DISCUSSION PAPER SERIES

DP19990
(v. 2)

EDUCATION, LIFESTYLES, AND HEALTH INEQUALITY

Jesus Bueren, Josep Pijoan-Mas and Dante Amengual

LABOUR ECONOMICS AND
MACROECONOMICS AND GROWTH

CEPR

ISSN 0265-8003

EDUCATION, LIFESTYLES, AND HEALTH INEQUALITY

Jesus Bueren, Josep Pijoan-Mas and Dante Amengual

Discussion Paper DP19990
First Published 04 March 2025
This Revision 12 July 2025

Centre for Economic Policy Research
187 boulevard Saint-Germain, 75007 Paris, France
2 Coldbath Square, London EC1R 5HL
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Labour Economics
- Macroeconomics and Growth

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Jesus Bueren, Josep Pijoan-Mas and Dante Amengual

EDUCATION, LIFESTYLES, AND HEALTH INEQUALITY

Abstract

We study the effect of lifestyles on the education gradient of life expectancy. We use panel data on health behavior and health outcomes to estimate latent lifestyle types and their impact on health dynamics. We find that the higher frequency of health-protective lifestyles among the more educated individuals explains almost 1/2 of the education gradient in life expectancy. To understand lifestyle formation, we build a life cycle model where lifestyles and education are jointly chosen early in life. These two investments are complementary because of the more educated's higher income and the higher yield of their health-protective behavior. Importantly, with these complementarities, individuals with lower costs of healthier lifestyles self-select into higher education. Quantitatively, we find the three mechanisms similarly important in explaining the correlation between education and healthy lifestyles. We also find that the increase in the college wage premium over the last decades has widened the education gradient in lifestyles, resulting in a 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 and 60% by the induced changes in the composition of college graduates and high school dropouts.

JEL Classification: E21, D15, I12, I14, I24, C38

Keywords: Health inequality, Education, Healthy lifestyle, Latent factor

Jesus Bueren - jesus.bueren@gmail.com
EUI

Josep Pijoan-Mas - pijoan@cemfi.es
CEMFI and CEPR

Dante Amengual - amengual@cemfi.es
CEMFI

Acknowledgements
The authors thank for their valuable feedback Alexander Bick, Raquel Fonseca, Elena Manresa, Pierre-Carl Michaud, Pedro Mira, Mariacristina de Nardi, Florian Pelgrin, and Eugenio Zucchelli. Josep Pijoan-Mas acknowledges financial support from Grant PID2022-142790NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU, and by Grant PRX23/00630 funded by MICIU/AEI/10.13039/501100011033, which has financed the visit to Princeton University where part of this research took place.

Education, Lifestyles, and Health Inequality*

Jesús Bueren
Católica-Lisbon

Josep Pijoan-Mas
CEMFI and CEPR

Dante Amengual
CEMFI

September 2026

Abstract

We study the effect of lifestyles on the education gradient in life expectancy. We use panel data on health behaviors and health outcomes to estimate latent lifestyle types and their effects on health dynamics. We find that the higher prevalence of health-protective lifestyles among more educated individuals explains 42% of the education gradient in life expectancy. To understand lifestyle formation, we build a life-cycle model in which lifestyles and education are jointly chosen early in life. These investments are complementary because the higher expected income of more educated individuals increases the value of adopting a health-protective lifestyle. This complementarity induces individuals with lower costs of healthier lifestyles to self-select into higher education, while a primitive correlation between education and lifestyle costs reinforces this sorting. Quantitatively, both the direct income channel and selection contribute substantially to the correlation between education and healthy lifestyles, although selection plays the larger role. We also find that widening labor-earnings inequality across education groups increases the education gradient in life expectancy between the 1930s and 1970s cohorts by about one year, with one-third of the increase driven by the direct effect of changing income prospects and two-thirds by induced changes in the composition of college graduates and high school dropouts.

