Understanding the Economic Consequences of Shifting Trends in Population Health

D. Goldman, D. Lakdawalla, P.-C. Michaud and Y. Zheng[1]
RAND Corporation

April 2008

Preliminary
Do Not Cite without Permission

Abstract

Much attention has focused on the economic consequences of increased life expectancy and demographic aging in the years to come. However, the economic burden of shifting trends in population health remains uncertain. Sustained increases in obesity, diabetes, and other diseases could reduce life expectancy—with a concomitant decrease in the annuity burden—but these savings may be offset by worsening functional status that increases health care spending, reduces labor supply, and increases public assistance. We attempt to quantify the economic consequences of shifting trends in population health in an international context by contrasting US and European health. We then use microsimulation to model medical care costs, labor supply, earnings, wealth, tax revenues, and government expenditures (including Social Security and income assistance). This simulation is used to compare outcomes under a variety of population health scenarios.

[1] This research was funded by Employee Benefits Security Administration US Department of Labor through Contract J-9-P-2-0033. Corresponding author: Pierre-Carl Michaud, michaud@rand.org, 1776 Main Street, Santa Monica, CA 90407-2138, United States, Tel: +1 (310) 393-0411 extension 7620. The authors are solely responsible for the content.

1. Introduction

Understanding how the current public and private pension system will respond to the changing demographic composition and health of the population is crucial to anticipating future policy needs. While we have clear ideas of the financial consequences of increased life expectancy and demographic aging in the years to come, the burden placed on Americans by shifting trends in population health remains uncertain. Mobility and environmental modifications have made it easier to function in society but the alarming increase in obesity and its immediate consequences such as diabetes and other health problems make it difficult to forecast their financial implications.

This paper examine how the interplay of these health and demographic trends will ultimately affect savings, retirement, and the burden on the public sector over the next 50 years. We extend the Future Elderly Model (FEM) developed by Goldman et al. (2004), a model currently describing population health dynamics in detail, to include economic outcomes such as financial wealth, pension claiming and labor force participation. The model can answer a series of "what-if" questions by forecasting health and economic outcomes in the future based on dynamic relationships that estimated using Health and Retirement Study (HRS) data.

We consider two main scenarios. First, we consider the effect of halting the current rise in obesity, diabetes, and hypertension, among 50 year-olds. We focus on these conditions, because they have attracted considerable policy attention in recent years. The second scenario considers the effect of raising the health of 50 year-old Americans to the level enjoyed by Europeans of the same age, but without affecting Americans' financial situation. In terms of the markers commonly collected in aging surveys, Europeans appear healthier than Americans (Banks et al., 2006).

The paper is structured as follow. We first describe current trends in health in the U.S. and compare the situation with that prevailing in Europe. In section 3, we then describe the model that is used to describe the long-term economic consequences of these trends. In Section 4, we present the scenarios we considered in this preliminary version of the paper and present some preliminary results.

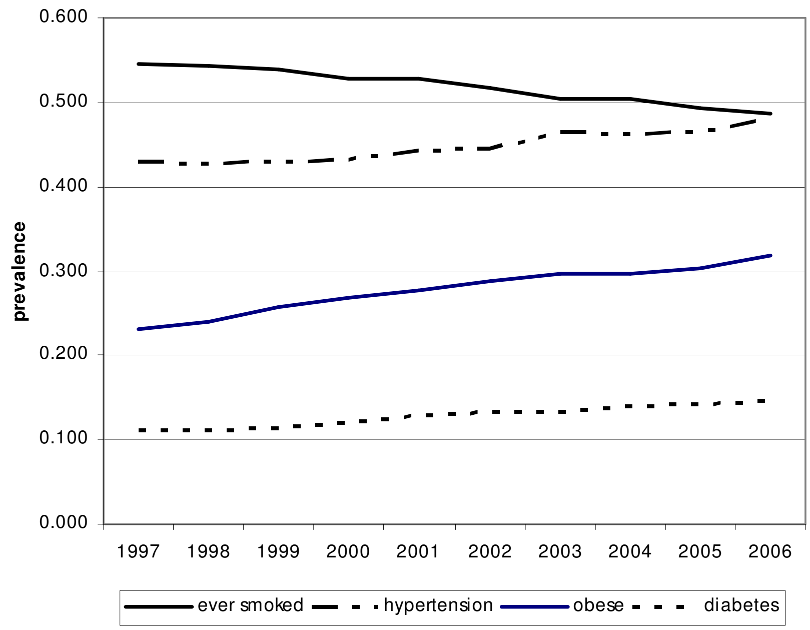
2. Health Trends in the U.S. and Cross-Country Comparison

The United States and other developed countries have experienced large gains in life-expectancy over the last century. For example, life expectancy at birth increased from 61 years in 1933 to 78 years in 2004.[2] The first 50 years of the 20th century were marked by a strong decline in infectious diseases which greatly decreased mortality rates, particularly for the young. In the second half of the century, medical technology has been a critical factor behind further improvements in life expectancy, by reducing mortality rates among older age groups. But over the last 50 years, chronic illnesses associated with more sedentary lifestyles have spread, somewhat mitigating those advances. These trends are particularly pronounced in the United States.

On the one hand, there are several important trends favorable for public health. Over the last 10 years, the prevalence of smoking has declined markedly. In addition, disability has generally receded in the elderly population in the 1980s and 1990s. On the other hand, however, is the large increase in the prevalence of obesity and its attendant consequences, including diabetes and hypertension (Lakdawalla et al., 2005). The next figure makes those recent trends clear by showing the prevalence of life-time smoking, obesity, diabetes and hypertension over the last 10 years among the elderly population. We use data from the National Heath Interview Surveys (NHIS) of 1997 to 2006 and focus on the population older than 50.

[2] Based on life tables collected in the Human Mortality Database (www.mortality.org).

Figure 1 U.S. Trends in Selected Health Outcomes



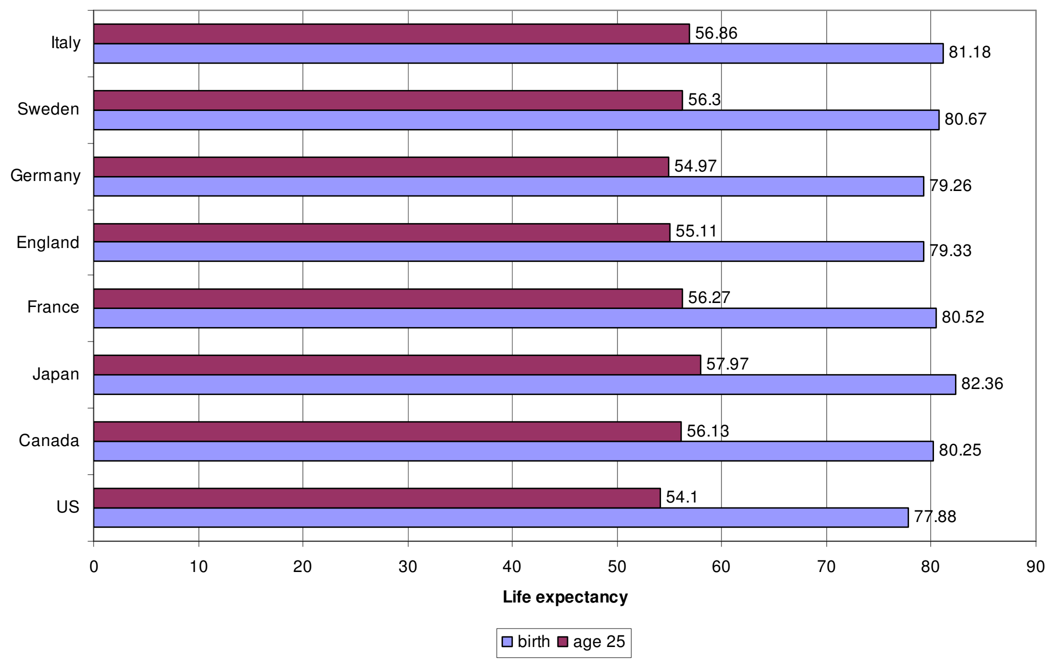
Source: National Health Interview Surveys (NHIS) 1997-2006, age 50+ population.

In 1997, the fraction of individuals 50 years of age or older who were obese (BMI of 30kg/m^2 or more) was 23.2%. In 2006, this fraction was 31.8%. Ruhm (2007) projects that if this trend continue, roughly one half of Americans between the ages of 45 and 64 will be obese by 2050, which places them at great risk of developing conditions such as diabetes, hypertension, heart disease and stroke. For example, only 5.5% of Americans aged 45-64 had Type 2 diabetes in 1980 according to the Centers for Disease Control and Prevention. But the prevalence has more than doubled in less than 25 years (10.2% in 2005).

On the other hand, the same graph makes clear that we continue to see a reduction in the prevalence of smoking, which will lead to less cancer and cardiovascular disease. As a consequence, future overal health trends remain uncertain, and will depend upon the interplay between favorable trends in smoking (and perhaps in elderly disability), and unfavorable trends in obesity and related conditions. At a minimum, the mix of diseases afflicting the elderly population is likely to change over the next 20 years. Hence, non-trivial effects on life-expectancy and disability are likely to occur.

While the future outlook for the US is unclear, the *current state* of US health is unquestionably poorer than that of other developed countries. The next figure first shows projected life-expectancy at birth and age 25 in the U.S. and a selected set of developed countries with similar economic standard of living. We select two ages, because it filters out the effect of infant mortality, which is higher in the US. Life expectancy at birth in the U.S. is 1.5-4.5 years shorter. At age 25, Americans can expect to live 54.1 years compared to 54.97 in Germany, 55.11 in England, and over 56 years in other countries.

Figure 2 Life Expectancy at Birth and Age 25 in Selected Countries



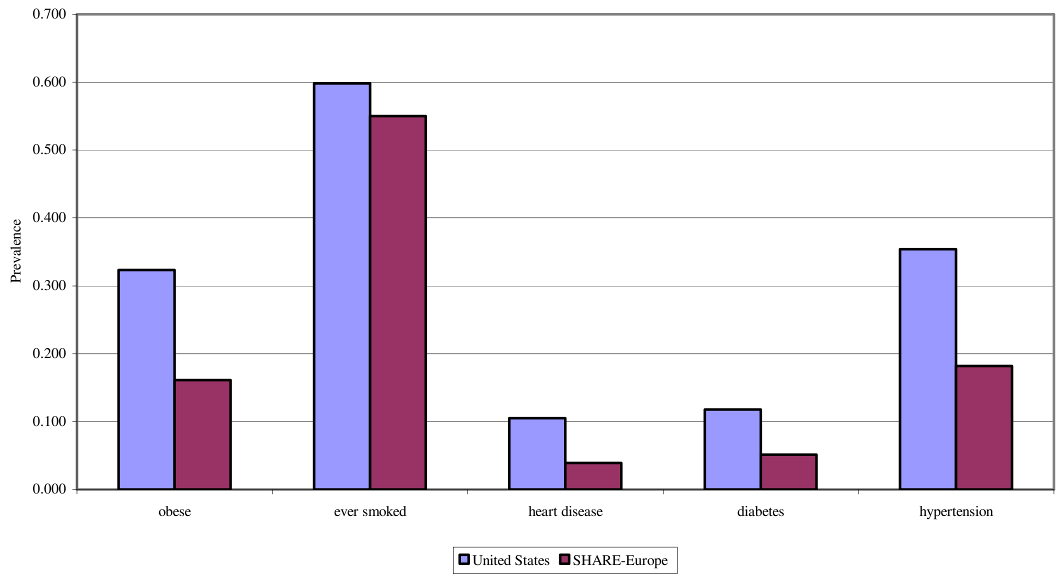
Source: Human Mortality Database, 2004 period life-table

Another way to compare health in the U.S. and Europe is to compare outcomes in so-called aging surveys, surveys that collect data on the age 50+ population, in a relatively consistent way. In the next figure, we compare the prevalence of obesity (using self-reported weight and height), life-time smoking, heart disease, diabetes and hypertension in the U.S. (using the HRS 2004 wave) and the Survey of Health Ageing and Retirement in Europe (SHARE 2004 wave) which covers 11 European countries.[3]

[3] For these figures, we use data from Austria, Belgium, Denmark, France, Germany, Italy, Spain, Sweden and Greece. More detail on the comparability of health measures in the Appendix, Table A.1.

We focus on a narrow age group around age 50 to control for differences in the age structure of these population (Europeans are on average 2-4 years older among the 50+ population).

Figure 3 Comparison of Key Health Indicators between the United States and European Countries in SHARE, Pop Aged 50-53



Source: HRS 2004 for U.S. and SHARE 2004 for Europe.

The differences are quite large in terms of obesity and associated illness. The prevalence of obesity in Europe is about half that in the U.S. Similar differences are observed for diseases such as heart disease, diabetes and hypertension. Lifetime prevalence of smoking is much more similar and in fact current smokers are more common in Europe than in the U.S. These differences, particularly those for chronic illness, do not necessarily reflect true differences in health (e.g. because of differences in diagnostic procedures) but they are suggestive. For example, Banks et al. (2006) show that even taking into account differences in rates of diagnosis, those living in England appear healthier than their American counterparts.

The relatively poor health of Americans and the uncertainty regarding projected trends in health raise important policy questions. First and foremost, what are the financial implications of those trends for Social Security, and for the economic well-being of the elderly? Second, does the uncertain direction of these trends translate into financial uncertainty for the government and for private entities, like pension plans and employers? Finally, what would be the long-term benefits of adopting healthier lifestyles in the U.S.? In the next section, we describe the model we used to answer these and related questions.

3. Microsimulation Model of Health and Economic Dynamics

3.1 Background

Current models of economic decisions in old age tend to either ignore health or have a very narrow view of health, often reduced to an indicator of whether health is self-reported "good" or "bad" (Gustman and Steinmeier, 1986; Rust and Phelan, 1997; French, 2005). Typically this constraint is imposed by computational limitations, rather than a modeling choice made by the analysts, since dynamic structural models of economic decisions make a more detailed treatment of health computationally prohibitive.

Dynamic structural models have the advantage of providing a clear and complete description of the system being modeled. Their drawback is the set of reductive simplifications needed in order to make such a complex model tractable. To provide a more detailed view of health dynamics than would be possible in a structural model, we propose a reduced-form approach to the problem. We model the evolution of various health and economic outcomes over time. In a sense, this resembles the approach taken by macroeconomists in the 1970s, who departed from rigid structural models of the economy towards vector autoregressive models that could be used to forecast GDP, inflation, etc (e.g. Sims, 1972). The added flexibility in such models allows one to consider a richer set of health as well as economic indicators.

While structural models impose significant computational requirements, reduced-form models impose different requirements. In particular, one needs estimates of behavioral responses to policy experiments; these are fed into the model to derive conclusions for how the entire system responds. A structural model generates its own behavioral responses, but a reduced-form model requires them as inputs.

Our strategy for estimating behavioral responses is to use the longitudinal nature of the data at our disposal. The trends we are analyzing are not new. US and European populations have been exposed to them for some time. This allows us to recover estimates of population responses. Hence, longitudinal data on these trends as well as associated trends in economic outcomes at the micro-level potentially provide an estimate of behavioral responses, which can be used to forecast future outcomes.

3.2 Functioning of the Model

3.2.1 Overview

The Future Elderly Model (FEM) was first developed at RAND to examine health and health care costs (Goldman et al., 2004). Initially, the model focused on the elderly Medicare population (age 65+), but we have now extended it to include all Americans aged 50 and older. The main defining characteristic of the model is that it considers real rather than synthetic cohorts in the analysis. The level of observation is at the individual level. This allows for more heterogeneity in behavior than would be allowed by a cell-based approach. Also, since the HRS interviews both respondent and spouse, we can link records to calculate household level outcomes such as net income and Social Security retirement benefits which depend on outcomes of both spouses.

