Health, Longevity, and Welfare Inequality of the Elderly*

Ray Miller[†] Neha Bairoliya[‡]

October 28, 2017

Abstract

We propose a framework to understand the distribution of individual well-being and its change over time with an application to the U.S. elderly population. Using data from the Health and Retirement Study, we estimate life-cycle dynamics and simulate individual outcome paths starting from age sixty. We use an expected utility framework and the simulated profiles to construct a measure of individual welfare that incorporates differences in consumption, leisure, health, and mortality. Our measure suggests substantial variation in welfare across individuals driven foremost by gaps in health and mortality followed by gaps in consumption. Incorporating the utility cost of living with poor health into elderly welfare substantially increases overall inequality. Elderly welfare inequality has increased over time due to growing gaps in consumption, health, and mortality. Disparity measures based on cross-sectional income or consumption at age sixty underestimate aggregate welfare inequality. Moreover, health at age sixty is a better indicator of individual well-being rank than income or consumption.

JEL classifications: D63, I14, I31, J14
Keywords: welfare, inequality, elderly, health, morbidity, aging, mortality

*Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Number P30AG024409 and by the National Institutes of Health (NIH, Grant No.: 5R01AG048037-02). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

[†]Corresponding author. Harvard Center for Population and Development Studies, Cambridge, MA, 02138. Email: hrmiller@hsph.harvard.edu

[‡]Harvard Center for Population and Development Studies

1 Introduction

Inequality has been a subject of great interest to researchers and policymakers on grounds of both fairness and potential consequences.[1] However, the most widely used disparity measures are often based on either income or consumption which provide an incomplete metric of social welfare inequality. Leisure, health, social interactions, political and natural environments, and other factors have all been linked to individual well-being. Moreover, strong socioeconomic gradients have been found in related metrics such as life expectancy (Chetty et al., 2016). Given the potential correlation across these factors, a more comprehensive understanding of social welfare and its distribution has significant implications for policy evaluation and prioritization.

We provide a framework to understand how economic circumstances, health, and mortality jointly influence the dispersion of welfare in a given population. Using standard expected utility theory and microsimulations from a model of life-cycle dynamics, we construct a measure of well-being at the individual level—measured as an ex-ante consumption equivalent. This allows us to analyze the entire distribution of welfare. Our measure is based on comparing expected lifetime utility across individuals of a given age. We incorporate differences in the uncertain evolution of consumption, leisure, health, and mortality over remaining life, providing a more complete measure of well-being than consumption or life expectancy alone.

We apply our methods to estimate the welfare distribution among sixty year olds in the U.S. using data from the Heath and Retirement Study (HRS). As our measure at sixty incorporates individual expectations about outcomes over the entirety of remaining life, it provides a useful single metric of ex-ante elderly well-being. For example, we intend to understand questions such as: how much better do we expect remaining life to be for the median sixty year old in the U.S., compared to the sixty year old who is the worst off? Moreover, how much of the difference in well-being is driven by expected gaps in consumption versus gaps in leisure or health? With these questions in mind, we refer to our measure of well-being as elderly welfare, though strictly speaking we are referring to ex-ante welfare at age sixty.

We conduct our analysis on multiple cohorts in the HRS to examine how the distribution of elderly welfare has changed over time. While income and consumption inequality have increased in the United States over the past three decades, the implications for the distribution of individual welfare are unclear.[2] Growing economic inequality may overstate welfare disparities if, for example, some of the effects are mitigated through improvements in public health and health equity. The opposite may be

[1]Fairness is tied to the importance of luck in determining well-being (see Rawls (1971); Dworkin (1985); Roemer (1998)). Inequality has been directly tied to a wide range of outcomes including education, crime, economic growth and mobility, civic engagement, and political influence and polarization. See Kenworthy (2008) for a comprehensive literature review.

[2]See Piketty and Saez (2014); Heathcote et al. (2010); Autor et al. (2008); Katz et al. (1999); Gottschalk et al. (1994) for evidence on income inequality and Attanasio and Pistaferri (2014); Attanasio et al. (2014, 2010); Cutler and Katz (1992) for consumption.

true if health gains accrue disproportionately to the financially well-off, making changes in welfare inequality larger than what would be suggested by economic variables alone.

The influence of health on welfare inequality is particularly relevant among the elderly where most health differences are concentrated (Deaton and Paxson, 1998). Compression of morbidity in the elderly U.S. population (Cutler et al., 2013) has been accompanied by evidence of a strong socioeconomic gradient in disability incidence rates in later life (Minkler et al., 2006). More generally, recent evidence suggests that there has been a widening gap in life expectancy and an increase in the socioeconomic gradient of mortality rates.[3] Our focus on the elderly is further motivated by the rapid aging of the U.S. population—more than 20% of people are estimated to be aged 65 and older by 2050 (Colby et al., 2015).

Our approach to welfare analysis can be summarized in three broad steps. First, we propose a "welfare model" for evaluating individual well-being using an expected utility framework. This model accounts for the impact of consumption, leisure, health, and mortality on well-being and provides a simple analytic decomposition of the contribution of each channel. Next, we propose a dynamic system of equations to approximate the joint evolution of outcomes over the elderly life-cycle. The parameters of the system are estimated using HRS data and a mix of multivariate probit and linear dynamic panel data estimators. Finally, using the estimated system and data from a subset of HRS respondents as initial conditions, we repeatedly simulate potential outcome paths. These paths are embedded in the welfare model to compute an ex-ante measure of well-being for each individual in our sample at age sixty.

We measure welfare of a *given* individual in consumption equivalents; how much consumption would have to increase/decrease across the remaining lifetime of a reference person to yield the same expected level of utility as that obtained by the current and potential future outcome bundles of the *given* individual. In our empirical application, we compare consumption equivalents computed for each individual at the age of sixty using the individual with the median utility ranking as our reference person. This measure incorporates all expected inequalities in outcomes across individuals over their remaining lives. It also accounts for welfare costs of uncertainty in outcomes after sixty, providing a useful metric of ex-ante elderly welfare.

The most salient findings of our analysis can be summarized as follows:

1. There is substantial variation in the ex-ante welfare of individuals at age sixty. The Gini coefficient for consumption-equivalent welfare in our benchmark cohort is 0.66. Those at the ninetieth percentile of the welfare distribution have 23 times higher welfare than those at the tenth percentile.

2. Health differences have important implications for the distribution of elderly well-being. Excluding the utility cost of living with poor health and morbidities lowers

[3] See, for example, Chetty et al. 2016; Currie and Schwandt 2016; National Academies of Sciences, Engineering, and Medicine 2015; Pijoan-Mas and Ríos-Rull 2014; Meara et al. 2008

our welfare Gini coefficient by 23%. It significantly under-predicts relative welfare for those in the top end of the distribution and over-predicts for those at the bottom. This is driven by a positive correlation between health, consumption, and mortality.

3. The largest drivers of elderly welfare inequality are health and mortality gaps followed by gaps in consumption. Differences in leisure play a comparatively minor role.

4. Welfare inequality among the elderly has increased over time due to growing gaps in consumption, health, and mortality. Compared to the cohort of individuals reaching age sixty between 1992-2001, the welfare Gini rose 9% for those reaching sixty between 2002-07 and 22% for those reaching between 2008-14.

5. Ignoring dynamic uncertainty and the persistence in outcomes over the life-cycle greatly underestimates welfare inequality. The Gini of age sixty flow utility is only 70% of that based on our dynamic welfare measure.

A key implication of our results is that cross-sectional distributions of income and consumption underestimate aggregate welfare inequality at age sixty. This occurs for two primary reasons. First, cross-sectional measures ignore dynamic uncertainty and the persistence of inequality over remaining life. Second, there is a positive correlation between health and consumption, implying those with high consumption also enjoy better health and longer lives on average. However, even in cases where economic outcomes provide a reasonable approximation to aggregate welfare inequality, our results suggest they may still provide a poor ranking of individual well-being. For example, the rank correlation between consumption and welfare is a relatively modest 0.56 for our benchmark cohort. Moreover, we find cross-sectional health utility at age sixty to be a better predictor of remaining lifetime welfare rank, despite the fact that it drastically underestimates aggregate welfare inequality.

Our paper builds on a large body of work attempting to extend measures of welfare beyond income (see Fleurbaey (2009) for an extensive review). Recent examples include Becker et al. (2005) who combine national income and expected longevity in a utility framework to examine the changes in cross-country inequality over time. Fleurbaey and Gaulier (2009) extend this work by examining level differences across countries and incorporating leisure, health-adjusted life expectancy, and aggregate inequality. Our welfare framework builds on recent work by Jones and Klenow (2016) who construct an alternate cross-country measure of economic well-being. Our work is different from these papers along many dimensions. Most notably, by using longitudinal data and estimating the joint dynamic process of outcomes, we are able to construct welfare at the individual as opposed to aggregate level. Moreover, we explicitly allow for health to affect individual welfare by mapping subjective and objective measures directly into utility. We also focus on the U.S. elderly and examine inequality evolution over birth cohorts as opposed to cross-sectional changes over time.

Our paper is also more broadly tied to the literature measuring economic well-being at older ages. Using HRS data, Hurd and Rohwedder (2007) find that income based poverty rates underestimate economic well-being of the elderly compared to consumption based measures. Crystal and Shea (1990) argue that despite the increased presence of social safety nets at older ages, economic inequalities are exacerbated with aging due to the accumulation of "economic advantages and disadvantages" over the entire life-course. Using longitudinal data, Crystal and Waehrer (1996) likewise find that income inequality rises within cohorts as they age, but also document considerable mobility in relative income position. More recently, Bosworth et al. (2016) document that income inequality has increased for the elderly over the past three decades, though more slowly than among the non-elderly perhaps due to wider availability of social safety nets at older ages. We contribute to this line of research by providing estimates of elderly well-being that incorporates consumption, leisure, and health into a single measure rooted in economic and health theory.

Several limitations to our approach warrant mentioning at the outset. First, we do not explicitly account for morbidity spillover effects such as the cost of caregiver time and the numerous costs associated with the loss of a spouse. Likewise, we abstract from other potentially important inputs into elderly welfare such as social interactions, bequests, and end-of-life care. Second, we estimate welfare based on common preferences. Considering heterogeneity across individual's preferences could reduce the welfare costs of inequality along some components. Finally, we assume institutions and relevant policies remain fixed moving forward. Significant anticipated changes to Social Security or Medicare programs in the future, for example, could alter our welfare measure, particularly for the younger cohorts we study.

The remainder of the paper is organized as follows. Section 2 outlines the theory including our models of welfare and life-cycle dynamics. Section 3 provides details of the data and empirical methods used in our analysis. Section 4 discusses our welfare results including robustness to alternate modeling assumptions. Finally, section 5 provides concluding remarks.

2 Theory

This section outlines the expected utility framework used in the welfare analysis and the life-cycle dynamics model used for estimating consumption, leisure, morbidity and mortality profiles.

2.1 Welfare model

Our welfare concept aims to compare well-being across individuals of a given age j. These individuals may differ along many dimensions due to their childhood environment, education, occupation, previous health behaviors, and numerous other factors. We define individual welfare based on observed outcomes at age j and the potential realization of outcomes in the future based on these multi-dimensional differences. Although individuals are heterogeneous, we make welfare comparisons through a common preference specification. These preferences are defined by an expected lifetime utility at age j given by:

$$E\left[\sum_{a=j}^{J}\psi_a\beta^{a-j}u\left(c_a,l_a,h_a\right)\right]$$

where c is consumption, l leisure, h health, and ψ survival probability. Expectations are taken with respect to the uncertainty in the evolution of consumption, leisure, health, and mortality probability after age j.

We use a consumption-equivalent variation (EV) measure to quantify welfare differences across individuals. This approach requires choosing a reference person as the basis for our welfare comparisons. Welfare λ_{ij} is then the proportion of the reference individual's consumption that must be maintained at every age starting from j (in all possible realizations of the world and holding health and leisure fixed) that would make them indifferent to facing the current and potential future outcome bundles of individual i. For example, if person i is relatively poor and unhealthy, we may have a welfare measure $\lambda_{ij} = 0.3$. This implies the reference individual would be ex-ante indifferent between maintaining 30% of their own consumption every period from age j or receiving the outcome bundle of person i at age j and facing person i's stochastic evolution of consumption, leisure, health, and mortality profiles moving forward. As this measure is based on potential outcomes over the remaining life, it encompasses the cross-sectional inequality in outcomes at age j as well as the likelihood of persistence and emergence of inequalities in future outcomes.

Let $U_{ij}\left(\lambda\right)$ denote the expected lifetime utility at age j from the outcome bundles of individual i if consumption is multiplied by a factor λ at each age and realization of the world:

$$U_{ij}\left(\lambda\right)=E\left[\sum_{a=j}^{J}\psi_{ia}\beta^{a-j}u\left(\lambda c_{ia},l_{ia},h_{ia}\right)\right].$$

The consumption-equivalent variation measure of welfare for individual i, λ_{ij}, is derived through the condition:

$$U_{mj}\left(\lambda_{ij}\right)=U_{ij}\left(1\right) \tag{1}$$

where U_{mj} refers to the expected lifetime utility from the outcome bundles of the reference individual.

In our benchmark model we assume that preferences are additively separable between consumption and leisure and non-separable between health and the consumption-leisure composite. The flow utility takes the following form:

$$u\left(c,l,h\right)=\phi\left(h\right)\left[\bar{u}+log\left(c\right)+\nu\left(l\right)\right]. \tag{2}$$

The additive separability between consumption and leisure allows for a simple decomposition of welfare effects but we check the robustness of our results to more general preferences. Our treatment of health in the utility function follows from a large literature on quality-adjusted life years (QALYs) going back to the works of Klarman and Rosenthal (1968); Fanshel and Bush (1970); Torrance et al. (1972); and Zeckhauser and Shepard (1976). The central assumption is that health utility is a function of both the length and quality of life. The QALY literature provides a framework to combine these two aspects of health in a single index. Accordingly, preferences over health are chosen such that period utility from whatever is regarded as the best possible health state or "full health" equals one. In our framework, health function $\phi(h) \in [0,1]$ scales the utility from consumption and leisure such that $\phi(h) = 1$ represents utility in the perfect health state and $\phi(h) = 0$ represents the dead state. At the same time, $\psi\phi(h)$ represents a measure of QALYs. For instance, $\psi\phi(h) = 1$ represents a year of life with no morbidity; a single QALY.

Under preferences given in (2), welfare condition (1) may be rewritten:

$$log\left(\lambda_{ij}\right) = \tilde{\psi}\left(U_{ij}\left(1\right) - U_{mj}\left(1\right)\right) \tag{3}$$

where $\tilde{\psi}$ is the reciprocal of discounted quality-adjusted life expectancy (QALE) of the reference individual:

$$\tilde{\psi} = \frac{1}{E\left[\sum_{a=j}^{J} \psi_{ma}\beta^{a-j}\phi\left(h_{ma}\right)\right]}.$$

Let u_{ia} denote flow utility unadjusted for health at age a given outcome bundles i: $\left[\bar{u} + log\left(c_{ia}\right) + \nu\left(l_{ia}\right)\right]$. Adding and subtracting the discounted sum of $E\left[\psi_{ia}\phi\left(h_{ia}\right)\right]E\left[u_{ia}\right]$ and $E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E\left[u_{ia}\right] - E\left[u_{ma}\right]\right)$ from the right-hand side of (3) yields the following additive decomposition of welfare:

$$log\left(\lambda_{ij}\right) = \tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left[\left(E\left[\psi_{ia}\phi\left(h_{ia}\right)\right] - E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\right)E\left[u_{ia}\right] + \Phi\right] \qquad QALE \tag{4}$$

$$+ \tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E\left[log\left(c_{ia}\right)\right] - E\left[log\left(c_{ma}\right)\right]\right) \qquad Cons. \tag{5}$$

$$+ \tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E\left[\nu\left(l_{ia}\right)\right] - E\left[\nu\left(l_{ma}\right)\right]\right). \qquad Leisure \tag{6}$$

where

$$\begin{aligned}\Phi = &\left(E\left[\psi_{ia}\phi\left(h_{ia}\right)u_{ia}\right] - E\left[\psi_{ia}\phi\left(h_{ia}\right)\right]E\left[u_{ia}\right]\right) \\ &- \left(E\left[\psi_{ma}\phi\left(h_{ma}\right)u_{ma}\right] - E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]E\left[u_{ma}\right]\right).\end{aligned}$$

The first term in (4) is the the difference in quality-adjusted life expectancy weighted by how much a healthy life year is worth—the expected flow utility from outcome bundles of individual i. The Φ term is an adjustment for the uncertainty of health and mortality over the life-cycle. Combined these provide the approximate contribution to welfare of health and mortality relative to the reference individual. The final two terms give the utility difference in consumption (5) and leisure (6) weighted by the quality-adjusted life expectancy of the reference individual. These provide the approximate contributions of consumption and leisure to welfare.

