Healthy Habits and Inequality*

Dante Amengual
CEMFI

Jesús Bueren
EUI

Josep Pijoan-Mas
CEMFI and CEPR

July 2024

** ** Conference draft ** **
(Please, do not circulate)

Abstract

The socio-economic gradient of health outcomes and life expectancy has widened over recent decades. We argue that differences in health behaviors across education groups are key to understanding this fact. We start by estimating discrete (latent) types of lifestyles and their impact on health dynamics by use of data on health behavior and health outcomes in the HRS and the PSID. We find that there are large gradients in life expectancy across types (8 years at age 50) and that the higher frequency of health-protective lifestyles among the more educated individuals explains 40% of the education gradient in life expectancy. Next, we build a life cycle model with idiosyncratic labor market and health risks. In the model, education and lifestyles are jointly chosen early in life by individuals who are heterogeneous in the utility cost of adopting protective lifestyles and acquiring education. Importantly, these two early-life investments are complements. We find that the more educated individuals choose healthier lifestyles partly because of their income advantage (life is more enjoyable with higher consumption possibilities), partly because of the higher yield of their health-protective lifestyle investments (a better lifestyle has a larger effect on health outcomes for them), and partly due to their better selection in terms of costs of adopting healthier behaviors. Finally, we find that the increase in the college wage premium over the last decades has widened the education gradient in lifestyles, resulting in one-year increase in the education gradient of life expectancy across cohorts born in the 1930s and 1970s. Of this increase, 40% is driven by the direct effect of wage changes, while 60% is due to changes in the composition of the set of individuals not graduating from high school.

*The authors thanks feedback from presentations held at Bristol University, CEMFI, EUI, HEC Montreal, and the 2024 BSE Summer Forum. Emails: amengual@cemfi.es, jesus.bueren@eui.eu, pijoan@cemfi.es

1 Introduction

Western economies have witnessed a sustained increase in income and health inequalities over recent decades. This trend is underpinned by two fundamental observations: a strong correlation between economic and health outcomes (Kitagawa and Hauser, 1973; Pijoan-Mas and Ríos-Rull, 2014; Chetty et al., 2016) and a widening educational gap in health outcomes (Preston and Elo, 1995; Meara et al., 2008; Case and Deaton, 2015). However, the precise mechanisms linking economic status and health outcomes are not well understood. Our study seeks to address this gap by examining the role played by lifestyle factors and health behaviors.

Lifestyle factors and health behaviors —like exercise habits, dietary patterns, or smoking and drinking— are important determinants of health outcomes. McGinnis and Foege (1993) seminal paper estimates that around half of the deaths that occurred in the US in 1990 resulted from risk factors arising from bad health behaviors. Many papers have followed showing that individuals who engage in healthier behaviors are more likely to experience positive health outcomes, while those with unhealthy habits are at greater risk of developing chronic diseases and experiencing premature mortality (Taylor et al., 2002; Li et al., 2018; Zaninotto et al., 2020). Research has also consistently shown that individuals with higher levels of education tend to adopt healthier lifestyles (Lantz et al., 1998; Cutler and Lleras-Muney, 2010; Polvinen et al., 2013). Hence, heterogeneous lifestyle choices may contribute to the educational gradient in health outcomes if both the education gradient in health behavior and the effect of health behavior on health outcomes are sufficiently large.

Our paper tackles three main questions. First, we want to measure the impact of different lifestyle habits on health dynamics, health inequalities, and economic outcomes. Second, we seek to understand the joint determination of education and lifestyle habits early in life, exploring why an education gradient exists in lifestyle habits. Finally, given these ingredients, we aim to quantify the effect of raising labor earnings inequality across education groups on the observed increase in the education gradients of lifestyle and health outcomes. In particular, we want to understand whether the observed deterioration of lifestyle habits by the high school dropouts is the result of the direct effect of their declining economic prospects or, rather, whether it is the result of changes in the composition of this shrinking population group.

Using data from the Health and Retirement Study (HRS) and the Panel Study of Income Dynamics (PSID), we exploit a rich array of health behavior indicators. These include preventive tests, substance abuse, and exercise habits, among others. To understand the determinants of these choices we would like to incorporate all this information into a structural model. However, three challenges arise. First, the observed health behaviors are imperfectly correlated across individuals and over time, which suggests either a noisy measure of these variables or frequent changes in different dimensions of behavior by individuals. Second, estimating the effects of each health behavior on health dynamics is difficult, as different health behaviors are hard to isolate from each other and their health effects in the long run are very hard to identify. Finally, the curse of dimensionality makes it impractical to consider so many different variables in a dynamic model. To address these issues, we make a novel contribution by developing a methodology to reduce the dimensionality of the data. In particular, we identify patterns in lifestyle behavior by assigning individuals to permanent types based on their health behaviors and health dynamics. Hence, the types summarize the propensity of individuals to engage in different health behaviors as well as their health trajectories. Our econometric procedure also estimates the long-run effect of lifestyle choices on health outcomes.

We consider a parsimonious representation of health behavior in two distinct lifestyles, which we label *protective* and *detrimental*. This representation delivers four main results. First, lifestyles have a strong effect on health dynamics. At age 50, there is an 8.6 years life expectancy gap between individuals with protective and detrimental lifestyles. Second, there is a strong correlation between lifestyle habits and education, with harmful behaviors being more prevalent among the less educated. This education gradient in health behaviors explains approximately 40% of the education gradient in life expectancy. Third, the life expectancy differences across lifestyles types are also large within education categories, and more so for college graduates than for high-school dropouts. In particular, the life expectancy gap between individuals with protective and detrimental lifestyles is 6.6 years among high school dropouts and 9.6 years among college graduates. This uncovers an important complementarity between education and lifestyle choices, which we discuss below. Finally, and equally important, we find an increasing dispersion in lifestyles across education groups for individuals born in more recent cohorts.

To understand the joint determination of education and health behavior choices, we propose a heterogeneous agents model comprising two distinct stages. In the first stage, individuals make a one-time early-life education and health behavior choice. In this stage, individuals exhibit heterogeneity in both their costs associated with education and their costs associated with engagement in protective health behaviors. In the second stage, during the working/retirement phase, individuals make consumption and savings choices in a standard life-cycle model with idiosyncratic labor income and health risks, with outcomes conditioned on their specific education and lifestyle choices made in the initial stage.

Our economic model incorporates complementarities between education and lifestyle investments because of two important reasons. First, an extra year of life is more valuable with higher consumption possibilities. This means that the value of a health-protective lifestyle is larger for the more educated and the value of education is larger for individuals expecting to live more years. Second, as implied by the results of our econometric model, the benefit in terms of better health transitions of investing in protective health behavior is larger for higher-educated individuals. This complementarity may arise due to the higher ability of more educated individuals to react to the results of preventive health care.[1] In addition, these two complementarities between health and

[1]For instance, it is well known that more educated individuals comply better with prescribed therapy (Gold and

education investments shape the selection of individuals into different education and lifestyle categories. Specifically, individuals facing lower costs of health investments self-select into higher education, and individuals facing lower costs of education self-select into protective lifestyles.

We calibrate our model to accurately replicate the joint distribution of education and lifestyle choices for cohorts born in 1930 and 1970, as well as to match the median wealth accumulation over the life cycle by individuals in different education and health behavior types. In the model, cohorts differ in their education-specific average labor earnings over the life cycle. Notably, our model effectively captures the overall wealth distribution and is able to explain 50% of the observed increase in life-expectancy inequalities across education groups between the 1930 and 1970 cohorts.

Our calibrated model allows to identify the mechanisms driving the observed education gradient in lifestyle behaviors. We find that the income advantage of the more educated explains 37% of the observed education gradient in health behavior and 1.3 years of the education gradient in life expectancy. The health advantage of the more educated has the same quantitative effect as the income advantage. Endogenous selection into college education (high school dropouts) of individuals with lower (higher) costs of investing in their health explains 42% of the observed education gradient in health behavior and 1.5 years of the education gradient in life expectancy.

Finally, when comparing individuals born into the 1930 and 1970 cohorts, we find that the direct effect of worse economic prospects for the high-school dropouts in the 1970 cohort generates a widening of the education gradient of health behaviors, which leads to an increase in the education gradient of life expectancy of 5.3 months. The increase in the college premium of labor earnings also induces a large increase in the share of college graduates and a decline in the share of high school dropouts. This large change in education choices across cohorts sets in motion a selection mechanism that explains 7.5 months of the increase in the education gradient of life expectancy, as the high school dropouts in 1970 contain a larger fraction of individuals for which it is more costly to invest in their health. All in all, the increase in income inequality across education groups explains an increase in the education gradient of life expectancy of 1.1 years (55% of the observed increase) through the widening of the gradient in lifestyles. This result connects to the *deaths of despair* literature. Case and Deaton (2017) argue that the worsening of labor market opportunities for white males without a high school degree has led to an increase in risky health behavior (in particular, the use of opioids) for this population group. This has damaged the life expectancy of the less educated and widened the education gradient in life expectancy. Our results broaden the scope of changes in health-related behavior beyond substance abuse and provide a quantitative exercise for this type of argument. However, instead of looking at changes in behavior during the life cycle, we look at the early life determinants of lifestyle choices, which arguably are more important for comparisons across cohorts. Critically, our model finds a stronger impact from selection mechanisms (the composition of the shrinking set of high school dropouts) than from the direct effect of changes in labor market opportunities.

McClung, 2006) or that they are more likely to use newer drugs (Lleras-Muney and Lichtenberg, 2005).

1.1 Related literature

Our estimation of latent types in health behavior relates to very recent literature estimating unobserved fixed effects on health dynamics. De Nardi et al. (2022) were the first to show that health dynamics are not markovian and that unobserved fixed effects are needed to explain long-run health dynamics. Contemporaneous work by Borella et al. (2024) and Hong et al. (2024) estimate unobserved fixed effects by k-means clustering of health trajectories. Different from these three papers, our study classifies individuals into types by use of both health behavior data and observed health trajectories. The use of health behavior data provides an interpretation of the unobserved fixed effects. Our econometric framework allows us to deal with the selection of survivors and, hence, exploit longer trajectories of health data for clustering. Finally, our economic model allows us to treat the unobserved fixed effects as endogenous and study their correlation with socio-economic variables.

The second stage of our model builds on a mature literature that employs life cycle models to quantify how heterogeneous health dynamics impact economic outcomes —see for instance De Nardi et al. (2010), French and Jones (2011), De Nardi et al. (2016), Ameriks et al. (2020), Nakajima and Telyukova (2023), or Bueren (2023)— and welfare —see Capatina (2015), Braun et al. (2019), Hosseini et al. (2021), and De Nardi et al. (2022). Importantly, the latter group of papers estimates significant welfare losses associated with adverse health episodes. The differential welfare losses of bad health across education groups are a key ingredient in our paper, as they shape the different incentives to invest in good health by individuals making different education choices.

Finally, the paper is also related to previous work featuring endogenous health dynamics. Fonseca et al. (2021), Ozkan (2023), and Hong et al. (2024) consider monetary health investments, but they struggle to find a strong causal effect of medical spending on health outcomes. The same can be said of quasi-experimental evidence, like Aron-Dine et al. (2013), Finkelstein et al. (2012), and Baicker et al. (2013). Furthermore, the socioeconomic gradient in health outcomes is also big in the UK or Germany, countries characterized by free universal health care and low out-of-pocket medical spending, see Boháček et al. (2021) or Mahler and Yum (2022). This justifies our focus on health behavior choices as a potential driver of the educational gradient of health outcomes. Health behavior choices have been previously studied by Cole et al. (2019), Mahler and Yum (2022), and Margaris and Wallenius (2023). These models allow agents to adjust their behavior over the life-cycle, taking education choices as given. They also assume the same utility costs of good health behavior across individuals in different education groups. Instead, our model acknowledges the heterogeneity of these costs and capitalizes on the persistence of health behavior, which is often established early in life, to model health behavior and education choices jointly. These features allow us to assess the extent to which selection drives differences in health inequality across education groups. In this sense, our approach is more similar to Hai and Heckman (2022), who study the education gradient of smoking.

The remainder of the paper is organized as follows: Section 3 and Section 4 present the econometric model used to identify lifestyles and its results, respectively. Section 5 outlines the economic model, while Section 6 details the calibration strategy. Section 7 presents the quantitative findings, followed by concluding remarks in Section 8.

2 Data on health behavior

Our data on health behavior collects binary information on whether individuals engage in heavy drinking (having more than 2 alcoholic beverages on the day that the respondent drinks), whether they smoke, whether they have taken a preventive cancer test (males: prostate cancer screening; females: mammography), cholesterol test, and flu shot in the last year, and whether they have an exercise habit (some time participating in sports or other exercise activity during the week before being interviewed). All six variables are available in the HRS, but only the smoking and drinking variables are available in the PSID.

These data show three important patterns. First, there is a clear education gradient on all measures of behavior, whereby the more educated groups contain a higher fraction of individuals who engage in healthier behaviors (see Columns 1 to 3 in Table 1). This gradient suggests that health behaviors may be an important factor behind the education gradient of life expectancy. Second, there is persistence over time in health behaviors, but, except for smoking, this persistence is not high. For instance, among individuals between 50 and 60 years of age, the 4-year autocorrelations of heavy drinking or a flu shot are 0.48 and 0.55, respectively (see Columns 8 and 9 in Table 1). This weak persistence is partly the result of a life-cycle pattern. As Table 1 shows, the incidence of healthier behavior increases with age (except for exercise habits, which decline). For instance, the incidence of smoking falls from 0.22 of the population aged 50 to 60 to 0.09 of the population aged 70 to 80, while the incidence of flu shots rises from 0.46 to 0.74. And third, these different measures of health behavior are correlated across individuals as one would expect, but the correlations are small (see Table 2). For instance, the correlation across individuals between 50 and 60 years of age between cholesterol test with flu shot and cancer test are 0.24 and 0.40, respectively, and only -0.06, -0.13, and 0.04 with drinking, smoking, and exercise habits.[2]

The positive but low correlation of different good health behaviors over time and across individuals suggests the need for a hidden factor model such that individuals of different types have a different propensity to engage in certain behaviors. Ideally, this propensity should change with age (as the incidence of behavior does) and with health (to allow for two-way relations between health and behavior). Finally, linking the types to observed health dynamics would serve to estimate the health consequences of behavior types, and to use this as a criterion to form the types.[3] We present

[2]The imperfect correlation over time and in the cross-section, facts two and three, have already been highlighted for healthy diet and exercise for a sample of nurses in the US, see Bairoliya et al. (2024).

[3]The estimated relationship between types and health dynamics is different from estimating the causal effect of each behavior on health transitions, which is difficult for two reasons. Firstly, a positive but low correlation of

Table 1: Mean health behavior and 4-year auto-correlation

	Mean					AC	
	HSD	HSG	CG	50-60	70-80	50-60	70-80
Drinking	0.07	0.08	0.07	0.13	0.05	0.48	0.42
Smoking	0.18	0.15	0.07	0.22	0.09	0.75	0.64
Cancer test	0.60	0.70	0.79	0.69	0.74	0.40	0.42
Cholesterol	0.74	0.81	0.86	0.75	0.85	0.33	0.29
Flu shot	0.57	0.60	0.66	0.46	0.74	0.55	0.58
Exercise	0.27	0.38	0.54	0.43	0.39	0.40	0.38

Notes: Data from the HRS. HSD: high-school dropout; HSG: high-school graduate; CG: college graduate; 50-60: sub-sample of individuals aged 50 to 60; 70-80: sub-sample of individuals aged 70 to 80. The last two columns show the autocorrelation (AC) of each health behavior with a 4-year lag.

Table 2: Cross correlation health behaviors

	Drinking	**Smoking**	**Cancer test**	**Cholesterol**	**Flu shot**	**Exercise**
Drinking	1.00	0.09	-0.00	-0.01	-0.03	0.02
Smoking	0.16	1.00	-0.09	-0.09	-0.08	-0.07
Cancer test	-0.08	-0.15	1.00	0.30	0.20	0.09
Cholesterol	-0.06	-0.13	0.40	1.00	0.23	0.06
Flu shot	-0.05	-0.07	0.19	0.24	1.00	0.02
Exercise	-0.00	-0.13	0.08	0.04	0.02	1.00

Notes: Data from the HRS. Upper diagonal: individuals aged between 70 and 80. Lower diagonal: individuals aged between 50 and 60.

such a factor model in the next Section.

