Pathways to disability and mortality in mid-ages: Estimates from the Health and Retirement Study Data[*]

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

DRAFT: please do not distribute

mailto:Lakshmi.Raut@ssa.gov

April 5, 2019

Abstract

The paper studies determinants of enrollment onto the Social Security's DI (Disability Insurance) and SSI (Supplemental Security Income) programs or death before reaching age 65. As individuals age, the homeostatic regulatory mechanism that controls physiological body systems—such as respiratory, cardiovascular, neuroendocrine, immune, and metabolic—becomes more and more frail in its ability to face

[*]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, had normal BMI, CES-D and other standard bio-markers (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.

This 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, Congressional Budget Office, and from the audience. I had many insightful discussions with Han Altae-Tran, MIT and my colleagues at the Indian Statistical Institute, Kolkata. I am grateful to all of them.

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

internal and external stressors. This leads to one-or-comorbid chronic diseases (such as diabetes, cancer, vascular, musculoskeletal, cognitive and mental disorders), DI-qualifying disabilities or death. The speed of progression through these health states depends on the rate of depletion of one's homeostatic health and frailty levels. Genetic and epigenetic factors comprised of internal and external environments, health care use, health related behavior and cognitive endowments modulate the depletion rate of homeostatic health over life cycle. These, in turn, determine the likelihood of various pathways through the health states. I use the Health and Retirement Study (HRS) data to estimate a multi-state time-to-event model of pathways through these health states. I use bio-markers such as BMI, CES-D, measures of cognitive health, and indicators of health related behaviors such as smoking, drinking, 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 the disability programs — Social Security Disability Insurance (DI) and the Supplemental Security Income (SSI). According to the biology of living organisms, individuals succumb to aging, and experience diseases and disabilities of various kinds. 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 of time 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 death through various health states—such as 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 is studied in this paper. The huge disability literature shows, however, that there is no general consensus on the definition of disability. 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 Hahn (1985); Altman (2001) for general discussions and Snyder et al. (2008) in the context of sports). The original antique definition is based on the ”medical model of disability”. According to this view, a serious disease or an injury causes impairments of the body structures, limiting ‘normal’ functional activities, which constitutes a disability. What is normal is controversial with bearings on social and political machinery to discriminate and demean others by the so-called normal people.

The medical definition has been modified later with the ”social model of disability”. This definition incorporates the social environment—involving family, community and workplace environment—in which an individual with a medical impairment operates to perform activities of daily living. The examples of some of the environments are the availability of accessible transportation and workplace accommodations (for more on this, see Albrecht and Verbrugge (2003); Marks (1997); Altman (2001)). This definition is further refined in the context of workers’ compensation or legal settlements with the “economic model” of disability. The economic model views disability as the inability to work gainfully or as a loss of productivity due to a disease or injury.

The definition of disability and its modifications over time for the DI and SSI programs uses a combination of the above definitions. The current definition of disability for the DI is specified in the Social Security Act, Title II, § 223(d), paragraph (1)A as

> “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] I use the above definition of disability in this paper. For details on how the above statutory definition of disability is used in the determination of disability, see Institute of Medicine (2007).

The above models do not provide a biological or a behavioral mechanism of disablement process from which policy implications for clinical practice and health care policy can be derived. In this direction, researchers introduced what is known as disablement models. 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, to functional limitations and finally to disability, briefly described below.

Pathology is an interruption of normal physiological processes 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 sport and physical recreation, reading, and distinct trips). The final stage is the disability, the definition of which depends on the purpose and involves a combination of all the above models of disability. The first disablement model 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

[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.

biological, environmental and behavioral risk factors affecting all four stages of the disablement process. 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 Üstün 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.

The disablement models are useful for conceptualization, diagnosis and record keeping of disabilities but limited for the study of the causes of disability in epidemiological and policy research. The starting point of the above disablement models is an onset of a chronic disease or an injury causing a disability. For policy research on disability and mortality, it is important to study the biomedical processes modulated by genetic, epigenetic and behavioral factors in the manifestation and prognosis of disabling diseases and on the risk of disabling injuries. While an injury as the starting point of disablement process serves well for certain purpose such as for workers' compensation in sports, construction and factories, a large proportion of disabilities in the mid ages are caused by diseases. Diseases, leading to disabilities — both developmental disabilities and late age disabilities — 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. 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 aging process and incidence of disease, disability and death over the lifespan?

I will not get into the details of the biomedical literature on these issues. See Raut (2019) for a more detailed analysis along this line. I briefly summarize the main findings from this literature along the lines of the above questions, and get guidance to formulate the statistical model to address our main issues. Similar to the literature of behavioral genetics of personality and intelligence, the *nature-nurture* controversy exists in the health literature: Is it all nature (i.e., all genetics or genome) or is it all nurture (i.e., all epigenetics or epigenome modulated by the environment and health related individual behaviors) that determines the

progression of health over the life span of an individual? The consensus so far is that it is neither the nature nor the nurture, it is a combination of both that determines health developments over one's life-span. The research so far found that certain genetic make-ups (i.e., certain sequences of DNA) predispose one to certain diseases, (Barondes (1999); Khoury et al. (2009); Bookman et al. (2011)), but the epigenetic inputs—especially at the very early stage of life, i.e. in the womb, but not the least at later stages of life—are also very important determinants of life expectancy and quality of life. The biomedical research so far has not found genes that are responsible for aging and age related diseases leading to early disability and mortality. The twenty-first century biomedical research emphasizes more on the epigenetic factors than the genetic factors to explain the pattern of health developments over the life-span.

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

Entringer, et al. (2013); DiLoreto and Murphy (2015); Shalev and Belsky (2016); Simons et al. (2016)).

More recently emerged second line of biomedical research on aging and aging related diseases explores the epigenetic (which literally means on top of genetic) mechanism for these life-cycle processes. The twenty-century emphasis that the full DNA mapping of human genome will be able to uncover fully the mechanism of human development pathways fell short of explaining why identical twins diverge so much in their gene expressions or phenotypes as they progress through their lives—the phenomenon is known as the *epigenetic drift* (see Boks et al., 2009). Although all cells in the body have the same genome (i.e., same set of instructions or genes that produce all different types of cells of the human body and brain), cells get epigenetic signals both internally and from other neighboring cells and environment outside the body and from health related individual inputs. The epigenetic signals are chemical molecules on top of the DNA sequence. A configuration of the collection of all these molecules on top of the DNA sequence is known as an *epigenome*. Coding of epigenome is more complex than the coding of genome. DNA methylation, noncoding RNA (ncRNA)-associated gene silencing, histone modifications and chromatin remodeling of DNA are the currently known main ways that initiate and sustain epigenetic change, which are responsible for aging process and incidence of age related diseases, disabilities and death. The internal and external environments of the body cells influence the epigenome, the changes in which over the life-span determine which genes are expressed, silenced, or mutated during cell divisions, and hence determine the development of health status and diseases as life progresses. Recent research is trying to develop instruments to measure the changes in major CpG islands (i.e., clusters of CpG sequences where most of the methylation of CpG subsequences occur) and associate the changes to aging process and incidence of age related diseases, see for instance, Esteller (2008); Barres and Zierath (2011); Alisch et al. (2012); Horvath (2013); Hannum et al. (2013).

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

Research along this line began with the striking findings of Barker (1990); Barker (1998) and later of Gluckman et al. (2008). They found strong associations between birth weight

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

Studies in social sciences find that low socio-economic status (SES) are associated with inflammation, metabolic dysregulation, and various chronic and age-related diseases such as type 2 diabetes, coronary heart disease, stroke, and dementia, and that low SES create epigenetic changes in individuals that lead to faster biological aging even after controlling for health-related behaviors such as diet, exercise, smoking, alcohol consumption, or having health insurance, see for evidence, Simons et al. (2016). The study by Karakus and Patton (2011) uses the Health and Retirement Studies data and after controlling for education, race, income, health risk indicators like BMI and smoking, functional limitations like gross motor index, health limitations for work, and income, they find depression at baseline leads to significantly higher risk for developing diabetes, heart problems, and arthritis and no significant effect on developing cancer during the 12 years follow-up period. Renna (2008) uses National Longitudinal Survey of Youth data to find no significant effect of alchohol 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.