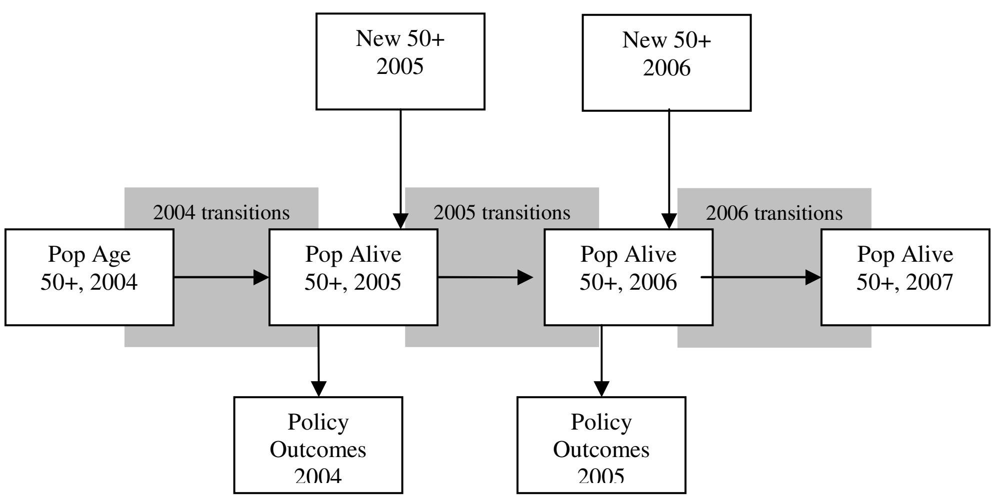
The model has three core components. First, there is an initial conditions model that predicts the financial and health outcomes of new cohorts of 50 year olds using data from the Health and Retirement Study (HRS) and trends calculated from other sources. This model allows us to "generate" cohorts as the simulation proceed forward and maintain focus on the age 50+ population in any given year.

The second component is a transition model which calculates the transition probabilities between various health states and financial outcomes. The probabilities depend on risk factors such as smoking, weight, age and education but also on lagged health states and financial outcomes. Hence, the model features a great deal of heterogeneity and allows for general feedback effects. These probabilities are estimated on longitudinal data from the Health and Retirement Study (HRS).

The last component aggregates outcomes into policy outcomes such as taxes, medical care costs, pension benefits paid, disability benefits. This last component takes account of program rules to the extent made possible by the outcomes being modeled. Because we have access to restricted data from Social Security records and Employers pension plans which is linked with the HRS, we are also able to model realistically retirement benefits that workers get once retired.

The next figure gives an overview of how the model works. We start in 2004 with an initial population aged 50+ taken from the HRS. We then predict outcomes using transition probabilities. Those who survive make it to the end of that year. At that point, we calculate policy outcomes for that year. We then enter the following year, at which point a new cohort of 50 year olds enters. This forms the age 50+ population for that year. This population then goes trough the transition model again. As in the preceding year, we calculate policy outcomes for this population in 2005. The same process is repeated until we reach the last year of the simulation.

Figure 4 Overview of the Future Elderly Model



3.2.2 Transition Model

The transition model tracks transition among states as a function of risk and demographic factors. Here we give a brief overview of the model and its estimation; the technical appendix provides much greater detail both on estimation and validation. The data come from the 1992 to 2004 biennial waves of the HRS. We consider two groups of outcomes: health and economic outcomes. Health conditions are based on the most prevalent conditions in the survey: hypertension, stroke, heart disease, lung disease, diabetes and cancer and come from questions of the form: "Has a doctor ever told you you had…." Consistent with the chronicity of these illnesses, we assume no recovery so that we model the time until diagnosis (after age 50).

Second, we consider measures of risk factors. We focus on smoking and obesity. We model transitions across three "obesity" states: not obese (BMI<27.5), overweight (27.5-30) and obese (BMI>30). This information is derived from self-reports on weight and height.[4] As for smoking, we model whether the respondent has never smoked, has ever smoked but has quitted or whether the respondent is still smoking at the time of the interview.

Third, we consider measures of disability commonly known as ADL (activities of daily living) limitations and IADL (instrumental activities of daily living) limitations. These are counts of positive answers to questions (5 for ADL limitations and 7 for IADL limitations) such as whether the respondent has trouble walking, getting out of bed, dressing up, etc. Based on the empirical distribution and some preliminary testing, we use whether the respondent has no IADL or ADL limitations, whether he has at most 2 IADL limitations but no ADL limitations (IADL are less severe), whether the respondent has 1 or 2 ADL limitations, or 3+ ADL limitations. This gives 4 mutually exclusive categories (based on the data). These are not assumed to be absorbing states such that we allow for recovery.

Finally, we observe mortality from exit interviews in the HRS. Mortality hazards derived from the HRS correspond closely with life-table probabilities (Adams et al., 2003; Kapteyn et al., 2006). We also track nursing home stays. The HRS initially did not

[4] We do not attempt to correct self-reports for measurement error (notoriously underreporting). This could be done in future work using for example estimated relationships between measured and self-reported outcomes from Cawley and Burkauser (2007).

sample from the nursing home population. However, it does follow respondents into nursing homes and also records transition back to independent or assisted living outside nursing homes.

We add to these outcomes a number of economic variables. We express all monetary units in terms of equivalent 2004 dollars using the Consumer Price Index (CPI). First, we track whether someone works for pay (any positive hours) and also earnings on the main job. We also track whether the respondent has health insurance on the job or derives coverage from other sources (other than Medicare). We do this for the population younger than age 65. After that age almost all respondents have insurance through Medicare. Someone can derive coverage from his employer, the employer of his/her spouse, from a private insurer or from a government assistance program such as Medicaid (if disabled). We also record whether someone has a pension on this job, and whether of the defined benefit (DB) or defined contribution (DC) type. If the respondent reports a DC pension, we use the self-report of the account balance. If respondent reports a DB pension, we record the earliest age at which the pension can be claimed as well as the number of years on the job and the normal retirement age on that plan. We also construct a variable recording whether someone is claiming a DB pension on the current job (quitting a job with a DB pension).

We model Social Security retirement benefit receipt using the self-report of the age at which benefits were first claimed. Because we have access to earnings records for respondents in the HRS, we also determine who is eligible upon reaching age 62 using quarters of coverage. We also construct the Average Indexed Monthly Earnings (AIME) for the initial interview and update the AIME in the simulation using a simplified rule as in French (2005). The AIME is the basis for computing benefits. We also model disability insurance (DI) benefit receipt and Supplemental Security Income (SSI) using self-reports from the HRS.

Finally, we construct a measure of net financial wealth using self-reports from the HRS and imputations performed by RAND (Hurd, Hoynes and Chand, 1998). Net financial wealth is defined as the value of financial assets (checking, savings, stocks, IRAs, Certificate deposits, bonds) plus the value of real assets (primary house, other real estate, other real assets) minus all debt (mortgage, home loans, credit cards, etc).

We make several restrictions on the way dynamic feedback effects are allowed. First, we only allow feedback from diseases where clinical knowledge is available that support such a link. For example, we allow hypertensive patients to have higher incidence risk of heart disease but we do not allow hypertensive patients to have higher risk of cancer. These restrictions are documented in the appendix and the methodology for deciding on which restrictions should be imposed is documented in Goldman et al. (2004).

Another important restriction we make is that economic outcomes do not have feedback effects on health. This is consistent with the findings from studies by Adams et al. (2003) and Michaud and van Soest (2008) that SES does not have a causal effect on health outcomes in this age range. Yet, a correlation between SES is observed because of two important factors: first feedback effects from health to economic factors, most notable trough the effect of health deterioration on the capacity to work and on health care spending. The second factor is predetermined (earlier) events or common factors (genetics, etc) that create a correlation between SES and health in this age group without there being any causal effect. These two factors are taken into account in the estimation. The next table gives an overview of the outcomes we consider.

Table 1 Outcomes in the Transition Model

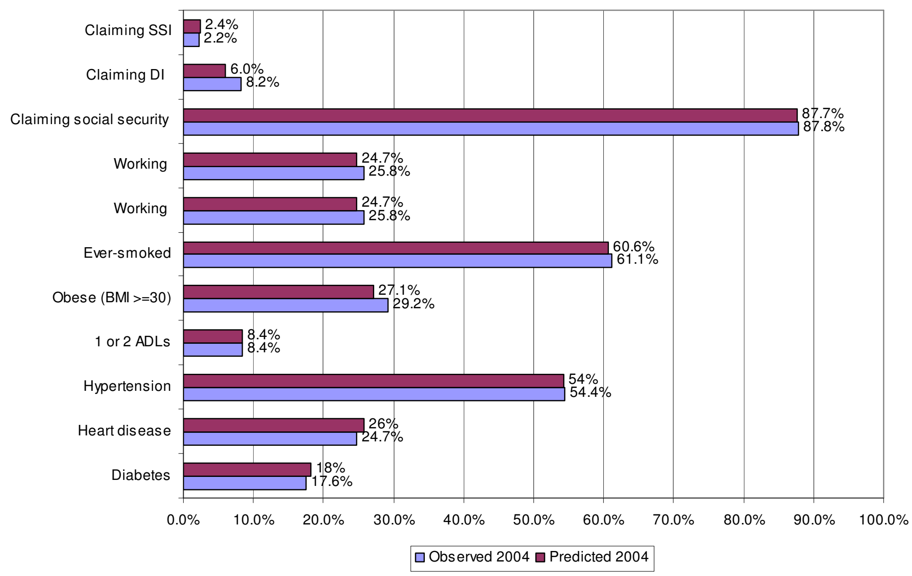
Health	**SES & Other**
Disease	**LFP & Benefit Status**
heart disease	working
hypertension	DB pension receipt
stroke	SS benefit receipt
lung disease	DI benefit receipt
cancer	Any Health insurance
diabetes	ssi receipt
Risk factors	**Financial Resources**
Smoking Status	financial wealth
never smoked	earnings
ever smoked	wealth positive
current smoker	
BMI Status	**nursing home residence**
normal	**death**
overweight	
obese	
Functional status	
No ADL	
iADL only	
1-2 ADL	
3+ ADL	

Notes: See Appendix for more details

Transition rates are allowed to differ across demographic and economic groups. In particular, we allow differences by gender, race and ethnicity, education, and marital status. Transition equations are estimated on 7 waves of the HRS. To assess the fit, we use estimated transition probabilities to simulate what 1992 HRS respondents would look like in 2004. We then compare to actual outcomes. We use a half the sample for estimation and the other half for simulation. The fit from the estimated transition equations is satisfactory. The next figure shows selected comparisons. More details on this exercise can be found in the appendix.

In general, the model fits the data quite well, with a close correspondence between predicted and actual outcomes in most areas including labor supply, claiming behavior, functional status, and diseases. Some areas remain for improvement, including obesity. Most notably though, the fit of the predictions is less than satisfactory for wealth. The main reason is the peculiar distribution of wealth. Models for the mean will then to lead to overprediction of the median, even when transformed using the log transformation for example. In future drafts, we will experiment with other transformations.

Figure 5 Comparison of Predicted and Observed Outcomes in 2004 for HRS Respondents First Observed in 1992



3.2.3 New Cohorts of 50 year-olds

Because the FEM simulations aim to represent outcomes for the age 50+ population, we need to predict how the population of 50 year old will look like in terms of health, demographics and economic outcomes in future years. The HRS makes this prediction difficult because the youngest members of the sample are age 50 in 2004. One way to accommodate those changes is to reassign current age 50 respondents who were not obese to be obese in the future. This way, the new age 50 population will have a higher obesity rate. But we cannot re-assign in a totally random fashion because obesity is correlated with other outcomes such as hypertension and diabetes. Hence, the new obese are more likely to come from those with hypertension than those without. In summary, we have two challenges. One is to accurately predict trends for each outcome over time. The other is to reassign outcomes in a way which is consistent with the correlations observed in the data.

We estimate the model jointly using maximum likelihood methods to preserve the correlation between outcomes. In the next table, we list the outcomes that represent the characteristics of the entering cohort that we need to model. We have 7 binary outcomes representing disease prevalence, but also labor force status, insurance status and positive wealth. The inclusion of this last indicator is necessary because of the observed spike in the distribution of net wealth at zero in the HRS. We also consider three ordered outcomes, BMI status, smoking status and functional status defined in a similar way as in the transition model. We consider 5 continuous outcomes. The average indexed monthly earnings (AIME) and the number of quarters of coverage are the basis for the computation of Social Security benefits. Earnings, financial wealth and DC wealth are necessary as covariates in the transition model. We also model characteristics of pension plans respondents have. First, whether they have a DB or DC plan on the current job. Then, the earliest age at which they are eligible and the normal retirement age. We group these around peaks in the empirical distribution. The most common early retirement age is 55 and the most common normal retirement age is 62. Finally, we make each of those outcomes a function of fixed characteristics such as race, education, gender and marital status. We also consider cancer, lung disease and stroke as fixed because their prevalence is very low in this population (age 50-53).

Table 2 Outcomes in the Model for New Cohorts

Binary	**Censored Discrete**
working for pay	Any db plan
positive wealth	Any dc plan
hypertension	**Censored Ordered**
heart disease	Early Age eligible DB
diabetes	<52
any health insurance	52-57
SRH fair/poor	58>
	Normal Age eligble DB
Ordered	<57
BMI status	57-61
normal	62-63
overweight	64>
obese	**Covariates**
Smoking status	hispanic
never smoked	black
ever smoked	male
current smoker	less HS
Functional status	college
No ADL	single
iADL only	widowed
1+ ADL	cancer
Continuous	lung disease
AIME (Nominal $USD)	stroke
quarters of coverage	
earnings	
wealth	
dc wealth	

Notes: More detail in appendix.

We then calculate trends from two sources. First, we use the method described in Goldman et al. (2004) to calculate trends in disease prevalence from the National Health Interview Surveys (NHIS). The estimated trends are relatively close to other independent estimates of those trends as documented in the appendix. For other outcomes, we use various sources also documented in the appendix. An overview of the trends used is given in the following table, which shows the prevalence of health conditions in future cohorts of 50-year olds until 2050 (relative to prevalence in 2004) .[5] Our base case

[5] Although more trends are documented in the appendix (DB/DC wealth for example), they are not implemented in this draft.

assumes continued increases in hypertension, diabetes, and (especially) obesity, and reductions in smoking and cardiovascular disease.

Table 3 Projected Trends in Health Outcomes

	Prevalence/Means relative to year 2004					
Outcome	2004	2010	2020	2030	2040	2050
Prevalence of binary outcomes						
hypertension	1	1.04	1.07	1.09	1.11	1.13
heart disease	1	0.95	0.91	0.88	0.85	0.83
diabetes	1	1.12	1.22	1.27	1.31	1.36
Prevalence of the highest category of ordered outcomes						
BMI status - obesity	1	1.11	1.34	1.48	1.63	1.8
Smoking status - smoking now	1	0.94	0.73	0.6	0.5	0.41

Notes: See Appendix. for details on trends.

The size of the entering cohort is adjusted to reflect population projections from Census by gender and race. We also post-stratify the size of the initial new cohort in 2004 to the Current Population Survey (CPS).

