Pathways to Disability and Death Before Disability in Mid-ages: Estimates from the Health and Retirement Study Data*

Lakshmi K. Raut
Social Security Administration
250 E St SW, #8092F
Washington, DC 20025

mailto:Lakshmi.Raut@ssa.gov

2020-11-24

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

*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) 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.

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 data on early childhood factors such as childhood SES, health and education, 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.

1 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; Altman, 2001; Albrecht and Verbrugge, 2003; Marks, 1997; Altman, 2001; Snyder et al., 2008). I use the following statutory definition of disability that the Social Security Administration uses 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.[1]

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

[1] 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.

was introduced by the sociologist Nagi, 1965, which he further refined in Nagi, 1976; Nagi, 1991. This model was extended by Verbrugge and Jette, 1994; Verbrugge, Latham, et al., 2017 who added biological, environmental and behavioral risk factors affecting all four stages of the disablement process.[2] 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 the 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

[2]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.

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

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. The biomedical literature finds that aging of the cells, i.e, cellular senescence, and age related diseases are associated with shortening of telomere length[3] and changes in global methylation[4], and that stress, smoking, drinking, chemical misuse, and diet are important modulators for these changes. For telomere mechanism, see for instance, Austad and Fischer, 2016; Blair et al., 1989; Vaupel, 2010; Zarulli et al., 2018, and for methylation mechanism, see for instance, Alisch et al., 2012; Barres and Zierath, 2011; Boks et al., 2009; Esteller, 2008; Hannum et al., 2013; Horvath, 2013

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; Barker, 1998 and later of Gluckman et al., 2008. They found strong association 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

[3]Telomeres are the caps at the end of chromosomes in a DNA sequence. They look like the plastic caps at the end of shoelaces.

[4]definition if possible

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.”

The finding of the above microbiology literature — that cellular stressors at various stages of the cell’s lifecycle, especially during the prenatal and postnatal development period are important determinants of the speed of cellular aging — is an important milestone in aging research. For public policy, however, it is important to find socioeconomic factors and health related behavioral factors that modulate the cellular stressors. Many studies in social sciences find that low socioeconomic 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 et al., 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. Luo and Waite, 2005 using the Health and Retirement Studies data found childhood SES and childhood health influence strongly later life health outcomes. 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 data on childhood SES, health, and education, biomarkers 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 econometric specifications and estimation methods followed in the paper. In Section 4, I describe the dataset and the creation variables for this study. In Section 6, I present the estimates of the transition probabilities for the overall population. In 6, I estimate the effect of childhood factors on mid-age initial health; and childhood factors, biomarkers and health behaviors on individual health trajectories. In In Section 7, I discuss the policy implications of the estimates. Section 8 concludes the paper.

2 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.[5] After reaching this state, the individual is not followed any further. A competing risk for getting on the rolls is death 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.

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.

[5] 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.

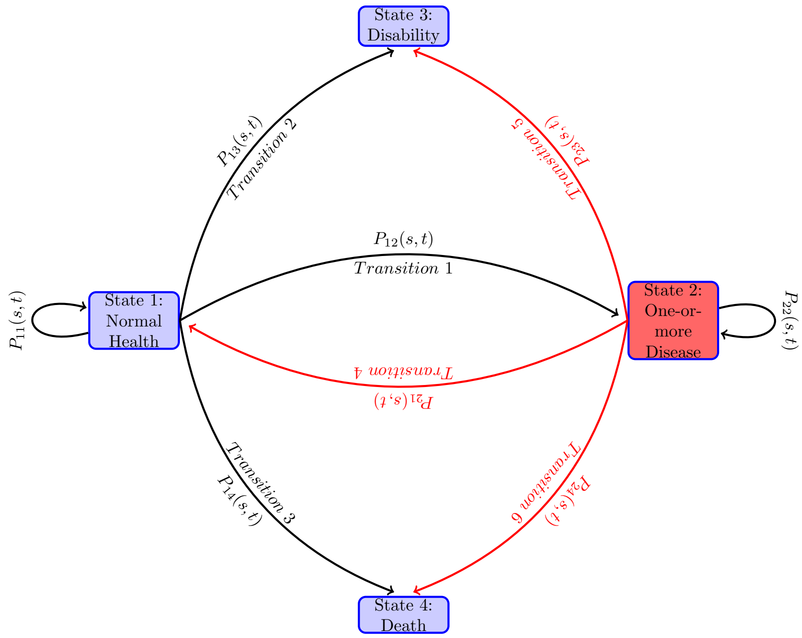

Figure 1: Path diagram of health trajectors.

The goal is to estimate the probability of getting onto the disability rolls or die before disability sequentially over time as an individual progresses over life experiencing health shocks or changing health 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 lifespans 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

$$P(s,t) \equiv \left(P_{hj}(s,t)\right)_{h,j=1...4}.$$

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

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

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[6] 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

[6] 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.

$$
\begin{aligned}
\lambda_{hj}(t) &= \lim_{\Delta t \to 0} \frac{P_{hj}(t, t+\Delta t) - P_{hj}(t,t)}{\Delta t}, \text{ for } j \in S \\
&= \lim_{\Delta t \to 0} \frac{P_{hj}(t, t+\Delta t)}{\Delta t}, \text{ for } j \neq h \\
\lambda_{hh}(t) &= \lim_{\Delta t \to 0} \frac{P_{hh}(t, t+\Delta t) - 1}{\Delta t}, \text{ for } j = h, \\
&= -\lim_{\Delta t \to 0} \frac{\sum_{j \neq h} P_{hj}(t, t+\Delta t)}{\Delta t} \\
&= -\sum_{j \neq h} \lambda_{hj}(t)
\end{aligned} \tag{2}
$$

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

$$
\Gamma(t) = \begin{pmatrix} -(\lambda_{12}(t) + \lambda_{13}(t) + \lambda_{14}(t)) & \lambda_{12}(t) & \lambda_{13}(t) & \lambda_{14}(t) \\ \lambda_{21}(t) & -(\lambda_{21} + \lambda_{23}(t) + \lambda_{24}(t)) & \lambda_{23}(t) & \lambda_{24}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3}
$$

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[7]

[7] 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.

$$\frac{\partial P\left(s,t\right)}{\partial t}=P\left(s,t\right)\cdot\Gamma\left(t\right) \tag{4}$$

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 4.

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 4. 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 $\hat{\Gamma}(t)$.[8] 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.

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 $\Lambda_{hj}\left(t\right) = \int_0^t \lambda_{hj}\left(u\right) du$. 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

$$P(s,t) = P(t_0,t_1)\cdot P(t_1,t_2)\cdot \ldots \cdot P(t_{m-1},t_m) = \prod_{i=1}^{m} P(t_{i-1},t_i) \tag{5}$$

Note that as $|t_{i-1} - t_i| \to 0$, the transition probability matrix $P(t_{i-1},t_i) \to P(t_{i-1},(t_{i-1}+$

[8] 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.

