The Future Adult Model: Technical Documentation

Dana P. Goldman, University of Southern California
Duncan Ermini Leaf, University of Southern California
Bryan Tysinger, University of Southern California

April 24, 2025

Contents

List of Figures

List of Tables

1 Functioning of the dynamic model

1.1 Background

The Future Elderly Model (FEM) is a microsimulation model originally developed out of an effort to examine health and health care costs among the elderly Medicare population (age 65+). A description of the previous incarnation of the model can be found in Goldman et al. (2004). The original work was founded by the Centers for Medicare and Medicaid Services and carried out by a team of researchers composed of Dana P. Goldman, Paul G. Shekelle, Jayanta Bhattacharya, Michael Hurd, Geoffrey F. Joyce, Darius N. Lakdawalla, Dawn H. Matsui, Sydne J. Newberry, Constantijn W. A. Panis and Baoping Shang.

Since then various extensions have been implemented to the original model. The most recent version of the FEM now projects health outcomes for all Americans aged 51 and older and uses the Health and Retirement Study (HRS) as a host dataset rather than the Medicare Current Beneficiary Survey (MCBS). The work has also been extended to include economic outcomes such as earnings,

labor force participation and pensions. This work was funded by the National Institute on Aging through its support of the RAND Roybal Center for Health Policy Simulation (P30AG024968), the Department of Labor through contract J-9-P-2-0033, the National Institutes of Aging through the R01 grant "Integrated Retirement Modeling" (R01AG030824) and the MacArthur Foundation Research Network on an Aging Society.

This document describes the Future Adult Model (FAM), the development of the model to forecast Americans aged 25 and older. FAM uses the Panel Survey of Income Dynamics (PSID) as the host dataset. In addition to modeling health, health care costs, and economic outcomes, FAM also models life events such as changes in marital status and childbearing. Development of FAM is suported by the National Institutes of Aging through the USC Roybal Center for Health Policy Simulation (5P30AG024968-13) and the MacArthur Foundation Research Network on an Aging Society.

1.2 Overview

The defining characteristic of the model is the modeling of real rather than synthetic cohorts, all of whom are followed at the individual level. This allows for more heterogeneity in behavior than would be allowed by a cell-based approach. Also, since the PSID interviews both respondent and spouse, we can link records to calculate household-level outcomes, which depend on the responses of both spouses.

The model has three core components:

- The replenishing cohort module predicts the economic and health outcomes of new cohorts of 25/26 year-olds. This module takes in data from the Panel Survey of Income Dynamics (PSID) and trends calculated from other sources. It allows us to "generate" cohorts as the simulation proceeds, so that we can measure outcomes for the age 25+ population in any given year.

- The transition module calculates the probabilities of transiting across various health states and financial outcomes. The module takes as inputs risk factors such as smoking, weight, age and education, along with lagged health and financial states. This allows for a great deal of heterogeneity and fairly general feedback effects. The transition probabilities are estimated from the longitudinal data in the PSID.

- The policy outcomes module aggregates projections of individual-level outcomes into policy outcomes such as taxes, medical care costs, and disability benefits. This component takes account of public and private program rules to the extent allowed by the available outcomes.

Figure 1 provides a schematic overview of the model. In this example, we start in 2014 with an initial population aged 25+ taken from the PSID. We then predict outcomes using our estimated transition probabilities (see section 3.1). Those who survive make it to the end of that year, at which point we calculate policy outcomes for the year. We then move to the following time period (two years later), when a replenishing cohort of 25 and 26 year-olds enters (see section 4). This entrance forms the new age 25+ population, which then proceeds through the transition model as before. This process is repeated until we reach the final year of the simulation.

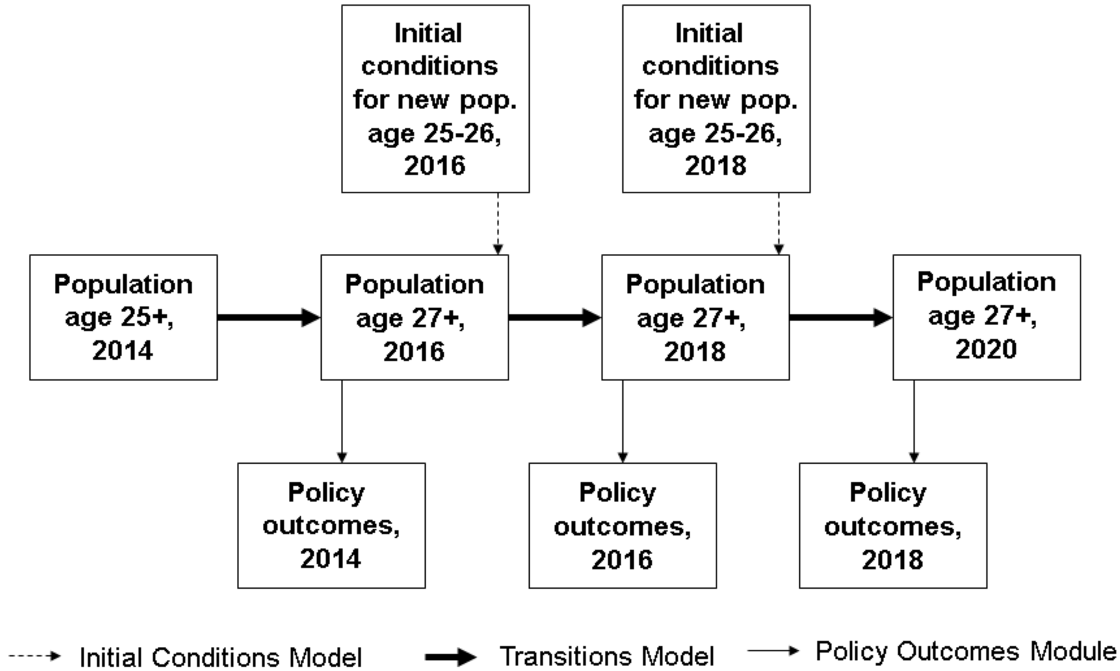

Figure 1: Architecture of the FAM

1.3 Comparison with other microsimulation models of health expenditures

The precursor to the FAM, the FEM, was unique among models that make health expenditure projections. It was the only model that projected health trends rather than health expenditures. It was also unique in generating mortality projections based on assumptions about health trends rather than historical time series.

FAM extends FEM to younger ages, adding additional dimensions to the simulation. Events over the life course, such as marital status and childbearing are simulated. Labor force participation is modeled in greater detail, distinguishing between out-of-labor force, unemployed, working part-time, and working full-time.

1.3.1 CBOLT Model

The Congressional Budget Office (CBO) uses time-series techniques to project health expenditure growth in the short term and then makes an assumption on long-term growth. They use a long term growth of excess costs of 2.3 percentage points starting in 2020 for Medicare. They then assume a reduction in excess cost growth in Medicare of 1.5% through 2083, leaving a rate of 0.9% in 2083. For non-Medicare spending they assume an annual decline of 4.5%, leading to an excess growth rate in 2083 of 0.1%.

1.3.2 Centers for Medicare and Medicaid Services

The Centers for Medicare and Medicaid Services (CMS) performs an extrapolation of medical expenditures over the first ten years, then computes a general equilibrium model for years 25 through 75 and linearly interpolates to identify medical expenditures in years 11 through 24 of their estimation. The core assumption they use is that excess growth of health expenditures will be one percentage point higher per year for years 25-75 (that is if nominal GDP growth is 4%, health care expenditure growth will be 5%).

1.3.3 MINT Model

Modeling Income in the Near Term (MINT) is a microsimulation model developed by the Urban Institute and others for the Social Security Administration to enable policy analysis of proposed changes to Social Security benefits and payroll taxes Smith and Favreault (2013). MINT uses the Survey of Income and Program Participation (SIPP) as the base data and simulates a range of outcomes, with a focus on those that will impact Social Security. Recent extensions have included health insurance coverage and out-of-pocket medical expenditures. Health enters MINT via self-reported health status and self-reported work limitations. MINT simulates marital status and fertility.

2 Data sources used for estimation

The Panel Survey of Income Dynamics is the main data source for the model. We estimate models for assigning characteristics for the replacement cohorts in Replenishing Conditions Module. These are summarized in Table 1. We estimate transition models for the entire PSID population in the Transition Model Module. Transitioned outcomes are described in Table 2.

2.1 Panel Survey of Income Dynamics

The Panel Survey of Income Dynamics (PSID), waves 1999-2019 are used to estimate the transition models. PSID interviews occur every two years. We create a dataset of respondents who have formed their own households, either as single heads of households, cohabitating partners, or married partners. These heads, wives, and ”wives” (males are automatically assigned head of household status by the PSID if they are in a couple) respond to the richest set of PSID questions, including the health questions that are critical for our purposes.

We use all respondents age 25 and older. When appropriately weighted, the PSID is representative of U.S. households. We also use the PSID as the host data for full population simulations that begin in 2009. Respondents age 25 and 26 are used as the basis for the synthetic cohorts that we generate, used for replenishing the sample in population simulations or as the basis of cohort scenarios.