3 An econometric model of health dynamics with latent types

We combine data from the HRS and the PSID to create an unbalanced panel of individuals $i = 1, \ldots, N$ followed for $t = 0, \ldots, T_i$ periods. For each individual and period we observe standard demographic variables: cohort of birth $c_i \in \{c_{10}, c_{30}, c_{50}, c_{70}, c_{90}\}$ (individuals born between 1900 and 1920, 1920 and 1940, *etcetera*), gender $s_i \in \{s_m, s_f\}$ (male, female), education $e_i \in \{\text{HSD}, \text{HSG}, \text{CG}\}$ (high school dropout, high school degree, college degree), and age $a_{it} \in \{25, 26, \ldots, 100\}$, plus a wide array of health-related variables, which we classify into two groups: health outcomes and health behaviour. The health outcome $h_{it} \in H \equiv \{h_g, h_b, h_d\}$ takes three values: good health (h_g), bad health (h_b), or dead (h_d), which is an absorbing state. We build this variable by use of the 5-category self-rated health variable (where the best two categories form the good health state) and the information on survival. The health behavior vector $\mathbf{z}_{it} = \{z_{1,it}, z_{2,it}, \ldots, z_{N_z,it}\}$ contains

different health-enhancing behaviors over time and across individuals complicates the isolation of singular behavior effects. Secondly, the impact of individual health behaviors on health outcomes may not manifest in observable health changes within a short period of time but may appear many years ahead. Our methodology shifts the focus from isolated behaviors to the holistic lifestyle patterns that shape long-run health trajectories.

information on N_z different categorical variables $z_{m,it} \in \{0,1\}$ describing whether individual i in period t does some particular action. These actions are whether the individual has taken a cancer test (prostrate or mammography), a cholesterol test, or a flu shot in the last year, whether the individual drinks (more than two drinks on the day she drinks), whether the individual smokes, and whether the individual is performs any type of physical activity.[4]

We assume that both the observed health behaviour $\mathbf{z}_{it}$ and health outcomes h_{it} depend on some unobserved time-invariant latent variable $y_i \in Y \equiv \{y_1, y_2, \ldots, y_{N_y}\}$, with $N_y < N_z$. We interpret the latent variable as lifestyle / health habit type —determined before the start of working life— that captures the idea that individuals differ in their propensity to undertake actions that are good for their health. This notion of an unobserved latent variable is important because in the data observed health behaviour is imperfectly correlated across individuals and over time. That it, we want the latent types to also affect health outcomes because we want the classification of individuals based on behavior to be meaningful in terms of health dynamics.

We aim to estimate the parameters of our econometric model by maximizing the probability of observing the joint sequence of health behaviours $\mathbf{z}_i^T$ and health outcomes $\mathbf{h}_i^T$ of each individual i conditional on the demographic variables and initial health. We can write the likelihood of the data as a mixture model:

$$p\left(\mathbf{z}_i^T, \mathbf{h}_i^T | \mathbf{a}_i^T, c_i, s_i, e_i, h_{i0}\right) = \sum_{y \in Y} p\left(\mathbf{z}_i^T | y, h_i^T, \mathbf{a}_i^T, c_i, s_i, e_i\right) p\left(\mathbf{h}_i^T | y, h_{i0}, \mathbf{a}_i^T, s_i, e_i\right) p\left(y | h_{i0}, a_{i0}, c_i, s_i, e_i\right) \quad (1)$$

where the elements in the right hand side, the probability of observing a sequence of health behaviours, the probability of observing a sequence of health outcomes, and the initial distribution of types, are explained in Section 3.1, 3.2, and 3.3 respectively.

3.1 Health behaviour

We assume that the probability of individual i in period t reporting the m^{th} behavior $(h_{m,it} = 1)$ depends only the health behaviour type y_i but also on age a_{it}, on health status h_{it}, and on an idiosyncratic shock $\varepsilon_{m,it}$. The idea is that the association of observed behavior such as smoking or cancer tests with types may differ over age and across health states. Instead, conditional on these variables, we impose that health behavior does not depend on cohort c_i, gender s_i, or education e_i. We do so because we want the definition of types to be stable across demographic groups. This makes the types comparable across groups, and it lets the variation in behavior across demographic groups to arise from their different distribution of types.

We assume that, conditional on y_i, a_{it}, and h_{it}, the shock $\varepsilon_{m,it}$ is iid across m, i, and t, and

[4] Some of the health outcome and health behavior variables for a given individual may be missing for some period t. Indeed, in the PSID we do not observe the variables for health protective behaviour (cancer test, cholesterol test, flu shot). We take missing observations into account under the assumption that they occur completely at random, but we abstract from them in the model description to simplify the exposition.

that it follows a standard normal distribution. Then, we can model the probability of individual i in period t reporting the m^{th} behavior ($h_{m,it} = 1$) as a probit model. In particular, let $z^*_{m,it} = z^*_m (y_i, a_{it}, h_{it})$ be a latent variable such that $z^*_{m,it} + \varepsilon_{m,it} > 0 \Rightarrow z_{m,it} = 1$.[5] Then, the probability of observing the m^{th} behavior is given by $\Phi\left(z^*_{m,it}\right)$ where $\Phi()$ is the cdf of the standard normal distribution, and the probability of observing a sequence $\mathbf{z}_i^T$ of health behaviour vectors $\mathbf{z}_{it}$ for individual i is given by,

$$p\left(\mathbf{z}_i^T | y_i, h_i^T, a_i^T\right) = \Pi_{t=1}^T \Pi_{m=1}^M \left[\Phi\left(z^*_{m,it}\right)\right]^{z_{m,it}} \left[1 - \Phi\left(z^*_{m,it}\right)\right]^{1-z_{m,it}} \tag{2}$$

3.2 Health dynamics

We assume that health dynamics for individual i depends on gender s_i, education e_i, health behaviour type y_i, and age a_{it}. The dependence of health dynamics on gender and education is meant to capture differences in health outcomes associated to these variables that are not captured by differences in health behaviour types across these demographic groups. The absence of cohort c_i from the set of conditioning variables is an identification assumption. We do not observe full lifespans for individuals in different cohorts. Rather, we combine information on health dynamics at old ages from individuals born in earlier cohorts with information on health dynamics at young ages from individuals born in later cohorts. This is standard in estimation of models of health dynamics with survey data, see for instance Pijoan-Mas and Ríos-Rull (2014). Hence, our assumption is that health dynamics of individuals of given gender and education are, conditional on type, identical across cohorts. However, it is important to note that this allows gender and education health dynamics to differ across cohorts due to the different composition of types.

We assume that conditional on gender s_i, education e_i, health behaviour type y_i, and age a_{it}, the evolution of health outcomes is markovian, that is, only depends on one lag of health outcomes. We model survival and health transition probabilities as a multinomial probit model.

3.3 Distribution of health types

The final element we need is the fraction of individuals of each type, that is, the prior probability of each individual of being of a given health behaviour type before observing the data on health behavior and health outcomes. In particular, we define $p\left(y_i | c_i, s_i, e_i, a_{i0}, h_{i0}\right)$ as the probability that individual i born into cohort c_i, of gender s_i, education e_i, first observed with age a_{i0} and health h_{i0} is of type y_i. One could add this term to the likelihood function and estimate it non-parametrically from the data. However, because the panel is not balanced and the first age of observation of many individuals is quite advanced (a_{i0} is large), there would an identification problem in that observed changes of health behavior $\mathbf{z}_{it}$ with age (for instance, the decline with age of the incidence of smoking) could not be separated into (a) changes over age of health behavior conditional on

[5] We model $z^*_m (y_i, a_{it}, h_{it})$ as a flexible low order polynomial on age.

type $p\left(\mathbf{z}_i^T|y_i,h_i^T,a_i^T\right)$ and (b) changes in the distribution of types with age $p\left(y_i|c_i,s_i,e_i,a_{i0},h_{i0}\right)$. Therefore, we exploit the model of health dynamics described above to obtain an expression for how the posterior distribution of types conditional on observables evolves with age. This leaves us with the need to only estimate the prior probabilities of each individual of being of each type at the initial age of 25.

In order to do so, note that we can write,

$$p\left(y_i|c_i,s_i,e_i,a_{it},h_{it}\right)=\frac{p\left(y_i,h_{it}|c_i,s_i,e_i,a_{it}\right)}{\sum_{h\in H}p\left(y_i,h|c_i,s_i,e_i,a_{it}\right)} \tag{3}$$

The joint probability $p\left(y_i,h_{it}|c_i,s_i,e_i,a_{it}\right)$ can be decomposed as,

$$p\left(y_i,h_{it}|c_i,s_i,e_i,a_{it}\right) = \sum_{h_{it-1}\in H}p\left(h_{it}|y_i,h_{it-1},s_i,e_i,a_{it-1}\right)p\left(y,h_{it-1}|c_i,s_i,e_i,a_{it-1}\right) \tag{4}$$

The first term in the right hand side describes the health dynamics and it has been discussed in Section 3.2. The second term in the right hand side is the same as the left hand side, just one period before. Hence, we can use equation (4) recursively up to age 25, which is our initial age, that is, up to $p\left(y_i,h_{it}|c_i,s_i,e_i,a_{it}=25\right)$. This term describes the joint probability of an individual in cohort c_i, of gender s_i and education e_i of being of health h_{it} and type y_i at age 25. We decompose this probability in two pieces:

$$p\left(y_i,h_{it}|c_i,s_i,e_i,a_{it}=25\right)=p(y_i|c_i,s_i,e_i,a_{it}=25,h_{it})p(h_{it}|c_i,s_i,e_i,a_{it}=25)$$

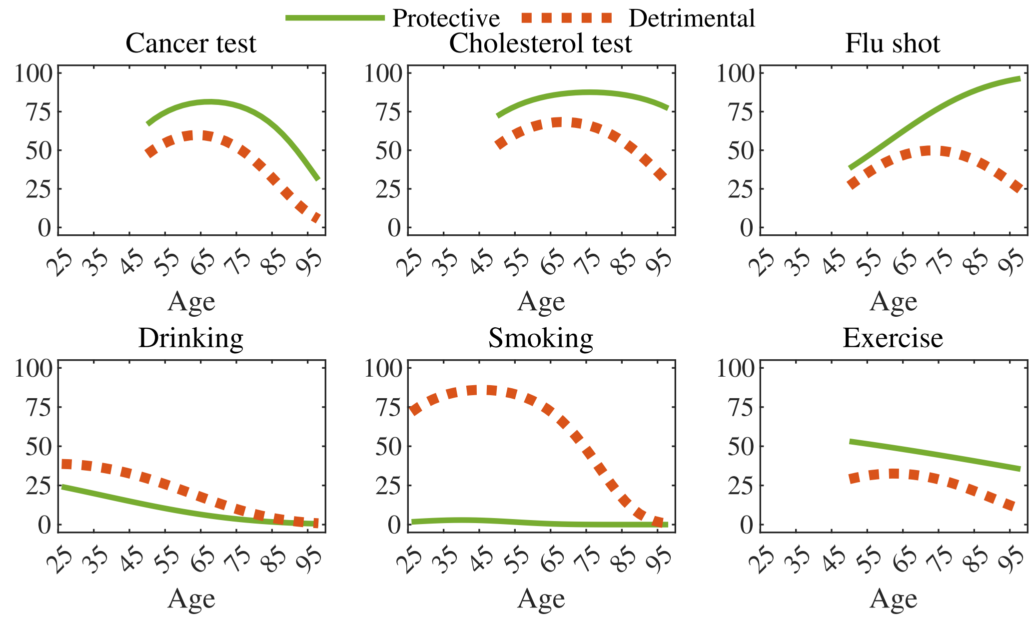
The second term in the right hand side describes the share of individuals of age 25 of given cohort c_i, gender s_i, and education e_i that have health h_{it}. This can be measured directly in the PSID for many but not all cohorts. We thus assume that conditional on education and gender, the probability of good and bad health at age 25 does not change across cohorts and we set it equal to the one that we observe in the PSID.[6] The first term in the right hand side describes the fraction of individuals of age 25, cohort c_i, gender s_i, education e_i, and health h_{it} that are of type y_i. We model this probability through a multinomial probit. That is, we define thresholds $y_{1,i}^*=y_1^*\left(c_i,s_i,e_i,h\right)$ and $y_{2,i}^*=y_2^*\left(c_i,s_i,e_i,h\right)$ that, given the realization of a shock $\varepsilon_{y,i}\sim N\left[0,1\right]$, separates individuals into types.

4 Results from the econometric model

We estimate the model using data from the PSID from 1999 to 2019 and the HRS between 1996 and 2018. If individual lifestyle types y were observed, the estimation of the model parameters by maximum likelihood would be straightforward. However, because we also have to classify indi-

[6] The probability of being in good health at age 25 varies between 77% for dropout females to 98% for male college graduates.

Figure 1: Health habits and health behavior types



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health.

viduals into types the estimation is challenging (in addition to the model parameters, we need to estimate thousands of latent variables, one for each individual). Therefore, we resort to Markov Chain Monte Carlo methods, which reduce the full likelihood into smaller simpler blocks, see Appendix A for details. In our main estimation, and in favor of parsimony, we choose $N_y = 2$, that is, we classify individuals into two latent groups. In what follows we describe the results.

4.1 Health behaviour

We start by showing how the estimated types are related to observed health behaviors. Figure 1 reports the probability of displaying each health behavior $z_{m,it}$ as a function of health type y_i, age a_{it}, and $h_{it} = h_g$ (the case $h_{it} = h_b$ is not too different). Individuals in one group, which we label *protective* (solid green line), have a higher likelihood of reporting health protective habits (cancer test, cholesterol test, flu shot, and exercise) and a lower probability of reporting health detrimental habits (smoking and drinking). For individuals in the other group, which we label *detrimental* (red dashed line), the probability of reporting all the protective habits is lower and their smoking probability is very high. We note that, within each type, the probability of displaying each health behavior changes with age and it does so differently across types. For instance, the probability of taking a flu shot is similar across the two groups at age 50. However, as they age, individuals classified as *protective* increase their probability of getting a flu shot while individuals classified as *detrimental* do not.

4.2 Health dynamics

Next, we examine the relationship between health behavior types and health dynamics. A good way of summarizing this information is by looking at the gradient in age-50 life expectancy across health behaviour types.[7] In Table 3 (second column) we show that, on average, *protective* types live for 8.6 more years than *detrimental* types. The effect of lifestyle on health dynamics is also large within each education category, and more so within the more educated: *protective* types live 6.6 more years than *detrimental* types among males without a high school degree, 6.8 more years among males with a high school degree, and 9.6 more years among males with a college degree (see columns 6, 6, and 8).

4.3 Distribution of health types

Our estimation assigns a different fraction of individuals to each type y depending on cohort c, gender s, and education e. For the 1970 cohort, 74.4% of individuals are classified as *protective* and 25.6% as *detrimental* (see first column in Table 3). The main finding here is that there is a large educational gradient of health behaviour types: the share of *protective* individuals grows from 44.3% to 93.6% as we move from high school dropout to college graduate (see columns 3, 5, and 7). These figures are large and reflect a strong correlation between education and lifestyle. In a sense, these numbers are not too surprising: it is well known that the incidence of smoking, drinking, and obesity declines with education, and Table 1 before shows how in our data the share of individuals taking cancer tests, cholesterol tests, and flu shots increases with education. It is therefore natural that our classification of individuals into types according to the observed health-related behaviour retains the education gradient of these latter variables. However, we highlight that our classification of individuals into types also uses longitudinal information on health dynamics *conditional on education*. That is, it also uses the fact that within education, *protective* types have better health dynamics than *detrimental* types.