$dt) = I + \Gamma(t)dt$. [9] 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 5 converges to a matrix called the *the product integral*[10] of the integrated hazard functions $\Lambda(s,t)$. This product integral is denoted as $\widetilde{\prod}_s^t (I + d\Lambda(u))$. 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.

$$P(s,t) = \widetilde{\prod}_s^t (I + d\Lambda(u)). \tag{6}$$

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 6 and then compute the matrix products.

There are broadly two types of statistical methods to get the estimates — parametric and semi-parametric methods. I will follow the widely used semi-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. Without covariates, the semi-parametric method is, in fact, nonparametric method.

3.1 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

[9] From the definition of transition intensity in Equations 2 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 $P(t, t+\Delta t) = I + \Gamma(t)\Delta t$.

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

functions

$$\lambda_{hj}\left(t;X\left(t\right)\right)=\lambda_{hj}^{0}\left(t\right)e^{\beta_{hj}'X(t)}. \tag{7}$$

In the above specification, $\lambda_{hj}^{0}(t)$ is known as the *baseline hazard function*. The specification of transition intensity in Equation 7 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.[11] I use this estimation method in this paper. First I describe the Aalen-Johansen-Flemming estimator for models with no covariates and then I describe the method for the general case with covariates.

3.1.1 Aalen-Johansen-Flemming estimator without covariates

To describe the Aalen-Johansen-Flemming estimator, I introduce some concepts and notation. For each individual $i, i = 1, 2, ..., n$ 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; 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 $\bar{N}_{hj}(t) = \sum_i^n N_{hj,i}(t)$, a counting process measuring the number of transitions of the type $h \to j$ in the sample at time t, $\bar{Y}_h(t) = \sum_i^n Y_{h,i}(t)$, a counting process measuring the number of individuals in the sample at risk for a transition at time t, and $\bar{M}_{hj}(t) = \sum_i^n M_{hj,i}(t)$. In any empirical study the data will be at the discrete times, say in ordered times $0 = t_0 < t_1 < ... < t_m$. At each time t_i, we calculate

$$\hat{\lambda}_{hj}(t_i) = \frac{\triangle\bar{N}_{hj}(t_i)}{\bar{Y}_h(t_i)}, j \neq h, \tag{8}$$

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

[11] Fleming gave the estimator for complete data, Aalen and Johansen gave the estimator for censored data.

$$
\begin{aligned}
\hat{\Lambda}_{hj}(t) &= \sum_{i:t_i \le t} \hat{\lambda}_{hj}(t_i), j \neq h, \\
\hat{\Lambda}_{hh}(t) &= -\sum \hat{\Lambda}_{hj}(t) \text{ and} \\
\hat{\Lambda}_{hj}(t) &= 0, \text{ for } h = 3,4; \quad j = 1,2,3,4
\end{aligned} \tag{9}
$$

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 6 as follows

$$
\hat{P}(s,t) = \widetilde{\prod_{s<u<t}} (I + d\hat{\Lambda}(u)) = \widetilde{\prod_{i:t_i \le t}} (I + [\hat{\Lambda}(t_i) - \hat{\Lambda}(t_{i-1})]). \tag{10}
$$

3.1.2 Aalen-Johansen-Flemming estimator with covariates

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

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

The estimates of cumulative intensities with covariates are obtained from Equation 8 by replacing, $\bar{Y}_h(t)$ with $\bar{Y}^{*}_{hj}(t)$.

Aalen-Johnson-Fleming estimator has nice statistical property. For instance, using Martingale calculus, it can be shown that the estimator is asymptotically unbiased and the normalized estimate is normally distributed (i.e., the central limit theorem holds for normalized parameter estimates) with an asymptotic estimable variance-covariance matrix see for details, Aalen, Borgan, et al., 2008; Andersen et al., 1993; Fleming and Harrington, 2005.

4 The Dataset and the Variables

4.1 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 status of the individuals in the sample over the survey years.

Survey Year	Alive:normal health	Alive: diseased	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 only 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.

4.2 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. I will discuss more about it in Section 6.

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.

Init.HLTH is a categorical variable denoting the initial health state of an respondent right before the respondent entered the Health and Retirement Study, taking value 1 or 2.

5 Transition probability estimates for the overall population

I have used the R package *mstate* that implemented most of the estimation methods with and without covariates that I described in Section 3 see, Wreede et al., 2010, for details on the package. In this section, I report the estimated transition probabilities for the overall population without any covariate. The notation $P_{ij}(t)$ denotes the probability of transiting to state j by time t starting from state$i$ at the base age 51 (the base period of this paper is 51). I report in the path diagram of Figure 2 these estimated probabilities for $t = 60$ and 65. I report these estimated probabilities for all other ages in Table 2 and plot in Figure 3 from two transitory states — state 1 (normal health status) in panel (a) and state 2 (diseased health status) in panel (b), both starting in the transient state at the base age of 51.

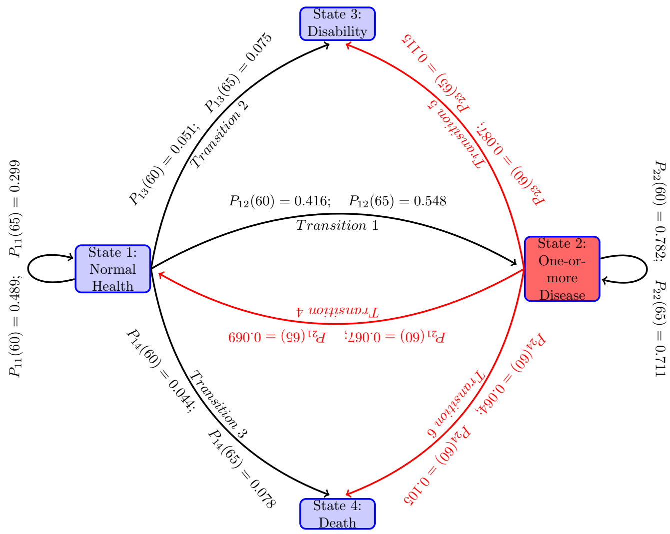

Figure 2: Path diagram of the estimated model for the overall population.

The estimated probabilities show some general characteristics of the population. For instance, an individual of age 51 with normal health has the probability of getting onto the disability rolls (i.e., transition $1 \rightarrow 3$) 0.051 by age 60 and 0.075 by age 65. Similarly, an individual of age 51 with one or more chronic diseased has the probability of getting onto the disability rolls (i.e., transition $2 \rightarrow 3$) 0.087 by age 60 and 0.115 by age 65, which are higher than for an individual with normal health. The latter is expected, but one can get a quantitative comparisons. Similarly the estimates of the competing risk of death before disability from these two transient health states can be read from the lower part of the figure. The transition probability estimates of the competing risk — death before disability — are important because it can give some quantitative ideas how the probabilities of getting onto disability may increase due to improvements in medical technology or changes in health related behaviors that may reduce the probabilities of the competing risks.

Figure 2 and Table 2 give more details of the nature of these transition probabilities age by age.