3.2.4 Policy Outcomes

The model outputs a number of relevant health and economic outcomes. First, we consider a set of health outcomes such as life-expectancy and disability free life expectancy at age 50 and medical expenditures. Average medical expenditures by disease by demographic groups are calculated from two sources. For those younger than age 65, we use the Medical Expenditure Panel Survey (MEPS) and include in medical expenditures but medical care costs and the cost of drugs. Since the MEPS is known to under-predict expenditures, we make an adjustment based on comparing MEPS average cost for those older than 65 with a more reliable source, the Medical Current Beneficiary Survey (MCBS). This adjustment results in a 15% upward revision in average cost. More detail can be found in the Appendix.

A second set of outcomes consist of revenues and expenditures by the Federal Government and Social Security for the age 50+ population. First, we compute Social Security retirement benefits for those predicted to receive such benefits. Since we have the AIME of both respondent and spouse, we can get quite precise estimation of the distribution of retirement benefits. We also compute disability insurance (DI) benefits and Supplemental Security Income (SSI). We compute DB pension income from average pension payments by tenure, last earnings and early and normal retirement ages using Pension Plan characteristics reported by Employers of HRS respondents. Using all these income flows, for the respondent and the spouse (if present), we compute net household income where taxes include Federal, State, City income taxes as well as the Social Security taxes (OASDI and Medicare). More detail in the appendix.

3.2.5 Simulation Methods

The simulation starts with the existing age 50+ population in the 2004 wave of the HRS. The microsimulation is stochastic, meaning that transitions are random conditional on deterministic relationships estimated from the HRS. There are two stochastic elements in the model: the simulation of new cohorts and the simulation of transitions. To generate new cohorts, we make use of the estimated joint distribution and use random draws from the correlated errors to generate a new cohort. Second, for each respondent in a given year, we calculate transition probabilities. We then draw random numbers to attribute new outcomes. This process is repeated a number of times to ensure that it is not dependent on any particular sequence of random numbers. We choose 10 replications for this draft. We simply average results over those replications when reporting the results. In future drafts we will use more replications and also assess the uncertainty in our results due to the finite number of draws we take.

4. Scenarios

4.1 Counterfactuals

We consider two scenarios in this preliminary version.

First, we simulate the model projecting trends for the new cohorts. We document these trends in the Appendix. For simplicity, we consider only the trends in hypertension, diabetes, and obesity. We then assume no trends─that is, that each entering cohort of future 50-year olds look just like those in the 2004 cohort─and re-simulate the models. This amounts to simulating the effect of halting the trends in these outcomes.

In a second scenario, we adjust the health of Americans such that its resembles that of Europeans using data from SHARE. We can do that using the estimated joint distribution for new cohorts. The numbers we use are given in the table below.

Table 4 Initial Conditions for Health in the U.S. and SHARE-Europe

Health outcomes	United States	SHARE-Europe	Ratio EU/US
obese	0.323	0.161	0.498
ever smoked	0.598	0.550	0.920
current smoker	0.249	0.299	1.200
heart disease	0.105	0.039	0.372
diabetes	0.118	0.051	0.434
stroke	0.025	0.015	0.598
lung disease	0.055	0.029	0.523
cancer	0.050	0.030	0.597
hypertension	0.354	0.182	0.514

4.2 Preliminary Results

4.2.1 Baseline estimates and projection

We first present results for the baseline scenario where projected trends are realized. Table 5 shows the results.

Table 5 Baseline projection for the population aged 50 and plus in the U.S.

	Year		
	2004	2030	2050
Population size (Million)	80.71	120.00	145.01
Prevalence of selected conditions			
obesity (BMI >=30) (%)	28.3%	41.4%	47.4%
over weight (25<=BMI<30) (%)	38.0%	39.1%	36.3%
Diabetes	17.5%	25.0%	28.8%
Heart disease	23.8%	28.2%	31.3%
Hypertension	52.0%	60.5%	64.8%
Labor participation among age <75			
Working (%)	45.9%	39.7%	42.0%
Government revenues from aged 50+ (Billion dollars)			
Federal personal income taxes	364.74	483.49	535.14

States personal income taxes	109.26	172.08	202.97
Local personal income taxes	71.62	112.45	132.91
Social security payroll taxes	120.10	130.67	149.55
Medicare payroll taxes	30.43	35.02	39.81
Government expenditures from aged 50+ (Billion dollars)			
Old Age and Survivors Insurance benefits (OASI)	393.16	1,248.03	1,617.63
Disability Insurance benefits (DI)	41.99	42.43	58.02
Supplementary Security Income (DI)	8.59	19.29	21.85
Medicare costs	296.55	564.68	785.07
Medicaid costs	91.46	125.26	208.96
Total medical costs for aged 50+ (Billion dollars)	971.28	1,604.40	2,157.39

Note: All dollars are in 2004 values

The population size at year 2050 is projected to be 145.0 million. The size of the population (not shown here) is projected to be 82.0 million at year 2050, very close to the Middle Series projection of the 2008 OASDI Trustees Report, which is 80.8 million. We also compare our results on government expenditures with other data sources whenever applicable. The estimated OASI outlays in 2004 are $393 billion, while the OASI benefits paid in 2004, released by Social Security, was $415 billion (Social Security Administration, Annual Statistical Supplement, 2005). Our estimation is lower for a good reason: we do not include children's benefits, which were about $17.3 billion at year 2004 (Social Security Administration, Annual Statistical Supplement, 2005). The Medicare costs are $297 billion, close to the government estimate of $308 billion (CMS, HI and SMI Trustee Report 2005). We miss the Medicare expenditure for those younger than age 50 and disabled. According to the National Health Expenditure (NHE) data, the estimated that Medicare expenditure for younger than age 55 was $21 billion in 2004. We under-estimate Medicaid costs relative the NHE estimate. We estimated that the costs for aged 50+ was $91.5 billion, while the NHE estimates that the costs for aged 55+ was $103 billion. We will investigate this issue further. In addition, we find that the estimated population size receiving DI and monthly DI benefits are both close to the Social Security Administration (SSA) data. Finally, we currently under-predict the SSI population. Our estimation of the SSI population among aged 65 and over is much lower than the SSA data, due to underreporting of SSI status in the HRS. We will account for this underreporting in future models.

On the revenue side, we compare the estimated federal personal income taxes with other data sources. since we only estimate taxes for aged 50 and over, we need to estimate what share of taxes come from the 50+ population. We use the March Supplement of the Current Population Survey, 2005 to estimate what proportion of federal personal income taxes come from the 50+ population. We should keep in mind that the estimated federal income taxes in CPS are based on self-reported income and are calculated using a tax simulation model. According to the 2003 CPS data, this proportion is 37%. While the total federal personal income taxes in year 2004 was $832 billion (Table 468, the 2008 Statistical Abstract, Census Bureau). Therefore in 2004 approximately $308 billion came from the 50+ population. Our estimation is higher: $364 billion.

4.2.2 Counterfactual scenarios – no obesity trend

We now move on to our two counterfactual scenarios. We first present results for the first scenario in which we shut down the trends in obesity related outcomes (obesity, diabetes, hypertension and heart disease). For simplicity we call this the "no obesity trend" scenario. Table 6 shows the results for the projected outcomes for aged 50 plus population in every year. All outcomes are presented as the percentage change relative to the baseline scenario. Table -7 shows the simulated lifetime outcomes for those who turn to age 50/52 at different time points – 2004, 2030, 2050. Both demographics (gender composition, race/ethnicity composition) and health status will change according to the projected trends shown in Table 3.

Table -6. Percentage change of the "no obesity trend" scenario among 50+

	Year	
	2030	2050
Population size	0.07%	0.48%
Prevalence of selected conditions		
Obesity (BMI >=30) (%)	-15.0%	-26.2%
Over-weight (25<=BMI<30) (%)	5.1%	11.8%
Diabetes	-8.9%	-17.4%
Heart disease	-1.2%	-3.0%
Hypertension	-3.1%	-5.1%
Labor participation among age <75		
Working (%)	0.2%	0.2%
Government revenues from aged 50+		
Federal personal income taxes	0.09%	0.84%

States personal income taxes	0.10%	0.58%
Local personal income taxes	0.10%	0.57%
Social security payroll taxes	0.22%	1.02%
Medicare payroll taxes	0.21%	1.08%
Government expenditures from aged 50+		
Old Age and Survivors Insurance benefits (OASI)	-0.04%	0.00%
Disability Insurance benefits (DI)	0.46%	-2.69%
Supplementary Security Income (DI)	-1.35%	-2.43%
Medicare costs	-1.42%	-3.50%
Medicaid costs	-3.48%	-4.95%
Total medical costs for aged 50+	-1.62%	-3.20%

Table -7. Percentage change of the "no obese trend" scenario for future aged 50

Year turning into age 50	Scenarios	Lifetime outcomes at age 50				
		Average years lived since age 50	Average disability-free years lived since age 50	Lifetime government revenues ($)	Lifetime government expenditures ($)	Lifetime total medical costs ($)
2004	Status-quo	31.91	29.39	193376	288479	241516
	No obese trend	31.91	29.39	193376	288479	241516
	Relative change (%)	0.0%	0.0%	0.0%	0.0%	0.0%
2030	Status-quo	32.93	29.28	188403	300919	251356
	No obese trend	33.21	30.04	189086	296988	241364
	Relative change (%)	0.8%	2.6%	0.4%	-1.3%	-4.0%
2050	Status-quo	33.87	29.78	182186	317821	266482
	No obese trend	34.55	30.89	185384	317172	256809
	Relative change (%)	2.0%	3.7%	1.8%	-0.2%	-3.6%

* Dollars are in 2004 values and are discounted by 3% per year

Note: "Government revenues" include personal income taxes at the federal, state and local level, as well as social security and Medicare payroll taxes; "Government expenditures" include outlays of the following government program: OASI, DI, SSI, Medicare, and Medicaid.

According to Table -6, under the "no obese trend" scenario, the prevalence of obesity at year 2050 will be 26% lower, relative to the status quo. Part of the obese population will be over-weight instead. Therefore the prevalence of "over-weight" under the "no obese trend" scenario will be higher. The prevalence of hypertension is 17% lower while the prevalence of heart disease and diabetes only decrease by a small percent: 3.0% and 5.1%, respectively.

As table -7 shows, the effects of "no obesity trend" on mortality and life expectancy are less evident than the effect on morbidity. Under the "no obese trend" scenario, the 2050 cohort aged 50-year olds will live 33.87 years on average, 2.0% higher than the baseline, but the years of life without disability will increase by a larger percentage: 3.7%.

There are also improvements of "no obese trends" on economic outcomes, since workers are now in better health. According to Table -6, labor participation rate among those aged younger than 75 at year 2050 is 0.2% higher than the baseline. Table -7 shows that reducing obesity and the related health conditions (diabetes and hypertension) will be a net-benefit to the government financial situation – lifetime government revenues increase while lifetime government expenditure decrease. And the reduction effect on lifetime medical expenditures is the most significant.

4.2.2 Counterfactual scenarios – European health at year 2004, with US trend up to to 2050

We now consider the second scenario which effectively gives the 2004 age 50-52 Americans the health of Europeans of the same age, but letting the health trends unchanged. Table 8 and Table 9 show the results. For simplicity we will denote this scenario as the "European health" scenario.

Table -8. Percentage change of the "European health" scenario among 50+

	Year	
	2030	2050
Population size	0.71%	2.08%
Prevalence of selected conditions		
Obesity (BMI >=30) (%)	-43.3%	-51.1%
Over weight (25<=BMI<30) (%)	-3.9%	0.6%
Diabetes	-30.3%	-40.2%
Heart disease	-17.0%	-20.2%
Hypertension	-19.7%	-20.9%
Labor participation		
Working (%)	0.5%	0.8%
Government revenues from aged 50+		
Federal personal income taxes	-0.17%	2.40%
States personal income taxes	0.25%	1.66%
Local personal income taxes	0.26%	1.67%
Social security payroll taxes	0.99%	2.12%
Medicare payroll taxes	0.91%	2.32%
Government expenditures from aged 50+		
Old Age and Survivors Insurance benefits (OASI)	0.64%	1.61%
Disability Insurance benefits (DI)	-15.74%	-20.77%
Supplementary Security Income (DI)	-6.57%	-4.11%
Medicare costs	-9.11%	-12.48%
Medicaid costs	-16.28%	-14.04%
Total medical costs for aged 50+	-9.65%	-11.32%

Table -9. Percentage change of the "European health" scenario for future aged 50

Year turning into age 50	Scenarios	Lifetime outcomes				
		Average years lived since age 50	Average disability-free years lived since age 50	Lifetime government revenues ($)	Lifetime government expenditures ($)	Lifetime total medical costs ($)
2004	Status-quo	31.91	29.39	193376	288479	241516
	European health	32.49	30.42	194304	276897	214049
	Relative change (%)	1.8%	3.5%	0.5%	-4.0%	-11.4%
2030	Status-quo	32.93	29.28	188403	300919	251356
	European health	33.87	31.01	192644	291135	219525
	Relative change (%)	2.8%	5.9%	2.3%	-3.3%	-12.7%
2050	Status-quo	33.87	29.78	182186	317821	266482
	European health	35.22	31.76	187394	309957	233724
	Relative change (%)	4.0%	6.7%	2.9%	-2.5%	-12.3%

* Dollars are in 2004 values and are discounted by 3% per year

Note: "Government revenues" include personal income taxes at the federal, state and local level, as well as social security and Medicare payroll taxes; "Government expenditures" include outlays of the following government program: OASI, DI, SSI, Medicare, and Medicaid.

According to table 8, under the "European health" scenario, there are significant improvements in health conditions. At year 2050, the prevalence of obesity is reduced by half, while the prevalence of diabetes is reduced by 40%, prevalence of hypertension and heart disease is reduced by 20%.

Because of much better health, there are also larger magnitudes of improvements in economic outcomes. Labor participation rate increases slightly (0.8% at year 2050). However, labor force participation rates in early ages increase more substantially such that the average retirement increases by close to 6 months. As for lifetime analysis, under the "European health" scenario, an aged 50 year olds in year 2050 will contribute 2.9% more to the government, while receiving 2.5% less. And the decrease in total medical spending is very large: 12.7%. Overall, this scenario shows large positive financial consequences for the government.

5. Preliminary Conclusions

In this paper, we attempt to construct a microsimulation model that can accurately predict the long-term financial consequences of trends in Population health for the U.S.

Our preliminary results suggest that trends in population health have important consequences for economic well-being as well as large financial consequences for governments. A number of improvements can be made to the model to deal with such issues as underreporting of program participation and problems related with forecasting the distribution of financial wealth. Furthermore, we plan to incorporate correlation in the behavior of spouses so as to replicate the distribution of household income in a more satisfactory manner.

References

Adams, P., M. Hurd, D. McFadden, A. Merrill and T. Ribeiro (2003): "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status" *Journal of Econometrics* (2003), 112, 3-56.

Banks, J., M. Marmot, J. Smith and Z. Oldfied (2006) : "Disease and Disadvantage in the United States and in England," Journal of the American Medical Association (JAMA) 295(17):2037-2045.

Bhattacharya, J., D. Cutler, D. Goldman, M. Hurd, G. Joyce, D. Lakdawalla, C. Panis and B. Shang Disability Forecasts and Future Medicare Costs, *Frontiers in Health Policy Research* 7, Ed. D. Cutler and A. Garber, Boston, MA; MIT Press, 2004.

Cawley, J. and R. Burkhauser (2006), "Beyond BMI: The Value of More Accurate Measures of Fatness and Obesity in Social Science Research," NBER Working Paper 12291, National Bureau of Economic Research, Cambridge MA.

