HEC MONTRÉAL
affiliée à l'Université de Montréal

Predictive Validity of Subjective Survival Probabilities and Its Implications for Selection Problem in Annuity Markets

par

Ainur Artay

Pierre-Carl Michaud
HEC Montréal
Directeur de recherche

Sciences de la gestion
Spécialisation Économie Financière Appliquée

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)

Résumé

Cette étude démontre la validité prédictive des perceptions subjectives du risque de mortalité, remettant en question la croyance selon laquelle les perceptions de risque des individus sont intrinsèquement biaisées. En utilisant les données de *Health and Retirement Study*, nous examinons si les probabilités de survie subjectives prédisent le risque de mortalité et si elles sont corrélées avec les horloges épigénétiques, des biomarqueurs du vieillissement qui sont fortement prédictifs de la mortalité. Nos résultats montrent que les probabilités de survie subjectives sont des prédicteurs significatifs de la mortalité et qu'elles sont fortement corrélées avec les horloges épigénétiques, qui sont des prédicteurs avérés du vieillissement et de la mortalité, particulièrement l'horloge *GrimAge*. Ces deux résultats suggèrent que les individus possédent des informations privées sur leur santé et leur bien-être qui sont plus informatives du risque de mortalité que les caractéristiques observables.

Abstract

This study presents evidence for the predictive validity of subjective mortality risk perceptions, challenging the belief that individuals' risk perceptions are inherently distorted. Using data from the Health and Retirement Study, we examine whether subjective survival probabilities are predictive of mortality risk and if they correlate with epigenetic clocks, biomarkers of aging that are highly predictive of mortality outcomes. Our findings demonstrate that subjective survival probabilities are significant predictors of mortality and that they strongly correlate with epigenetic clocks, which are proven predictors of aging and mortality, particularly the GrimAge clock. Both of these results suggest that individuals possess private information about their health and well-being that is above and over informative of mortality risk than observable characteristics.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I want to thank my advisor, Professor Pierre-Carl Michaud. Working with a master firsthand is the best education one can get, and it is the surest way to raise the bar. I am incredibly fortunate to have had his guidance, invaluable insights, and support.

I want to extend my gratitude to the two most important people in my life, my husband Ross and my daughter Gemma, for showering me with encouragement, endless joy, and love during my thesis journey.

1 Introduction

The uncertainty surrounding individual longevity presents a significant financial risk known as longevity risk. This risk emerges from the potential of outliving one's retirement savings due to a lifespan that exceeds the expectation. Empirical evidence of longevity risk is found in demographic studies that reveal heterogeneity in life expectancy, influenced by factors such as genetics, lifestyle, and socio-economic status. Longevity risk is a critical concern not only for individuals planning for retirement but also for institutions managing pension schemes, and for government policymakers.

One strategy to mitigate longevity risk is through the purchase of life annuities (hereafter referred to simply as annuities), which are financial instruments designed to provide a consistent stream of income for the duration of an individual's life in exchange for a lump-sum payment. However, the decision to annuitize is heavily influenced by individual differences in life expectancy, which affect the perceived fairness of annuity pricing. If the mortality risk perceived by an individual is higher than the risk estimates used by insurers to price annuities, akin to a price loading, this can lead to a perception of unfair pricing by the potential annuitant. Given that agents are price elastic (Boyer et al. (2020)), annuities perceived as overpriced may not be considered optimal, thus lowering their take-up rate.

Indeed, empirical evidence consistently demonstrates suboptimal take-up rates for annuities despite their apparent benefits, a phenomenon termed 'the annuity puzzle' by Modigliani (1986). The optimality of full annuitization was first formally demonstrated by Yaari (1965), and later expanded upon by Davidoff et al. (2005), who showed that annuitization should remain optimal even under a general set of assumptions, with exceptions including the presence of bequest motives and market imperfections. Benartzi et al. (2011) offer numerous behavioral and economic factors contributing to market's divergence from this optimality, including presence of adverse selection, which arises when individuals posses private information about their mortality risk. This study examines subjective mortality risk perceptions to

understand if they are predictive of the actual risk and potentially lead to adverse selection.

The demand for annuities is a function of individuals' risk perceptions. If the subjective risk perceptions are not correlated with the objective mortality risk, it suggests that these perceptions are largely based on biases, which are non-informative of the actual outcomes, thereby not contributing to adverse selection. Conversely, if, within a risk class, these risk perceptions do predict actual mortality outcomes, it may indicate that those opting for annuities have disproportionately higher longevity risk than average. This can result in an adverse selection spiral, whereas insurers, recognizing the higher risk profile of annuity purchasers, adjust their pricing models accordingly. The resultant increase in annuity prices can deter individuals at lower mortality risk from purchasing these products, thereby exacerbating the issue of adverse selection. This dynamic ultimately may lead to a market equilibrium characterized by higher annuity prices and lower overall take-up rates (Einav and Finkelstein, 2011; Boyer et al., 2020).

The predictive validity of the subjective risk perceptions can be examined data on mortality and individuals' risk perceptions. Indeed, previous studies have consistently shown that people's own estimates of their survival chances are predictive of their actual mortality. Hurd and McGarry (2002) demonstrated this using subjective survival probabilities (SSPs) from the Health and Retirement Study (HRS) panel data on the older US population. They found that those who survived in the panel reported survival probabilities approximately 50% greater at baseline than those who died. Using data from the same study, the HRS, Smith et al. (2001) concluded that actual mortality outcomes were "signaled" through the reported lower longevity expectations. Similar findings were reported examining data on populations of European countries (Post and Hanewald, 2013; Delavande and Rohwedder, 2011).

The objective of this thesis is to understand whether individual differences in subjective risk perceptions are predictive of actual mortality. To achieve this, the study empirically tests the correlation of SSP, first, with observed deaths using data from the HRS and second, with epigenetic clocks, using newly available data on from

the Venous Blood Study (VBS) within the HRS. We refer to the first part of our study as *the mortality risk analysis* and the second part – as *the epigenetic clock analysis*.

Epigenetic clocks serve as biomarkers of aging, integrating genetic and epigenetic information to estimate an individual's biological age. Unlike traditional genetic markers that focus on DNA sequence variations, epigenetic clocks capture dynamic changes in methylation influenced by both genetic and environmental, lifestyle factors, such as smoking and stress.In essence, they are values derived from complex calculations of DNA methylation levels — a biochemical process influencing gene expression without altering the DNA sequence — across specific genomic regions. Epigenetic clocks provide a comprehensive understanding of an individual's aging process, extending beyond chronological age.

Despite their recent emergence, epigenetic clocks have proven valuable in aging and health studies by accurately estimating age-related diseases and mortality. In healthcare, they find applications in predicting disease risk, prognosis, and treatment responses. In longevity research, they offer insights into factors influencing lifespan. In public health, they inform interventions for healthy aging and mitigate age-related health disparities. Beyond medicine, epigenetic clocks extend to forensic applications, estimating age in criminal investigations.

In studies on SSP, integrating epigenetic clocks provides a new perspective on factors influencing individuals' perceptions of their mortality risk. Comparing subjective and biological age estimates explores alignment or divergence between beliefs and objective biological indicators, assessing SSP accuracy and reliability. This insight is valuable for financial decision-making and retirement planning implications.

On the downside, using epigenetic clocks for actuarial purposes introduces significant costs associated with the collection and computation of DNA data, limiting the feasibility of their widespread adoption in insurance practices. A central objective of this thesis is to investigate the potential of SSP in eliciting individuals' biological age. A robust correlation of SSP with epigenetic clocks would suggest that individuals possess awareness of their biological age affected by life circumstances when evalu-

ating survival chances, and establish SSP as a more economical proxy for insights provided by epigenetic clocks.

This brings to the main objective of this thesis, which is investigating and providing insights into the information content and predictive validity of SSP using mortality and epigenetic data. The derived insights are important for several reasons. First, they are informative in designing better annuity plans and prices that ensure financial security throughout individuals' retirement. Second, they guide policymakers in shaping retirement, social security, and healthcare policies based on more realistic perceptions of longevity. By knowing how individuals perceive their longevity, policies can be more effective, particularly given the context of aging population. Finally, examining risk perceptions identifies differences in how various groups think about longevity, informing education to reduce biases and encourage better retirement planning decisions.

The rest of this thesis is organized as follows: Section 2 reviews the existing literature on the topic. Sections 3 and 4 describe the data used in our analyses and outline the methodology we use to test our hypotheses, respectively. In Section 5, we report our results and discuss their implications. Lastly, we present our concluding remarks in Section 6.

2 Literature Review

This review aims to consolidate existing research findings. First, we consider the dual perspectives on annuities: the theoretical prediction of their value as an optimal financial solution for the majority of retirees and the contrasting empirical evidence regarding their low adoption rate. Following this, we explore various rational and behavioral factors proposed to explain this discrepancy, focusing on one of the factors – the adverse selection. As the driving force behind the adverse selection, we discuss asymmetric information and the role of subjective survival probabilities in creating it. Finally, we provide an overview of epigenetic clocks as predictors of mortality risk.

2.1 The Annuity Puzzle

The vast literature focused on identifying the value and the optimal level of annuitization starts with seminal paper of Yaari (1965), where he introduces uncertainty over one's longevity to the standard life cycle hypothesis. This uncertainty fundamentally alters the consumption-savings decision of individuals and emphasizes the related risks of premature depletion of one's resources and underconsumption signified by a positive net wealth at the end of life. Annuities play a pivotal role in this uncertainty framework by providing a guaranteed stream of income until the individual's death.

Yaari (1965) concludes that in complete annuity markets with actuarially fair prices, a risk-averse consumer without a bequest motive should fully annuitize their wealth as it ensures the maximum possible constant lifetime consumption. Later studies have demonstrated the optimality of positive annuitization even under conditions that are less strict than Yaari's. Davidoff et al. (2005) find that in complete markets settings, full annuitization still emerges as the optimal strategy, whereas in incomplete markets, the optimal consumption path substantially deviates from the income streams offered by annuities, thereby making partial annuitization optimal.

Poterba et al. (1996), Mitchell et al. (1999) conduct both empirical and theoretical investigations into the annuity markets, using the "wealth equivalence" method

grounded in utility maximization. Their findings suggest that even in scenarios where insurance loads exceed a quarter of the price, an individual, given plausible parameters for the utility function, would still experience a more favorable outcome through annuitization. The value of an annuity contract can be measured by the money's worth ratio (MWR), a ratio of the expected present value of the annuity payments to the initial premium paid for it. A MWR equal to one indicates a fair price for the annuity, while a ratio below one suggests that the annuity is actuarially overpriced.

A study by James and Song (2001) of annuity markets in a range of high and middle-income countries finds that, when discounting at the risk-free rate, the MWR for annuitants is greater than 95% in most countries and sometimes greater than 100%. The MWRs for the average population member are lower but still exceed 90% in most cases. Mitchell et al. (1999) estimate that the average annuity policy payouts in the US are valued at 80-85 cents per dollar of premium paid, which is lower than the 90-95 cents per dollar often cited in the literature. The authors offer a number of possible explanations for this discrepancy, including the possibility that the mortality rates of the population pool used by annuity companies are too low. Other studies, such as those by McCarthy and Mitchell (2004), Finkelstein and Poterba (2004), support the hypothesis of adverse selection in the annuity markets. Investigating the Canadian market, Milevsky and Wu Shao (2010) estimate the MWRs to be around 100% and conclude that it is "a fairly good deal for the annuity purchaser".

Despite the demonstrated value of annuities and significant increase in life expectancies, empirical studies on the actual uptake of annuities demonstrate that only around a small 10% of the population owns annuities (James and Song, 2001; Rusconi, 2008; Milevsky and Wu Shao, 2010; Boyer et al., 2020).

2.2 Explanations of the Annuity Puzzle

Existing literature provides a range of explanations that can be broadly categorized as rational and behavioral factors. Benartzi et al. (2011) propose both rational and

behavioral explanations to the puzzle, offering a holistic perspective on why retirees might be reluctant to annuitize their savings, despite the apparent financial benefits. On the behavioral dimension, the authors discuss experimental evidence and studies that support the influence of biases such as loss aversion, framing effects, and mental accounting on annuitization decisions. Added to these biases, rational considerations such as bequest motives, concerns about inflation, and liquidity preference to cover medical expenses offer a more comprehensive view of the factors influencing retirees' decisions regarding annuitization.

Gong and Webb (2008) discuss mortality heterogeneity as an explanation. Analyzing the variances in life expectancy across diverse socioeconomic strata in the US, they unveil substantial mortality heterogeneity characterized by individuals from lower socioeconomic backgrounds largely experiencing lower life expectancy. In the context of distributional equality, the authors argue that annuitization might not be as optimal for this portion of the population as compared to their counterparts from higher socioeconomic echelons, who generally enjoy longer life spans. This is because the former are more likely to be advantageously selected into the annuity pool and, consequently, risk receiving diminished lifetime benefits.

Individuals' valuation of annuities is heterogeneous as well. Brown et al. (2017) investigate the factors driving this heterogeneity through randomized experimental studies and find that, depending on existing differences in cognitive constraints, consumers face varying levels of challenge when valuing annuities. This variation translates into heterogeneity of perceived annuity value. By controlling for the status quo, the authors eliminate the effects of endowment bias and liquidity constraints. They conclude that individuals are reluctant to enter into an annuity transaction if they have difficulty ascertaining its value. Such reluctance regarding difficult-to-value transactions generally serves as a protective behavior. Similarly, Brown and Finkelstein (2011) find that consumers seem to show limited interest in long-horizon insurance products like life annuities, as they experience specific challenges when making decisions about long-term, probabilistic outcomes.

Fairness of annuity prices is a major factor that determines the demand. The

pricing of annuities, and their MWR, is primarily influenced by two key factors: the prevailing term structure of interest rates and the probabilities of mortality (Dickson et al., 2009). Empirical analysis finds the price sensitivity to changes in rates to be "sluggish," which leaves the mortality rates to be the significant determinant of prices (Charupat et al., 2016). A high MWR can be an indicator of adverse selection, whereby the insurer prices annuities based on the mortality risk of a population that is higher than the mortality risk of annuitants actually selected into the pool. Individuals who expect to have a longer-than-average lifespan often find annuities more attractive because they expect to receive the annuity payments over a longer period of time, maximizing the benefits from their investment (Rothschild and Stiglitz, 1976). Because of its upward impact on the prices, the existence of adverse selection may make annuities not worthwhile for individuals who expect to live shorter lives.

2.3 Asymmetric Information and Adverse Selection

Adverse selection is a form of precontractual asymmetric information. In the context of annuity markets, it occurs because individuals with the highest longevity have the greatest incentive to purchase annuities. We discuss evidence of adverse selection in the annuity markets in depth because it leads to the main focus of this paper – the contribution of subjective survival probabilities to informational asymmetry in the annuity markets.

The annuity market provides an important field for studying asymmetric information because differentiating between the effects of adverse selection and moral hazard presents a considerable challenge, with each carrying distinct welfare and policy implications. In comparison to other insurance markets, moral hazard is presumed to be low in the annuity markets as annuities are unlikely to induce substantial efforts towards extending life. If annuities yield negligible moral hazard, conducting tests for asymmetric information in this market essentially constitutes testing for adverse selection (Finkelstein and Poterba, 2004).

Cohen and Siegelman (2010) examine the body of empirical work on adverse selection within insurance markets, including the annuity market. They state that the existing literature, based on analysis of variations in the mortality rates between annuitants and the broader population across different nations, points to a tendency of annuitants having longer lifespans than non-annuitants, thus suggesting the presence of asymmetric information favoring annuitants.

The tendency of annuitants living longer lives may also be explained by variation in risk preference – people who have less risky behaviors live longer and are more likely to buy annuities. Cutler et al. (2008) propose that the effect of risk preference on risk occurrence may provide a potential unifying explanation for the observed heterogeneity of selection and demand across insurance markets. In an attempt to understand the effects of risk preference on demand for insurance, Einav et al. (2007) find that risk preference varies significantly across markets, and that it is an equally important determinant of annuity demand as the variation in risk itself. Their evidence reinforces the standard hypothesis about the effect of asymmetric information.