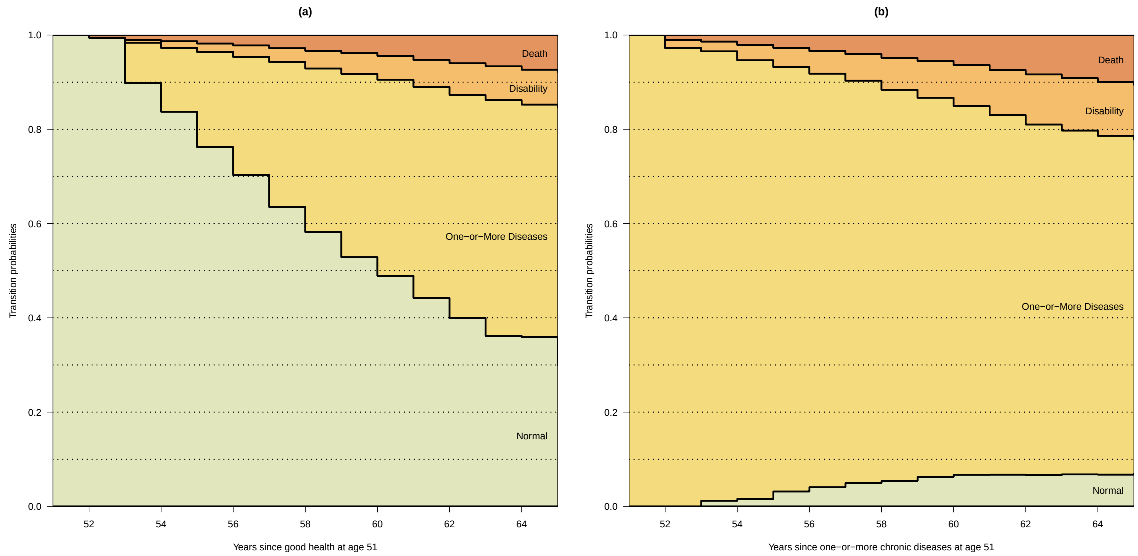

Figure 3: Transition probabilities (a) from normal health state, (b) from one-or-more diseased health state

Table 2: Estimated transition probabilities by duration of stay.

Age	$1 \to 1$	$2 \to 2$	$1 \to 2$	$2 \to 1$	$1 \to 3$	$2 \to 3$	$1 \to 4$	$2 \to 4$
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

From the transition probabilities in Table 2, one can compute the probability of various transition paths. For instance, suppose one is interested in estimating the risk of disability by age 65 for a person who is in normal health state at age 60. How does that probability compare with the probability of one who is in diseased health state at age 60? Or one is interested in estimating such probabilities for an individual who joined the Health and Retirement Study at age 60. These can be obtained if we can compute $P(60,65)$. To compute this transition probability matrix $P(60,65)$ from the computed probabilities in Table 2, note that applying the Chapman-Kolmogorov Equation 1, one has $P(51,65) = P(51,60) \cdot P(60,65)$. This implies, $P(60,65) = [P(51,60)]^{-1} \cdot P(51,65)$, (assuming that the inverse exists).

The components of the estimated transition probability matrix $P(60,65)$ for the transitions in Table 2 turn out to have values: $1 \rightarrow 1 = 0.579$, $2 \rightarrow 2 = 0.877$, $1 \rightarrow 2 = 0.374$, $2 \rightarrow 1 = 0.038$, $1 \rightarrow 3 = 0.021$, $2 \rightarrow 3 = 0.034$, $1 \rightarrow 4 = 0.026$, $2 \rightarrow 4 = 0.051$. Compare these probabilities with the corresponding probabilities in the last row of Table 2.

6 Childhood factors and mid-age health outcomes

Molecular biology literature mentioned in the introduction points out that the stressors of the body cells are important determinants of the nature of cell divisions during early development and later life health outcomes.[12] While those stressors cannot be directly observed or measured, many socioeconomic factors modulate those stressors and cell developments, and thus affect later life health outcomes. Furthermore, early life health developments together with health related behaviors are important determinants of later life health outcomes. Health behaviors are partly determined by cognitive and non-cognitive skills. Education level thus can affect health behaviors and health developments in later life. Education also determines earnings, which determine health related expenditures and thus health outcomes.

The HRS dataset does not have prenatal or postnatal data on individuals. It has a few variables on childhood socioeconomic status, which are correlates of the stressors of the cell developments. How does one quantify childhood SES? There is no consensus on what exactly constitutes cSES. Some studies use different sets of variables to represent cSES. For instance, Heckman and Raut, 2016 and a few other studies use parents' education as a measure childhood SES in modeling attainment of college degree. Luo and Waite, 2005

[12] Genetic make-up also controls gene expressions for producing proteins that create diseases but the epigenetic factors creating the stressors are important as well.

used Father's and Mother's education and the Family financial well-being as regressors without aggregating them into a single measure to examine how these variables affect a measure of mid-age health outcomes for the HRS sample. It is useful to have a single measure of SES. Some studies used the latent variable approach to come with a statistically defined measure of cSES. For instance, Vable et al., 2017 used Mplus software and a number of variables from the HRS dataset to create their cSES measure. Similar to their approach, I use the latent variable statistical procedure IRT on a set of parental characteristics during childhood of the respondents as described in Section 4.2. I will validate this aggregate measure using three Logit regression models as described in the next subsection.

Childhood health status (cHLTH) is an important factor for later life health outcomes and educational attainments. cSES influences the stressors of the cells environment and thus will affect cHLTH. Apart from cSES, other factors such as nutrition and pediatric health care are important factors. We do not have data on those. In the next subsection I will specify a Logit model of cHLTH with childhood socioeconomic status together with other observable characteristics.

Cognitive skills or Education level is an important factor for later health outcomes as it determines various health related choices made by an individual throughout life, and also an important determinant of earnings, and types of employment with or without covered health insurance that one does. Similar to many studies, I use a binary education level, College+, which takes value 1 if the individual has at least a college degree, and 0 otherwise. Many factors determine College+ such as innate IQ, family background, preschool inputs, prenatal and postnatal stressors for brain development, the childhood health status, and mother's time input. See, Heckman, 2008 and Raut, 2018 for recent literature on the biology of brain development and the role of socioeconomic factors, and Heckman and Raut, 2016 for a Logit model of college completion in which a IQ mesure, family background measured with parents' education, preschool inputs and non-cognitive skills play important roles. HRS does not have data on many variables. I use childhood socioeconomic status and cHLTH, together with a few other demographic variables as regressors in the Logistic specification in the next subsection.

I examine two types of mid-age health outcomes — initial health status at one's early *50s* (Init.HLTH), the health status that a respondent had when first participated in the Health and Retirement Study, and the other one is the pathway through the health states that the individual will traverse starting from the initial health state. Both types of health outcomes are modeled as function of childhood factors, cSES, cHLTH, College+. The subsection 6.2 below has the first model, and the subsection 6.2 following that has the second model. In subsection 6.3, I estimate the second model including biomarkers and health behaviors together with the childhood factors.