2.2 Life-cycle dynamics model

Our expected lifetime utility approach to welfare requires the knowledge of all possible life-cycle path realizations for an individual. As only the realized outcome path is observable in any longitudinal data set, we estimate a life-cycle dynamics model to approximate the joint evolutionary process of consumption, leisure, health, and mortality over time. In our application to the U.S. elderly population, we use the model to simulate counterfactual outcome paths for each individual.

We define leisure as time not spent in the labor market. As our empirical focus is on individuals nearing the end of working life, we limit labor considerations to the extensive margin by modeling an absorbing retirement decision (R).[4] Health at time t is defined by state vector $h_t = \{M_t, s_t\}$ where M is a vector of indicators for a number of absorbing morbidity conditions (e.g. ever diagnosed with diabetes, heart disease, etc.) and s is a measure of general health (e.g. self-rated health).

Probit regression models are used to estimate the probabilities of entering the absorbing retirement, morbidity, and death states. Standard linear dynamic panel models are used to estimate the life-cycle evolution of consumption and general health. Before we lay out our dynamic model in detail, some clarification on notation is in order. Included in all of our model equations are dummy indicators for age (μ), an equation specific linear time trend (t), and a 2008 indicator (P) to help control for the influence of the great recession on outcomes. In modeling absorbing states, we include a vector of observed time invariant individual characteristics (X). In the linear model for general health we include a time invariant individual unobserved "health endowment" (ν) and in the consumption model an unobserved "SES endowment" (π).[5] These endowments are modeled as fixed effects with no restriction on their correlation with other model regressors. While the unobserved endowments are identified from the panel structure of the linear models, we also include them as additional independent predictors in all absorbing state models.[6]

[4] An extended model may include the intensive margin, partial retirement, and/or reentry into the workforce but this comes with additional model complexity. Moreover, we find relatively small effects of leisure on welfare in our empirical analysis and retirement is likely to be the first-order leisure effect for this age group.

[5] The unobserved individual effect helps maintain the appropriate variance in health and consumption across time by effectively acting as a person specific drift in the auto-regressive processes.

[6] As many individuals never enter a given absorbing state in the data, it is not possible to estimate unique unobserved fixed effects for each individual for each absorbing state. As these estimates would be required for our simulations, we do not include additional unobserved fixed effects in absorbing

Conditional on being alive at time $t-1$, survival to the following period of life for individual i of age a is modeled as:

$$\psi_{iat} = I\left(\sum_{k=1}^{K_\psi} \gamma_k^\psi h_{i,t-k} + \mu_a^\psi + \delta^\psi t + \zeta^\psi P_t + \beta^\psi X_i + \alpha^\psi \nu_i + \kappa^\psi \pi_i + \epsilon_{iat}^\psi > 0\right) \tag{7}$$

where $I(.)$ is an indicator function, h a lagged health state vector, and ϵ^ψ an *iid* random shock with standard normal distribution. Coefficient vector γ_k^ψ allows the k^{th} lag of the health state vector to effect the current probability of survival. Including multiple lags allows the onset of morbidities to have differential effects over time. For example, the recent onset of diabetes may alter the probability of death more than if an individual has been living with diabetes for an extended period of time.

Following a similar logic, the probability of realizing a given morbidity state is allowed to depend on multiple lags of all other morbidities. Specifically, conditional on survival to time t, each morbidity condition $m_q \in M$ is modeled as:

$$m_{qiat} = I\left(\sum_{k=1}^{K_m} \gamma_{qk}^m M_{i,t-k}^{q'} + \mu_{qa}^m + \delta_q^m t + \zeta_q^m P_t + \beta_q^m X_i + \alpha^m \nu_i + \kappa^m \pi_i + \epsilon_{qiat}^m > 0\right) \tag{8}$$

where $I(.)$ is an indicator function, $M^{q'}$ a lagged vector of all $q' \neq q$ morbidities, and ϵ^m is a random shock. While ϵ^m is assumed independent across individuals and time, we allow an individual's shocks to be contemporaneously correlated across morbidity states (i.e. $cov\left(\epsilon_{qiat}^m, \epsilon_{q'iat}^m\right) \neq 0$). Contemporaneous morbidity shocks are assumed to follow a standard multivariate normal distribution with an $M \times M$ covariance matrix given by Σ.

Given the realizations of morbidity states at time t, the evolution of general health takes the following form:

$$s_{iat} = \sum_{k=1}^{K_s} \gamma_k^s s_{i,t-k} + \sum_{k=0}^{K_d} \gamma_k^{sm} M_{i,t-k} + \mu_a^s + \delta^s t + \zeta^s P_t + \nu_i + \epsilon_{iat}^s \tag{9}$$

where ν_i is the unobserved health endowment and ϵ^s an *iid* shock distributed $N\left(0, \sigma_s^2\right)$. The inclusion of lagged values of general health incorporates the persistence in general health shocks over the life-course. Morbidities are allowed to influence general health through lags as well as contemporaneously.

Turning to labor supply, conditional on working at time $t-1$, retirement the following period is modeled as:

$$R_{iat} = I\left(\sum_{k=1}^{K_r} \gamma_k^r h_{i,t-k} + \mu_a^r + \delta^r t + \zeta^r P_t + \beta^r X_i + \alpha^r \nu_i + \kappa^r \pi_i + \epsilon_{iat}^r > 0\right) \tag{10}$$

where ϵ^r is an *iid* shock drawn from a standard normal distribution. Similar to survival, lagged values of health (both general and specific morbidities) are allowed to influence

state models.

the probability of retirement. This is important given the evidence that health effects the retirement decision (Currie and Madrian, 1999).

Lastly, period t consumption depends on lagged consumption, retirement status, and contemporaneous and lagged values of health:

$$log\left(c_{iat}\right) = \sum_{k=1}^{K_c} \gamma_k^c log\left(c_{i,t-k}\right) + \sum_{k=0}^{K_{ch}} \gamma_k^{ch} h_{i,t-k} + \gamma_{cr} R_{it} + \mu_a^c + \delta^c t + \zeta^c P_t + \pi_i + \epsilon_{iat}^c \quad (11)$$

where π_i is the unobserved SES endowment and ϵ^c an iid shock distributed $N\left(0, \sigma_c^2\right)$. This specification is consistent with evidence that consumption declines with retirement (Hall, 2009). Allowing health to influence the evolution of consumption is also important given the evidence that health impacts economic outcomes, particularly at older ages (Smith, 1999). In contrast, the effects of economic status on health appear concentrated during childhood and young adulthood when health trajectories are being established (Smith, 1999).

3 Data and methods

Equipped with our theoretical framework, our empirical analysis involves three broad steps.

1. We use data from the HRS to estimate the parameters of the life-cycle dynamics model. Here we use data on all individuals aged fifty and older from all available waves of the HRS from 1992-2014.

2. Using the parameter estimates and age sixty data as initial conditions, we repeatedly simulate remaining life-cycle shocks to mortality, health, consumption, and leisure for a sub-sample of the HRS respondents. This simulation sample includes all individuals with age sixty data and requisite lagged data for simulations.

3. We embed the simulated data within our expected utility framework to construct a measure of ex-ante welfare at age sixty for each individual in our simulation sample.

Our choice of age sixty for welfare comparisons is primarily driven by empirical considerations. It provides a large enough sub-sample for analysis across three broad birth cohorts included in the HRS after accounting for the sampling design of the survey and lagged data requirements of our dynamic model. We compute the distribution of elderly welfare within a birth cohort as well as compare welfare distributions across cohorts to examine how it has changed over time. The remainder of this section details the data used in our analysis, estimation and simulation of the dynamic model, and calibration of parameters used in the welfare model.

3.1 Data

The HRS is a longitudinal panel study surveying individuals in the U.S. over the age of fifty and their spouses on a biennial basis. The study consists of five primary birth cohorts—the initial HRS cohort (born 1931-1941), AHEAD cohort (born before 1924), Children of Depression (born 1924-1930), War Babies (born 1942-1947), and Baby Boomers (born 1948-1959).[7] The core survey was conducted on alternate years starting from 1992 for the initial HRS cohort with the other cohorts added over subsequent waves of the survey. As such, a model period corresponds to be two calendar years and individuals are grouped in two-year age intervals. To be clear, our welfare measure is constructed for each individual at age sixty *or* sixty-one, though our model makes no distinction between the two and for brevity we refer to this as welfare at age sixty.

We use the cleaned RAND HRS data file (v.P), available through the HRS website, to obtain data on health, work, and other individual characteristics from 1992 to 2014.[8] We define retired individuals as those reporting less than 500 annual hours of work in the current or any previous survey wave.[9] As fixed characteristics (X_i) in our morbidity, mortality, and retirement models we use indicators for gender, education level, race, census division, census occupation code, and birth cohort.

Our primary morbidity measures include eight binary indicators for ever having been diagnosed by a doctor with the following health problems—(1) high blood pressure or hypertension; (2) diabetes or high blood sugar; (3) cancer or a malignant tumor of any kind except skin cancer; (4) chronic lung disease except asthma such as chronic bronchitis or emphysema; (5) heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems; (6) stroke or transient ischemic attack (TIA); (7) emotional, nervous, or psychiatric problems; and (8) arthritis or rheumatism. As a final measure of morbidity, we include an indicator for ever reported difficulty with any activity of daily living (ADL). Difficulty with ADLs are a commonly used morbidity metric among the elderly and include activities such as walking across the room, bathing, and getting dressed. Finally, as our general health measure (s) we use self-rated health status reported on a five-point scale from poor (one) to excellent (five). Self-rated health has been shown to be predictive of mortality, even after controlling for other health conditions and socioeconomic characteristics (Idler and Benyamini, 1997).

Consumption data comes from the Consumption and Activities Mail Survey (CAMS). From 2001, the CAMS was sent to a random sub-sample of HRS participants during off years of the core survey. We use the cleaned RAND CAMS data file (v.D2) containing annual consumption data from 2001-2011. An estimate of total household consump-

[7]Baby Boomers are split into two groups by the HRS (early and mid) but we group them together as very few mid Baby Boomers have the lagged data required for estimation of the dynamic model or simulations.

[8]Data available at `http://hrsonline.isr.umich.edu`.

[9] We combine data on weekly hours worked and weeks worked per year to estimate annual hours worked.

tion is provided inclusive of durables, non-durables, housing, and transportation. We subtract out-of-pocket health spending to create an adjusted measure of household consumption.[10] We use this adjusted household consumption divided by the number of household members as our individual consumption measure.[11] As consumption data is only available in between the core HRS waves, we merge each CAMS wave with the HRS core data from the previous wave.[12] One of the biggest challenges to our analysis is that consumption data is missing for all waves before 2000 and after 2010 and for about 80% of the HRS sample from 2000-2010. However, closely related data is available across all survey waves including detailed measures of wealth and income. We use this additional information to address the missing consumption data issue by using the multiple imputation method proposed by Honaker and King (2010) for cross-sectional time-series data (see appendix A for details).

The pooled sample used to estimate the dynamic model includes all individuals born prior to 1960 and aged fifty and over at the time of the survey. This consists of 35,889 unique individuals and 216,626 total individual-year observations. Table 1 shows descriptive statistics for modeled outcomes for each cohort in the HRS (additional descriptives shown in appendix Table 12). Incidence of each morbidity state was substantial among respondents, allowing for relatively precise estimates of their effects on dynamic processes. However, there was still significant variation in incidence rates across morbidity states. For example, in the HRS cohort, over 50% of observations reported arthritis while only 6% reported having suffered a stroke. In terms labor supply, share of retired individuals ranged from 26% in the most recent Baby Boomer cohort to 94% in the (much older) AHEAD cohort. Annual real consumption averaged between $23-$29,000 across cohorts.

Modeling the evolution of outcomes in our dynamic framework allows us to disentangle the age and cohort effects present in the observed data. This enables us to isolate welfare differences across individuals and over time. The dependency structure of our life-cycle dynamics model is motivated by the observed correlations in consumption, health, and labor supply found in the data (see Figure 1). There are positive associations of varying strength across morbidities highlighting the importance of modeling their evolution jointly. Morbidity states have a strong negative correlation with self-rated health and a more modest positive association with retirement. Consumption is positively associated with self-rated health and negatively associated with retirement and all morbidities except cancer. Cancer is the clear outlier with the weakest co-morbidity correlations and a small positive correlation with consumption.

The positive relationship between self-rated health and consumption—and negative relationship with morbidities—suggests using consumption as the sole basis of a well-being metric could understate the inequality among the elderly. On the other hand, the negative association between consumption and retirement suggests a possible over-

[10]Health spending includes health insurance, medication, health services, and medial supplies.

[11]We use the CPI-U to convert all waves to 2010 dollars.

[12]This is the recommended procedure for use of the RAND CAMS data file and is also consistent with the time structure of our dynamic model.

Table 1: Estimation sample descriptive statistics by cohort

	AHEAD	CODA	HRS	WB	BB
Individuals	7,758	4,231	10,498	3,615	9,787
Observations	36,896	27,522	88,450	26,805	36,953
Age (mean)	81.59	75.02	64.44	59.63	56.08
Hypertension (%)	54.31	57.01	49.28	45.97	42.94
Diabetes (%)	14.74	18.14	16.42	16.65	16.07
Cancer (%)	17.38	17.39	12.14	10.46	7.38
Lung disease (%)	9.79	10.40	8.85	7.03	6.62
Heart disease (%)	35.92	30.01	20.47	15.85	12.75
Stroke (%)	15.21	12.00	6.02	5.11	3.43
Psyche problem (%)	11.47	11.86	11.66	16.52	18.85
Arthritis (%)	55.42	60.64	53.62	48.88	41.29
Difficulty with ADLs (%)	39.34	28.25	20.38	20.51	16.51
Self-rated health (mean)	2.87	3.06	3.26	3.35	3.33
Retired (%)	93.90	88.45	64.09	48.05	31.58
Annual consumption ($1000s, mean)	23.52	26.02	28.21	29.30	27.15

Notes: Mean and percentage estimates use base year sampling weights. Children of the Depression denoted by CODA, War Babies by WB, and Baby Boomers by BB. Consumption is reported in real 2010 dollars.

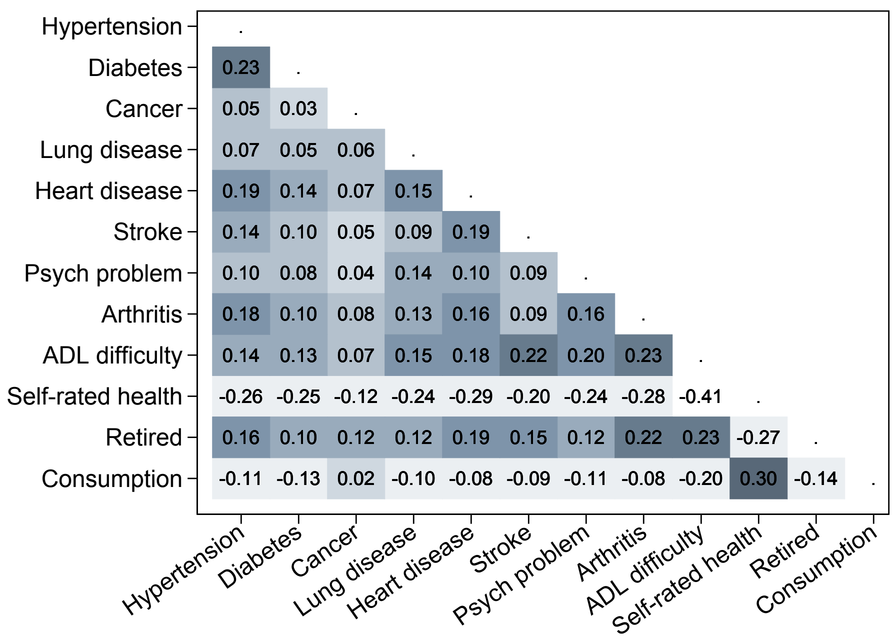

Figure 1: Outcome correlations

statement of welfare inequality as those with low consumption may enjoy more leisure. Our welfare model allows us to gauge the relative strength of these channels.