Another evidence of the presence of adverse selection in annuity markets comes from Finkelstein and Poterba (2004), who examine the structure of annuity contracts. They find that those anticipating living longer exhibit a greater tendency to opt for "back-loaded" policies and are less likely to select annuities featuring guaranteed survivor benefits. This observation indicates that annuitants' choice of contracts is influenced by their often objective expectation to outlive the horizons estimated by insurers.

With regards to testing for asymmetric information, the positive correlation test as described in Chiappori and Salanie (2000) is the most commonly used test. The test rejects the null hypothesis of symmetric information if a significant positive correlation exists between annuity demand and mortality risk. One of the limitations of the test, as discussed by Finkelstein and McGarry (2006) and Chiappori et al. (2006), lies in its inability to hold in the presence of unobserved heterogeneity in risk preferences, i.e. if individuals have private information about characteristics

other than risk type, such as risk aversion, and these characteristics affect insurance demand. Finkelstein and Poterba (2014) attempt to overcome this limitation by introducing a concept of "unused observables," which are individual characteristics that are correlated with both the risk and the demand for insurance, yet not used by the insurers when designing contract menus. The study discusses annuitants' place of residence being an "unused observable" until the insurance companies started using postal codes as a pricing variable. These insights suggest that this kind of selection could be happening in different ways across various insurance markets and that understanding how insurance companies choose what information to use in pricing is important for addressing the adverse selection and low demand for annuities.

Why is addressing adverse selection important? The presence of adverse selection in annuity markets diminishes welfare (Einav et al., 2010; Einav et al., 2007; Einav and Finkelstein, 2011). Einav et al. (2007) find that, within the UK annuity market, asymmetric information at the guarantee margin results in a welfare reduction of approximately 2 percent of annual premiums, compared to an optimal, symmetric information benchmark. Their evidence also suggests that government mandates, which are the conventional remedy for adverse selection issues, do not necessarily improve the asymmetric information equilibrium, implying that achieving welfare improvements through compulsory social insurance might be more challenging in reality than a simple theory suggests. Moreover, adverse selection may potentially cause a "death spiral," a situation where an adverse selection cycle leads to rapidly increasing premiums and decreasing numbers of insured individuals, eventually resulting in a market failure.

2.4 Subjective Survival Probabilities

Private information about mortality risk plays a significant role in determining the demand for annuities. In this context, subjective survival probabilities — individuals' own assessments of their likelihood of surviving to a certain age — can be seen as a consolidation of private information that individuals hold regarding their health,

lifestyle, family medical history, and other unobservable factors influencing mortality risk in a single number. Understanding subjective survival probabilities is crucial as they influence individuals' decisions concerning insurance, savings, retirement, and other aspects of financial planning. Depending on their accuracy in predicting mortality risk, they hold the potential to introduce asymmetric information into the annuity markets and determine the equilibrium outcome.

As noted by Paté-Cornell (1996), "... uncertainties in decision and risk analyses can be divided into two categories: those stemming from variability in known (or observable) populations and, therefore, represent randomness in samples (aleatory uncertainties), and those arising from a fundamental lack of knowledge about underlying phenomena (epistemic uncertainties or ambiguities)." Theoretically, subjective survival probabilities, as assessments of uncertain events, could be more accurate and nuanced than the probabilities calculated based on general population data, or, conversely, they might be less accurate if predominantly shaped by individual biases. In the existing literature on SSPs, there is greater evidence of predictive validity about mortality risk.

Hurd and McGarry (1995), Hurd et al. (1998), Hurd and McGarry (2002), pioneering a series of empirical analysis of SSPs, are based on the HRS. The respondents of the study, who were individuals aged 50 years and older, were asked to assess their chances of surviving to 75 or 85 years of age. The authors find that the SSPs in the HRS are predictive of actual mortality, even when controlling for factors such as health status, socioeconomic status, and lifestyle choices. They find that individuals with higher socioeconomic status tend to give higher probabilities of survival, while individuals who smoke or have poorer health status give lower probabilities. Moreover, they find that extreme probabilities about survival, such as zero or 100%, also significantly correlate with actual mortality. These findings lead the authors to conclude that individuals' self-assessments of their chances of survival are generally accurate and consistent with other factors that influence mortality, making the SSPs a great potential tool that can be used in models of intertemporal decision-making under uncertainty. This predictive validity stems from individuals generally drawing

accurate inferences from their health status, survival experiences of acquaintances, and other factors. These insights bear significant implications for both policymakers, who can leverage this information to design strategies promoting healthy aging and longevity, and researchers, who can design improved studies on aging and mortality.

Building upon these foundational works that have significantly advanced our understanding of SSPs and their correlation with mortality, our study further extends this research. First, in our analysis we use the latest HRS data, spanning 15 waves from 1992 to 2020. This dataset includes a substantially higher rate of observed deaths among respondents (51.5%) compared to the study conducted by Hurd and McGarry (2002), which was based on the first two waves only (1.6%). The greater incidence of death provides robustness to our findings and extends the validation of SSPs' predictive validity over a longer term. Additionally, our study differs in its methodological approach by presenting results across five different specifications with progressively added covariates, allowing us to observe the dynamics of mortality predictors. Lastly, we employ dummy variables for different age, income, and wealth subgroups, rather than using continuous values, to capture the potential differential impacts of these subgroups on mortality and the predictive power of SSPs. These data and methodological extensions aim to contribute to the existing literature by providing additional robustness of the results and new insights.

Bissonnette et al. (2017) compare subjective mortality expectations with objective ones and examine the impact of the discrepancies on savings and consumption decisions. Using HRS data over a 16-year period, they develop an econometric model to compare subjective and objective mortality hazards, taking into account the rounding effect in SSPs. Their study finds that individuals, especially certain subgroups such as smokers, black individuals, more educated respondents, and younger cohorts, tend to be slightly optimistic regarding their survival prospects. This optimism or pessimism in survival expectations influences individuals' consumption paths, with pessimistic individuals consuming wealth more quickly. The paper also demonstrates that misperceptions in survival probabilities could result in significant welfare losses and affect decisions related to annuitization. These findings underscore

the importance of accurate mortality expectations in economic models and addressing discrepancies in survival perceptions.

Hamermesh (1985) study, based on 500 white, married couples aged 62-69, finds that SSPs exhibit greater variation than actuarial probabilities. He finds that individuals show awareness of current life table improvements, yet they don't necessarily extrapolate this information when determining their subjective life expectancies. Instead, individuals largely base their longevity projections on idiosyncratic factors such as their relatives' longevity (Hamermesh and Hamermesh, 1983).

A similar conclusion is reached by Baji and Bíró (2018), who explore the trajectories of subjective survival probabilities following various health shocks. Using HRS longitudinal data, they observe that individuals update their probabilities after health shocks and that the reaction to shocks varies depending on factors such as shock severity, availability of treatments, and individual coping mechanisms. Specifically, they find that individuals diagnosed with cancer exhibit a return to pre-diagnosis health measures and survival probabilities, indicating adaptation to the health shock. In contrast, those experiencing a stroke or heart attack show persistent effects on survival probabilities and incomplete recovery in self-reported health measures, suggesting long-term health consequences.

Ai et al. (2017) extend their research question beyond the interaction of survival probabilities and health state transitions, examining the impact of the two on retirees' optimal annuitization decisions. They introduce concepts of morbidity expansion and contraction, which are defined as extended lifespan spent in more or less healthy states, respectively. They find different outcomes for wealthier and poorer slices of the population. When faced with morbidity contraction, wealthier retirees tend to increase their demand for annuities, and, conversely, decrease the demand when morbidity is expanding. In contrast, poorer retirees are less responsive to health changes, which can be partially explained by the general low propensity to annuitize due to the consumption floor provided by governmental subsidies. This study, therefore, underscores the impact of SSPs, updated on health shocks, on preventative decisions such as annuitization in the face of change in longevity perspectives.

The insights of the research presented in this section mark a substantial contribution to the understanding of subjective survival probabilities and hold significant implications for refining behavioral models and developing more effective public policies.

2.5 The Epigenetic Clocks

Epigenetic clocks serve as a biomarker of aging based on DNA methylation process (DNAmAge), which results from both intrinsic processes, governed by individuals' genetic profile, and extrinsic factors, such as environmental influences. It has emerged as one of the most promising measures to estimate biological age, demonstrating robust evidence of predictive validity. As individuals age, DNA methylation patterns undergo dynamic changes, with alterations in certain sites linearly correlated with chronological age, forming the foundation for the development of epigenetic clocks (Duan et al., 2022). Originally developed to estimate chronological age, epigenetic clocks have evolved to reveal the disparities between chronological and epigenetic age. These disparities are indicative of accelerated aging or delayed aging, thereby offering a useful insight into an individual's health status and susceptibility to age-related diseases (Horvath, 2013).

Due to the unavailability of data until recently, research into the correlation between SSPs and epigenetic clocks is practically absent. Consequently, this section of the literature review is dedicated to providing an overview of epigenetic clocks, with a particular emphasis on those that have demonstrated a strong correlation with mortality risk. Ultimately, our research interest lies in determining whether there is a correlation between SSPs and epigenetic clocks. Establishing such a correlation would suggest that individuals possess an awareness of their pace of biological aging, which is a better predictor of mortality than chronological age, and the proximity of their mortality. This would also serve as a robust validation of the predictive validity of SSPs. The application of such findings to the analysis of annuity markets could prove instrumental, offering a valuable tool for designing more efficient annuity

contracts, as well as public policy.

Horvath's epigenetic clock is a prominent multi-tissue age estimator, well-supported by research in areas like cancer, Alzheimer's disease, and aging. It uses DNA methylation data from a wide range of samples and demonstrates a strong correlation of 0.96 between with chronological age, and a median absolute difference (MAD) of only 3.6 years. This clock is notable for its versatility and accuracy across various tissues, although it has limitations with cultured cells (Horvath, 2013).

In contrast, Hannum's epigenetic clock is a single-tissue estimator, developed using methylation data from adult whole blood samples. It shows a correlation of 0.96 with chronological age, and a MAD of 3.9 years. While it offers high accuracy for adult blood samples, it is less applicable to non-blood tissues and children, indicating a more specialized focus compared to Horvath's comprehensive model (Hannum et al., 2013).

Levine et al. (2018) introduced the PhenoAge clock, which combines clinical measures with DNA methylation data, providing a broader perspective on biological aging. This clock exhibits a robust correlation with numerous morbidity and mortality indicators, surpassing other models in mortality prediction. GrimAge clock is developed by Lu et al. (2019) and uses methylation levels at specific DNA sites linked to aging-related characteristics such as smoking, body mass index, and disease history. Demonstrating exceptional precision in estimating time to death, predictive capabilities of GrimAge outperforms its predecessors. GrimAge clock's potential as an effective instrument for evaluating individual health and forecasting mortality risk holds substantial implications for academia, clinical settings, and public health.

The utility of epigenetic clocks in predicting mortality risk highlights their potential role in addressing the issues related to the annuity puzzle. If a correlation between SSPs and epigenetic clocks is established, it would validate the predictive accuracy of SSPs and contribute to the understanding of the role of subjective survival probabilities in the annuity markets. By providing a more accurate measure of individual mortality risk, epigenetic clocks could help in designing more efficient annuity contracts and improving the welfare of retirees.

This literature review has explored the theoretical and empirical perspectives on annuities, the various factors influencing annuitization decisions, and the role of adverse selection in the annuity markets. The role of subjective survival probabilities and the potential application of epigenetic clocks in predicting mortality risk have also been highlighted, laying the foundation for further exploration of their correlation and implications for the annuity markets.

3 Data

3.1 The HRS Dataset

The Health and Retirement Study (HRS) is a national panel survey of older individuals and their spouses in the USA conducted by the Institute for Social Research (ISR) at the University of Michigan. The survey spans 15 waves from 1992 to 2020 and includes respondents from seven distinct cohorts, each representative of the USA population. The data from each survey year are collected and maintained by the ISR. A user-friendly longitudinal version of the data is prepared by the RAND Corporation, which improves data quality by incorporating enhancements and imputations, addressing issues such as nonresponse and inconsistencies.

Our study is based on the most recent release of the longitudinal data, the RAND HRS Longitudinal File 2020 (V1) from March 2023, HRS Epigenetic clock data, released in November 2020 and consisting of values for 13 epigenetic clocks developed from the 2016 HRS Venous Blood Study (VBS) data, and HRS Tracker file, which facilitates the use of HRS data within and across waves. The data used in this study are all publicly available.

Figure 3.1.1 displays the distribution of HRS respondents by cohort and wave, outlining the demographic characteristics of each cohort.

1. Initial HRS cohort, born 1931 to 1941. This cohort was first interviewed in 1992 and subsequently every two years.

2. AHEAD cohort, born before 1924, initially a separate study (The Study of Assets and Health Dynamics Among the Oldest Old). This cohort was first interviewed in 1993 and subsequently in 1995, 1998, and subsequently every two years.

3. Children of Depression (CODA) cohort, born 1924 to 1930. This cohort was first interviewed in 1998 and subsequently every two years.

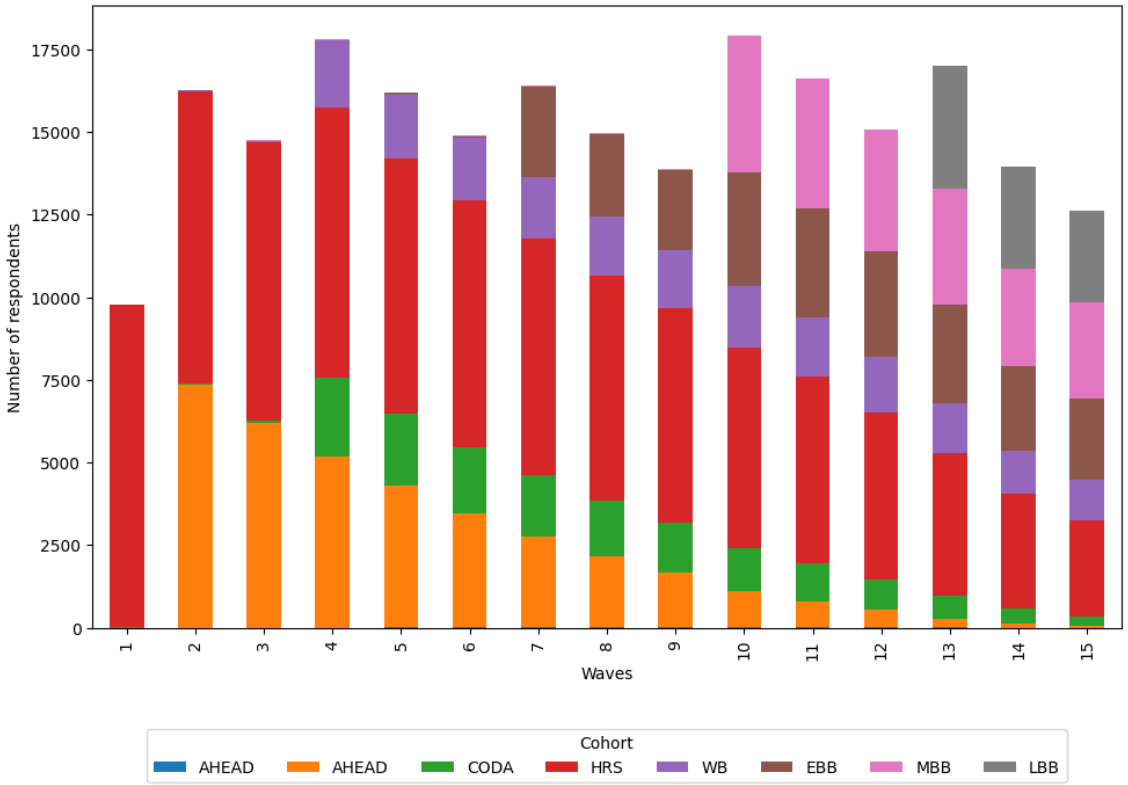
4. War Baby (WB) cohort, born 1942 to 1947. This cohort was also first interviewed in 1998 and subsequently every two years.