A. Kapteyn, A. van Soest and J. Smith (2007): Vignettes and Self-Reported Work Disability in the US and The Netherlands," *American Economic Review* 97(1):461-473.

French, E. (2005), "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior," *Review of Economic Studies*

Goldman, D., M. Hurd, P. Shekelle, S. Newberry, C. Panis, B. Shang, J. Bhattacharya, G. Joyce and D. Lakdawalla (2004). "Health status and medical treatment of the future elderly: Final Report." RAND Corporation Technical Report TR-169. Santa Monica, CA: RAND Corporation.

Goldman, D.P., B. Shang, J. Bhattacharya, A. M.Garber, M. Hurd, G.F. Joyce, D.N. Lakdawalla, C. Panis, and P.G. Shekelle (2005). "Consequences Of Health Trends And Medical Innovation For The Future Elderly," *Health Affairs*, W5, pp. R5-R17.

Gustman, A.L., and T. L. Steinmeier (1986). "A Structural Retirement Model," *Econometrica*, Vol. 54, No. 3. (May, 1986), pp. 555-584.

Hoynes, H., Hurd, M. and H. Chand (1998): Household Wealth of the Elderly under Alternative Imputation Procedures, in Inquiries in the Economics of Aging, David Wise (ed.), University of Chicago Press, Chicago, pp. 229-257.

Kapteyn, A., Michaud, P.-C., Smith, J. P., & Van Soest, A. (2006). *Effects of attrition and non-response in the Health and Retirement Study* (Working Paper No. WR-407). Santa Monica, CA: RAND Corporation

Lakdawalla DN, Goldman DP, Shang B. The health and cost consequences of obesity among the future elderly. *Health Aff (Millwood).* 2005;24 Suppl 2:W5R30-41.

Mokdad AH, Ford ES, Bowman BA, et al. Diabetes trends in the U.S.: 1990-1998. *Diabetes Care.* Sep 2000;23(9):1278-1283.

Michaud, P.-C. and A. van Soest (2008): "Health and Wealth of Elderly Couples: Causality Tests using Dynamic Panel Data Models", forthcoming *Journal of Health Economics*.

Rust, J. and C. Phelan, (1997), "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets" *Econometrica*, 781-832 Vol. 65 (4). Ruhm, Christopher J., "Current and Future Prevalence of Obesity and Severe Obesity in the United States", *Forum for Health Economics and Policy*, Vol. 10, No. 2 (Obesity), Article 6, 2007, 1-26.

Sims, C. (1972): "Money, Income and Causality", *American Economic Review* 62 540-52.

Appendix

Data Sources

Health and Retirement Study

The Health and Retirement Study (HRS), waves 1992-2004 are used to estimate the transition model. We use the dataset created by RAND (RAND HRS) as our basis for the analysis. We use all cohorts in the analysis and consider sampling weights whenever appropriate.

Social Security Covered Earnings Files

To get information on Social Security entitlements of respondents, we match the HRS data to the Social Security Covered Earnings files of 1992, 1993, 1998 and 2004 which provides information on earning histories of respondents as well as their entitlement to future Social Security benefits. We then construct the Average indexed monthly earnings (AIME) from these earning histories. The AIME is constructed by first indexing using the National Wage Index to the wage level when the respondent turns age 60. If this occurs after 2008, we project the evolution of the NWI using the average annual rate of change of the last 20 years (2.9% nominal). We then take the 35 highest years (if less than 35 years available remaining ones are zero earning years) and take the average. We then convert back this annual amount on a monthly basis and convert back to $2004 U.S. dollars using the CPI. Quarters of coverage, which determine eligibility to Social Security are defined as the sum of posted quarters to the file. A worker is eligible to Social Security if he has accumulated at least 40 quarters of coverage. A worker roughly accumulates a quarter of coverage for every $4000 of coverage earnings up to a maximum of 4 per year.

European Data for New Cohorts

In order to construct a scenario for the health of 50 year old Americans resembles that of Europeans, we use the Survey of Health, Ageing and Retirement in Europe (SHARE) which mimics closely the HRS in its questionnaire as well as design. In its first wave in 2004, SHARE sampled the age 50+ population in 10 European countries. We use release 1 of the data.

Data related to health conditions is relatively similar in the two datasets.

Table A.1 Comparison of Questions on Chronic Health Conditions in HRS and SHARE

	HRS	SHARE
Question	Has a doctor ever told you that you have …	Has a doctor ever told you that you had any of the conditions on this card? Please tell me the number or numbers of the
Heart disease	...a heart attack, coronary heartdisease, angina, congestive heart failure, or other heart problems?	...A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure
Hypertension	...high blood pressure or hypertension?	...High blood pressure or hypertension
Stroke	...a stroke?	...A stroke or cerebral vascular disease
Diabetes	...diabetes or high blood sugar?	...Diabetes or high blood sugar
Lung disease	...chronic lung disease such aschronic bronchitis or emphysema?	...Chronic lung disease such as chronic bronchitis or emphysema
Cancer	...cancer or a malignant tumor, excluding minor skin cancers?	...Cancer or malignant tumour, including leukaemia or lymphoma, but excluding minor skin cancers

Notes: Extract from Questionnaire from each survey.

Table A.2 Data sources and methods for projecting trends for future Cohorts

	Data source	Projection method	Directly obtained from other sources
Chronic conditions	National Health Interview Survey 1997-2006 cohort-mortality rate from www.mortality.org	Assume no recovery	
Cancer Diabetes Heart Hypertension Lung Stroke	Assumed annual mortality improvement rate for year 2005-2030: 0.8% per year Assume relative risks of mortality for each condition: rr = 2 for cancer, diabetes and heart rr = 1.5 for hypertension, lung and stroke	Use synthetic cohort approach to estimate age-specific incidence rate for each condition, using NHIS 1997-2006 Baseline prevalence is obtained from the NHIS 2003-2005 pooled data Use Markov model to model the transition into a certain condition or die from 2005 to 2030	There are other forecasts (Honeycutt, 2003, Mainous 2007) for the trends of diabetes in the U.S population; we compare their forecasts to ours and they are reasonably close
Over-weight and obese	Prevalence of over-weight and obese for aged 46-56 from year 2001 to 2030, generated by Ruhm upon request		Ruhm, Christopher J., "Current and Future Prevalence of Obesity and Severe Obesity in the United States", *Forum for Health Economics and Policy*, Vol. 10, No. 2 (Obesity), Article 6, 2007, 1-26.
Ever-smoked and smoking now	1) Status quo - Tobacco control policies will be frozen in place as of the beginning of 2006, with excise tax rates assumed to be adjusted for inflation. 2) "Extreme scenario": - No new initiation under the age of 18 - A doubling of the quit rate for adults under 40 - Prevalence is cut in half by 2010 3) "Super extreme scenario" - under this scenario there is an immediate smoking cessation in year 2006 and there is no new initiation at any age. (In all four scenarios it is assumed that no one initiates smoking after age 24).		Forecast of prevalence of ever-smoked and smoking now for aged 45-54 from year 2005 to 2025, by David Levy (2006)
Any DB from current job	Prevalence of DB entitlement from current job among aged 50-55, in HRS 1992 and 2004		Historical trends of DB participation rates among all

		persons by different birth cohorts and by age, by Poberta 2007 (a)
Any DC from current job	Prevalence of DC entitlement from current job among aged 50-55, in HRS 1992 and 2004	Forecast of DC participation rates among all persons by different birth cohorts and by age, by Poberta 2007 (b)
Hispanic Non-Hispanic black Population size 50-52	Projection of population from US census Bureau, Interim projection consistent with 2000 census (2004) Projection of population from US census Bureau, middle series, final projection consistent with 1990 census (2000) Since the interim projection consistent with 2000 census doesn't provide projection for all race/ethnicity categories, we cannot obtain the projection of non-Hispanic black population. As a result I turn to the final projection consistent with 1990 census and find out what proportion of the black population is non-Hispanic and the proportion is approximately 95%. http://www.census.gov/population/www/projections/natdet-D5.html	

Table A.3 Projected Trends in the U.S.

	Outcome	Prevalence/Means relative to year 2004					
		2004	2010	2020	2030	2040	2050
Prevalence of binary outcomes							
1	working for pay	1	1	1	1	1	1
2	positive wealth	1	1	1	1	1	1
3	hypertension	1	1.04	1.07	1.09	1.11	1.13
4	heart disease	1	0.95	0.91	0.88	0.85	0.83
5	diabetes	1	1.12	1.22	1.27	1.31	1.36
6	any health insurance	1	1	1	1	1	1
7	SRH fair/poor	1	1	1	1	1	1
Prevalence of the highest category of ordered outcomes							
8	BMI status - obesity	1	1.11	1.34	1.48	1.63	1.8
9	Smoking status - smoking now	1	0.94	0.73	0.6	0.5	0.41
10	Functional status - 1 or more ADLs	1	1	1	1	1	1
Means of continuous outcomes							
11	AIME	1	1	1	1	1	1
12	quarters of coverage	1	1	1	1	1	1
Means of censored continuous outcomes							
13	log earnings	1	1	1	1	1	1
14	log net wealth	1	1	1	1	1	1
15	log dc wealth	1	1	1	1	1	1
Prevalence of censored discrete outcomes							
16	Any db plan	1	0.89	0.72	0.59	0.48	0.39
17	Any dc plan	1	1.14	1.41	1.56	1.56	1.56
Prevalence of censored ordered outcomes							
18	Early Age eligible DB	1	1	1	1	1	1
19	Normal Age eligble DB	1	1	1	1	1	1

Notes: See Table A.1. for details on trends.

The next table gives the set of outcomes considered for the transition model along with descriptive statistics and the population at risk when estimating the relationships.

Table A.4 Outcomes in the Transition Model

Health outcomes				**Economic Outcomes**			
	Type	mean/fraction	At risk		Type	mean fraction	At risk
Disease				**LFP & Benefit Status**			
heart disease	incidence	0.043	undiagnosed	working	prevalence	0.413	age<75
hypertension	incidence	0.080	undiagnosed	DB pension receipt	incidence	0.117	ligible & not receivin
stroke	incidence	0.017	undiagnosed	SS benefit receipt	incidence	0.525	eligible not receiving
lung disease	incidence	0.019	undiagnosed	DI benefit receipt	prevalence	0.075	age<65 & eligible
cancer	incidence	0.025	undiagnosed	Any Health insurance	prevalence	0.897	age<65
diabetes	incidence	0.026	undiagnosed	ssi receipt	prevalence		if disabled or age>=6
Risk factors				**Financial Resources**		median $USD 2004	
Smoking Status	ordered		all	financial wealth	log		all positive wealth
never smoked		0.400		earnings	log		all working
ever smoked		0.440		wealth positive	prevalence		all
current smoker		0.160					
BMI Status	ordered		all	**nusring home res**	prevalence	0.014	all
normal		0.369		**death**	incidence	0.055	all
overweight		0.390					
obese		0.241					
Functional status	ordered		all				
No ADL		0.815					
iADL only		0.035					
1-2 ADL		0.107					
3+ ADL		0.043					

Notes: Statistics unweighted on sample 1992-2004. Statistic for incidence variable is the average biannual incidence rate.

The next two tables document restrictions placed on the transition model

Table A.5 Restrictions on Health Outcomes

	Outcome at (t)									
	heart	blood pressure	stroke	lung disease	diabetes	cancer	disability	mortality	Smoking	BMI
Prevalence of Contiions at t-1										
heart			x				x	x	x	x
blood pressure	x		x				x	x	x	x
stroke							x	x	x	x
lung disease							x	x	x	x
diabetes	x	x	x				x	x	x	x
cancer			x				x	x	x	x
disability								x	x	x
Economics Outcomes at t-1										
claimed DI										
claimed SS										
claimed DB										
work										
earnings										
wlth = 0										
wealth										
nursing home stay										

Notes: x denotes a parameter which is allowed to be estimated. Other controls include initial conditions, demographics, dc and dc plan

Table A.6 Restrictions on Economic Outcomes

	Outcome at (t)									
	Any HI	Di claim	SS claim	Db claim	SSI claim	Nursing H	Work	Earnings	Wlth non-zero	Wealth
Prevalence of Contiions at t-1										
heart	x	x	x	x	x	x	x	x	x	x
blood pressure	x	x	x	x	x	x	x	x	x	x
stroke	x	x	x	x	x	x	x	x	x	x
lung disease	x	x	x	x	x	x	x	x	x	x
diabetes	x	x	x	x	x	x	x	x	x	x
cancer	x	x	x	x	x	x	x	x	x	x
disability	x	x	x	x	x	x	x	x	x	x
Economics Outcomes at t-1										
claimed DI	x	x	x	x	x		x	x	x	x
claimed SS	x			x	x		x	x	x	x
claimed DB			x		x		x	x	x	x
claimed SSI					x					
work	x	x	x		x		x	x	x	x
earnings	x	x	x	x	x		x	x	x	x
wlth non zero	x	x	x	x	x	x	x	x	x	x
wealth	x	x	x	x	x	x	x	x	x	x
nursing home stay					x	x			x	x

Notes: x denotes a parameter which is allowed to be estimated. Other controls include initial conditions, demographics, dc and dc plan

We also include a set of other controls. A list of such controls is given in the next table along with descriptive statistics.

Table A.7 Descriptive Statistics on Exogenous Controls

Control variable	mean	sd	min	max
age spline <75	63.6	8.373	40	75
age spline >75	1.219	3.326	0	34
widowed t-1	0.162	0.368	0	1
black	0.128	0.334	0	1
hispanic	0.063	0.244	0	1
less than high school	0.264	0.441	0	1
college education	0.370	0.483	0	1
male	0.449	0.497	0	1
widowed baseline	0.118	0.322	0	1
single at baseline	0.118	0.323	0	1
logaime/10 at baseline	0.673	0.152	-0.116	0.883
log quarters/10 at baseline	0.428	0.089	0	0.524
self-reported health poor baseline	0.221	0.415	0	1
anydb baseline	0.145	0.353	0	1
nra 57-61	0.038	0.191	0	1
nra 62-63	0.026	0.159	0	1
nra 64>	0.057	0.232	0	1
any dc at baseline	0.162	0.369	0	1
log dc wealth /100 baseline	0.004	0.012	-0.048	0.086

Notes: unweighed statistics for baseline sample

Since we have a stock sample from the age 50+ population, each respondent goes through an individual-specific series of intervals. Hence, we have an unbalanced panel over the age range starting from 50 years old. Denote by j_{i0} the first age at which respondent i is observed and j_{iT_i} the last age when he is observed. Hence we observe outcomes at ages $j_i = j_{i0},..., j_{iT_i}$.