3.2 Estimation and simulation of life-cycle dynamics

The dependency structure in models (7)-(11) does not introduce simultaneity into the system allowing for equation-by-equation estimation for all models with *iid* shocks. Self-rated health is treated as a continuous measure allowing models for general health (9) and consumption (11) to follow the structure of a standard linear dynamic panel data model with lagged dependent variables and individual level fixed effects. It is well known that OLS estimates of such models suffer from Nickell (1981) bias. As a bias correction, we use the bootstrap-based method proposed by Everaert and Pozzi (2007).[13] Our modeling approach assumes shocks to be serially uncorrelated. Allowing for a single period lag (two calendar years) of health on consumption ($K_{ch} = 1$) and two lags (four years) for all other variables ($K_c = K_s = K_d = 2$) is sufficient to meet this requirement.[14]

Turning to absorbing states, models for mortality (7) and retirement (10) are estimated independently using standard probit regressions. The evolution of morbidities described by equation (8) follows a multivariate probit structure with correlated shocks. Given our large number of observations and binary morbidity outcomes, we estimate the multivariate probit model via the computationally feasible method of using a chain of bivariate probit estimators proposed by Mullahy (2016). Finally, we use two lags of independent variables in the absorbing state models for consistency with the linear models for health and consumption ($K_\psi = K_m = K_r = 2$).

3.2.1 Estimation results

Select estimation results are provided in Figure 2 while the full set of results are shown in Tables 13-15 in appendix B. The first two columns in Figure 2 provide the estimated contemporaneous association between morbidity states, self-rated health, and consumption. The final two columns provide the association between current health states and the odds of retirement/mortality the following period.

Outcomes evolve through the system in interdependent ways. Take the case of heart disease and consumption. Heart disease has a small contemporaneous direct association with log consumption with a point estimate of 0.007. However, heart disease is associated with a decline in self-rated health of 0.324. A unit decline in self-rated health in turn lowers consumption by 0.039 log points. The net contemporaneous impact of the onset of heart disease is thus a decline in consumption of about 2% ($0.007 + 0.324 \times 0.039$). Moreover, these relationships continue to evolve dynamically throughout the system over time (see impulse response Figure 11 in appendix B which demonstrates the dynamic relationships for heart disease).

As with heart disease, each morbidity has a strong negative association with self-rated health. In turn, lower self-rated health is associated with a significant increase

[13]We implement the bootstrap with De Vos et al. (2015) Stata routine *xtbcfe*.

[14]First order autocorrelation was tested for consumption and general health using the approach of Born and Breitung (2016) and implemented in Stata with Wursten et al. (2016).

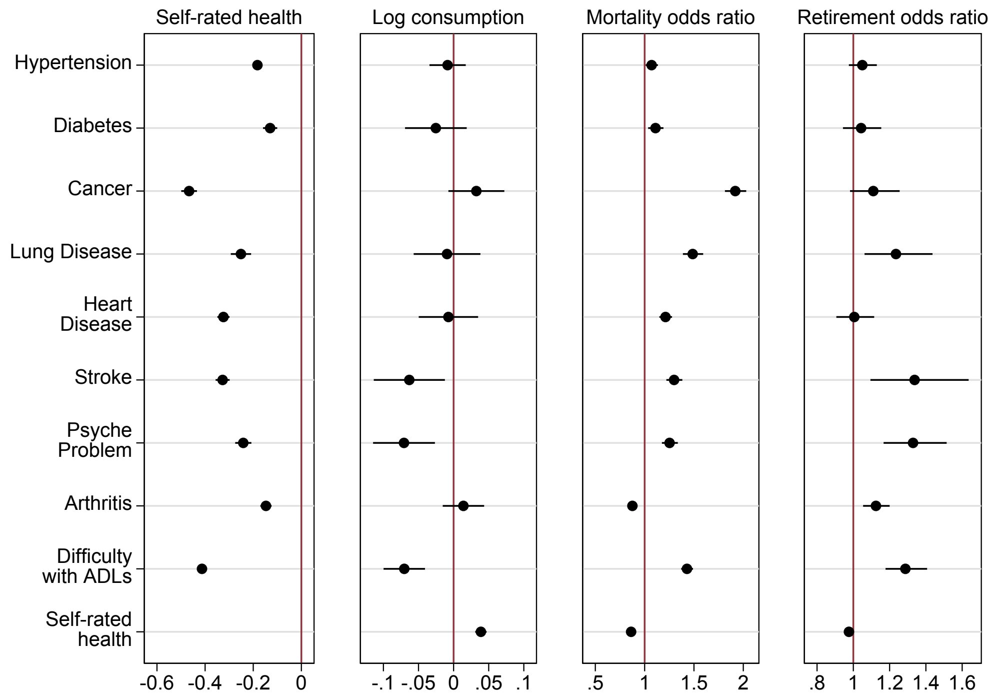

Figure 2: Select estimation results for self-rated health, consumption, retirement, and mortality models

Notes: Dependent variables across columns. Contemporaneous associations reported for self-rated health and consumption. Lagged associations reported for retirement and mortality. Spikes indicate 95% confidence intervals.

in the probability of death and retirement and a decrease in contemporaneous consumption. Independent of self-rated health, individual morbidity states have diverse associations with other outcomes. Arthritis, for example, is the only morbidity not directly associated with mortality, though it does have a positive independent relationship with the probability of retirement. In contrast, a recent stroke is associated with an increased probability of death, retirement, and a loss of consumption independently of the effect through self-rated health.

3.2.2 Simulations

We use the estimated dynamic models to construct expected remaining lifetime utility for each sixty year old in our sample. Specifically, using age sixty data as initial conditions, we simulate the remaining life outcomes and associated utility 5,000 times for each individual and average across simulations to obtain a measure of expected lifetime utility.[15] Note that as the HRS began in 1992, age sixty data is not available for the older AHEAD or CODA cohorts so our simulations are limited to the initial

[15]Initial conditions also include unobserved endowments ν and π estimated from models (9) and (11) using the prediction method of De Vos et al. (2015).

HRS, War Babies, and Baby Boomers cohorts. Moreover, the HRS is structured such that individuals were initially surveyed at various ages implying age sixty data is not available for all individuals within the younger three cohorts.

Limiting our sample to those with the requisite data leaves a simulation sample of 6,544 individuals from the initial HRS cohort (reaching age sixty between 1992-2001), 2,547 War Babies (2002-2007), and 3,437 Baby Boomers (2008-2014).[16] Our aggregate welfare statistics are estimated separately for each of the three cohorts to analyze the change in inequality over time (we also check robustness of results to fixed two-year birth cohorts). Table 2 provides a summary of initial conditions in the simulation sample. By most measures, there was an average decline in age sixty health over cohorts as well as a fall in the share retired. Average age sixty consumption increased for War Babies but declined for Baby Boomers, presumably due to the timing of the great recession which hit when Baby Boomers were in their late fifties.

Table 2: Simulation sample initial conditions by cohort

	HRS	WB	BB
Individuals	6,544	2,547	3,437
Age (mean)	60.00	60.00	60.00
Hypertension (%)	40.14	46.71	50.93
Diabetes (%)	11.93	16.75	20.01
Cancer (%)	7.63	10.97	9.59
Lung disease (%)	6.87	7.51	8.29
Heart disease (%)	14.51	15.65	16.06
Stroke (%)	3.53	5.25	4.37
Psyche problem (%)	9.79	16.95	21.89
Arthritis (%)	46.48	51.38	52.42
Difficulty with ADLs (%)	15.37	22.29	22.29
Self-rated health (mean)	3.34	3.31	3.28
Retired (%)	47.17	46.87	42.07
Annual consumption ($1000s, mean)	29.10	30.10	27.56

Notes: Mean and percentage estimates use base year sampling weights. War Babies denoted by WB and Baby Boomers by BB. Consumption is reported in real 2010 dollars.

A comparison between the mean and standard deviations of the simulated life-cycle profiles and those based on available data is shown by cohort in Figures 12-16 in appendix B. Overall, the simulations match the available aggregated data well suggesting our life-cycle dynamics model provides a reasonable approximation to the underlying data generating processes.

[16] Those born after 1953 do not have the requisite data for simulations, leaving the Baby Boomer simulation cohort as those born 1948-1953.

3.3 Calibration of welfare model

Analysis using our welfare model requires calibration of preference parameters. These include parameters of the functions $\phi(h)$ and $\nu(l)$ mapping health states and leisure into flow utility. Additional parameters include the discount rate β and flow intercept $\bar{u}$. Here we detail our calibration strategy and estimates.

We assume health utility depends linearly on our health state vector: $\phi(h_t) = \gamma h_t$. However, we bound $\phi(h_t) \in [0,1]$ to be consistent with our QALY framework. We use the Health Utilities Index Mark 3 (HUI3) instrument as the conceptual basis of our health utility function (see Horsman et al. (2003) for a detailed discussion on the HUI3). This choice is motivated by two main features of the HUI3. First, it provides a comprehensive description of health status that has been shown to be responsive to changes in health over time (Barr et al., 1997; Furlong et al., 2001; Blanchard et al., 2003). Second, it provides a direct estimate of QALYs which is our preferred measure of health in this analysis.

The HUI3 questionnaire was included as a module for a subset of approximately 1,200 of the participants in the 2000 wave of the HRS with the associated utility scores available through the HRS website.[17] We use the HUI multi-attribute utility score (*hui3ou*) in our analysis (see Furlong et al. (1998); Feeny et al. (2002) for details on construction). It is a health related quality of life measure where death is defined by a score of zero and perfect health by a score of one.[18]

The vector of utility weights γ is estimated by regressing the HUI3 utility score on self-rated health and all morbidity indicators. Table 3 provides the linear regression results.[19] Self-rated health has a strong and highly significant positive association with utility. The estimated weight implies a one unit increase in self-rated health improves flow utility by 9.0 percentage points. Conditions such as hypertension, diabetes, and cancer have little independent effect on health utility after controlling for their association with self-rated health and other co-morbidities. Other morbidities such as stroke and arthritis have larger (and statistically significant) independent negative effects. The most influential morbidity indicator is difficulty with ADLs, which lowers utility an estimated 17.9 percentage points.

Leisure is normalized to one for retired individuals. We assume workers supply 2,000 annual hours to the labor market and set associated leisure to $0.66 = 1 - (2000/5,840)$ where 5,840 = 16 hours a day × 365 days. Preferences over leisure are defined by $\nu(l) = -\frac{\theta\epsilon}{1+\epsilon}(1-l)^{\frac{1+\epsilon}{\epsilon}}$, where ϵ is a constant Frisch elasticity of labor supply (the elasticity of labor supply with respect to wage, holding the marginal utility of consumption fixed). Empirical studies of the Frisch elasticity vary considerably, with estimates ranging from 0.5 to nearly 2 (Chetty, 2012; Hall, 2009). We follow Jones and Klenow (2016) and choose a benchmark value of $\epsilon = 1$ while testing the sensitivity of our results to alternate values. Likewise, we choose a discount factor $\beta = 0.98$ corresponding to an

[17] Researcher contribution file HUI3 (v.1.0).

[18] Negative scores are possible and represent health states that are worse than death.

[19] Results are insensitive to use of a Tobit regression.

Table 3: Estimated health utility weights

Measure	Weight	SE
Self-rated health	0.090	0.007
Hypertension	0.010	0.013
Diabetes	-0.005	0.018
Cancer	0.013	0.018
Lung disease	-0.034	0.023
Heart disease	-0.031	0.015
Stroke	-0.088	0.023
Psych problem	-0.085	0.021
Arthritis	-0.059	0.013
Diff with ADL	-0.179	0.017
Constant	0.543	0.028

Notes: Results from regression of HUI3 score on self-rated health and morbidities. SE denotes standard error. $R^2 = 0.48$. N = 1,089.

annual discounting of one percent. Note that there is additional implicit discounting in the expected utility framework due to mortality.

The standard first-order condition for the labor-leisure decision equates the marginal cost and benefit of leisure: $u_l = w\,(1-\tau)\,u_c$, where w is the wage rate and τ is the marginal tax rate. Given our function forms, the implied disutility weight on labor supply $\theta = w\,(1-\tau)\,(1-l)^{-1/\epsilon}\,/c$. Using earnings and hours worked data from the HRS and a marginal tax rate of 0.38 from 2002 (Barro and Redlick, 2011), we calculate the implied θ for each working sixty year old in the sample. Selecting the median of these values yields our benchmark $\theta = 7.56$.[20]

Finally, we set the intercept in flow utility $\bar{u}$ so that the median value of remaining life for sixty year olds in our simulation sample is \$50,000 per QALY.[21] In a review of the literature, Ryen and Svensson (2015) estimate mean and median values across studies of approximately \$98,000 and \$32,000. Traditional values in the U.S. often range from \$50,000 to \$100,000 (Kaplan and Bush, 1982). Using \$50,000 as our benchmark and normalizing consumption to thousands of 2010 dollars gives $\bar{u} = -0.385$.[22]

4 Welfare results

Our simulation sample for welfare analysis includes three birth cohorts—the initial HRS cohort, War Babies, and Baby Boomers. We use the initial HRS cohort as our

[20] As noted by Jones and Klenow (2016), this calibration strategy implicitly invokes wedges (i.e. labor market frictions) to explain individual deviations from the static first-order condition.

[21] The value of life per QALY at age j is given by $VOL_j/E\left[\sum_{a=j}^{J}\psi_a\beta^{a-j}\phi\,(h_a)\right]$ where $VOL_j = U_{ij}\,(1)\,c_j$.

[22] It is possible for an individual to obtain negative expected remaining lifetime utility in this framework but this occurs for less than 0.4% of our sample.

benchmark group as it is the earliest of the three and contains the longest panel of available data. The reference person for all welfare calculations is the individual with the age sixty initial conditions that yield the median expected lifetime utility within the HRS cohort. The same reference person is used to calculate welfare for the later War Babies and Baby Boomers cohorts. This approach allows for direct comparison of welfare across cohorts as the reference person is held fixed. At the end of this section, we check robustness of welfare estimates to the choice of reference individual as well as other modeling assumptions.

4.1 Elderly welfare inequality

We begin by examining the distribution of our consumption-equivalent measure of welfare across the sample of sixty year olds from the initial HRS cohort. The first row in Table 4 provides different summary measures of welfare inequality in our fully specified "benchmark" model. In order to assess the importance of morbidity on our welfare calculations, we also provide measures where we exclude the utility cost of less than perfect health from preferences (i.e. $\phi(h) = 1 \, \forall h$). The results from this model are labeled as "no health" in the table.

Table 4: Summary measures of welfare inequality at age sixty for initial HRS cohort

Measure	Gini	10/50 ratio	90/50 ratio	ρ
Benchmark λ	0.667	0.226	5.215	-
No health λ ($\phi(h) = 1$)	0.515	0.319	3.172	0.966

Notes: Estimates use base year sampling weights. No health measure removes health from flow utility. Spearman's rank correlation between the two welfare measures denoted by ρ.

There is substantial variation in welfare across individuals—the benchmark Gini coefficient is 0.667. Moreover, welfare at the tenth percentile of the distribution is only 22.6% of the median welfare while that of the ninetieth percentile is over five folds higher than the median. Ignoring the utility costs of poor health largely preserves the rank ordering of welfare ($\rho = 0.966$) but significantly under-estimates the inequality—the Gini coefficient is under-estimated by about 23%. Moreover, this morbidity bias occurs at both ends of the distribution. For example, relative to our benchmark measure, welfare at the tenth percentile increases to 31.9% of the median while that at the ninetieth falls to just over three fold.

The difference in welfare measures between the two models suggests substantial and varied individual utility costs of living with poor health and morbidities among the elderly. Figure 3 plots remaining life expectancy at age sixty against the ratio of QALE to life expectancy for each individual in the initial HRS cohort.[23] The positive correlation implies those with higher life expectancy also expect better health over remaining life. For example, those with a remaining life expectancy of 29 years have

[23]Life expectancy at age j defined by $E\left[\sum_{a=j}^{J} \psi_a\right]$ and QALE as $E\left[\sum_{a=j}^{J} \psi_a \phi(h_a)\right]$.

a quality-adjusted life expectancy of about 25 *healthy* life years—or about 85%. In contrast, those at the bottom end of the distribution expect greater utility losses from poor health (with considerably more variability). For example, those with a remaining life expectancy of 10 years expect anywhere from about 2 to 6 quality-adjusted life years.