5. Early Baby Boomer (EBB) cohort, born 1948 to 1953. This cohort was first interviewed in 2004.

6. Mid Baby Boomer (MBB) cohort, born 1954 to 1959. This cohort was first interviewed in 2010.

7. Late Baby Boomer (LBB) cohort, born 1960 to 1965. This cohort was first interviewed in 2016.

Figure 3.1.1: The HRS Cohorts, by Wave



We use two distinct samples in our analyses of mortality risk and epigenetic clocks.

In the mortality risk analysis, we restrict our sample to the Initial HRS cohort. Rationale for this is twofold. First, this cohort has been tracked for the longest

duration within the HRS survey, providing us with a substantial sample size. Second, at the onset of the survey in 1992, these individuals were 50-61 years old, enabling us to record a higher number of observed deaths as the survey progressed, and consequently improving the statistical power of our analysis.

On the other hand, in the second part of our analysis, the examination of correlation between SSP and epigenetic clocks, we do not restrict our sample to a specific cohort, but instead use the HRS data on individuals from all 7 cohorts. However, limitation comes from the DNA data, which is available only for a representative subsample of the overall HRS sample.

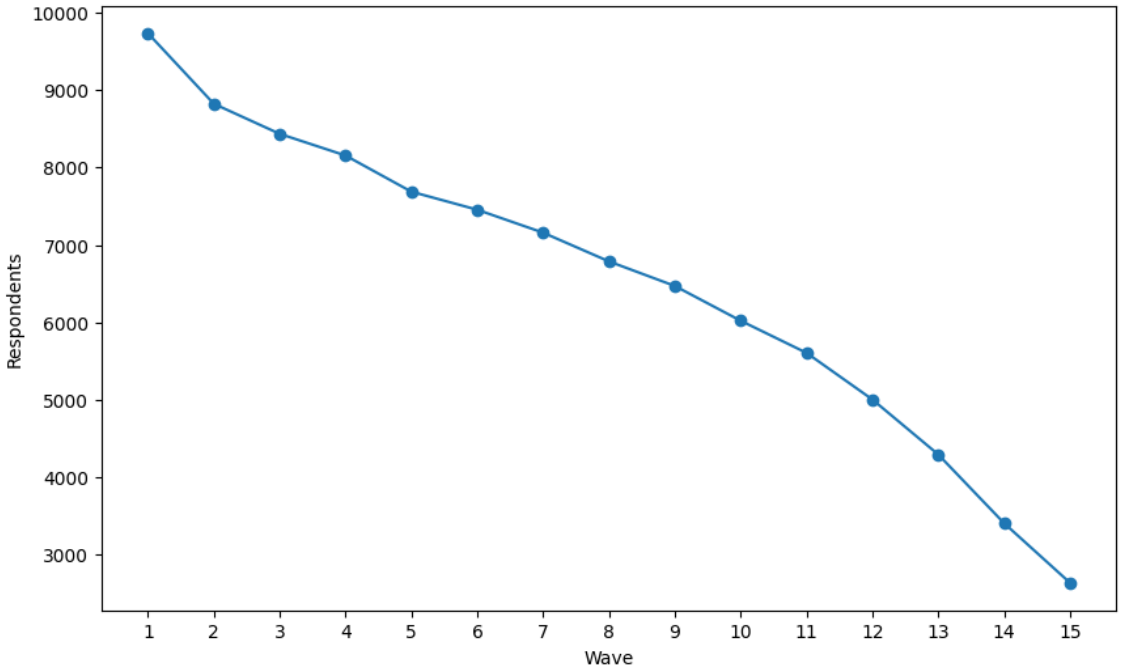
3.2 Data for Mortality Risk Analysis

We construct our data in this section in a long form so that our the number of observations is constructed by the number of individuals observed in each year of the study[1]. Figure 3.2.2 illustrates the shrinking sample size with the progress in waves. In wave 1, the sample consists of 9,734 individuals. Of these, 2,639 remain by the end of the wave 15. while the remainder exited the study prematurely either because they have deceased or stopped responding. Roughly half of the whole sample, 4,719 individuals (or 51.5%), is confirmed to have deceased. The HRS confirms deaths via an exit interview, where a close relative reports the death. Additionally, the HRS cross-references its respondents with the National Death Index, a centralized death record database maintained by the National Center for Health Statistics, providing an additional verification method for deaths. We assume that the remaining 2,376 individuals (or 24.4%) who have prematurely dropped out from the study were not deceased during the period of our observation, as the HRS does not report them as being dead.

The basic descriptive statistics on socio-demographic characteristics of our sample is presented in Table 3.2.1, and the breakdown of these statistics by gender. Our

[1]In order to better interpret the results of our regression, to get the conditional mortality risk for any given year, rather than wave, we fill the data for years in between the waves with the information from the previous wave.

Figure 3.2.1: Evolution of the Sample Size, by Wave



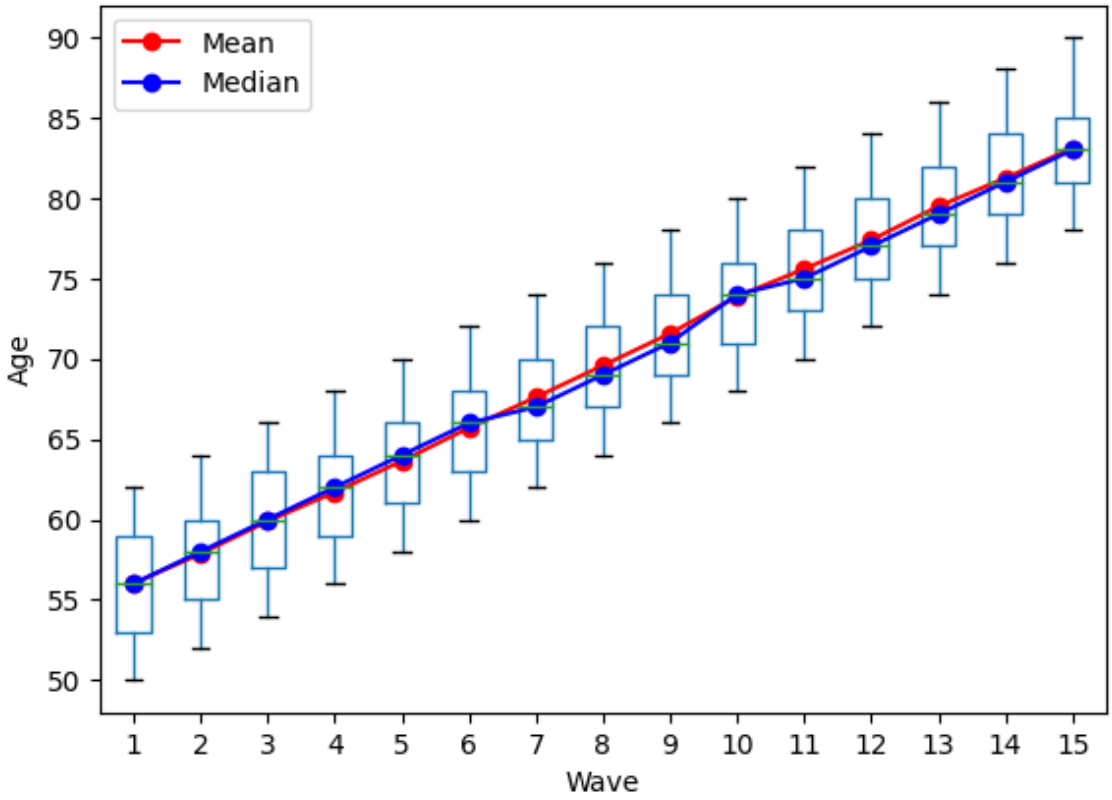
sample is predominantly female, 52.24%. An average individual in our sample has at least high school education, report household income of approximately $55,000 US dollars, and has net wealth close to $364,000 US dollars. The mean income and wealth indicators are inflated by outliers in the sample, making the median values more relevant. The median individual has a household income and net wealth of $35,150 and $133,600, respectively.

The evolution of the age distribution of our sample is presented in Figure 3.2.3. Mean and median overlap almost perfectly throughout all waves, while the right tail becomes longer with the progression of the years, indicating on skewness caused by fewer individuals surviving to older ages.

Our sample suggests that, on average, mothers tend to outlive fathers, with average life expectancies being 77 and 71 years, respectively. This aligns with general life expectancy trends in life tables. The median mother lived to at least 75 years but did not survive to 85, whereas the median father did not survive to either age.

Figure 3.2.3 illustrates the distribution of actual and perceived mortality outcomes

Figure 3.2.2: Evolution of the Sample Age, by Wave



by gender and race. Women and white respondents, on average, assign higher probabilities to their survival chances, which might reflect the actual life expectancies of the people surrounding them.

Compared to women, men are more likely to give a 0% response for both SSP 75 and SSP 85, while women are more likely to give a 100% response. Compared to white individuals, non-white individuals are more likely to respond with extreme probabilities for both probabilities to survive to 75 and 85. White individuals and women are more likely to give 50% response.

When comparing actual survival rates with predictions of surviving to at least 75 and 85 years, we observe notable differences. The mean values of SSP 75 and SSP 85 are only slightly lower their respective median values, suggesting a slightly left-skewed distributions. The mean SSP 75 exceeds the mean SSP 85 by nearly 20%, while the actual difference between the proportions of individuals who lived to at

Table 3.2.1: Mortality Risk Analysis Sample:
Descriptive Statistics

	Mean
Male	47.76%
Black	17.56%
Hispanic	9.53%
Education	
Less than high-school	32%
High-school	32%
College	36%
Household income	$54,742
bottom 10%	*$9,900*
Median	*$35,151*
top 10%	*$105,000*
Household wealth	$364,188
bottom 10%	*$1,000*
Median	*$133,600*
top 10%	*$789,595*
Mother's longevity	76.76
Mother survived to 75	*64.08%*
Mother survived to 85	*35.01%*
Father's longevity	71.41
Father survived to 75	*46.12%*
Father survived to 85	*19.44%*
SSP 75	58.79
1st quartile	*40.00%*
Median	*60.00%*
3rd quartile	*90.00%*
SSP 85	39.77
1st quartile	*0.00%*
Median	*40.00%*
3rd quartile	*70.00%*
Actual survival	75.41
Survived to 75	*61.48%*
Survived to 85	*13.41%*

least 75 and those who reached at least 85 is 48%, suggesting an optimistic bias in predictions.

SSP 85 exhibits a wider range than that of a SSP 75, indicating a considerable variation in individuals' perceived risks over a longer-term outcome. Overall, SSP 75 aligns fairly well with actual outcomes, but SSP 85 tends to overestimate the true

survival rates.

3.3 Data for Epigenetic Clock Analysis

In 2016, the HRS carried out the Venous Blood Study (VBS). This study gathered DNA methylation data from a sample of 4,018 individuals. The sample is selected non-randomly, but is representative of the broader HRS sample. 3,329 of these individuals (or 83%) were matched with the HRS sample that includes individuals from all cohorts of age at least 50 when entering the study. The not matched sample might be explained by the presence of younger individuals in the VBS sample.

Table 3.2.2 presents the descriptive statistics of the merged sample (further referred as the epigenetic clocks sample).

The DNA data from VBS was used to construct 13 epigenetic clocks using machine learning algorithms which were initially designed by epidemiology and genetics researchers[2]. The HRS staff constructed the clocks with the guidance of some of the researchers who originally developed the clocks, including M.Levine, S.Horvath, K.Sugden, to ensure data reliability.

Each epigenetic clock is designed with a specific application in mind. Some aim to predict chronological age, while others are calibrated to predict specific health outcomes, lifespan, or disease risk. The objective of each clock dictates the training data on which the algorithms focus and the scale of the clock[3]. The scale of a clock may also be influenced by the methodological approaches, statistical techniques, and reference populations used during their development.

As discussed previously, we narrow our focus to five epigenetic clocks: GrimAge, Hannum, Levine, Horvath, and Horvath Skin, all of which have shown a strong correlation with mortality risk. While the prediction objectives of these clocks differ

[2]Details discussed in the Chapter 2. Literature Review.

[3]Every clock is derived from a distinct set of CpG sites, specific locations in the DNA that are prone to methylation under specific circumstances. The selection of these sites to train a model is based on the objective of the study, its correlation with age or age-related outcomes. Given that different clocks might prioritize particular CpG sites or genomic regions, this can lead to variations in scale when the algorithms are applied.

Table 3.3.2: The Epigenetic Clock Analysis Sample: Descriptive Statistics

	Mean
Age	70.60
Male	46%
Black	18%
Hispanic	14%
Education	
Less than high-school	*0.23%*
High-school	*0.27%*
College	*0.50%*
Household income	68,440
bottom 10%	*$10,800*
Median	*$40,200*
top 10%	*$143,846*
Household wealth	$393,721
bottom 10%	*$0*
Median	*$140,000*
top 10%	*$919,200*
Mother's longevity	78.43
Mother survived to 75	*68.64%*
Mother survived to 85	*40.73%*
Father's longevity	72.83
Father survived to 75	*50.32%*
Father survived to 85	*23.55%*
SSP 75	49.05
1st quartile	*15%*
Median	*50%*
3rd quartile	*80%*
SSP 85	43.24
1st quartile	*10%*
Median	*50%*
3rd quartile	*75%*
Actual survival	74.13
Survived to 75	*45.87%*
Survived to 85	*15.35%*

in some instances, they are all measured in age years[4]. For our regression analysis, we standardize each clock by subtracting its mean and dividing by the standard deviation. This standardization enables us to interpret the clock estimates in terms of standard deviations and facilitates comparisons across the various clocks.

Table 3.2.1 summarizes the objectives and scale of the epigenetic clocks in our analysis, while Table 3.2.2 presents the descriptive statistics. Figure 3.2.1 presents their correlation heatmap.