We first start with discrete outcomes which are absorbing states (e..g. disease diagnostic, mortality, benefit claiming). Record as $h_{i,j_i,m}$ =1 if the individual outcome m as of age j_i. We assume the individual-specific component of the hazard can be decomposed in a time invariant and variant part. The time invariant part is composed of the effect of observed characteristics x_i and permanent unobserved characteristics specific to outcome m , $\eta_{i,m}$. The time-varying part is the effect of previously diagnosed outcomes $h_{i,j_i-1,-m}$, (other than the outcome m) on the hazard.[6] We assume an index of the

[6] With some abuse of notation, $j_i - 1$ denotes the previous age at which the respondent was observed.

form $z_{m,j_i} = x_i\beta_m + h_{i,j_i-1,-m}\gamma_m + \eta_{i,m}$. Hence, the latent component of the hazard is modeled as

$$\begin{aligned} h^*_{i,j_i,m} &= x_i\beta_m + h_{i,j_i-1,-m}\gamma_m + \eta_{i,m} + a_{m,j_i} + \varepsilon_{i,j_i,m}, \\ m &= 1,\dots,M_0,\ j_i = j_{i0},\dots,j_{iT_i},\ i = 1,\dots,N \end{aligned} \quad . \tag{1}$$

We approximate a_{m,j_i} with an age spline. After several specification checks, a node at age 75 appears to provide the best fit. This simplification is made for computational reasons since the joint estimation with unrestricted age fixed effects for each condition would imply a large number of parameters.
The outcome, conditional on being at risk, is defined as

$$\begin{aligned} h_{i,j_i,m} &= \max(I(h^*_{i,j_i,m} > 0), h_{i,j_i-1,m}) \\ m &= 1,\dots,M_0,\ j_i = j_{i0},\dots,j_{iT_i},\ i = 1,\dots,N \end{aligned} \quad . \tag{2}$$

As mentioned in the text we consider 8 outcomes which are absorbing states. The occurrence of mortality censors observation of other outcomes in a current year. Mortality is recorded from exit interviews.

We have two other three other types of outcomes.

First, we have binary outcomes which are not absorbing state. We specify latent indices as in (1) for these outcomes as well but where the lag dependent outcome also appears as a right-hand side variable. This allows for state-dependence.
Second, we have ordered outcomes. These outcomes are also modeled as in (1) recognizing the observation rule is a function of unknown thresholds ς_m. Similarly to binary outcomes, we allow for state-dependence by including on the right-hand side
The third outcomes we consider are censored outcomes, earnings and financial wealth. Earnings are only observed when individuals work and similarly log wealth is only observed when individuals have non-zero financial wealth. For these, we consider two part models where the latent variable is specified as in (2.1) but model probabilities only when censoring does not occur. In total, we have M outcomes.

Unobserved Heterogeneity

The term $\varepsilon_{i,j_i,m}$ is a time-varying shock specific to age j_i. We assume that this last shock is normally distributed, and uncorrelated across diseases. Unobserved difference η_{im} are persistent over time and are allowed to be correlated across diseases $m = 1,\dots,M$. We assume that these have a normal distribution with covariance matrix Ω_η.

Likelihood and Initial Condition Problem

The parameters $\theta_1 = (\{\beta_m, \gamma_m, \varsigma_m\}_{m=1}^{M}, vech(\Omega_\eta))$, can be estimated by maximum simulated likelihood. Given the normality distribution assumption on the time-varying unobservable, the joint probability of all time-intervals until failure, right-censoring or death conditional on the individual frailty is the product of normal univariate probabilities. Since these sequences, conditional on unobserved heterogeneity, are also independent across diseases, the joint probability over all disease-specific sequences is simply the product of those probabilities.

For a given respondent with frailty η_i observed from initial age j_{i0} to a last age j_{T_i}, the probability of the observed health history is (omitting the conditioning on covariates for notational simplicity)

$$l_i^{-0}(\theta;\eta_i,h_{i,j_{i0}}) = \left[\prod_{m=1}^{M-1}\prod_{j=j_{i1}}^{j_{T_i}} P_{ij,m}(\theta;\eta_i)^{(1-h_{ij-1,m})(1-h_{ij,M})}\right]\times\left[\prod_{j=j_{i1}}^{j_{T_i}} P_{ij,M}(\theta;\eta_i)\right] \tag{2.3}$$

We make explicit the conditioning on $h_{i,j_{i0}} = (h_{i,j_{i0},0},...,h_{i,j_{i0},M})'$, we have limited information on outcomes prior to this age.

To obtain the likelihood of the parameters given the observables, it remains to integrate out unobserved heterogeneity. The complication is that $h_{i,j_{i0},-m}$, the initial outcomes in each hazard is not likely to be independent of the common unobserved heterogeneity term which needs to be integrated out. A solution is to model the conditional probability distribution $p(\eta_i \mid h_{i,j_{i0}})$ (Wooldridge, 2004). Implementing this solution amounts to including initial outcomes at baseline each hazard. This is equivalent to writing

$$\eta_i = \Gamma h_{i0} + \alpha_i$$
$$\alpha_i \sim N(0,\Omega_\alpha)$$

Therefore, this allows for permanent differences in outcomes due to differences in baseline outcomes. The likelihood contribution for one respondent's sequence is therefore given by

$$l_i(\theta;h_{i,j_{i0}}) = \int l_i(\theta;\alpha_i,h_{i,j_{i0}})dF(\alpha_i) \tag{2.4}$$

To estimate the model, we make use of maximum simulated likelihood. We replace (2.4) with a simulated counterpart based on R draws from the distribution of α. We them optimize over this simulated likelihood using the BFGS algorithm. For this preliminary report, we could not obtain convergence of the joint estimator. So we assumed the distribution of α_i to be degenerate. The following tables give parameter estimates.

Table A.8a

	Status at time t							
	Mortality	Living in nursing home	Heart disease	Stroke	Cancer	Hypertension	Diabetes	Lung disease
Demographics								
Non-Hispanic black	0.0519	-0.1481	-0.1848	0.0059	-0.0455	0.1459	0.0664	-0.268
	1.68	-2.42	-5.09	0.13	-1.09	3.99	1.73	-5.66
Hispanic	-0.0938	-0.5576	-0.2368	-0.0754	-0.1426	0.0353	0.1781	-0.2992
	-2.04	-4.87	-4.71	-1.18	-2.39	0.8	3.65	-4.49
Less than high school	0.0247	0.0155	0.0641	-0.0008	0.015	0.0521	0.0873	0.0411
	0.98	0.31	2.17	-0.02	0.44	1.78	2.59	1.14
Some college and above	-0.034	0.0771	0.0177	0.0003	0.0425	-0.0105	0.0042	-0.0352
	-1.33	1.47	0.67	0.01	1.46	-0.42	0.13	-1.03
Male	0.2521	0.1048	0.1217	0.0336	0.1454	-0.15	0.1155	-0.0799
	10.04	2.09	4.51	0.94	4.8	-5.73	3.63	-2.32
Status at time t-1								
Age spline <75	0.0218	0.0434	0.0223	0.0225	0.0247	0.0209	0.0114	0.0181
	10.15	7.75	10.39	7.6	10.21	10.56	4.66	6.81
Age spline >75	0.0491	0.0526	0.0252	0.0193	-0.0026	-0.0011	-0.0098	-0.0036
	19.45	12.38	7.11	4.74	-0.62	-0.3	-2.04	-0.76
Widowed	0.0091	0.2085	-0.0089	0.0057	-0.0006	0.1394	0.0367	0.0003
	0.22	2.84	-0.17	0.09	-0.01	2.73	0.58	0
Heart disease	0.1737	0.0535		0.1546				
	5.02	0.8		2.94				
Stroke	0.1889	0.4141						
	4.26	5.82						
Cancer	0.6064	-0.2007		0.0289				
	15.93	-1.74		0.39				
Hypertension	0.1564	0.1553	0.2093	0.152				
	4.27	2.24	5.25	2.86				
Diabetes	0.2038	0.2417	0.1507	0.2301		0.2656		
	4.56	2.68	2.78	3.67		4.31		
Lung disease	0.3729	-0.1795						
	8.28	-1.52						
HI cov -gov/emp/other	0.0703	-0.1595	0.0176	0.1285	0.0885	0.1095	0.0637	0.1255
	1.18	-1.18	0.32	1.57	1.32	2.1	1.07	1.75
Overweight	-0.3063	-0.1858	0.014	-0.0529	-0.0675	0.1335	0.1713	0.0435
	-9.86	-3.01	0.37	-1.11	-1.64	3.81	3.49	0.92
Obese	-0.4722	-0.3305	0.0515	-0.0823	-0.0318	0.2943	0.4055	0.1553
	-10.02	-3.31	0.98	-1.21	-0.54	5.76	6.67	2.32
Smoking now	0.0417	0.0081	0.1231	0.0473	0.0282	-0.0288	-0.1896	0.2032
	0.97	0.08	2.35	0.71	0.5	-0.59	-3.33	3.65
IADL only	0.2324	0.6579						
	5.26	9.16						
ADL 1 or 2	0.3262	0.5476						
	11.2	10.35						
ADL 3-5	0.7445	0.9149						
	18.83	12.6						
R live in nursingh ome at interview		2.1234						
		21.17						

Table A8b

Transition in mortality, nursing home residency, and chronic conditions (coefficients/z statistic) - Cont.

	Status at time t							
	Mortality	Living in nursing home	Heart disease	Stroke	Cancer	Hypertension	Diabetes	Lung disease
Initial conditions								
Initial-Heart disease	0.0316	-0.0963		-0.0184	0.0744	0.021	0.0664	0.1393
	0.84	-1.29		-0.31	2.14	0.59	1.8	3.67
Initial-Stroke	-0.0147	-0.2459	0.0087		0.0224	0.0871	0.0459	-0.2152
	-0.27	-2.63	0.14		0.36	1.21	0.72	-2.79
Initial-Cancer	-0.2942	0.0892	0.1339	-0.0404		0.0053	-0.0535	0.1357
	-6.29	0.68	3.27	-0.46		0.13	-0.99	2.75
Initial-Hypertension	-0.0184	-0.133	0.0072	0.0578	0.052		0.1843	0.0366
	-0.5	-1.91	0.18	1.09	1.98		6.97	1.21
Initial-Diabetes	0.122	0.0666	0.0909	0.0174	-0.0606	-0.0623		-0.0098
	2.48	0.67	1.47	0.25	-1.34	-0.86		-0.2
Initial-Lung disease	-0.0866	0.0867	0.1558	0.1376	0.1032	0.0077	0.0423	
	-1.64	0.63	3.3	2.56	2.03	0.16	0.77	
Initial-HI cov -gov/emp/other	-0.0493	0.0752	0.0032	-0.0886	-0.0767	-0.0804	-0.0714	0.0086
	-1.03	0.62	0.06	-1.35	-1.37	-1.72	-1.38	0.14
Initial-Overweight	0.0692	-0.0499	0.0191	0.0427	0.0841	0.0629	0.2563	-0.0742
	2.25	-0.83	0.51	0.9	2.04	1.8	5.26	-1.58
Initial-Obese	0.1949	0.1114	0.059	0.0807	0.0603	0.0215	0.4339	-0.0776
	4.11	1.14	1.1	1.17	0.99	0.41	7.05	-1.14
Initial-Eversmoked	0.1207	-0.0138	0.0425	0.0223	0.0757	-0.0371	0.0437	0.2687
	4.94	-0.3	1.66	0.66	2.64	-1.51	1.46	7.4
Initial-Smoking now	0.1747	0.0558	0.0404	0.107	0.0685	0.0946	0.2379	0.2609
	4.31	0.6	0.8	1.69	1.28	2.02	4.47	4.77
Initial-IADL only	0.0824	0.1129	0.0891	0.0394	-0.1096	-0.0223	-0.0767	0.0899
	1.92	1.34	1.92	0.64	-1.85	-0.49	-1.37	1.57
Initial-ADL 1 or 2	0.0254	-0.0491	0.148	0.0699	0.0592	-0.0111	0.0519	0.1702
	0.8	-0.81	3.89	1.52	1.33	-0.26	1.2	3.79
Initial-ADL 3-5	-0.1358	-0.1693	0.0851	-0.1355	0.1547	-0.2298	-0.0602	0.2073
	-1.88	-1.43	0.72	-0.89	1.22	-1.53	-0.45	1.64
Initial-Working	-0.1778	-0.2191	0.0285	-0.0388	0.0253	-0.0301	-0.0613	-0.0928
	-6.06	-3.13	0.98	-1	0.78	-1.08	-1.8	-2.47
Initial-HH wealth not zero	0.0655	0.0293	0.0217	-0.0345	-0.0362	-0.0169	-0.0129	0.1399
	1.53	0.39	0.4	-0.53	-0.54	-0.3	-0.22	2.23
Initial-Log(HH wealth in 1000s transformed)/100	-2.2278	-6.4713	-2.0971	-2.046	2.7951	-1.1835	-1.8181	-5.3152
	-3.26	-5.33	-2.59	-2	2.89	-1.45	-1.93	-5.49
Initial-Sidowed	0.0556	0.0601	-0.0043	0.0421	-0.0061	-0.1167	0.0526	0.0106
	1.25	0.81	-0.07	0.62	-0.09	-2	0.75	0.14
Initial-Single	0.0708	0.2985	-0.0233	-0.1106	0.0913	-0.0324	0.0027	0.0521
	2.16	4.54	-0.64	-2.16	2.24	-0.92	0.06	1.18
Initial-LogAIME	-0.0341	-0.1939	0.3492	-0.566	0.2864	0.8954	0.1429	0.2258
	-0.16	-0.53	1.43	-1.89	1.06	3.8	0.49	0.73
Initial-Log quarters worked	0.3318	0.0224	-0.2349	0.8041	-0.4964	-1.0566	-0.3346	-0.2189
	1.06	0.04	-0.64	1.79	-1.22	-2.98	-0.77	-0.48
Initial-Poor health	0.1296	0.0261	0.1546	0.1207	-0.0079	0.042	0.1064	0.1827
	5.28	0.54	5.06	3.31	-0.22	1.28	3.07	5.02
Initial-Any DB pension	-0.0541	-0.0009	-0.0664	-0.1611	0.0971	0.0578	0.0313	-0.0627
	-0.55	0	-0.89	-1.27	1.33	0.92	0.38	-0.6
Initial-DB normal elig age 56-69	0.0893	0.1029	0.0826	0.0785	-0.0953	-0.0675	0.0472	0.0894
	0.78	0.29	0.94	0.52	-1.06	-0.89	0.49	0.73
Initial-DB normal elig age 60-61	0.0801	-0.1843	0.0858	0.2169	-0.1042	0.0329	0.1558	0.13
	0.66	-0.42	0.91	1.44	-1.06	0.41	1.55	1.03
Initial-DB normal elig age >=62	0.1026	0.1052	0.1058	0.1212	-0.0718	0.0043	0.0794	0.0903
	0.95	0.31	1.29	0.87	-0.86	0.06	0.88	0.79
Initial-Any DC pension	-0.1119	0.2787	-0.0761	0.0806	0.0768	0.1011	-0.0958	-0.0097
	-1.53	1.85	-1.37	1.02	1.3	2.19	-1.47	-0.13
Initial-Log(DC wealth in 1000s)/100	1.6333	-12.5426	1.9567	-2.108	-3.4726	-2.584	1.5475	-0.2327
	0.76	-2.15	1.22	-0.87	-1.98	-1.86	0.81	-0.11
Constant	-3.8392	-5.3402	-3.5948	-3.7593	-3.8255	-2.955	-3.1932	-3.6307
	-24.83	-13.9	-22.99	-17.76	-21.26	-20.33	-18.11	-18.59
N	59,569	56,227	45,564	53,084	50,846	32,313	49,243	51,980

Data source: HRS 1992-2004, only those with linked social security earnings included

Table A.8c

Transition of health insurance coverage (among <65), economic outcomes (Probit models) (coefficients/t statistic)