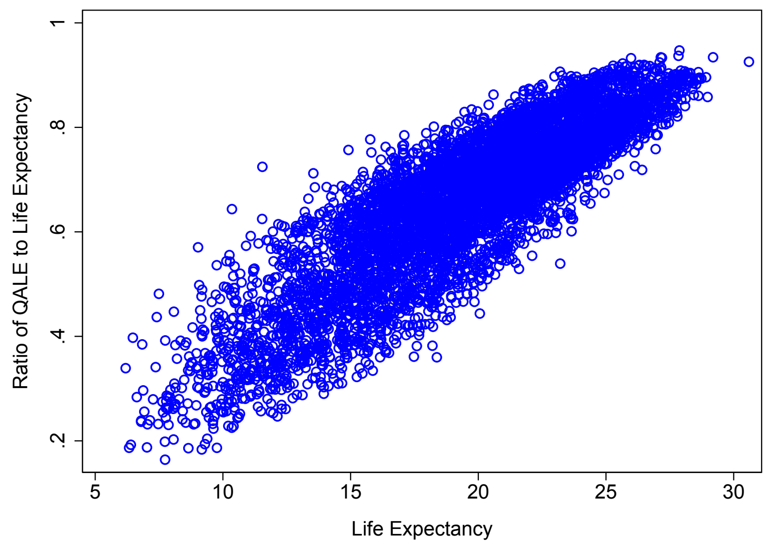

Figure 3: Life expectancy and quality-adjusted life expectancy (QALE) at age sixty

We turn now to the welfare impact of QALE relative to consumption and leisure. Table 5 shows welfare and its decomposition averaged within each decile of the welfare distribution. Consumption and QALE are the largest source of welfare loss for those in the bottom of the distribution. Low consumption and quality-adjusted life expectancy costs the bottom decile an average of 69.7 and 113.7 log points in welfare relative to the median individual. Higher leisure adds a comparatively modest 1.6 log points to welfare. In contrast, the top of the distribution experiences the highest consumption and quality-adjusted life expectancy, though marginally less leisure.

Recall that welfare for each individual is estimated using potential outcome bundles of a reference person. While the reference individual has the median expected lifetime utility by definition, their expected levels of consumption, leisure, and quality-adjusted life expectancy could be somewhat arbitrary. Comparing average log point gaps across deciles provides a more robust examination of the strength of the relative components across the welfare distribution. For example, the log point gap in consumption between the highest and lowest decile is $174.2 = 104.5 + 69.7$. The analogous gaps for leisure and QALE are -7.6 and 248.4. Overall, these gaps suggest the strongest driver of welfare inequality are differences in health and mortality followed by consumption differences.

In order to see the pattern of health and consumption driving the welfare gaps, Figure 4 plots quality-adjusted life expectancy against annual consumption at age sixty

Table 5: Mean welfare by decile of distribution

			Decomposition		
Welfare Decile	Mean λ	Mean log λ	Consumption	Leisure	QALE
Lowest	0.170	-1.818	-0.697	0.016	-1.137
2nd	0.285	-1.269	-0.295	-0.005	-0.968
3rd	0.432	-0.847	-0.125	-0.019	-0.703
4th	0.619	-0.485	0.009	-0.036	-0.458
5th	0.859	-0.156	0.132	-0.043	-0.245
6th	1.174	0.156	0.236	-0.044	-0.037
7th	1.631	0.483	0.331	-0.046	0.198
8th	2.401	0.866	0.465	-0.052	0.452
9th	3.874	1.334	0.641	-0.056	0.750
Highest	14.687	2.332	1.045	-0.060	1.347

Notes: Estimates use base year sampling weights.

for individuals in the top and bottom deciles of the welfare distribution. A majority of sixty year olds in the highest decile had a QALE of over 15 years. There was more substantial variation in annual consumption in the group with values ranging from around \$20,000 to more than \$100,000. This suggests health as the major driver of welfare at the top end of the distribution. A majority of individuals in the lowest decile had annual consumption under \$30,000 and a QALE of less than 10 years. However, some had relatively higher QALE but low welfare due to very low consumption.

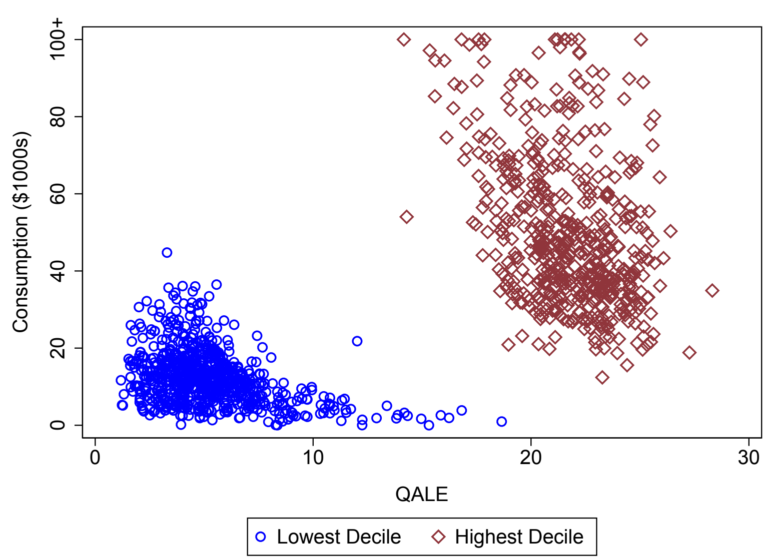

Figure 4: Age sixty consumption and QALE by decile of welfare distribution

Finally, Figure 5 plots average expected life-cycle profiles for select welfare deciles to gain a sense of the differences across individuals. Consumption, leisure, and health gaps are largest at age sixty and gradually decline as individuals age. However, substantial gaps remain for consumption and health even among individuals who survive into their late nineties. Over the entire remaining life-cycle, the gaps in consumption are relatively small between the first and fifth deciles compared to the much higher average consumption in the top decile. In contrast, the health and leisure gaps are substantially larger between the bottom and middle deciles with a smaller difference between the middle and the top.

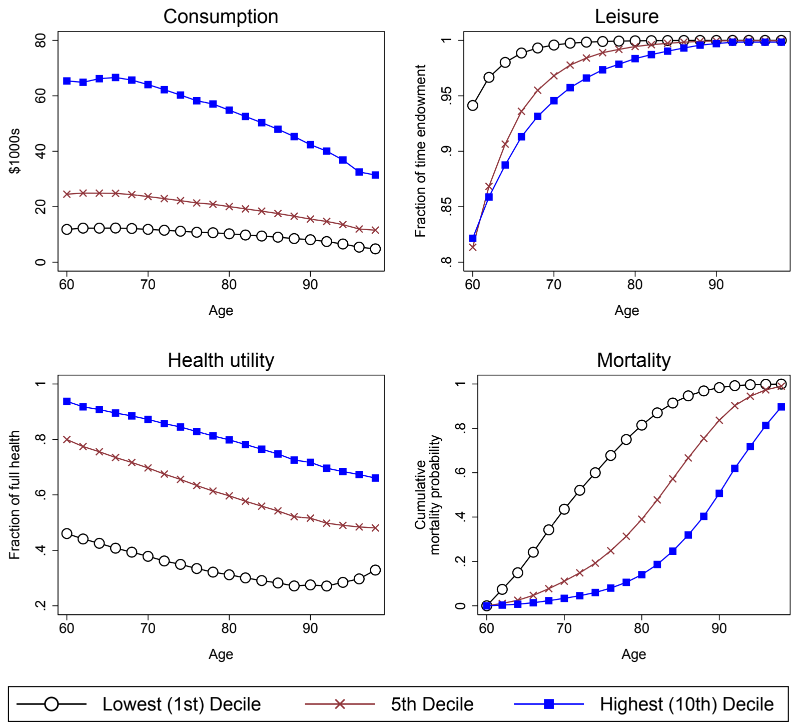

Figure 5: Average life-cycle profiles by select deciles of welfare distribution

Notes: Consumption, leisure, and health profiles are expected values conditional on survival.

4.2 Welfare over cohorts

We next examine how welfare has changed across time by comparing individuals in the initial HRS cohort (reaching age sixty between 1992-2001) to War Babies (2002-2007) and Baby Boomers (2008-2014). Note that aggregate welfare differences across cohorts stem from two sources—the distribution of age sixty initial conditions, and a time trend and cohort specific intercept in all modeled outcome processes. Table 6 provides summary welfare measures for each cohort/time period.

Comparing War Babies to the HRS cohort, welfare decreased 2.4% at the median but increased 27.5% on average, suggesting an unequal shift in the welfare distribution. This pattern was maintained but stronger for Baby Boomers with median welfare

Table 6: Summary welfare measures by cohort

Cohort	Median λ	Mean λ	Gini	10/50 ratio	90/50 ratio
HRS	1.000	2.612	0.667	0.226	5.215
War Babies	0.976	3.331	0.726	0.216	6.368
Baby Boomers	0.928	5.598	0.813	0.207	7.563

Notes: Estimates use base year sampling weights.

declining 4.9% while the mean increased 68.1% over War Babies. Turning to the disparity measures provides further evidence of a sustained increase in welfare inequality over time. Relative to the HRS cohort, the welfare Gini rose 8.8% for War Babies and 21.9% for Baby Boomers.[24] At the bottom end of the distribution, the welfare of the tenth percentile declined from 22.6% of the median to 21.6% among War Babies and 20.7% among Baby Boomers. The welfare at the top also pulled further away from the center—the 90/50 ratio increased by 45.0% between the HRS and Baby Boomer cohorts.

Table 7 provides the decomposition of mean log welfare for each of the three cohorts. The gain in mean welfare for War Babies over the HRS cohort was driven by an average increase of 1.3 log points from consumption and 1.1 log points from quality-adjusted life expectancy. There was also a 0.3 log point decline in average welfare due to delayed retirement and consequently less leisure. Comparing Baby Boomers to War Babies, the decomposition reveals that while health and longevity continued to improve—average QALE contribution to welfare rose 11.1 log points—consumption declines cost average welfare 11.3 log points. The consumption decline is presumably driven by the timing of the 2008 recession which hit when Baby Boomers were in their late fifties (see Figure 18 in appendix B).

Table 7: Welfare decomposition by cohort

		Decomposition		
Cohort	Mean log λ	Consumption	Leisure	QALE
HRS	0.059	0.174	-0.034	-0.080
War Babies	0.081	0.187	-0.037	-0.069
Baby Boomers	0.066	0.074	-0.050	0.042

Notes: Estimates use base year sampling weights.

Figure 6 provides the distribution of log welfare, expected remaining lifetime consumption, life expectancy, and QALE at age sixty across cohorts for closer examination. The welfare distribution became flatter and more skewed over time demonstrating the rise in welfare inequality. Compared to the HRS cohort, welfare improved for the top end of the War Babies distribution but declined somewhat for the bottom end. Simi-

[24]The pattern of increased inequality over time also holds when examining two-year birth cohorts (see appendix Figure 17).

larly, the distribution of welfare for Baby Boomers increased slightly over War Babies at the top but a fatter left tail implies a more substantial welfare decline for the bottom end. Expected remaining lifetime consumption follows a similar pattern as the welfare distribution with initial gains concentrated at the top end for War Babies followed by substantial declines concentrated in the bottom end for Baby Boomers.

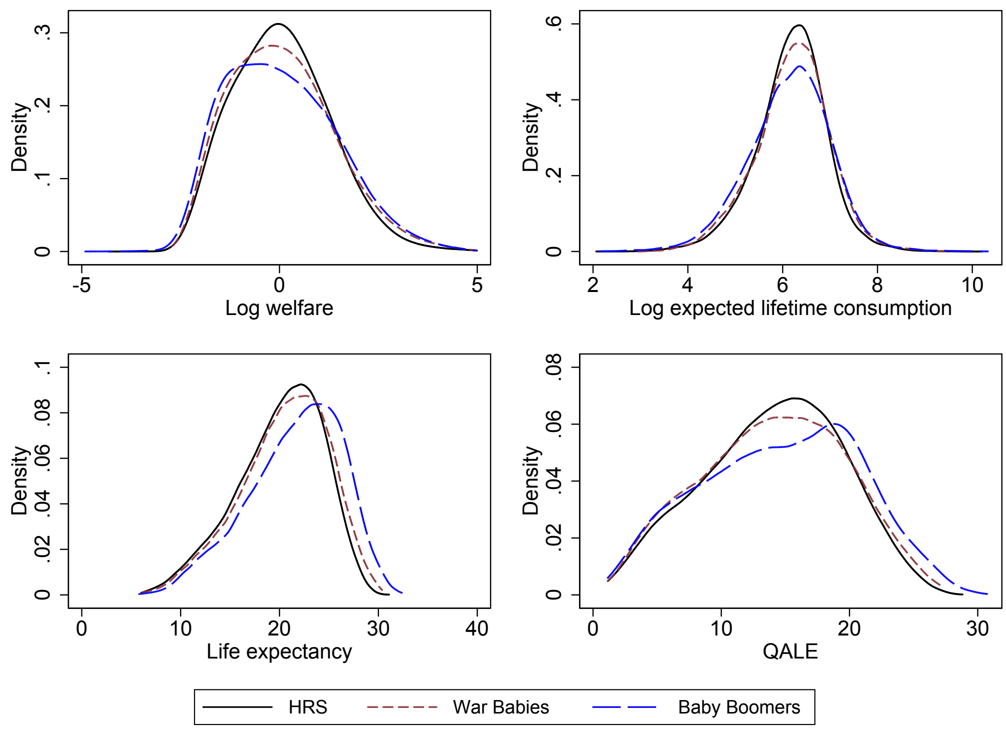

Figure 6: Distribution of welfare, consumption, and life expectancy over cohorts

While life expectancy has shown broad improvements over time, the distribution has increased in skewness for War Babies and Baby Boomers. This implies mortality gains have disproportionately benefited those in the top of the distribution. This is consistent with the existing evidence of increasing socioeconomic gradients in mortality. When adjusting life expectancy for quality of health, the distribution becomes more disperse. The fattening left tail in QALE relative to previous cohorts shows that life expectancy gains are even outweighed by health losses at the bottom end of the distribution. Overall, cohort results demonstrate an increase in welfare inequality driven primarily by a combination of increasing gaps in consumption and quality-adjusted life expectancy.

4.3 Comparison with other measures of well-being

Our welfare measure incorporates inequality of various components of well-being into a single metric. Moreover, our measure captures the static welfare effect of each component at age sixty as well as their expected joint dynamic influence throughout remaining life. As a comparison, Table 8 provides inequality statistics across alternate measures of well-being for the age sixty population in the initial HRS cohort. The final column provides Spearman's rank correlation coefficient between our welfare measure and each alternative measure.

Table 8: Comparing measures of inequality at age sixty

Measure	Gini	10/50 ratio	90/50 ratio	ρ
Welfare (λ)	0.667	0.226	5.215	-
Income	0.519	0.228	2.971	0.508
Consumption	0.435	0.353	2.649	0.568
Health utility	0.121	0.610	1.221	0.772
Flow utility	0.465	0.347	3.022	0.789

Notes: Estimates for initial HRS cohort using base year sampling weights. Income, consumption, and health utility are cross-sectional measures at age sixty. Flow utility is calculated using cross-sectional consumption, leisure, and health along with our benchmark preferences. Spearman's rank correlation between λ and each measure denoted by ρ.

Cross-sectional income inequality is lower than welfare inequality, though income does well predicting welfare at the bottom end of the distribution—the 10/50 ratio is similar in the two measures. However, the rank correlation of 0.50 between welfare and income is quite low. So while income may provide a reasonable measure of aggregate inequality, relative income and welfare can be quite different at the individual level. Cross-sectional consumption at age sixty provides a somewhat better ranking of individual welfare, but under-estimates welfare inequality substantially more than income. The improved rank correlation is perhaps unsurprising as consumption directly enters preferences used in our welfare model. Age sixty health alone severely under-estimates aggregate welfare inequality, although it provides a better individual ranking than income or consumption. This speaks to the substantial influence of health and mortality in determining the distribution of our welfare measure.

The final row of Table 8 provides an estimate of welfare incorporating age sixty consumption, leisure, and health into our benchmark flow utility specification (2) but ignoring the subsequent life-cycle dynamics. Incorporating all three components in a static framework provides the best ranking of individual welfare, but ignoring the dynamics substantially under-estimates welfare inequality. For example, the Gini of age sixty flow utility is only 70% of that based on our dynamic welfare measure.

Figure 7 further illustrates the nuanced relationship between health and economic outcomes by plotting quality-adjusted life expectancy against the ratio of consumption to welfare at age sixty. There is a clear negative correlation between the two measures with consumption over-predicting welfare for those of poor health and under-predicting for those of good health. This pattern is consistent with the positive correlation between consumption and health found in the raw data. However, there is also substantial variation in the plotted relationship explaining the relatively modest rank correlation between welfare and consumption. A key takeaway here is even in cases where economic outcomes provide a reasonable approximation to aggregate welfare inequality, they may still fail to be an adequate welfare measure at the individual level.