Table 3.3.3: Epigenetic Clocks: Summary

	Year developed	Prediction objective	Scale
GrimAge	2019	Time to death, age-related conditions	Age
Hannum	2013	Chronological age	Age
Levine	2018	Phenotypic age, risk of morbidity	Age
Horvath	2013	Chronological age, tissues	Age
Horvath Skin	2018	Skin and blood age	Age

GrimAge, specifically designed to provide insights into the aging process, has a mean of 68.75 years with a standard deviation of 8.41 years [5]. Its range spans from 49.03 to 99.61 years, a spectrum of biological ages closer to the chronological ages of the sample. The Hannum clock, another predictor of biological age, has a lower mean of 55.12 years and exhibits the most extensive range, from 25.06 to 107.79

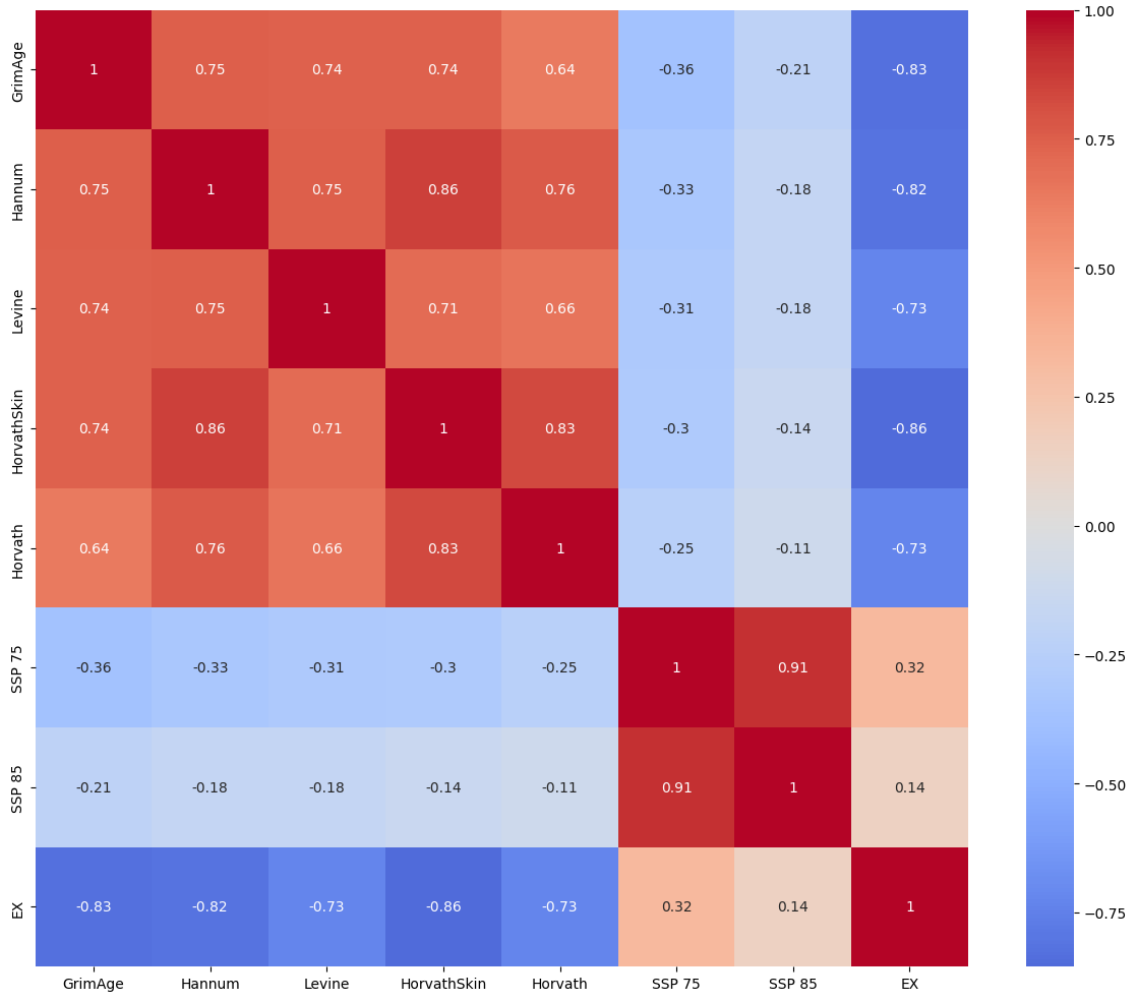
[4]Other clocks available in the HRS dataset, including Yang, Zhang, Bocklandt, Garagnani, and DunedinPoAm38, use different scales, such as risk scores.

[5]GrimAge is an epigenetic clock designed to predict time-to-death, however, its value is interpreted as the "biological age" of an individual in the context of how many years they appear to have left, given average life expectancies and health outcomes

Table 3.3.4: Epigenetic Clocks: Descriptive Statistics

	Size	Mean	Std. dev.	Min	25%	50%	75%	Max
GrimAge	3,329.0	68.75	8.41	49.03	62.14	68.49	74.84	99.61
Hannum	3,329.0	55.12	9.01	25.06	48.47	54.37	61.10	107.79
Levine	3,329.0	57.97	9.92	28.66	51.00	57.43	64.23	101.68
Horvath	3,329.0	66.20	9.34	23.31	59.66	65.80	72.37	106.47
Horvath Skin	3,329.0	70.23	8.58	36.97	63.58	69.71	76.57	101.29
Actual age	3,329.0	70.60	9.40	52.00	62.00	69.00	78.00	100.00

Figure 3.3.1: Correlation Heatmap



years, hinting at a diverse age distribution of the respondents. Levine's clock, which presents a more holistic perspective on aging by capturing phenotypic age, has a mean of 57.97 years. Its range, from 28.66 to 101.68 years, is comparable to that of Hannum's, suggesting similar variability in the cohort's biological aging. The Horvath clock, renowned for its versatility across various tissues, has a mean age of 66.20 years, closely paralleling GrimAge. Its distribution ranged from 23.31 to 106.47 years. Lastly, the Horvath Skin clock, designed specifically for skin and blood tissues, has the highest mean age of 70.23 years, with its range extending from 36.97 to 101.29 years.

The observed variations in the means and distributions across these clocks

highlight the complex nature of epigenetic aging, as well as the distinct ability of each clock in capturing different aspects of aging process.

3.4 Construction of Variables

In our mortality risk analysis the outcome variable is a binary variable representing mortality status. It equals 1 for a respondent who dies in year t, given that they survived the previous year, t-1.

Our primary explanatory variable in the mortality risk analysis and the dependent variable in the epigenetic clock analysis is the self-reported probabilities to survive to a certain age which capture their perception of mortality risk. In the HRS dataset these variables are – the probabilities of surviving to age 75 and 85 given during the first interview round, which we denoted as SSP 75 and SSP 85, respectively. These variables were derived from asking the HRS respondents the following question: "What is the percent chance, with 0 meaning absolutely no chance and 100 meaning absolutely certain, that you will live to be at least 75 (85)?" The values of SSP 75 and SSP 85 are rescaled to be between 0 and 1.

For SSP 75 and SSP 85, there are missing values. We generate dummy variables to account for these missing values, and impute the original missing values in SSP 75 and SSP 85 with zeros. We include these indicator variables for missing values as controls together with the SSP variables in our regression model. As Hurd and McGarry (2002) discuss, these missing values are often non-random. Individuals might choose to withhold a survival probability if they perceive a higher mortality risk, even if they don't explicitly express this.

Reported probabilities fluctuate based on the known risk factors like smoking habit and are aligned with the risks associated with certain socio-demographic groups. For instance, subjective survival probabilities generally increase with income, wealth, and education. They tend to be lower for male and non-white individuals.

In a similar way, we construct dummy variables for focal responses, which equal one if an SSP value is 0%, 50%, or 100%. These responses could arise from genuine

uncertainty or confidence, cognitive simplicity, or the inherent challenge of predicting one's mortality. As presented in Figure 3.2.2, there's a significant clustering of responses at these focal points. The reasons for selecting extreme responses (0% or 100%) might differ from those for choosing 50%. Extreme values might stem from an individual's strong perception based on private knowledge or overly pessimistic or optimistic (biased) perception. On the other hand, a 50% response might indicate genuine belief in an even chance of survival to age 75 or 85, or a difficulty of assigning a probability to a remote uncertain event (an epistemic uncertainty).

An individual's self-reported health can greatly influence their subjective survival probabilities. In the HRS this variable is constructed by asking the respondents to self-rate their general health status from "1" (Excellent) to "5" (Poor). We hold the rating "Poor" as a reference group and create dummy variables for other ratings. Self-reported health is likely pivotal in forming subjective survival probabilities, and existing research emphasizes its connection to both stated probabilities and actual mortality outcomes.

The HRS collects extensive data on health, cognition, economic status, and family relations. We incorporate some of these variables as controls in our analysis. All of the health variables are constructed as dummy variables that equal 1 if a doctor has diagnosed the respondent with a specific condition, like cancer or diabetes.

Obesity and overweight dummies are constructed using reported and, where available, physically measured BMI data, and the Centre for Disease Control's BMI interpretation guide for adults[6]. An individual is obese if their BMI equals or greater than 30, while they are overweight if their BMI is between 25-29.

Individuals' educational attainment and economic status is also relevant in the context of subjective survival probabilities and mortality risk. The expectation is that, other things being equal, individuals with more education and higher wealth and income may be able to make better health related decisions, as well as have better access to healthcare, affecting both individuals' perceived and actual survival outcomes.

[6]https://www.cdc.gov/healthyweight/assessing/bmi/adult$_b$*mi*/*index.html*

We construct two dummy variables for individuals with at least high school and at lest college education, holding the third group of individuals with less than high school education as a reference group. Parents education is controlled by a dummy variable that equals one if either parent has at least 12 years education.

Income and wealth factors are accounted by dummies for individual's household income and net welath depending on the quartile they fall into. We hold individuals in the bottom quartile of income and wealth as the reference group.

The value of household income is the sum of earnings, Social Security, pensions, annuity, disability and other government incomes. The household wealth value is the sum of the values of individuals' primary residence, other real estate owned, vehicles, businesses, IRA's stocks, chequing and saving accounts less any debt and mortgage owed.

Individuals' parents' longevity is accounted for using separate dummy variables for mother and father. They equal one if wither parents' age of death, if dead, or age in the last wave, if alive, is equal or greater than 75.

4 Methodology

Our study is structured around two distinct analyses. In the first part, we use a logistic regression to assess whether subjective survival probabilities can accurately predict mortality status. We test this model using five specifications of explanatory variables. In the second part, we use an linear regression (OLS) to test the correlation of subjective survival probabilities with epigenetic clocks. Similarly, this model is tested in five covariate specification.

4.1 Mortality Risk Analysis

To understand the link between how well the individuals' perception about their survival chances and the actual mortality outcomes, we use a logistic regression, as shown in equation (4.1.1), as our econometric model. This model is chosen for several reasons. Firstly, our dependent variable, the mortality status $d_{i,t}$, is binary (either dead or alive). Logistic regression is designed to handle such binary outcomes, making it a natural fit. Secondly, the model allows us to account for various factors that might influence mortality, such as age, health status, lifestyle habits, etc., represented by the covariate matrix X_i. Econometrically speaking, logistic regression assumes that the log odds of the outcome is a linear combination of the predictors. This means that as we change the value of our predictors (like the subjective survival probability), the rate of increase or decrease in the odds of the outcome (being dead) is constant. Additionally, for our results to be reliable, we assume no perfect multicollinearity among our predictors, the absence of extreme outliers, and that our observations are independent of each other.

In our logistic regression model addressing attrition, individuals exiting due to attrition at a specific age are treated as having right-censored observations. This implies assigning them zero probabilities of experiencing the event, mortality, from their current age up to the attrition age, and subsequently, they are excluded from the logistic regression as they never transition to the event, i.e. they never turn 1. This treatment assumes that, given the set of covariates, attrition is uncorrelated

with mortality. This assumption is justified by the alignment of overall mortality probabilities by age with life-tables.

$$\log\left(\frac{p(d_{i,t}=1\,|\,d_{i,t-1}=0)}{1-p(d_{i,t}=1\,|\,d_{i,t-1}=0)}\right)=\beta_0+\rho SSP_i+X_i\beta \tag{4.1.1}$$

where:

- $d_{i,t}$ is individual i's mortality status in year t and $d_{i,t-1}$ is the mortality status in year $t-1$;
- SSP_i is individual i's subjective survival probability given during at the first survey year;
- β_0 is the intercept that interpreted as the log odds of being dead when all predictors are zero;
- ρ is the coefficient estimator for SSP_i, and represents the change in the log odds of being dead for a one-unit increase in the subjective survival probability, holding other variables constant;
- $X_{i,t}$ is matrix of covariates for individual i at time t (detailed specifications of the covariates provided later);
- β is the vector of coefficients associated with the covariates in $X_{i,t}$, and represent the change in the log odds of being dead for a one-unit increase in the respective covariate, holding other variables constant.

4.1.1 Marginal Effects

While the coefficients in a logistic regression, such as ρ), provide insights on the log odds scale, they might not be straightforward to interpret. Marginal effects, on the other hand, give us the change in the probability of the event occurring for a one-unit change in the predictor, holding other factors constant. In simpler terms, it tells us how much the chance of an event, eg. dying, changes for a small change in our variable of interest, an increase in subjective survival probability. Thus, using marginal effects allows us to directly interpret our results in a more intuitive manner.

The marginal effect of the predictor, SSP_i, in our logistic regression is represented by equation (4.1.2).

$$\frac{\partial p(d_{i,t} = 1 \mid d_{i,t-1} = 0)}{\partial SSP_i} = \rho \times p(d_{i,t} = 1 \mid d_{i,t-1} = 0) \times (1 - p(d_{i,t} = 1 \mid d_{i,t-1} = 0)) \tag{4.1.2}$$

4.1.2 Specifications

We apply the nesting approach to our model, gradually expanding the set of control variables. This approach allows us to observe how new sets of variables influence our main predictor, SSP_i, and explore the robustness and sensitivity of our findings. We estimate the model in seven distinct specifications presented below [7].

We start with a set of basic demographic characteristics and technical variables in Specification 1, and progressively incorporate more detailed information about the respondent's economic status, parents' longevity, lifestyle, objective and perceived health status. As discussed in previous chapters, previous works have established a significant correlation between the SSP variables and self-reported health. We introduce the self-reported health variable as the final step in Specification 5 in order to examine its potential confounding effect on the relationships between SSP and other variables in the model.

This methodology serves two primary purposes. First, it allows us to identify at which stage, if at all, our main findings become sensitive to the inclusion of additional controls. If our main results remain largely consistent across specifications, it may imply robustness of the results. Second, by including variables step-by-step, we can also explore the explanatory power of each group of variables.

Specification 1. We start by controlling for a basic set of socio-demographic variables, missing and focal SSP values:

[7] We provide explain construction of these variables in detail in Chapter 3. Data)

- dummies for being in late 50s (55-59) and for being older than 60 (reference group is being in early 50s (50-54))
- gender and race dummies (for male, black and hispanic)
- dummies for educational attainment (for having at least high-school and at least college education)
- a dummy for parents' years of education (for having more than 12 years of education (equivalent to high school level)
- dummies for missing SSP and SSP of values 0, 0.5 or 1.

Specification 2. We introduce dummies for parents' longevity and economic status:

- for having mother lived to or beyond 75 years old
- for having father lived to or beyond 75 years old
- for having income in 2nd, 3rd and top quartile
- for having wealth in 2nd, 3rd and top quartile.

Specification 3. We introduce lifestyle variables:

- a dummy for ever being a smoker
- a dummy for being obese and overweight.

Specification 4. We introduce a set of health variables:

- dummies for ever being diagnosed with cancer, diabetes, stroke, a heart disease, a lung disease, a psychological condition and having high-blood pressure.

Specification 5. We introduce the self-reported health variable:

- dummies for rating own heath as fair, good, very good and excellent.

4.2 Epigenetic Clock Analysis

In this part of our analysis we want to test the relationship between individuals' perceived survival chances and their epigenetic aging process that we assume to be as

shown by equation (4.2.1). We use a linear regression model illustrated in equation (4.2.2), a robust and straightforward method, to test this relationship. We test this model for each of the five selected epigenetic clocks individually.

The choice of this model is grounded in several rationales. First, our dependent variable, SSP, as well as our primary predictor, the epigenetic clock, are continuous in nature. Linear regression is well-suited to capture and quantify relationships between such variables. Second, the model provides a framework to control for multiple covariates that might have confounding effect. This is presented by the covariate matrix X_i, that includes variables from socio-demographic to health indicators, ensuring that the relationship of our interest is not spurious, but instead reflective of an underlying association.

From an econometric standpoint, linear regression presumes that the expected value of the dependent variable, SSP_i, is a linear combination of the predictors. This linear relationship implies that a unit change in the predictor, an epigenetic clock, will result in a consistent change in the expected value of the SSP, given other predictors remain constant.

$$SSP_i = \beta_0 + \theta EpigeneticClock_i + X_i\beta_1 + \epsilon_i \tag{4.2.1}$$

$$E(SSP_i|EpigeneticClock_i, X_i) = \beta_0 + \theta EpigeneticClock_i + X_i\beta_1 \tag{4.2.2}$$

Where:

- SSP_i is individual i's subjective probability to survive to at least 75 or 85,
- $EpigeneticClock_i$ is one of the five epigenetic clocks for individual i,
- X_i is a vector of other characteristics of individual i (X_i represents seven distinct specifications, which we described later),
- β_0 is the constant term,
- θ and β_1 are the sets of coefficient estimates for the epigenetic clock and the

set of covariates.

4.2.1 Specifications

We test the model in (4.2.2) in 5 different covariate specifications, which are the same as described in the 4.1 Mortality Risk Analysis section.