I adopt all the above views and formulate a statistical model of disablement process. I postulate that as individuals age, 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 death. I draw from the microbiology literature that study the genetic and epigenetic mechanism for aging and timing and severity of aging related diseases, disability and death. I use available bio-markers (such as BMI, CES-D, cognition) and health related behaviors such as smoking, drinking, 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 most 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. See Crowther and Lambert (2017) for a similar reasoning.

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 specification and estimation issues. In Section 4, I describe the subset of the Health and Retirement Survey dataset that I use for this study . In Section 5, I present the estimates and discuss the results. Section 6 concludes the paper.

2 The Model

Two ways to model event history with multiple events—as a multivariate counting process, (see for instance, Commenges, 2015, Section 3.7, Andersen, Borgan, et al., 1993, Section II.4 and Aalen, Borgan, and Gjessing (2008)) or as a multistate stochastic process $X(t), t \in T$, where at each time point t, the random variable $X(t)$ takes a value from a finite number of health states in S. I follow the second approach.

Let $T = [51, 65]$ and unit of time is a year. In the state space S we denote states by $1 =$ "healthy or normal health", $2 =$ "ill with one or more chronic disease", $3 =$ "disabled with DI-or SSI-qualifying disability" and $4 =$ "Death". Sometimes we 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

$$P_{hj}(s,t) = Prob(X(t) = j | X(s) = h), \tag{1}$$

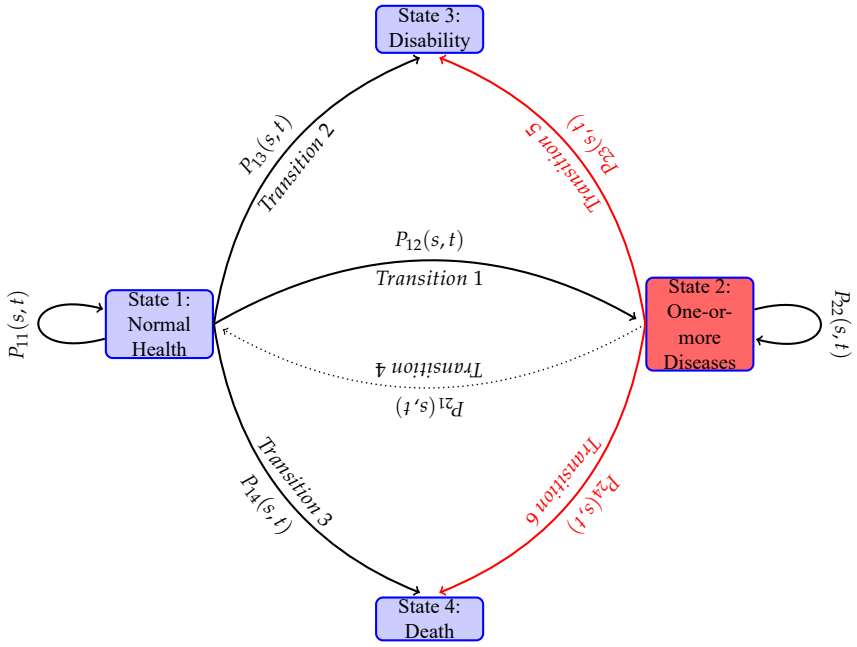
for all $h, j \in S$, $s, t \in T$, $t \geq s$. Denote the matrix of transition probabilities by

$$P(s,t) \equiv \left(P_{hj}(s,t)\right)_{h,j=1\ldots p}. \tag{2}$$

An individual may be in any of the health states in S at time t, the probability of which, known as the occupation probability, depends on the occupation probabilities of various health states in previous periods and the transition probability matrix. Let $\pi_j(t)$ be the occupation probability of an individual in health state j at time t. Alternatively, occupation probability can be viewed as the proportion of population of age t who are in health state j. 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,

$$\pi(t) = \pi'(s)P(s,t), 0 \leq s < t. \tag{3}$$

The following path diagram describes various pathways that individuals may go through in their mid-ages.



The transition probabilities of the stochastic process satisfies Chapman-Kolmogorov equation

$$P(s,t) = P(s,u) \cdot P(u,t), \text{ for all } s,u,t \in T \text{ with } s < u < t \tag{4}$$

I assume that the transition probabilities $P(s,t)$ are absolutely continuous in s and t. A *transition intensity*—also known as the *hazard rate* in the survival analysis literature when there is only one event of interest and the *cause-specific hazard rate* in the competing risk analysis[2]—of the health process X_t from health state h to health state j at time t is the derivative

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} P_{hj}(t, t+\Delta t)/\Delta t \tag{5}$$

While $P_{hj}(s,t)$ is an unconditional probability, the transition intensity or the hazard rate $\alpha_{hj}(t)\Delta t$ is the conditional (instantaneous) probability of an individual experiencing the event j in the 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 (which in our case is the chronological age) and other characteristics and the path through various health states that is followed by the individual to be at 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 it is in at time s, and the time s and t, but does not depend on the path that he followed to come to state h at time s. "The future health condition depends on the current health condition but not how one came to the current health state." Some of the other cases are explained in Section 3.

Denote by $\alpha(t) = \left(\alpha_{hj}(t)\right)_{h,j=1,\ldots,p}$, the matrix of transition intensities. It can be shown[3] that the Chapman-Kolmogorov equation leads to the following Kolmogorov forward equation

$$\frac{\partial P(s,t)}{\partial t} = P(s,t) \cdot \alpha(t) \tag{6}$$

Note that $P(s,s) = I$ for all $s \in T$, which can be taken as the initial condition of the above system of differential equations. Thus, given intensities $\alpha(t)$, the solution $P(s,t)$ of the above differential equation defines a continuous time Markov chain and conversely, given absolutely continuous transition probabilities $P(s,t)$ for a continuous time Markov chain, one has the Kolmogorove forward equation Eq. (6). For the empirical work, I characterize

[2]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.

[3]Note that $P(s,t+\Delta t) = P(s,t) \cdot P(t,t+\Delta t)$, i.e., $\frac{P(s,t+\Delta t)-P(s,t)}{\Delta t} = P(s,t)\frac{(P(t,t+\Delta t)-I)}{\Delta t}$, taking limit $\Delta t \to 0$ on both sides, and utilizing the definition in equation (5) we have $\frac{\partial P(s,t)}{\partial t} = P(s,t) \cdot \alpha(t)$.

the Markov chain of pathways through various health states with the intensity matrix $\alpha(t)$ and in our empirical specification, I parameterize $\alpha(t)$ as function of covariates and estimate these using the HRS data set. I then use these estimates of the intensity matrix to estimate the transition probabilities and study their properties.

From the Fundamental Theorem of ordinary differential equations, we know that there exists a solution $P(s,t)$ to the system of ordinary differential equations in Eq. (6). In general, it is not possible to find analytical solution of the Kolmogorov forward equation. However, given the special structure for the intensity matrix and the assumption that there is no transition to normal health from the diseased health state and the disability health state, we can solve analytically[4] the Kolmogorove forward equation Eq. (6). The matrix of transition intensities in Eq. (6) becomes the following,

$$\alpha(t) = \begin{pmatrix} -(\lambda_{12}(t) + \lambda_{13}(t) + \lambda_{14}(t)) & \lambda_{12}(t) & \lambda_{13}(t) & \lambda_{14}(t) \\ 0 & -(\lambda_{23}(t) + \lambda_{24}(t)) & \lambda_{23}(t) & \lambda_{24}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{7}$$

and the system of differential equations in Eq. (6) becomes,

$$\frac{\partial P_{11}(s,t)}{\partial t} = -(\lambda_{12}(t) + \lambda_{13}(t) + \lambda_{14}(t))P_{11}(s,t) \tag{8}$$

$$\frac{\partial P_{12}(s,t)}{\partial t} = \lambda_{12}(t) P_{11}(s,t) - (\lambda_{23}(t) + \lambda_{24}(t))P_{12}(s,t) \tag{9}$$

$$\frac{\partial P_{13}(s,t)}{\partial t} = \lambda_{13}(t) P_{11}(s,t) + \lambda_{23}(t) P_{12}(s,t) \tag{10}$$

$$\frac{\partial P_{14}(s,t)}{\partial t} = \lambda_{14}(t) P_{11}(s,t) + \lambda_{24}(t) P_{12}(s,t) \tag{11}$$

$$\frac{\partial P_{22}(s,t)}{\partial t} = -(\lambda_{23}(t) + \lambda_{24}(t)) P_{22}(s,t) \tag{12}$$

$$\frac{\partial P_{23}(s,t)}{\partial t} = \lambda_{23}(t) P_{22}(s,t) \tag{13}$$

$$\frac{\partial P_{24}(s,t)}{\partial t} = \lambda_{24}(t) P_{22}(s,t) \tag{14}$$

with the initial condition, $P(s,s) = I$. Note that Eq. (8), Eq. (12) - Eq. (14) are differential equations of the form $dy(t)/dt = \lambda(t)y(t)$, which has the solution $y(t) =$

[4] I follow Andersen and Perme (2008).