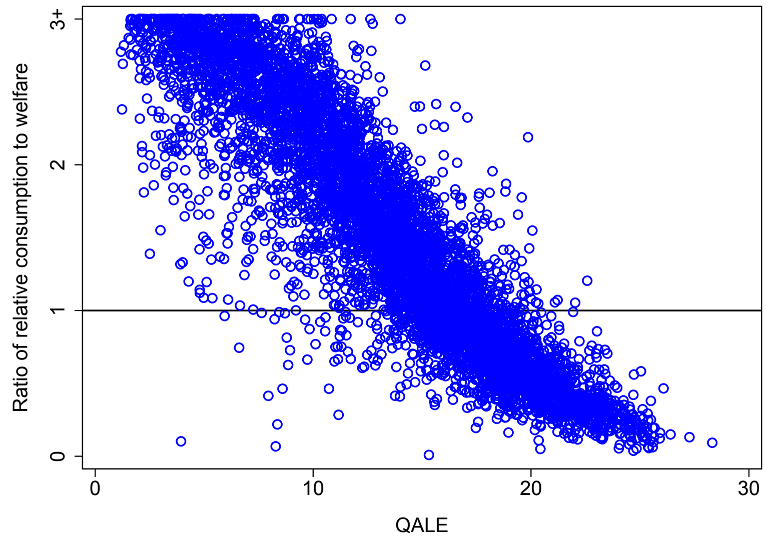

Figure 7: Variation in relationship between age sixty consumption, welfare, and QALE

Notes: Ratio of relative consumption to welfare given by $(c_i/c_m)/\lambda$ where c is age sixty consumption. Plot includes HRS cohort only.

Finally, Table 9 provides Gini coefficients by cohort for age sixty distributions of our welfare and other measures. Welfare inequality has grown significantly more than age sixty cross-sectional income, consumption, or health inequality. This implies disparity measures based on economic outcomes such as income or consumption have become worse estimates of welfare inequality over time (as inequality has increased) at least partially due to growing gaps in health and mortality.

Table 9: Gini coefficients of welfare measures over cohorts

Cohort	Welfare (λ)	Income	Consumption	Health utility	QALE
HRS	0.667	0.519	0.435	0.121	0.216
War Babies	0.726	0.485	0.438	0.133	0.228
Baby Boomers	0.813	0.505	0.458	0.139	0.238

Notes: Estimates use base year sampling weights. Income, consumption, and health utility are cross-sectional measures at age sixty. QALE is quality-adjusted life expectancy at age sixty.

4.4 Robustness

Table 10 provides sensitivity results estimated under alternate modeling assumptions from our benchmark specification. While the magnitude of inequality measures are somewhat sensitive to underlying assumptions, the finding that welfare inequality is substantial and has grown over time is quite robust across specifications. Moreover, the rank correlation between welfare and alternate well-being measures remain relatively stable.

Table 10: Robustness results

Measure	λ 10/50	λ 90/50	Gini by cohort HRS	WB	BB	ρ
Benchmark	0.226	5.215	0.667	0.726	0.813	0.568
Compensating variation	0.033	3.442	0.561	0.594	0.619	0.549
Reference 90th %tile	0.340	3.312	0.503	0.553	0.618	0.568
Age 70	0.198	8.777	0.790	0.837	-	0.575
$100k per QALY	0.059	15.336	0.848	0.886	0.937	0.485
$\beta = 0.90$	0.244	4.040	0.583	0.635	0.691	0.608
$\epsilon = 0.5$	0.199	6.186	0.706	0.762	0.842	0.567
$\epsilon = 2$	0.232	5.552	0.680	0.739	0.824	0.567
$\theta = 15.5$	0.221	5.728	0.692	0.752	0.837	0.567

Notes: Estimates use base year sampling weights. War Babies denoted by WB and Baby Boomers by BB. Spearman's rank correlation between welfare and cross-sectional consumption at age sixty denoted by ρ.

Our benchmark welfare measure is calculated in terms of consumption *equivalent* variation (EV). Alternatively, we could use the inverse of a *compensating* variation (CV) measure—what share of individual i's consumption would the median individual need to be ex-ante compensated to make them indifferent to receiving the current and potential future outcome bundles of individual i. Consumption-compensating variation satisfies:

$$log\left(\lambda_{ij}^{CV}\right) = \frac{U_{ij}(1) - U_{mj}(1)}{E\left[\sum_{a=j}^{J} \psi_{ia}\beta^{a-j}\phi(h_{ia})\right]}$$

where the only difference with the EV measure is the denominator is now the QALE for individual i as opposed to the median individual. The new measure effectively weights the benefit/loss of a consumption change by the flow utility of individual i as opposed to the median. This implies gaps in QALE are more detrimental to welfare for those below the median and less beneficial for those above.

The second row of Table 10 shows the sensitivity of inequality measures to the choice of consumption variation metric. Welfare at the tenth percentile is 22.6% of the median based on the baseline EV measure, compared to only 3.3% using CV. CV welfare estimates are also substantially lower at the top end of the distribution netting an overall lower Gini coefficient. However, the CV based measure still finds welfare inequality increased over cohorts, was higher than cross-sectional income or consumption inequality, and maintained only a modest rank correlation with consumption.

The third row of Table 10 shows sensitivity of results when using the individual at the ninetieth percentile of welfare in the HRS cohort as the reference person instead of the median. The magnitude of inequality measurements are sensitive to the reference individual chosen—the welfare Gini falls to a lower but still substantial 0.50. This puts aggregate welfare inequality closer in line with cross-sectional income and consumption at age sixty. Nonetheless, individual rank correlation does not change (by definition) and a substantial increase in inequality across cohorts remains.

The fourth row of Table 10 shows welfare results using age seventy data as initial conditions in our simulations. Relative to the age sixty benchmark, there is an increase in inequality at both ends of the welfare distribution for the HRS cohort. While Baby Boomers lack the requisite data for estimating age seventy welfare, there remains an increase in welfare inequality between HRS and War Baby cohorts. Additionally, the Gini coefficient for the older CODA cohort at age seventy is estimated to be 0.775—less than the HRS cohort. There is only a small increase in the rank correlation between welfare and consumption at age seventy compared to the age sixty benchmark.

Next we examine the impact of assuming a higher monetary value per QALY to calibrate the flow intercept $\bar{u}$. Ryen and Svensson (2015) document substantial variation across estimates of willingness to pay for a QALY, most notably with conversions based on revealed preferences of the value of statistical life (VSL) averaging 5-7 times higher than those based directly on stated preferences.[25] As a robustness check, we double our target to \$100,000 per QALY, which aligns more closely with VSL studies. The change results in a higher $\bar{u}$ placing additional weight on health and longevity differences in the welfare calculations. Inequality is substantially higher across all measurements but continues to increase across cohorts. The rank correlation with consumption also falls to 0.48.

The final four rows in Table 10 indicate sensitivity of other preference parameters. With a lower time discount rate β, our measure indicates somewhat less welfare inequality as differences in mortality and future consumption and health declines are less important. However, the pattern of main results hold. Main conclusions are also insensitive to alternate values of the Frisch elasticity of labor supply ϵ or altering the disutility weight on labor supply θ such that the first order condition holds for the sixty year old at the 75th percentile of the distribution (as opposed to our benchmark choice of the median).

Finally, we examine the robustness of results to a more general form of flow utility given by:

$$u(c,l,h) = \phi(h)\left[\bar{u} + \frac{c^{1-\gamma}}{1-\gamma}\left(1-(1-\gamma)\frac{\theta\epsilon}{1+\epsilon}(1-l)^{\frac{1+\epsilon}{\epsilon}}\right)^{\gamma} - \frac{1}{1-\gamma}\right]$$

which reduces to our benchmark case with $\gamma = 1$. These preferences follow those proposed by Trabandt and Uhlig (2011) and Jones and Klenow (2016) which maintain a constant Frisch elasticity of labor supply. With $\gamma > 1$ there is more curvature over consumption and the welfare cost of consumption inequality increases. However, leisure and consumption become less substitutable implying welfare inequality may be reduced if the inputs are strongly negatively correlated across individuals.

We examine sensitivity of results to increases in curvature to $\gamma = 1.5$ and $\gamma = 2$. However, with higher curvature over consumption than the benchmark, it is no longer

[25] The VSL studies reviewed by Ryen and Svensson (2015) are by definition measuring value of length of life, while stated preference studies elicited willingness to pay for pure quality of life improvements, pure length of life, or a mixture of both.

Table 11: Robustness results

Measure	EV 10/50 ratio by cohort HRS	WB	BB	CV 90/50 ratio by cohort HRS	WB	BB	ρ
$\gamma = 1$	0.226	0.216	0.207	3.442	3.945	4.245	0.568
$\gamma = 1.5$	0.172	0.169	0.162	4.466	5.155	5.388	0.502
$\gamma = 2$	0.182	0.185	0.168	5.225	6.167	5.896	0.450

Notes: Estimates use base year sampling weights. War Babies denoted by WB and Baby Boomers by BB. Spearman's rank correlation between EV measure of welfare and cross-sectional consumption at age sixty denoted by ρ.

possible to calculate EV welfare for those at the very top of the distribution as no amount of consumption increase would provide the same expected life-time utility to the median individual. Likewise, using the CV measure is not possible for the worst off as no amount of consumption would be enough to compensate the median individual. Under these feasibility considerations, Table 11 provides the 10/50 welfare ratio based on the EV measure and the 90/50 ratio based on CV for alternate curvatures.

The 90/50 ratio monotonically increases with the curvature parameter suggesting welfare inequality may be under-predicted in the benchmark case. The change in 10/50 ratio is non-monotonic but both robustness experiments suggest increased inequality relative to the benchmark as well. Looking more closely across cohorts, with the highest curvature the 10/50 ratio no longer declines for War Babies relative to the HRS cohort. However, there continues to be a substantial increase in dispersion indicated by the 90/50 ratio. In contrast, the 90/50 ratio falls for Baby Boomers relative to War Babies while the decline in 10/50 ratio is larger than the benchmark. While these patterns are inconclusive, they suggest the increase in inequality across cohorts could be somewhat muted with higher curvature. Finally, there is some decline in the rank correlation with consumption as leisure and consumption are less substitutable in welfare and are negatively correlated across individuals.

5 Conclusion

We propose and estimate an individual measure of welfare incorporating heterogeneity and uncertainty in future consumption, leisure, health, and mortality at age sixty. Our measure broadly indicates that inequality is larger and has increased more rapidly than suggested by other welfare metrics such as income or consumption. We also find health and mortality gaps are more important than consumption in explaining welfare inequality among the elderly in our sample, with leisure playing a comparatively minor role. Moreover, health at age sixty is a better indicator of individual well-being rank than income or consumption.

It is important to recognize that we have opted to model life-cycle dynamics as a statistical process to be estimated directly from the data. Alternatively, modeling explicit dynamic maximization of lifetime utility would allow for a richer set of counterfactual policy analyses when outcomes are endogenous. The major difficulty with the latter approach arises from specifying and solving an intertemporal model of endogenous savings, labor supply, and multivariate health behaviors and investments. While our model falls short of a fully specified structural model, the equations can be viewed as approximations of the underlying decision rules mapping state variables to individual choices.

While we rely on direct estimation of outcome dynamics, we are limited by the length of our longitudinal data, particularly for the more recent cohorts we analyze, requiring assumptions on the potential future paths of consumption and health. Incorporating longer panel data in our life-cycle models would potentially improve the fit of the model providing more precise predictions. While there are some limitations to our approach, the framework provides ample opportunities for insight and further analyses. As our framework allows estimation of welfare for each individual, it is possible to compare welfare distributions across various sub-groups of the population (see appendix B for examples of welfare breakdowns by education, region, gender, and race). While we focus on welfare at age sixty, changes in welfare can also be calculated and analyzed over the elderly life-course, for example comparing our measure with welfare at age seventy or eighty. Our measure could also be used as an outcome in designed or natural experiments, for example to examine the effect of healthcare policy on the distribution of welfare. Moreover, our framework could be extended in multiple directions to examine additional cohorts, younger ages, or welfare inequality differences across countries.

References

Attanasio, O., Hurst, E., and Pistaferri, L. (2014). The evolution of income, consumption, and leisure inequality in the United States, 1980–2010. In *Improving the Measurement of Consumer Expenditures*, pages 100–140. University of Chicago Press.

Attanasio, O. and Pistaferri, L. (2014). Consumption inequality over the last half century: Some evidence using the new PSID consumption measure. *The American Economic Review*, 104(5):122–126.

Attanasio, O. P., Battistin, E., and Padula, M. (2010). *Inequality in living standards since 1980: Income tells only a small part of the story*. Government Institutes.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in US wage inequality: Revising the revisionists. *The Review of Economics and Statistics*, 90(2):300–323.

Barr, R., Petrie, C., Furlong, W., Rothney, M., and Feeny, D. (1997). Health-related quality of life during post-induction chemotherapy in children with acute lymphoblastic leukemia in remission. *International Journal of Oncology*, 11(2):333–339.

Barro, R. J. and Redlick, C. J. (2011). Macroeconomic effects from government purchases and taxes. *The Quarterly Journal of Economics*, 126(1):51–102.

Becker, G. S., Philipson, T. J., and Soares, R. R. (2005). The quantity and quality of life and the evolution of world inequality. *The American Economic Review*, 95(1):277–291.

Blanchard, C., Feeny, D., Mahon, J. L., Bourne, R., Rorabeck, C., Stitt, L., and Webster-Bogaert, S. (2003). Is the health utilities index responsive in total hip arthroplasty patients? *Journal of Clinical Epidemiology*, 56(11):1046–1054.

Born, B. and Breitung, J. (2016). Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews*, 35(7):1290–1316.

Bosworth, B., Burtless, G., and Zhang, K. (2016). Later retirement, inequality in old age, and the growing gap in longevity between rich and poor. *Economic Studies at Brookings*.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *JAMA*, 315(16):1750–1766.

Colby, S. L., Ortman, J. M., et al. (2015). Projections of the size and composition of the US population: 2014 to 2060. *US Census Bureau*, 9.

Crystal, S. and Shea, D. (1990). The economic well-being of the elderly. *Review of Income and Wealth*, 36(3):227–247.

Crystal, S. and Waehrer, K. (1996). Later-life economic inequality in longitudinal perspective. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 51(6):S307–S318.

Currie, J. and Madrian, B. C. (1999). Health, health insurance and the labor market. *Handbook of Labor Economics*, 3:3309–3416.

Currie, J. and Schwandt, H. (2016). Inequality in mortality decreased among the young while increasing for older adults, 1990–2010. *Science*, 352(6286):708–712.

Cutler, D. M., Ghosh, K., and Landrum, M. B. (2013). Evidence for significant compression of morbidity in the elderly US population. *NBER Working Paper No. 19268*.

Cutler, D. M. and Katz, L. F. (1992). Rising inequality? Changes in the distribution of income and consumption in the 1980s. *NBER Working Paper No. 3964*.

De Vos, I., Everaert, G., Ruyssen, I., et al. (2015). Bootstrap-based bias correction and inference for dynamic panels with fixed effects. *Stata Journal*, 15(4):986–1018(33).

Deaton, A. S. and Paxson, C. H. (1998). Aging and inequality in income and health. *The American Economic Review*, 88(2):248–253.

Dworkin, R. (1985). *A matter of principle*. OUP Oxford.

Everaert, G. and Pozzi, L. (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control*, 31(4):1160–1184.

Fanshel, S. and Bush, J. W. (1970). A health-status index and its application to health-services outcomes. *Operations research*, 18(6):1021–1066.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., Denton, M., and Boyle, M. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical care*, 40(2):113–128.

Fleurbaey, M. (2009). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, 47(4):1029–1075.

Fleurbaey, M. and Gaulier, G. (2009). International comparisons of living standards by equivalent incomes. *The Scandinavian Journal of Economics*, 111(3):597–624.

Furlong, W., Feeny, D., Torrance, G., Goldsmith, C., DePauw, S., Zhu, Z., Denton, M., Boyle, M., et al. (1998). Multiplicative multi-attribute utility function for the health utilities index mark 3 (hui3) system: a technical report. Technical report, Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada.

Furlong, W. J., Feeny, D. H., Torrance, G. W., and Barr, R. D. (2001). The health utilities index (hui®) system for assessing health-related quality of life in clinical studies. *Annals of Medicine*, 33(5):375–384.

Gottschalk, P., Moffitt, R., Katz, L. F., and Dickens, W. T. (1994). The growth of earnings instability in the US labor market. *Brookings Papers on Economic Activity*, 1994(2):217–272.

Hall, R. E. (2009). Reconciling cyclical movements in the marginal value of time and the marginal product of labor. *Journal of Political Economy*, 117(2):281–323.