4.2.2 Sensitivity Analysis

As part of our sensitivity analysis, in addition to the model presented in equation (4.2.2), we test a model that does not include any of the five epigenetic clocks and that includes all five clocks in aggregate. We present these scenarios in equations (4.2.3) and (4.2.4), respectively, and refer to them as "baseline" and "collective" models (we refer to the model represented by equation (4.2.2) as an "individual" model).

$$E(SSP_i|X_i) = \beta_0 + X_i\beta_1 \tag{4.2.3}$$

$$E(SSP_i|EpigeneticClock_{j,i}, X_i) = \beta_0 + \sum_{j=1}^{5} \theta_j EpigeneticClock_{j,i} + X_i\beta_1 \tag{4.2.4}$$

Where:

- $EpigeneticClock_{j,i}$ is clock j (one of the five clocks in our analysis) for individual i,
- θ_j is the coefficient estimate for clock j,
- the other variables remain as previously defined in model (4.2.2).

By including the analysis of a model that excludes the epigenetic clocks is a we establishes a baseline, that allows us to observe the relationships among other covariates and SSP and determine potential confounders. Comparing the performance of models with and without the clocks can help us to establish contribution of the epigenetic clocks more accurately and ensure the robustness of our findings.

Including all epigenetic clocks in one model, in addition to analyzing each individually, allows for a simultaneous assessment of their unique contributions to the outcome while controlling for the effects of the others. This approach offers a perspective on the combined influence of the clocks and highlights the distinct significance of each[8].

Lastly, we standardize the epigenetic clock variables in our regression analyses, rescaling them to have a mean of zero and a standard deviation of one, in order to make our results directly comparable across the clocks and enable a more straightforward interpretation.

[8]This model can be extended further by using Lasso regression and evaluating out-of-sample mean squared error (MSE), thereby improving the predictive power of epigenetic clocks for subjective mortality risk.

5 Results and Discussion

5.1 Mortality Risk Analysis

The results of our logistic regression of mortality risk on subjective survival probabilities, as well as their marginal effects, are presented in Table 5.1.1 (full table of results is presented in Tables A1-A4 of the Appendix). The analysis in this section will be primarily focused on the discussion of the results on SSP 75 as the general behavior trends of SSP 85 and the covariates being similar with the difference being the smaller magnitude of SSP 85 effect.

Table 5.1.1: SSP 75 and SSP 85

	Dependent variable: Mortality Risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
SSP 75	-0.523***	-0.399***	-0.373***	-0.128*	0.010
	(0.072)	(0.072)	(0.072)	(0.072)	(0.074)
Marginal effect	-0.01***	-0.008***	-0.007***	-0.002*	0.0
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	195,263	195,263	193,799	193,799	193,799
SSP 85	-0.261***	-0.160**	-0.135*	-0.020	0.073
	(0.073)	(0.073)	(0.073)	(0.073)	(0.073)
Marginal effect	-0.005***	-0.003**	-0.003*	-0.0	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	195,624	195,624	194,160	194,160	194,160
Note:				*p<0.1; **p<0.05; ***p<0.01	

5.1.1 Subjective Survival Probabilities

In the presented table, the SSPs are on a continuous scale ranging from 0 to 1, representing the probability of surviving in percentage points. The mortality risk is a binary variable that takes values of 0 or 1 to indicate an individual's alive or dead status; the average mortality risk is in the table presented in percentage points.

The overall significance of the estimates of SSP 75 indicates that it is informative of mortality risk, with higher subjective survival probabilities linked to lower mortality

risk, even after controlling for a wide set of socio-demographic, economic, and health conditions. The interpretation of this finding could be that individuals' subjective expectations about their survival are informed by their personal, possibly private, health information and overall well-being, which might not be fully captured by the observed covariates in the study. This supports the idea that SSP is not merely an optimistic or pessimistic attitude but reflects some intrinsic knowledge about one's health that is predictive of actual survival chances.

The average mortality risk, that is, the average probability of death in the following year for individuals aged 51-61 years, given they were alive in the previous year, equals roughly 2%. In other words, on average, about 2 individuals out of 100, or 200 individuals in our initial sample of almost 10,000, die every year during the period from 1992 to 2020.

In our analysis, a consistent negative relationship between SSP and mortality risk is observed in all model specifications except the last. This indicates that individuals who estimate a higher likelihood of surviving to age 75 are, in reality, less likely to die in the subsequent year. The coefficient estimate for SSP varies: it is -0.523 in the simplest model specification (1) and changes to 0.010 in our most comprehensive model (5), which incorporates all covariates. This range in coefficient values reflects the influence of additional factors accounted for in different model specifications. Notably, the -0.523 coefficient in model (1) indicates a substantial decrease in the log odds of mortality, by 0.523, when SSP increases from 0 to 1[9].

A more intuitive way of interpreting this effect is by looking at the marginal effect of SSP on the log odds of dying. The marginal effect of SSP on the log odds of mortality risk ranges from -0.01 to 0.0 in specifications 1 to 5, respectively. This means that when SSP increases from 0 to 1, the log odds of mortality decrease by 1 percentage points. Given that the mean mortality risk in our sample is 0.0195 (or 1.95%), a reduction of 1 percentage points is quite substantial, representing nearly half of the mean mortality risk in our sample.

[9]An increase by 1 unit represents a shift from 0% to 100% in the perceived chance of surviving to at least 75

The predictive power of SSP was reported by previous works on mortality risk, including among others, Hurd and McGarry (2002), which we discuss in Chapter 2 of this thesis, Bisconti and Bergeman (1999) who indicate that individuals' SSP is associated with lower mortality risk over a 10-year period.

These results suggest that individuals might have private information about their own survival prospects that is not fully captured by observable measures. We will now discuss the effects of the covariates included in our model to gain insights not only about the factors influencing mortality risk directly, but also about how these factors change the explanatory power of SSP. Such sensitivity analysis may help us better understand the mechanisms through which individuals construct their SSP.

5.1.2 Self-Reported Health

The estimate for SSP remained significant even after controlling for most of the measurable objective health and behavior indicators that could potentially summarize one's physical and psychological state. However, it stops being a significant predictor as soon as indicators for individuals' self-reported health are added. To get a more accurate picture of the confounding effect of self-reported health on the power of SSP, we additionally tested a model in a specification where, unlike specification 5, we include self-reported health dummies without including the dummies for lifestyle and objective health measures. We refer to this specification as to specification (6) in Table 5.1.2, where we present the marginal effects of the variables.

Our observation is that the marginal effect of SSP on mortality risk is non-significant when controlling for self-reported health dummies, regardless of the additional inclusion of objective health and lifestyle variables.

Although on the surface this finding might appear obvious as both SSP and self-rated health are individuals' perceptions, the evidence is insightful in several respects.

First, it is possible that SSP and self-reported health contain largely overlapping information regarding mortality risk. When we include both variables in the same model, it appears that the unique contribution of SSP is overshadowed by the self-

reported health indicators, which could be a more immediate and explicit indicator of health status.

Second, self-reported health may act as a mediator between SSP and mortality risk, meaning that individuals' perceptions of their survival probabilities could be significantly influenced by their own assessments of their health. If self-reported health captures much of the variation in mortality risk, once it is controlled for, the direct effect of SSP on mortality risk is no longer distinguishable.

Lastly, it is likely that not only the self-reported health accurately captures information about an individual's objective health, but that it also contains information over and above what the objective health measures and SSP contains. This may include aspects such as mental well-being, lifestyle choices, and other day-to-day challenges, which are hard to measure and observe, and which individuals might not fully consider when constructing their SSP (Idler and Benyamini, 1997).

The marginal effect estimates for self-rated health indicators, when compared to the reference group of individuals who rated their health as very bad, show a negative association with mortality risk, indicating that better self-rated health is significantly associated with lower mortality risk. Improvement in individuals' perception of their health from very bad to fair, and further to excellent ratings, is associated with a reduction in the mortality risk by 0.2 percentage points (statistically significant at the 5%) and 1.2 percentage points (statistically significant at the 1%), respectively.

5.1.3 Missing and Focal Values

We account for the potential impact of missing SSP by including a dummy variable that equals one when an SSP value is not reported. Similar to the previous works on SSP using HRS data, our results suggest that missing SSP values are not random and, in fact, that they are significantly and positively with an increase in mortality risk.

This finding suggests that individuals who do not report an SSP may differ systematically from those who do, potentially indicating a form of self-selection that correlates with mortality risk. A potential explanation for this might be that those

Table 5.1.2: Self-Reported Health

	Dependent variable: Mortality Risk (mean=0.0195)		
	(4)	(6)	(5)
SSP 75	-0.002*	-0.0014	0.0
	(0.001)	(0.001)	(0.001)
Smoker	0.008***		0.008***
	(0.001)		(0.001)
Obesity	-0.001		-0.001
	(0.001)		(0.001)
Overweight	-0.002**		-0.002*
	(0.001)		(0.001)
Cancer	0.01***		0.009***
	(0.001)		(0.001)
Diabetes	0.013***		0.011***
	(0.001)		(0.001)
Stroke	0.009***		0.008***
	0.002)		(0.002)
Heart disease	0.006***		0.005***
	(0.001)		(0.001)
Lung disease	0.009***		0.008***
	(0.001)		(0.001)
High blood pres.	0.004***		0.003***
	(0.001)		(0.001)
Psych condition	0.005***		0.003***
	(0.001)		(0.001)
SRH: fair		-0.0058***	-0.002**
		(0.001)	(0.001)
SRH: good		-0.0124***	-0.006***
		(0.001)	(0.001)
SRH: very good		-0.0170***	-0.009***
		(0.001)	(0.001)
SRH: excellent		-0.0213***	-0.012***
		(0.001)	(0.002)
Observations	193,799	193,799	193,799
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5.1.3: Socio-Demographic Predictors

	Dependent variable: Mortality Risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
Age 55-59	-0.01***	-0.011***	-0.011***	-0.012***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age >60	0.008***	0.008***	0.008***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.005***	0.007***	0.005***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Black	0.003***	-0.001	-0.0	-0.0	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.009***	-0.011***	-0.01***	-0.008***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
College	-0.007***	-0.003***	-0.002**	-0.002**	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High school	-0.004***	-0.001*	-0.001	-0.0	0.0
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Parents' edu >12	-0.002**	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Note:					*p<0.1; **p<0.05; ***p<0.01

who do not provide an SSP estimate are aware of the private circumstances that reduce their survival chance.

Focal SSP values, which are those reported as 0, 0.5, or 1, also show a positive association with mortality risk. This result may indicate accuracy of the systematic heuristic thinking adopted by individuals who report focal probabilities.

The fact that both missing and focal SSP values are significant predictors of mortality risk indicates the complexity of how individuals perceive and report their survival chances. The missing SSP values suggest there might be unobserved heterogeneity among respondents, while the focal SSP values imply that overly simplified or rounded assessments of survival probability are informative of mortality risk.

5.1.4 Socio-Demographic Predictors

Socio-demographic variables such as age, race, and education level have been shown to influence mortality risk. Our analysis confirms these trends, with older age categories and certain racial demographics having higher mortality risks. The results for marginal effects of the variables is presented in Table 5.1.3.

Age is a fundamental determinant of mortality risk, which is demonstrated by the significance and persistence of its estimates. However, our results suggest that, depending on the grouping, age may have a non-linear relationship with the mortality risk. A potential explanation could be that individuals in these groups may belong to distinct risk profiles due to a variety of biological, social, economic and behavioral factors (eg. variation in the stress levels of individuals of pre- and post-retirement).

Being *male* increases the mortality risk of an individual in our sample by 0.5-0.7 percentage points. This observation aligns with the well-documented phenomenon in demographic and health studies, where male individuals have a shorter life expectancy compared to females.

Black individuals in the sample have a higher risk of mortality compared to non-black individuals. This finding is in line with existing literature that often shows disparities in health outcomes by race, with black individuals typically facing higher mortality rates. However, this association diminishes and becomes statistically non-significant once income and wealth are controlled for. This change indicates that the initial higher mortality risk associated with being black may be largely attributable to differences in income and wealth rather than race itself.

The consistent and strong negative association between being *hispanic* and mortality risk suggests that Hispanic individuals in the sample have a lower risk of mortality compared to the reference group of white individuals. This phenomenon, often referred to as the "Hispanic paradox," suggests that despite potential socioeconomic disadvantages, Hispanic populations tend to have lower mortality rates and better health outcomes in certain respects compared to other ethnic groups (Franzini et al., 2001). Various hypotheses have been proposed to explain this paradox, including

cultural factors, dietary habits, stronger social networks, and the "healthy migrant effect," where immigrants may be healthier than the average population of their home countries. The strength and persistence of the effect suggest that the lower mortality risk for Hispanics is not fully explained by the covariates included in the model.

The negative coefficients for *educational attainment* suggest that higher levels of education are associated with a reduced risk of mortality. The marginal effect sizes indicate that the reduction in mortality risk is stronger for individuals with at least a college education compared to those with at least a high school education. Education has a protective effect on mortality due to a variety of factors such as better health behaviors, greater access to healthcare, a higher likelihood of engaging in preventive health measures. Much of these factors are also linked to individuals' wealth and income, presence of which in our model shrinks the effect of education significantly. Similarly, the effect of parents' education is significant until the model controls for individuals' economic status.

5.1.5 Economic Predictors

Income and wealth influence the ability to access healthcare services, maintain a healthy lifestyle, live in safer environments, and manage stress — all of which can contribute to longevity. Results for income and wealth predictors are presented in Table 5.1.4.

We observe that income has a more persistent and stronger effect on mortality outcomes than wealth. This could be because, unlike income, individuals' wealth are more likely to be tied up in illiquid assets, in which case accessibility of health and preventative care reduces. Compared to being in the bottom quartile, being in the top quartile of income is associated with the reduction of mortality risk by 0.7 - 1.0 percentage points. These effects remain statistically significant at 1% across all model specifications, indicating a robust association between income and mortality. Wealth indicators, except for being in the top quartile (Q4) of wealth, stop having a significant association with mortality risk when health variables are introduced.

Table 5.1.4: Economic Predictors

	Dependent variable: Mortality Risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
Income in Q2		-0.005***	-0.005***	-0.004***	-0.003***
		(0.001)	(0.001)	(0.001)	(0.001)
Income in Q3		-0.007***	-0.007***	-0.006***	-0.005***
		(0.001)	(0.001)	(0.001)	(0.001)
Income in Q4		-0.01***	-0.01***	-0.009***	-0.007***
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q2		-0.003***	-0.002**	-0.001	0.0
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q3		-0.004***	-0.003***	-0.002	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q4		-0.006***	-0.004***	-0.002*	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Note:					*p<0.1; **p<0.05; ***p<0.01

5.1.6 Parents' Longevity

As we show in Table 5.1.5, having longer lived parents is significantly and persistently associated with a reduction in of one's mortality risk. The effect might be due to inherited genetic predispositions towards a longer life, or other confounding factors that might have helped the parents live longer in the first place, such as socioeconomic status or behavioral factors. For instance, if an individual's parents' longevity is partially attributed to their better healthcare access, lifestyle choices and habits, the individual is more likely to have been nurtured to have similar attributes. Controlling for parents' longevity in our model decreases the effect of SSP on mortality risk by 10%, from 1.0 to 0.9 percentage points.

5.1.7 Lifestyle and Objective Health Indicators

Smoking has a significant and persistent association with mortality risk, which have been confirmed by the vast body of medical literature. Conversely, the effect of being overweight or obese is not consistent in our model. The statistical significance of the effect, and even the sign, changes, which might be because substantial confounding

Table 5.1.5: Parents' Longevity

	Dependent variable: Mortality Risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
Mother lived >75		-0.003***	-0.003***	-0.002***	-0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
Father lived >75		-0.002***	-0.002***	-0.002***	-0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
Note:					*p<0.1; **p<0.05; ***p<0.01

effect of other variables in the model, in particular objective and subjective health measures.

Health conditions are significant predictors of mortality risk, among them being diagnosed with diabetes and cancer having the strongest positive effect, followed by having had a stroke and lung disease.

Together the lifestyle and objective health measures are responsible for a decrease in the explanatory power of SSP by 70% (from 0.7 to 0.2 percentage points). As a reference, adding self-reported health indicators alone, without the lifestyle variables, decreases the predictive power of SSP by 80% (from 0.7 to 0.14), suggesting it may have a stronger link with SSP.