$exp\left[\int_s^t \lambda(u)du\right]$. Furthermore, Eq. (10) and Eq. (11) are of the form $dy(t)/dt = g(t)$ where the function $g(t)$ does not depend on y, but are funciton of other transition probabilities. The solution for this is easily derived to be $y(t) = exp\left[\int_s^t g(u)du\right]$. The remaining Eq. (9), is of the form

$$dy(t)/dt + \lambda(t)y(t) = g(t) \tag{15}$$

which is more difficult to solve. Various techniques are available to solve this type of differential equations. I use the integrating factor technique. Define the integrating factor $exp\left[\int_s^t \lambda(u)du\right]$, multiply both sides of (Eq. (15)) with this integrating factor and then simplifying the left hand side using the multiplicative rule for differentiation, one gets $\frac{d}{dt}y(t)exp\left[\int_s^t \lambda(u)du\right] = g(t)exp\left[\int_s^t \lambda(u)du\right]$. This has the solution,

$$y(t)exp\left[\int_s^t \lambda(u)du\right] = \int_s^t \left[g(u)exp\left[\int_s^u \lambda(\tau)d\tau\right]\right]du.$$

Dividing both sides by the integrating factor, and utilizing the split, $exp\left[\int_s^t \lambda(u)du\right] = exp\left[\int_s^u \lambda(u)du + \int_u^t \lambda(u)du\right]$, one gets $y(t) = \int_s^t \left[g(u)exp\left[-\int_u^t \lambda(\tau)d\tau\right]\right]du$. Now substituting the original values for $g(t)$ and $\lambda(t)$, and noting that the exponential term is indeed the already solved $P_{22}(u,t)$ (see Eq. (20) below), we have solutions for all the equations as follows

$$P_{11}(s,t) = \exp\left(-\int_s^t (\lambda_{12}(u) + \lambda_{13}(u) + \lambda_{14}(u))\,du\right) \tag{16}$$

$$P_{12}(s,t) = \int_s^t P_{11}(s,u)\,\lambda_{12}(u)\,P_{22}(u,t)\,du \tag{17}$$

$$P_{13}(s,t) = \int_s^t \lambda_{13}(u)\,P_{11}(s,u)\,du + \int_s^t \left[P_{11}(s,u)\lambda_{12}(u)\int_u^t P_{22}(u,\tau)\,\lambda_{23}(\tau)\,d\tau\right]du \tag{18}$$

$$P_{14}(s,t) = \int_s^t \lambda_{14}(u)\,P_{11}(s,u)\,du + \int_s^t \left[P_{11}(s,u)\lambda_{12}(u)\int_u^t P_{22}(u,\tau)\,\lambda_{24}(\tau)\,d\tau\right]du \tag{19}$$

$$P_{22}(s,t) = \exp\left(-\int_s^t (\lambda_{23}(u) + \lambda_{24}(u))\,du\right) \tag{20}$$

$$P_{2h}(s,t) = \int_s^t \lambda_{2h}(u)\,P_{22}(s,u)\,du, h = 3,4 \tag{21}$$

The above formulas for the transition probabilities have interesting interpretations. If we view state 1 as "alive" and states 2, 3 and 4 (the states that an individual can move to from

state 1) as the competing causes of death, then by the definition of transition probabilities in Eq. (1), $P_{11}(0,t)$ is the probability that the individual is alive at time t, i.e., did not die from any of the competing causes 2, 3 or 4 of death before age t. That is, $P_{11}(0,t)$, is nothing but the survival function of the competing risk literature. Denote by $\lambda_1(t) = \lambda_{12}(t) + \lambda_{13}(t) + \lambda_{14}(t)$, the combined risk or the hazard rate of exiting state 1 at time t. Then $\int_0^t \lambda_1(u)\, du$ is the cumulative hazard (also known as the cumulative risk or the integrated hazard) of dying by time t. The Eq. (16) is nothing but the well-known formula in the survival analysis that survival function $S(t) = exp[-A(t)]$. In $P_{11}(s,t)$ the individual begins state 1 at time s in stead of at time 0, and is more relevant in the multi-state models as the individual could be in some other health state before moving to health state at time s. In what follows, I will sometimes denote a transition $j \to h$ by a Greek letter or when the reference of the from state j and the to-state h is important, I will denote it by jh. Furthermore, I will often abbreviate a transition probability $P_{jh}(0,t)$ as $P_{hj}(t)$.

The interpretation of the other transition probabilities are slightly more complex in the multi-state context as there are multiple pathways to move from one state to another state and not all other states that one moves to are absorbing states. For instance, the transition probability $P_{12}(s,t)$ is by definition the probability of an individual being in health state 2 (i.e., have one-or-more diseases) at time t given that he was in health state 1 at time s. The formula for it in Eq. (17) means that for an individual in health state 1 at time s to be in health state 2 at time t, he has to be in health state 1 up to a time u, $(s < u < t)$, the probability of which is $P_{11}(s,u)$, and make an instaneous transition at time u (i.e., during $[u, u+du)$ from state 1 to state 2, the probability of which is $\lambda_{12}(u)du$ and remain in state 2 during the remaining time u through t, the probability of which is $P_{22}(u,t)$). Moreover, this transition time u takes any of the mutually exclusive values between s to t and thus we need to integrate over these mutually exclusive values of u between s and t, which is represented in the formula in Eq. (17).

$P_{13}(s,t)$, the probability of a person who is in normal health, i.e. in state 1 at time s to be on the DI rolls, i.e., in state 3 at time t has two parts—corresponding to the two mutually exclusive paths that can lead to this: First, he can be in good health state until time u $(s < u < t)$ (the probability of which is $P_{11}(s,u)$) and then transit to the disability health state 3 at time u with probability $\lambda_{13}(u)du$. Second, from state 1 at time s, he moves to state 2 at time u $(s < u < t)$ (the probability of which is $P_{12}(s,u)$) and then transit to the disability health state 3 at time u with probability $\lambda_{13}(u)$. Once arrived on the on he

stays there until t with probability 1 (as it is an absorbing state) Since this is true for any value of u, we integrate over u to get the probability. I denote the probability of the direct path $1 \to 3$ as $Q_{12}(s,t)$ and the probability of the path $1 \to 2 \to 3$ as $Q_{123}(s,t)$. Similar is the interpretation of the two components for $P_{14}(s,t)$ in Eq. (19).

These qualitative properties of our disablement model can be studied from analytical solutions of these transition probabilities. In general it is not possible to get analytical solutions. However, I derive it in the next subsection under the assumption that the intensity matrix $\alpha(t)$ is time constant, i.e., independent of time t.

2.1 Time constant intensity process and solution of the transition probabilities

To get more insight about the interdependence of the transition probabilities and their effects on each other, I consider the time constant case for which the transition probabilities can be solved analytically. Assume that $\lambda_{hj}(t)$ is constant over time, i.e., independent of t, and denote it by λ_{hj}. Denote by $\lambda_1 \equiv \lambda_{12} + \lambda_{13} + \lambda_{14}$, which is the intensity of exiting state 1 from any of the three competing risks of exits from the state and the transition intensity of state 2 by $\lambda_2 \equiv \lambda_{23} + \lambda_{24}$.

