Genetic Epidemiology*. https://doi.org/10.1002/gepi.20571.

[19.] Bugliari, Delia, Nancy Campbell, Chris Chan, Orla Hayden, Michael Hurd, Regan Main, Joshua Mallett, et al. 2016. "RAND HRS Data Documentation, Version P." Technical report. RAND Center for the Study of Aging. https://www.rand.org/content/dam/rand/www/external/labor/aging/dataprod/randhrs1992_2014v2.pdf.

[20.] Cai, Li, Kilchan Choi, Mark Hansen, and Lauren Harrell. 2016. "Item Response Theory." IRT. *Annual Review of Statistics and Its Application* 3 (1): 297–321. https://doi.org/10.1146/annurev-statistics-041715-033702.

[21.] Case, Anne, and Angus Deaton. 2015. "Rising Morbidity and Mortality in Midlife Among White Non-Hispanic Americans in the 21st Century." *Proceedings of the National Academy of Sciences* 112 (49): 15078–83. https://doi.org/10.1073/pnas.1518393112.

[22.] Conti, Rena M., Ernst R. Berndt, and Richard G. Frank. 2009. "Early Retirement and Di/Ssi Applications: Exploring the Impact of Depression." CESD, HRS, Disability. In *Health at Older Ages: The Causes and Consequences of Declining Disability Among the Elderly*, 381–408. NBER Chapters. National Bureau of Economic Research, Inc. http://www.nber.org/chapters/c11120.pdf.

[23.] DiLoreto, Race, and Coleen T. Murphy. 2015. "The Cell Biology of Aging." Biology of Aging. Edited by William Bement. *Molecular Biology of the Cell* 26 (25): 4524–31. https://doi.org/10.1091/mbc.e14-06-1084.

[24.] Embretson, S. E., and S. P. Reise. 2000. *Item Response Theory*. IRT, DFA. Item Response Theory for Psychologists. Taylor & Francis. https://books.google.com/books?id=9Xm0AAAAQBAJ.

[25.] Epel, Elissa S., Elizabeth H. Blackburn, Jue Lin, Firdaus S. Dhabhar, Nancy E. Adler, Jason D. Morrow, and Richard M. Cawthon. 2004. "Accelerated Telomere Shortening in Response to Life Stress." *Proceedings of the National Academy of Sciences* 101 (49): 17312–5. https://doi.org/10.1073/pnas.0407162101.

[26.] Esteller, Manel. 2008. "Epigenetics in Cancer." Cancer, Epigenetics. *New England Journal of Medicine* 358 (11): 1148–59. https://doi.org/10.1056/nejmra072067.

[27.] Fisher, Gwenith G, and Lindsay H Ryan. 2017. "Overview of the Health and Retirement Study and Introduction to the Special Issue." HRS. Edited by Mo Wang. *Work, Aging and Retirement* 4 (1): 1–9. https://doi.org/10.1093/workar/wax032.

[28.] Fleming, Thomas R. 1978. "Nonparametric Estimation for Nonhomogeneous Markov Processes in the Problem of Competing Risks." Duration, mult-state, counting process. *The Annals of Statistics* 6 (5): 1057–70. http://www.jstor.org/stable/2958608.

[29.] Fleming, Thomas R., and David P. Harrington. 2005. *Counting Processes and Survival Analysis*. Survival Analysis, Stochastic Process. Wiley.

[30.] Gill, Richard D. 2005. “Product-Integration.” In *Encyclopedia of Biostatistics*. American Cancer Society. https://doi.org/10.1002/0470011815.b2a11058.

[31.] Gill, Richard D., and Soren Johansen. 1990. “A Survey of Product-Integration with a View Toward Application in Survival Analysis.” *The Annals of Statistics* 18 (4): 1501–55. http://www.jstor.org/stable/2241874.

[32.] Gluckman, Peter D., Mark A. Hanson, Cyrus Cooper, and Kent L. Thornburg. 2008. “Effect of in Utero and Early-Life Conditions on Adult Health and Disease.” Early Childhood events, Health, Epigenetics. *New England Journal of Medicine* 359 (1): 61–73. https://doi.org/10.1056/nejmra0708473.