5.2 Epigenetic Clock Analysis

We start this section by comparing the explanatory power of the models that don't and do include epigenetic clocks. Table 5.2.1 presents the results for adjusted R-squared for models (1) with no epigenetic clocks, represented by the equation (4.2.3), (2) with all five epigenetic clocks included collectively represented by the equation (4.2.4), and (3) with each clock included individually, represented by equation (4.2.2). All models are presented in 5 specification of covariates.

The number of variables included in our models varies across these three settings, depending on the inclusion of epigenetic clocks and whether they are included collectively or individually. We use adjusted R-squared to compare the explanatory

Table 5.2.1: Comparison of Adjusted R-Squared Results

	Model Specifications				
	(1)	(2)	(3)	(4)	(5)
(1) Without clocks	0.143	0.165	0.164	0.168	0.175
(2) With all clocks	0.209	0.222	0.226	0.237	0.248
(3) Individual clocks					
GrimAge	0.206	0.219	0.223	0.233	0.243
Hannum	0.189	0.202	0.208	0.216	0.227
Levine	0.184	0.197	0.203	0.209	0.218
Horvath Skin	0.179	0.193	0.199	0.208	0.219
Horvath	0.167	0.182	0.188	0.195	0.205
Observations	3,329	3,329	3,329	3,319	3,319

power of the models, as it provides a more accurate measure of fit by accounting for the number of variables included in the model.

The first setting, which excludes the clocks, serves as our baseline. In this model, the covariates account for 14.3% to 17.5% of the variation in SSP.

Including all epigenetic clocks in the model improves the adjusted R-squared in the first specification that controls only for socio-demographic characteristics by 6.6% (or 46% of of the original power). However, the increase is less pronounced, at 7.7% (or 42% of the original power), in the last specification that includes variables on individuals' objective and subjective health, lifestyle and parents' longevity. This suggests that the epigenetic clocks and health measures share similar informational content that explains the variation in SSP.

Interestingly, the incremental explanatory power provided by individual clocks is closely comparable to that of the combined clocks. The effect of GrimAge is particularly noteworthy. The model's explanatory power with GrimAge alone is only 5% less than when all clocks are included. This implies that adding the other four clocks additional to GrimAge provides minimal additional benefit. Compared to the baseline, GrimAge alone improves the model's explanatory power by 6.3-6.8% across our basic (1) and most comprehensive (5) model specifications, respectively. For context, the covariates on health, lifestyle and parents' longevity collectively increase the model's explanatory power by 6.4%.

Continuing the hierarchy of the strongest predictors among the epigenetic clocks, individually, Hannum adds an additional 5.2% to the baseline model's explanatory power. Levine and Horvath Skin clocks have similar effects, each improving the adjusted R-squared by 4.3% and 4.4%, respectively. Lastly, Horvath's contribution is 3%.

While the clocks offer the most explanatory power when used together, the individual contributions of most clocks are largely similar, leading to an antagonistic cumulative effect.

We now discuss the results of the models that include individual epigenetic clocks.

Tables 5.2.2 and 5.2.3 present the coefficient estimates of the epigenetic clocks from the models where the clocks are included collectively and individually, respectively.

Table 5.2.2: Aggregate Model Results

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
GrimAge	-0.076***	-0.074***	-0.073***	-0.079***	-0.079***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Hannum	-0.034***	-0.032***	-0.032***	-0.031***	-0.032***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Levine	-0.014	-0.014	-0.014	-0.012	-0.011
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Horvath Skin	0.002	0.002	0.001	-0.004	-0.007
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Horvath	0.006	0.005	0.005	0.004	0.002
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Note:					*p<0.1; **p<0.05; ***p<0.01

When we include all epigenetic clocks in the same model, the results are statistically significant only for the GrimAge and Hannum clocks, both of which show a strong negative correlation with SSP (at the 1% level). We standardized the epigenetic clock values, therefore the result for GrimAge suggests that, holding everything else equal, a one standard deviation increase in GrimAge is associated with a 7.6-7.9% decrease in the SSP (equivalent to 16% of the mean SSP). The

estimate for Hannum is nearly twice as small as the estimate for GrimAge.

This indicates that individuals who perceive their chances of survival to be lower indeed have a greater epigenetic clocks, suggesting that they might be aware of their biological age. Given that epigenetic clocks are strongly and positively correlated with mortality outcomes (Lu et al., 2019; Horvath, 2013; Levine et al., 2018; Hannum et al., 2013), the significant negative relationship between SSPs and these clocks provides important evidence of the predictive power of SSPs regarding individuals' survival outcomes

Table 5.2.3: Single-Clock Model Results

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
GrimAge	-0.105***	-0.102***	-0.100***	-0.110***	-0.113***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Hannum	-0.089***	-0.086***	-0.084***	-0.091***	-0.094***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Levine	-0.079***	-0.076***	-0.075***	-0.078***	-0.080***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Horvath Skin	-0.079***	-0.077***	-0.076***	-0.083***	-0.087***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Horvath	-0.062***	-0.059***	-0.059***	-0.064***	-0.067***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Note:					*p<0.1; **p<0.05; ***p<0.01

In contrast to the aggregate model, the results from the single-clock model analysis, where each epigenetic clock was included separately, demonstrate significant estimates at the 1% level for all clocks, across all models. This indicates that the information content of these clocks largely overlaps. Each clock, while distinct in its specific measurement, appears to capture a similar underlying factors of an individual's biological age, which is significantly associated with subjective survival probabilities.

GrimAge shows the strongest negative correlation with SSPs, which is aligned with the fact that it was specifically developed to predict longevity. The coefficients

range from -0.105 to -0.113, indicating that a one standard deviation increase in GrimAge is associated with a decrease in SSPs, ranging from 10.5% to 11.3% (21-23% of the mean SSP). It is followed by Hannum, which has an effect of 8.9 - 9.4% (18-19% of the mean SSP). Horvath shows the smallest effect of the five clocks which ranges from 6.2-6.7% (12.7-13.7% of the mean SSP).

5.3 Implications and Applications of Findings

In the broader context of annuity markets and public policy, the findings of this thesis, specifically, the predictive validity of SSPs regarding mortality risk and their strong correlation with epigenetic clocks are important findings for several reasons.

First, as discussed in the initial chapters of this thesis, the predictive power of SSPs about mortality may indicate the presence of asymmetric information in annuity markets. Our findings suggest that the informational content of individuals' SSPs is predictive of mortality, over and above the combined effect of socio-economic status, genetic factors (approximated by parental longevity), lifestyle, and objective health characteristics. Moreover, this information is private, being reflected through individuals' perception about their longevity and health status.

Biological age, as measured by the epigenetic clocks, is likely to be a more precise predictor of life expectancy than chronological age alone. However, there are significant legal and economical challenges associated with obtaining DNA data and its use. Given that SSPs are significantly correlated with epigenetic clocks, they could serve as a proxy for biological age. This could be used to improve economic models, in particular estimating the longevity risk, as well as inform public policy on retirement age, social security, and healthcare planning.

The insights have practical applications, especially in the context of actuarial considerations and pricing insurance products. This involves incorporating SSP into a basic pricing model to create alternative fair pricing distributions. This process allows for a comparison between fair pricing distributions generated from different mortality models, considering the presence or absence of subjective risk. Integrating

SSPs, which have shown a strong negative correlation with mortality risk, refines these fair pricing distributions to better align with individuals' subjective perceptions of mortality risk. This method can improve our understanding of how subjective risk influences pricing dynamics, and offer insights for designing more accurate and fair pricing structures.

An additional application of the findings involves dissecting the components of subjective beliefs that are uncorrelated with epigenetic clocks from those that exhibit correlation. This separation allows for examination of their distinct impacts on market equilibrium, particularly in the context of adverse selection. By isolating the subjective beliefs portion that is unrelated to epigenetic clocks and contrasting it with those correlating with clocks, one can gain insights into the differential influences on market dynamics. This analytical approach contributes to understanding how individuals' private health information, reflected in their subjective beliefs and biological age estimations, collectively shape market equilibrium. Such insights are important for refining models of adverse selection.

Lastly, the differences in the accuracy of SSPs could reveal systematic prejudices held by certain socio-demographic groups, possibly due to their misinformation or misperceptions about personal health. This could influence retirement planning and financial decision-making. Policymakers could leverage these findings to better understand and course-correct behaviors related to retirement savings and annuity purchases.

6 Conclusion

Some economists believe that individuals are unlikely to accurately quantify probabilities of uncertain future events. This is attributed to the fact that, although individuals may possess a somewhat articulated internal scale for assessing risks, their perceptions are often biased (Machina, 1990). However, previous research on individuals' perceptions about their mortality risk has provided evidence that these assessments can indeed have significant predictive power regarding actual mortality risks, suggesting that they may not be as biased as previously thought. This thesis contributes by further exploring the predictive power of subjective survival probabilities in relation to actual mortality risk.

Our findings reveal a significant negative correlation between subjective survival probabilities and mortality risk, even after accounting for a comprehensive set of socio-demographic, economic, and health variables. This suggests that individuals' subjective risk perceptions may be based on their private health information and overall well-being, which are not fully captured by observable characteristics. The robustness of subjective survival probabilities as a predictor of mortality risk was further confirmed by its significant association with mortality, even when controlling for self-reported health.

We also observe that the missing values of subjective survival probabilities have a non-random nature and are positively correlated with mortality risk, aligning with existing literature. This emphasizes complexity of how individuals perceive and report their survival chances, with those not providing a survival probability potentially being aware of factors reducing their survival odds. Additionally, our findings suggest that income has a stronger effect on mortality than wealth, and confirm the presence of the "Hispanic paradox".

Our epigenetic clock analysis added a new dimension to the study. The strong correlation between subjective survival probabilities and epigenetic clocks, in particular the GrimAge clock, indicates the potential of subjective survival probabilities as predictors of individuals' biological age and actual mortality outcomes. The costliness

associated with collecting DNA data and calculating epigenetic clocks highlights the significance and practicality of this finding. Using SSP in predictive models is economically feasible, offering valuable information about individuals' biological age.

This study findings have important implications for annuity markets and public policy. The predictive validity of SSP in predicting the mortality risk suggests the presence of asymmetric information in annuity markets. Understanding this could improve models predicting longevity risk, which is crucial for designing and pricing annuity products. Moreover, subjective survival probabilities could serve as a proxy for biological age and help to formulate policies related to retirement savings and annuity purchases. Recognizing the heterogeneity in the accuracy of subjective risk perceptions across different demographic groups could help policymakers better understand prejudices that are common to certain groups and offer targeted education to reduce those distortions.

In summary, this thesis contributes to understanding the information content of SSP and emphasizes the value of integrating them into economic and policy models, potentially leading to more efficient and equitable outcomes in the financial and public sectors. Future research focusing on the exploration of residuals, achieved by incorporating within the model additional subjective variables, such as beliefs and attitudes, or other unexplored dimensions, could unveil valuable insights and further enrich our understanding of subjective risk perceptions.

References

Ai, J., Brockett, P. L., Golden, L. L., & Zhu, W. (2017). Health state transitions and longevity effects on retirees' optimal annuitization. *Journal of Risk and Insurance*, *84*(SI), 319–343.

Baji, P., & Bíró, A. (2018). Adaptation or recovery after health shocks? evidence using subjective and objective health measures. *Health Economics*, *27*(5), 850–864.

Benartzi, S., Previtero, A., & Thaler, R. H. (2011). Annuitization puzzles. *The Journal of Economic Perspectives*, *25*(4), 143–164.

Bissonnette, L., Hurd, M. D., & Michaud, P. C. (2017). Individual survival curves comparing subjective and observed mortality risks. *Health Economics*, *26*(12), e285–e303.

Boyer, M. M., Box-Couillard, S., & Michaud, P.-C. (2020). Demand for annuities: Price sensitivity, risk perceptions, and knowledge. *Journal of Economic Behavior and Organization*, *180*, 883–902.

Brown, J. R., & Finkelstein, A. (2011). Insuring long-term care in the united states. *The Journal of Economic Perspectives*, *25*(4), 119–141.

Brown, J. R., Kapteyn, A., Luttmer, E. F. P., & Mitchell, O. S. (2017). Cognitive constraints on valuing annuities. *Journal of the European Economic Association*, *15*(2), 429–462.

Charupat, N., Kamstra, M. J., & Milevsky, M. A. (2016). The sluggish and asymmetric reaction of life annuity prices to changes in interest rates. *The Journal of Risk and Insurance*, *83*(3), 519–555.

Chiappori, P., Jullien, B., Salanié, B., & Salanié, F. (2006). Asymmetric information in insurance: General testable implications. *The RAND Journal of Economics*, *37*(4), 783–798. http://www.jstor.org/stable/25046274

Chiappori, P., & Salanie, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, *108*(1), 56–78.

Cohen, A., & Siegelman, P. (2010). Testing for adverse selection in insurance markets. *Journal of Risk and Insurance*, *77*, 39–84.

Cutler, D. M., Finkelstein, A., & McGarry, K. (2008). Preference heterogeneity and insurance markets: Explaining a puzzle of insurance. *The American Economic Review*, *98*(2), 157–162.

Davidoff, T., Brown, J. R., & Diamond, P. A. (2005). Annuities and individual welfare. *The American Economic Review*, *95*(5), 1573–1590.

Delavande, A., & Rohwedder, S. (2011). Individuals' uncertainty about future social security benefits and portfolio choice. *Journal of Applied Econometrics*, *26*(3), 498–519.

Dickson, D., Hardy, M., & Waters, H. (2009). *Actuarial mathematics for life contingent risks*. Cambridge University Press.

Duan, R., Fu, Q., Sun, Y., & Li, Q. (2022). Epigenetic clock: A promising biomarker and practical tool in aging. *Ageing Research Reviews*, *81*, 101743. https://doi.org/10.1016/j.arr.2022.101743

Einav, L., & Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *The Journal of Economic Perspectives*, *25*(1), 115–138.

Einav, L., Finkelstein, A., & Cullen, M. R. (2010). Estimating welfare in insurance markets using variation in prices. *The Quarterly Journal of Economics*, *125*(3), 877–921.

Einav, L., Finkelstein, A., & Schrimpf, P. (2007). *The welfare cost of asymmetric information: Evidence from the u.k. annuity market* (tech. rep. Working Papers 13228). National Bureau of Economic Research, Inc.

Finkelstein, A., & Poterba, J. (2004). Adverse selection in insurance markets: Policyholder evidence from the u.k. annuity market. *Journal of Political Economy*, *112*(1), 183–208.

Finkelstein, A., & Poterba, J. (2014). Testing for asymmetric information using "unused observables" in insurance markets: Evidence from the u.k. annuity market. *The Journal of Risk and Insurance*, *81*(4), 709–734.

Finkelstein, A., & McGarry, K. (2006). Multiple dimensions of private information: Evidence from the long-term care insurance market. *American Economic Review*, *96*(4), 938–958. https://doi.org/10.1257/aer.96.4.938

Franzini, L., Ribble, J. C., & Keddie, A. M. (2001). Understanding the hispanic paradox. *Ethnicity & Disease*, *11*(3), 496–518.

Gong, G., & Webb, A. (2008). Mortality heterogeneity and the distributional consequences of mandatory annuitization. *The Journal of Risk and Insurance*, *75*(4), 1055–1079.

Hamermesh, D. S. (1985). Expectations, life expectancy, and economic behavior. *The Quarterly Journal of Economics*, *100*(2), 389–408. https://doi.org/10.2307/1885388

Hamermesh, D. S., & Hamermesh, F. W. (1983). Does perception of life expectancy reflect health knowledge? *American Journal of Public Health*, *73*(8), 911–914.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, *49*(2), 359–367.

Horvath, S. (2013). Dna methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115.

Hurd, M. D., McFadden, D., & Gan, L. (1998). Subjective survival curves and life cycle behavior. In D. Wise (Ed.), *Inquiries in the economics of aging* (pp. 259–305). University of Chicago Press.

Hurd, M. D., & McGarry, K. (1995). Evaluation of the subjective probabilities of survival in the health and retirement study. *Journal of Human Resources*, *30*, s268–92.

Hurd, M. D., & McGarry, K. (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal*, *112*(482), 966–985.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*(1), 21–37.