For the time homogeneous case becomes, it is straightforward to calculate all that the transition probabilities from the formula, which are given in Eq. (22)

$$
\begin{aligned}
P_{11}(t) &= \exp(-\lambda_1 t), \\
P_{22}(s,t) &= \exp(-\lambda_2 (t-s)), \\
P_{12}(t) &= \frac{\lambda_{12}}{\lambda_2 - \lambda_1}\left[\exp(-\lambda_1 t) - \exp(-\lambda_2 t)\right] \\
P_{13}(t) &= \frac{\lambda_{13}}{\lambda_1}\left[1 - \exp(-\lambda_1 t)\right] + \frac{\lambda_{12}\lambda_{23}}{\lambda_2}\left[\frac{\exp(-\lambda_2 t) - \exp(-\lambda_1 t)}{\lambda_2 - \lambda_1} + \frac{1 - \exp(-\lambda_1 t)}{\lambda_1}\right] \\
P_{14}(t) &= \frac{\lambda_{14}}{\lambda_1}\left[1 - \exp(-\lambda_1 t)\right] + \frac{\lambda_{12}\lambda_{24}}{\lambda_2}\left[\frac{\exp(-\lambda_2 t) - \exp(-\lambda_1 t)}{\lambda_2 - \lambda_1} + \frac{1 - \exp(-\lambda_1 t)}{\lambda_1}\right] \\
P_{2h}(t) &= \frac{\lambda_{2h}}{\lambda_2}\left(1 - \exp(-\lambda_2 t)\right), h = 3, 4.
\end{aligned}
\tag{22}
$$

In Eq. (22), the first term on the right hand side of $P_{13}(t)$ is the probability of directly transiting to state 3 from 1, i.e. $Q_{13}(t)$ and the second term is the probability of the indirect path $1 \to 2 \to 3$, i.e., $Q_{123}(t)$ in the notation of the previous section. Similar is the case for

$P_{14}(t)$. It is possible to caluclulate these probabilities separately for our time homogeneous case.

Notice how the transition probabilities are interdependent and how say $P_{13}(t)$ changes when say there is a reduction in λ_{24} due to discoveries of medical technology that reduces the probability of death from diseased state.

In the next section, I estimate the constant transition intensities from our data set using maximum likelihood procedure and study dynamics of the transition probabilities for various groups.

3 The data set and the variables

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. RAND created many variables from the original HRS data for ease of use. I create my dataset and all the variables with a few exceptions mentioned below from the RAND HRS dataset version P. The details of the Rand HRS version P can be found in Bugliari et al. (2016).

As mentioned in the introduction, I define the disability health state to be 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, how-

ever, 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 received disability before the first survey year 1992 and I also drop the spouses in the sample who were not born between 1931 to 1941, that is the respondents in our sample are between age 51 to 61 and not disabled or dead in 1992. I ended up with the final sample size of 9,493 for this analysis.

Table 1 and Table 2 provide a few characteristics of the data.

Table 1: Percentage distribution of the pooled sample population by health status by age

Age	#obs	Percentage distriubtion of HealthStatus			
		Normal	With Diseases	Disabled	Died at age
51	945	47.62	52.38	0.00	0.00
52	936	47.65	52.35	0.00	0.00
53	1906	42.71	56.19	0.73	0.37
54	1850	41.95	57.08	0.76	0.22
55	2791	39.59	58.47	1.29	0.64
56	2684	37.30	61.07	0.97	0.67
57	3572	35.39	62.46	1.12	1.04
58	3469	32.92	64.72	1.33	1.04
59	4240	31.04	66.11	1.75	1.11
60	4182	29.94	67.62	1.51	0.93
61	4894	27.42	69.55	1.45	1.57
62	4080	25.51	70.20	2.16	2.13
63	4746	23.68	72.48	1.85	1.98
64	3905	21.95	75.29	0.72	2.05
65	4564	20.86	76.40	0.66	2.08

Source: The author.

3.1 Variables

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 college and above (does not include some college), i.e., has a college degree and more and taking

Table 2: Distribution of the sample by health status in various survey rounds

Age	#obs	Percentage Distribution of HealthStatus			
		Normal	With Diseases	Disabled	Died in period
1992	9493	39.65	60.35	0.00	0.00
1994	9493	34.16	62.73	1.75	1.36
1996	9198	30.17	66.72	1.52	1.59
1998	7461	27.48	69.09	1.81	1.62
2000	5791	25.38	71.14	1.49	1.99
2002	4106	22.80	74.35	1.32	1.53
2004	2437	20.68	75.91	1.40	2.01
2006	785	17.83	79.75	0.38	2.04

Source: The author.

value 0 otherwise.

Age: This is a time varying covariate not perfectly correlated with the duration of occupancy in a health state that the respondent is currently in. No perfect correlation because individuals enter a health state not all at the same age. Thus the transition intensity may have separate effect of age as compared to the duration dependence built in the baseline hazard function. Even at the beginning of the study period, i.e. in 1992, we have individuals begin their initial health states of their health trajectories at different ages (more precisely ages 51 to 61 years).

cesd: I used a 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 dynamic IRT (item response theory) on these responses and used the first dominant factor's loadings for each individual, and define the variable behav_prev to take value 1 if the factor loading of the variable in a survey year is above the mean of the factor loading score and 0 otherwise. This makes the measure to be on the same footing for all individuals and time periods. For the years 1992 and other years with missing values, I imputed it as the first future non-missing value.

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 for all the years.

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.

4 Econometric parameterization and Estimation

Two types of models exist to study the effects of covariates X on the transition probabilities parameterized through transition intensities - (1) more recently introduced Aalan additive model (Aalen, 1989; Aalen, Borgan, and Fekjær, 2001) and (2) most widely used multiplicative Cox regression model, which I use in this paper.

For statistical formulation, for each individual i, define the random process, $Y_{h,i}(t) = I(X_i(t_-) = h)$, i.e., it takes value 1 at time t if the individual i is in health state $h, h = 1, 2$, i.e. the individual is in risk for exit from any of the competing causes of exit from that health state, and $T^*_{h,i}$ as the completed duration if already exited health state h because of any cause before age 65, otherwise the time from entry point to age 65, i.e., the censoring time. For each possible transition, $h \rightarrow j$ and individual i who are in the health state h, define a counting process, $N_{hj,i}(t) = \#$ of transitions of type that the individual i made during the time interval $[0, t]$. I assume that when there is a transition to a non-absorbing state, the individual becomes new individual who begins a new process from time $t = 0$ from that state with covariate values reset to the values that exist at the beginning of this episode.

Analogous to survival analysis, two approaches are followed to estimate a multiplicative regression, i.e., a Cox regression model. One is the parametric approach which assumes a parametric form for the baseline intensity $\lambda^0_{hj}(t)$ such as the Exponential distribution, Weibull distribution, or more flexible exponentiated Weibull distribution[5] and then maximize the likelihood function

$$L(\theta) = \prod_i \prod_{h \rightarrow j, h \neq j, \substack{h=1,2 \\ j=2,3,4}} \left(\prod_t \lambda_{hj,i}(t|X_{h,i})^{\triangle N_{hj,i}(t)} \right) \exp\left(- \int_0^{T^*_{h,i}} \lambda_{hj,i}(u|X_{h,i}) \right) du \tag{23}$$

Even without covariates, computation of the parameters of the extended Weibull or even the Weibull model is difficult.