A very interesting question is how has the distribution of health types evolved across education groups over time. In Panels (a) to (c) of Figure 2 we report the distribution of types by education group from the 1910 to 1990 cohorts. Within the high school dropouts, the *detrimental* types increased monotonically from 40% in the 1930 cohort to 75% in the 1990 cohort. This implies a severe deterioration in the lifestyle of individuals in the least educated group, which reverses a slight improvement in the type distribution between the 1910 and the 1930 cohorts. In contrast, among college educated individuals there is a smaller change, with a slight increase in the share of *protective* and a slight decline of the *detrimental*. All in all, this implies that the educational gradient in lifestyles has widened remarkably between the 1930 and the 1990 cohort. As we will see in the next Section, this will generate an increasing life expectancy gap across education groups.

[7] Figure 11 in the Appendix shows the actual health transitions and survival functions by health behaviour type

Table 3: LE at age 50 across education and lifestyles: males born in 1970s

	All		HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	29.3	100.0	24.9	100	28.0	100.0	32.8	7.9	4.7	3.2
PRO	74.4	31.5	44.3	28.5	69.0	30.1	93.6	33.4	4.9		
DET	25.6	22.9	55.7	22.0	31.0	23.3	6.4	23.8	1.9		
Δ_y	48.8	8.6	-11.4	6.6	37.9	6.8	87.2	9.6	3.0		

Notes: This table reports the share of male individuals of each lifestyle and the life expectancy (LE) at age 50 for different population groups. Column (a) corresponds to the counterfactual LE gradient when the distribution of behavior types among high school dropouts is the same as for college-educated individuals. Column (b) is the difference between the actual LE gradient and the counterfactual in column (a), that is, it corresponds to the gradient explained by difference in lifestyles across education groups for given health dynamics.

4.4 Decompositions

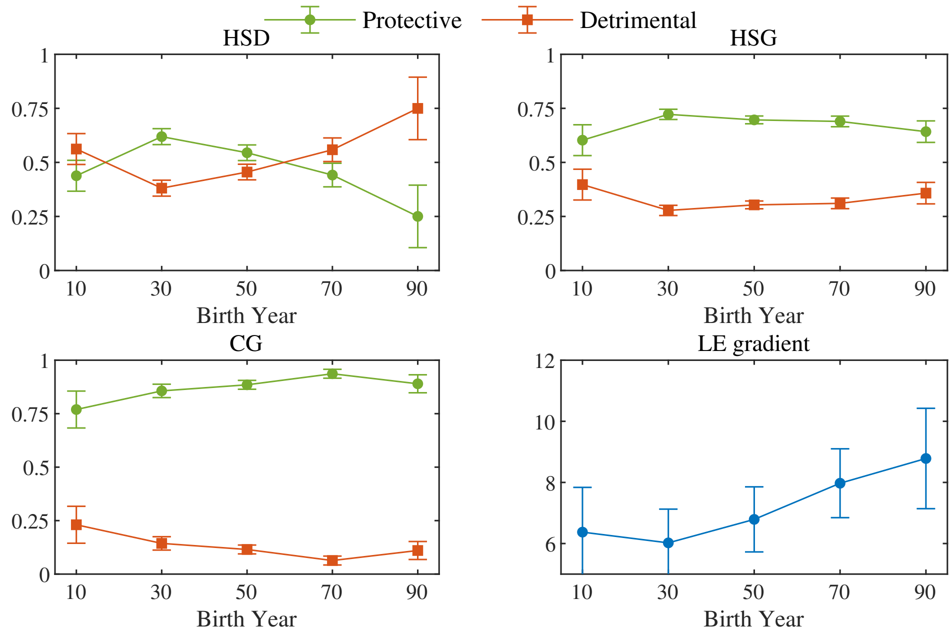
Combining all previous results, we can show that the different type composition across education groups explains an important fraction of the educational gradient of life expectancy, which it is 7.8 years for the 1970 cohort. In particular, if males without a high school degree had a distribution of health behaviour types as the males with a college degree, their life expectancy would rise in 4.7 years. Hence, the different health dynamics across education groups for fixed distribution of lifestyle behavior explains 4.7 out of 7.8 years of the life expectancy gradient. The difference, 3.2 years or 40% of the gradient, is explained by the different distribution of health behaviour types across education groups.

Finally, in Panel (d) of Figure 2 we report the age-50 predicted education gradient in life expectancy for different cohorts. In our estimation, different cohort have different life expectancy because of different composition of latent types as described in Panels (a) to (c) of Figure 2, but health dynamics conditional on type are identical across cohorts. This allows us to infer health dynamics at old ages of younger cohorts, for which most individuals are still alive today. Our finding show a growing education gradient in life expectancy: from 6.8 years in the 1930 cohort to 9.1 years in the 1990 cohort, which follows a 0.7 year decline in the gradient between the 1910 and the 1930 cohorts.

4.5 Health inequality and economic inequality

The final empirical question is whether the health behavior types correlate with some economic outcomes in adult life. In this Section we show that wealth accumulation is positively linked to health behavior types. To do so, we need to recover the wealth distribution across the unobserved health behavior types. For this purpose, we model the observed wealth distribution as a mixture model, see Appendix C for details. We present selected moments of the estimated wealth distribution conditional on age, education and type in scatter plots in Figure 3. As it is well known, wealth

Figure 2: Distribution of types by education and cohort (males)



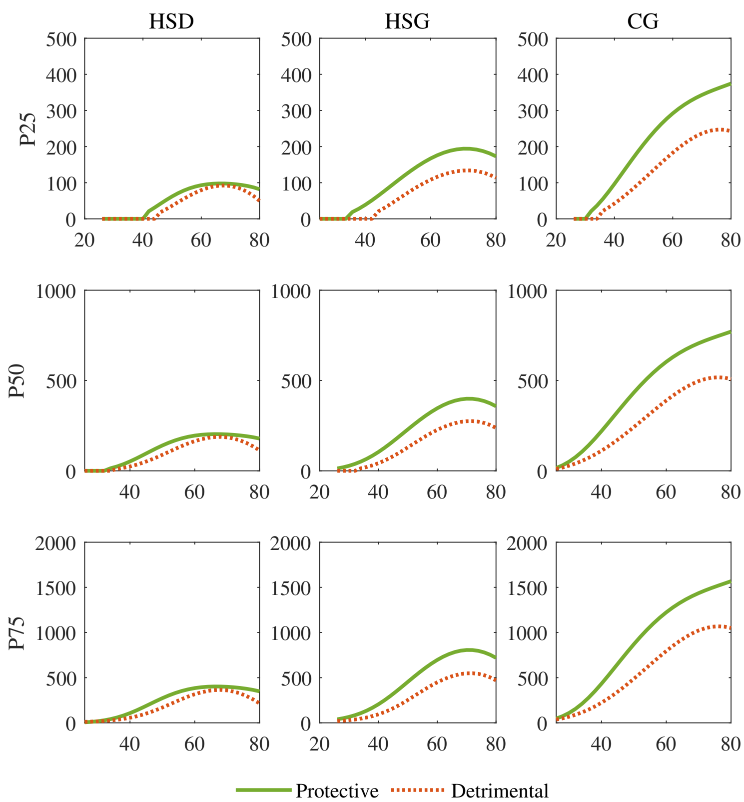
Notes: The first three panels display the fraction of individuals of each behavior type in each different cohort for a given education group. The bottom-right panel reports the estimated life expectancy difference between college graduates and high school dropouts in each cohort.

accumulation is positively correlated with education. What is interesting of our results is that, within each education category, wealth accumulation is also stronger for better health types. We would like to highlight two results. First, wealth accumulation is stronger for the *protective* type. Second, the difference in wealth accumulation across types is especially apparent within college educated individuals, and much smaller (or null) within the other two education categories.

5 An economic model

Our economic model considers two distinct life stages. The *early life* stage is a static problem where young individuals belonging to a cohort c choose their education $e \in \{\text{HSD}, \text{HSG}, \text{CG}\}$ and their lifestyle or health-related behaviour $y \in \{\text{DET}, \text{PRO}\}$. This problem is a stand-in for choices and investments made by either parents in early childhood or young adults before entering the labor market. The objective is to maximize the expected value of starting working life with a given type (education and lifestyle) minus the (idiosyncratic) costs of choosing each type. This stage serves to account for the observed correlation between education and health types. The *adult life* stage is a dynamic life-cycle consumption-saving problem under uncertainty in both labor market and health outcomes where individuals differ in type (education and lifestyle). This stage serves to link inequality in health and economic outcomes, and provides the value of starting life in each type, which is used in the *early life* stage.

Figure 3: Wealth distribution across lifestyles



5.1 Stage 1: early life

Let $V_0^{\mathrm{c,e,y}}$ be the value of starting working life with a type (e, y) for individuals in cohort c. Before entering the labor market, young individuals choose their type by solving

$$\max_{\mathrm{e,y}} \left\{ V_0^{\mathrm{c,e,y}} - \tau_e - \tau_y \right\}$$

where τ_e and τ_y represent the cost of choosing an education e and a lifestyle y, respectively. These education and health behavior costs are heterogeneous in the population.[8] This formulation imposes that education and healthy habit choices are taken together at a young age and never change. We want to comment on a few things about this assumption. First, there is empirical evidence on the early-age adoption of health-related habits.[9] Second, these habits tend to be very persistent, as exemplified by the small effects of interventions to change them during adulthood.[10] Third, we

[8] The heterogeneous costs of education τ_e may arise due to many different factors, such as differences in family background (Hauser and Featherman, 1976), distance to quality education centers (Card, 1995), or taste (Willis and Rosen, 1979). However, because τ_e does not directly affect earnings, we should not think of it as labor market ability in the manner of Keane and Wolpin (1997). Less is known about the heterogeneous costs of healthy habits choices. A recent literature relates brain characteristics of young adolescents with later-in-life health risk behavior (Xiang et al., 2023), which highlights the importance of individual-level variation in τ_y.

[9] For instance, see Farrell and Fuchs (1982) or Hai and Heckman (2022) for the early adoption of smoking and how it correlates with later decisions of college education.

[10] See Conner and Norman (2017) and references therein.

note that, according to our econometric specification in Section 3, a permanent choice of a latent lifestyle type $y \in \{\text{DET}, \text{PRO}\}$ still allows for changes in observed behavior (like smoking, exercise, or preventive tests) over the life-cycle and across health states, see for instance Figure 1. And fourth, while some individuals might change their lifestyles over the life cycle, to understand differences in lifestyles across cohorts focusing on early life choices seems the most relevant margin.

Going into details, we normalize to 0 the cost choosing high-school dropout education and detrimental lifestyle ($\tau_{\text{HSD}} = 0$ and $\tau_{\text{DET}} = 0$). The cost of protective behavior is assumed to be normally distributed: $\tau_{\text{PRO}} \sim N(\mu_{\text{PRO}}, \sigma_{\text{PRO}})$. The cost of graduating from high-school and the cost of college education are assumed to be $\tau_{\text{HSG}} = \mu_{\text{HSG}} + \sigma_{\text{HSG}}\epsilon_e$ and $\tau_{\text{CG}} = \mu_{\text{CG}} + \sigma_{\text{CG}}\epsilon_e$, respectively with $\epsilon \sim N(0,1)$.

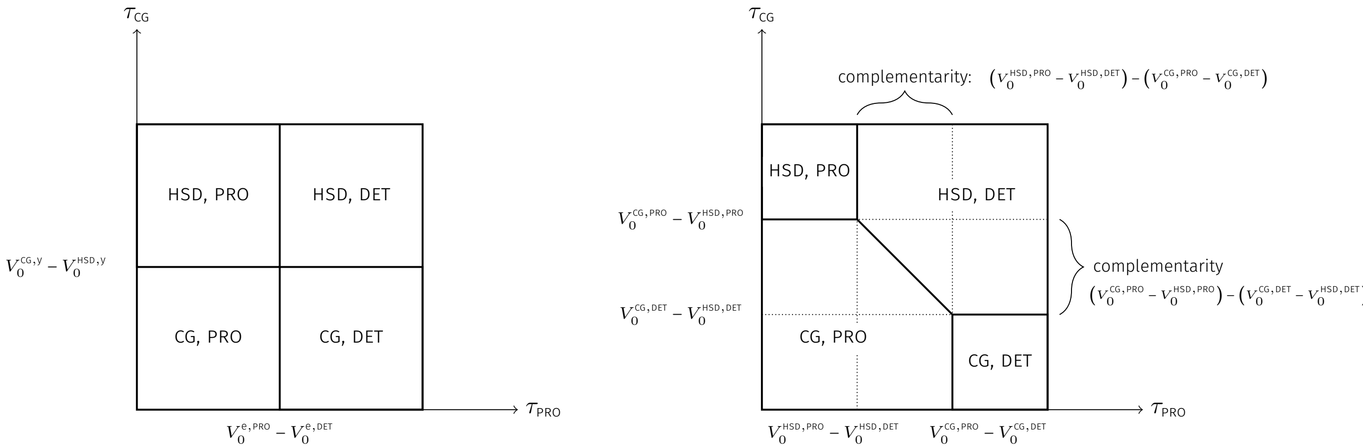
In this framework, the choices of education and lifestyle are independent from each other if and only if $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} = V_0^{\text{HSG,PRO}} - V_0^{\text{HSG,DET}} = V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$, that is, if the value of pursuing a protective lifestyle is the same no matter what is the chosen education level (or, likewise, the value of choosing higher education is the same no matter what is the lifestyle choice). Such a situation is illustrated in Panel (A) of Figure 4 for the particular case of only two education categories (we drop out $e = \text{HSG}$). In this situation, an individual chooses a protective lifestyle if and only if $V_0^{\text{e,PRO}} - V_0^{\text{e,DET}} > \tau_{\text{PRO}}$. Therefore, individuals with τ_{PRO} to the left of $V_0^{\text{e,PRO}} - V_0^{\text{e,DET}}$ choose a protective lifestyle while individuals with τ_{PRO} to the right choose a detrimental lifestyle. Likewise, an individual chooses a college education if and only if $V_0^{\text{CG,y}} - V_0^{\text{HSD,y}} > \tau_{\text{CG}}$. Therefore, indivuals with τ_{CG} below $V_0^{\text{CG,y}} - V_0^{\text{HSD,y}}$ choose a college education while individuals τ_{CG} above choose a high-school dropout education.

In contrast, whenever $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$ there are complementarities between education and lifestyle choices. We plot this situation in Panel (B) of Figure 4. In this case, the value of pursuing each lifestyle depends on the chosen education (and the value of pursuing a particular education depends on the chosen lifestyle). In particular, individuals with an intermediate cost of protective behavior (τ_{PRO}) choose a protective lifestyle and a college education when the cost of studying (τ_{CG}) is low. However, when the cost of studying is high, they choose a lower level of education ($e = \text{HSD}$) but also they switch to detrimental behavior, as the value of a protective lifestyle is lower when not going to college. Panel (B) of Figure also illustrates the pattern of selection (in terms of τ_{PRO} and τ_{CG}) generated by the complementarity of education and lifestyle investments. No matter what is the distribution of the population over τ_{PRO} and τ_{CG}, the average value of τ_{PRO} will be lower for college graduates than for high-school dropouts and the average value of τ_{CG} will be lower for protective types than for detrimental types.

Next, we move to discuss the modeling of the adult life, which generates the value functions $V_0^{\text{c,e,y}}$ that are used in this early life stage.

5.2 Stage 2: adult life

Figure 4: Early life choices of education and lifestyle



(A) Independent choices (B) Complementarity

Notes. Eary life choices of education and lifestyle for the case with only two education categories ($\tau_{\text{HSG}} \to \infty$). Panel (A) displays the choices for the case in which $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} = V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$, while Panel (B) displays the choices for the case $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$.

5.2.1 Demographics, preferences, and shocks

The model period corresponds to two years. Individuals live for at most T periods but survival is stochastic every period. During the first $R-1$ periods of life people are exposed to health shocks, medical expenditure shocks, and labor income shocks. Individuals retire at age R, when they start receiving a retirement pension instead of stochastic labor income.

Preferences over consumption flows c_t are described by a standard CRRA period utility:

$$u(c_t) = \frac{(c_t/\bar{n}_t)^{1-\sigma} - 1}{1-\sigma} + \bar{b},$$

where σ is the risk-aversion, $\bar{n}_t$ is an age specific household size and $\bar{b}$ is a positive term to ensure that individuals in our model value their life. In the period when they die, individuals also derive utility from leaving a bequest of size k_t:

$$v(k_t) = b_0 \frac{(k_t + b_1)^{1-\sigma} - 1}{1-\sigma}$$

where b_0 drives the strength of the bequest motive and b_1 the extent to which preferences for bequest increase with wealth.