6.1 Models of childhood factors, initial mid-age health and validation of the cSES measure

In this subsection, I estimate three sets of Logit regression models for cHLTH and College+ and Init.HLTH. and validate this paper's measure of cSES on a number of regression models. In each set, I have two specifications of Logistic regression models: in one model, I include the cSES measure that I have created in this paper, and in the second model, I include in its palce three family background variables used in Luo and Waite, 2005, — Father's Education, Mother's Education and Father's job situation during the respondent's childhood — Two models in each set have common other regressors as shown in Table 3. I then examine if the coefficient estimates and their significance levels of the common covariates of the models are similar. If they are similar, then the single measure cSES of the paper is validated. I have also calculated the pseudo R^2 defined as $R^2 = (1 - deviance/nulldeviance)$. It turns out to be the case that the parameter estimates of the common regressors mostly do not differ in statistical significance levels and numerical magnitudes. The R^2 for the models with the regressor cSES is slightly higher or close to the R^2 of the competing models. Therefore, the measure cSES constructed in the paper is validated with respect to these three Logistic regression models.

From the statistically significant parameter estimates of the variable cSES in the models with cSES as a regressor, we see that cSES has positive effect on child health, and on the probability of college completion and the probability better (i.e., normal as compared to diseased) Init.HLTH.

A better childhood health leads to a higher probability of college completion and a higher probability of being in normal health in one's early *50s*. An education level of at least a college degree also has a significant positive effect on the probability of being in normal health in one's early *50s*.

Furthermore, the estimates show that White has higher probabilities for better childhood health, attaining college degree and more, and better initial health outcomes in one's early *50s* and Female has lower probability of completing college andlower probability of

Table 3: Effects of childhood factors, race and sex on childhood health, college education and initial health in early 50s.

	cHLTH		College+		Init.HLTH	
	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.220^{***}	0.162	-2.205^{***}	-3.981^{***}	-1.028^{***}	-1.132^{***}
	(0.053)	(0.091)	(0.087)	(0.153)	(0.063)	(0.098)
White	0.282^{***}	0.221^{***}	0.227^{**}	-0.076	0.226^{***}	0.193^{**}
	(0.053)	(0.066)	(0.077)	(0.089)	(0.057)	(0.067)
Female	-0.022	-0.012	-0.537^{***}	-0.571^{***}	-0.218^{***}	-0.177^{***}
	(0.044)	(0.053)	(0.057)	(0.063)	(0.044)	(0.050)
Childhood SES	0.841^{***}		1.328^{***}		0.222^{***}	
	(0.054)		(0.058)		(0.051)	
Father's Education		0.038^{***}		0.109^{***}		0.021^{*}
		(0.009)		(0.011)		(0.009)
Mother's Education		0.029^{**}		0.139^{***}		-0.000
		(0.010)		(0.013)		(0.009)
Father's Job		0.447^{***}		0.714^{***}		0.042
		(0.096)		(0.084)		(0.078)
Childhood Health			0.342^{***}	0.407^{***}	0.206^{***}	0.217^{***}
			(0.064)	(0.077)	(0.048)	(0.057)
College					0.149^{*}	0.137^{*}
					(0.059)	(0.064)
R^2	0.026	0.018	0.085	0.136	0.010	0.008
Num. obs.	9601	7457	9601	7457	9601	7457

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

remaining in good health in their early *50s*.

It is possible that after controlling for cSES, the variable White might become less significant for health related behaviors. I cannot control for health behaviors in the models of this subsection as the HRS data does not have data on health behaviors prior to the survey years. I will examine this with the second measure of health risks, i.e., probabilities of following different health trajectories starting at age *51* in the next two subsections.

6.2 Childhood factors, and mid-age health pathways

I examine the effects of childhood factors, race and sex on the probability of various health pathways. The parameter estimates are shown in Table 4.

Table 4: Effects of childhood factors, race and sex on health transitions in mid ages.

	$1 \to 2$	$1 \to 3$	$1 \to 4$	$2 \to 1$	$2 \to 3$	$2 \to 4$
White	0.022	−0.086	-0.630^{**}	0.320^{**}	-0.323^{***}	-0.233^{*}
	(0.061)	(0.250)	(0.223)	(0.113)	(0.098)	(0.094)
Female	0.032	−0.377	−0.411	-0.407^{***}	−0.042	-0.627^{***}
	(0.044)	(0.200)	(0.213)	(0.079)	(0.088)	(0.086)
Childhood SES	-0.096^{*}	−0.113	−0.501	0.005	−0.201	-0.657^{***}
	(0.048)	(0.223)	(0.299)	(0.090)	(0.114)	(0.130)
Childhood Health	0.293^{***}	−0.107	-2.404^{***}	0.351^{***}	−0.055	-1.583^{***}
	(0.054)	(0.208)	(0.278)	(0.091)	(0.093)	(0.093)
College+	−0.103	-0.970^{**}	−0.465	0.072	-0.972^{***}	−0.244
	(0.056)	(0.336)	(0.354)	(0.103)	(0.183)	(0.145)
R^2	0.010	0.004	0.039	0.007	0.008	0.056
Num. events	1918	105	94	667	512	564
Num. obs.	3580	3580	3580	8045	8045	8045

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

To understand the quantitative significance of these estimates, say the parameter estimate of -0.97 in Table 4 for the parameter of College+ in the model for transition $1 \to 3$ (transition from normal health to disability) is that given all other factors equal, an individual of a given age with normal health will have the *hazard ratio* of 0.38 ($\equiv exp(-0.97)$), which means that the probability of a college graduate of normal health at age 51 becoming disabled at a given later age is 0.38 times the probability of a person of normal health at age

51 without a college degree becoming disabled at that later given age. In Cox proportational hazard models, the hazard ratio is time constant, i.e., the ratio is same for all future ages.

From the table we see that childhood health has the most significant effect for transition $1 \rightarrow 4$ and Numerically and statistically most significant effect is for the transition $1 \rightarrow 4$ with a proportional hazard ratio of 0.09, i.e., the probability of an individual in normal health at age 51 dying before disability is who had experienced good childhood health is 0.09 times the probability dying before disability of an individual who had bad childhood health. A better good childhood health leads to lower probability of death before disability and a higher probability recover from diseased health state. A better childhood health also has a higher probability becoming diseased, but given that it also leads to higher probability recovery, comparing the magnitudes of the parameter estimates $1 \rightarrow 2$ and $2 \rightarrow 1$, we see that better childhood health leads to lower net transition from $1 \rightarrow 2$.

An individual from the disadvantaged childhood SES has a higher probability of disease incidence and death before disability after they become diseased.

As far as getting onto disability, only variable that has significant effect is college+.