The PSID continually adds new cohorts that are descendents (or new partners/spouses of descendents). Consequently, updating the simulation to include more recent data is straightforward.

2.2 Health and Retirement Study

The Health and Retirement Study (HRS), waves 1998-2012 are pooled with the PSID for estimation of mortality and widowhood models. The HRS has a similar structure to the PSID, with interviews occurring every two years. The HRS data is harmonized to the PSID for all relevant variables. We use the dataset created by RAND (RAND HRS, version O) as our basis for the analysis. We use all cohorts in the analysis. When appropriately weighted, the HRS in 2010 is representative of U.S. households where at least one member is at least 51. Compared to the PSID, the HRS includes more older Hispanics and interviews more respondents once they have entered nursing homes.

2.3 Medical Expenditure Panel Survey

The Medical Expenditure Panel Survey (MEPS), beginning in 1996, is a set of large-scale surveys of families and individuals, their medical providers (doctors, hospitals, pharmacies, etc.), and employers across the United States. The Household Component (HC) of the MEPS provides data from individual households and their members, which is supplemented by data from their medical providers. The Household Component collects data from a representative sub sample of households drawn from the previous year's National Health Interview Survey (NHIS). Since NHIS does not include the institutionalized population, neither does MEPS: this implies that we can only use the MEPS to estimate medical costs for the non-elderly (25-64) population. Information collected during household interviews include: demographic characteristics, health conditions, health status, use of medical services, sources of medical payments, and body weight and height. Each year the household survey includes approximately 12,000 households or 34,000 individuals. Sample size for those aged 25-64 is about 15,800 in each year. MEPS has comparable measures of social-economic (SES) variables as those in PSID, including age, race/ethnicity, educational level, census region, and marital status. We estimate expenditures and utilization using 2007-2010 data.

See Section 5.1 for a description. FAM also uses MEPS 2001-2003 data for QALY model estimation.

2.4 Medicare Current Beneficiary Survey

The Medicare Current Beneficiary Survey (MCBS) is a nationally representative sample of aged, disabled and institutionalized Medicare beneficiaries. The MCBS attempts to interview each respondent twelve times over three years, regardless of whether he or she resides in the community, a facility, or transitions between community and facility settings. The disabled (under 65 years of age) and oldest-old (85 years of age or older) are over-sampled. The first round of interviewing was conducted in 1991. Originally, the survey was a longitudinal sample with periodic supplements and indefinite periods of participation. In 1994, the MCBS switched to a rotating panel design with limited periods of participation. Each fall a new panel is introduced, with a target sample size of 12,000 respondents and each summer a panel is retired. Institutionalized respondents are interviewed by proxy. The MCBS contains comprehensive self-reported information on the health status, health care use and expenditures, health insurance coverage, and socioeconomic and demographic characteristics of the entire spectrum of Medicare beneficiaries. Medicare claims data for beneficiaries enrolled in fee-for-service plans are also used to provide more accurate information on health care use and expenditures. MCBS years 2007-2010 are used for estimating medical cost and enrollment models. See section 5.1 for discussion.

3 Estimation

In this section we describe the approach used to estimate the transition model, the core of the FAM, and the initial cohort model which is used to rejuvenate the simulation population.

3.1 Transition model

We consider a large set of outcomes for which we model transitions. Table 5 gives the set of outcomes considered for the transition model along with descriptive statistics and the population at risk when estimating the relationships.

Since we have a stock sample from the age 25+ population, each respondent goes through an individual-specific series of intervals. Hence, we have an unbalanced panel over the age range starting from 25 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 $j_i = j_{i0}, \ldots, j_{iT_i}$.

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 has occurred 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 that are constant over the entire life course and initial conditions $h_{i,j_0,-m}$ (outcomes other than the outcome m) that are determined before the first age in which each individual is observed The time-varying part is the effect of previously diagnosed outcomes $h_{i,j_i-1,-m}$, on the hazard for m.[1] We assume an index of the form $z_{m,j_i} = x_i\beta_m + h_{i,j_i-1,-m}\gamma_m + h_{i,j_0,-m}\psi_m$. Hence, the latent component of the hazard is modeled as

$$h^*_{i,j_i,m} = x_i\beta_m + h_{i,j_i-1,-m}\gamma_m + h_{i,j_0,-m}\psi_m + a_{m,j_i} + \varepsilon_{i,j_i,m}, \tag{1}$$

$$m = 1, \ldots, M_0,\ j_i = j_{i0}, \ldots, j_{i,T_i},\ i = 1, \ldots, N$$

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. We approximate a_{m,j_i} with an age spline with knots at ages 35, 45, 55, 65, and 75. 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 absorbing outcome, conditional on being at risk, is defined as

$$h_{i,j_i,m} = \max\{I(h^*_{i,j_i,m} > 0), h_{i,j_i-1,m}\}$$

The occurrence of mortality censors observation of other outcomes in a current year.

A number of restrictions are placed on the way feedback is allowed in the model. Table 7 documents restrictions placed on the transition model. We also include a set of other controls. A list of such controls is given in Table 8 along with descriptive statistics.

We have five other types of outcomes:

1. First, we have binary outcomes which are not an absorbing state, such as starting smoking. 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.

2. 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 the lagged outcome on the right-hand side.

3. The third type of outcomes we consider are censored outcomes, such as financial wealth. For wealth, there are a non-negligible number of observations with zero and negative wealth. For these, we consider two part models where the latent variable is specified as in (1) but model probabilities only when censoring does not occur. In total, we have M outcomes.

4. The fourth type of outcomes are continuous outcomes modeled with ordinary least squares. For example, we model transitions in log(BMI). We allow for state-dependence by including the lagged outcome on the right-hand side.

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

5. The final type of models are categorical, but without an ordering. For example, an individual can transition to being out of the labor force, unemployed, or working (either full- or part-time). In situations like this, we utilize a multinomial logit model, including the lagged outcome on the right-hand side.

The parameters $\theta_1 = \left(\{\beta_m, \gamma_m, \psi_m, \varsigma_m\}_{m=1}^{M}, \right)$, can be estimated by maximum 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 initial conditions $h_{i,j_0,-m}$ is the product of normal univariate probabilities. Since these sequences, conditional on initial conditions, are also independent across diseases, the joint probability over all disease-specific sequences is simply the product of those probabilities.

For a given respondent 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)

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

We use the -0 superscript to make explicit the conditioning on $\mathbf{h}_{i,j_{i0}} = (h_{i,j_{i0},0}, \ldots, h_{i,j_{i0},M})'$. We have limited information on outcomes prior to this age. The likelihood is a product of M terms with the mth term containing only $(\beta_m, \gamma_m, \psi_m, \varsigma_m)$. This allows the estimation to be done seperately for each outcome.

3.1.1 Further Details on Specific Transition Models

This section describes the modeling strategy for particular outcomes.

Employment Status Ultimately, we wish to simulate if an individual is out of the labor force, unemployed, working part-time, or working full-time at time t. We treat the estimation of this as a two-stage process. In the first stage, we predict if the individual is out of the labor force, unemployed, or working for pay using a multinomial logit model. Then, conditional on working for pay, we estimate if the individual is working part- or full-time using a probit model.

Earnings We estimate last calendar year earnings models based on the current employment status, controlling for the prior employment status. Of particular concern are individuals with no earnings, representing approximately twenty-five percent of the unemployed and seventy-eight percent of those out of the labor force. This group is less than 0.5% of the full- and part-time populations. We use a two-stage process for those out of the labor force and unemployed. The first stage is a probit that estimates if the individual has any earnings. The second stage is an OLS model of log(earnings) for those with non-zero earnings. For those working full- or part-time, we estimate OLS models of log(earnings).

Relationship Status We are interested in three relationship statuses: single, cohabitating, and married. In each case, we treat the transition from time t to time $t+1$ as a two-stage process. In the first stage, we estimate if the individual will remain in their current status. In the second stage, we estimate which of the two other states the individual will transition to, conditional on leaving their current state.

Childbearing We estimate the number of children born in two-years separately for women and men. We model this using an ordered probit with three categories: no new births, one birth, and two births. Based on the PSID data, we found the exclusion of three or more births in a two-year period to be appropriate.

Body mass index We model body mass index (BMI) log-linearly as a function of sex, age, race/ethnicity, marital status, economic status as a child, self-reported health as a child, and lags of BMI. To account for the persistence of BMI within the same person over time, we explored the use of multi-year lags of BMI in the prediction model. Previous work has shown that inclusion of 2-year lags results in BMI projections that have high internal validity with PSID, and high external validity compared to the Behavioral Risk Factor Surveillance System (Tysinger, 2020). Compared to 2-year lags of BMI, including 4-year lags resulted in slightly better predictions at the mean and right tail. Including 6-year lags of BMI did not result in noticeably better predictions relative to 4-year lags. As such, the BMI prediction model employs 4-year lags.