Following the empirical model in Section 3, health h_t can be either good (h_g) or bad (h_b) and it evolves according to the age-dependent Markov chain $\Gamma_t^{\text{e,y}}(h)$, which depends on education e and health type behavior y. The survival probability $s_t^{\text{e,y}}(h)$ depends on health h, but also (possibly) on education e and health-behavior type y.

Every period of their working life, individuals receive an exogenous income which we model in two components. First, there is an employment shock $\ell_t \in \{0,1\}$ that determines if the individual has the chance of working in the labor market ($\ell_t = 1$) or not ($\ell_t = 0$). We model $\text{Prob}(\ell_t = 1|e,t,h_t,c)$ as an i.i.d shock conditional con age, health, education, and cohort. This component aims to capture the higher probability of individuals in bad health of being non-employed, which is at the core of the health gradient of labor income, and also differences in employment across education and cohorts. Second, conditional on working, individuals receive labor income given by,

$$\log w_t^{\mathrm{c,e}}(h_t,\xi_t,\epsilon_t) = \omega_t^{\mathrm{c,e}}(h_t) + \xi_t + \epsilon_t, \tag{5}$$

where $\omega_t^{\mathrm{c,e}}(h)$ is a deterministic component depending on cohort, age, education, and health, while ξ_t and ϵ_t are persistent and transitory shocks. The initial value of the persistent component ξ_0 is drawn from a normal distribution with mean zero and variance $\sigma^2_{\xi_0}$. The stochastic persistent component is assumed to follow a Gaussian AR(1) process with persistence ρ_ξ and variance of the innovations σ^2_ξ. The transitory component ϵ_t is Gaussian white noise with mean 0 and variance σ^2_ϵ.

Finally, medical expenses are given by,

$$\log m_t^{\mathrm{e}}(h_t,\zeta_t) = \mu_t^{\mathrm{e}}(h_t) + \zeta_t$$

where $\mu_t^{\mathrm{e}}(h_t)$ is a deterministic component, as a function of age t, education e and health h_t, while ζ_t is an i.i.d gaussian white noise process with mean zero and variance $\sigma^{\mathrm{e}}_{\zeta,t}(h)$.

5.2.2 Taxation and social transfers

We model the tax system as follows. Working households pay payroll taxes, which include the Medicare tax (τ_{MCR}) and the Social Security tax (τ_{ss}). The latter only affects earnings below w_{ss}. There is a progressive labor income tax $T(w)$ which we specify as Heathcote et al. (2020)

$$T(w) = w - a_{\tau 0} w^{1-a_{\tau 1}}$$

We represent several existing mean-tested programs in a stylized way through a public safety-net program. This program guarantees every household a minimum income floor $\underline{x}$. Retirees receive Social Security benefits. In practice, these payments depend on an individual's history of earnings. To capture the existing variation in pension benefits without increasing computational costs, we approximate the benefits using the following approach. First, we divide individuals into groups based on their labor force participation just before retirement, their last draw of the persistent productivity shock and on their education and health behavior type. Then, for each group, we compute average earnings over the 17 model periods (34 years) with the highest earnings. Then we apply the Social Security benefits formula to these average earnings.

5.2.3 The optimization problem

At the beginning of the period, working-age individuals of type (c,e,y) and age t learn their cash in hand x_t, labor force status ℓ_t, persistent component of productivity ξ_t (conditional on participating in the labor force), and health state h_t. All these variables form the state of the individual: x_t is payoff relevant in the current period and the other variables serve to predict next period outcomes. Based on this information, individuals decide on consumption c_t and savings k_{t+1}. At the end of the period, there are new realizations of the shocks for survival, health, labor force participation, productivity (persistent and transitory), and medical expenses. The timing for retired individuals is similar, with the difference that there are no employment or labor earnings shocks.

The optimization problem for working age individuals ($t < R$) can be written in recursive form as:

$$V_t^{\mathrm{c,e,y}}(x,h,\xi) = \max_{c,k'} \left\{ u(c) + \beta s_t^{\mathrm{e,y}}(h) \sum_{h'} \Gamma_t^{\mathrm{e,y}}(h)\, \mathbb{E}_{\ell,\xi,\zeta,\epsilon}[V_{t+1}^{\mathrm{c,e,y}}(x',h',\xi')] + \beta^{T-t}\left(1 - s_t^{\mathrm{e,y}}(h)\right) v(k') \right\}$$

$$\begin{aligned}
&\text{s.t.} \\
&c + k' = x \\
&x' = \min\left\{ w_{t+1}^{\mathrm{c,e,y}}(h',\xi',\epsilon')\ell' - Tax + (1+r)k' - m_{t+1}^{\mathrm{e}}(h',\zeta'),\, \underline{x} \right\} \\
&Tax = T\left(w_{t+1}^{\mathrm{c,e,y}}(h',\xi',\epsilon')\ell'\right) + \tau_{MCR} w_{t+1}^{\mathrm{c,e,y}}(h',\xi',\epsilon')\ell' + \tau_{ss}\min\{w_{t+1}^{\mathrm{e,y}}(h',\xi',\epsilon')\ell', w_{ss}\}
\end{aligned}$$

The optimization problem for retired individuals ($t \geq R$) is analogous, with ξ_{R-1} instead of ξ in the state space, no social security taxes, and deterministic pension $p^{\mathrm{c,e,y}}(\xi_{R-1})$ instead of labor earnings.

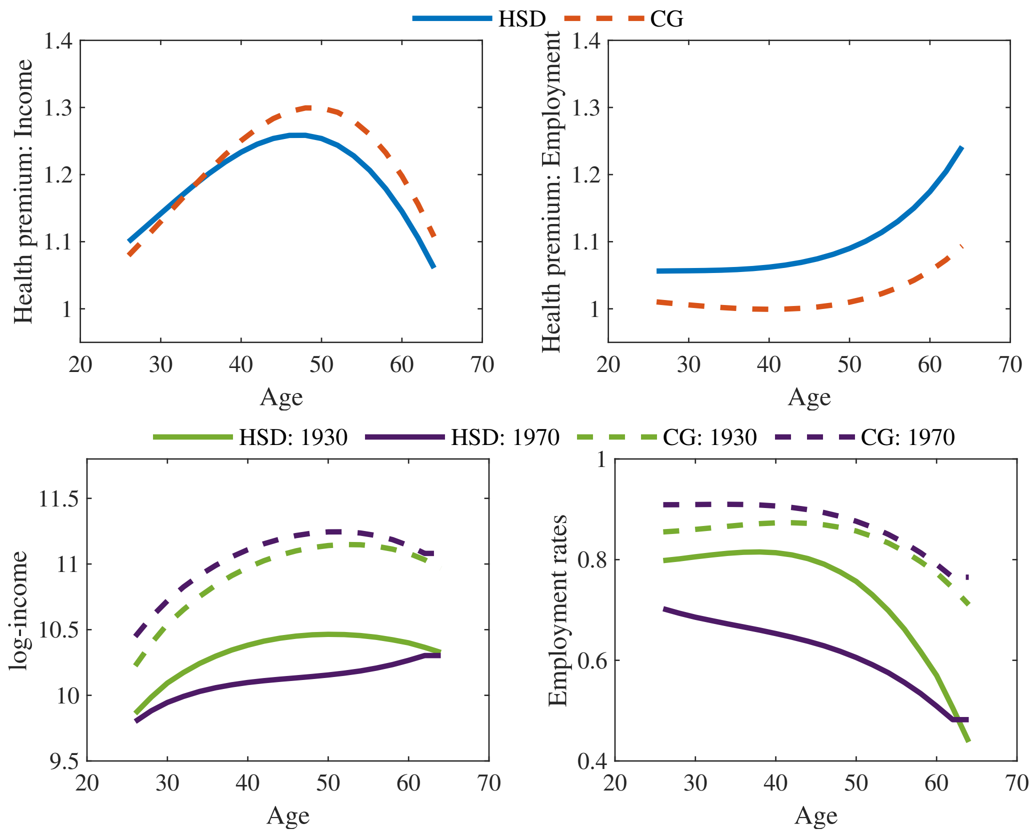
6 Calibration

The calibration of our model has three distinct parts: the parameters related to the life cycle model of adults (Section 6.1), the parameter $\bar{b}$ driving the value of life, which does not affect the outcomes in the life cycle model (Section 6.2), and the parameters shaping unobserved heterogeneity in the early life stage (Section 6.3).

6.1 Stage 2: adult life

The survival and health processes are taken from Section 3.2. We fix the interest rate at 4%, the risk aversion parameter σ at 1, and the discount factor β at 0.98. The calibration of taxes and social security parameters is standard and explained in Appendix D.

Figure 5: Deterministic Income Component and Employment Rates



Notes: The top two panels report the ratios of the income and employment rate between good and bad health for two education groups (CG and HSD). This is obtained from PSID. The bottom two panels report (log) income and employment rates for two education groups (CG and HSD) in two different cohorts (1930 and 1970). This is obtained from Census.

6.1.1 Income process

We use decennial census data spanning from 1940 to 2020 and PSID data covering the years 1999 to 2019 to calibrate the deterministic component of income process and labor force participation. In this calibration, we consider that both the deterministic component of income and the likelihood of being employed are function of a cubic in age and health, which are further interacted with education and cohorts.

We define non-employment as having an annual wage below 3,770 USD per year. We regress log-income and employment status against a cubic age, cohort, and education using Census data. Since health status is not directly observable in Census data, we resort to imputing it using PSID data. To accomplish this, we regress log-income and employment status against a cubic age and health, interacting with education, using the PSID data. Then, we assume that the health effect observed in the PSID coincides with that in the census data and remains constant across different cohorts.

The upper panels of Figure 5 report the ratio of income (left panel) and employment rates

Table 4: Parameters of the stochastic component of income

	$\sigma^2_{\xi_0}$	ρ_ξ	$\sigma^2_\xi \times 10^{-2}$	σ^2_ϵ
HSD	0.18	0.96	2.37	0.16
HSG	0.16	0.98	2.33	0.14
CG	0.18	0.99	3.76	0.13

Notes: Estimated parameters for the stochastic process of wages.

(right panel) of individuals in good health versus bad health for college and high-school dropouts. The figure shows that, when employed, individuals in good health earn on average 20% higher wages than individuals in bad health and that there is little variation across education groups. The figure also shows that, for college graduates, the employment rates are not very different across health states until age 50 and that a health premium of around 5% develops after that. Instead, for high school dropouts the health premium of employment is larger, reaching up to 20% at age 60. Overall, the labor market losses of bad health are larger for the less educated.

The bottom-left panel of Figure 5 shows the average wage age profile across education for individuals born in 1930 and 1970. The figure reveals large increases in the college premium driven by both an increase in college wages and decreases in the wages of high school dropouts. Moreover, the bottom-right panel of the figure also shows a larger dispersion in employment rates across education groups for more recent cohorts.

We estimate the stochastic component of income by using the model described in equation (5) with PSID data. For this purpose, we estimate the model using Bayesian methods. Table 4 provides the estimated parameter values.

6.1.2 Matching wealth trajectories

Given the parameters discussed above, we estimate the remaining model parameters using the simulated method of moments. We do so by minimizing the sum of squared differences between median assets by education, age and lifestyle for individuals born in 1930s in the data and in the model. The set of parameters to be estimated is $\{\underline{x}, b_0, b_1\}$.

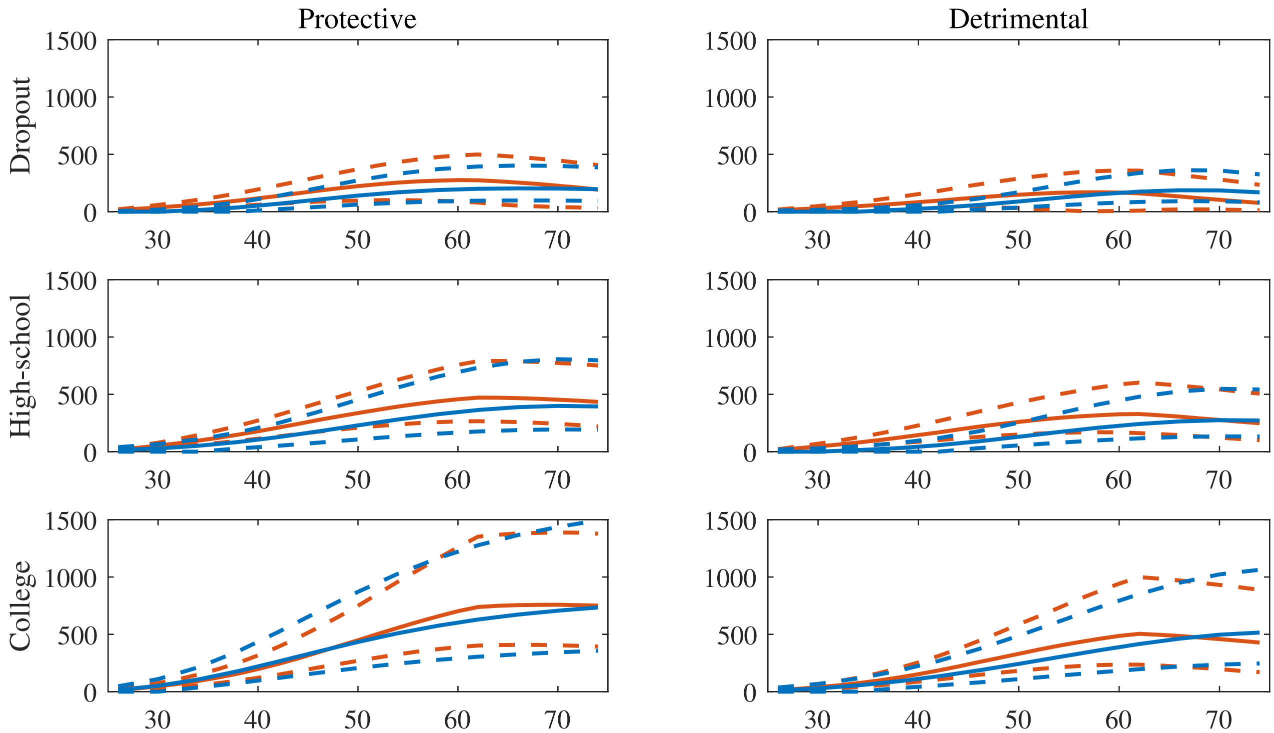
The third column of Table 5 reports our estimated parameters. The estimated bequest parameters b_0 and b_1 are 8.86 and 376.64, respectively. In a model with risk aversion equal to 1, these values imply that in the period before certain death, the bequest motives becomes active at \$42,500 and the marginal propensity to bequeath is 90 cents out of every additional dollar for bequests. The income floor is estimated at \$17,400 dollar. These parameters fall in the range of parameters estimated by the previous literature.

Figure 6 compares the median wealth from our model (lines) and data (diamonds). Overall, our model is able to generate the fact that higher-educated and health protective individuals accumulate more wealth. It is worth noting that although not specifically targeted, the model is able to match

Table 5: Internally calibrated parameters

Parameter	Description	Value
$\underline{x}$	income floor	17.40
b_0	bequest motive: marginal utility	8.86
b_1	bequest motive: non-homoteticity	376.64
b	value of life	0.94

Figure 6: Model fit: wealth distribution model (lines) versus data (scatter)



the overall distribution, as indicated by the 25th and 75th quantiles in both the data (squares) and the model (dashed lines).