---

[5]

$$\lambda^0_{hj}(t) = \frac{\theta_{hj} \cdot \frac{\nu_{hj}}{\sigma_{hj}} \left(\frac{t}{\sigma_{hj}}\right)^{\nu_{hj}-1} \exp[-\left(\frac{t}{\sigma_{hj}}\right)^{\nu_{hj}}] \left(1 - \exp[-\left(\frac{t}{\sigma_{hj}}\right)^{\nu_{hj}}]\right)^{\theta_{hj}-1}}{1 - \left(1 - \exp[-\left(\frac{t}{\sigma_{hj}}\right)^{\nu_{hj}}]\right)^{\theta_{hj}}}$$

The above transition intensities provide flexible forms for the baseline hazards and nests Weibull when $\theta_{hj} = 1$, and exponential hazard when $\theta_{hj} = 1$ and $\nu_{hj} = 1$ (taking $\lambda_{hj} = 1/\sigma_{hj}$).

4.1 Estimates of Time Constant Transition Intensities

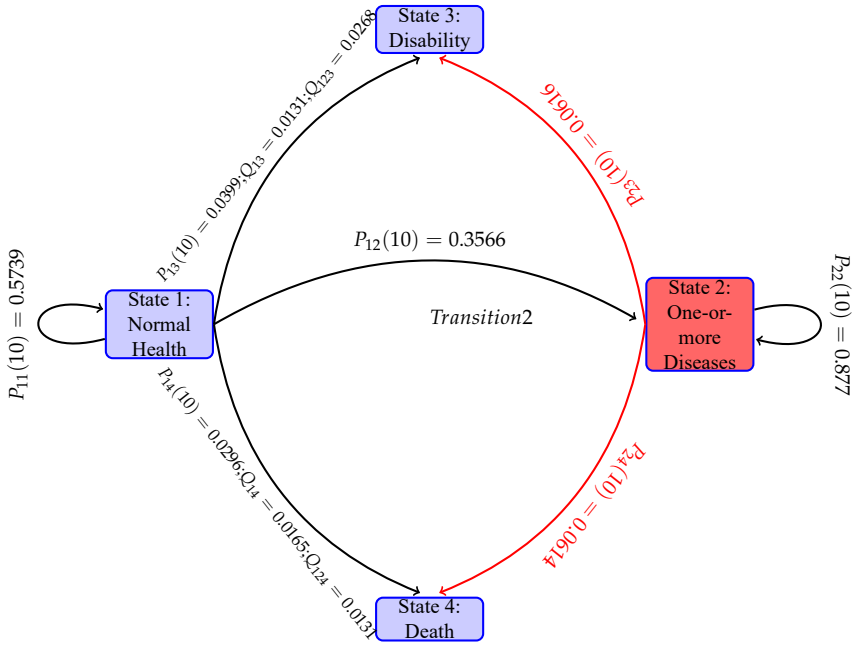
For time constant hazard, $\lambda_{hj}(t) = \lambda_{hj}$, i.e. for exponential model, close-form solution can be derived, which after simplification is given by

$$\hat{\lambda}_{hj} = \frac{\overline{N}_{hj}}{\bar{T}_h^*}; \quad V(\hat{\lambda}_{hj}) = \frac{\lambda_{hj}^2}{\overline{N}_{hj}}; \quad \widehat{s.e}(\hat{\lambda}_{hj}) = \frac{\sqrt{\overline{N}_{hj}}}{\bar{T}_h^*}$$

where $\overline{N}_{hj} = \sum_{i=1}^{n_h} \int_0^t N_{hj,i}(t)\, dt$ is the total number of transitions of type $h \to j$, n_h is the total number of individuals in the health state h, and the common denominators in all the expression $\bar{T}_h^* = \sum_{i=1}^{n_h} T_{h,i}^*$ is the total of all completed transition times and censor times of individuals who are in health state h, $(h = 1, 2)$ in the extended sample. For derivation of these estimators, see Moore, 2016, pp.22.

I first report the estimates of time constant transition intensities for the overall sample and illustrate the interdependence of the transition probabilities and how they evolve over time. Then I compare these estimates of transition intensities and transition probabilities for a selected few groups in Table 4.

Probabilities of following various paths for the estimated time constant parameters: $\hat{\lambda}_{12} = 0.04990, \hat{\lambda}_{13} = 0.00349, \hat{\lambda}_{14} = 0.00215, \hat{\lambda}_{23} = 0.00657, \hat{\lambda}_{24} = 0.00656$.



To get an idea about these transition probabilities over time, I plot transition probabilities

out of state 1 in Figure 1 panel(a) and out of state 2 in Figure 1 panel(b).

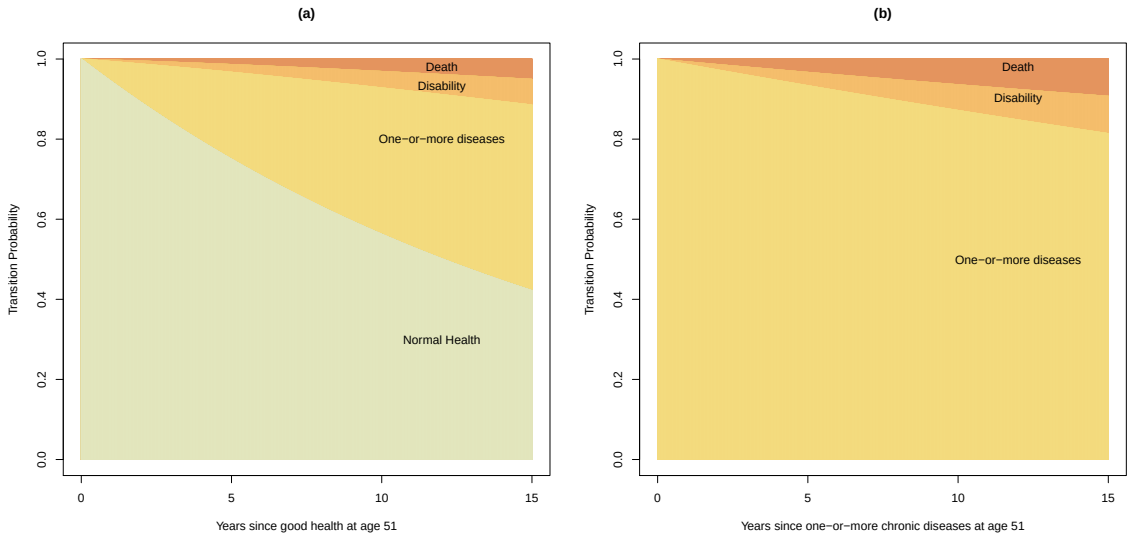


Figure 1: Transition probabilities (a) from normal health state, (b) from one-or-more diseased health state

It will be interesting to compute the distributions of waiting times T_{13} and T_{123} and compare them empirically. Parameter estimates and selected transition probabilities at 10 years are given in Table 3.

4.2 Estimations of Time Non-homogeneous Non-parametric Transition Intensities

Without covariates, the maximization of the likelihood function Eq. (23) the non-parametric maximum likelihood estimate of the integrated hazard function known as the *Nelson-Aalen estimator*

$$\begin{aligned}
\hat{\Lambda}_{hj}(t) &= \int_0^t \frac{d\bar{N}_{hj}(u)}{\bar{Y}_h(u)} du = \sum_{i:t_i \leq t} \frac{\triangle \bar{N}_{hj}(t_i)}{\bar{Y}_h(t_i)}, h \neq j, h = 1,2 \;\; j = 1,...4, \\
\hat{\Lambda}_{hh}(t) &= -\sum \hat{\Lambda}_{hj}(t), h = 1,2, \\
\hat{\Lambda}_{hj}(t) &= 0 \text{ for all other } h, j \text{ combinations}
\end{aligned}$$

The slope of the integrated hazard $\hat{\Lambda}_{hj}(t)$ at time t gives the estimate of transition intensity $\hat{\lambda}_{hj}(t)$.