3.1.2 Inverse Hyperbolic Sine Transformation

One problem fitting the wealth distribution is that it has a long right tail and some negative values. We use a generalization of the inverse hyperbolic sine transform (IHT) presented in MacKinnon and Magee (1990). First denote the variable of interest y. The hyperbolic sine transform is

$$y = \sinh(x) = \frac{\exp(x) - \exp(-x)}{2} \tag{2}$$

The inverse of the hyperbolic sine transform is

$$x = \sinh^{-1}(y) = h(y) = \log(y + (1 + y^2)^{1/2})$$

Consider the inverse transformation. We can generalize such transformation, first allowing for a shape parameter θ,

$$r(y) = h(\theta y)/\theta \tag{3}$$

Such that we can specify the regression model as

$$r(y) = x\beta + \varepsilon, \varepsilon \sim \mathrm{N}(0, \sigma^2) \tag{4}$$

A further generalization is to introduce a location parameter ω such that the new transformation becomes

$$g(y) = \frac{h(\theta(y + \omega)) - h(\theta\omega)}{\theta h'(\theta\omega)} \tag{5}$$

where $h'(a) = (1 + a^2)^{-1/2}$.

We specify (4) in terms of the transformation g. The shape parameters can be estimated from the concentrated likelihood for θ, ω. We can then retrieve β, σ by standard OLS.

Upon estimation, we can simulate

$$\tilde{g} = x\hat{\beta} + \sigma\tilde{\eta}$$

where η is a standard normal draw. Given this draw, we can retransform using (5) and (2)

$$h(\theta(y + \omega)) = \theta h'(\theta\omega)\tilde{g} + h(\theta\omega)$$

$$\tilde{y} = \frac{\sinh\left[\theta h'(\theta\omega)\tilde{g} + h(\theta\omega)\right] - \theta\omega}{\theta}$$

The included estimates table (estimates_FAM.xml) gives parameter estimates for the transition models.

4 Model for replenishing cohorts

We first discuss the empirical strategy, then present the model and estimation results. The model for replinishing cohorts integrates information coming from trends among younger cohorts with the joint distribution of outcomes in the current population of age 25 respondents in the PSID.

4.1 Model and estimation

Assume the latent model for $\mathbf{y}_i^* = (y_{i1}^*, \ldots, y_{iM}^*)'$,

$$\mathbf{y}_i^* = \mu + \varepsilon_i,$$

where ε_i is normally distributed with mean zero and covariance matrix $\mathbf{\Omega}$. It will be useful to write the model as

$$\mathbf{y}_i^* = \mu + \mathbf{L}_\Omega \eta_i,$$

where $\mathbf{L}_\Omega$ is a lower triangular matrix such that $\mathbf{L}_\Omega \mathbf{L}_\Omega' = \mathbf{\Omega}$ and $\eta_i = (\eta_{i1}, \ldots, \eta_{iM})'$ 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.

The vector μ can depend on some variables which have a stable distribution over time $\mathbf{z}_i$ (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

$$\mu_i = \mathbf{z}_i \beta$$

and the whole analysis is done conditional on $\mathbf{z}_i$.

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. We also fix to zero the correlation between selected outcomes (say earnings) and their selection indicator. Hence, we consider two-part models for these outcomes. Because some parameters are naturally bounded, we also re-parameterize the problem to guarantee an interior solution. In particular, we parameterize

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

and estimate the $(\delta_{m,m}, \xi_{m,n}, \gamma_k)$ instead of the original parameters. The parameter values are estimated using the *cmp* package in Stata (Roodman, 2011). Table 9 gives parameter estimates for the indices while Table 10 gives parameter estimates of the covariance matrix in the outcomes.

4.2 Trends for replenishing cohorts

Using the jointly estimated models previously described, we then assign outcomes to the replenishing cohorts, imposing trends for some health, risk factor, and social outcomes. We currently impose trends on BMI, education, number of children, marital status, hypertension, and smoking status for these 25-26 year olds. These trends are estimated using the National Health Interview Survey (health and risk factors) or the American Community Survey (social outcomes). All trends are halted after 2029. The trends are shown in Table 11, Table 12 and Table 13.

5 Government revenues and expenditures

This gives a limited overview of how revenues and expenditures of the government are computed.

5.1 Medical costs estimation

In the FAM, a cost module links a person's current state–demographics, economic status, current health, risk factors, and functional status to 4 types of individual medical spending. The FAM models: total medical spending (medical spending from all payment sources), Medicare spending[2], Medicaid spending (medical spending paid by Medicaid), and out of pocket spending (medical spending by the respondent). These estimates are based on pooled weighted least squares regressions of each type of spending on risk factors, self-reported conditions, and functional status, with spending inflated to constant dollars using the medical component of the consumer price index. We use the 2007-2010 Medical Expenditure Panel Survey for these regressions for persons not Medicare eligible, and the 2007-2010 Medicare Current Beneficiary Survey for spending for those that are eligible for Medicare. Those eligible for Medicare include people eligible due to age (65+) or due to disability status. Comparisons of prevalences and question wording across these different sources are provided in Tables 3 and 4, respectively.

In the baseline scenario, this spending estimate can be interpreted as the resources consumed by the individual given the manner in which medicine is practiced in the United States during the post-part D era (2006-2010). Models are estimated for total, Medicaid, out of pocket spending, and for the Medicare spending.

Since Medicare spending has numerous components (Parts A and B are considered here), models are needed to predict enrollment. In 2004, 98.4% of all Medicare enrollees, and 99%+ of aged enrollees, were in Medicare Part A, and thus we assume that all persons eligible for Medicare take Part A. We use the 2007-2010 MCBS to model take up of Medicare Part B for both new enrollees into Medicare, as well as current enrollees without Part B. Estimates are based on weighted probit regression on various risk factors, demographic, and economic conditions. The PSID starting population for the FAM does not contain information on Medicare enrollment. Therefore another model of Part B enrollment for all persons eligible for Medicare is estimated via a probit, and used in the first year of simulation to assign initial Part B enrollment status. Estimation results are shown in estimates table. The MCBS data over represents the portion enrolled in Part B, having a 97% enrollment rate in 2004 instead of the 93.5% rate given by Medicare Trustee's Report. In addition to this baseline enrollment probit, we apply an elasticity to premiums of -0.10, based on the literature and simulation calibration for actual uptake through 2009 (Atherly et al., 2004; Buchmueller, 2006). The premiums are computed using average Part B costs from the previous time step and the means-testing thresholds established by the ACA.

Since 2006, the Medicare Current Beneficiaries Survey (MCBS) contains data on Medicare Part D. The data gives the capitated Part D payment and enrollment. When compared to the summary data presented in the CMS 2007 Trustee Report, the 2006 per capita cost is comparable between the MCBS and the CMS. However, the enrollment is underestimated in the MCBS, 53% compared to 64.6% according to CMS.

A cross-sectional probit model is estimated using years 2007-2010 to link demographics, economic status, current health, and functional status to Part D enrollment - see the estimates table. To account for both the initial under reporting of Part D enrollment in the MCBS, as well as the CMS

[2] We estimate annual medical spending paid by specific parts of Medicare (Parts A, B, and D) and sum to get the total Medicare expenditures.

prediction that Part D enrollment will rise to 75% by 2012, the constant in the probit model is increased by 0.22 in 2006, to 0.56 in 2012 and beyond. The per capita Part D cost in the MCBS matches well with the cost reported from CMS. An OLS regression using demographic, current health, and functional status is estimated for Part D costs based on capitated payment amounts.

The Part D enrollment and cost models are implemented in the Medical Cost module. The Part D enrollment model is executed conditional on the person being eligible for Medicare, and the cost model is executed conditional on the enrollment model leading a true result, after the Monte Carlo decision. Otherwise the person has zero Part D cost. The estimated Part D costs are added with Part A and B costs to obtain total Medicare cost, and any medical cost growth assumptions are then applied.

6 Implementation

The FAM is implemented in multiple parts. Estimation of the transition and cross sectional models is performed in Stata. The replenishing cohort model is estimated in Stata using the CMP package (Roodman, 2011). The simulation is implemented in C++ for speed and flexibility. Currently, the simulation is run on Linux, Windows, and Mac OS X.

To match the two year structure of the PSID data used to estimate the transition models, the FAM simulation proceeds in two year increments. The end of each two year step is designed to occur on July 1st to allow for easier matching to population forecasts from Social Security. A simulation of the FAM proceeds by first loading a population representative of the age 25+ US population in 2009, generated from PSID. In two year increments, the FAM applies the transition models for mortality, health, working, wealth, earnings, and benefit claiming with Monte Carlo decisions to calculate the new states of the population. The population is also adjusted by immigration forecasts from the US Census Department, stratified by race and age. If incoming cohorts are being used, the new 25/26 year olds are added to the population. The number of new 25/26 year olds added is consistent with estimates from the Census, stratified by race. Once the new states have been determined and new 25/26 year olds added, the cross sectional models for medical costs are performed. Summary variables are then computed. Computation of medical costs includes the persons that died to account for end of life costs. To reduce uncertainty due to the Monte Carlo decision rules, the simulation is performed multiple times (typically 100), and the mean of each summary variable is calculated across repetitions.