6.2 Value of a Statistical Life (VSL)

The VSL comes from the estimated wage premium for a given probability of a fatal accident in risky jobs. This literature delivers numbers in the range of \$1 to \$7 million to save one life. We want the model to deliver this same marginal rate of substitution between income and survival probability. Using the value function expressed in Section 5.2.3 we can obtain the total differential:

$$\frac{\partial V_t^{\mathrm{c,e,y}}(x,h,\ell,\xi)}{\partial x}dx + \frac{\partial V_t^{\mathrm{c,e,y}}(x,h,\ell,\xi)}{\partial s_t^{\mathrm{e,y}}(h)}ds_t^{\mathrm{e,y}}(h) = 0$$

Table 6: Parameters of education and lifestyle costs

Parameter	Description	Value	Parameter	Description	Value
μ_{HSG}	Average cost of HSG education	5.38	σ_{HSG}	Dispersion in cost of HSG education	2.99
μ_{CG}	Average cost of CG education	27.37	σ_{CG}	Dispersion in cost of CG education	24.53
μ_{PRO}	Average cost of PRO lifestyle	2.58	σ_{PRO}	Dispersion in cost of PRO lifestyle	6.07

Notes: Estimated parameters for the distributions of the utility costs τ_{PRO}, τ_{CG}, and τ_{HSG}.

relating changes in cash-on-hand x and survival probabilities $s_t^{\text{e,y}}(h)$ that leave individuals indifferent. Rearranging we obtain,

$$-\frac{dx}{ds_t^{\text{e,y}}(h)} = \frac{\partial V_t^{\text{c,e,y}}(x,h,\ell,\xi,\zeta)}{\partial s_t^{\text{e,y}}(h)} \left[\frac{\partial V_t^{\text{c,e,y}}(x,h,\ell,\xi)}{\partial x}\right]^{-1}$$

Hence, for an individual of type (c,e,y) with state variables (x,h,ℓ,ξ) at age t to accept an increase in his survival probability in say 1%, he would require $0.01 \times dx/ds_t^{\text{e,y}}(h)$ units of income. Putting 100 identical agents together we would have one death on average in exchange of $dx/ds_t^{\text{e,y}}(h)$ units of income. Hence, this expression gives the model equivalent of the VSL. Because the empirical estimates of a VSL typically come from blue-collar jobs, we want the model to deliver a VSL for the average high school dropout of 35 years of age. We target a VSL of 1,000,000 for a high-school graduate aged 40 born in the 1930s. This identifies the parameter $\bar{b}$.

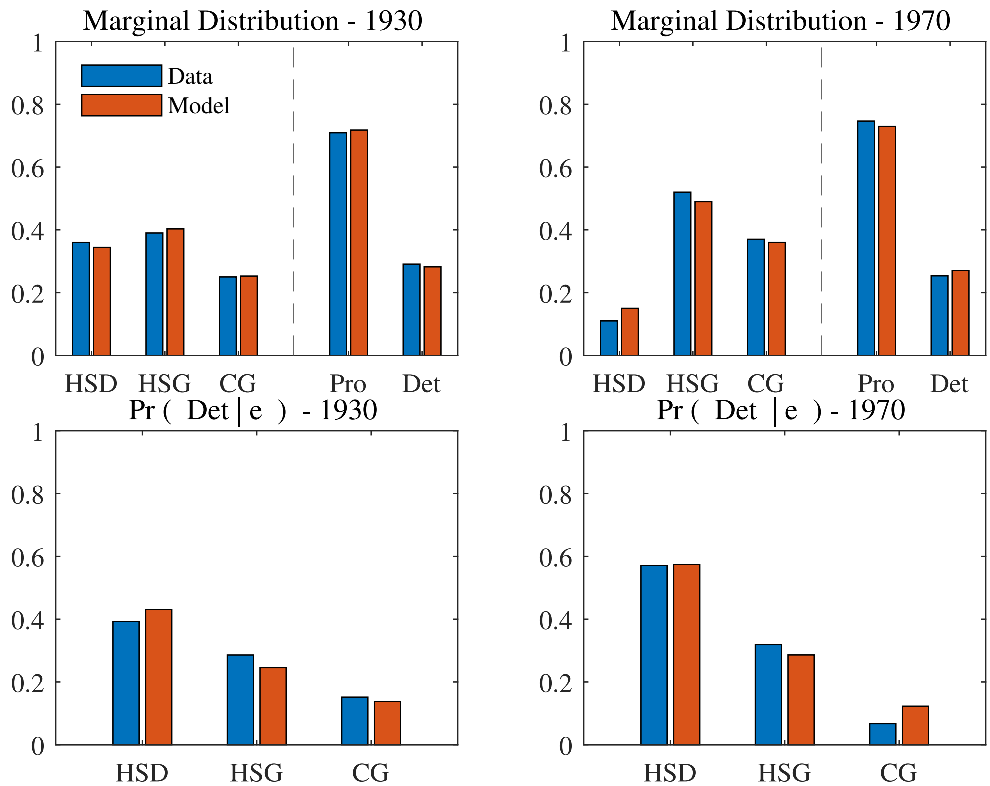
6.3 Stage 1: Early life

In order to calibrate the parameters of the early-in-life model, we require the model to match the joint distribution of education and lifestyles for two different cohorts: 1930 and 1970. The value functions for each cohort vary differently in terms of education and lifestyle due to the different paths of wages and employment rates over the life cycle, with the expected real labor earnings of high school dropouts being slightly lower for the 1970s cohort, the expected real labor earnings of college graduates being larger for the 1970 cohort, and hence the college premium being larger for the 1970 cohort.

6.3.1 Identification and results

We have 6 parameters to estimate: $(\mu_{\text{CG}}, \mu_{\text{HSG}}, \mu_{\text{PRO}}, \sigma_{\text{CG}}, \sigma_{\text{HSG}}, \sigma_{\text{PRO}})$. Given that we require the model to match the joint distribution of (e,y) for two different cohorts we have 10 moment conditions ($3 \times 2 - 1$ targets per year). In general, one may think that the average cost parameters $(\mu_{\text{CG}}, \mu_{\text{HSG}}, \mu_{\text{PRO}})$ are identified by the average share of individuals in each education and life-style group, while the dispersion parameters $(\sigma_{\text{CG}}, \sigma_{\text{HSG}}, \sigma_{\text{PRO}})$ are identified by the changes across cohorts of the share of individuals in each education and lifestyle category as the wage paths of different

Figure 7: Model Fit: First-Stage



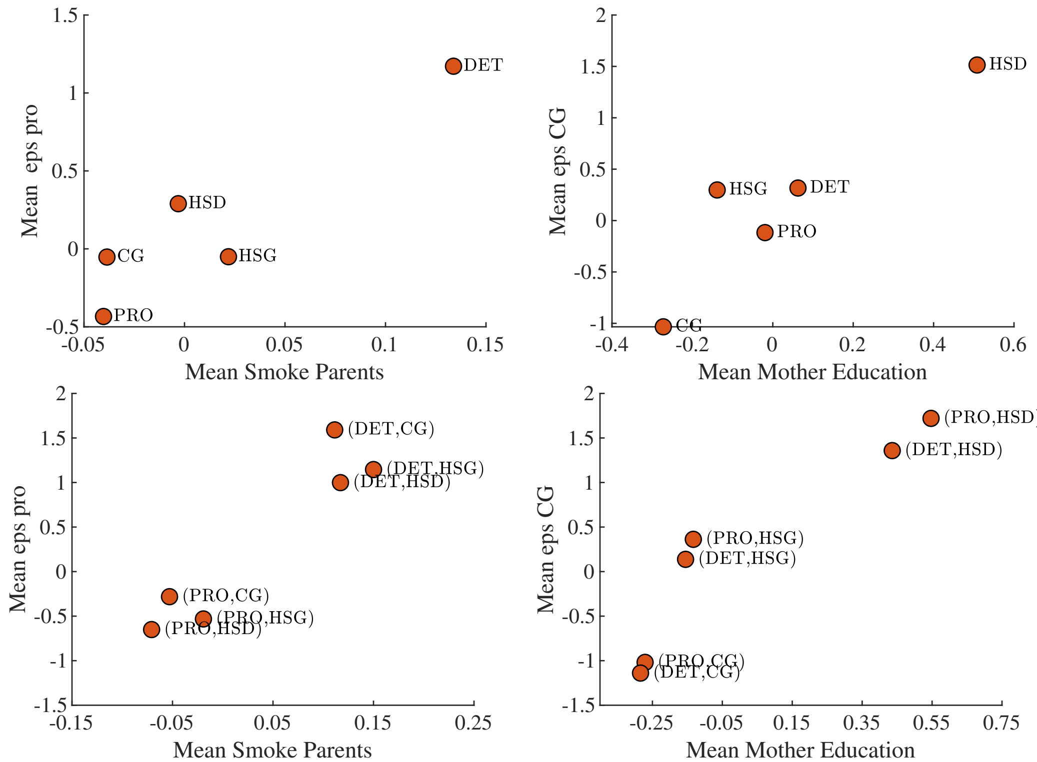
Notes: The top two panels report the marginal distribution of education and lifestyle types for two different cohorts. The bottom panel reports the distribution of lifestyle types conditional on education choices for the same two cohorts.

cohorts change.[11] However, by targeting the joint distribution of education and lifestyles, these parameters must also shape the level and changes in the distribution of lifestyles conditional on education.

The upper panels of Figure 7 shows that the model is able to match well the marginal distributions of education and health behavior as well as their changes across cohorts. Moreover, the lower panels of Figure 7 shows that the model is also able to match the level and changes in the distribution of lifestyles conditional on education. In particular, the model reproduces well the fact that less educated individuals tend to invest less in their health. In the 1930s cohort, the proportion of individuals with *detrimental* lifestyles among high-school dropouts is 39% in the data and 43% in the model, while for college graduates these figure is 15.1% in the data and 13.8% in the model. Furthermore, given the observed changes in wage and employment trajectories by education, the model reproduces well the fact that the less educated have worsened their health-related behavior. On the other hand, the model is less successful at matching the improvements in health behavior of the higher-educated. The proportion of individuals with detrimental lifestyles among high-school dropouts has increased by 17.8 percentage points in the data and 14.3 percentage points in the

[11] This extends the identification strategy of Heathcote et al. (2010), whose first stage contains a college education choice but does not consider a lifestyle choice.

FIGURE 8: shocks and parental characteristics



Notes: The top two panels report the average τ_{PRO} (top left) and the average τ_{CG} (top right) for each education and lifestyle category in the calibrated model (Y-axis) against the average share of smoking parents (top left) and average share of high-school dropout mothers (top right) for the same groups in the data (X-axis). The bottom two panels report the same information for each education-lifesytle group.

model, while the proportion of individuals with detrimental lifestyles among college graduates has declined by 8.5 percentage points in the data and 1.5 percentage points in the model. Consequently, the increase in the life-expectancy gradient is estimated at 1.1 years in the model and 1.9 years in the data. Hence, the calibrated model, with only wage and employment changes, accounts for around 60% of the overall increase in the life-expectancy gradient between college graduates and high-school dropouts between the 1930 to 1970 cohorts.

6.3.2 Validation

We cannot observe τ_{PRO} and τ_{CG} in the data, but we can use the rich data in the HRS to look for proxies of these costs. In particular, in the HRS we have data on whether the main carer smoked when the respondent was a kid and on parental education (whether the mother did not finish high school). Taking these two variables as imperfect proxies for τ_{PRO} and τ_{CG} respectively, we compare the pattern of selection of τ_{PRO} and τ_{CG} within each education and lifestyle group in the model with the one observed in the data.

First, the model predicts that individuals who choose a *protective* lifestyle have on average lower τ_{PRO} and individuals who choose college education have on average lower τ_{CG}. We can see this in the Y axes on the top two panels in Figure 8. Likewise, in the data individuals classified as *protective* are less likely to have smoking parents than individuals classified as *detrimental*, and more educated individuals are less likely to have a high school dropout mother, see the X axes on the top two panels in Figure 8.

Second, the model predicts that individuals who choose a *protective* lifestyle also have on average lower τ_{CG} and individuals who study college education have on average lower τ_{PRO}. This is due to the complementarity of the two investments, and it is clearly apparent in the Y axes of the top two panels in Figure 8. In the data, we also find that individuals classified as *protective* are less likely to have a high school dropout mother than those classified as *detrimental*, and that college educated individuals are less likely to have smoking parents than high school dropouts. Only the high-school graduates do not conform with the model predictions, as the share of them with smoking parents, while being above the one of college graduates it is also above the one of high school dropouts.

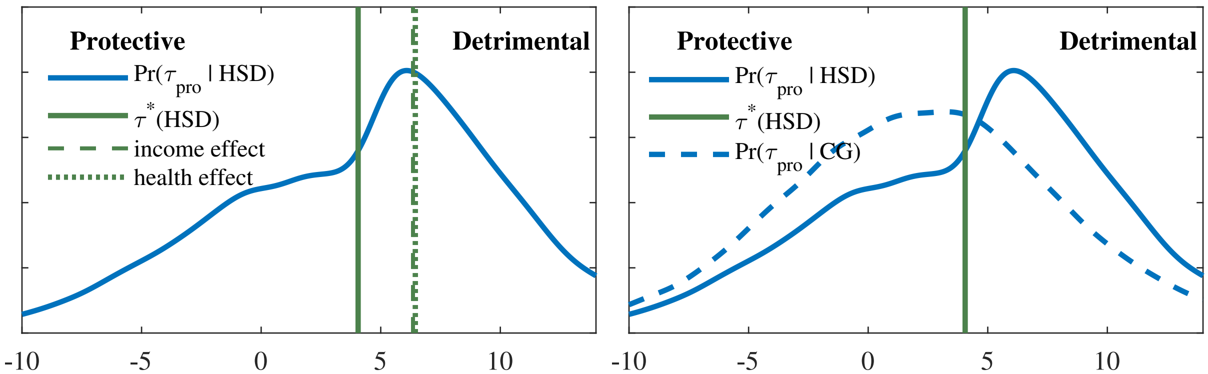
Third, within each lifestyle choice, the least educated individuals have lower τ_{PRO}, that is, they are better selected in terms of the utility cost of good lifestyle, see the Y axis in the bottom-left panel of Figure 8. The reason is that, due to the complementarity of investments, what leads a low-educated individual to choose a *protective* lifestyle is a low cost of *protective* lifestyle compared to the more educated individuals. Analogously, what leads a highly educated individual to choose a *detrimental* lifestyle is the high cost of *protective* lifestyle compared to the less educated individuals. In the data, we see that these patterns are also present: among the individuals classified as *protective*, the parents of those who go to college are more likely to smoke than the parents of those that drop out from high school. Among the individuals classified as *detrimental*, the parents of those who go to college are also more likely to smoke than the parents of those that drop out from high school, see the X axis in the bottom-left panel of Figure 8.

And fourth, within education choice, the individuals with *detrimental* lifestyle have lower τ_{CG} than individuals with *protective* lifestyle, see the Y axis in the bottom-right panel of Figure 8. That is, due to the complementarity of investments, *detrimental* individuals are better selected in terms of education costs. The pattern in the data is similar. Both within high school dropouts and within high school graduates, the mothers of individuals classified as *detrimental* are less likely to be high-school dropouts, see the X axis in the right panel of Figure 8. Among college graduates, the likelihood of high school dropout mothers is virtually the same across lifestyle groups.

7 Counterfactual results

Given our calibrated model, we aim to answer two questions. First, why is there an education gradient of lifestyles (Section 7.1). Second, what has been the effect of the rise in the education wage premium on the increase in education gradient of life expectancy across cohorts (Section 7.2).

Figure 9: Decomposition of lifestyle choices conditional on education (1970 cohort)



Notes: In the left panel the blue line describes the density function of τ_{PRO} for individuals who choose HSD education. The solid thick green vertical line reports the threshold τ_{PRO} that separates HSD individuals into *detrimental* and *protective*. The dashed green vertical lines report this same threshold if HSD individuals had, respectively, the income prospects and health yield of *protective* lifestyle as CG individuals. The dashed blue line in the right panel describes the density function of τ_{PRO} for individuals who choose CG education.

7.1 The education gradient of lifestyles

The model incorporates three mechanisms that can generate and educational gradient of lifestyles. Firstly, higher expected income among the more educated encourages healthier behavior as life becomes more valuable with higher consumption possibilities. Secondly, as detailed in Section 3, the estimated health transitions imply a higher yield of a *protective* lifestyle investment for the more educated (the *protective* lifestyle has a larger effect on health outcomes for the more educated). Finally, given the two previous complementarities, individuals facing lower costs of protective behavior (lower τ_{PRO}) are more likely to pursue higher education. This means that low τ_{PRO} individuals will be more frequent among the highly-educated.