James, E., & Song, X. (2001). Annuities markets around the world: Money's worth and risk intermediation. *Employee Benefits*.

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., Hou, L., Baccarelli, A. A., Stewart, J. D., Li, Y., Whitsel, E. A., Wilson, J. G., Reiner, A. P., Aviv, A., Lohman, K., Liu, Y., Ferrucci, L., & Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, *10*(4), 573–591.

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., Assimes, T. L., Ferrucci, L., & Horvath, S. (2019). Dna methylation grimage strongly predicts lifespan and healthspan. *Aging*, *11*(2), 303–327.

Machina, M. J. (1990). *Choices under uncertainty: Problems solved and unsolved* (P. B. Hammond & R. Coppock, Eds.). National Academy Press.

McCarthy, D., & Mitchell, O. S. (2004). Annuities for an ageing world. In E. Fornero & E. Luciano (Eds.), *Developing an annuity market in europe* (pp. 13–48). Edward Elgar Publishing.

Milevsky, M. A., & Wu Shao, L. (2010). *Annuities and their derivatives: The recent canadian experience* (Working Paper No. 2010-16). Pension Research Council.

Mitchell, O. S., Poterba, J. M., Warshawsky, M. J., & Brown, J. R. (1999). New evidence on the money's worth of individual annuities. *American Economic Review*, *89*(5), 1299–1318.

Modigliani, F. (1986). Life cycle, individual thrift, and the wealth of nations. *The American Economic Review*, *76*(3), 297–313. http://www.jstor.org/stable/1813352

Paté-Cornell, M. E. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering and System Safety*, *54*, 95–111.

Post, T., & Hanewald, K. (2013). Longevity risk, subjective survival expectations, and individual saving behavior. *Journal of Economic Behavior & Organization*, *86*(100), 200–220.

Poterba, J. M., Venti, S. F., & Wise, D. A. (1996). How retirement saving programs increase saving. *Journal of Economic Perspectives*, *10*(4), 91–112.

Rothschild, M., & Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, *90*(4), 629–649.

Rusconi, R. (2008). *National annuity markets: Features and implications* (OECD Working Papers on Insurance and Private Pensions No. 24). OECD Publishing.

Smith, V. K., Taylor, D. H., & Sloan, F. A. (2001). Longevity expectations and death: Can people predict their own demise? *The American Economic Review*, *91*(4), 1126–1134.

Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, *32*(2), 137–150.

Appendix

Table A1.1: Logistic Regression Result for SSP 75

	Dependent variable: mortality status (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
SSP 75	-0.523***	-0.399***	-0.373***	-0.128*	0.010
	(0.072)	(0.072)	(0.072)	(0.072)	(0.074)
Missing SSP	-0.562***	-0.442***	-0.409***	-0.273***	-0.206**
	(0.081)	(0.082)	(0.084)	(0.085)	(0.085)
Focal SSP	0.193***	0.158**	0.157**	0.156**	0.151**
	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)
Age >60	0.416***	0.407***	0.414***	0.320***	0.321***
	(0.037)	(0.038)	(0.038)	(0.038)	(0.038)
Age in [55, 59]	-0.534***	-0.556***	-0.569***	-0.645***	-0.658***
	(0.079)	(0.079)	(0.080)	(0.080)	(0.080)
Male	0.289***	0.362***	0.283***	0.310***	0.315***
	(0.033)	(0.034)	(0.035)	(0.036)	(0.036)
Black	0.138***	-0.024	-0.007	-0.016	-0.049
	(0.043)	(0.045)	(0.046)	(0.047)	(0.047)
Hispanic	-0.461***	-0.596***	-0.550***	-0.431***	-0.459***
	(0.066)	(0.067)	(0.068)	(0.068)	(0.068)
High school	-0.194***	-0.071*	-0.040	-0.019	0.021
	(0.041)	(0.042)	(0.043)	(0.043)	(0.043)
College	-0.370***	-0.142***	-0.112**	-0.098**	-0.038
	(0.045)	(0.048)	(0.048)	(0.048)	(0.049)
Parents' edu >12	-0.086**	-0.057	-0.060	-0.046	-0.031
	(0.037)	(0.037)	(0.038)	(0.038)	(0.038)
Income in Q2		-0.244***	-0.249***	-0.202***	-0.161***
		(0.045)	(0.045)	(0.046)	(0.046)
Income in Q3		-0.372***	-0.360***	-0.297***	-0.246***
		(0.051)	(0.051)	(0.051)	(0.052)
Income in Q4		-0.517***	-0.510***	-0.460***	-0.383***
		(0.059)	(0.059)	(0.060)	(0.061)
Wealth in Q2		-0.140***	-0.108**	-0.029	-0.006

Table A1.1 continued from previous page

	Dependent variable: mortality risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
		(0.045)	(0.046)	(0.046)	(0.046)
Wealth in Q3		-0.216***	-0.180***	-0.080	-0.046
		(0.052)	(0.052)	(0.053)	(0.053)
Wealth in Q4		-0.296***	-0.228***	-0.110*	-0.073
		(0.057)	(0.057)	(0.058)	(0.058)
Mother lived >75		-0.149***	-0.140***	-0.104***	-0.109***
		(0.034)	(0.034)	(0.035)	(0.035)
Father lived >75		-0.120***	-0.115***	-0.091***	-0.099***
		(0.034)	(0.034)	(0.034)	(0.034)
Smoker			0.455***	0.410***	0.406***
			(0.038)	(0.038)	(0.038)
Obesity			0.116***	-0.029	-0.058
			(0.042)	(0.044)	(0.044)
Overweight			-0.049	-0.079**	-0.078**
			(0.039)	(0.040)	(0.040)
Cancer				0.536***	0.490***
				(0.068)	(0.068)
Diabetes				0.663***	0.589***
				(0.048)	(0.048)
Stroke				0.485***	0.414***
				(0.082)	(0.083)
Heart disease				0.342***	0.255***
				(0.049)	(0.050)
Lung disease				0.505***	0.422***
				(0.064)	(0.064)
High blood pres.				0.214***	0.167***
				(0.036)	(0.037)
Psych condition				0.244***	0.177***
				(0.060)	(0.061)
SRH: fair					-0.128**
					(0.062)
SRH: good					-0.309***
					(0.064)
SRH: very good					-0.459***

Table A1.1 continued from previous page

	Dependent variable: mortality risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
					(0.071)
SRH: excellent					-0.629***
					(0.079)
Constant	-3.568***	-3.207***	-3.574***	-4.083***	-3.843***
	(0.061)	(0.068)	(0.077)	(0.082)	(0.093)
Observations	195263	195263	193799	193799	193799

Note: *p<0.1; **p<0.05; ***p<0.01

Table A1.2: Logistic Regression Result for SSP 85

	Dependent variable: mortality status (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
SSP 85	-0.261***	-0.160**	-0.135*	-0.020	0.073
	(0.073)	(0.073)	(0.073)	(0.073)	(0.073)
Missing SSP	0.391***	0.305***	0.270***	0.167**	0.125*
	(0.069)	(0.070)	(0.071)	(0.072)	(0.072)
Focal SSP	0.293***	0.240***	0.228***	0.126**	0.093*
	(0.055)	(0.055)	(0.055)	(0.055)	(0.056)
Age >60	0.423***	0.416***	0.421***	0.323***	0.324***
	(0.037)	(0.037)	(0.038)	(0.038)	(0.038)
Age in [55, 59]	-0.536***	-0.564***	-0.577***	-0.652***	-0.664***
	(0.079)	(0.079)	(0.080)	(0.080)	(0.080)
Male	0.289***	0.368***	0.292***	0.320***	0.325***
	(0.033)	(0.034)	(0.035)	(0.036)	(0.036)
Black	0.156***	-0.023	-0.008	-0.019	-0.055
	(0.043)	(0.045)	(0.046)	(0.047)	(0.047)
Hispanic	-0.436***	-0.556***	-0.512***	-0.397***	-0.427***
	(0.065)	(0.066)	(0.067)	(0.068)	(0.068)
High school	-0.197***	-0.070*	-0.037	-0.012	0.029
	(0.041)	(0.042)	(0.043)	(0.043)	(0.043)
College	-0.369***	-0.128***	-0.098**	-0.087*	-0.026
	(0.045)	(0.048)	(0.048)	(0.048)	(0.049)
Parents' edu >12	-0.081**	-0.045	-0.049	-0.036	-0.022
	(0.037)	(0.037)	(0.037)	(0.038)	(0.038)
Income in Q2		-0.252***	-0.257***	-0.209***	-0.165***
		(0.045)	(0.045)	(0.046)	(0.046)
Income in Q3		-0.378***	-0.366***	-0.303***	-0.249***
		(0.050)	(0.051)	(0.051)	(0.052)
Income in Q4		-0.521***	-0.516***	-0.467***	-0.388***
		(0.059)	(0.059)	(0.060)	(0.060)
Wealth in Q2		-0.143***	-0.111**	-0.031	-0.005
		(0.045)	(0.046)	(0.046)	(0.046)
Wealth in Q3		-0.225***	-0.189***	-0.085	-0.049
		(0.052)	(0.052)	(0.053)	(0.053)
Wealth in Q4		-0.302***	-0.235***	-0.112*	-0.072

Table A1.2 continued from previous page

	Dependent variable: mortality risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
		(0.057)	(0.057)	(0.058)	(0.058)
Mother lived >85		-0.259***	-0.254***	-0.214***	-0.219***
		(0.036)	(0.036)	(0.037)	(0.037)
Father lived >85		-0.245***	-0.233***	-0.208***	-0.217***
		(0.045)	(0.046)	(0.046)	(0.046)
Smoker			0.449***	0.407***	0.405***
			(0.038)	(0.038)	(0.038)
Obesity			0.117***	-0.032	-0.063
			(0.042)	(0.043)	(0.044)
Overweight			-0.052	-0.080**	-0.079**
			(0.039)	(0.039)	(0.040)
Cancer				0.536***	0.485***
				(0.067)	(0.067)
Diabetes				0.650***	0.574***
				(0.048)	(0.048)
Stroke				0.494***	0.414***
				(0.082)	(0.082)
Heart disease				0.337***	0.247***
				(0.049)	(0.049)
Lung disease				0.513***	0.427***
				(0.063)	(0.064)
High blood pres.				0.216***	0.168***
				(0.036)	(0.037)
Psych condition				0.254***	0.182***
				(0.060)	(0.061)
SRH: fair					-0.141**
					(0.062)
SRH: good					-0.327***
					(0.063)
SRH: very good					-0.473***
					(0.070)
SRH: excellent					-0.649***
					(0.079)
Constant	-3.840***	-3.449***	-3.799***	-4.170***	-3.868***

Table A1.2 continued from previous page

	Dependent variable: mortality risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
	(0.056)	(0.061)	(0.071)	(0.074)	(0.087)
Observations	195624	195624	194160	194160	194160

Note: *p<0.1; **p<0.05; ***p<0.01

Table A1.3: Marginal Effects for SSP 75

	Dependent variable: mortality status (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
SSP 75	-0.01***	-0.008***	-0.007***	-0.002*	0.0
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Missing SSP	0.011***	0.008***	0.008***	0.005***	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Focal SSP	0.004**	0.003**	0.003**	0.003**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age in [55, 59]	-0.01***	-0.011***	-0.011***	-0.012***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age >60	0.008***	0.008***	0.008***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.005***	0.007***	0.005***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Black	0.003***	-0.001	-0.0	-0.0	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.009***	-0.011***	-0.01***	-0.008***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
College	-0.007***	-0.003***	-0.002**	-0.002**	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High school	-0.004***	-0.001*	-0.001	-0.0	0.0
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Parents' edu >12	-0.002**	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Income in Q2		-0.005***	-0.005***	-0.004***	-0.003***
		(0.001)	(0.001)	(0.001)	(0.001)
Income in Q3		-0.007***	-0.007***	-0.006***	-0.005***
		(0.001)	(0.001)	(0.001)	(0.001)
Income in Q4		-0.01***	-0.01***	-0.009***	-0.007***
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q2		-0.003***	-0.002**	-0.001	0.0
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q3		-0.004***	-0.003***	-0.002	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q4		-0.006***	-0.004***	-0.002*	-0.001

Table A1.3 continued from previous page

	Dependent variable: mortality risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
		(0.001)	(0.001)	(0.001)	(0.001)
Mother lived >75		-0.003***	-0.003***	-0.002***	-0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
Father lived >75		-0.002***	-0.002***	-0.002***	-0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
Smoker			0.009***	0.008***	0.008***
			(0.001)	(0.001)	(0.001)
Obesity			0.002***	-0.001	-0.001
			(0.001)	(0.001)	(0.001)
Overweight			-0.001	-0.002**	-0.002*
			(0.001)	(0.001)	(0.001)
Cancer				0.01***	0.009***
				(0.001)	(0.001)
Diabetes				0.013***	0.011***
				(0.001)	(0.001)
Stroke				0.009***	0.008***
				(0.002)	(0.002)
Heart disease				0.006***	0.005***
				(0.001)	(0.001)
Lung disease				0.009***	0.008***
				(0.001)	(0.001)
High blood pres.				0.004***	0.003***
				(0.001)	(0.001)
Psych condition				0.005***	0.003***
				(0.001)	(0.001)
SRH: fair					-0.002**
					(0.001)
SRH: good					-0.006***
					(0.001)
SRH: very good					-0.009***
					(0.001)
SRH: excellent					-0.012***
					(0.002)

Note: *p<0.1; **p<0.05; ***p<0.01

Table A1.4: Marginal Effects for SSP 85

	Dependent variable: mortality status (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
SSP 85	-0.005***	-0.003**	-0.003*	-0.0	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Missing SSP	0.007***	0.006***	0.005***	0.003**	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Focal SSP	0.006***	0.005***	0.004***	0.002**	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age in [55, 59]	-0.01***	-0.011***	-0.011***	-0.012***	-0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age >60	0.008***	0.008***	0.008***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.005***	0.007***	0.005***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Black	0.003***	-0.0	-0.0	-0.0	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.008***	-0.011***	-0.01***	-0.007***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High school	-0.004***	-0.001	-0.001	-0.0	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
College	-0.007***	-0.002***	-0.002**	-0.002*	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Parents' edu >12	-0.002**	-0.001	-0.001	-0.001	-0.0
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mother lived >85		-0.005***	-0.005***	-0.004***	-0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Father lived >85		-0.005***	-0.004***	-0.004***	-0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Income in Q2		-0.005***	-0.005***	-0.004***	-0.003***
		(0.001)	(0.001)	(0.001)	(0.001)
Income in Q3		-0.007***	-0.007***	-0.006***	-0.005***
		(0.001)	(0.001)	(0.001)	(0.001)
Income in Q4		-0.01***	-0.01***	-0.009***	-0.007***
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q2		-0.003***	-0.002**	-0.001	-0.0

Table A1.4 continued from previous page

	Dependent variable: mortality risk (mean=0.0195)				
	(1)	(2)	(3)	(4)	(5)
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q3		-0.004***	-0.004***	-0.002	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Wealth in Q4		-0.006***	-0.004***	-0.002*	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Smoker			0.009***	0.008***	0.008***
			(0.001)	(0.001)	(0.001)
Obesity			0.002***	-0.001	-0.001
			(0.001)	(0.001)	(0.001)
Overweight			-0.001	-0.002**	-0.002**
			(0.001)	(0.001)	(0.001)
Cancer				0.01***	0.009***
				(0.001)	(0.001)
Diabetes				0.012***	0.011***
				(0.001)	(0.001)
Stroke				0.009***	0.008***
				(0.002)	(0.002)
Heart disease				0.006***	0.005***
				(0.001)	(0.001)
Lung disease				0.01***	0.008***
				(0.001)	(0.001)
High blood pres.				0.004***	0.003***
				(0.001)	(0.001)
Psych condition				0.005***	0.003***
				(0.001)	(0.001)
SRH: fair					-0.003**
					(0.001)
SRH: good					-0.006***
					(0.001)
SRH: very good					-0.009***
					(0.001)
SRH: excellent					-0.012***
					(0.001)