	Status at time t					
	Any Health insurance	Claiming DI	Claiming OASI	Claiming DB pension	Claiming SSI	Working
Demographics						
Non-Hispanic black	-0.1218	-0.0231	-0.1107	0.0868	0.0719	-0.0156
	-3.3	-0.44	-2.35	1.33	1.33	-0.61
Hispanic	-0.317	-0.2229	-0.1156	0.0925	-0.0091	-0.031
	-7.53	-2.93	-1.88	0.94	-0.13	-0.92
Less than high school	-0.1617	0.0713	0.0137	0.0087	0.0875	-0.0984
	-4.89	1.49	0.32	0.13	1.75	-4.2
Some college and above	0.092	-0.1242	-0.2151	-0.1436	-0.0595	0.0857
	2.91	-2.62	-6.01	-3.14	-1.08	4.72
Male	-0.0814	-0.0008	-0.2833	0.0168	0.2028	0.0279
	-2.5	-0.02	-7.53	0.35	4.04	1.41
Status at time t-1						
Age spline <75	-0.0055	-0.0033	0.4615	0.1091		-0.0396
	-1.58	-0.62	19.56	17.99		-17.41
Age 60-61			2.3964			-0.2819
			15.99			-11.26
Age 62-64			1.8689			-0.1325
			17.16			-5.62
Age spline 51-64					-0.0202	
					-3.6	
Age spline 65-74					-0.1293	
					-8.83	
Age spline >75					-0.0781	
					-1.42	
Widowed	-0.0247	0.3578	0.128	0.0852	-0.0793	0.1444
	-0.28	3.12	1.35	0.61	-0.58	3.03
Heart disease	0.1173	0.2726	0.0013	0.0466	0.1478	-0.0789
	1.53	3.5	0.02	0.48	1.89	-1.99
Stroke	0.0479	0.1055	-0.0761	0.047	0.2124	-0.1834
	0.35	0.83	-0.7	0.24	1.97	-2.31
Cancer	0.1701	0.1797	0.0248	-0.1728	0.1542	0.0096
	1.54	1.53	0.28	-1.38	1.35	0.2
Hypertension	0.0363	0.0017	0.0944	-0.029	-0.0489	0.0001
	0.67	0.02	1.75	-0.37	-0.59	0
Diabetes	0.0834	0.1013	0.1412	-0.0732	0.1217	-0.0438
	1.09	1.05	1.92	-0.71	1.33	-1
Lung disease	0.2241	0.0218	0.1093	-0.0256	-0.0276	-0.1391
	2.03	0.2	1.19	-0.18	-0.26	-2.37
HI cov -gov/emp/other	1.398	0.0989	-0.2184	0.1393	-0.0106	0.0058
	39.64	1.46	-3.53	0.8	-0.15	0.17
Overweight	0.0382	-0.0191	-0.022	0.0004	0.0017	-0.0309
	0.82	-0.28	-0.44	0.01	0.02	-1.13
Obese	0.0432	-0.0181	0.0454	-0.0352	0.0264	-0.0284
	0.7	-0.2	0.68	-0.36	0.29	-0.76
Smoking now	-0.0086	-0.0402	0.0308	0.0746	-0.1308	-0.039
	-0.15	-0.53	0.48	0.78	-1.72	-1.07
IADL only	-0.1051	0.2721	0.0198	0.1304	0.1489	-0.0797
	-1.63	3.07	0.22	1	1.67	-1.69
ADL 1 or 2	0.0608	0.3893	-0.0852	0.2724	0.1624	-0.2105
	1.17	6.82	-1.46	2.74	2.76	-5.66
ADL 3-5	0.3047	0.305	-0.2253	0.5019	-0.0595	-0.5591
	2.7	3.09	-2.13	1.96	-0.6	-5.29
Claiming SSDI	0.7162	3.0149	-0.9343	-0.1163	-0.0752	-0.8271
	8.08	53.74	-15.78	-0.44	-1.11	-11.86
Lag of Claiming SSI	0.1495				3.0797	
	1.54				49.11	
Claiming OASI	-0.1118			-0.3313		-0.211
	-2.01			-4.35		-7.29
Working	-0.2216	-0.2645	0.3553		-0.1858	1.3433
	-3.69	-2.47	4.16		-1.47	38.08
Log(Earnings in 1000s)/100	9.0155	-6.6858	-21.6108	25.574	-2.7794	5.6478
	5.02	-2.13	-9.13	13.15	-0.72	5.43
Claiming DB			0.016			-0.199
			0.2			-4.76

Table A.8d

Transition of health insurance coverage (among <65), economic outcomes (Probit models) (coefficients/t statistic) - Cont.

	Status at time t					
	Any Health insurance	Claiming DI	Claiming OASI	Claiming DB pension	Claiming SSI	Working
Initial conditions						
Initial-Heart disease	0.0129	-0.1606	0.0342	0.0545	-0.0793	-0.0235
	0.14	-1.78	0.43	0.47	-0.88	-0.5
Initial-Stroke	-0.1433	-0.1466	-0.087	-0.5284	-0.1648	-0.0045
	-0.9	-0.92	-0.6	-1.74	-1.19	-0.05
Initial-Cancer	-0.1399	-0.1843	-0.0738	0.2558	-0.1299	-0.0292
	-1.11	-1.28	-0.67	1.65	-0.9	-0.49
Initial-Hypertension	-0.0459	0.1046	-0.0683	0.1609	0.1157	-0.0402
	-0.81	1.35	-1.22	1.97	1.4	-1.25
Initial-Diabetes	0.0055	-0.0456	-0.1618	0.1529	-0.0661	0.0071
	0.06	-0.42	-1.86	1.19	-0.65	0.14
Initial-Lung disease	-0.3374	0.0222	-0.1697	-0.214	0.1303	0.0637
	-2.75	0.17	-1.54	-1.14	1.07	0.9
Initial-HI cov -gov/emp/other	0.3787	-0.1209	-0.0261	0.1132	-0.0436	0.0301
	10.29	-1.91	-0.46	0.64	-0.68	0.95
Initial-Overweight	-0.0312	0.0134	0.0274	-0.0244	-0.0254	0.0472
	-0.67	0.2	0.55	-0.35	-0.36	1.73
Initial-Obese	-0.0691	0.1153	-0.0396	0.0343	0.0017	0.0332
	-1.1	1.3	-0.58	0.35	0.02	0.87
Initial-Eversmoked	0.1069	0.0658	0.0224	0.0855	-0.0354	-0.0195
	3.35	1.38	0.63	1.89	-0.67	-1.06
Initial-Smoking now	-0.1567	0.1621	0.0072	-0.076	0.2689	-0.0537
	-2.74	2.15	0.12	-0.84	3.54	-1.54
Initial-IADL only	0.1152	-0.0726	0.0487	-0.0556	0.0321	0.0121
	2.22	-0.95	0.8	-0.57	0.44	0.36
Initial-ADL 1 or 2	0.0356	0.1145	-0.0632	0.0993	0.104	-0.077
	0.65	1.89	-1.05	0.86	1.72	-1.95
Initial-ADL 3-5	-0.0202	0.2508	-0.5465		-0.5786	-0.3018
	-0.09	1.17	-0.63		-1.93	-1.21
Initial-Working	-0.109	0.1201	-0.1502		-0.0438	0.3743
	-2.98	2.17	-3.23		-0.73	16.31
Initial-HH wealth not zero	-0.2313	0	0.0123	-0.004	-0.0244	0.2019
	-4.28	0	0.16	-0.03	-0.34	4.6
Initial-Log(HH wealth in 1000s transformed)/100	7.0043	-4.6308	0.0498	3.1035	-5.0596	-7.198
	7.64	-3.36	0.04	1.69	-3.58	-11.36
Initial-Sidowed	-0.0748	-0.3924	-0.2022	-0.1176	0.1039	-0.0883
	-0.75	-2.86	-1.82	-0.68	0.67	-1.5
Initial-Single	-0.0922	-0.0008	-0.1913	0.0227	0.1991	-0.0163
	-2.5	-0.02	-4.2	0.37	3.69	-0.65
Initial-LogAIME	0.9656	-0.3229	-0.3529	-0.1865	-2.4914	-0.0788
	3.1	-0.66	-0.96	-0.35	-5.09	-0.4
Initial-Log quarters worked	-0.9064	2.8925	4.5022	-0.4762	3.4458	1.5861
	-1.97	3.72	8.03	-0.55	4.58	5.15
Initial-Poor health	-0.0318	0.2854	-0.0673	0.0289	0.18	-0.205
	-0.86	5.98	-1.41	0.38	3.58	-7.84
Initial-Any DB pension	0.3319	-0.1279	0.0898		-0.2715	-0.0453
	3.95	-1.03	1.06		-1.42	-1.14
Initial-DB normal elig age 56-69	0.1348	0.0407	-0.0964	-0.1657	0.1811	0.072
	1.25	0.27	-1.02	-2.71	0.83	1.55
Initial-DB normal elig age 60-61	-0.0119	0.2432	0.0103	-0.2214	0.5518	0.0372
	-0.11	1.65	0.1	-3.29	2.66	0.73
Initial-DB normal elig age >=62	-0.0742	0.1985	-0.0363	-0.2268	0.3358	0.0714
	-0.79	1.49	-0.4	-3.89	1.67	1.64
Initial-Any DC pension	0.0766	0.1506	0.0353	-0.1823	0.1847	0.1435
	1.53	2.05	0.58	-2.59	2.11	4.71
Initial-Log(DC wealth in 1000s)/100	6.9601	-5.4994	1.0676	4.8753	-3.2247	-3.5639
	3.94	-2.28	0.6	2.36	-1.08	-3.94
Constant	-0.1509	-2.8852	-31.2817	-8.2647	-0.9181	0.7245
	-0.71	-8.78	-19.94	-18.38	-2.7	5.08
N	28,741	28,741	7,953	8,449	56,227	43,807

Data source: HRS 1992-2004, only those with linked social security earnings included

Table A.8e

Ordered probits for the transition of BMI, smoking, and functional status (coefficient/t statistic)

	Status at time t		
	BMI status	Smoking status	Functional status
Demographics			
Non-Hispanic black	0.0469	-0.0496	0.0603
	2.34	-1.49	2.91
Hispanic	0.0328	-0.0236	0.0968
	1.26	-0.53	3.57
Less than high school	-0.0093	-0.0149	0.0565
	-0.55	-0.53	3.21
Some college and above	-0.054	-0.0712	-0.0009
	-3.68	-2.82	-0.05
Male	0.0284	0.0256	0.0478
	1.84	1	2.81
Status at time t-1			
Age spline <75	-0.0026	-0.0184	0.0074
	-1.24	-9.22	5.67
Age spline >75	-0.0177	-0.0138	0.0548
	-5.52	-2.89	25.99
Widowed	-0.0502	0.0588	-0.03
	-1.59	1.07	-0.93
Heart disease	0.004	0.0872	0.1206
	0.15	1.9	4.43
Stroke	-0.143	0.0506	0.4184
	-3.33	0.72	11.2
Cancer	0.0045	-0.0367	0.0683
	0.13	-0.6	1.86
Hypertension	0.1122	-0.056	0.0387
	4.7	-1.37	1.47
Diabetes	-0.0267	-0.0523	0.1166
	-0.81	-0.93	3.51
Lung disease	0.043	-0.0396	0.232
	1.06	-0.7	6.04
HI cov -gov/emp/other	0.0192	-0.1863	-0.0129
	0.64	-4.13	-0.39
Overweight	1.7343	-0.1599	-0.0643
	88.14	-4.64	-2.74
Obese	3.4757	-0.2906	0.0205
	119.9	-5.83	0.63
Smoking now	-0.2441	2.0343	0.0159
	-8.35	65.44	0.5
IADL only	-0.0006	-0.035	0.7002
	-0.02	-0.61	23.62
ADL 1 or 2	0.0274	-0.0001	1.0132
	1.15	0	50.95
ADL 3-5	0.0242	-0.1183	1.9788
	0.55	-1.6	53.76
Initial conditions			
Initial-Heart disease	-0.0496	-0.0097	-0.0438
	-1.58	-0.19	-1.44
Initial-Stroke	0.0458	-0.1476	-0.1869
	0.85	-1.66	-3.97
Initial-Cancer	-0.0168	0.044	-0.0298
	-0.4	0.6	-0.68
Initial-Hypertension	-0.0393	-0.0487	0.0302
	-1.6	-1.14	1.14
Initial-Diabetes	0.036	0.0389	0.0398
	0.94	0.6	1.07
Initial-Lung disease	-0.0575	0.0803	-0.0401
	-1.19	1.19	-0.89
Initial-HI cov -gov/emp/other	-0.0408	0.0003	0.0042
	-1.51	0.01	0.14

Table A.8f

Ordered probits for the transition of BMI, smoking, and functional status (coefficient/t statistic) - Cont.

	Status at time t		
	BMI status	Smoking status	Functional status
Initial-Overweight	0.7157	0.0883	0.0662
	37.22	2.56	2.83
Initial-Obese	1.3966	0.1293	0.1736
	49.51	2.53	5.27
Initial-Eversmoked	0.0264	5.2217	0.0134
	1.84	109.11	0.83
Initial-Smoking now	0.091	0.9255	0.0925
	3.28	28.52	3.03
Initial-IADL only	-0.0739	0.0103	0.0769
	-2.67	0.22	2.7
Initial-ADL 1 or 2	0.0127	0.0511	0.3073
	0.5	1.23	13.99
Initial-ADL 3-5	0.0932	0.0843	0.3567
	1.2	0.57	5.53
Initial-Working	0.0257	-0.0413	-0.1438
	1.56	-1.51	-7.75
Initial-HH wealth not zero	0.0528	-0.0507	0.0407
	1.67	-1.04	1.36
Initial-Log(HH wealth in 1000s transformed)/100	-1.1539	-1.5232	-4.6995
	-2.44	-1.96	-9.64
Initial-Sidowed	0.0625	-0.0068	0.0064
	1.74	-0.11	0.18
Initial-Single	-0.0077	0.0382	0.0176
	-0.37	1.17	0.79
Initial-LogAIME	-0.0213	-0.0334	-0.4719
	-0.15	-0.13	-3.25
Initial-Log quarters worked	0.1207	0.1699	0.3976
	0.56	0.45	1.83
Initial-Poor health	-0.0357	0.0201	0.3067
	-1.96	0.67	17.81
Initial-Any DB pension	0.0242	0.0064	-0.1006
	0.67	0.1	-1.89
Initial-DB normal elig age 56-69	-0.0025	-0.0896	0.085
	-0.06	-1.18	1.35
Initial-DB normal elig age 60-61	0.0016	-0.0435	0.1196
	0.03	-0.54	1.8
Initial-DB normal elig age >=62	0.003	-0.1328	0.0819
	0.07	-1.9	1.39
Initial-Any DC pension	0.0342	0.0194	0.0149
	1.21	0.41	0.41
Initial-Log(DC wealth in 1000s)/100	-0.9376	-0.6248	-2.7517
	-1.12	-0.44	-2.36
Year of birth	0.0136		
	6.24		
Cutoff 2	27.2977	1.0009	1.7297
	6.34	6.98	18.64
Cutoff 3	30	5.6966	1.9332
	6.92	39.19	20.81
Cutoff 4			2.992
			31.94
N	56,227	56,227	56,227

BMI status: BMI < 25kg/m2; BMI >= 25 & BMI < 30; BMI >=30

Functional status: healthy; IADL only; 1-2 ADLs, 3 or more ADLs

Smoking status: Never-smoked, ever-smoked, and smoking now

Data source: HRS 1992-2004, only those with linked social security earnings included

Table A.8g

OLS regression of Log(Earnings in 1000s) conditional on working (coefficient/t statistic)

	Status at time t
	Log(earnings/1000)
Demographics	
Non-Hispanic black	0.02
	1.3
Hispanic	0
	0.08
Less than high school	-0.03
	-1.79
Some college and above	0.12
	8.62
Male	0.13
	9.05
Status at time t-1	
Age spline <75	0
	1.63
Heart disease	-0.11
	-3.2
Stroke	-0.13
	-1.63
Cancer	-0.01
	-0.32
Hypertension	0.01
	0.52
Diabetes	-0.01
	-0.31
Lung disease	-0.12
	-2.37
HI cov -gov/emp/other	0.05
	1.75
Overweight	0.01
	0.52
Obese	0.01
	0.21
Smoking now	0.01
	0.26
IADL only	-0.02
	-0.67
ADL 1 or 2	-0.09
	-2.53
ADL 3-5	-0.33
	-2.62
Claiming SSDI	-0.4
	-3.19
Lag of Claiming SSI	-0.15
	-0.98
Claiming OASI	-0.44
	-19.06
Claiming DB	-0.44
	-10.61
Working	-0.83
	-26.41
Log(Earnings in 1000s)	0.459
	56.63
Initial conditions	
Initial-Stroke	0.18
	1.8

Table A.8h

OLS regression of Log(Earnings in 1000s) conditional on working (coefficient/t statistic) - Cont.