The white has lower probability of death before disability both from the normal health state and from the diseased health state and also a lower probability of getting onto disability or recovering after being diseased. Female has a significantly lower probability of death before disability with a hazard ratio of 0.53. Furthermore, they also have a much lower probability of recovering from diseased health state with a hazard ratio 0.67. The favorable effects of white race and female sex could be because their quality of health, measured by biomarkers of health, and the health related living standards, not necessarily a genetic effect. I examine it in the subsection 6.3.

6.3 Health behaviors, biomarkers and mid-age health pathways

In this subsection, I include biomarkers and health related behaviors in the model of the previous subsection. The parameter estimates are shown in Table 5.

Note that the favorable effects of the White race variable has almost wiped out when we controlled for the health biomarkers and health related behaviors. Similarly, the Female sex has now much lower effect on the transition probability from diseased health state to death before disability. It could be that the whites and females have better health reflected in the controlled biomarker variables and the parameter estimates for white and sex variables in Table 4 captured those effects.

Numerically and statistically most significant effect in Table 5 is CES-D – measuring depression and stress. A higer level of CES-D of an individual of normal health at age 51 will have significantly higher probabilities becoming diseased and even higher probability of getting onto disability but no significant effect on death before disability; from the diseased health state, the individual will have much smaller chance of going back to normal health and significantly higher probability of getting onto disability, or die before disability.

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.

Table 6: Mean values of variables for various groups

sample	cSES	cHLTH	College.	CESD	cogtot	BMI	Prev_care	Smoke	Drink	Ex
Non-white	0.164	0.597	0.125	-0.281	21.130	0.711	0.377	0.599	0.664	
White	0.305	0.676	0.188	-0.353	24.492	0.615	0.517	0.629	0.750	
Male	0.284	0.666	0.217	-0.365	23.288	0.690	0.475	0.729	0.804	
Female	0.270	0.654	0.138	-0.316	24.276	0.585	0.500	0.528	0.667	
cSES: poor	0.000	0.612	0.113	-0.319	23.178	0.653	0.460	0.628	0.698	
cSES: rich	1.000	0.786	0.339	-0.386	25.314	0.587	0.562	0.610	0.821	
cHLTH: poor	0.174	0.000	0.126	-0.297	22.824	0.626	0.339	0.637	0.684	
cHLTH: good	0.329	1.000	0.201	-0.357	24.210	0.639	0.566	0.615	0.757	
No college degree	0.221	0.639	0.000	-0.322	23.232	0.646	0.466	0.637	0.710	
College degree+	0.535	0.757	1.000	-0.417	26.515	0.581	0.592	0.558	0.835	

Table 7: Polychoric correlations of childhood factors with other variables.

cFactors	cSES	cHLTH	College	CESD	cogtot	BMI	Prev_care	Smoke	Drink	Exercise
White	0.257	0.121	0.148	-0.203	0.360	-0.148	0.201	0.045	0.144	0.081
Female	-0.027	-0.021	-0.191	0.128	0.134	-0.174	0.038	-0.327	-0.260	-0.069
cSES	1.000	0.294	0.459	-0.200	0.270	-0.105	0.151	-0.028	0.232	0.166

cFactors	cSES	cHLTH	College	CESD	cogtot	BMI	Prev_care	Smoke	Drink	Exercise
cHLTH	0.294	1.000	0.186	-0.164	0.163	0.022	0.344	-0.035	0.133	0.366
College+	0.459	0.186	1.000	-0.294	0.396	-0.095	0.175	-0.114	0.227	0.106

As I have shown above, 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,??sample which are detrimental, but these did not offset the other conducive effects mentioned above.

7 Policy implications

The probabilities of disability or death before disability for a representative individual from the three groups are shown in Figure 4 and for selected ages 55, 60 and 65 are shown in Table 8. Due to a policy suppose an individual moves from group 1 to group 2, we can see how this persons transition probabilities would dramatically improve. Furthermore, if social policies can improve their health related living styles so that he can be a person with characteristics of group 3, the probabilities improve even more. Compare the above probabilities and the plots with estimated probabilities of and the plot for the overall population in Table 2 and Figure 3.

Furthermore, if social policies can make it possible for a person who would have been of type 1 to become a type 2 person, his likelihood of having a normal health would increase from 0.3096 to 0.4440, which will also reduce the rate of death before disability and the rate of disability enrollment in the population even further, as these probabilities are lower for an individual with normal health compared to the corresponding probabilities of an individual of diseased health state at age 51.

Table 5: Cox regression estimates of the effects of childhood factors, middle age health status and health behaviors.

	$1 \to 2$	$1 \to 3$	$1 \to 4$	$2 \to 1$	$2 \to 3$	$2 \to 4$
White	0.026	0.250	−0.596	0.261*	−0.146	−0.151
	(0.065)	(0.315)	(0.415)	(0.119)	(0.105)	(0.143)
Female	0.061	−0.411	−0.311	−0.458***	−0.084	−0.398**
	(0.047)	(0.233)	(0.419)	(0.084)	(0.098)	(0.130)
Childhood SES	−0.127**	0.140	0.302	−0.095	−0.129	−0.021
	(0.049)	(0.243)	(0.455)	(0.091)	(0.118)	(0.153)
Childhood Health	−0.003	−0.030	−1.429***	0.185*	−0.085	−0.588***
	(0.055)	(0.249)	(0.367)	(0.093)	(0.098)	(0.123)
College	−0.055	−0.632	−1.105	0.052	−0.694***	−0.409
	(0.059)	(0.396)	(0.772)	(0.108)	(0.194)	(0.224)
CES-D	0.661***	2.251***	−0.904	−0.677***	1.234***	0.487*
	(0.115)	(0.341)	(1.158)	(0.204)	(0.156)	(0.226)
Total cognitive scores	0.002	−0.081**	−0.002	0.010	−0.029**	0.010
	(0.005)	(0.027)	(0.037)	(0.010)	(0.009)	(0.014)
BMI	0.038***	0.026	0.033	−0.063***	0.021*	−0.002
	(0.005)	(0.028)	(0.035)	(0.009)	(0.009)	(0.013)
Behavior: Preventive care	0.305***	0.710**	−2.881**	−0.438***	0.248**	−1.456***
	(0.046)	(0.222)	(1.015)	(0.081)	(0.094)	(0.152)
Behavior: Smoking	0.084	0.244	2.107**	0.024	0.325**	0.893***
	(0.047)	(0.223)	(0.733)	(0.087)	(0.103)	(0.163)
Behavior: Drinking	0.114*	−0.240	0.730	0.067	−0.165	−0.358**
	(0.057)	(0.251)	(0.623)	(0.101)	(0.100)	(0.133)
Behavior: Exercising	−0.071	−0.922***	−0.669	0.684***	−0.567***	−0.877***
	(0.069)	(0.251)	(0.424)	(0.141)	(0.101)	(0.127)
R^2	0.044	0.028	0.024	0.026	0.033	0.041
Num. events	1824	93	31	639	486	266
Num. obs.	3203	3203	3203	7145	7145	7145