FAM simulation takes as inputs assumptions regarding the normal retirement age, real medical cost growth, and interest rates. The default assumptions are taken from the 2010 Social Security Intermediate scenario, adjusted for no price increases after 2010. Therefore simulation results are in real 2009 dollars.

Different simulation scenarios are implemented by changing any of the following components: incoming cohort model, transition models, interventions that adjust the probabilities of specific transition, and changes to assumptions on future economic conditions.

6.1 Intervention Module

The intervention module can adjust characteristics of individuals when they are first read into the simulation "init_interventions" or alter transitions within the simulation "interventions." At present, init_interventions can act on chronic diseases, ADL/IADL status, program participation, and some demographic characteristics. Interventions within the simulation can currently act on mortality, chronic diseases, and some program participation variables.

Interventions can take several forms. The most commonly used is an adjustment to a transition probability. One can also delay the assignment of a chronic condition or cure an existing chronic condition. Additional flexibility comes from selecting who is eligible for the intervention. Some examples might help to make the interventions concrete.

- Example 1: Delay the enrollment into Social Security OASI by two years. In this scenario claiming of Social Security benefits is transitioned as normal. However, if a person is predicted to claim their benefits, then that status is not immediately assigned, but is instead assigned two years later.

- Example 2: Cure hypertension for those with no other chronic diseases. In this scenario any individual with hypertension (including those who have had hypertension for many years) is cured (hypertension status is set to 0), as long as they do not have other chronic diseases. This example uses the individual's chronic disease status as the eligibility criteria for the intervention.

- Example 3: Reduce the incidence of hypertension for half of men aged 55 to 65 by 10% in the first year of the simulation, gradually increasing the reduction to 20% after 10 years. This example begins to show the flexibility in the intervention module. The eligibility criteria are more complex (half of men in a specific age range are eligible) and the intervention changes over time. Mathematically, the intervention works by acting on the incidence probability, ρ. In the first year of the simulation, the probability is replaced by $(1 - 0.5 * 0.1)\,\rho = 0.95\rho$. The binary outcome is then assigned based on this new probability. Thus, at the population level, there is a 5% reduction in incidence for men aged 55 to 65, as desired. After 10 years, the probability for this eligible population becomes $(1 - 0.5 * 0.2)\,\rho = 0.9\rho$.

More elaborate interventions can be programmed by the user.

7 Model development

This section gives some historical background about decisions and developments that led up to the current state of the FAM.

7.1 Quality adjusted life years

We use the generalized risk-adjusted cost-effectiveness framework (GRACE) to calculate quality-adjusted life years (GRA-QALYs) using a two-parameter expo-utility function, $W(H) = c - e^{-bH^a}$. (Lakdawalla and Phelps, 2022; Mulligan et al., 2024).

H denotes a health index, scaled [0,1]. In this context, H is the Visual Analog Scale (VAS), a measure of self-reported overall health with a range of 0-100 (0=worst health, 100=best health). VAS is part of the EQ-5D and is available from the MEPS for the period 2001-2003.

a, b are parameters empirically estimated by Mulligan and colleagues (2024), where $a = 2.1760$, $b = 2.6152$. By convention, utility is zero when the health state is zero, which is achievable by setting c equal to 1. $W(H)$ is rescaled by dividing by $W(1)$, such that $W(0) = 0$ and $W(1) = 1$.

We employ a multi-step estimation approach to VAS projections. Using MEPS 2001-2003, we use ordinary least squares to model VAS, using independent variables that are common between MEPS and PSID (disability, chronic condition status, obesity classification, marital status). We use these coefficients in PSID to predict VAS, then develop a VAS prediction model in PSID, and subsequently project VAS in FAM.

The GRACE framework includes an optional disability adjustment–we use a value of 0.851, which is the age- and gender-adjusted utility of the healthy U.S. population from the literature (Pickard et al., 2019)

7.2 Adjustment of BMI

Due to potential bias in self-reported BMI, we adjusted the PSID self-reported BMI distribution to match the NHANES measured BMI distribution, following a method described by Ward and colleagues (Ward et al., 2019). We calculated mean BMI by quantiles of self-reported BMI, mean BMI by quantiles of measured BMI, and the difference between the two (the "bias" amount). We then smoothed the bias with a cubic spline, and used this amount to adjust the self-reported BMI by quantile. Adjustments were made by sex, age, and year. BMI in PSID pre/post-adjustment is shown in Table 6.

8 Validation

We perform cross-validation and external corroboration exercises. Cross-validation is a test of the simulation's internal validity that compares simulated outcomes to actual outcomes. External corroboration compares model forecasts to others' forecasts.

8.1 Cross-validation

The cross-validation exercise randomly samples half of the PSID respondent IDs for use in estimating the transition models. The respondents not used for estimation, but who were present in the PSID sample in 1999, are then simulated from 1999 through 2019. Demographic, health, and economic outcomes are compared between the simulated ("FAM") and actual ("PSID") populations.

Worth noting is how the composition of the population changes in this exercise. In 1999, the sample represents those 25 and older. Since we follow a fixed cohort, the age of the population will increase to 45 and older in 2019. This has consequences for some measures in later years where the eligible population shrinks.

8.1.1 Demographics

Mortality and demographic measures are presented in Tables 14 and 15. Mortality incidence is comparable between the simulated and observed populations. Demographic characteristics do not differ between the two.

8.1.2 Health Outcomes

Binary health outcomes are presented in Table 16. FAM underestimates the prevalence of ADL and IADL limitations compared to the crossvalidation sample. Binary outcomes, like cancer, diabetes, heart diesease, and stroke do not differ. FAM underpredicts hypertension and lung disease compared to the crossvalidation sample.

8.1.3 Health Risk Factors

Risk factors are presented in Table 17. BMI is not statistically different between the two samples. Current smoking is not statistically different, but more individuals in the crossvalidation sample report being former smokers.

8.1.4 Economic Outcomes

Binary economic outcomes are presented in Table 18. FAM underpredicts claiming of federal disability and overpredicts Social Security retirement claiming. Supplemental Security claiming is not statistically different between FAM and the crossvalidation sample. Working for pay is not statistically different.

On the whole, the crossvalidation exercise is reassuring. There are differences that will be explored and improved upon in the future.

8.2 External Corroboration

Finally, we compare FAM population forecasts to Census forecasts of the US population.. Here, we focus on the full PSID population (25 and older) and those 65 and older. For this exercise, we begin the simulation in 2009 and simulate the full population through 2049. Population projections are compared to the 2012 Census projections for years 2012 through 2049. See resuts in Table 19. By 2049, FAM forecasts for 25 and older remain within 2% of Census forecasts.

9 Baseline Forecasts

In this section we present baseline forecasts of the Future Adult Model. The figures show data from the PSID for the 25+ population from 1999 through 2009 and forecasts from the FAM for the 25+ population beginning in 2009.

9.1 Disease Prevalence

Figure 2 depicts the six chronic conditions we project for men. And Figure 3 depicts the historic and forecasted values for women.

Figure 4 shows historic and forecasted levels for any ADL difficulties, three or more ADL difficulties, any IADL difficulties, and two or more IADL difficulties for men 25 and older. Figure 5 shows historic and forecasted levels for any ADL difficulties, three or more ADL difficulties, any IADL difficulties, and two or more IADL difficulties for women 25 and older.

10 Acknowledgments

The Future Elderly Model and Future Adult Model has been developed by a large team over the last decade. Jay Bhattacharya, Eileen Crimmins, Christine Eibner, Étienne Gaudette, Geoff Joyce, Darius Lakdawalla, Pierre-Carl Michaud, and Julie Zissimopoulos have all provided expert guidance. Adam Gailey, Baoping Shang, and Igor Vaynman provided programming and analytic support during the first years of FEM development at RAND. Jeff Sullivan then led the technical development for several years. More recently, the University of Southern California research programming team