To gauge the importance of each mechanism, we start by showing graphically the choice of lifestyle conditional on an educational choice. The solid blue line in the left panel of Figure 9 represents the distribution of protective behavior costs (τ_{PRO}) for individuals who choose to drop out of high-school in the 1970 cohort. The vertical solid green line represents $V^{70,\text{HSD},\text{PRO}} - V^{70,\text{HSD},\text{DET}}$, the value of choosing a *protective* lifestyle over a *detrimental* lifestyle for those choosing to dropout of high school. Individuals with $\tau_{\text{PRO}} < V^{70,\text{HSD},\text{PRO}} - V^{70,\text{HSD},\text{DET}}$ opt for protective behavior, while the rest choose detrimental behavior. The integral of the distribution of τ_{PRO} between minus infinity and the $V^{70,\text{HSD},\text{PRO}} - V^{70,\text{HSD},\text{DET}}$ line represents the fraction of dropouts adopting protective behavior.

Next, we conduct a series of counterfactual experiments. In these experiments, we keep individuals' education choices fixed and observe how their health investments would differ under various scenarios. In the first scenario, we simulate the behavior of high-school dropouts if they were to have the income prospects of college graduates. Due to the higher consumption possibilities, the increase in expected earnings leads to higher values for both y types, which we denote as $\widetilde{V}^{70,\text{HSD},\text{PRO}}$ and $\widetilde{V}^{70,\text{HSD},\text{DET}}$. However, this gain is higher for y = PRO, as life ex-

Table 7: Decomposition: 1970s cohort

	Pr(PRO|HSD)	Pr(PRO|CG)	ΔPr(PRO)	LE(HSD)	LE(CG)	ΔLE
Model	0.43	0.88	0.45	24.6	31.9	7.3
Same lifestyle	0.88	0.88	0.00	27.7	31.9	4.2
Health	0.60	0.88	0.28	25.9	31.9	6.0
Income	0.60	0.88	0.28	25.9	31.9	6.0
Selection	0.62	0.88	0.26	26.5	31.9	5.8

Notes: Model predictions and counterfactual exercises.

pectancy is higher for this type and the higher consumption flow is enjoyed for more years. Thus, $\widetilde{V}^{70,\text{HSD},\text{PRO}} - \widetilde{V}^{70,\text{HSD},\text{DET}} > V^{70,\text{HSD},\text{PRO}} - V^{70,\text{HSD},\text{DET}}$, which means that the threshold value that prompts individuals to adopt *protective* health behavior shifts to the right. The dashed vertical line in Figure 9 marks the value $\widetilde{V}^{70,\text{HSD},\text{PRO}} - \widetilde{V}^{70,\text{HSD},\text{DET}}$. This indicates that income largely influences why high-school dropouts tend to adopt more detrimental behavior. If faced with the same expected income as college graduates, the proportion of high-school dropouts choosing detrimental behavior would decrease from 57% to 40%, reducing the life-expectancy gap in 1.1 years (from 7.5 to 6.4 years, a 15% reduction).

In the second scenario, we simulate the behavior of high-school dropouts if they were to face the same relative gains from *protective* health transitions as college graduates. For this purpose, we first compute a counterfactual health transition model for dropout protective that matches the relative gains from protective behavior to detrimental behavior that we observed for the college graduates. The improved health transitions for dropout protective result in higher values of $\widehat{V}^{70,\text{HSD},\text{PRO}}$. The vertical dotted line in the left panel of Figure 9 represents the value $\widehat{V}^{70,\text{HSD},\text{PRO}} - V^{70,\text{HSD},\text{DET}}$ in this counterfactual scenario. If faced with the relative gains from protective behavior as college graduates, the proportion of high-school dropouts choosing detrimental behavior would decrease from 57% to 40%, reducing the life-expectancy gap as much as the income effect.

Lastly, in the third scenario we compute the fraction of individuals that choose *detrimental* type among high-school dropouts in the case that high-school dropouts had the same distribution of τ_{PRO} as college graduates. The right panel in Figure 9 displays the distribution of the cost of protective behavior for both high-school dropouts and college (dashed line). It shows that the mass of individuals below the $V^{70,\text{HSD},\text{PRO}} - V^{70,\text{HSD},\text{DET}}$ threshold (vertical line) is larger for the distribution of protective costs faced by the college individuals. If faced with the distribution of costs τ_{PRO} of college-educated individuals, the proportion of detrimental behavior would decrease from 57% to 38%. Therefore, selection would account for approximately 17% of the gradient in life expectancy.

Table 8: Decomposition

	Δ_cPr(CG)	Δ_cPr(HSD)	Δ_cPr(PRO)	Δ_cPr(PRO|CG)	Δ_cPr(PRO|HSD)	$\Delta_c\Delta_e$LE
Data	0.12	-0.25	0.04	0.08	-0.18	1.92
Model	0.11	-0.19	0.02	0.01	-0.14	1.06
Income				0.01	-0.05	0.44
Selection				0.00	-0.09	0.63

Notes: Model predictions and counterfactual exercises.

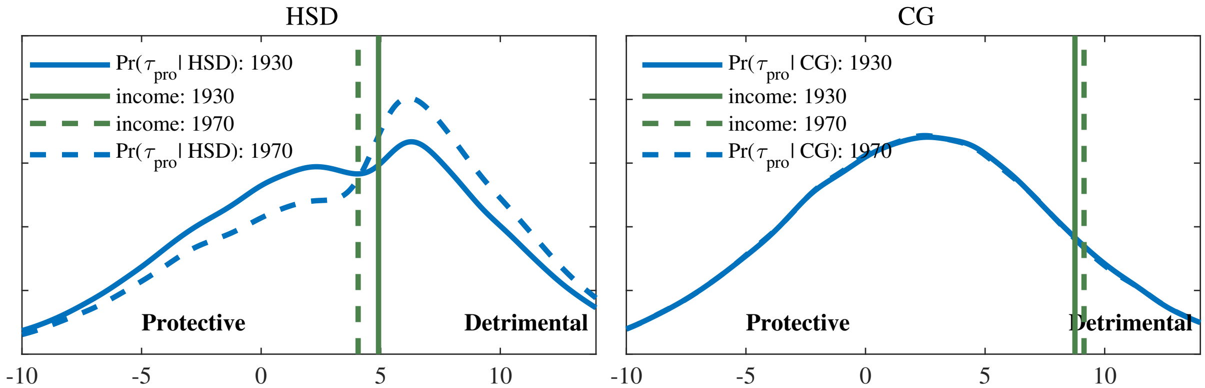
7.2 Changes over time

The econometric model estimated in Section 4 indicates an increase in the life expectancy gap between college graduates and high school dropouts of 1.9 years across the cohorts from the 1930s to the 1970s. This increase is driven by lower (higher) health investments among high school dropouts (college graduates). Our calibrated model, as discussed above, replicates 60% of this observed increase. The question here is how much of this increase is due to the direct effect of changes in individuals' income prospects and how much is due to changes in the composition of individuals in each education group.

To quantify the income effect while controlling for selection, we fix the distribution of health behavior costs faced by individuals in different education categories to the one in 1930. We then analyze how their health investments would have changed solely due to changes in income. In Figure 10, the solid line represents the distribution of the cost of protective behavior faced by individuals in 1930 for high-school dropouts (left panel) and college graduates (right panel). The vertical solid line indicates the threshold value at which individuals of that cohort decided to switch from a protective to a detrimental lifestyle. The dashed vertical line illustrates that for high-school dropouts, the threshold value decreased across cohorts due to declines in income and consequently, individuals dropping out of high school are less willing to invest in their health. Increasing divergence in income prospects across education groups account for 42% of the increase in the life expectancy gradient for individuals born in the 1930s and 1970s.

The dashed line in Figure 10 illustrates the distribution of behavior costs in the 1970s. Driven by changes in income across cohorts and education categories, compared to the distribution in 1930, the cost distribution worsened for high school dropouts born in the 1970s but did not quantitatively change for the college graduates. To quantify the selection effect while controlling for income, we fix the income prospects faced by individuals in different education categories to the one in 1930. We then analyze how their health investments would have changed solely due to changes in the distribution of health behavior costs. We find that if only the distribution of health behavior cost would have changed, the life-expectancy gradient would have increased 0.63 year or 60% of the increase in the life expectancy gradient.

Figure 10: Cost of protective lifestyle across cohorts



Notes: In the left panel the blue line describes the density function of τ_{PRO} for individuals in the 1930 cohort who chose HSD education. The solid green vertical line reports the threshold τ_{PRO} that separates HSD individuals into *detrimental* and *protective* in the 1930 cohort. The dashed green vertical line reports this same threshold if the income prospects of HSD individuals were as in the 1970 cohort. The dashed blue line describes the density function of τ_{PRO} for individuals who choose HSG education in the 1970 cohort. The right panel reports the same information for CG individuals.

We note that other authors (Cutler et al., 2011; Novosad et al., 2022) also find a decline over time in the life expectancy of the least educated controlling for selection in different ways. Our findings are consistent with these results, but they show a strong selection effect too.

8 Conclusions

In this paper, we propose a latent variable model to characterize how health behavior influences health dynamics across different education groups. Our findings indicate that health behavior can be parsimoniously summarized into two lifestyles: protective and detrimental. We observe that individuals with higher levels of education tend to more frequently choose protective behavior, and differences in lifestyles explain 40% of the variations in life expectancy across education groups. Additionally, conditional on behavior, we find that a protective lifestyle has a greater impact on extending life expectancy for college graduates than for dropouts. Finally, we identify an increasing life-expectancy gradient across education groups between the 1930s and 1970s, driven by worsening lifestyles among the less educated and improved lifestyles among the more educated.

We introduce a heterogeneous agents model comprising two distinct stages. Initially, individuals make a one-time decision regarding education and lifestyle, with the value of each choice given by the second stage of the model. In the second stage, agents address a consumption savings problem subject to income and health risks, as modeled in the econometric framework.

This model enables us to explain the connection between income and health inequality. Health and education decisions are shown to be complementary due to two key factors. Firstly, higher income increases the value of life, leading to greater returns from investing in health. Secondly, as reflected in our calibration process where we integrate the health dynamics from the econometric model, we observe greater returns to health investment for college-educated individuals. Driven by these complementarities, individuals facing higher costs of maintaining protective health behaviors are more likely to self-select into lower education categories.

We calibrate the model to match savings, education, and lifestyle choices across cohorts, and then we use it to understand why lower-educated individuals tend to choose unhealthier lives. Our analysis reveals that lower income and diminished returns in health outcomes from protective behavior largely account for the disparities observed across education groups. However, selection also matters.

Finally, the model is able to explain 50% of the increase in health inequality across the 1930s and the 1970s. 40% of the deterioration in life expectancy among the less educated is driven by worsening economic conditions, and 60% is attributed to selection effects. All improvements in lifestyle among college graduates are explained by improvements in economic conditions.

The model cannot fully account for the increase in the life-expectancy gradient observed in the data. Factors such as peer influence, segregation, genetic predispositions, and variations in intergenerational mobility across cohorts are likely significant drivers of health behavior choices made by individuals, which we abstract from in our current analysis. These avenues offer promising directions for future research.

References

Ameriks, J., J. Briggs, A. Caplin, M. Shapiro, and C. Tonetti (2020): "Long-Term Care Utility and Late in Life Saving," *Journal of Political Economy*, 128, 2375–2451. (Cited on page 4.)

Aron-Dine, A., L. Einav, and A. Finkelstein (2013): "The RAND Health Insurance Experiment, Three Decades Later," *Journal of Economic Perspectives*, 27, 197–222. (Cited on page 4.)

Baicker, K., S. Taubman, H. Allen, M. Bernstein, J. Gruber, J. P. Newhouse, E. C. Schneider, B. Wright, A. M. Zaslavsky, and A. Finkelstein (2013): "The Oregon Experiment - Effects of Medicaid on Clinical Outcomes," *New England Journal of Medicine*, 368, 1713–1722. (Cited on page 4.)

Bairoliya, N., R. Miller, and V. M. Nygaard (2024): "Exercise or Extra Fries? How Behaviors Impact Health Over the Life Cycle," Mimeo. (Cited on page 5.)

Boháček, R., J. Bueren, L. Crespo, P. Mira, and J. Pijoan-Mas (2021): "Inequality in Life Expectancies across Europe and the US," *Health Economics*, 30, 1871–1885. (Cited on page 4.)

Borella, M., F. Bullano, M. De Nardi, B. Krueger, and E. Manresa) (2024): "Health Inequality and Health Types," Mimeo. (Cited on page 4.)

Braun, A., K. Kopecky, and T. Koreshkova (2019): "Old, Frail, and Uninsured: Accounting for Puzzles in the US Long-Term Care Insurance Market," *Econometrica*, 87, 981–1019. (Cited on page 4.)

Bueren, J. (2023): "Long-Term Care Needs and Savings in Retirement," *Review of Economic Dynamics*, 49, 201–224. (Cited on page 4.)

Capatina, E. (2015): "Life-cycle Effects of Health Risk," *Journal of Monetary Economics*, 74, 67–88. (Cited on page 4.)

Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by L. N. Christofides, E. K. Grant, and R. Swidinsky, Toronto: University of Toronto Press, 201–222. (Cited on page 14.)

Case, A. and A. Deaton (2015): "Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century," *Proceedings of the National Academy of Sciences*, 112, 15078–15083. (Cited on page 1.)

——— (2017): "Mortality and Morbidity in the 21st Century," *Brooking Papers on Economic Activity*, Spring, 397–443. (Cited on page 3.)

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016): "The association between income and life expectancy in the United States, 2001-2014," *JAMA*, 315, 1750–1766. (Cited on page 1.)

Cole, H. L., S. Kim, and D. Krueger (2019): "Analysing the effects of insuring health risks: On the trade-off between short-run insurance benefits versus long-run incentive costs," *The Review of Economic Studies*, 86, 1123–1169. (Cited on page 4.)

Conner, M. and P. Norman (2017): "Health Behaviour: Current Issues and Challenges," *Psychology & Health*, 32, 895–906. (Cited on page 14.)

Cutler, D. M., F. Lange, E. Meara, S. Richards-Shubik, and C. J. Ruhm (2011): "Rising educational gradients in mortality: The role of behavioral risk factors," *Journal of Health Economics*, 30, 1174–1187. (Cited on page 29.)

Cutler, D. M. and A. Lleras-Muney (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29, 1–28. (Cited on page 1.)

De Nardi, M., E. French, and J. Jones (2010): "Why Do The Elderly Save? The Role Of Medical Expenses," *Journal of Political Economy*, 118, 39–75. (Cited on page 4.)

——— (2016): "Medicaid Insurance in Old Age," *American Economic Review*, 106, 3480–3520. (Cited on page 4.)

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2022): "The Lifetime Costs of Bad Health," NBER Working Paper 23963. (Cited on pages 4 and 43.)

Farrell, P. and V. Fuchs (1982): "Schooling and Health: the Cigarette Connection," *Journal of Health Economics*, 1, 217–230. (Cited on page 14.)

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, and K. Baicker (2012): "The Oregon Health Insurance Experiment: Evidence from the First Year," *Quarterly Journal of Economics*, 127, 1057–1106. (Cited on page 4.)

Fonseca, R., P.-C. Michaud, T. Galama, and A. Kapteyn (2021): "Accounting for the rise of health spending and longevity," *Journal of the European Economic Association*, 19, 536–579. (Cited on page 4.)

French, E. and J. Jones (2011): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79, 693–732. (Cited on page 4.)

Gold, D. T. and B. McClung (2006): "Approaches to Patient Education: Emphasizing the Long-Term Value of Compliance and Persistence," *The American Journal of Medicine*, 119, 32S–37S. (Cited on page 2.)

Hai, R. and J. Heckman (2022): "The Causal Effects of Youth cigarette Addiction and Education," NBER Working Paper 30304. (Cited on pages 4 and 14.)

Hauser, R. M. and D. L. Featherman (1976): "Equality of Schooling: Trends and Prospects," *Sociology of Education*, 49, 99–120. (Cited on page 14.)

Heathcote, J., K. Storesletten, and G. Violante (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, 118, 681–722. (Cited on page 23.)