	Status at time t
	Log(earnings/1000)
Initial-Heart disease	0.07
	1.75
Initial-Cancer	0.03
	0.62
Initial-Hypertension	-0.02
	-0.97
Initial-Diabetes	0
	0.11
Initial-Lung disease	0.1
	1.6
Initial-HI cov -gov/emp/other	0.06
	2.18
Initial-Overweight	0
	0.2
Initial-Obese	-0.01
	-0.39
Initial-Eversmoked	0
	0.05
Initial-Smoking now	-0.01
	-0.26
Initial-IADL only	-0.06
	-2.41
Initial-ADL 1 or 2	0.03
	0.9
Initial-ADL 3-5	0.36
	1.18
Initial-Working	0.18
	7.91
Initial-HH wealth not zero	-0.05
	-1.5
Initial-Log(HH wealth in 1000s)	0.0133
	2.74
Widowed	0.05
	1.18
Initial-Sidowed	-0.07
	-1.43
Initial-Single	0.06
	3.23
Initial-LogAIME	0.339
	21.42
Initial-Log quarters worked	-0.406
	-16.26
Initial-Poor health	-0.02
	-0.87
Initial-Any DB pension	0.14
	5.35
Initial-DB normal elig age 56-69	-0.03
	-0.92
Initial-DB normal elig age 60-61	-0.01
	-0.28
Initial-DB normal elig age >=62	-0.08
	-2.71
Initial-Any DC pension	-0.02
	-0.81
Initial-Log(DC wealth in 1000s)/100	3.4
	5.77
Constant	1.38
	13.06
N	18,123

Data source: HRS 1992-2004, only those with linked social security earnings included

Table A.8I

OLS regression of household wealth (in 1000s) (coefficient/t statistic)

	Status at time t Household wealth in 1000s
Demographics	
Non-Hispanic black	-31.87
	-2.87
Hispanic	-40.37
	-2.72
Less than high school	-0.59
	-0.06
Some college and above	37.96
	4.61
Male	6.96
	0.75
Status at time t-1	
Age spline <75	8.44
	6.08
Age spline >75	7.77
	4.48
Heart disease	-35.17
	-2.29
Stroke	-17.9
	-0.75
Cancer	16.93
	0.87
Hypertension	7.66
	0.57
Diabetes	1.65
	0.09
Lung disease	5.17
	0.23
HI cov -gov/emp/other	23.02
	1.36
Claiming SSDI	-0.54
	-0.02
Lag of Claiming SSI	-6.24
	-0.25
Claiming OASI	-29.1
	-2.18
Claiming DB	50.53
	2.41
R live in nursingh ome at interview	-47.5
	-0.94
Overweight	-11.66
	-1
Obese	1.61
	0.1
Smoking now	-13.14
	-0.79
IADL only	-16.38
	-0.86
ADL 1 or 2	-17.35
	-1.31
ADL 3-5	-22.86
	-0.92
Working	-12.23
	-0.99
Initial conditions	
Initial-Heart disease	30.98
	1.77
Initial-Stroke	8.37
	0.28
Initial-Cancer	-30.36
	-1.29

Table A.8J

OLS regression of household wealth (in 1000s) (coefficient/t statistic) - Cont.

	Status at time t Household wealth in 1000s
Initial-Hypertension	-20.44
	-1.47
Initial-Diabetes	5.96
	0.28
Initial-Lung disease	-20.19
	-0.75
Initial-HI cov -gov/emp/other	7.02
	0.46
Initial-Overweight	-14.67
	-1.26
Initial-Obese	-29.39
	-1.73
Initial-Eversmoked	-19.98
	-2.51
Initial-Smoking now	-2.01
	-0.13
Initial-IADL only	-5.72
	-0.37
Initial-ADL 1 or 2	3.15
	0.22
Initial-ADL 3-5	-9.12
	-0.21
Initial-Working	-52.74
	-4.4
Widowed	-45.69
	-2.63
Initial-Sidowed	33.68
	1.72
Initial-Single	-2.81
	-0.25
Initial-Poor health	-10.13
	-1
Initial-Any DB pension	-12.12
	-0.57
Initial-DB normal elig age 56-69	0.28
	0.01
Initial-DB normal elig age 60-61	-18.51
	-0.69
Initial-DB normal elig age >=62	-0.27
	-0.01
Initial-Any DC pension	2.37
	0.22
fraime	0
	0.37
frq	-0.17
	-1.26
Init. of Individual earnings in 1000s-max 200	1.67
	7.78
Lag of Individual earnings in 1000s-max 200	-0.12
	-0.53
Individual DC wlth in 1000s	0.44
	8.46
Init. of HH wlth in 1000s	0.62
	62.9
Lag of HH wlth in 1000s	0.44
	85.12
rbyr	7.78
	6.39
Constant	-15476.98
	-6.4
N	56,227

Data source: HRS 1992-2004, only those with linked social security earnings included

Goodness-of-Fit

To judge the goodness-of-fit of the model, we estimated parameters on half the estimation sample and simulated outcomes of 1992 HRS respondents up to 2004. We then compared simulated and actual outcomes in 2004.

Table A.9 Comparison of Predicted and Observed Outcomes for HRS Cohort in 2004

Outcomes	1992	2004	
	Observed	Observed	Simulated
Survival	1	0.879	0.861
Chronic condition prevalence (%)			
Cancer	5.3%	15.4%	15.2%
Diabetes	7.4%	17.6%	18.3%
Heart disease	10.1%	24.7%	25.7%
Hypertension	30.7%	54.4%	54.3%
Lung disease	5.4%	12.5%	11.6%
Stroke	2.3%	7.0%	7.4%
Any chronic condition	43.8%	72.6%	76.3%
3+ chronic conditions	3.2%	15.6%	14.1%
Functional status			
IADL only	7.3%	2.4%	2.8%
1 or 2 ADLs	7.9%	8.4%	8.4%
3 or more ADLs	0.0%	2.2%	1.5%
Nursing home residency			
BMI status			
Obese (BMI >=30)	22.2%	29.2%	27.1%
Overweight (25 <=BMI<30)	39.9%	38.5%	39.1%
Smoking status			
Ever-smoked	63.2%	61.1%	60.6%
Smoking now	24.7%	13.2%	12.2%
Binary economic outcomes (%)			
Working	70.1%	25.8%	24.7%
Claiming social security	0.2%	87.8%	87.7%
Claiming DI (among those younger than 65	5.1%	8.2%	6.0%
Claiming Supplemental security income	4.1%	2.2%	2.4%
Continuous economic outcomes (in 2004 dollars)			
Mean annual earnings if working	38,505	28,687	29,019
Median HH wealth (in 1000s)	117,000	224,700	270,568

Notes: Estimation of the model is based on half of the HRS cohort is 1992. The other half is used for simulation.

Model for New Cohorts

Table A.10 Descriptive Statistics Initial Conditions for Estimation (1992) and Simulation (2004)

	mean/fraction 92	mean/fraction 04	selection
Binary			
working for pay	0.710	0.677	all
positive wealth	0.968	0.983	all
hypertension	0.319	0.335	all
heart disease	0.093	0.084	all
diabetes	0.076	0.108	all
any health insurance	0.858	0.885	all
SRH fair/poor	0.197	0.222	all
Ordered			
BMI status			
normal	0.351	0.315	all
overweight	0.397	0.412	all
obese	0.252	0.273	all
Smoking status			
never smoked	0.357	0.428	all
ever smoked	0.340	0.314	all
current smoker	0.303	0.258	all
Functional status			
No ADL	0.826	0.866	all
1 iADL	0.080	0.105	all
2+ ADL	0.094	0.030	all
Continuous			
AIME (Nominal $USD)	1232.4	1711.2	all
quarters of coverage	73.9	84.2	all

Table A.11 Descriptive Statistics Initial Conditions for Estimation (1992) and Simulation (2004) - Continued

Censored Continuous (median, '000 $USD)			
	mean/fraction 92	mean/fraction 04	selection
earnings	29.98	29.98	if working
wealth	171.25	171.25	if positive
dc wealth	19.23	19.23	if dc plan
Censored Discrete			
Any db plan	0.403	0.322	if working
Any dc plan	0.351	0.442	if working
Censored Ordered			
Early Age eligible DB			if db plan
<52	0.201	0.268	
52-57	0.571	0.557	
58>	0.228	0.176	
Normal Age eligible DB			if db plan
<57	0.168	0.222	
57-61	0.264	0.360	
62-63	0.166	0.195	
64>	0.402	0.224	
Covariates			
hispanic	0.088	0.145	
black	0.157	0.135	
male	0.453	0.474	
less HS	0.234	0.123	
college	0.369	0.586	
single	0.178	0.181	
widowed	0.042	0.014	
cancer	0.041	0.055	
lung disease	0.050	0.045	
stroke	0.023	0.022	

Notes: Characteristics in sample matched to records. Unweighted.

Information Available and Empirical Strategy

Ideally, we need information on

$$f_t(y_{i1},...,y_{iM}) = f_t(y_i) \tag{1}$$

where t denotes calendar time, and $y_i = (y_{i1},...,y_{iM})$ is a vector of outcomes of interest whose probability distribution at time *t* is $f_t()$. Information on how the joint distribution evolves over time is not available. Trends in conditional distributions are rarely reported either.

Generally, we have from published or unpublished sources good information on trends for some moments of each outcome (say a mean or a fraction). That is, we have information on

$$g_{t,m}(y_{i,m}) \tag{2}$$

where $g_{t,m}()$ denotes the marginal probability distribution of outcome *m* at time *t.*
For example we know from the NHIS repeated cross-sections that the fraction obese is increasing by roughly 2% a year among 50 year olds. In statistical jargon this means we have information on how the mean of the marginal distribution of y_{im}, a indicator variable that denotes whether someone is obese, is evolving over time.

We also have information on the joint distribution at one point in time, say year t_0. For example, we can estimate the joint distribution on age 50 respondents in the 1992 wave of the HRS, $f_{t_0}(y_i)$.

We make the assumption that only some part of $f_t(y_i)$ evolves over time. In particular, we will model the marginal distribution of each outcome allowing for correlation across these marginals. The correlations will be assumed fixed while the mean of the marginals will be allowed to change over time.

Model

Assume the latent model for $y_i^* = (y_{i1}^*,...,y_{iM}^*)'$,

$$y_i^* = \mu + \varepsilon_i \tag{3}$$

where ε_i is normally distributed with mean zero and covariance matrix Ω. It will be useful to write the model as

$$y_i^* = \mu + L_\Omega \eta_i \tag{4}$$

where L_Ω is a lower triangular matrix such that $L_\Omega L_\Omega' = \Omega$ and $\eta_i = (\eta_{i1},...,\eta_{iM})'$ are standard normal. We observe $y_i = \Gamma(y_i^*)$ which is a non-invertible mapping for a subset of the M outcomes. For example, we have binary, ordered and censored outcomes for which integration is necessary.

Because the mapping is non-invertible, integration needs to be performed to calculate the likelihood contributions $L_i(\theta \mid y_i)$. Integration needs to be done over a large number of dimensions. We will use maximum simulated likelihood to estimate the parameters of the model . The estimator is given by

$$\theta_{MSL} = \arg\max_{\theta=(\mu,\Omega)} \tfrac{1}{N}\sum_{i=1}^{N} \log \tfrac{1}{R}\sum_{r=1}^{R} \tilde{\Pr}(y_i \mid \theta)_r \tag{5}$$

where $\frac{1}{R}\sum_{r=1}^{R}\tilde{\Pr}(y_i \mid \theta)_r$ is a consistent estimate of $\tilde{\Pr}(y_i \mid \theta)$. This estimator is consistent if both N, R tend to infinity. In practice, one can vary R to assess the bias of the estimator for smaller R. It is asymptotically efficient for $R/\sqrt{N}$ tending to infinity.

Specification of Conditional Mean

The vector μ can depend on some variables which have a stable distribution over time z_i (say race, gender and education). This way, estimation preserves the correlation with these outcomes without having to estimate their correlation with other outcomes. Hence, we can write

$$\mu_i = z_i\beta \tag{6}$$

and the whole analysis is done conditional on z_i.

Identification

For binary and ordered outcomes, we fix $\Omega_{m,m}=1$ which fixes the scale. Also we fix the location of the ordered models by fixing thresholds as $\tau_0 = -\infty, \tau_1 = 0, \tau_K = +\infty$ where K denotes the number of categories for a particular outcome. Because some of the binary outcomes are rare, we fix correlations to zero between two outcomes if both fraction positive are below 10%. Furthermore, we fix to zero the correlation between selected outcomes (say earnings) and their selection indicator (work). Hence, we consider two-part models for these outcomes.

Simulation

For exposition, we order the observed outcomes as binary, ordered, continuous and finally censored. The GHK simulator can be used to simulate $\Pr(y_i \mid \theta)$.

We start with the first outcome y_{i1}^*, a discrete outcome.