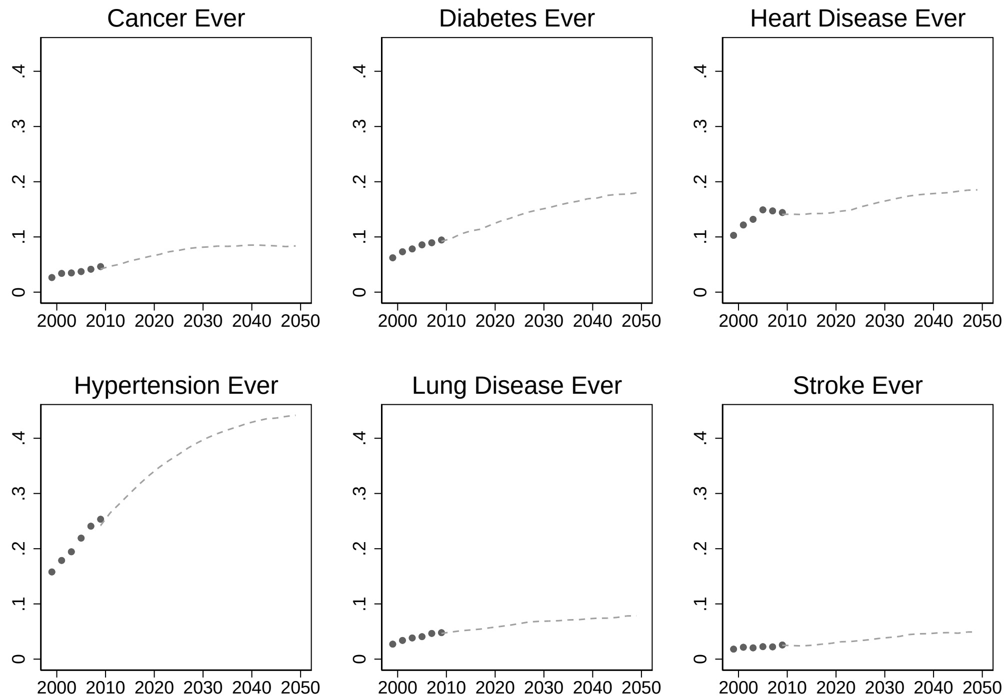

Figure 2: Historic and Forecasted Chronic Disease Prevalence for Men 25+

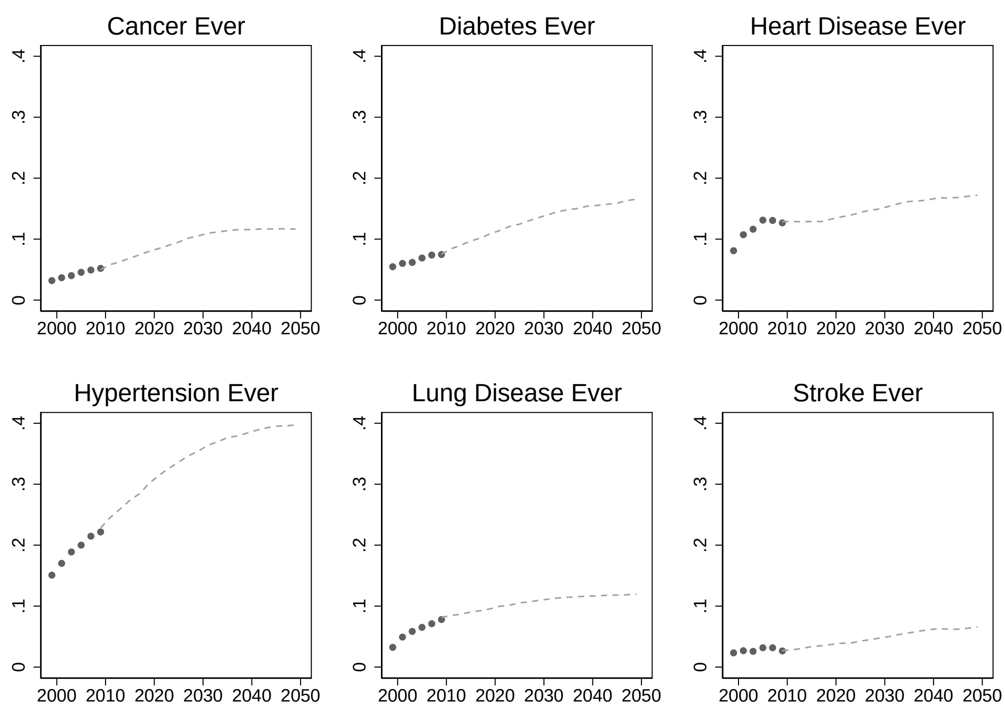

Figure 3: Historic and Forecasted Chronic Disease Prevalence for Women 25+

has supported model development, including FAM development. These programmers include Patricia St. Clair, Laura Gascue, Henu Zhao, and Yuhui Zheng. Barbara Blaylock, Malgorzata Switek,

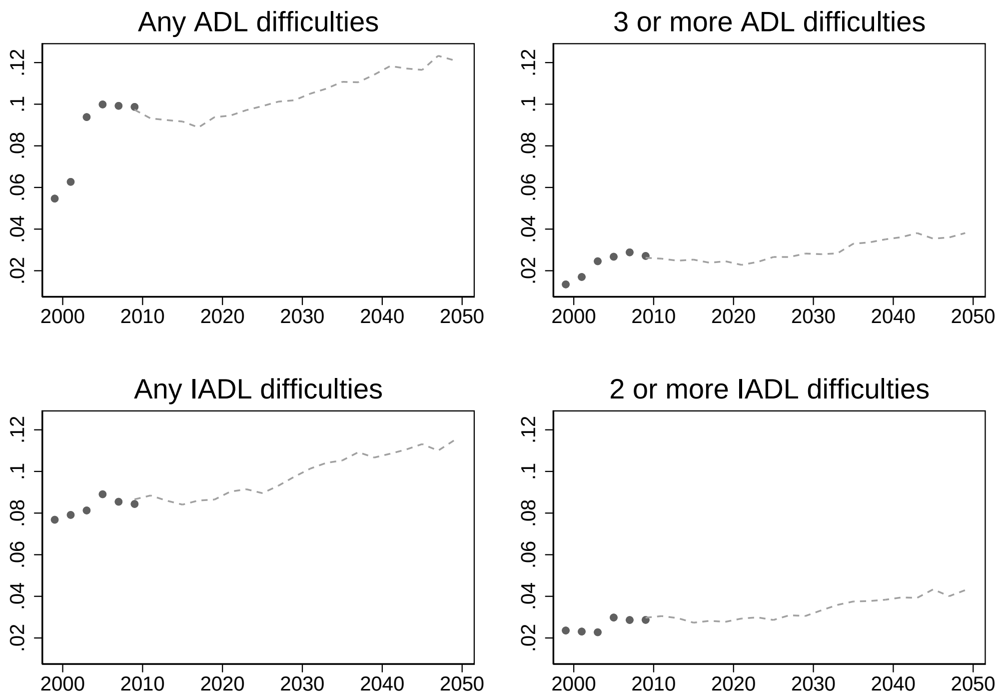

Figure 4: Historic and Forecasted ADL and IADL Prevalence for Men 25+

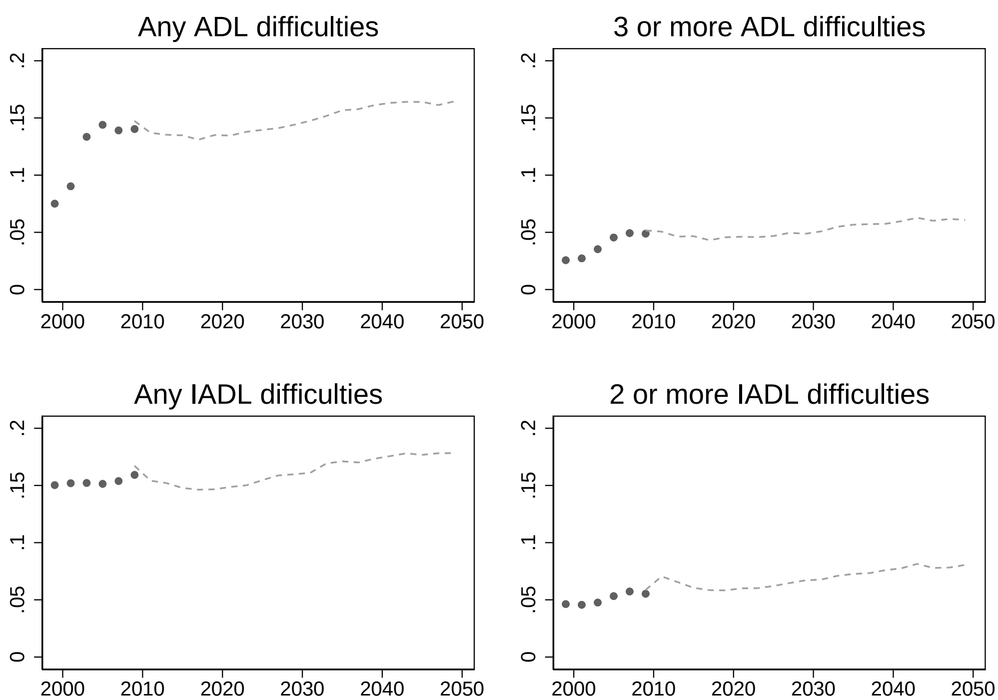

Figure 5: Historic and Forecasted ADL and IADL Prevalence for Women 25+

and Wendy Cheng have greatly aided model development while working as research assistants at USC.