Table 3: Transition Probabilities for transitions $i \rightarrow j$ by duration of stay in each state 1 and 2 for the model with time constant intensities

duration	$1 \rightarrow 1$	$2 \rightarrow 2$	$1 \rightarrow 2$	$1 \rightarrow 3$	$2 \rightarrow 3$	$1 \rightarrow 4$	$2 \rightarrow 4$
1	0.9460	0.9870	0.0482	0.0036	0.0065	0.0023	0.0065
2	0.8949	0.9741	0.0932	0.0072	0.0130	0.0047	0.0129
3	0.8465	0.9614	0.1351	0.0110	0.0193	0.0073	0.0193
4	0.8008	0.9488	0.1742	0.0149	0.0256	0.0101	0.0256
5	0.7575	0.9365	0.2105	0.0189	0.0318	0.0130	0.0317
6	0.7166	0.9243	0.2443	0.0230	0.0379	0.0161	0.0378
7	0.6779	0.9122	0.2757	0.0271	0.0439	0.0193	0.0439
8	0.6413	0.9003	0.3048	0.0313	0.0499	0.0226	0.0498
9	0.6066	0.8886	0.3317	0.0356	0.0558	0.0261	0.0557
10	0.5739	0.8770	0.3566	0.0399	0.0616	0.0296	0.0614
11	0.5429	0.8655	0.3797	0.0443	0.0673	0.0332	0.0672
12	0.5135	0.8543	0.4009	0.0487	0.0729	0.0369	0.0728
13	0.4858	0.8431	0.4204	0.0531	0.0785	0.0407	0.0784
14	0.4596	0.8321	0.4384	0.0576	0.0840	0.0445	0.0839
15	0.4347	0.8213	0.4548	0.0621	0.0895	0.0484	0.0893

Source: The author.

Using the Nelson-Aalan estimates for the integrated hazard functiond, I estimate the transition probabilities as follows

$$\hat{P}(t) = \prod_{0<u<t} \left(I + d\hat{\Lambda}(u)\right) = \prod_{i:t_i \leq t} \left(I + \left[\hat{\Lambda}(t_i) - \hat{\Lambda}(t_{i-1})\right]\right).$$

This estimator is known as *Aalen-Johansen estimator*. I use the R package, *mstate*, developed and described by the authors in Wreede et al. (2010) for the estimation of the parameters and their standard errors?

Figure 2 panel(a) shows the probabilities of a representative individual (i.e., one with the mean value of all the regressors) of age 50 to remain in normal health, contact one-or-more diseases, become disabled or die as the years pass by. Similarly, Figure 2 panel(b) shows the corresponding probabilities for an individual of age 50 who is in the health state of one-or more diseases.

Table 4: Maximum Likelihood Estimates of the Time Constant Intensities and Transition Probabilities at 10 years for Various Groups

sample	n	λ_{12}	λ_{13}	λ_{14}	λ_{23}	λ_{24}	$P_{13}(10)$	$P_{23}(10)$	$P_{14}(10)$	$P_{24}(10))$
overall	11095	0.050	0.003	0.002	0.007	0.007	0.040	0.062	0.030	0.061
White	8900	0.050	0.003	0.002	0.006	0.006	0.037	0.057	0.026	0.056
Non-White	2195	0.050	0.005	0.004	0.009	0.009	0.051	0.079	0.046	0.080
Female	5841	0.052	0.003	0.002	0.007	0.005	0.037	0.062	0.023	0.048
Male	5254	0.048	0.004	0.003	0.007	0.008	0.042	0.061	0.037	0.077
College	1967	0.043	0.001	0.001	0.003	0.004	0.015	0.027	0.017	0.041
No college	9128	0.052	0.004	0.002	0.007	0.007	0.046	0.068	0.033	0.065
bmi $>$ 25	6980	0.057	0.004	0.002	0.007	0.006	0.042	0.064	0.029	0.057
bmi $<=$ 25	4113	0.042	0.003	0.002	0.006	0.008	0.037	0.055	0.031	0.072
Smoker	6891	0.051	0.004	0.003	0.008	0.009	0.044	0.070	0.041	0.079
Nonsmoker	4092	0.049	0.003	0.001	0.005	0.003	0.030	0.047	0.011	0.032
Drinker	4421	0.052	0.003	0.001	0.004	0.003	0.030	0.042	0.017	0.032
Non-drinker	6674	0.048	0.004	0.003	0.008	0.009	0.047	0.073	0.038	0.079
Exercise	8604	0.050	0.003	0.001	0.005	0.003	0.033	0.052	0.018	0.030
No exercise	2490	0.049	0.007	0.006	0.010	0.017	0.071	0.088	0.074	0.150

Source: Author's calculation.
Note: All hazard intensities are significant at 1 percent level.

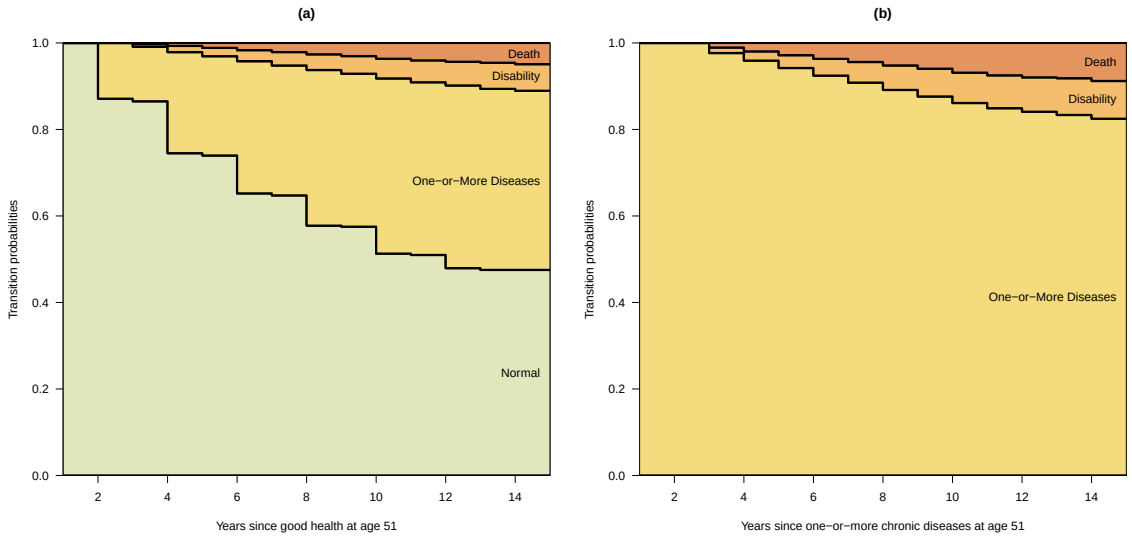


Figure 2: Transition probabilities (a) from normal health state and (b)from one-or-more diseased health state

Table 5: Transition probabilities for transition $i \rightarrow j$ by duration of stay in state 1 and 2 from the semi-parametric multistate model

duration	$1 \rightarrow 1$	$2 \rightarrow 2$	$1 \rightarrow 2$	$1 \rightarrow 3$	$2 \rightarrow 3$	$1 \rightarrow 4$	$2 \rightarrow 4$
0	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.8711	1.0000	0.1289	0.0000	0.0000	0.0000	0.0000
3	0.8651	0.9767	0.1259	0.0062	0.0127	0.0027	0.0106
4	0.7451	0.9591	0.2337	0.0143	0.0213	0.0069	0.0195
5	0.7397	0.9420	0.2296	0.0193	0.0296	0.0114	0.0283
6	0.6522	0.9245	0.3057	0.0255	0.0387	0.0167	0.0369
7	0.6476	0.9082	0.3003	0.0310	0.0477	0.0211	0.0441
8	0.5776	0.8913	0.3598	0.0363	0.0566	0.0263	0.0520
9	0.5752	0.8762	0.3537	0.0407	0.0643	0.0304	0.0596
10	0.5130	0.8611	0.4050	0.0455	0.0701	0.0366	0.0688
11	0.5098	0.8491	0.3993	0.0505	0.0760	0.0404	0.0749
12	0.4793	0.8409	0.4223	0.0549	0.0795	0.0435	0.0796
13	0.4752	0.8335	0.4186	0.0604	0.0849	0.0458	0.0816
14	0.4752	0.8247	0.4142	0.0615	0.0873	0.0491	0.0880
16	0.4752	0.8247	0.4142	0.0615	0.0873	0.0491	0.0880

Source: The author.