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

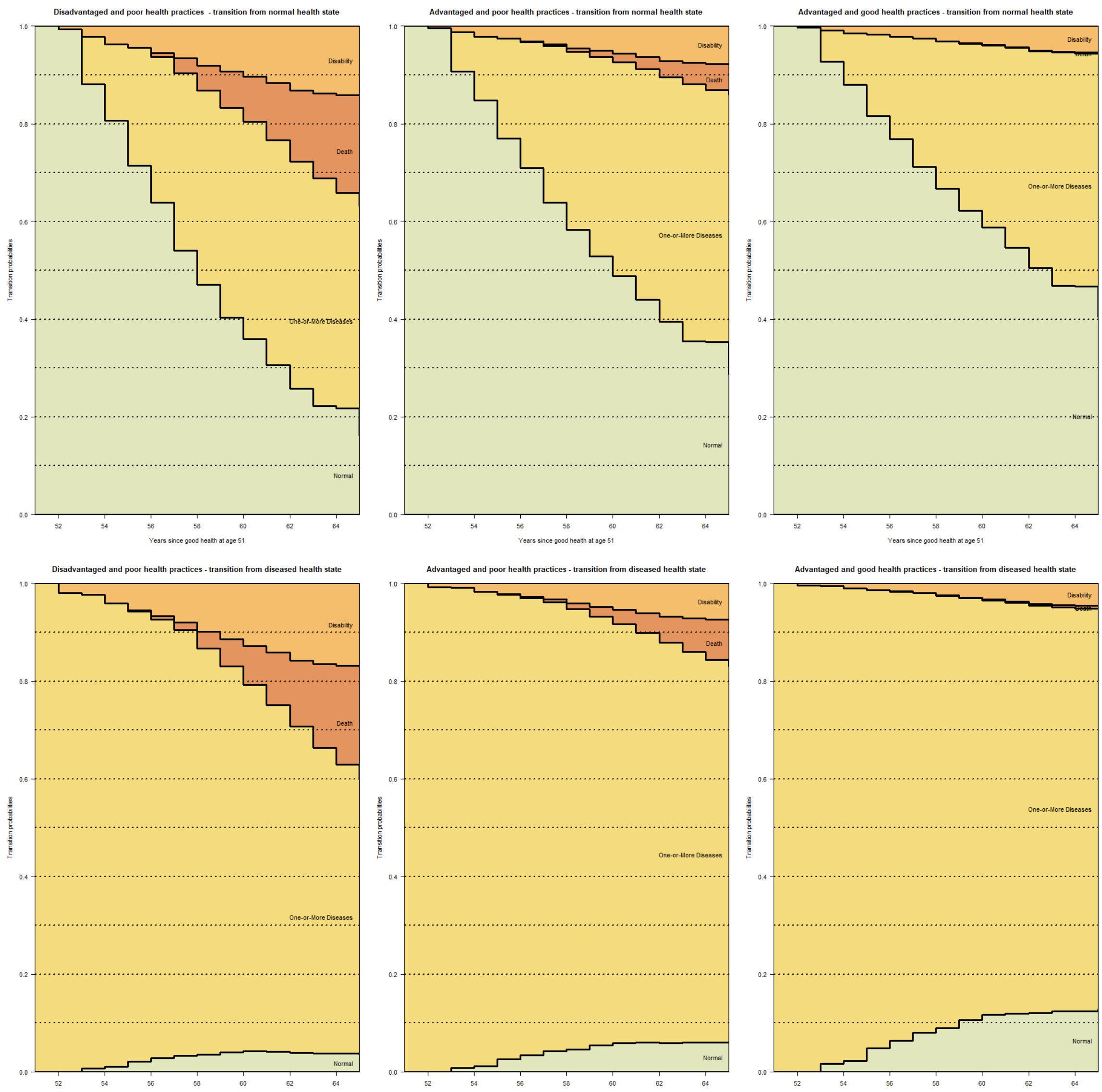

Figure 4: Transition probabilities (a) from normal health state, (b) from one-or-more diseased health state

Table 8: Cox regression estimates of the effects of childhood factors, middle age health status and health behaviors.

Type t: age a	$P_{11}(a)$	$P_{22}(a)$	$P_{12}(a)$	$P_{21}(a)$	$P_{13}(a)$	$P_{23}(a)$	$P_{14}(a)$	$P_{24}(a)$
1: 55	0.714	0.921	0.241	0.021	0.044	0.055	0.000	0.003
2: 55	0.769	0.951	0.205	0.026	0.025	0.022	0.000	0.001
3: 55	0.817	0.938	0.166	0.048	0.018	0.013	0.000	0.000
1: 60	0.360	0.750	0.444	0.043	0.103	0.129	0.093	0.079
2: 60	0.489	0.857	0.437	0.059	0.056	0.054	0.018	0.030
3: 60	0.588	0.848	0.373	0.117	0.039	0.033	0.001	0.002
1: 65	0.162	0.565	0.469	0.035	0.143	0.171	0.226	0.229
2: 65	0.287	0.770	0.573	0.060	0.078	0.075	0.062	0.095
3: 65	0.404	0.817	0.539	0.129	0.054	0.047	0.003	0.007

Childhood factors for mid-age health? cSES appears to be the most significant factor for mid age health. Many factors during childhood of an individual affect the individual's cSES. From policy perspective, let us examine which factors are most important. To that end, 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 = 0.59, family moved for financial reason = 0.49, family usually got financial help during childhood = 0.40, father unemployed during childhood = 0.43, father's usual occupation during childhood = 0.73, father's education = 0.95, mother's education = 0.84.

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.

cSES can be improved by improving college+, which in turn can improve Init.HLTH and the cSES of their children, i.e., it has positive effect on the individual's health outcomes, it also has intergenerational positive effects as it improves cSES cHLTH and probability of college completion and Init.HLTH of their children. It is also important to improve the child health which can have both positive effect on probability of his completing college and Init.HLTH.

8 Conclusion

In this paper, I study 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.

The paper finds that the early childhood socioeconomic status (cSES) of an individual is an important determinant of mid-age health pathways. From policy perspective, the papers finds 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.

After controlling for health behaviors, and early childhood factors, the paper finds that the often attributed significant good effects for women and whites disappeared. The paper finds that the 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.

References

[1] Aalen, O. O., Ø. Borgan, and H. K. Gjessing (2008). Survival and Event History Analysis. Springer New York. DOI: 10.1007/978-0-387-68560-1 (cit. on p. 15).

[2] Aalen, O. O. and S. Johansen (1978). An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations, *Scandinavian Journal of Statistics*, **5**, no. 3, 141–150 (cit. on p. 14).

[3] Albrecht, G. L. and L. M. Verbrugge (2003). "The global emergence of disability", *Handbook of Social Studies in Health and Medicine*. Ed. by G. L. Albrecht, R. Fitzpatrick, and S. C. Scrimshaw. London: Sage, 293–307 (cit. on p. 3).