11 Tables

Economic Outcomes	**Health Outcomes**	**Other Outcomes**
Work Status	BMI Category	Education
Earnings	Smoking Category	Partnered
Wealth	Hypertension	Partner Type
		Health Insurance

Table 1: Estimated outcomes in replenishing cohorts module

Economic Outcomes	**Health Outcomes**	**Marital Status**	**Other Outcomes**
Social Security Claiming	Mortality	Exit Single	Insurance Type
Disability Claiming	Heart Disease	Exit Cohabitation	
Non-Zero Capital Income	Cancer	Exit Married	
Capital Income (if non-zero)	Hypertension	Single to Married	
Non-zero Government Transfers	Diabetes	Cohabitation to Married	
Government Transfers (if non-zero)	Lung Disease	Married to Cohabitation	
Non-zero Wealth	Start Smoking		
Wealth (if non-zero)	Stop Smoking		
Labor Force Status (out, unemployed, working)	ADL Status		
Employed Full- or Part-time (if working)	IADL Status		
Any Earnings (if Unemployed)	Births/Paternity		
Any Earnings (if Not in Labor Force)	Self-reported Health		
Earnings (if Full-time)	BMI		
Earnings (if Part-time)	Partner Death		
Earnings (if Unemployed and any)			
Earnings (if Not in Labor Force and any)			

Table 2: Estimated outcomes in transitions module

					Prevalence %			
Source (years, ages)	Cancer	Heart Diseases	Stroke	Diabetes	Hypertension	Lung Disease	Overweight	Obese
HRS (2006-2010, 50-64)	8%	13%	4%	16%	45%	6%	37%	37%
HRS (2006-2010, 65+)	19%	30%	11%	22%	62%	11%	39%	27%
MCBS (2007-2010, 65+)	19%	41%	11%	24%	68%	17%	38%	25%
MEPS (2007-2010, 25-49)	2%	6%	1%	4%	17%	4%	34%	30%
MEPS (2007-2010, 50-64)	6%	16%	4%	14%	46%	7%	36%	34%
MEPS (2007-2010, 65+)	14%	37%	13%	21%	68%	11%	37%	26%
NHIS (2007-2010, 25-49)	2%	6%	1%	4%	16%	4%	34%	31%
NHIS (2007-2010, 50-64)	7%	14%	3%	13%	41%	7%	36%	35%
NHIS (2007-2010, 65+)	17%	32%	9%	19%	61%	10%	36%	27%
PSID (2007-2011, 25-49)	1%	5%	1%	3%	10%	3%	34%	28%
PSID (2007-2011, 50-64)	5%	15%	2%	11%	31%	8%	39%	28%
PSID (2007-2011, 65+)	13%	35%	9%	18%	49%	13%	39%	22%

Table 3: Health condition prevalences in survey data (self-reported)

	Survey			
Disease	PSID/HRS	NHIS	MEPS	MCBS
Cancer	Has a doctor ever told you that you have cancer or a malignant tumor, excluding minor skin cancers?	Have you ever been told be a doctor or other health professional that you had cancer or a malignanacy of any kind? (WHEN RECODED, SKIN CANCERS WERE EXCLUDED)	List all the conditions that have bothered (the person) from (START time) to (END time) CCS codes for the conditions list are 11-21, 24-45	Has a doctor ever told you that you had any (other) kind of cancer malignancy, or tumor other than skin cancer?
Heart Diseases	Four separate questions were asked about whether ever told by a doctor or other health professional that had: CHD, Angina, MI, other heart problems.	Four separate questions: Has a doctor or other health professional ever told you that you ... had congestive heart failure? coronary heart disease? angina? heart attack?	Have you ever been told by a doctor or health professional that you have CHD; Angina; MI; other heart problems	Six separate questions were asked about whether ever told by a doctor that had: Angina or MI; CHD; other heart problems (included four questions)
Stroke	Have you EVER been told by a doctor or other health professional that you had a stroke?	Has a doctor or other health professional ever told you that you ...had a stroke?	If Female, add: [Other than during pregnancy,] Have you ever been told by a doctor or health professional that you have a stroke or TIA (transient ischemic attack)	[Since (PREV< SUPP. RD. INT. DATE),] has a doctor (ever) told (you/SP) that (you/he/she) had a stroke, a brain hemorrhage, or a cerebrovascular accident?
Diabetes	Has a doctor ever told you that you have diabetes or high blood sugar?	If Female, add: [Other than during pregnancy,] Have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?	If Female, add: [Other than during pregnancy,] Have you hever been told by a doctor or health professional that you have diabetes or sugar diabetes?	Has a doctor (ever) told (you/SP) that (you/he/she) had diabetes, high blood sugar, or sugar in (your/his/her) urine? [DO NOT INCLUDE BOERDERLINE PREGNANCY, OR PREDIABETEIC DIABETES.]
Hypertension	Has a doctor ver told you that you have high blood pressure or hypertension?	Have you EVER been told by a doctor or other health professional that you had Hypertension, also called high blood pressure?	Have you EVER been told by a doctor or other health professional that you had Hypertension, also called high blood pressure?	Has a doctor (ever) told (you/SP) that (you/he/she) (still) (had) (have/has) hypertension, sometimes called high blood pressure?
Lung Disease	Has a doctor ever told you that you have chronic lung disease such a schronic bronchitis or emphysema? [IWER: DO NOT INCLUDE ASTHMA]	Question 1: During the PAST 12 MONTHS, have you ever been told by a doctor or other health professional that you had chronic bronchitis? Question 2: Have you EVER been told by a docotor or other health professional that you had emphysema?	List all the conditions that have bothered (the person) from (START time) to (END time) CCS codes for the conditions list are 127, 129-312	Has a doctor (ever) told (you/SP) that (you/he/she) had emphysema, asthma, or COPD? [COPD=CHRONIC OBSTRUCTIVE PULMONARY DISEASE.]
Overweight	Self-reported body weight and height			
Obese				

Table 4: Survey questions used to determine health conditions

			Type	At risk	Mean/fraction
Disease	heart disease		biennial incidence	undiagnosed	0.01
	hypertension		biennial incidence	undiagnosed	0.04
	stroke		biennial incidence	undiagnosed	0.01
	lung disease		biennial incidence	undiagnosed	0.01
	cancer		biennial incidence	undiagnosed	0.01
	diabetes		biennial incidence	undiagnosed	0.01
Risk Factors	Smoking Status	never smoked	ordered	all	0.51
		ex smoker	ordered	all	0.30
		current smoker	ordered	all	0.19
	Log BMI		continuous	all	3.31
	ADL Status	no ADLs	ordered	all	0.90
		1 ADL	ordered	all	0.05
		2 ADLS	ordered	all	0.02
		3+ ADLS	ordered	all	0.03
	IADL Status	no IADLs	ordered	all	0.89
		1 IADL	ordered	all	0.07
		2+ IADLs	ordered	all	0.04
LFP & Benefits	Employment Status	out of labor force	prevalence	all	0.26
		unemployed	prevalence	all	0.06
		part time	prevalence	all	0.17
		full time	prevalence	all	0.51
	SS benefit receipt		biennial incidence	eligible & not receiving	
	DI benefit receipt		prevalence	eligible & age < 65	0.04
	Any health insurance		prevalence	age < 65	0.84
	SSI receipt		prevalence	all	0.02
Family	Marital status	single	prevalence	all	0.28
		cohabitating	prevalence	all	0.09
		married	prevalence	all	0.62
	Childbearing	no children	biennial incidence	female	0.91
		1 child	biennial incidence	female	0.08
		2 children	biennial incidence	female	0.00
Financial	financial wealth		median	all non-zero wealth	56.00
Resources	earnings		median	working full time	17.50
	earnings		median	working part time	40.38
($K 2009)	wealth non-zero		prevalence	all	0.95

Table 5: Outcomes in the transition model. Estimation sample is PSID 1999-2019 waves.

	Mean BMI			Obesity Prevalence %		
Ages	Measured NHANES	Self-reported PSID	Adjusted PSID	Measured NHANES	Self-reported PSID	Adjusted PSID
25-49	28.69	27.59	28.98	34%	28%	36%
50-64	29.44	27.92	29.45	40%	28%	39%
65+	28.57	26.86	28.44	34%	22%	34%

Table 6: Comparison of BMI variables NHANES and PSID, 2007-2011. Sample includes ages 25+.

Value at time $T-1$	Outcome at time T																			
	Heart disease	hypertension	stroke	Lung disease	diabetes	cancer	disability	mortality	Smoking status	BMI	Any HI	DI Claim	SS Claim	DB Claim	SSI Claim	Nursing Home	Work	Earnings	Nonzero Wealth	Wealth
Heart disease			✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blood pressure	✓		✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Stroke							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lung disease							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Diabetes	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cancer			✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Disability							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Claimed DI											✓	✓	✓	✓	✓		✓	✓	✓	✓
Claimed SS											✓			✓	✓		✓	✓	✓	✓
Claimed DB													✓		✓		✓	✓	✓	✓
Claimed SSI															✓					
Work											✓	✓	✓		✓		✓	✓	✓	✓
Earnings											✓	✓	✓	✓	✓		✓	✓	✓	✓
Nonzero wealth											✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wealth											✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Nursing home stay															✓	✓			✓	✓

Table 7: Restrictions on transition model. ✓indicates that an outcome at time $T-1$ is allowed in the transition model for an outcome at time T.