Compare the non-parametric time non-homogeneous transition probability estimates in Table 5 with the parametric time-homogeneous estimates in Table 4 for the overall sample without covariates. They are very close to each other.

5 Estimations of Time Non-homogeneous Semi-parametric Cox Regression Models of Transition Intensities

With covariates, Andersen, Borgan, et al. (1993) derive the following generalized Cox partial likelihood to get parameter estimates of the Cox regression models

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

and the cumulative baseline hazard is estimated by

$$\begin{aligned}
\hat{\Lambda}_{hj}\left(t|X_h^0\right) &= \int_0^t \frac{Y_h(u)\, d\bar{N}_{hj}(u)}{\hat{S}_{hj}^0\left(\hat{\beta}_{hj}, u\right)} = \sum_{i:t_i \le t} \frac{\triangle \bar{N}_{hj}(t_i)}{\hat{S}_{hj}^0\left(\hat{\beta}_{hj}, t_i\right)}, h \neq j, h = 1,2,\ j = 1,..4 \\
\hat{\Lambda}_{hh}\left(t|X_h^0\right) &= -\sum \hat{\Lambda}_{hj}\left(t|X_h^0\right), h = 1,2, \\
\hat{\Lambda}_{hj}\left(t|X_h^0\right) &= 0 \text{ for all other } h, j \text{ combinations}
\end{aligned}$$

where $\hat{S}_{hj}^0\left(\hat{\beta}_{hj}, u\right) = \sum_{i=1}^{n_h} Y_{hj,i}(t) \exp\left(\hat{\beta}_{hj}' X_{h,i}^0\right)$. The Aalen-Johansen estimator of the transition probabilities is

$$\hat{P}\left(s,t|X^0\right) = \prod_{s<u<t}\left(I + d\hat{\Lambda}\left(u|X^0\right)\right) = \prod_{i:t_i \le t}\left(I + \left[\hat{\Lambda}\left(t_i|X^0\right) - \hat{\Lambda}\left(t_{i-1}|X^0\right)\right]\right).$$

For the Cox regression parameter estimates, I have used both the R package *mstate* (see, Wreede et al., 2010 for details) and also used the SAS procedure *phreg* (both produced the same estimates) and used the *mstate* package to estimate all the transition probabilities (SAS does not have readily available procedure for this purpose). The parameter estimates are shown in Table 6 and Table 7.

Table 6: Estimates of Cox regression models separately for each transition

	1->2	1->3	1->4	2->3	2->4
White	-0.0143	-0.3263	-0.7321^{**}	-0.3650^{***}	-0.4121^{***}
	(0.0628)	(0.2308)	(0.2650)	(0.1012)	(0.1006)
Female	0.0479	-0.2105	-0.5196^{*}	-0.0644	-0.5104^{***}
	(0.0463)	(0.1904)	(0.2499)	(0.0911)	(0.0922)
Age	-0.0823^{***}	-0.0568	0.0058	-0.1132^{***}	-0.0499^{***}
	(0.0081)	(0.0324)	(0.0421)	(0.0143)	(0.0135)
AIC	25259.8561	1740.3244	1063.5163	8073.3477	8057.7550
R^2	0.0249	0.0016	0.0031	0.0102	0.0081
Max. R^2	0.9992	0.3756	0.2537	0.6949	0.6935
Num. events	1602	112	69	476	475
Num. obs.	3583	3695	3652	6856	6855
Missings	0	0	0	0	0
PH test	0.0000	0.0125	0.0772	0.0000	0.0000

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

With only demographic covariates (most that can be done with the Administrative data) the parameter estimates in Table 6 show that significantly lower risks of transitions $1 \rightarrow 4$; $2 \rightarrow 3$; $2 \rightarrow 4$ for whites and $1 \rightarrow 4$ and $2 \rightarrow 4$ for women. This may entail that the genetic make-up of being white or female sex yield favorable genetic predisposition to have better health outcomes and longer life. This is a misleading inference as we will see next that when we control for epigenetic factors that the biomedical literature pointed out to have significant effects on aging process and health outcomes, the above effects disappear.

Table 7 shows the Cox regression coefficient estimates of the effects of various factors on the risk of having transitions $h \rightarrow j$ from health status $h = 1, 2$ to health status $j, j = 2, 3, 4$.

These estimates show that the parameter estimates for the demographic covariates in Table 6 are biased as they are capturing the effects of excluded epigenetic and behavioral factors in that model. After controlling for these epigenetic and behavioral factors, the significance of those effects disappear. Furthermore, women show significantly lower probability of transition from diseased health states onto the disability health state.

Most important factors in Table 7 are cesd, measuring depression and stress and college graduation or higher level of education, with positive effect on all transitions with the exception of no effect on transition from normal health to death.

Other important factors are smoking, with significant adverse effect on transitions, and exercising three or more times regularly has significant favorable effect on most transitions.

To understand the significance of these estimates, say the parameter estimate of -1.0293 for the parameter of $behav\_vigex$ in the model for transition $1 \rightarrow 3$ is that given all other factors equal, an individual of a given age with normal health will have a $0.36 \equiv exp(-1.0293)$ times lower risk of becoming disabled before age 65 if the person does moderately vigorous exercise compared to the probability of one who does not. From the diseased health state, the probability becoming disabled before age 65 is $0.55 \equiv exp(-0.5974)$ times lower compared to the corresponding probability of one who does not. The effect of other parameters can be easily read from the table.

6 Conclusion

In this paper, I study the determinants of falling chronically ill with one-or-more diseases, the likelihood of getting on the Social Security's DI (Disability Insurance) program or the

Table 7: Estimates of Cox regression models separately for each transition with health measures

	1->2	1->3	1->4	2->3	2->4
White	0.0086	0.0038	−0.3675	−0.1056	−0.1622
	(0.0671)	(0.2939)	(0.5095)	(0.1102)	(0.1624)
Female	0.0772	−0.3895	−0.4448	−0.2038*	−0.2308
	(0.0511)	(0.2390)	(0.4995)	(0.1037)	(0.1508)
College	−0.1768**	−0.9687*	−0.7994	−0.6165**	−0.5416*
	(0.0641)	(0.4333)	(0.7701)	(0.1950)	(0.2635)
Age	−0.0907***	−0.0849*	−0.1386*	−0.1076***	−0.1102***
	(0.0085)	(0.0378)	(0.0693)	(0.0150)	(0.0215)
cesd	0.5065***	1.8802***	0.0940	1.2030***	0.4952*
	(0.1092)	(0.3641)	(1.0818)	(0.1631)	(0.2526)
cogtot	−0.0112	−0.0662*	0.0192	−0.0328***	−0.0028
	(0.0058)	(0.0260)	(0.0440)	(0.0096)	(0.0160)
bmi	0.0398***	−0.0088	0.0141	0.0158	−0.0233
	(0.0055)	(0.0315)	(0.0536)	(0.0095)	(0.0170)
behav_prev	0.3828***	0.7397**	−19.7473***	0.1402	−1.5869***
	(0.0487)	(0.2270)	(0.2451)	(0.0974)	(0.1845)
behav_smoke	0.0436	0.2626	2.5071*	0.4057***	0.8133***
	(0.0504)	(0.2264)	(1.0096)	(0.1093)	(0.1817)
behav_drink	0.0307	−0.3579	0.5318	−0.4993***	−0.2507
	(0.0497)	(0.2414)	(0.4481)	(0.1118)	(0.1518)
behav_vigex	−0.2467***	−1.0293***	−0.9911	−0.5974***	−1.0397***
	(0.0702)	(0.2535)	(0.5155)	(0.1058)	(0.1447)
AIC	23182.4941	1370.5925	309.5095	7250.9572	3212.5472
R^2	0.0689	0.0302	0.0159	0.0471	0.0415
Max. R^2	0.9993	0.3529	0.0990	0.7050	0.4406
Num. events	1500	95	23	446	207
Num. obs.	3239	3334	3262	6165	5926
Missings	344	361	390	691	929
PH test	0.0000	0.0034	0.1191	0.0000	0.0000

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

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 general consensus in the biomedical literature is that much of an individual's later life health outcomes is programmed at an early stage of life—as early as the prenatal stage, most importantly right after the conception stage. The programming is strongly modulated by the epigentic inputs created by the environment in mother's womb. The genetic predisposition also matters. But epigentic factors modulate quite strongly the programming for later life developments in cognitive and non-cognitive health. The most important epigenetic factor is stress of any kind— psychological, financial, social and chemical—and other significant factors are diet, smoking, substance use, and exercising. These modulating factors are important throughout life, with stronger effects imparted in early stages of life.