Heathcote, J., Perri, F., and Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. *Review of Economic Dynamics*, 13(1):15–51.

Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581.

Honaker, J., King, G., Blackwell, M., et al. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45(7):1–47.

Horsman, J., Furlong, W., Feeny, D., and Torrance, G. (2003). The health utilities index (hui®): concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, 1(1):54.

Hurd, M. D. and Rohwedder, S. (2007). Economic well-being at older ages: Income- and consumption-based poverty measures in the HRS. *RAND Corporation Working Paper WR-410*.

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

Jones, C. I. and Klenow, P. J. (2016). Beyond GDP? Welfare across countries and time. *The American Economic Review*, 106(9):2426–2457.

Kaplan, R. M. and Bush, J. W. (1982). Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychology*, 1(1):61.

Katz, L. F. et al. (1999). Changes in the wage structure and earnings inequality. *Handbook of Labor Economics*, 3:1463–1555.

Kenworthy, L. (2008). *Jobs with Equality*. Oxford University Press.

Klarman, H. E. and Rosenthal, G. D. (1968). Cost effectiveness analysis applied to the treatment of chronic renal disease. *Medical Care*, 6(1):48–54.

Meara, E. R., Richards, S., and Cutler, D. M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs*, 27(2):350–360.

Minkler, M., Fuller-Thomson, E., and Guralnik, J. M. (2006). Gradient of disability across the socioeconomic spectrum in the United States. *New England Journal of Medicine*, 355(7):695–703.

Mullahy, J. (2016). Estimation of multivariate probit models via bivariate probit. *Stata Journal*, 16(1):37–51.

National Academies of Sciences, Engineering, and Medicine (2015). *The growing gap in life expectancy by income: Implications for federal programs and policy responses.* National Academies Press, Washington, DC.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Pijoan-Mas, J. and Ríos-Rull, J.-V. (2014). Heterogeneity in expected longevities. *Demography*, 51(6):2075–2102.

Piketty, T. and Saez, E. (2014). Inequality in the long run. *Science*, 344(6186):838–843.

Rawls, J. (1971). *A theory of justice*. Harvard university press.

Roemer, J. E. (1998). *Equality of opportunity*. Harvard University Press.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

Ryen, L. and Svensson, M. (2015). The willingness to pay for a quality adjusted life year: a review of the empirical literature. *Health Economics*, 24(10):1289–1301.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall, London.

Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 13(2):144.

Torrance, G. W., Thomas, W. H., and Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health Services Research*, 7(2):118.

Trabandt, M. and Uhlig, H. (2011). The Laffer curve revisited. *Journal of Monetary Economics*, 58(4):305–327.

Wursten, J. et al. (2016). Xtqptest: Stata module to perform Born & Breitung bias-corrected LM-based test for serial correlation. *Statistical Software Components*.

Zeckhauser, R. and Shepard, D. (1976). Where now for saving lives? *Law and Contemporary Problems*, 40(4):5–45.

A Imputation of consumption and other missing data

The CAMS collected consumption data for approximately 20% of the HRS sample starting from 2001. In order to estimate our dynamic panel models and construct simulated life-cycle paths for the remaining sample, we multiply impute their consumption data. We use the computationally attractive EM-bootstrapping algorithm allowing for cross-sectional time-series data proposed by Honaker and King (2010) and implemented through the freely available Amelia II software program (Honaker et al., 2011). This approach provides m separate complete datasets in which all analyses are conducted independently. Results are then combined into a single estimate.[26] We set $m = 12$ but test the sensitivity of results to higher values of m.

There are two primary assumptions underlying the proposed imputation method. First, the complete data is assumed be multivariate normal. While this may seem somewhat restrictive, it has been shown that multivariate normal imputation models provide an adequate approximation to the true underlying distribution in a variety of settings, even in the presence of categorical or mixed data (Schafer, 1997). Second is the standard required assumption that data is missing at random (MAR)—any nonrandom pattern of missingness can be accounted for by the observed data included in the model. Note this is less restrictive than the requirement data be missing completely at random (MCAR). In practice, we know that missing data is not at random, at least for years falling outside of the CAMS window (1992-1998 and 2012-2014). However, by including a rich set of related covariates in the imputation model, we argue that missing data can be treated as MAR in the statistical sense. While there is no way to empirically test this assumption, we run a number of diagnostic tests to check the credibility of the imputation model in search of any obvious deficiencies.

Variables from the RAND HRS data file (v.P) included in our imputation model are age (AGEY_E), aged squared, number of household members (HHRES), total wealth (ATOTA), wealth squared, log household income (ITOT), log income squared, and dummy indicators for labor force status (LBRF), gender (RAGENDER), race (RARACEM), education (RAEDUC), marital status (MSTAT), census division (CENDIV), 1980 census occupation code for longest reported tenure (JLOCC), self-reported health (SHLT), ADLs (ADLA), and eight doctor diagnosed health conditions (HIBPE, DIABE, CANCRE, LUNGE, HEARTE, STROKE, PSYCHE, ARTHRE). The model also included our constructed indicator for retirement and hours worked. In order to allow for the time-series structure of the data, lags and leads of consumption, wealth, income, and hours worked are included in the imputation model. While we are primarily imputing consumption data, Amelia II also provides imputed values for all other missing variables included in the model.[27]

A useful check of the viability of the imputation model is to compare the distributions of the imputed values against the observed data. While there is no need for these distributions to be the same, the comparison gives a sense of the plausibility of imputations (Honaker et al., 2011). Figure 8 plots the density of observed and imputed values of consumption. The imputed values are taken as the mean across the m imputed datasets. The comparison suggests no unusual pattern in the distribution of imputed values, providing cursory support of model plausibility.

Another diagnostic tool proposed by Honaker et al. (2011) is *overimputing.* While

[26] Assuming asymptotically normally distributed statistics implies a simple average across datasets (Rubin, 2004).

[27] If the observed data used in the imputation model has a poorly behaved likelihood, the convergence of the EM algorithm could be sensitive to the staring values chosen. We found no evidence of local convergence issues using the overdispersed start values diagnostic test proposed by Honaker et al. (2011).

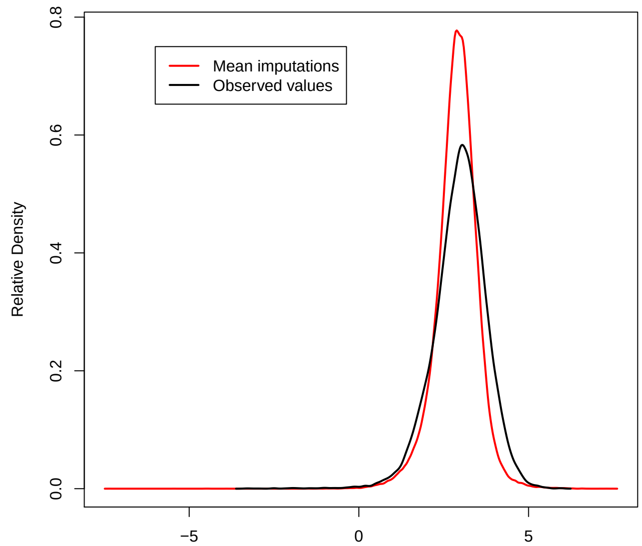

Figure 8: Distributions of observed and imputed values of consumption

it is impossible to examine if the imputed values are close to the missing values they are attempting to recover, *observed* values can be used to test the accuracy of the imputation process. Overimputing sequentially treats each of the observed consumption values as if they were missing and then imputes their values several hundred times. This provides a mean imputed value and confidence interval that can be compared to the actual observed data. Figure 9 plots all observed consumption values against the mean of their imputed values and the associated 95% confidence interval. A visual inspection of the diagnostic plot suggests the model does fairly well predicting values other than the lowest values. However, few individuals lie in this extreme end of the distribution—less than 0.3% of the observations fall below zero ($1,000 annual consumption). Honaker et al. (2011) suggest a good imputation model should have at least 90% of the confidence intervals containing the true values (i.e. 90% of the confidence intervals should cross the $y = x$ line). In our case, 94% of the observed values are within the confidence bounds.

As a final examination of the imputation model we try to get a sense of how it predicts missing values in a time series. While it is infeasible to examine the imputed time trends for each individual in the sample, Figure 10 provides time series for a random sub-set of ten individuals with at least one observed consumption value. The mean of the imputed values are plotted in red with 95% confidence bounds (based on 100 imputations). The isolated black points without bounds are observed data. Broadly, the imputed values fall in line with the observed data and no egregious outliers emerge. Note that prior to wave five (2000) and after wave ten (2010) all values are

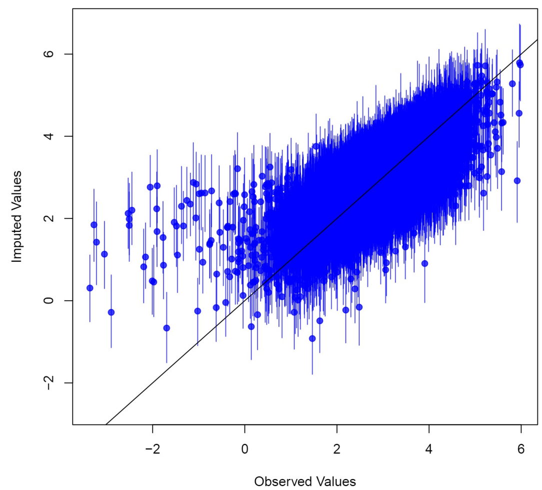

Figure 9: Overimputed values of consumption

imputed as these waves are outside of our CAMS data window.

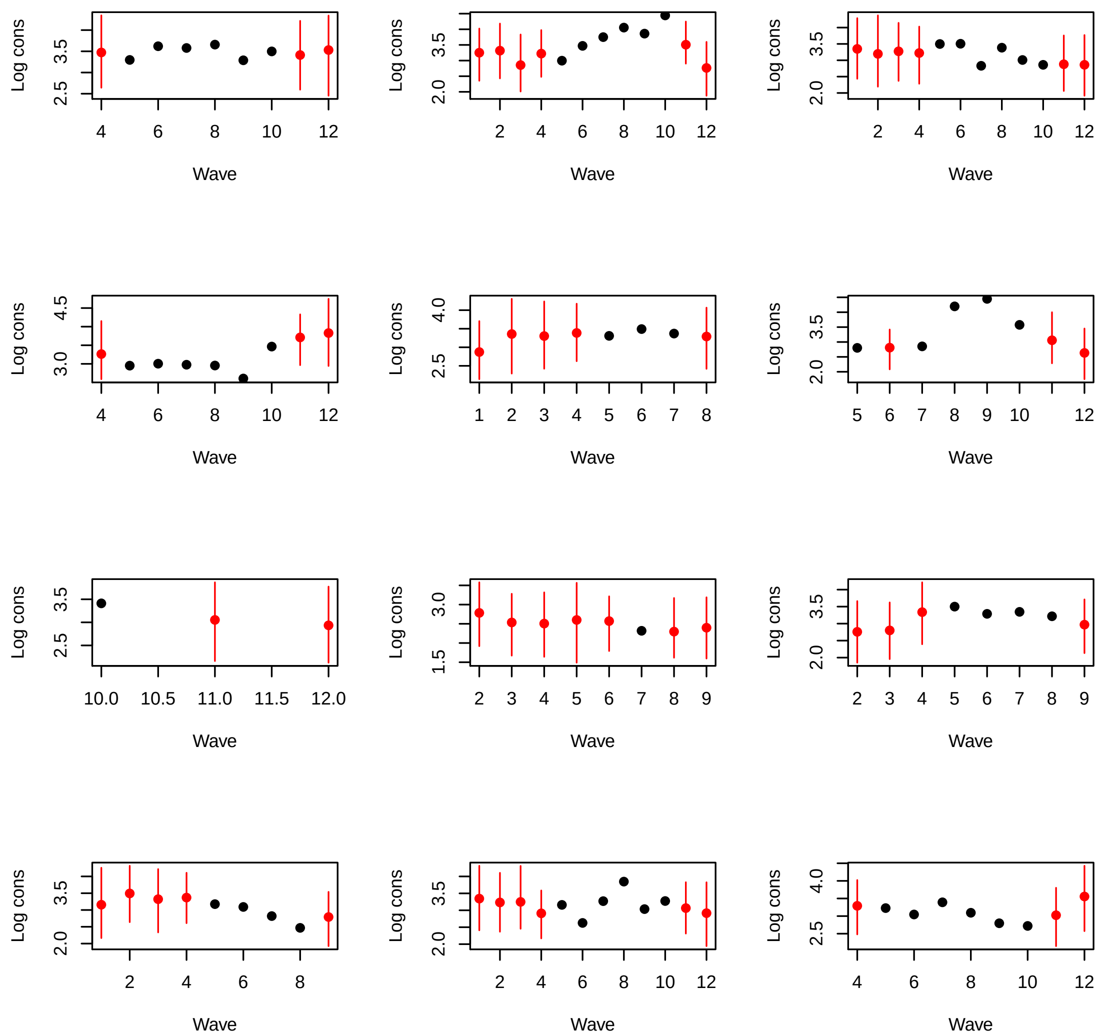

Figure 10: Observed and imputed consumption over time for a random sub-sample

B Additional Figures and Tables

Table 12: Additional estimation sample descriptive statistics by cohort

	AHEAD	CODA	HRS	WB	BB
Male (%)	37.69	43.87	45.73	45.58	46.02
Education (%)					
<HS	38.62	30.98	26.17	18.39	14.98
HS	31.19	32.23	33.80	30.73	24.66
Some College	16.79	17.84	20.20	24.94	28.94
College	13.40	18.95	19.83	25.94	31.42
Race (%)					
White	91.01	88.67	87.27	85.45	80.88
Black	7.27	8.08	9.45	9.81	10.75
Other	1.71	3.25	3.28	4.74	8.37
Census division (%)					
New England	6.07	5.25	5.36	3.78	4.37
Mid Atlantic	14.04	12.24	14.42	11.51	11.84
EN Central	18.40	17.73	15.34	16.83	18.35
WN Central	9.44	9.47	8.91	8.48	7.38
S Atlantic	17.95	20.65	20.63	26.54	20.19
ES Central	4.93	5.73	6.21	5.59	6.72
WS Central	10.42	9.18	9.12	9.94	9.66
Mountain	4.63	5.34	5.80	5.44	7.85
Pacific	14.10	14.36	14.10	11.79	13.59
Not US	0.02	0.04	0.12	0.11	0.07

Notes: Mean and percentage estimates use base year sampling weights. War Babies denoted by WB and Baby Boomers by BB.