Note: *p<0.1; **p<0.05; ***p<0.01

Table A2.1: Single-Clock Model for GrimAge

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
GrimAge	-0.105***	-0.102***	-0.100***	-0.110***	-0.113***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Age in [55, 59]	0.009	0.010	0.010	0.017	0.017
	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
Age >60	-0.034**	-0.021	-0.022	-0.005	-0.011
	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)
Male	0.009	0.010	0.009	0.009	0.006
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Black	0.105***	0.105***	0.110***	0.113***	0.117***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Hispanic	-0.068***	-0.058***	-0.061***	-0.067***	-0.056***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
College	0.091***	0.077***	0.075***	0.078***	0.063***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
High school	0.042***	0.034**	0.033**	0.032**	0.023
	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)
Parents' edu >12	0.043***	0.043***	0.041***	0.038***	0.036***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Income in Q2	0.038**	0.030**	0.029*	0.028*	0.018
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Income in Q3	0.075***	0.068***	0.065***	0.061***	0.047***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Income in Q4	0.093***	0.083***	0.082***	0.076***	0.059***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Wealth in Q2	-0.009	-0.008	-0.007	-0.016	-0.019
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Wealth in Q3	0.016	0.013	0.012	0.001	-0.005
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Wealth in Q4	0.014	0.017	0.014	0.002	-0.005
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Mother lived >75			0.026**	0.022**	0.018

Table A2.1 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
			(0.011)	(0.011)	(0.011)
Father lived >75			0.038***	0.034***	0.036***
			(0.010)	(0.010)	(0.010)
Smoker				0.028***	0.029***
				(0.011)	(0.011)
Obesity				-0.003	0.007
				(0.013)	(0.013)
Overweight				0.006	0.008
				(0.013)	(0.013)
Cancer				-0.004	0.009
				(0.025)	(0.025)
Diabetes				-0.066***	-0.052***
				(0.018)	(0.018)
Stroke				0.019	0.036
				(0.032)	(0.032)
Heart disease				-0.042**	-0.020
				(0.019)	(0.019)
Lung disease				-0.057**	-0.041
				(0.028)	(0.028)
High blood pres.				-0.019*	-0.006
				(0.011)	(0.012)
Psych condition				-0.053***	-0.030
				(0.018)	(0.018)
SRH: fair					0.014
					(0.026)
SRH: good					0.086***
					(0.026)
SRH: very good					0.101***
					(0.027)
SRH: excellent					0.143***
					(0.029)
Missing SSP		-0.071**	-0.069**	-0.056*	-0.044
		(0.029)	(0.029)	(0.029)	(0.029)

Table A2.1 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
Focal SSP		-0.155***	-0.150***	-0.133***	-0.115***
		(0.025)	(0.025)	(0.025)	(0.025)
Constant	0.350***	0.376***	0.343***	0.362***	0.284***
	(0.018)	(0.019)	(0.020)	(0.023)	(0.033)
Observations	3329	3329	3329	3319	3319
R^2	0.209	0.223	0.227	0.240	0.251
Adjusted R^2	0.206	0.219	0.223	0.233	0.243
Residual Std. Error	0.297	0.295	0.294	0.292	0.290
F Statistic	58.479***	55.766***	51.252***	35.747***	33.346***
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A2.2: Single-Clock Model for Hannum

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
Hannum	-0.089***	-0.086***	-0.084***	-0.091***	-0.094***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Age in [55, 59]	0.001	0.002	0.002	0.008	0.008
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Age >60	-0.054***	-0.041**	-0.041**	-0.028*	-0.034**
	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)
Male	-0.012	-0.011	-0.012	-0.007	-0.010
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Black	0.080***	0.080***	0.087***	0.088***	0.091***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Hispanic	-0.042**	-0.032*	-0.036**	-0.042**	-0.030*
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
College	0.106***	0.091***	0.088***	0.089***	0.074***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
High school	0.047***	0.039**	0.038**	0.038**	0.028*
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Parents' edu >12	0.047***	0.048***	0.045***	0.044***	0.042***

Table A2.2 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Income in Q2	0.040***	0.032**	0.031**	0.030**	0.021
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Income in Q3	0.079***	0.071***	0.068***	0.065***	0.052***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Income in Q4	0.111***	0.101***	0.099***	0.094***	0.077***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Wealth in Q2	-0.001	-0.001	0.000	-0.010	-0.013
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Wealth in Q3	0.020	0.016	0.016	0.002	-0.004
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Wealth in Q4	0.023	0.026	0.023	0.007	0.000
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Mother lived >75			0.029**	0.025**	0.022*
			(0.011)	(0.011)	(0.011)
Father lived >75			0.041***	0.037***	0.039***
			(0.010)	(0.010)	(0.010)
Smoker				-0.010	-0.010
				(0.011)	(0.011)
Obesity				0.001	0.012
				(0.013)	(0.013)
Overweight				0.006	0.008
				(0.013)	(0.013)
Cancer				-0.004	0.008
				(0.025)	(0.025)
Diabetes				-0.060***	-0.045**
				(0.018)	(0.018)
Stroke				0.020	0.037
				(0.032)	(0.032)
Heart disease				-0.044**	-0.022
				(0.019)	(0.019)
Lung disease				-0.067**	-0.051*
				(0.029)	(0.029)

Table A2.2 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
High blood pres.				-0.020*	-0.007
				(0.012)	(0.012)
Psych condition				-0.049***	-0.025
				(0.018)	(0.019)
SRH: fair					0.015
					(0.027)
SRH: good					0.087***
					(0.027)
SRH: very good					0.103***
					(0.028)
SRH: excellent					0.145***
					(0.029)
Missing SSP		-0.073**	-0.070**	-0.057*	-0.046
		(0.029)	(0.029)	(0.029)	(0.029)
Focal SSP		-0.158***	-0.153***	-0.138***	-0.119***
		(0.025)	(0.025)	(0.025)	(0.025)
Constant	0.342***	0.369***	0.333***	0.371***	0.292***
	(0.019)	(0.019)	(0.021)	(0.023)	(0.033)
Observations	3329	3329	3329	3319	3319
R^2	0.193	0.206	0.212	0.223	0.234
Adjusted R^2	0.189	0.202	0.208	0.216	0.227
Residual Std. Error	0.300	0.298	0.297	0.295	0.293
F Statistic	52.715***	50.661***	46.880***	32.556***	30.491***
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A2.3: Single-Clock Model for Levine

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
Levine	-0.079***	-0.076***	-0.075***	-0.078***	-0.080***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Age in [55, 59]	-0.005	-0.003	-0.003	0.001	0.000

Table A2.3 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Age >60	-0.070***	-0.055***	-0.055***	-0.046***	-0.052***
	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)
Male	-0.022**	-0.020*	-0.021**	-0.019*	-0.022**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Black	0.105***	0.105***	0.110***	0.113***	0.117***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Hispanic	-0.039**	-0.030*	-0.034**	-0.038**	-0.026
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
College	0.105***	0.090***	0.088***	0.090***	0.076***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
High school	0.044***	0.036**	0.035**	0.035**	0.026*
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Parents' edu >12	0.052***	0.052***	0.050***	0.049***	0.049***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Income in Q2	0.039**	0.031**	0.030**	0.030*	0.021
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Income in Q3	0.076***	0.068***	0.066***	0.063***	0.051***
	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)
Income in Q4	0.110***	0.100***	0.098***	0.094***	0.079***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Wealth in Q2	-0.011	-0.010	-0.009	-0.019	-0.022
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Wealth in Q3	0.014	0.011	0.011	-0.002	-0.008
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Wealth in Q4	0.015	0.018	0.015	0.001	-0.005
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Mother lived >75			0.025**	0.022*	0.019
			(0.011)	(0.011)	(0.011)
Father lived >75			0.043***	0.040***	0.042***
			(0.010)	(0.010)	(0.010)
Smoker				-0.007	-0.007
				(0.011)	(0.011)

Table A2.3 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
Obesity				0.006	0.016
				(0.013)	(0.013)
Overweight				0.009	0.011
				(0.013)	(0.013)
Cancer				-0.001	0.010
				(0.025)	(0.025)
Diabetes				-0.053***	-0.039**
				(0.018)	(0.018)
Stroke				0.017	0.034
				(0.032)	(0.032)
Heart disease				-0.041**	-0.020
				(0.019)	(0.019)
Lung disease				-0.056*	-0.040
				(0.029)	(0.029)
High blood pres.				-0.016	-0.003
				(0.012)	(0.012)
Psych condition				-0.044**	-0.022
				(0.018)	(0.019)
SRH: fair					0.016
					(0.027)
SRH: good					0.087***
					(0.027)
SRH: very good					0.097***
					(0.028)
SRH: excellent					0.135***
					(0.029)
Missing SSP		-0.066**	-0.064**	-0.053*	-0.042
		(0.030)	(0.029)	(0.030)	(0.030)
Focal SSP		-0.158***	-0.152***	-0.140***	-0.122***
		(0.025)	(0.025)	(0.025)	(0.026)
Constant	0.350***	0.377***	0.343***	0.372***	0.295***
	(0.019)	(0.019)	(0.021)	(0.023)	(0.034)

Table A2.3 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
Observations	3329	3329	3329	3319	3319
R^2	0.187	0.202	0.207	0.216	0.225
Adjusted R^2	0.184	0.197	0.203	0.209	0.218
Residual Std. Error	0.301	0.299	0.298	0.297	0.295
F Statistic	50.864***	49.176***	45.512***	31.175***	28.977***
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A2.4: Single-Clock Model for Horvath Skin

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
Horvath Skin	-0.079***	-0.077***	-0.076***	-0.083***	-0.087***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Age in [55, 59]	-0.001	0.001	0.001	0.006	0.007
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Age >60	-0.059***	-0.045***	-0.045***	-0.031*	-0.036**
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Male	-0.022**	-0.020*	-0.021**	-0.017	-0.020*
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Black	0.098***	0.099***	0.105***	0.107***	0.110***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Hispanic	-0.042**	-0.033*	-0.037**	-0.043**	-0.031*
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
College	0.108***	0.093***	0.090***	0.091***	0.076***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
High school	0.048***	0.040**	0.038**	0.038**	0.028*
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Parents' edu >12	0.051***	0.051***	0.048***	0.047***	0.045***
Income in Q2	0.042***	0.033**	0.032**	0.032**	0.022
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
Income in Q3	0.080***	0.073***	0.070***	0.066***	0.053***
	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)
Income in Q4	0.114***	0.104***	0.102***	0.096***	0.079***

Table A2.4 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Wealth in Q2	-0.002	-0.002	-0.000	-0.010	-0.013
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Wealth in Q3	0.020	0.017	0.017	0.003	-0.003
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Wealth in Q4	0.021	0.024	0.021	0.006	-0.001
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Mother lived >75			0.032***	0.028**	0.024**
			(0.011)	(0.011)	(0.011)
Father lived >75			0.041***	0.038***	0.039***
			(0.011)	(0.010)	(0.010)
Smoker				-0.011	-0.011
				(0.011)	(0.011)
Obesity				0.003	0.014
				(0.013)	(0.013)
Overweight				0.009	0.012
				(0.013)	(0.013)
Cancer				-0.005	0.008
				(0.025)	(0.025)
Diabetes				-0.065***	-0.051***
				(0.018)	(0.018)
Stroke				0.020	0.037
				(0.032)	(0.032)
Heart disease				-0.037*	-0.015
				(0.019)	(0.019)
Lung disease				-0.068**	-0.052*
				(0.029)	(0.029)
High blood pres.				-0.021*	-0.008
				(0.012)	(0.012)
Psych condition				-0.046**	-0.022
				(0.018)	(0.019)
SRH: fair					0.020
					(0.027)

Table A2.4 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
SRH: good					0.092***
					(0.027)
SRH: very good					0.106***
					(0.028)
SRH: excellent					0.150***
					(0.029)
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Missing SSP		-0.073**	-0.070**	-0.058*	-0.046
		(0.030)	(0.030)	(0.030)	(0.030)
Focal SSP		-0.162***	-0.156***	-0.142***	-0.122***
		(0.025)	(0.025)	(0.025)	(0.025)
Constant	0.341***	0.369***	0.331***	0.367***	0.283***
	(0.019)	(0.019)	(0.021)	(0.023)	(0.033)
Observations	3329	3329	3329	3319	3319
R^2	0.183	0.197	0.203	0.215	0.226
Adjusted R^2	0.179	0.193	0.199	0.208	0.219
Residual Std. Error	0.302	0.299	0.298	0.297	0.295
F Statistic	49.376***	47.861***	44.470***	30.997***	29.118***
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A2.5: Single-Clock Model for Horvath

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
				Note:	*p<0.1; **p<0.05; ***p<0.01
Horvath	-0.062***	-0.059***	-0.059***	-0.064***	-0.067***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Age in [55, 59]	-0.012	-0.010	-0.010	-0.006	-0.006
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Age >60	-0.087***	-0.071***	-0.070***	-0.059***	-0.065***
	(0.016)	(0.017)	(0.016)	(0.017)	(0.017)
Male	-0.021**	-0.020*	-0.021*	-0.016	-0.019*

Table A2.5 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Black	0.115***	0.115***	0.121***	0.124***	0.128***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Hispanic	-0.032*	-0.022	-0.027	-0.031*	-0.020
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
High school	0.052***	0.044***	0.042***	0.042***	0.033**
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
College	0.113***	0.098***	0.095***	0.096***	0.081***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Parents' edu >12	0.060***	0.060***	0.057***	0.056***	0.055***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Income in Q2	0.043***	0.035**	0.034**	0.034**	0.024
	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)
Income in Q3	0.087***	0.078***	0.075***	0.073***	0.060***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Income in Q4	0.124***	0.113***	0.111***	0.107***	0.091***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Wealth in Q2	-0.016	-0.016	-0.014	-0.024	-0.027*
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Wealth in Q3	0.011	0.008	0.008	-0.005	-0.011
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Wealth in Q4	0.009	0.013	0.010	-0.004	-0.011
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Mother lived >75			0.029**	0.026**	0.022*
			(0.011)	(0.011)	(0.011)
Father lived >75			0.045***	0.042***	0.043***
			(0.011)	(0.011)	(0.011)
Smoker				-0.011	-0.011
				(0.011)	(0.011)
Obesity				0.004	0.014
				(0.013)	(0.013)
Overweight				0.005	0.007
				(0.013)	(0.013)

Table A2.5 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
Cancer				-0.005	0.007
				(0.025)	(0.025)
Diabetes				-0.060***	-0.046**
				(0.018)	(0.018)
Stroke				0.019	0.037
				(0.033)	(0.033)
Heart disease				-0.041**	-0.019
				(0.019)	(0.019)
High blood pres.				-0.016	-0.003
				(0.012)	(0.012)
Lung disease				-0.060**	-0.043
				(0.029)	(0.029)
Psych condition				-0.037**	-0.014
				(0.019)	(0.019)
SRH: fair					0.028
					(0.027)
SRH: good					0.099***
					(0.027)
SRH: very good					0.111***
					(0.028)
SRH: excellent					0.150***
					(0.030)
Missing SSP		-0.063**	-0.061**	-0.049*	-0.037
		(0.030)	(0.030)	(0.030)	(0.030)
Focal SSP		-0.160***	-0.154***	-0.142***	-0.122***
		(0.026)	(0.025)	(0.026)	(0.026)
Constant	0.337***	0.366***	0.328***	0.360***	0.272***
	(0.019)	(0.019)	(0.021)	(0.024)	(0.034)
Observations	3329	3329	3329	3319	3319
R^2	0.171	0.186	0.193	0.202	0.213
Adjusted R^2	0.167	0.182	0.188	0.195	0.205
Residual Std. Error	0.304	0.301	0.300	0.299	0.297

Table A2.5 continued from previous page

	Dependent variable: SSP 75 (mean=0.49)				
	(1)	(2)	(3)	(4)	(5)
F Statistic	45.494***	44.562***	41.566***	28.686***	26.891***
Note:					*p<0.1; **p<0.05; ***p<0.01