Heathcote, J., K. Storesletten, and G. L. Violante (2020): "Optimal progressivity with age-dependent taxation," *Journal of Public Economics*, 189, 104074. (Cited on page 17.)

Holter, H. A., D. Krueger, and S. Stepanchuk (2019): "How do tax progressivity and household heterogeneity affect Laffer curves?" *Quantitative Economics*, 10, 1317–1356. (Cited on page 43.)

Hong, J., J. Pijoan-Mas, and V. Ríos-Rull (2024): "Health, Consumption, and Inequality," Mimeo. (Cited on page 4.)

Hosseini, R., K. A. Kopecky, and K. Zhao (2021): "How important is health inequality for lifetime earnings inequality?" FRB Atlanta Working Paper. (Cited on page 4.)

Keane, M. and K. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522. (Cited on page 14.)

Kitagawa, E. M. and P. M. Hauser (1973): *Differential mortality in the United States: A study in socioeconomic epidemiology*, Harvard University Press. (Cited on page 1.)

Lantz, P. M., J. S. House, J. M. Lepkowski, D. R. Williams, R. P. Mero, and J. Chen (1998): "Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults," *JAMA*, 279, 1703–1708. (Cited on page 1.)

Li, Y., A. Pan, D. D. Wang, X. Liu, K. Dhana, O. H. Franco, S. Kaptoge, E. Di Angelantonio, M. Stampfer, W. C. Willett, et al. (2018): "Impact of healthy lifestyle factors on life expectancies in the US population," *Circulation*, 138, 345–355. (Cited on page 1.)

Lleras-Muney, A. and F. R. Lichtenberg (2005): "Are the More Educated More Likely to Use New Drugs?" *Annales d'Economie et de Statistique*, 79/80, 671–696. (Cited on page 3.)

Mahler, L. and M. Yum (2022): "Lifestyle behaviors and wealth-health gaps in Germany," *Available at SSRN 4034661.* (Cited on page 4.)

Margaris, P. and J. Wallenius (2023): "Can Wealth Buy Health? A Model of Pecuniary and Non-Pecuniary Investments in Health," *Journal of the European Economic Association.* (Cited on page 4.)

McGinnis, J. M. and W. H. Foege (1993): "Actual Causes of Death in the United States," *JAMA*, 270, 2207–2212. (Cited on page 1.)

Meara, E. R., S. Richards, and D. M. Cutler (2008): "The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000," *Health affairs*, 27, 350–360. (Cited on page 1.)

Nakajima, M. and I. A. Telyukova (2023): "Medical Expenses and Saving in Retirement: The Case of U.S. and Sweden," Mimeo. (Cited on page 4.)

Novosad, P., C. Rafkin, and S. Asher (2022): "Mortality Change among Less Educated Americans," *American Economic Journal: Applied Economics*, 14, 1–34. (Cited on page 29.)

Ozkan, S. (2023): "Preventive vs. Curative Medicine: A Macroeconomic Analysis of Health Care over the Life Cycle," St Louis FED Working Paper # 2023-025B. (Cited on page 4.)

Pijoan-Mas, J. and J.-V. Ríos-Rull (2014): "Heterogeneity in expected longevities," *Demography*, 51, 2075–2102. (Cited on pages 1 and 8.)

Polvinen, A., R. Gould, E. Lahelma, and P. Martikainen (2013): "Socioeconomic differences in disability retirement in Finland: the contribution of ill-health, health behaviours and working conditions," *Scandinavian Journal of Public Health*, 41, 470–478. (Cited on page 1.)

Preston, S. H. and I. T. Elo (1995): "Are educational differentials in adult mortality increasing in the United States?" *Journal of Aging and Health*, 7, 476–496. (Cited on page 1.)

Taylor, D. H., V. Hasselblad, S. J. Henley, M. J. Thun, and F. A. Sloan (2002): "Benefits of Smoking Cessation for Longevity," *American Journal of Public Health*, 92, 990–996. (Cited on page 1.)

Willis, R. and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, s7–s36, pt. 2. (Cited on page 14.)

Xiang, S., T. Jia, C. Xie, W. Cheng, B. Chaarani, T. Banaschewski, G. J. Barker, A. L. W. Bokde, C. Büchel, S. Desrivières, H. Flor, A. Grigis, P. A. Gowland, R. Brühl, J.-L. Martinot, M.-L. P. Martinot, F. Nees, D. P. Orfanos, L. Poustka, S. Hohmann, J. H. Fröhner, M. N. Smolka, N. Vaidya, H. Walter, R. Whelan, H. Garavan, G. Schumann, B. J. Sahakian, T. W. Robbins, and J. Feng (2023): "Association between vmPFC Gray Matter Volume and Smoking Initiation in Adolescents," *Nature Communications*, 14, 4684. (Cited on page 14.)

Zaninotto, P., J. Head, and A. Steptoe (2020): "Behavioural risk factors and healthy life expectancy: evidence from two longitudinal studies of ageing in England and the US," *Scientific Reports*, 10, 6955. (Cited on page 1.)

Appendix A: Estimation details of the health model

In this appendix, we explain the details of the estimation of the health model in section 4.

Individuals in the sample can be aged (a) between 25 ($t = 1$) and 99 ($t = 38$). Thus each time period consists of two years. We observe an unbalanced panel of individuals $i = 1, \ldots, N$ followed for $t_i = \underline{t_i}, \ldots, \bar{t_i}$. For each individual, we observe M categorical variables describing habits related to health behaviors which also evolve across time ($z_{1,i,t}, z_{2,i,t}, \ldots, z_{M,i,t}$) provided the individual is alive and interviewed. On top, we observe health status which can be good ($h_{it} = g$), bad ($h_{it} = b$) or dead ($h_{it} = D$). We also observe a set of time invariant covariates ($X_i = \{e_i, g_i, c_i\}$): education (e_i), sex (s_i) and cohort (c_i).

We assume that there is one source of unobserved heterogeneity in the population: lifestyle (y). Health behavior is assumed to have a finite number possible groups $y \in \{1, \ldots, Y\}$. We assume that individual lifestyle is fixed across time.

Let's denote β the parameters associated to health transitions, γ the parameters associated to health behaviors and δ the parameters associated to the probability of having of lifestyle y. We can write the likelihood of the data as a mixture model:

$$p(h_i^T, z_i^T | h_{\underline{t_i}}, X_i, \beta, \gamma, \delta) = \sum_{y=1}^{K} p(h_i^T, z_i^T | \beta, \gamma, X_i, y) p(y | h_{\underline{t_i}}, X_i, \beta, \delta) \tag{6}$$

$$= \sum_{y=1}^{K} p(z_i^T | h_i^T, \gamma, X_i, y) p(h_i^T | \beta, X_i, y) p(y | h_{\underline{t_i}}, X_i, \beta, \delta) \tag{7}$$

Each component of equation 7 is described in section 3.

A.1 Gibbs sampler

We can write the complete data likelihood as:

$$p(\boldsymbol{h}^T, \boldsymbol{z}^T, \hat{\boldsymbol{y}} | h_{\underline{t_i}}, X_i, \beta, \gamma, \delta) = \prod_{i=1}^{n} \prod_{y=1}^{K} \Big(p(z_i^T | h_i^T, \gamma, X_i, y) p(h_i^T | h_{\underline{t_i}}, \beta, X_i, y) p(y | h_{\underline{t_i}}, X_i, \beta, \delta) \Big)^{\mathbb{1}\{\hat{y}_i = y\}},$$

where $\hat{y}_i$ denotes the lifestyle assigned to individual i and $\boldsymbol{h}^T, \boldsymbol{z}^T$ and $\hat{\boldsymbol{y}}$ denote health status, health behaviors, and assigned lifestyles for every individual in the sample, respectively.

We can now write the joint posterior distribution as:

$$p(\beta,\gamma,\delta|X_i,\boldsymbol{h}^T,\boldsymbol{z}^T,\hat{\boldsymbol{y}}) \propto \prod_{i=1}^{n}\prod_{y=1}^{K}\Big(p(z_i^T|h_i^T,\gamma,X_i,y)p(h_i^T|\beta,X_i,y)p(y|h_{\underline{t_i}},X_i,\beta,\delta)\Big)^{\mathbb{1}\{\hat{y}_i=y\}}$$
$$p(\beta)p(\gamma)p(\delta),$$

where $p(\beta), p(\gamma)$, and $p(\delta)$ denote the priors associated to each parameters.

Block 1: transition parameters

The first block is given by

$$p(\beta|X_i,\boldsymbol{h}^T,\boldsymbol{z}^T,\hat{\boldsymbol{y}},\gamma,\delta) \propto \prod_{y=1}^{K}\prod_{i:\hat{y}_i=y} p(h_i^T|h_{\underline{t_i}},\beta,X_i,y)p(y|h_{\underline{t_i}},X_i,\beta,\delta)p(\beta)$$

We partition the β parameters into those driving survival (β^s) and those driving health transitions given survival(β^h) and compute the posterior distribution of each given the other:

$$p(\beta^s|X_i,\boldsymbol{h}^T,\boldsymbol{z}^T,\hat{\boldsymbol{y}},\gamma,\delta,\beta^h) \propto \prod_{y=1}^{K}\prod_{i:\hat{y}_i=y}\prod_{t=\underline{t}_i}^{\bar{t}_i-1} p(\mathbb{I}_{h_{i,t+1}\neq D}|h_{i,t},\beta^s,y,X_i)p(y|h_{\underline{t_i}},X_i,\beta,\delta),$$
$$p(\beta^h|X_i,\boldsymbol{h}^T,\boldsymbol{z}^T,\hat{\boldsymbol{y}},\gamma,\delta,\beta^s) \propto \prod_{y=1}^{K}\prod_{i:\hat{y}_i=y}\prod_{t=\underline{t}_i}^{\bar{t}_i-1} (p(h_{i,t+1}|h_{i,t},\beta^{h|s},y,X_i))^{\mathbb{I}_{h_{i,t+1}\neq D}}p(y|h_{\underline{t_i}},X_i,\beta,\delta),$$

Using flat priors and data augmentation, the posterior of β^s and $\beta^{h|s}$ is normally distributed (weighted probit model).

Block 2: health behaviors parameters

The second block is given by:

$$p(\gamma|\boldsymbol{h}^T,\boldsymbol{z}^T,\hat{y},X_i) \propto \prod_{y=1}^{K}\prod_{i:\hat{y}_i=y}\prod_{t=\underline{t}_i}^{\bar{t}_i} p(z_{i,t}|h_{i,t},\gamma,X_i,y)p(\gamma)$$

As in the previous block, using flat priors and data augmentation, the posterior of γ is normally distributed (probit model).

Block 3: mixture weights parameters

The third blocks is given by:

$$p(\delta|\beta,\gamma,\hat{y},\boldsymbol{h}^T,X_i) \propto \prod_{y=1}^{K}\prod_{i:\hat{y}_i=y} p(y|h_{\underline{t_i}},X_i,\beta,\delta)p(\delta)$$

we model $p(y|h_1,X_i,\delta)$ as a multinomial probit model and sample δ using a Metropolis algorithm with flat priors in $p(\delta)$

Block 4: latent lifestyles

The posterior distribution of the latent lifestyles is given by:

$$p(\hat{y}_i = y|\beta,\gamma,\delta,\boldsymbol{h}^T,\boldsymbol{z}^T,X_i) \propto p(\boldsymbol{h}^T|\beta,y,X_i)p(\boldsymbol{z}^T|\boldsymbol{h}^T,\gamma,y)p(y|h_{\underline{t_i}},X_i,\delta),$$

this is true for every $y = 1,\ldots,Y$. Thus, we can compute the actual probability as:

$$p(\hat{y}_i = y|\beta,\gamma,\delta,h^T,z^T,X_i) = \frac{p(\boldsymbol{h}^T|\beta,y,X_i)p(\boldsymbol{z}^T|\boldsymbol{h}^T,\gamma,y)p(y|h_{\underline{t_i}},X_i,\delta)}{\sum_{\tilde{y}=1}^{Y} p(\boldsymbol{h}^T|\beta,\tilde{y},X_i)p(\boldsymbol{z}^T|\boldsymbol{h}^T,\gamma,\tilde{y})p(\tilde{y}|h_{\underline{t_i}},X_i,\delta)}$$

Appendix B: Simulated Behaviors

In this appendix, we assess the ability of the econometric model to capture the variation in the incidence of different health behaviors observed in the data. For this purpose, we simulate data from the econometric model and compute average probabilities of reporting a given health behavior, the autocorrelation structure, and the cross-correlation with other health behaviors.

For the reader's convenience, Tables 9 and 13 correspond to Tables 1 and 2 in the main text. Tables 10, 11, and 12 report the same statistics as in Table 9 using simulated data when the econometric model is set with two, thee, and four groups, respectively. The first five columns in the tables show that the average incidence of the different health behaviors across education and age is perfectly captured for all the econometric models. The last two columns show that for the econometric model, it is harder to capture the autocorrelation structure in the data. Nevertheless, even the most parsimonious model is able to capture the large autocorrelation structure of smoking. In order to match the persistence over time, the econometric model needs a larger number of groups. The last two columns in Table 12 show correlations close to those in the data.

Tables 14, 15, and 16 report the cross-correlation of health behaviors when the econometric model is set with 2, 3, and 4 groups, respectively. Compared with Table 13, even the econometric model with just two groups is able to capture the cross-correlations well.

Table 9: Mean health behavior and 4-year auto-correlation: Data

	Mean					AC	
	HSD	HSG	CG	50-60	70-80	50-60	70-80
Drinking	0.07	0.08	0.07	0.13	0.05	0.48	0.42
Smoking	0.18	0.15	0.07	0.22	0.09	0.75	0.64
Cancer test	0.60	0.70	0.79	0.69	0.74	0.40	0.42
Cholesterol	0.74	0.81	0.86	0.75	0.85	0.33	0.29
Flu shot	0.57	0.60	0.66	0.46	0.74	0.55	0.58
Exercise	0.27	0.38	0.54	0.43	0.39	0.40	0.38

Notes: Data from the HRS. HSD: high-school dropout; HSG: high-school graduate; CG: college graduate; 50-60: sub-sample of individuals aged 50 to 60; 70-80: sub-sample of individuals aged 70 to 80. The last two columns show the autocorrelation (AC) of each health behavior with a 4-year lag.

Table 10: Mean health behavior and 4-year auto-correlation: 2 Groups

	Mean					AC	
	HSD	HSG	CG	50-60	70-80	50-60	70-80
Drinking	0.07	0.08	0.08	0.13	0.04	0.03	0.01
Smoking	0.18	0.15	0.07	0.21	0.09	0.67	0.34
Cancer test	0.66	0.70	0.73	0.69	0.72	0.02	0.05
Cholesterol	0.79	0.80	0.82	0.75	0.84	0.03	0.04
Flu shot	0.63	0.61	0.61	0.47	0.72	0.03	0.06
Exercise	0.35	0.40	0.44	0.43	0.38	0.05	0.02

Notes: Data from the HRS. HSD: high-school dropout; HSG: high-school graduate; CG: college graduate; 50-60: sub-sample of individuals aged 50 to 60; 70-80: sub-sample of individuals aged 70 to 80. The last two columns show the autocorrelation (AC) of each health behavior with a 4-year lag.

Table 11: Mean health behavior and 4-year auto-correlation: 3 Groups

	Mean					AC	
	HSD	HSG	CG	50-60	70-80	50-60	70-80
Drinking	0.08	0.08	0.07	0.13	0.05	0.07	0.02
Smoking	0.17	0.15	0.07	0.21	0.09	0.68	0.48
Cancer test	0.63	0.70	0.76	0.69	0.72	0.15	0.12
Cholesterol	0.77	0.80	0.84	0.75	0.84	0.13	0.09
Flu shot	0.58	0.61	0.65	0.47	0.73	0.17	0.27
Exercise	0.34	0.40	0.44	0.43	0.38	0.05	0.02

Notes: Data from the HRS. HSD: high-school dropout; HSG: high-school graduate; CG: college graduate; 50-60: sub-sample of individuals aged 50 to 60; 70-80: sub-sample of individuals aged 70 to 80. The last two columns show the autocorrelation (AC) of each health behavior with a 4-year lag.