Table 13: Dynamic model estimates for morbidities

Variable	Hypertension		Diabetes		Cancer		Lung disease		Heart disease		Stroke		Psych		Arthritis	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	SE	SE
Hyper			0.281	0.035	-0.043	0.039	0.086	0.041	0.132	0.033	0.100	0.041	0.176	0.037	0.102	0.033
Diab	0.266	0.051			0.072	0.047	0.007	0.053	0.037	0.045	0.056	0.053	-0.009	0.052	0.063	0.046
Cancer	-0.007	0.052	0.032	0.055			0.084	0.060	-0.089	0.052	0.017	0.059	0.012	0.061	0.048	0.052
Lung	0.063	0.061	0.045	0.059	0.107	0.059			0.265	0.052	0.020	0.064	0.156	0.061	0.161	0.065
Heart	0.100	0.044	0.092	0.041	0.029	0.042	0.222	0.042			0.185	0.041	0.118	0.042	0.133	0.042
Stroke	0.081	0.068	-0.021	0.064	-0.015	0.060	0.037	0.063	0.076	0.056			0.296	0.052	-0.026	0.064
Psych	0.111	0.054	0.064	0.053	-0.072	0.060	0.103	0.057	0.127	0.048	0.171	0.054			0.267	0.055
Arthritis	0.087	0.030	0.001	0.034	-0.012	0.035	0.143	0.038	0.085	0.030	0.001	0.037	0.121	0.036		
ADL	0.075	0.033	0.031	0.036	0.002	0.036	0.143	0.036	0.085	0.030	0.207	0.033	0.265	0.032	0.185	0.037
Lag Hyper			0.030	0.034	0.065	0.039	-0.072	0.040	0.044	0.033	0.056	0.041	-0.095	0.037	0.019	0.033
Lag Diab	-0.074	0.055			-0.088	0.050	-0.096	0.056	0.115	0.047	0.083	0.055	0.043	0.055	-0.050	0.049
Lag Cancer	-0.002	0.056	0.001	0.059			0.058	0.064	0.127	0.055	-0.014	0.063	0.056	0.065	0.029	0.056
Lag Lung	-0.127	0.067	-0.055	0.065	0.022	0.064			-0.095	0.058	0.056	0.069	0.005	0.067	-0.053	0.072
Lag Heart	-0.031	0.047	0.010	0.043	-0.008	0.043	-0.056	0.044			-0.014	0.042	-0.044	0.044	-0.047	0.045
Lag Stroke	-0.035	0.076	0.036	0.070	0.016	0.066	-0.027	0.069	0.071	0.061			-0.193	0.059	0.034	0.070
Lag Psych	-0.056	0.058	-0.074	0.057	0.060	0.064	0.038	0.060	-0.056	0.051	-0.052	0.057			-0.130	0.059
Lag Arthre	-0.030	0.030	0.018	0.034	0.059	0.035	-0.005	0.036	0.031	0.030	0.001	0.037	0.001	0.035		
Lag ADL	-0.065	0.036	0.035	0.038	-0.001	0.038	-0.024	0.038	0.017	0.032	-0.097	0.035	-0.058	0.034	-0.073	0.042
Time	0.038	0.005	0.019	0.006	0.008	0.006	0.031	0.007	-0.006	0.005	-0.017	0.006	0.003	0.006	-0.012	0.005
2008	0.077	0.024	0.124	0.024	0.058	0.025	0.054	0.029	0.093	0.022	0.099	0.028	0.042	0.028	0.051	0.026
CODA	-0.039	0.035	-0.027	0.039	-0.015	0.036	-0.022	0.041	-0.031	0.033	0.005	0.037	0.057	0.039	-0.160	0.036
HRS	-0.087	0.049	-0.049	0.054	-0.081	0.052	-0.097	0.058	-0.015	0.046	-0.023	0.053	0.054	0.055	-0.190	0.050
War Babies	-0.088	0.066	0.014	0.073	-0.075	0.072	-0.119	0.081	0.024	0.064	0.054	0.077	0.194	0.076	-0.094	0.067
Boomers	-0.203	0.081	-0.016	0.089	-0.080	0.088	-0.174	0.101	0.043	0.079	0.044	0.095	0.259	0.094	-0.090	0.082
Black	0.184	0.023	0.065	0.022	-0.024	0.023	-0.173	0.027	-0.125	0.021	0.039	0.025	-0.189	0.026	0.006	0.021
Other race	0.021	0.033	0.189	0.035	-0.185	0.045	-0.115	0.047	-0.103	0.037	-0.146	0.050	-0.045	0.042	-0.038	0.034
HS grad	-0.026	0.019	-0.059	0.021	-0.000	0.021	-0.054	0.023	0.021	0.019	0.044	0.023	-0.065	0.022	-0.036	0.020
Some college	-0.057	0.022	-0.058	0.025	0.039	0.025	-0.023	0.028	0.039	0.022	0.062	0.028	-0.011	0.026	-0.012	0.023
College grad	-0.086	0.026	-0.110	0.030	0.036	0.030	-0.152	0.037	-0.021	0.027	0.046	0.034	-0.033	0.032	-0.068	0.027
Female	0.030	0.016	-0.117	0.018	-0.201	0.018	-0.046	0.021	-0.179	0.016	-0.060	0.020	0.123	0.019	0.155	0.016
Health FE	-0.044	0.014	-0.190	0.016	-0.081	0.018	-0.308	0.020	-0.157	0.015	-0.146	0.019	-0.259	0.018	-0.078	0.014
SES FE	-0.067	0.022	-0.099	0.022	0.055	0.026	-0.047	0.035	-0.019	0.024	-0.053	0.032	-0.067	0.027	0.042	0.022
Constant	-1.373	0.091	-1.419	0.103	-1.938	0.106	-1.488	0.124	-1.541	0.094	-2.263	0.124	-1.478	0.109	-1.257	0.091

Notes: Multivarite probit results with dependent variable across columns. Regressions also include dummies for age, occupation, and census division.

Table 14: Dynamic model estimates ADLs, self-rated health, mortality, retirement, and consumption

	ADLs		Self-rated health		Mortality		Retirement		Consumption	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Hyper	0.071	0.031	-0.183	0.009	0.068	0.029	0.048	0.037	-0.009	0.013
Diab	0.109	0.040	-0.130	0.015	0.105	0.036	0.042	0.052	-0.025	0.022
Cancer	0.112	0.044	-0.466	0.017	0.652	0.029	0.104	0.063	0.032	0.020
Lung	0.236	0.049	-0.251	0.022	0.397	0.035	0.211	0.077	-0.009	0.024
Heart	0.130	0.034	-0.324	0.013	0.192	0.027	0.005	0.053	-0.007	0.022
Stroke	0.464	0.047	-0.327	0.015	0.261	0.032	0.291	0.103	-0.063	0.026
Psych	0.392	0.044	-0.241	0.017	0.226	0.033	0.285	0.067	-0.071	0.022
Arthritis	0.224	0.026	-0.147	0.012	-0.133	0.029	0.118	0.033	0.014	0.015
ADL			-0.413	0.010	0.357	0.021	0.252	0.045	-0.070	0.015
Health					-0.147	0.009	-0.024	0.011	0.039	0.004
Lag Hyper	-0.005	0.031	0.069	0.013	-0.030	0.029	-0.015	0.038	-0.005	0.012
Lag Diab	0.033	0.042	0.065	0.015	0.055	0.037	-0.029	0.056	0.001	0.019
Lag Cancer	-0.025	0.047	0.302	0.017	-0.433	0.031	-0.067	0.069	-0.005	0.021
Lag Lung	-0.002	0.054	0.117	0.022	-0.091	0.038	-0.086	0.086	0.002	0.020
Lag Heart	-0.035	0.036	0.162	0.017	-0.023	0.028	0.052	0.057	0.009	0.019
Lag Stroke	-0.194	0.053	0.196	0.019	-0.045	0.035	-0.190	0.116	0.011	0.023
Lag Psych	-0.131	0.047	0.148	0.020	-0.120	0.035	-0.120	0.072	0.027	0.020
Lag Arthre	0.078	0.026	0.067	0.014	0.058	0.028	-0.071	0.034	-0.009	0.013
Lag ADL			0.187	0.008	-0.090	0.021	-0.178	0.053	0.018	0.017
Time	-0.029	0.005	-0.042	0.007	0.024	0.004	-0.030	0.006	-0.001	0.009
2008	0.058	0.023	-0.045	0.007	0.149	0.019	0.163	0.029	0.014	0.007
CODA	0.083	0.029			-0.062	0.025	0.131	0.076		
HRS	0.087	0.042			-0.095	0.037	0.119	0.088		
War Babies	0.107	0.059			-0.257	0.055	0.186	0.103		
Boomers	0.170	0.073			-0.431	0.069	0.190	0.116		
Black	0.088	0.019			0.028	0.018	0.036	0.024		
Other race	0.018	0.034			-0.119	0.035	-0.070	0.038		
HS grad	-0.097	0.017			0.039	0.016	-0.032	0.024		
Some college	-0.051	0.020			0.046	0.020	-0.041	0.027		
College grad	-0.083	0.025			0.055	0.024	-0.032	0.031		
Female	0.012	0.015			-0.193	0.014	0.106	0.018		
Health FE	-0.441	0.015			-0.168	0.017	-0.191	0.021		
SES FE	-0.112	0.021			-0.088	0.021	-0.038	0.026		
Constant	-0.529	0.083			-2.286	0.177	-0.816	0.124		
Lag Health			0.239	0.004	0.005	0.008	-0.004	0.011	0.002	0.004
Lag2 Health			0.134	0.003						
Retired									-0.049	0.014
Lag Con									0.170	0.005
Lag2 Con									0.082	0.004

Notes: Dependent variable across columns. Multivariate probit results reported for ADLs as dependent outcome. Standard probit results reported for mortality and retirement as dependant outcomes. Linear dynamic panel estimates reported for self-rated health and consumption as outcomes. All regressions also include dummies for age. Regressions for ADLs, mortality, and retirement also include dummies for occupation and census division. Regression for self-rated health also includes second lags for all morbidities.

Table 15: Morbidity shock covariance matrix (Σ)

	Hyper	Diabetes	Cancer	Lung	Heart	Stroke	Psych	Arthritis	ADLs
Hyper	1.00	0.26	0.04	0.08	0.28	0.29	0.14	0.09	0.09
Diabetes	0.26	1.00	0.07	0.05	0.10	0.14	0.06	0.03	0.07
Cancer	0.04	0.07	1.00	0.12	0.02	0.05	0.11	0.05	0.13
Lung	0.08	0.05	0.12	1.00	0.22	0.10	0.17	0.10	0.19
Heart	0.28	0.10	0.02	0.22	1.00	0.28	0.16	0.10	0.14
Stroke	0.29	0.14	0.05	0.10	0.28	1.00	0.21	0.10	0.39
Psych	0.14	0.06	0.11	0.17	0.16	0.21	1.00	0.15	0.29
Arthritis	0.09	0.03	0.05	0.10	0.10	0.10	0.15	1.00	0.26
ADLs	0.09	0.07	0.13	0.19	0.14	0.39	0.29	0.26	1.00

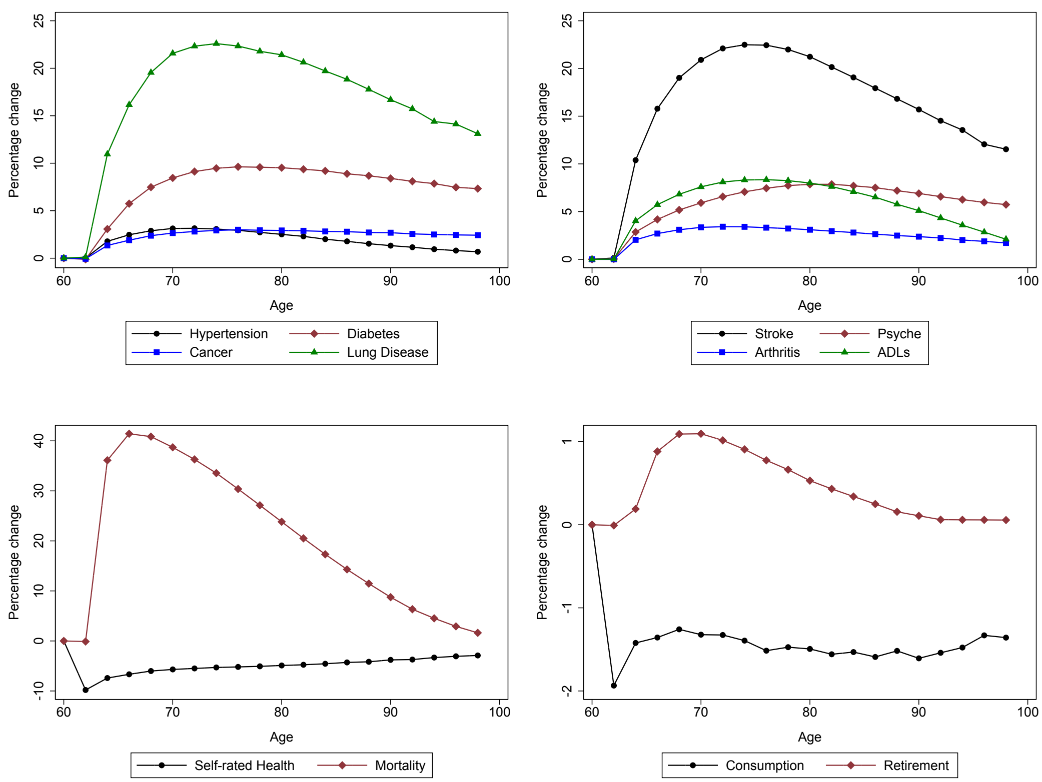

Figure 11: Impulse response to incidence of heart disease at age 62

Notes: Results plot percentage difference in expected outcomes with the exogenous onset of heart disease at age sixty-two relative to remaining without heart disease at sixty-two. Sample includes all individuals in the simulation sample without heart disease at age sixty. Expected outcomes are conditional on survival.

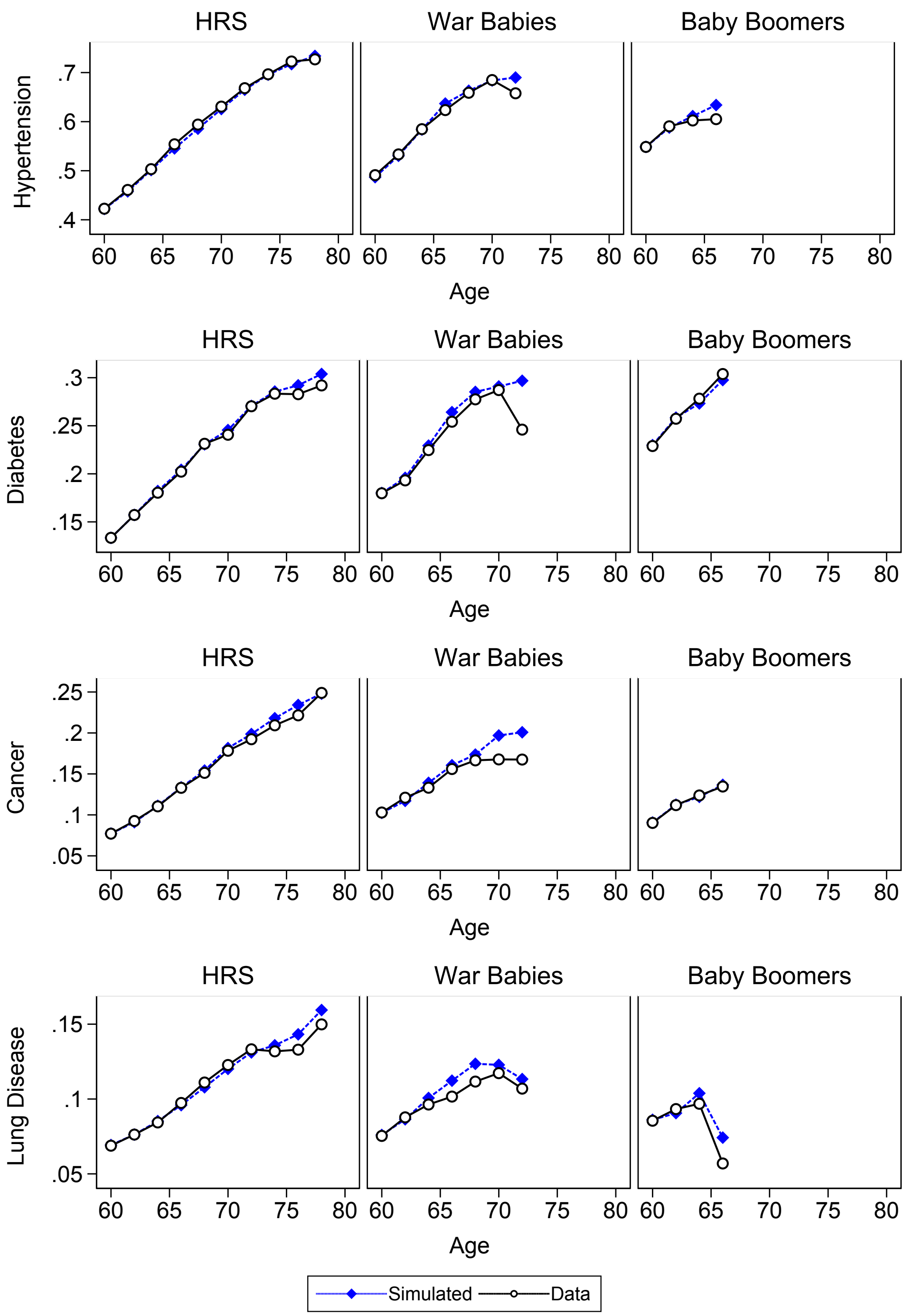

Figure 12: Mean of life-cycle morbidity profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