1. A draw of η_{i1} consistent with observed choice y_{i1} is

$$\tilde{\eta}_{i1} = \Phi^{-1}[\tilde{u}_{i1}\Phi(\tfrac{\bar{c}_{i1}-\mu_{i1}}{L_{\Omega,11}}) + (1-\tilde{u}_{i1})\Phi(\tfrac{\underline{c}_{i1}-\mu_{i1}}{L_{\Omega,11}})] \tag{7}$$

   where $\bar{c}_{i1} = \begin{cases} +\infty & \text{if } y_{i1} = 1 \\ 0 & \text{if } y_{i1} = 0 \end{cases}, \underline{c}_{i1} = \begin{cases} 0 & \text{if } y_{i1} = 1 \\ -\infty & \text{if } y_{i1} = 0 \end{cases}$ and $\tilde{u}_{i1}$ is a uniform draw. The bounds are slightly different for ordered outcomes where thresholds are also estimated. In particular we have

$$\bar{c}_{i1} = \tau_k, \underline{c}_{i1} = \tau_{k-1} \text{ if } y_{i1} = k$$

   where τ_k are parameters to be estimated.
2. The probability of that first outcome is $\tilde{\Pr}(y_{i1} \mid \theta) = \Phi(\frac{\bar{c}_{i1}-\mu_i}{L_{\Omega,11}}) - \Phi(\frac{\underline{c}_{i1}-\mu_i}{L_{\Omega,11}})$

3. Now a draw of η_{i2} consistent with y_{i2} and the draw $\tilde{\eta}_{i1}$ is given by

$$\tilde{\eta}_{i2} = \Phi^{-1}[\tilde{u}_{i2}\Phi(\tfrac{\bar{c}_{i2}-\mu_{i2}-L_{\Omega,21}\tilde{\eta}_{i1}}{L_{\Omega,22}}) + (1-\tilde{u}_{i2})\Phi(\tfrac{\underline{c}_{i2}-\mu_{i2}-L_{\Omega,21}\tilde{\eta}_{i1}}{L_{\Omega,22}})]$$

4. Then the probability is given by

$$\tilde{\Pr}(y_{i1}, y_{i2} \mid \theta) = \tilde{\Pr}(y_{i1} \mid \theta)[\Phi(\tfrac{\bar{c}_{i2}-\mu_{i2}-L_{\Omega,21}\tilde{\eta}_{i1}}{L_{\Omega,22}}) - \Phi(\tfrac{\underline{c}_{i2}-\mu_{i2}--L_{\Omega,21}\tilde{\eta}_{i2}}{L_{\Omega,22}})] \quad (8)$$

5. Cycle trough 3 and 4 until end of discrete outcomes. Denote by $m_0 - 1$ the number of discrete outcomes.
6. An error consistent with the first continuous outcome is

$$\tilde{\eta}_{im_0} = \frac{y_{i,m_0} - \mu_{i,m_0} - \sum_{s=1}^{m_0-1} L_{\Omega,m_0,s}\tilde{\eta}_{is}}{L_{\Omega,m_0,m_0}}$$

7. The probability is $\tilde{\Pr}(y_{i,m_0} \mid \theta) = \frac{1}{L_{\Omega,m_0,m_0}}\phi(\frac{y_{i,m_0} - \mu_{i,m_0} - \sum_{s=1}^{m_0-1} L_{\Omega,m_0,s}\tilde{\eta}_{is}}{L_{\Omega,m_0,m_0}})$
8. Hence $\tilde{\Pr}(y_{i1},..., y_{i,m_0} \mid \theta) = \tilde{\Pr}(y_{i1},..., y_{i,m_0-1} \mid \theta)\tilde{\Pr}(y_{i,m_0} \mid \theta)$
9. Cycle trough 6 to 8 until reach $m_1 - 1$, the last continuous outcome.
10. Denote by m_1 the first censored outcome. Denote by y_{ij} the binary outcome that records whether y_{im_1} can be observed. A draw consistent with y_{i,m_1} is given by

$$\tilde{\eta}_{im_1} = \Phi^{-1}[\tilde{u}_{im_1}] \text{ if } y_{ij} = 0$$

and

$$\tilde{\eta}_{im_1} = \frac{y_{i,m1} - \mu_{i,m1} - \sum_{s=1}^{m_1-1} L_{\Omega,m_1,s}\tilde{\eta}_{is}}{L_{\Omega,m_1,m_1}} \text{ if } y_{ij} = 1$$

If y_{im1} is continuous and given by a draw similar to (7) if a binary outcome.

11. The probability is then

$$\tilde{\Pr}(y_{i,m_1} \mid \theta) = \left[\frac{1}{L_{\Omega,m_1,m1}}\phi(\frac{y_{i,m1} - \mu_{i,m_1} - \sum_{s=1}^{m_1-1} L_{\Omega,m1,s}\tilde{\eta}_{is}}{L_{\Omega,m_1,m_1}})\right]^{I(y_{i,j}=1)}$$

For continuous and cumulative normal similar to (8) for discrete.

12. Cycle 10-11 until reach M.
13. Repeat 1-9 R times and calculate $\frac{1}{R}\sum_{r=1}^{R}\tilde{\Pr}(y_i \mid \theta)_r$.
14. Repeat for each $i = 1,..., N$

Random Draws

We use draws from Halton sequences to generate uniform random draws (see Train, 2003). Note that draws $\left\{\left\{\left\{u_{im,r}\right\}_{m=1,...,M}\right\}_{r=1,...,R}\right\}_{i=1,...,N}$ are kept fixed trough estimation. For the first past, we used 75 draws along each dimension.

Reparametrization

Because some parameters are naturally bounded, we reparametrize the problem to guarantee an interior solution. In particular, we parametrize

$$\Omega_{m,m} = \exp(\delta_m),\ m = m_0 - 1, \ldots, M$$
$$\Omega_{m,n} = \tanh(\xi_{m,n})\sqrt{\Omega_{m,m}\Omega_{n,n}},\ m,n = 1,\ldots,M$$
$$\tau_{m,k} = \exp(\gamma_{m,k}) + \tau_{k-1},\ k = 2,\ldots,K_m - 1, m \text{ ordered}$$

and estimate the $(\delta_{m,m}, \xi_{m,n}, \gamma_k)$ instead of the original parameters.

Simulation-Forward Preserving Correlation

The latent model is written as

$$y_i^* = \mu + L_\Omega \eta_i$$

Each marginal as a mean change equal to $E(y \mid \mu) = (1+\tau)g(\mu)$ where τ is the percent change in the outcome and $g()$ is a non-linear but monotone mapping. Since it is invertible, we can find the vector μ^* where $\mu^* = g^{-1}(E(y \mid \mu)/(1+\tau))$. We use these new intercepts to simulate new outcomes.

Table A.12 Parameter Estimates for Latent Model (Conditional Mean and Thresholds)

	hypertension coefficient	heart disease coefficient	diabetes coefficient	hlt insurance coefficient	self-rate hlth coefficient	BMI status coefficient	smoking status coefficient	Func status coefficient	work coefficient	positive wlth coefficient
black	0.673	-0.112	0.637	-0.171	0.257	0.151	0.167	-0.052	0.077	-1.124
hispan	0.238	-0.290	0.512	-0.888	0.580	-0.277	-0.332	-0.331	0.328	-0.868
hsless	0.000	0.372	0.055	-0.365	0.649	0.008	0.287	0.679	-0.805	-0.282
college	-0.166	-0.186	0.039	0.366	-0.392	-0.182	-0.191	-0.421	0.316	0.762
male	0.170	0.247	0.195	-0.024	0.137	0.158	0.382	-0.018	0.337	0.275
single	-0.011	0.013	0.006	-0.342	0.164	-0.039	0.230	0.076	0.040	-0.505
widowed	0.177	0.031	0.020	-0.424	0.137	-0.075	0.140	0.056	0.151	-1.526
lunge	-0.051	0.199	0.065	0.287	0.378	-0.125	0.830	0.702	-0.212	-0.115
cancre	0.064	0.329	0.013	-0.060	0.676	-0.049	-0.016	0.247	-0.104	-0.751
stroke	0.805	1.306	0.733	-0.146	1.373	0.408	-0.101	1.059	-0.729	-0.113
Constant	-0.620	-1.457	-1.602	1.472	-0.997	0.552	0.077	-1.127	0.240	2.754

	log AIME coefficient	log quarters coefficient	log earnings coefficient	log net wlth coefficient	log dc wlth coefficient	any dc plan coefficient	any db plan coefficient	db early ret age coefficient	db normal ret age coefficient
black	-0.012	-0.002	0.000	-0.015	-0.009	-0.573	0.254	0.282	0.117
hispan	-0.024	-0.005	-0.003	-0.011	-0.014	-0.633	-0.919	-0.176	-1.456
hsless	-0.098	-0.043	-0.004	-0.011	-0.016	-0.385	-0.017	0.206	0.076
college	0.015	-0.001	0.005	0.008	0.008	0.183	0.221	-0.347	-0.367
male	0.112	0.047	0.005	0.000	0.006	0.147	0.221	0.010	0.167
single	0.006	0.004	-0.001	-0.012	-0.007	-0.228	-0.038	-0.193	-0.041
widowed	-0.012	-0.001	-0.001	-0.006	0.004	-0.546	-0.152	-0.306	-0.748
lunge	-0.005	0.000	0.000	-0.007	-0.002	0.565	-0.383	0.818	0.278
cancre	0.021	0.018	0.000	0.000	0.001	0.497	-0.207	-1.053	-0.168
stroke	-0.016	-0.007	0.001	-0.001	-0.001	-0.545	-0.242	-0.088	-0.019
Constant	0.670	0.428	0.030	0.050	0.010	-0.137	-0.646	0.980	1.036

Thresholds for ordered outcome	BMI coefficient	noking status coefficient	functional tus coefficient	db early ret ge coefficient	lb normal ret ge coefficient
thres 2	1.088	0.917	0.937	1.350	0.961
thres 3					1.490

Table A.13 Parameter Estimates: Covariance Matrix

Covariance matrix

	hypertension	heart disease	diabetes	hlt insurance	self-rate hlth	BMI status	smoking status	Func status	work	positive wlth
hypertension	1.00000									
heart disease	0.21794	1.00000								
diabetes	0.40057	0.06424	1.00000							
hlt insurance	0.01561	0.00000	-0.01355	1.00000						
self-rate hlth	0.33147	0.29938	0.40623	0.00143	1.00000					
BMI status	0.31394	0.14999	0.25930	0.00130	0.12028	1.00000				
smoking status	-0.01968	0.00295	-0.02120	-0.04206	0.04959	-0.16650	1.00000			
Func status	0.04853	0.14068	0.18471	-0.02790	0.56352	0.09503	0.06146	1.00000		
work	0.02340	-0.20047	-0.14983	0.16229	-0.37670	-0.02793	-0.06327	-0.37912	1.00000	
positive wlth	0.13044	0.00000	0.07869	0.00000	0.01406	-0.01406	-0.22470	0.09888	0.40038	1.00000
log AIME	0.04240	0.00725	-0.07339	0.10089	-0.13375	-0.06438	0.03868	-0.12157	0.30105	0.05583
log quarters	0.05037	0.00808	-0.05538	0.06371	-0.09476	-0.06199	0.08901	-0.08747	0.29318	-0.01143
log earnings	-0.03156	-0.05095	-0.10864	0.17057	-0.04078	-0.02353	-0.05361	0.00736	0.00000	0.00941
log net wlth	-0.11234	-0.11798	-0.11941	0.10959	-0.05539	-0.09587	-0.08670	-0.16131	-0.03728	0.00000
log dc wlth	-0.10852	-0.08475	-0.04731	0.06500	-0.10484	0.00815	-0.05135	-0.12966	0.09232	-0.01388
any dc plan	-0.16965	-0.12914	-0.11053	0.02845	-0.09796	-0.05579	-0.08583	0.03741	0.00000	0.00746
any db plan	0.07748	-0.06320	-0.03164	0.20374	-0.07232	0.00921	-0.09924	-0.03597	0.00000	0.12491
db early ret age	0.00365	-0.06018	-0.16666	-0.02320	0.18265	-0.19525	-0.01477	0.24162	-0.20216	0.27097
db normal ret age	-0.10478	-0.14876	-0.12448	-0.15468	-0.01682	0.02839	0.03297	-0.16667	0.12254	-0.23927

	log AIME	log quarters	log earnings	log net wlth	log dc wlth	any dc plan	any db plan	db early ret age	db normal ret age
log AIME	-4.40987								
log quarters	1.65267	-5.62343							
log earnings	0.26801	0.07117	-9.71822						
log net wlth	0.13979	0.07418	0.17235	-7.74280					
log dc wlth	0.27395	0.14323	0.61582	0.20899	-7.53380				
any dc plan	0.15263	0.07295	0.39823	0.05628	0.93289	1.00000			
any db plan	0.01544	-0.06372	0.30512	0.02089	0.15811	0.07350	1.00000		
db early ret age	-0.10807	-0.04090	-0.37548	-0.05549	-0.39958	-0.09736	0.00000	1.00000	
db normal ret age	0.08630	0.16115	-0.20542	-0.11818	-0.22688	-0.09801	0.00000	0.25021	1.00000

Note: Shaded figures are fixed before estimation

Government Revenue and Expenditures

This gives a limited overview of how revenues and expenditures of the government are computed. These functions are based on 2004 rules but we include predicted changes in program rules such changes based on year of birth (e.g. Normal retirement age).

We cover the following revenues and expenditures:

Revenues	Expenditures
Federal Tax	Social Security Retirement benefits
State and City Tax	Social Security Disability benefits
Social Security Tax	Supplementary Security Income (SSI)
Medicare Tax	Medical Care Costs

Social Security Benefits

Workers with 40 quarters of coverage and age 62 are eligible to their retirement benefit. The benefit is calculated based on the Average Indexed Monthly Earnings (AIME) and the age at which benefits are first received. If an individuals claims at his normal retirement age (NRA) (65 for those born prior to 1943, 66 for those between 1943 and 1957, and 67 thereafter), he receives his Primary Insurance Amount (PIA) as a monthly benefit. The PIA is a piece-wise linear function of the AIME. If a worker claims prior to his NRA, his benefit is lower than his PIA. If he retires after the NRA, his benefit is higher. While receiving benefits, earnings are taxed above a certain earning disregard level prior to the NRA. An individual is eligible to half of his spouse's PIA, properly adjusted for the claiming age, if that his higher than his/her own retirement benefit. A surviving spouse is eligible to the deceased spouse's PIA. Since we assume prices are constant in our simulations, we do not adjust benefits for the COLA (Cost of Living Adjustment) which usually follows inflation.

DI Benefits

Workers with enough quarters of coverage and under the normal retirement age are eligible for their PIA (no reduction factor) if they are judged disabled (which we take as the predicted outcome of DI receipt) and earnings are under a cap called the Substantial Gainful Activity (SGA) limit. This limit was $9720 in 2004. We ignore the 9 month trial period over a 5 year window in which the SGA is ignored.

SSI Benefits

SSI payments are based on two eligibility test: an income test and a resource test. This test differs for singles and couples. The income test was $637 (monthly) for singles and $956 for couples in 2004 (so called Federal Benefit Rate). The resource criterion is $2000 for singles and $3000 for couples. Once these two tests are met, a household is eligible to the FBR. SSI benefits decrease 1 for 1 with income so that only the difference between income and the FBR is paid out. Since we predict SSI receipt, we do not model eligibility.

Taxes

We consider Federal, State and City taxes paid at the household level. We also calculate Social Security taxes and Medicare taxes. HRS respondents are linked to their spouse in the HRS simulation. We take program rules from the OECD's Taxing Wages Publication for 2004. Households have basic and personal deductions based on marital status and age (>65). Couples are assumed to file jointly. Social Security benefits are partially taxed. The amount taxable increases with other income from 50% to 82%. Low income elderly have access to a special tax credit and the earned income tax credit is applied for individuals younger than 65. We calculate state and city taxes for someone living in Detroit Michigan. The OECD chose this location because it is generally representative of average state and city taxes paid in the U.S. Since Social Security administrative data cannot be used jointly with Geocode information in the HRS, we apply these hypothetical taxes to all respondents.

At the state level, there is a basic deduction for each member of the household ($3100) and taxable income is taxed at a flat rate of 4%. At the city level, there is a small deduction of 750$ per household member and the reminder is taxed at a rate of 2.55%. There is however a tax credit that decreases with income (20% on the first 100$ of taxes paid, 10% on the following 50$ and 5% on the remaining portion).

We calculate taxes paid by the employee for Old-Age Social Insurance (SS benefits and DI) and Medicare (Medicaid and Medicare). It does not include the equivalent portion paid by the employer. OASI taxes of 6.2% are levied on earnings up to $97,500 (2004 cap) while the Medicare tax (1.45%) is applied to all earnings.

Demographic Adjustments

We make two adjustments to the weighting in the Health and Retirement Study to match population counts from the Census. First, we post-stratify the HRS sample by 5 year age groups, gender and race and rebalance weights using the 2004 Current Population Survey (CPS). The CPS is itself matched to the decennial Census. Since we deleted some cases from the data and only considered the set of respondents with matched Social Security records, this takes account of selectivity based on these characteristics. We do this for both new cohort and host data set. The second adjustment we make is to scale up weights for future new cohorts using population projections from the Census Bureau. Again, we do this by race and gender.