Table 12: Mean health behavior and 4-year auto-correlation: 4 Groups

	Mean					AC	
	HSD	HSG	CG	50-60	70-80	50-60	70-80
Drinking	0.07	0.08	0.07	0.13	0.05	0.26	0.24
Smoking	0.17	0.15	0.07	0.21	0.09	0.68	0.48
Cancer test	0.62	0.70	0.76	0.69	0.73	0.14	0.13
Cholesterol	0.77	0.80	0.84	0.75	0.84	0.13	0.09
Flu shot	0.58	0.61	0.65	0.47	0.73	0.21	0.28
Exercise	0.34	0.40	0.44	0.43	0.38	0.05	0.02

Notes: Data from the HRS. HSD: high-school dropout; HSG: high-school graduate; CG: college graduate; 50-60: sub-sample of individuals aged 50 to 60; 70-80: sub-sample of individuals aged 70 to 80. The last two columns show the autocorrelation (AC) of each health behavior with a 4-year lag.

Table 13: Cross correlation health behaviors: data

	Drinking	**Smoking**	**Cancer test**	**Cholesterol**	**Flu shot**	**Exercise**
Drinking	1.00	0.09	-0.00	-0.01	-0.03	0.02
Smoking	0.16	1.00	-0.09	-0.09	-0.08	-0.07
Cancer test	-0.08	-0.15	1.00	0.30	0.20	0.09
Cholesterol	-0.06	-0.13	0.40	1.00	0.23	0.06
Flu shot	-0.05	-0.07	0.19	0.24	1.00	0.02
Exercise	-0.00	-0.13	0.08	0.04	0.02	1.00

Notes: Data from the HRS. Upper diagonal: individuals aged between 70 and 80. Lower diagonal: individuals aged between 50 and 60.

Table 14: Cross correlation health behaviors: 2 groups

	Drinking	**Smoking**	**Cancer test**	**Cholesterol**	**Flu shot**	**Exercise**
Drinking	1.00	0.08	-0.01	-0.03	-0.03	0.01
Smoking	0.15	1.00	-0.13	-0.12	-0.13	-0.06
Cancer test	-0.03	-0.15	1.00	0.04	0.03	0.02
Cholesterol	-0.03	-0.12	0.03	1.00	0.04	0.02
Flu shot	-0.01	-0.06	0.02	0.02	1.00	0.04
Exercise	-0.01	-0.14	0.02	0.04	0.01	1.00

Notes: Data from the HRS. Upper diagonal: individuals aged between 70 and 80. Lower diagonal: individuals aged between 50 and 60.

Table 15: Cross correlation health behaviors: 3 groups

	Drinking	**Smoking**	**Cancer test**	**Cholesterol**	**Flu shot**	**Exercise**
Drinking	1.00	0.08	-0.03	-0.04	-0.06	0.00
Smoking	0.13	1.00	-0.09	-0.06	-0.07	-0.07
Cancer test	-0.12	-0.13	1.00	0.11	0.17	0.04
Cholesterol	-0.11	-0.11	0.15	1.00	0.16	0.02
Flu shot	-0.10	-0.06	0.15	0.14	1.00	0.06
Exercise	-0.03	-0.13	0.05	0.06	0.03	1.00

Notes: Data from the HRS. Upper diagonal: individuals aged between 70 and 80. Lower diagonal: individuals aged between 50 and 60.

Table 16: Cross correlation health behaviors: 4 groups

	Drinking	**Smoking**	**Cancer test**	**Cholesterol**	**Flu shot**	**Exercise**
Drinking	1.00	0.07	0.01	0.00	-0.02	0.03
Smoking	0.14	1.00	-0.09	-0.07	-0.07	-0.07
Cancer test	-0.05	-0.14	1.00	0.11	0.18	0.04
Cholesterol	-0.04	-0.12	0.15	1.00	0.16	0.04
Flu shot	-0.04	-0.06	0.16	0.16	1.00	0.07
Exercise	-0.01	-0.14	0.06	0.07	0.04	1.00

Notes: Data from the HRS. Upper diagonal: individuals aged between 70 and 80. Lower diagonal: individuals aged between 50 and 60.

Appendix C: Obtaining the wealth distribution conditional on age, education, and type

We model the observed wealth distribution as a mixture model. In order to separate the mass point at zero wealth and the distribution of wealth conditional on positive wealth we proceed in two stages. First, we write the distribution of (positive) wealth conditional on observables as:

$$p(w_{i,t}|e_i, a_{it}, z_i^T, h_i^T, w_{i,t} > 0) = \sum_{y \in Y} p(w_{i,t}|y, e_i, a_{it}, z_i^T, h_i^T, w_{i,t} > 0)p(y|e_i, z_i^T, h_i^T),$$

The first term in the right hand side is the conditional probability of observing wealth $w_{i,t}$. We assume that wealth conditional on age a, education e, and type y is lognormally distributed, that is, $\log p(w|y, e, a, w > 0) \sim N(\mu^1(y, e, a), \sigma^1(y, e))$. This implies that we are imposing that given a, e, and y, wealth is independent from h_i^T and z_i^T.

The second term in the right hand side gives the conditional distribution of types, which we have estimated above, see Section 3.3. Hence, we only need to estimate $\mu^1(y, e, a)$ and $\sigma^1(y, e)$ for the sample of male individuals with positive asset holdings.[12]

Second, we can similarly write the probability of reporting zero (or negative) assets conditional on observables as

$$p(w_{it} = 0|e_i, a_{it}, z_i^T, h_i^T) = \sum_{y \in Y} p(w_{it} = 0|y, e_i, a_{it}, z_i^T, h_i^T)p(y|e_i, z_i^T, h_i^T),$$

where the first term on the right hand side is modelled as a probit, that is, it is given by $\Phi\left(w^*\left(y, e_i, a_{it}\right)\right)$, where the threshold $w^*\left(y, e_i, a_{it}\right)$ is modelled as a flexible low order polynomial on age. As above, we assume that this probability does not depends on h_i^T or z_i^T.

Appendix D: First step estimation details

D.1 Income process

The labor income process is modeled as the sum of a deterministic and stochastic component:

$$\begin{aligned}
&\log w_t^{\mathrm{c,e}}(h_t, \xi_t, \epsilon_t) = \omega_t^{\mathrm{c,e}}(h_t) + \xi_t + \epsilon \\
&\xi_{t+1} = \rho_\xi \xi_t + \nu_t, \nu_t \sim N(0, \sigma_\nu^2) \\
&\epsilon \sim N(0, \sigma_\epsilon^2) \\
&\xi_0 \sim (0, \sigma_{\xi,0}^2)
\end{aligned}$$

[12] We model $\mu^1(y, e, a)$ as a flexible low-order polynomial on age and $\sigma^1(y, e)$ non-parametrically.

We propose a Gibbs algorithm to compute the posterior distribution of all parameters using Bayesian methods.

1. Sample parameters of the deterministic component: multivariate normal.
2. Sample persistent shocks: Kalman smoother.
3. Sample persistent component parameters : normal posterior for ρ_{xi} and inverse gamma for σ^2_ν.
4. Sample initial distribution of shocks: Metropolis.
5. Sample variance of the transitory component: inverse gamma.

D.2 Medical shocks

To estimate the mean of health cost distribution $\mu^e_t(h_t)$, we run an OLS regression of log out-of-pocket expenditures in the last two years in HRS on a cubic in age, health, health interacted with age and education interacted with age and individual fixed effects. In order to compute the education fixed effects, we regress the residuals of the previous regression on education dummies. In order to estimate $\sigma^e_{\zeta,t}(h)$, we regress the squared residuals from the previous regression on a cubic in age, health, health interacted with age, education, and education interacted with age.

D.3 Tax system

We follow De Nardi et al. (2022) and set the Medicare and Social Security tax rates to 2.9% and 12.4%, respectively. We use the Social Security rules for 2018, and therefore we set the maximum taxable income for Social Security to w_{ss} =$113,700.

For the progressive tax labor income tax function, we follow Holter et al. (2019) estimates of the tax progressivity for families without children:

$$T(y) = y - 0.873964 \times y^{1-0.108002}$$

Appendix E: Extra figures and tables

Figure 11: Health transitions across education and lifestyles

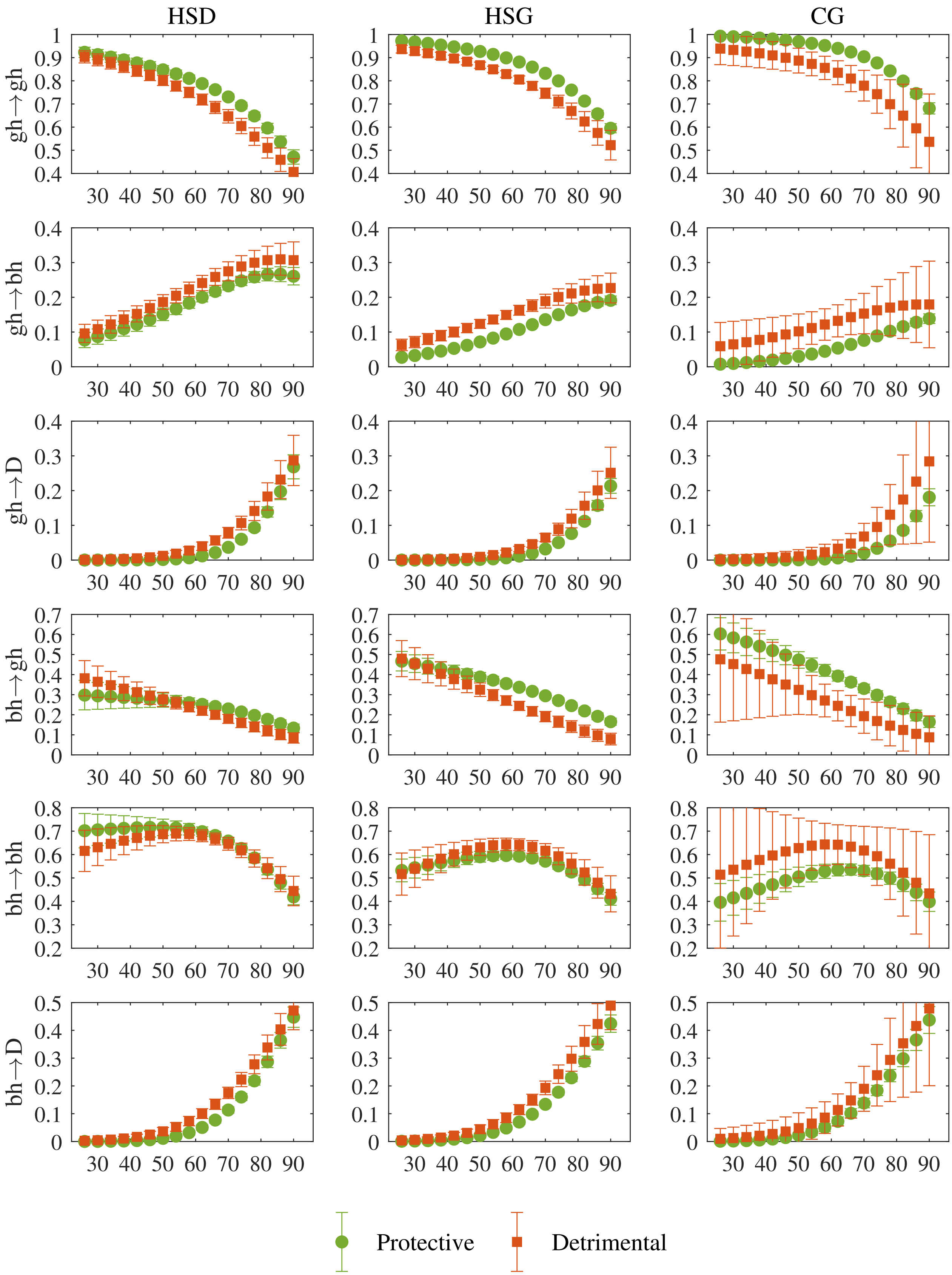
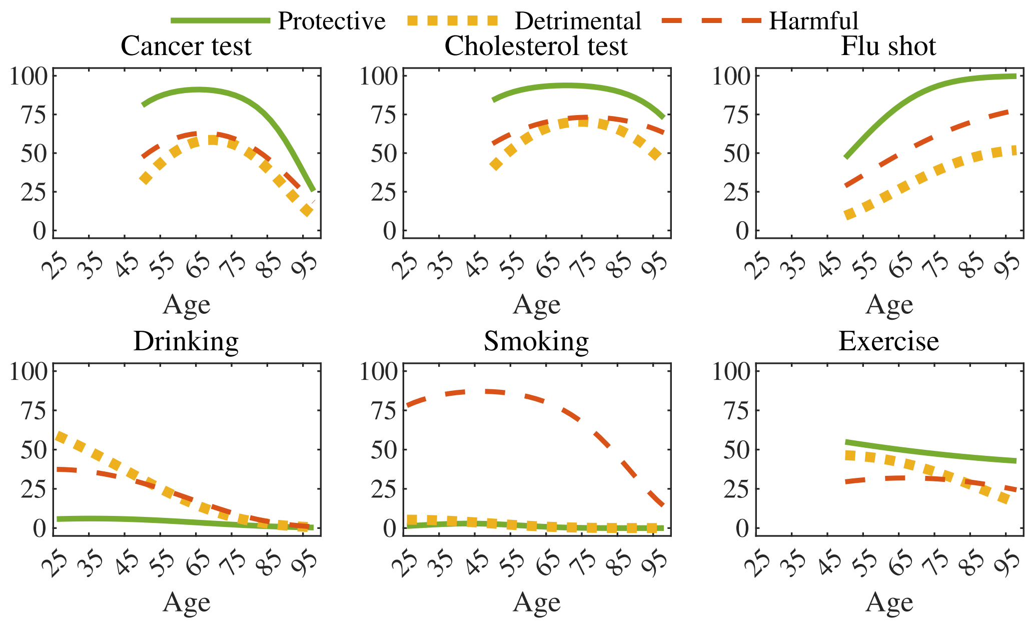
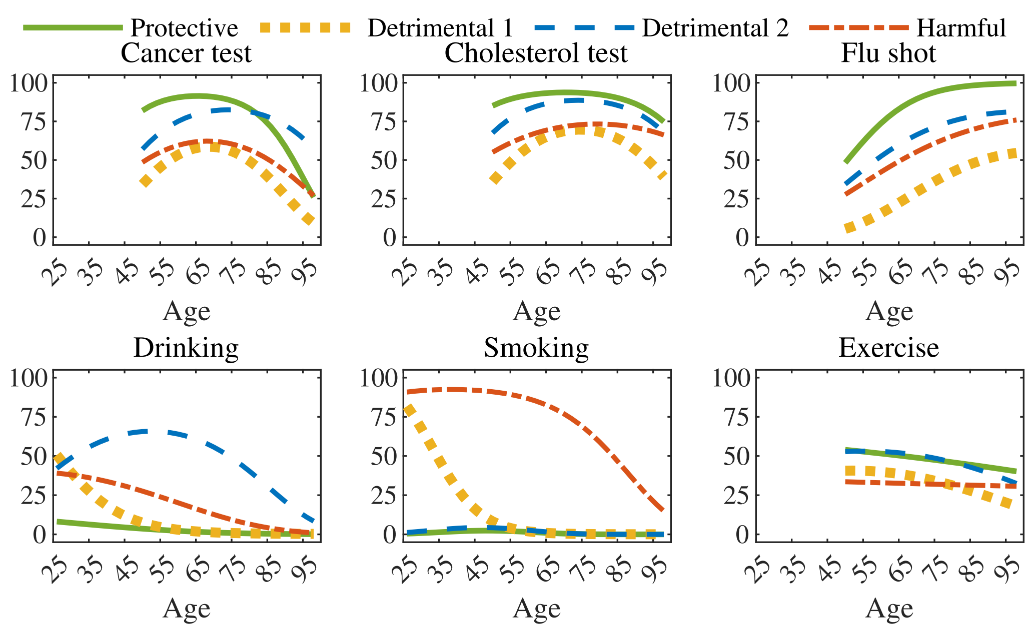


Figure 12: Health habits and health behavior types: 3 groups



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health.

Figure 13: Health habits and health behavior types: 4 groups



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health.