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

I first estimate the model with only demographic variables as covariates, and found that whites have significantly lower risk of transiting from good health to death and from diseased health state to disability or death. The women found to have significantly lower

risk of death than men, both from normal health and diseased health, and have no difference with men in becoming disabled. These are commonly observed phenomena found in other studies. These are the effects one can study using only the Administrative data.

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

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

References

[1] Aalen, O. O. (1989). A linear regression model for the analysis of life times, *Statistics in Medicine*, **8**, no. 8, 907–925. DOI: 10.1002/sim.4780080803 (cit. on p. 20).

[2] Aalen, O. O., Ø. Borgan, and H. Fekjær (2001). Covariate Adjustment of Event Histories Estimated from Markov Chains: The Additive Approach, *Biometrics*, **57**, no. 4, 993–1001. DOI: 10.1111/j.0006-341X.2001.00993.x (cit. on p. 20).

[3] 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. 9).

[4] 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).

[5] 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. 7).

[6] 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).

[7] Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993). Statistical Models Based on Counting Processes. Springer-Verlag, New York (cit. on pp. 9, 25).

[8] Andersen, P. K. and M. P. Perme (Sept. 2008). Inference for outcome probabilities in multi-state models, *Lifetime Data Analysis*, **14**, no. 4, 405–431. DOI: 10.1007/s10985-008-9097-x (cit. on p. 12).

[9] 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. 6).

[10] Barbara, M. A., Y. Abdilla, and J. Calleja-Agius (May 2017). An Introduction to Epigenetics, *Neonatal Network*, **36**, no. 3, 124–128. DOI: 10.1891/0730-0832.36.3.124 (cit. on p. 8).

[11] Barker, D. J. P. (May 2007). The origins of the developmental origins theory, *Journal of Internal Medicine*, **261**, no. 5, 412–417. DOI: 10.1111/j.1365-2796.2007.01809.x (cit. on p. 8).

[12] Barker, D. J. P. (1990). The fetal and infant origins of adult disease. *BMJ: British Medical Journal*, **301**, no. 6761, 1111 (cit. on p. 7).

[13] 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. 7).

[14] Barondes, S. (1999). Molecules and Mental Illness. Scientific American Library (cit. on p. 6).

[15] 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. 7).

[16] 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. 6).

[17] 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. 7).

[18] Bookman, E. B., K. McAllister, E. Gillanders, K. Wanke, D. Balshaw, J. Rutter, J. Reedy, D. Shaughnessy, T. Agurs-Collins, D. Paltoo, and et al. (2011). Gene-environment interplay in common complex diseases: forging an integrative model-recommendations from an NIH workshop, *Genetic Epidemiology*, n/aâ€“n/a. DOI: 10.1002/gepi.20571 (cit. on p. 6).

[19] 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. RAND Center for the Study of Aging (cit. on p. 16).

[20] Commenges, D. (2015). Dynamical biostatistical models. Boca Raton: CRC Press (cit. on p. 9).

[21] 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. 9).

[22] Crowther, M. J. and P. C. Lambert (2017). Parametric multistate survival models: Flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences, *Statistics in Medicine*, **36**, no. 29, 4719–4742. DOI: 10.1002/sim.7448 (cit. on p. 9).

[23] DiLoreto, R. and C. T. Murphy (Dec. 2015). The cell biology of aging, *Molecular Biology of the Cell*, **26**, no. 25. Ed. by W. Bement, 4524–4531. DOI: 10.1091/mbc.e14-06-1084 (cit. on p. 7).

[24] Epel, E. S., E. H. Blackburn, J. Lin, F. S. Dhabhar, N. E. Adler, J. D. Morrow, and R. M. Cawthon (2004). Accelerated telomere shortening in response to life stress, *Proceedings of the National Academy of Sciences*, **101**, no. 49, 17312–17315. DOI: 10.1073/pnas.0407162101 (cit. on p. 6).

[25] Esteller, M. (Mar. 2008). Epigenetics in Cancer, *New England Journal of Medicine*, **358**, no. 11, 1148–1159. DOI: 10.1056/nejmra072067 (cit. on p. 7).

[26] 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).

[27] 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. 7, 8).

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

[29] Hannum, G. et al. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates, *Molecular Cell*, **49**, no. 2, 359–367. DOI: https://doi.org/10.1016/j.molcel.2012.10.016 (cit. on p. 7).

[30] Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains, *Experimental Cell Research*, **37**, no. 3, 614–636. DOI: 10.1016/0014-4827(65)90211-9 (cit. on p. 6).

[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. 7).

[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 pp. 4, 5).

[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] Kanherkar, R. R., N. Bhatia-Dey, and A. B. Csoka (Sept. 2014). Epigenetics across the human lifespan, *Frontiers in Cell and Developmental Biology*, **2**, no. 3, 124–128. DOI: 10.3389/fcell.2014.00049 (cit. on p. 8).

[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. 8).

[36] Khoury, M. J. et al. (June 2009). Genome-Wide Association Studies, Field Synopses, and the Development of the Knowledge Base on Genetic Variation and Human Diseases, *American Journal of Epidemiology*, **170**, no. 3, 269–279. DOI: 10.1093/aje/kwp119 (cit. on p. 6).

[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] Moore, D. F. (2016). Applied Survival Analysis Using R. Springer International Publishing. DOI: 10.1007/978-3-319-31245-3 (cit. on p. 21).

[39] 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).

[40] 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).

[41] 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).

[42] 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. 5).

[43] Raut, L. K. (2017). Exits from Disability: Estimates from a Competing Risk Model, *Social Security Bulletin* (cit. on p. 11).

[44] Raut, L. K. (2019). Progression of health, mortality and morbidy in the aging process: genetics, epigenetics and fetal programming, *(mimeo) University of Chicago* (cit. on p. 5).

[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. 8).

[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. 8).

[47] Shalev, I. and J. Belsky (2016). Early-life stress and reproductive cost: A two-hit developmental model of accelerated aging?, *Medical Hypotheses*, **90**, 41–47. DOI: https://doi.org/10.1016/j.mehy.2016.03.002 (cit. on p. 7).

[48] Shalev, I., S. Entringer, P. D. Wadhwa, O. M. Wolkowitz, E. Puterman, J. Lin, and E. S. Epel (2013). Stress and telomere biology: A lifespan perspective, *Psychoneuroendocrinology*, **38**, no. 9, 1835–1842. DOI: https://doi.org/10.1016/j.psyneuen.2013.03.010 (cit. on p. 6).

[49] 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: https://doi.org/10.1016/j.socscimed.2015.12.001 (cit. on pp. 7, 8).

[50] 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).

[51] 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).

[52] Steffick, D. E. (2000). Documentation of affective functioning measures in the Health and Retirement Study, *Ann Arbor, MI: University of Michigan* (cit. on p. 18).

[53] Thornburg, K. L., J. Shannon, P. Thuillier, and M. S. Turker (2010). "In Utero Life and Epigenetic Predisposition for Disease", *Epigenetics and Cancer, Part B*. Elsevier, 57–78. DOI: 10.1016/b978-0-12-380864-6.00003-1 (cit. on p. 8).

[54] Üstün, T. B., 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. 5).

[55] Vaupel, J. W. (Mar. 2010). Biodemography of human ageing, *Nature*, **464**, no. 7288, 536–542. DOI: 10.1038/nature08984 (cit. on p. 6).

[56] 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).

[57] 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).

[58] World Health Organization (2001). *International Classification of Functioning, Disability and Health*. Tech. rep. Geneva: World Health Organization (cit. on p. 5).

[59] 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 pp. 23, 26).

[60] 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. 6).