[4] Alisch, R. S., B. G. Barwick, P. Chopra, L. K. Myrick, G. A. Satten, K. N. Conneely, and S. T. Warren (Feb. 2012). Age-associated DNA methylation in pediatric populations, *Genome Research*, **22**, no. 4, 623–632. DOI: 10.1101/gr.125187.111 (cit. on p. 5).

[5] Altman, B. M. (2001). "Definitions, Models, Classifications, Schemes, and Applications", *Handbook of Disability Studies*. SAGE Publications, Inc., 97–122. DOI: 10.4135/9781412976251.n4 (cit. on p. 3).

[6] An, X. and Y.-F. Yung (2014). Item response theory: what it is and how you can use the IRT procedure to apply it, *SAS Institute Inc. SAS364-2014* (cit. on p. 18).

[7] Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). Statistical Models Based on Counting Processes. Springer-Verlag, New York (cit. on p. 15).

[8] Austad, S. N. and K. E. Fischer (2016). Sex Differences in Lifespan, *Cell Metabolism*, **23**, no. 6, 1022–1033. DOI: 10.1016/j.cmet.2016.05.019 (cit. on p. 5).

[9] Barker, D. J. P. (1990). The fetal and infant origins of adult disease. *BMJ: British Medical Journal*, **301**, no. 6761, 1111 (cit. on p. 5).

[10] Barker, D. J. P. (Aug. 1998). In utero programming of chronic disease, *Clinical Science*, **95**, no. 2, 115–128. DOI: 10.1042/cs0950115 (cit. on p. 5).

[11] Barres, R. and J. R. Zierath (Apr. 2011). DNA methylation in metabolic disorders, *The American Journal of Clinical Nutrition*, **93**, no. 4, 897S–900S. DOI: 10.3945/ajcn.110.001933 (cit. on p. 5).

[12] Bedirhan, O. T., S. Chatterji, N. Kostanjsek, J. Rehm, C. Kennedy, J. Epping-Jordan, S. Saxena, M. von Korff, and C. Pull (May 2010). Developing the World Health Organization Disability Assessment Schedule 2.0, *Bulletin of the World Health Organization*, **88**, no. 11, 815–823. DOI: 10.2471/blt.09.067231 (cit. on p. 4).

[13] Blair, S. N., I. Kohl Harold W., J. Paffenbarger Ralph S., D. G. Clark, K. H. Cooper, and L. W. Gibbons (Nov. 1989). Physical Fitness and All-Cause Mortality: A Prospective Study of Healthy Men and Women, *JAMA*, **262**, no. 17, 2395–2401. DOI: 10.1001/jama.1989.03430170057028 (cit. on p. 5).

[14] Boks, M. P., E. M. Derks, D. J. Weisenberger, E. Strengman, E. Janson, I. E. Sommer, R. S. Kahn, and R. A. Ophoff (Aug. 2009). The Relationship of DNA Methylation with Age, Gender and Genotype in Twins and Healthy Controls, *PLoS ONE*, **4**, no. 8. Ed. by J. Najbauer, e6767. DOI: 10.1371/journal.pone.0006767 (cit. on p. 5).

[15] Bugliari, D., N. Campbell, C. Chan, O. Hayden, M. Hurd, R. Main, J. Mallett, C. McCullough, E. Meijer, M. Moldoff, P. Pantoja, S. Rohwedder, and P. St.Clair (2016). *RAND HRS Data Documentation, Version P*. Tech. rep. (cit. on p. 16).

[16] Cai, L., K. Choi, M. Hansen, and L. Harrell (June 2016). Item Response Theory, *Annual Review of Statistics and Its Application*, **3**, no. 1, 297–321. DOI: 10.1146/annurev-statistics-041715-033702 (cit. on p. 18).

[17] Case, A. and A. Deaton (Nov. 2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century, *Proceedings of the National Academy of Sciences*, **112**, no. 49, 15078–15083. DOI: 10.1073/pnas.1518393112 (cit. on pp. 5, 6).

[18] Conti, R. M., E. R. Berndt, and R. G. Frank (July 2009). "Early Retirement and DI/SSI Applications: Exploring the Impact of Depression", *Health at Older Ages: The Causes and Consequences of Declining Disability among the Elderly*. NBER Chapters. National Bureau of Economic Research, Inc, 381–408 (cit. on p. 6).

[19] Embretson, S. and S. Reise (2000). Item Response Theory. Item Response Theory for Psychologists. Taylor & Francis (cit. on p. 18).

[20] Esteller, M. (Mar. 2008). Epigenetics in Cancer, *New England Journal of Medicine*, **358**, no. 11, 1148–1159. DOI: 10.1056/nejmra072067 (cit. on p. 5).

[21] Fisher, G. G. and L. H. Ryan (Dec. 2017). Overview of the Health and Retirement Study and Introduction to the Special Issue, *Work, Aging and Retirement*, **4**, no. 1. Ed. by M. Wang, 1–9. DOI: 10.1093/workar/wax032 (cit. on p. 16).

[22] Fleming, T. R. (1978). Nonparametric Estimation for Nonhomogeneous Markov Processes in the Problem of Competing Risks, *The Annals of Statistics*, **6**, no. 5, 1057–1070 (cit. on p. 14).

[23] Fleming, T. R. and D. P. Harrington (2005). Counting Processes and Survival Analysis. Wiley (cit. on p. 15).

[24] Gill, R. D. (2005). "Product-integration", *Encyclopedia of Biostatistics*. American Cancer Society. DOI: 10.1002/0470011815.b2a11058 (cit. on p. 13).

[25] Gill, R. D. and S. Johansen (1990). A Survey of Product-Integration with a View Toward Application in Survival Analysis, *The Annals of Statistics*, **18**, no. 4, 1501–1555 (cit. on p. 13).

[26] Gluckman, P. D., M. A. Hanson, C. Cooper, and K. L. Thornburg (July 2008). Effect of In Utero and Early-Life Conditions on Adult Health and Disease, *New England Journal of Medicine*, **359**, no. 1, 61–73. DOI: 10.1056/nejmra0708473 (cit. on pp. 5, 6).

[27] Hahn, H. (1985). Toward a politics of disability: Definitions, disciplines, and policies, *The Social Science Journal* (cit. on p. 3).

[28] Hannum, G. et al. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates, *Molecular Cell*, **49**, no. 2, 359–367. DOI: 10.1016/j.molcel.2012.10.016 (cit. on p. 5).

[29] Heckman, J. J. (July 2008). Schools, Skills and Synapses, *Economic Inquiry*, **46**, no. 3, 289–324. DOI: 10.1111/j.1465-7295.2008.00163.x (cit. on p. 24).

[30] Heckman, J. J. and L. K. Raut (2016). Intergenerational long-term effects of preschool-structural estimates from a discrete dynamic programming model, *Journal of Econometrics*, **191**, no. 1, 164–175. DOI: 10.1016/j.jeconom.2015.10.001 (cit. on pp. 23, 24).

[31] Horvath, S. (Dec. 2013). DNA methylation age of human tissues and cell types, *Genome Biology*, **14**, no. 10, 3156. DOI: 10.1186/gb-2013-14-10-r115 (cit. on p. 5).