[33.] Hahn, Harlan. 1985. “Toward a Politics of Disability: Definitions, Disciplines, and Policies.” *The Social Science Journal*. https://www.independentliving.org/docs4/hahn2.html.

[34.] Hannum, Gregory, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, SriniVas Sadda, Brandy Klotzle, et al. 2013. “Genome-Wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates.” Biological Age measure. *Molecular Cell* 49 (2): 359–67. https://doi.org/10.1016/j.molcel.2012.10.016.

[35.] Hayflick, L. 1965. “The Limited in Vitro Lifetime of Human Diploid Cell Strains.” *Experimental Cell Research* 37 (3): 614–36. https://doi.org/10.1016/0014-4827(65)90211-9.

[36.] Horvath, Steve. 2013. “DNA Methylation Age of Human Tissues and Cell Types.” Biological Age measure. *Genome Biology* 14 (10): 3156. https://doi.org/10.1186/gb-2013-14-10-r115.

[37.] Institute of Medicine. 2007. *Improving the Social Security Disability Decision Process*. Edited by John D. Stobo, Michael McGeary, and David K. Barnes. Washington, DC: The National Academies Press. https://doi.org/10.17226/11859.

[38.] Juster, F. Thomas, and Richard Suzman. 1995. “An Overview of the Health and Retirement Study.” HRS, overview. *The Journal of Human Resources* 30: S7. https://doi.org/10.2307/146277.

[39.] Kanherkar, Riya R., Naina Bhatia-Dey, and Antonei B. Csoka. 2014. “Epigenetics Across the Human Lifespan.” *Frontiers in Cell and Developmental Biology* 2 (3): 124–28. https://doi.org/10.3389/fcell.2014.00049.

[40.] Kaplan, E. L., and Paul Meier. 1958. “Nonparametric Estimation from Incomplete Observations.” *Journal of the American Statistical Association* 53 (282): 457–81. https://doi.org/10.1080/01621459.1958.10501452.

[41.] Karakus, Mustafa C., and Lisa C. Patton. 2011. “Depression and the Onset of Chronic Illness in Older Adults: A 12-Year Prospective Study.” *The Journal of Behavioral Health Services & Research* 38 (3): 373–82. https://doi.org/10.1007/s11414-011-9234-2.

[42.] Khoury, Muin J., Lars Bertram, Paolo Boffetta, Adam S. Butterworth, Stephen J. Chanock, Siobhan M. Dolan, Isabel Fortier, et al. 2009. "Genome-Wide Association Studies, Field Synopses, and the Development of the Knowledge Base on Genetic Variation and Human Diseases." *American Journal of Epidemiology* 170 (3): 269–79. https://doi.org/10.1093/aje/kwp119.

[43.] Luo, Ye, and Linda J. Waite. 2005. "The Impact of Childhood and Adult SES on Physical, Mental, and Cognitive Well-Being in Later Life." *The Journals of Gerontology: Series B* 60 (2): S93–S101. https://doi.org/10.1093/geronb/60.2.s93.

[44.] Marks, Deborah. 1997. "Models of Disability." *Disability and Rehabilitation* 19 (3): 85–91. https://doi.org/10.3109/09638289709166831.

[45.] Nagi, Saad Z. 1965. "Some Conceptual Issues in Disability and Rehabilitation." In *Sociology and Rehabilitation*, edited by M B Sussman. 100-113: American Sociological Association. https://ci.nii.ac.jp/naid/10015569425/en/.

[46.] ———. 1976. "An Epidemiology of Disability Among Adults in the United States." *The Milbank Memorial Fund Quarterly. Health and Society* 54 (4): 439. https://doi.org/10.2307/3349677.

[47.] ———. 1991. "Disability in America: Toward a National Agenda for Prevention." In, edited by Andrew M. Pope and Alvin R. Tarlov, 309–27. Washington, DC: National Academies Press. https://doi.org/10.17226/1579.