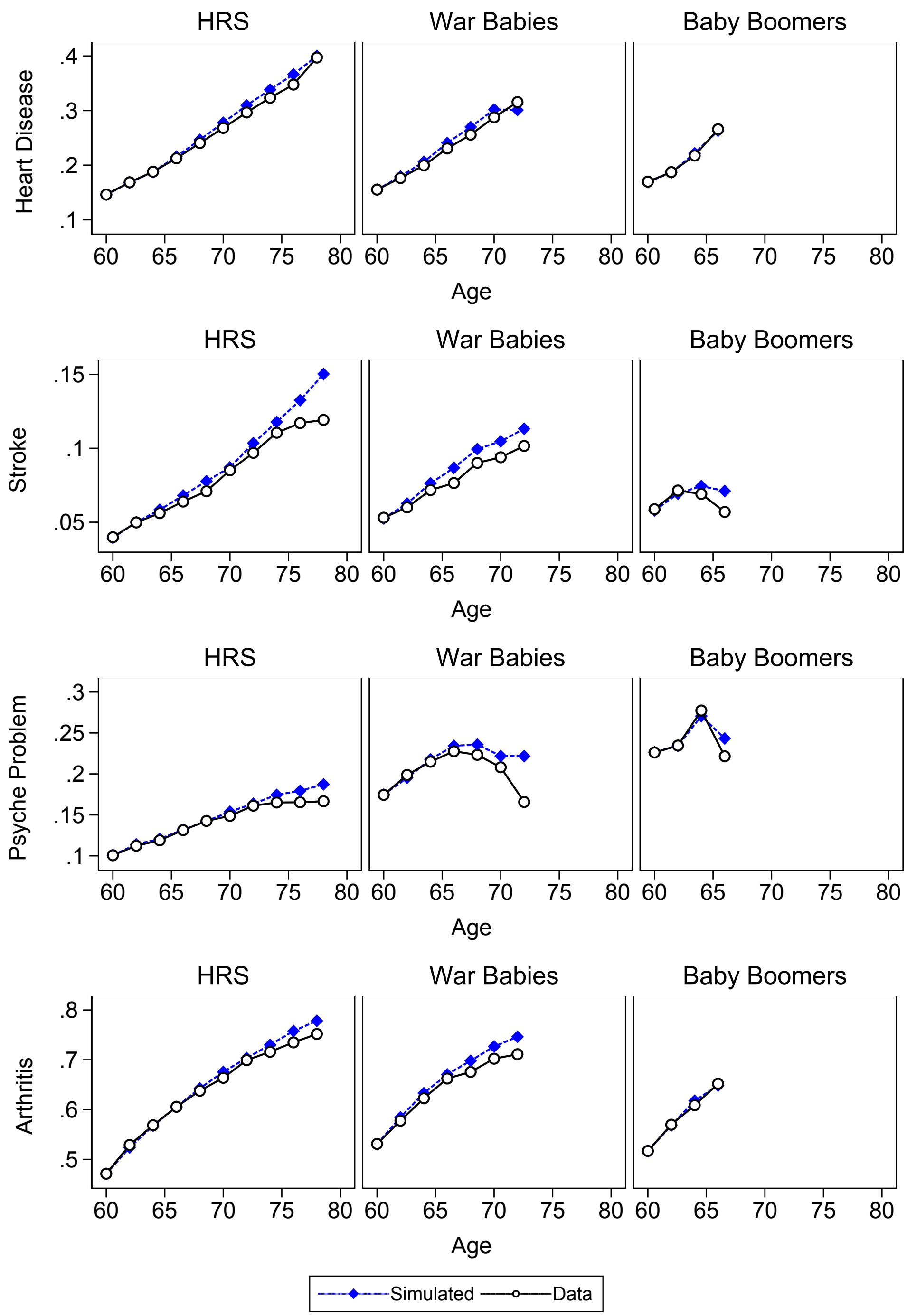

Figure 13: Mean of life-cycle morbidity profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

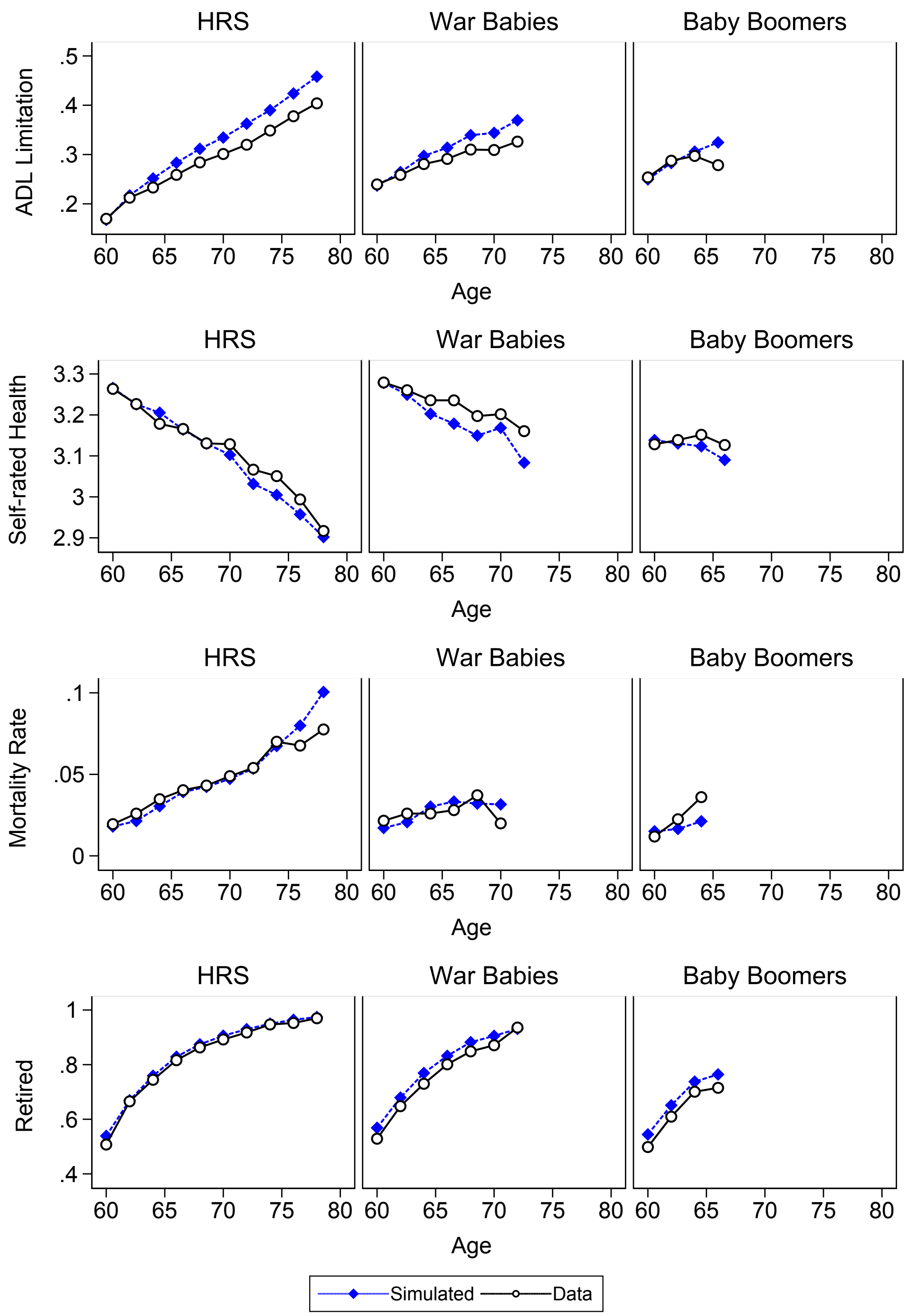

Figure 14: Mean of life-cycle health, mortality, and retirement profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

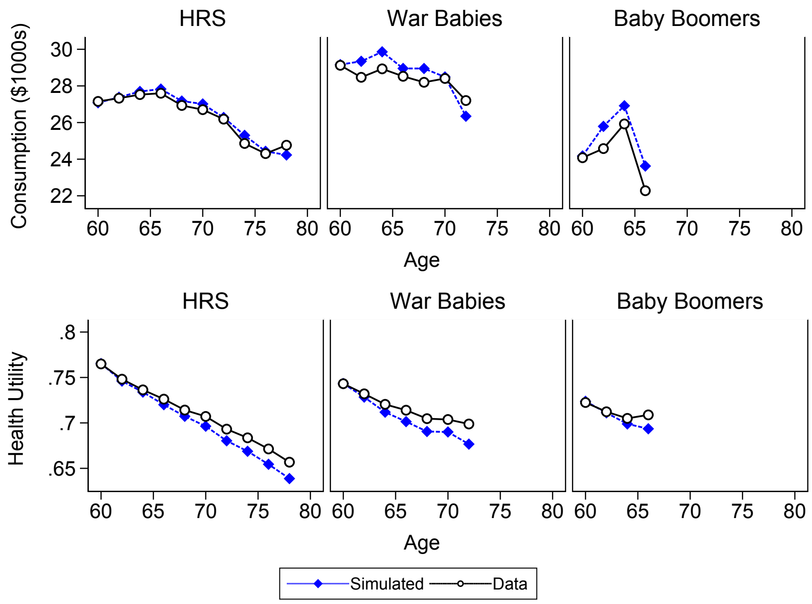

Figure 15: Mean of life-cycle consumption and health utility profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

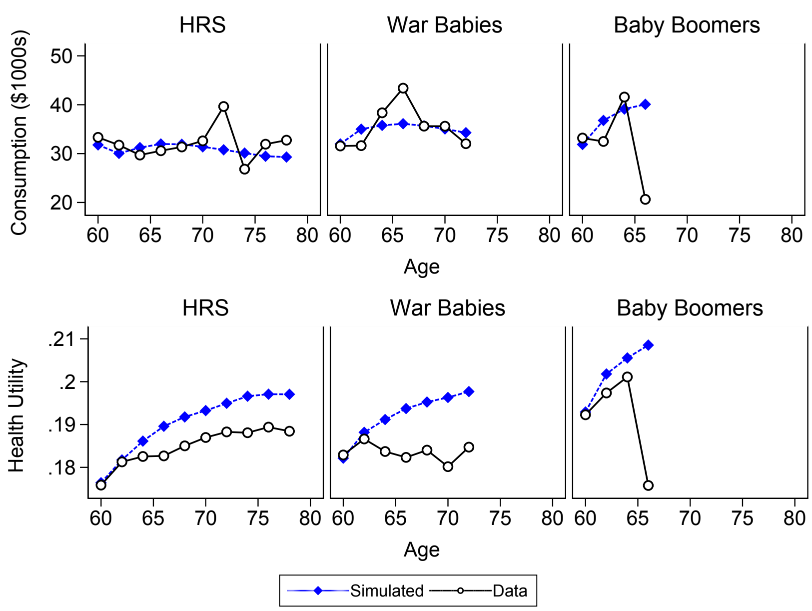

Figure 16: Standard deviation of consumption and health utility life-cycle profiles by cohort

Notes: "Data" plots standard deviation of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of standard deviations of simulated outcome for each observation in the data (i.e. the mean of standard deviations calculated for each of the 5,000 simulation runs using only person-year observations that also appear in the data).

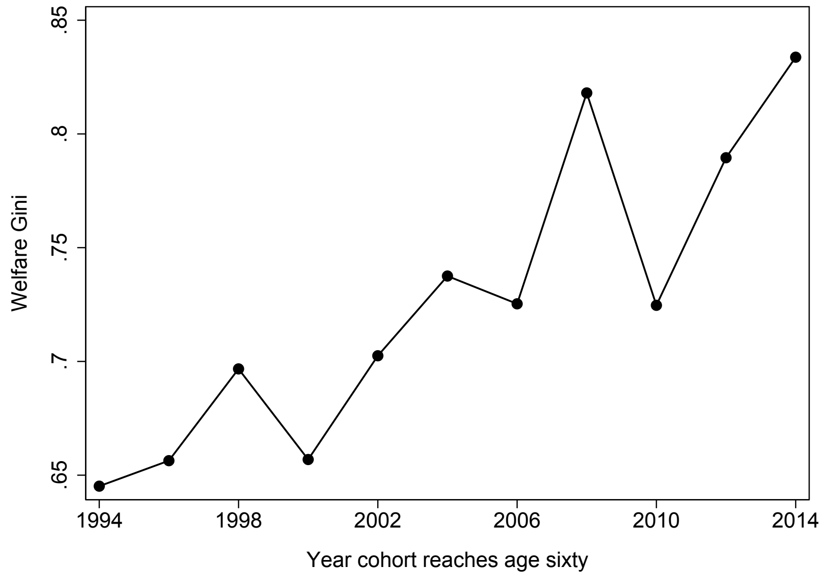

Figure 17: Welfare (λ) Gini by two-year birth cohort

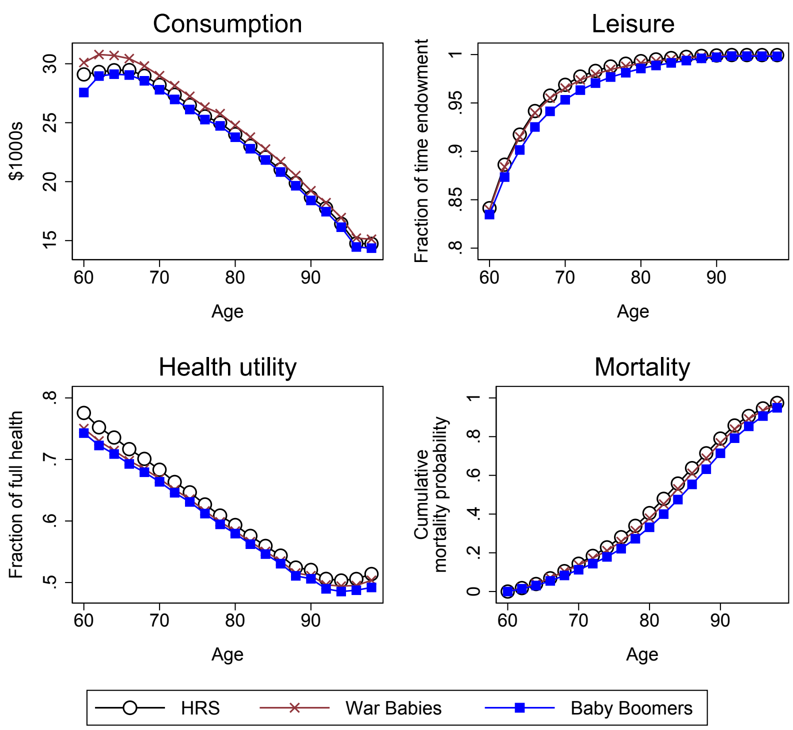

Figure 18: Average life-cycle profiles by cohort

Table 16: Welfare decomposition in HRS cohort and welfare Gini for each cohort by select characteristics

	Median λ	Mean $\log \lambda$	Decomposition			Welfare Gini by cohort		
			Cons.	Leisure	QALY	HRS	War	Boomers
Education								
<HS	0.436	-0.744	-0.249	-0.015	-0.480	0.532	0.549	0.640
HS grad	1.022	0.029	0.134	-0.031	-0.075	0.567	0.616	0.639
Some college	1.295	0.305	0.331	-0.042	0.015	0.616	0.703	0.766
College grad	2.538	0.956	0.658	-0.059	0.357	0.647	0.687	0.795
Gender								
Male	0.908	-0.023	0.215	-0.054	-0.184	0.654	0.674	0.766
Female	1.100	0.129	0.139	-0.018	0.008	0.673	0.752	0.832
Race								
White	1.103	0.148	0.229	-0.036	-0.046	0.661	0.721	0.810
Black	0.461	-0.676	-0.237	-0.019	-0.419	0.594	0.539	0.743
Other	0.764	-0.153	-0.101	-0.039	-0.013	0.674	0.640	0.727

Notes: Estimates use base year sampling weights.

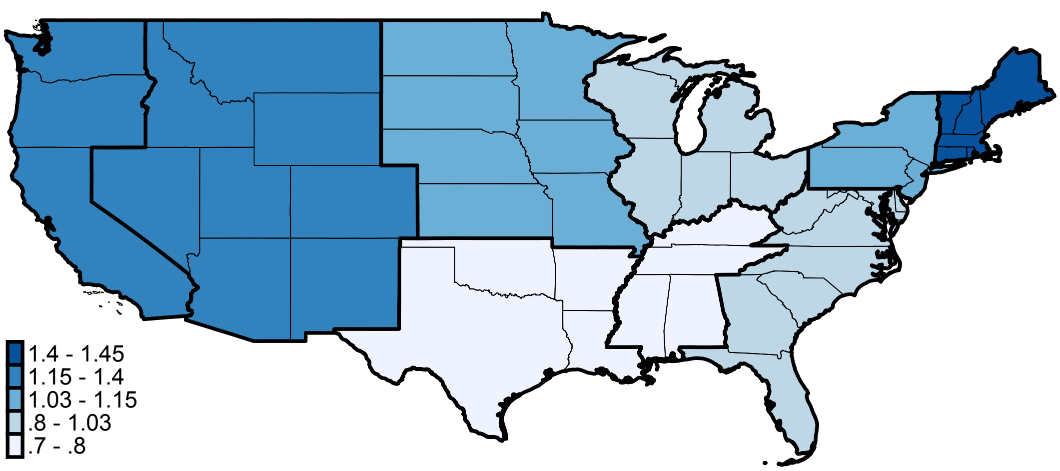

Figure 19: Median welfare by census division for HRS cohort