Control variable	Mean	Standard deviation	Minimum	Maximum
Non-hispanic black	0.112	0.315	0	1
Hispanic	0.127	0.333	0	1
Single	0.343	0.475	0	1
Cohabitating	0.0540	0.226	0	1
Married	0.603	0.489	0	1
Less than high school	0.133	0.339	0	1
High school/GED/some college/AA	0.552	0.497	0	1
College graduate	0.210	0.407	0	1
More than college	0.105	0.307	0	1
Doctor ever - heart disease	0.141	0.348	0	1
Doctor ever - hypertension	0.243	0.429	0	1
Doctor ever - stroke	0.0287	0.167	0	1
Doctor ever - chronic lung disease	0.0676	0.251	0	1
Doctor ever - cancer	0.0496	0.217	0	1
Doctor ever - diabetes	0.0870	0.282	0	1
Never smoked	0.473	0.499	0	1
Former smoker	0.346	0.476	0	1
Current smoker	0.181	0.385	0	1
No ADL limitations	0.868	0.338	0	1
1 ADL limitation	0.0597	0.237	0	1
2 ADL limitations	0.0262	0.160	0	1
3 or more ADL limitations	0.0458	0.209	0	1
No IADL limitations	0.866	0.341	0	1
1 IADL limitation	0.0859	0.280	0	1
2 or more IADL limitations	0.0480	0.214	0	1
$25 < \text{BMI} < 30$	0.341	0.474	0	1
$30 < \text{BMI} < 35$	0.216	0.412	0	1
$35 < \text{BMI} < 40$	0.0879	0.283	0	1
$\text{BMI} > 40$	0.0676	0.251	0	1
Any Social Security income LCY	0.200	0.400	0	1
Any Disability income LCY	0.0388	0.193	0	1
Any Supplemental Security Income LCY	0.0189	0.136	0	1
Any health insurance LCY	0.877	0.329	0	1
Out of labor force	0.318	0.466	0	1
Unemployed	0.0617	0.241	0	1
Working part-time	0.177	0.381	0	1
Working full-time	0.444	0.497	0	1
Earnings in 1000s capped at 200K	34.01	40.03	0	200
Wealth in 1000s capped at 2 million	270.0	457.3	-1974	2000

Table 8: Desciptive statistics for variables in 2009 PSID ages 25+ sample used as simulation stock population

Covariate	Education level	Partnered	Partnership type	Weight status	Smoking status	Hypertension	In labor force	Number of biological children
Non-hispanic black	-0.31	-0.70	-0.57	0.38	-0.38	0.20	0.12	0.31
Hispanic	-0.04	0.02	-0.14	0.27	-0.53	-0.08	-0.09	0.22
Male	-0.25	0.06	-0.12	-0.16	0.25	0.10	0.46	-0.45
Less than HS/GED	0.00	-0.36	-0.10	0.03	0.76	0.09	-0.19	0.52
College	0.00	0.09	0.07	-0.43	-0.73	-0.13	0.17	-0.82
Beyond college	0.00	0.24	-0.02	-0.80	-1.06	-0.20	-0.25	-1.15
R's mother less than high school	-0.33	-0.07	-0.01	0.00	0.00	0.00	-0.05	0.14
R;s mother some college	0.30	-0.17	0.12	0.00	0.00	0.00	-0.06	-0.07
R's mother college graduate	0.58	-0.27	-0.01	0.00	0.00	0.00	0.07	-0.17
R's father less than high school	-0.15	-0.07	0.04	0.00	0.00	0.00	-0.04	-0.04
R;s father some college	0.31	-0.22	0.09	0.00	0.00	0.00	0.02	-0.20
R's father college graduate	0.73	-0.24	0.03	0.00	0.00	0.00	-0.03	-0.24
Poor as a child	-0.20	0.04	-0.01	0.00	0.00	0.00	-0.10	0.09
Wealthy as a child	-0.05	-0.06	-0.05	0.00	0.00	0.00	-0.04	0.10
Fair or poor health before age 17	-0.17	-0.08	-0.02	0.00	0.00	0.00	-0.19	-0.06
Age 25 or 26	-0.15	-0.17	-0.23	-0.12	-0.05	-0.15	-0.07	-0.33
Constant	1.45	0.94	0.80	0.48	0.13	-1.82	1.01	0.64

Table 9: Parameter estimates for latent model: conditional means and thresholds. Sample is respondents age 25-30 in 2005-2011 PSID waves

	Education level	Partnered	Partnership type	Weight status	Smoking status	Hypertension	In labor force	Number of biological children
Education level	1.000							
Partnered	-0.107	1.000						
Partnership type	0.098	0.000	1.000					
wtstateadj	0.144	-0.011	0.061	1.000				
Smoking status	0.012	-0.127	-0.204	-0.019	1.000			
Hypertension	0.009	-0.084	0.066	0.324	0.015	1.000		
In labor force	0.102	-0.157	-0.025	-0.008	0.002	-0.012	1.000	
Number of biological children	0.030	0.346	0.260	0.045	-0.001	0.027	-0.176	1.000

Table 10: Parameter estimates for latent model: parameterized covariance matrix. Sample is respondents age 25-30 in 2005-2011 PSID waves

Year	Hypertension	Overweight	Obese 1	Obese 2	Obese 3	Never Smoked	Former Smoker	Current Smoker
2009	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2010	1.00	1.00	1.03	1.01	1.01	1.00	1.00	0.99
2011	0.99	1.00	1.06	1.01	1.03	1.01	0.99	0.98
2012	0.99	1.00	1.08	1.01	1.04	1.01	0.99	0.98
2013	1.00	1.00	1.10	1.02	1.06	1.01	0.99	0.97
2014	1.00	1.01	1.12	1.02	1.07	1.02	0.98	0.96
2015	1.00	1.00	1.14	1.03	1.08	1.02	0.98	0.95
2016	1.01	1.02	1.17	1.03	1.10	1.03	0.98	0.94
2017	1.01	1.03	1.19	1.04	1.11	1.03	0.97	0.94
2018	0.98	1.09	1.20	1.05	1.13	1.03	0.97	0.93
2019	0.98	1.09	1.23	1.06	1.14	1.04	0.97	0.92
2020	0.99	1.09	1.24	1.08	1.16	1.04	0.96	0.91
2021	0.99	1.09	1.26	1.09	1.17	1.04	0.96	0.91
2022	0.98	1.09	1.28	1.11	1.19	1.05	0.95	0.90
2023	0.98	1.09	1.29	1.13	1.20	1.05	0.95	0.89
2024	0.98	1.08	1.30	1.15	1.22	1.05	0.95	0.88
2025	0.98	1.07	1.31	1.18	1.24	1.06	0.94	0.87
2026	0.99	1.06	1.32	1.21	1.25	1.06	0.94	0.87
2027	0.99	1.04	1.34	1.24	1.27	1.06	0.94	0.86
2028	0.99	0.98	1.43	1.25	1.28	1.07	0.93	0.85
2029	1.01	0.97	1.46	1.26	1.30	1.07	0.93	0.84
2030	1.01	0.97	1.46	1.26	1.30	1.07	0.93	0.84
2031	1.01	0.97	1.46	1.26	1.30	1.07	0.93	0.84
2032	1.01	0.97	1.46	1.26	1.30	1.07	0.93	0.84
2033	1.01	0.97	1.46	1.26	1.30	1.07	0.93	0.84
2034	1.01	0.97	1.46	1.26	1.30	1.07	0.93	0.84
2035	1.01	0.97	1.46	1.26	1.30	1.07	0.93	0.84

Table 11: Health and risk factor trends for replenishing cohorts, prevalences relative to 2009

Year	Less than HS	HS Grad	College Grad	Graduate School
2009	1.00	1.00	1.00	1.00
2010	0.98	0.99	1.02	1.04
2011	0.96	0.98	1.03	1.08
2012	0.93	0.97	1.05	1.12
2013	0.91	0.96	1.06	1.17
2014	0.89	0.95	1.08	1.21
2015	0.87	0.94	1.09	1.26
2016	0.85	0.93	1.11	1.31
2017	0.83	0.92	1.12	1.36
2018	0.81	0.91	1.14	1.41
2019	0.79	0.90	1.15	1.46
2020	0.77	0.88	1.16	1.51
2021	0.75	0.87	1.18	1.57
2022	0.73	0.86	1.19	1.63
2023	0.71	0.85	1.20	1.68
2024	0.69	0.84	1.21	1.74
2025	0.67	0.83	1.23	1.80
2026	0.66	0.81	1.24	1.87
2027	0.64	0.80	1.25	1.93
2028	0.62	0.79	1.26	1.99
2029	0.60	0.78	1.27	2.06
2030	0.60	0.78	1.27	2.06
2031	0.60	0.78	1.27	2.06
2032	0.60	0.78	1.27	2.06
2033	0.60	0.78	1.27	2.06
2034	0.60	0.78	1.27	2.06
2035	0.60	0.78	1.27	2.06

Table 12: Education trends for replenishing cohorts, prevalences relative to 2009