[48.] Pope, Andrew M., and Alvin R. Tarlov. 1991. *Disability in America: Toward a National Agenda for Prevention*. Edited by Andrew M. Pope and Alvin R. Tarlov. National Academies Press. https://doi.org/10.17226/1579.

[49.] Raut, Lakshmi K. 2017. "Exits from Disability: Estimates from a Competing Risk Model." Published. *Social Security Bulletin* 77 (3): 15–38. https://www.ssa.gov/policy/docs/ssb/v77n3/v77n3p15.html.

[50.] ———. 2019. "Pathways to Disability and Mortality in Mid-Ages: Estimates from the Health and Retirement Study Data." Working: SSA. *Draft Presented at the 2019 SGE Annual Conference*.

[51.] Renna, Francesco. 2008. "Alcohol Abuse, Alcoholism, and Labor Market Outcomes: Looking for the Missing Link." *ILR Review* 62 (1): 92–103. https://doi.org/10.1177/001979390806200105.

[52.] Seib, Charrlotte, Eliza Whiteside, Kathryn Lee, Janice Humphreys, Tiet Hanh Dao Tran, Lisa Chopin, and Debra Anderson. 2014. "Stress, Lifestyle, and Quality of Life in Midlife and Older Australian Women: Results from the Stress and the Health of Women Study." *Women's Health Issues* 24 (1): e43–e52. https://doi.org/10.1016/j.whi.2013.11.004.

[53.] Shalev, Idan, and Jay Belsky. 2016. "Early-Life Stress and Reproductive Cost: A Two-Hit Developmental Model of Accelerated Aging?" Ageing Biology,

Genetics,Neuroscience. *Medical Hypotheses* 90: 41–47. https://doi.org/10.1016/j.mehy.2016.03.002.

[54.] Shalev, Idan, Sonja Entringer, Pathik D. Wadhwa, Owen M. Wolkowitz, Eli Puterman, Jue Lin, and Elissa S. Epel. 2013. "Stress and Telomere Biology: A Lifespan Perspective." *Psychoneuroendocrinology* 38 (9): 1835–42. https://doi.org/10.1016/j.psyneuen.2013.03.010.

[55.] Simons, Ronald L., Man Kit Lei, Steven R. H. Beach, Robert A. Philibert, Carolyn E. Cutrona, Frederick X. Gibbons, and Ashley Barr. 2016. "Economic Hardship and Biological Weathering: The Epigenetics of Aging in a U.s. Sample of Black Women." Ageing Biology, Genetics,Neuroscience. *Social Science & Medicine* 150: 192–200. https://doi.org/10.1016/j.socscimed.2015.12.001.

[56.] Snyder, Alison R., John T. Parsons, Tamara C. Valovich McLeod, R. Curtis Bay, Lori A. Michener, and Eric L. Sauers. 2008. "Using Disablement Models and Clinical Outcomes Assessment to Enable Evidence-Based Athletic Training Practice, Part I: Disablement Models." *Journal of Athletic Training* 43 (4): 428–36. https://doi.org/10.4085/1062-6050-43.4.428.

[57.] Sonnega, A., J. D. Faul, M. B. Ofstedal, K. M. Langa, J. W. Phillips, and D. R. Weir. 2014. "Cohort Profile: The Health and Retirement Study (HRS)." *International Journal of Epidemiology* 43 (2): 576–85. https://doi.org/10.1093/ije/dyu067.

[58.] Steffick, Diane E. 2000. "Documentation of Affective Functioning Measures in the Health and Retirement Study." *Ann Arbor, MI: University of Michigan*. http://hrsonline.isr.umich.edu/sitedocs/userg/dr-005.pdf.

[59.] The US Burden of Disease Collaborators. 2018. "The State of US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among US States." *JAMA* 319 (14): 1444–72. https://doi.org/10.1001/jama.2018.0158.