[32] Institute of Medicine (2007). Improving the Social Security Disability Decision Process. Ed. by J. D. Stobo, M. McGeary, and D. K. Barnes. Washington, DC: The National Academies Press. DOI: 10.17226/11859 (cit. on p. 4).

[33] Juster, F. T. and R. Suzman (1995). An Overview of the Health and Retirement Study, *The Journal of Human Resources*, **30**, S7. DOI: 10.2307/146277 (cit. on p. 16).

[34] Kaplan, E. L. and P. Meier (1958). Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, **53**, no. 282, 457–481. DOI: 10.1080/01621459.1958.10501452 (cit. on pp. 12, 13).

[35] Karakus, M. C. and L. C. Patton (Feb. 2011). Depression and the Onset of Chronic Illness in Older Adults: A 12-Year Prospective Study, *The Journal of Behavioral Health Services & Research*, **38**, no. 3, 373–382. DOI: 10.1007/s11414-011-9234-2 (cit. on p. 6).

[36] Luo, Y. and L. J. Waite (Mar. 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**, no. 2, S93–S101. DOI: 10.1093/geronb/60.2.s93 (cit. on pp. 6, 23, 25).

[37] Marks, D. (Jan. 1997). Models of disability, *Disability and Rehabilitation*, **19**, no. 3, 85–91. DOI: 10.3109/09638289709166831 (cit. on p. 3).

[38] Nagi, S. Z. (1965). "Some conceptual issues in disability and rehabilitation", *Sociology and Rehabilitation*. Ed. by M. B. Sussman. 100-113: American Sociological Association (cit. on p. 4).

[39] Nagi, S. Z. (1976). An Epidemiology of Disability among Adults in the United States, *The Milbank Memorial Fund Quarterly. Health and Society*, **54**, no. 4, 439. DOI: 10.2307/3349677 (cit. on p. 4).

[40] Nagi, S. Z. (Jan. 1991). "Disability in America: Toward a national agenda for prevention", ed. by A. M. Pope and A. R. Tarlov. Washington, DC: National Academies Press. Chap. Disability concepts revised: Implication for prevention, 309–327. DOI: 10.17226/1579 (cit. on p. 4).

[41] Pope, A. M. and A. R. Tarlov (Jan. 1991). Disability in America: Toward a national agenda for prevention. Ed. by A. M. Pope and A. R. Tarlov. National Academies Press. DOI: 10.17226/1579 (cit. on p. 4).

[42] Raut, L. K. (2017). Exits from Disability: Estimates from a Competing Risk Model, *Social Security Bulletin*, **77**, no. 3, 15–38 (cit. on pp. 8, 10).

[43] Raut, L. K. (June 2018). Long-term Effects of Preschool on School Performance, Earnings and Social Mobility, *Studies in Microeconomics*, **6**, no. 1-2, 24–49. DOI: 10.1177/2321022218802023 (cit. on p. 24).

[44] Raut, L. K. (2019). Pathways to disability and mortality in mid-ages: Estimates from the Health and Retirement Study Data, *Draft Presented at the 2019 SGE Annual Conference* (cit. on p. 12).

[45] Renna, F. (Oct. 2008). Alcohol Abuse, Alcoholism, and Labor Market Outcomes: Looking for the Missing Link, *ILR Review*, **62**, no. 1, 92–103. DOI: 10.1177/001979390806200105 (cit. on p. 6).

[46] Seib, C., E. Whiteside, K. Lee, J. Humphreys, T. H. D. Tran, L. Chopin, and D. 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**, no. 1, e43–e52. DOI: 10.1016/j.whi.2013.11.004 (cit. on p. 6).

[47] Simons, R. L., M. K. Lei, S. R. Beach, R. A. Philibert, C. E. Cutrona, F. X. Gibbons, and A. Barr (2016). Economic hardship and biological weathering: The epigenetics of aging in a U.S. sample of black women, *Social Science & Medicine*, **150**, 192–200. DOI: 10.1016/j.socscimed.2015.12.001 (cit. on p. 6).

[48] Snyder, A. R., J. T. Parsons, T. C. Valovich McLeod, R. Curtis Bay, L. A. Michener, and E. 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**, no. 4, 428–436. DOI: 10.4085/1062-6050-43.4.428 (cit. on p. 3).

[49] Sonnega, A., J. D. Faul, M. B. Ofstedal, K. M. Langa, J. W. Phillips, and D. R. Weir (Mar. 2014). Cohort Profile: the Health and Retirement Study (HRS), *International Journal of Epidemiology*, **43**, no. 2, 576–585. DOI: 10.1093/ije/dyu067 (cit. on p. 16).

[50] Steffick, D. E. (2000). Documentation of affective functioning measures in the Health and Retirement Study, *Ann Arbor, MI: University of Michigan* (cit. on p. 19).

[51] The US Burden of Disease Collaborators (Apr. 2018). The State of US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among US States, *JAMA*, **319**, no. 14, 1444–1472. DOI: 10.1001/jama.2018.0158 (cit. on p. 5).

[52] Vable, A. M., P. Gilsanz, T. T. Nguyen, I. Kawachi, and M. M. Glymour (Oct. 2017). Validation of a theoretically motivated approach to measuring childhood socioeconomic circumstances in the Health and Retirement Study, *PLOS ONE*, **12**, no. 10. Ed. by A. Fraser, e0185898. DOI: 10.1371/journal.pone.0185898 (cit. on p. 24).

[53] Vaupel, J. W. (Mar. 2010). Biodemography of human ageing, *Nature*, **464**, no. 7288, 536–542. DOI: 10.1038/nature08984 (cit. on p. 5).

[54] Verbrugge, L. M. and A. M. Jette (1994). The disablement process, *Social Science & Medicine*, **38**, no. 1, 1–14. DOI: 10.1016/0277-9536(94)90294-1 (cit. on p. 4).

[55] Verbrugge, L. M., K. Latham, and P. J. Clarke (Feb. 2017). Aging With Disability for Midlife and Older Adults, *Research on Aging*, **39**, no. 6, 741–777. DOI: 10.1177/0164027516681051 (cit. on p. 4).

[56] World Health Organization (2001). *International Classification of Functioning, Disability and Health*. Tech. rep. Geneva (cit. on p. 4).

[57] Wreede, L. C. de, M. Fiocco, and H. Putter (Sept. 2010). The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models, *Computer Methods and Programs in Biomedicine*, **99**, no. 3, 261–274. DOI: 10.1016/j.cmpb.2010.01.001 (cit. on p. 20).

[58] Zarulli, V., J. A. Barthold Jones, A. Oksuzyan, R. Lindahl-Jacobsen, K. Christensen, and J. W. Vaupel (2018). Women live longer than men even during severe famines and epidemics, *Proceedings of the National Academy of Sciences*, **115**, no. 4, E832–E840. DOI: 10.1073/pnas.1701535115 (cit. on p. 5).