Year	No Children	One Child	Two Children	Three Chidren	Four or More Children	Partnered	Married
2009	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2010	1.01	1.00	0.99	0.98	0.98	1.00	0.98
2011	1.01	0.99	0.98	0.97	0.95	0.99	0.96
2012	1.02	0.99	0.97	0.95	0.93	0.99	0.94
2013	1.03	0.98	0.96	0.93	0.90	0.99	0.91
2014	1.03	0.98	0.95	0.91	0.88	0.99	0.89
2015	1.04	0.97	0.94	0.90	0.86	0.98	0.87
2016	1.05	0.97	0.92	0.88	0.84	0.98	0.85
2017	1.05	0.96	0.91	0.87	0.82	0.98	0.82
2018	1.06	0.95	0.90	0.85	0.79	0.98	0.80
2019	1.07	0.95	0.89	0.83	0.77	0.98	0.78
2020	1.07	0.94	0.88	0.82	0.75	0.98	0.76
2021	1.08	0.94	0.87	0.80	0.73	0.98	0.73
2022	1.09	0.93	0.86	0.79	0.72	0.97	0.71
2023	1.09	0.93	0.85	0.77	0.70	0.97	0.69
2024	1.10	0.92	0.84	0.76	0.68	0.97	0.66
2025	1.11	0.92	0.83	0.75	0.66	0.97	0.64
2026	1.11	0.91	0.82	0.73	0.64	0.98	0.62
2027	1.12	0.90	0.81	0.72	0.63	0.98	0.60
2028	1.13	0.90	0.80	0.70	0.61	0.98	0.57
2029	1.13	0.89	0.79	0.69	0.59	0.98	0.55
2030	1.13	0.89	0.79	0.69	0.59	0.98	0.55
2031	1.13	0.89	0.79	0.69	0.59	0.98	0.55
2032	1.13	0.89	0.79	0.69	0.59	0.98	0.55
2033	1.13	0.89	0.79	0.69	0.59	0.98	0.55
2034	1.13	0.89	0.79	0.69	0.59	0.98	0.55
2035	1.13	0.89	0.79	0.69	0.59	0.98	0.55

Table 13: Social trends for replenishing cohorts, prevalences relative to 2009

	2001			2007			2013			2019		
	FAM	PSID		FAM	PSID		FAM	PSID		FAM	PSID	
Outcome	mean	mean	p	mean	mean	p	mean	mean	p	mean	mean	p
Died	0.014	0.018	0.040	0.020	0.023	0.262	0.025	0.026	0.796	0.031	0.032	0.804

Table 14: Crossvalidation of 1999 cohort: Mortality in 2001, 2007, 2013, and 2019

	2001			2007			2013			2019		
	FAM	PSID		FAM	PSID		FAM	PSID		FAM	PSID	
Outcome	mean	mean	p	mean	mean	p	mean	mean	p	mean	mean	p
Age on July 1st	48.965	49.196	0.316	53.082	53.379	0.192	56.939	57.960	0.000	60.803	62.684	0.000
Black	0.102	0.090	0.005	0.103	0.088	0.001	0.102	0.092	0.045	0.101	0.094	0.199
Hispanic	0.075	0.072	0.347	0.080	0.086	0.198	0.087	0.096	0.091	0.094	0.091	0.697
Male	0.464	0.467	0.661	0.463	0.463	0.995	0.460	0.458	0.749	0.456	0.459	0.752

Table 15: Crossvalidation of 1999 cohort: Demographic outcomes in 2001, 2007, 2013, and 2019

	2001			2007			2013			2019		
	FAM	PSID		FAM	PSID		FAM	PSID		FAM	PSID	
Outcome	mean	mean	p	mean	mean	p	mean	mean	p	mean	mean	p
Any ADLs	0.080	0.070	0.006	0.108	0.126	0.000	0.130	0.142	0.056	0.151	0.204	0.000
Any IADLs	0.100	0.115	0.001	0.114	0.130	0.003	0.136	0.170	0.000	0.156	0.210	0.000
Cancer	0.036	0.036	0.959	0.063	0.052	0.002	0.089	0.073	0.000	0.120	0.125	0.396
Diabetes	0.064	0.062	0.417	0.096	0.088	0.072	0.131	0.122	0.096	0.168	0.176	0.241
Heart Disease	0.098	0.110	0.007	0.129	0.152	0.000	0.159	0.173	0.027	0.194	0.214	0.012
Hypertension	0.176	0.168	0.126	0.267	0.252	0.024	0.354	0.329	0.002	0.443	0.458	0.134
Lung Disease	0.037	0.038	0.795	0.062	0.058	0.227	0.083	0.091	0.125	0.105	0.124	0.002
Stroke	0.019	0.021	0.330	0.027	0.031	0.102	0.035	0.035	0.995	0.045	0.057	0.009

Table 16: Crossvalidation of 1999 cohort: Binary health outcomes in 2001, 2007, 2013, and 2019

	2001			2007			2013			2019		
	FAM	PSID		FAM	PSID		FAM	PSID		FAM	PSID	
Outcome	mean	mean	p	mean	mean	p	mean	mean	p	mean	mean	p
BMI	28.393	28.160	0.010	29.148	28.964	0.069	29.681	29.310	0.001	30.072	29.940	0.316
Current smoker	0.181	0.201	0.001	0.152	0.167	0.014	0.130	0.146	0.007	0.108	0.119	0.059
Ever smoked	0.471	0.511	0.000	0.469	0.525	0.000	0.461	0.531	0.000	0.452	0.532	0.000

Table 17: Crossvalidation of 1999 cohort: Risk factor outcomes in 2001, 2007, 2013, and 2019

	2001			2007			2013			2019		
	FAM	PSID		FAM	PSID		FAM	PSID		FAM	PSID	
Outcome	mean	mean	p	mean	mean	p	mean	mean	p	mean	mean	p
Claiming SSDI	0.016	0.023	0.001	0.020	0.033	0.000	0.029	0.049	0.000	0.029	0.066	0.000
Claiming OASI	0.184	0.199	0.009	0.209	0.218	0.157	0.265	0.284	0.016	0.339	0.386	0.000
Claiming SSI	0.017	0.014	0.135	0.015	0.016	0.580	0.014	0.017	0.157	0.015	0.017	0.296
Working for pay	0.665	0.684	0.004	0.637	0.657	0.006	0.605	0.598	0.407	0.552	0.501	0.000

Table 18: Crossvalidation of 1999 cohort: Binary economic outcomes in 2001, 2007, 2013, and 2019

Year	Census 25+	FAM Minimal 25+	Census 65+	FAM Minimal 65+
2009	202.1	202.0	39.6	39.4
2011	206.6	205.5	41.4	40.1
2013	211.0	209.0	44.7	42.8
2015	215.9	213.2	47.7	45.6
2017	220.9	217.6	50.8	48.2
2019	225.5	222.2	54.2	50.9
2021	229.8	226.0	57.7	53.9
2023	233.9	229.9	61.4	56.5
2025	238.0	233.8	65.1	60.1
2027	241.9	237.8	68.4	63.7
2029	245.7	241.6	71.4	67.3
2031	249.3	245.4	73.8	70.2
2033	252.9	248.7	75.5	71.6
2035	256.0	251.4	77.3	73.6
2037	259.2	253.9	78.8	73.8
2039	262.6	256.9	79.4	74.1
2041	265.8	259.8	79.9	73.8
2043	269.0	262.8	80.4	74.9
2045	272.2	265.8	81.3	76.0
2047	275.3	268.8	82.2	76.9
2049	278.4	271.4	83.2	77.4

Table 19: Population forecasts: Census compared to FAM

References

Atherly, A., Dowd, B. E., and Feldman, R. (2004). The effect of benefits, premiums, and health risk on health plan choice in the medicare program. *Health services research*, 39(4p1):847–864.

Buchmueller, T. (2006). Price and the health plan choices of retirees. *Journal of Health Economics*, 25(1):81–101.

Goldman, D. P., Shekelle, P. G., Bhattacharya, J., Hurd, M., and Joyce, G. F. (2004). Health status and medical treatment of the future elderly. Technical report, DTIC Document.

Lakdawalla, D. N. and Phelps, C. E. (2022). A guide to extending and implementing generalized risk-adjusted cost-effectiveness (grace). *European Journal of Health Economics*, 23(3):433–451.

MacKinnon, J. G. and Magee, L. (1990). Transforming the dependent variable in regression models. *International Economic Review*, pages 315–339.

Mulligan, K., Baid, D., Doctor, J. N., Phelps, C. E., and Lakdawalla, D. N. (2024). Risk preferences over health: Empirical estimates and implications for medical decision-making. *Journal of Health Economics*, 94:102857.

Pickard, A. S., Law, E. H., Jiang, R., Pullenayegum, E., Shaw, J. W., Xie, F., Oppe, M., Boye, K. S., Chapman, R. H., Gong, C. L., Balch, A., and Busschbach, J. J. (2019). United states valuation of eq-5d-5l health states using an international protocol. *Value in Health*, 22(8):931–941.

Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *Stata Journal*, 11(2):159–206(48).

Smith, K. and Favreault, M. (2013). A primer on modeling income in the near term, version 7. Technical report, The Urban Institute.

Tysinger, B. (2020). Validating risk factor and chronic disease projections in the future adult model. *International Journal of Microsimulation*, 13:54–69.

Ward, Z. J., Bleich, S. N., Cradock, A. L., Barrett, J. L., Giles, C. M., Flax, C., Long, M. W., and Gortmaker, S. L. (2019). Projected u.s. state-level prevalence of adult obesity and severe obesity. *New England Journal of Medicine*, 381(25):2440–2450.