[60.] Thornburg, Kent L., Jackilen Shannon, Philippe Thuillier, and Mitchell S. Turker. 2010. "In Utero Life and Epigenetic Predisposition for Disease." Early Childhood events, Health, Epigenetics. In *Epigenetics and Cancer, Part B*, 57–78. Elsevier. https://doi.org/10.1016/b978-0-12-380864-6.00003-1.

[61.] Vable, Anusha M., Paola Gilsanz, Thu T. Nguyen, Ichiro Kawachi, and M. Maria Glymour. 2017. "Validation of a Theoretically Motivated Approach to Measuring Childhood Socioeconomic Circumstances in the Health and Retirement Study." IRT, SES, HRS. Edited by Abigail Fraser. *PLOS ONE* 12 (10): e0185898. https://doi.org/10.1371/journal.pone.0185898.

[62.] Vaupel, James W. 2010. "Biodemography of Human Ageing." *Nature* 464 (7288): 536–42. https://doi.org/10.1038/nature08984.

[63.] Verbrugge, Lois M., and Alan M. Jette. 1994. "The Disablement Process." Disablement Process. *Social Science & Medicine* 38 (1): 1–14. https://doi.org/10.1016/0277-9536(94)90294-1.

[64.] Verbrugge, Lois M., Kenzie Latham, and Philippa J. Clarke. 2017. “Aging with Disability for Midlife and Older Adults.” *Research on Aging* 39 (6): 741–77. https://doi.org/10.1177/0164027516681051.

[65.] World Health Organization. 2001. “International Classification of Functioning, Disability and Health.” Technical report. Geneva: World Health Organization. https://psychiatr.ru/download/1313?view=name=CF_18.pdf.

[66.] Wreede, Liesbeth C. de, Marta Fiocco, and Hein Putter. 2010. “The Mstate Package for Estimation and Prediction in Non- and Semi-Parametric Multi-State and Competing Risks Models.” Duation Analysis. *Computer Methods and Programs in Biomedicine* 99 (3): 261–74. https://doi.org/10.1016/j.cmpb.2010.01.001.

[67.] Zarulli, Virginia, Julia A. Barthold Jones, Anna Oksuzyan, Rune Lindahl-Jacobsen, Kaare Christensen, and James W. Vaupel. 2018. “Women Live Longer Than Men Even During Severe Famines and Epidemics.” *Proceedings of the National Academy of Sciences* 115 (4): E832–E840. https://doi.org/10.1073/pnas.1701535115.

Appendix

(#tab:table7) Effects of childhood factors, race and sex on childhood health, college education and initial health in early 50’s.

Table 4: Effects of childhood factors, race and sex on childhood health, college education and initial health in early 50’s.

	cHLTH (1)	cHLTH (2)	College+(1)	College+(2)
Intercept	0.7419***	0.1619	-1.9915***	-3.9812***
	(0.0580)	(0.0907)	(0.0978)	(0.1527)
White	0.2981***	0.2210***	0.1430	-0.0761
	(0.0579)	(0.0665)	(0.0831)	(0.0888)

Female
-0.0775
-0.0124
-0.5452***
-0.5707***

(0.0482)
(0.0528)
(0.0601)
(0.0628)
Childhood SES
0.3201***

1.0714***

(0.0321)

(0.0404)

Father’s Education

0.0381***

0.1094***

(0.0093)

(0.0113)
Mother’s Education

0.0289**

0.1395***

(0.0098)

(0.0127)
Father’s Job

0.4467***

0.7138***

(0.0959)

(0.0838)
Childhood Health

0.4533***
0.4071***

(0.0738)
(0.0772)
AIC
10239.7368
8586.1769
7190.0004
6482.8162
BIC
10268.0358
8627.6783
7225.3741
6531.2346
Log Likelihood
-5115.8684
-4287.0884
-3590.0002
-3234.4081
Deviance
10231.7368
8574.1769
7180.0004
6468.8162
Num. obs.
8732
7457
8732
7457
p < 0.001; *p < 0.01;* p < 0.05