JEL classification: E21, D15, I12, I14, I24, C38
Keywords: Health Inequality, Healthy Lifestyles, Education, Latent Types

*The authors thank for their valuable feedback Alexander Bick, Raquel Fonseca, Lukas Mahler, Elena Manresa, Pierre-Carl Michaud, Pedro Mira, Mariacristina de Nardi, Florian Pelgrin, Eugenio Zucchelli and the participants from presentations held at Bristol University, Católica Lisbon, CEMFI, CUNEF, EUI, HEC Montreal, Princeton University, St Louis FED, Stony Brook University, UAB, UQAM, the 2024 BSE Summer Forum in Barcelona, and the 2024 EEA Annual Meetings in Rotterdam. Emails: jbueren@ucp.pt, pijoan@cemfi.es, amengual@cemfi.es. Josep Pijoan-Mas acknowledges financial support from Grant PID2022-142790NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU, and from the Mobility Program of the MICIU, which has financed the visit to Princeton University where part of this research took place.

1 Introduction

The US economy has 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 gradient 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 health behaviors and lifestyle factors.

Health behaviors —like exercise habits, dietary patterns, or substance abuse— are important determinants of health outcomes. McGinnis and Foege (1993)'s 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. Since then, many papers have shown that individuals who engage in healthier behaviors are more likely to experience positive health outcomes and lower risk of mortality (e.g. Taylor et al., 2002; Li et al., 2018; Zaninotto et al., 2020). This literature has also consistently shown that individuals with higher levels of education tend to adopt healthier lifestyles (e.g. Lantz et al., 1998; Cutler and Lleras-Muney, 2010).

Our paper tackles three main questions. First, we measure the impact of different lifestyles —understood as the collection of different health behaviors— on health dynamics and health inequalities. Second, we study the joint determination of education and lifestyles early in life, the role of selection, and why there is an education gradient in lifestyles. Finally, we quantify the effect of rising labor earnings inequality across education groups on the observed increase in the education gradients of lifestyles and health outcomes.

To address the first question, we use data from the Health and Retirement Study (HRS) and the Panel Study of Income Dynamics (PSID), which contain a rich array of health behavior indicators —including preventive tests, substance abuse, and exercise habits. The use of these data presents three challenges. First, the observed health behaviors are imperfectly correlated across individuals and over time, which suggests the existence of a noisy relationship between individual incentives and different health behaviors. 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 are likely to appear in the long run. Finally, it is unfeasible to consider so many different variables in a dynamic model that tries to understand the determinants of health behaviors. To address these issues, we make a contribution by developing a novel methodology to reduce the dimensionality of the data. In particular, we identify patterns in health behavior by assigning individuals to permanent types based on both their health behaviors and their health dynamics. These types, which we call lifestyles, summarize the propensity of individuals to engage in different health behaviors over time as well as the resulting health trajectories. Consequently, our novel econometric procedure allows us to estimate the long-run effect of lifestyles on health

outcomes, which is a key ingredient for our model of endogenous health dynamics.

We consider a parsimonious representation of health behavior consisting of two lifestyles, which we label *protective* and *detrimental*. This representation delivers four main results. First, lifestyles have a strong association with health dynamics: at age 50, males classified as *protective* live, on average, 9.2 years longer than those classified as *detrimental*. Second, lifestyle choices vary substantially with education, with *protective* individuals being more frequent among the more educated. We find that differences in lifestyle composition across education groups account for 42% of the education gradient in life expectancy. Third, life expectancy differences between lifestyles remain large within every education group: 8.4 years among high school dropouts, 7.9 years among high school graduates, and 8.9 years among college graduates. Finally, the education gradient in lifestyles increases across cohorts, as the prevalence of *protective* lifestyles declines among the least educated and increases modestly among the most educated.

To understand the joint determination of education and lifestyles, our second question, we develop a heterogeneous-agents model with two stages. Early in life, individuals and their families simultaneously choose education and lifestyle. In this stage, individuals exhibit heterogeneity in both their utility costs associated with education and their utility costs associated with acquiring a health-protective lifestyle. These costs may be correlated, which would capture, in reduced form, common early-life factors that may influence both investments. In the second stage, during working life and retirement, individuals make consumption and saving decisions in a standard life-cycle model with idiosyncratic income and health risks, conditional on the education and lifestyle choices made in the first stage.

The model features a complementarity between education and lifestyle choices that is standard in models of education and health investments (e.g. Becker, 2007). This means that the value of a healthy lifestyle is larger for the more educated, and the value of education is larger for individuals expecting to live more years. This complementarity generates endogenous selection: even when the costs of the two investments are independently distributed in the population, individuals with lower costs of adopting a *protective* lifestyle are more likely to acquire more education. Correlated cost heterogeneity reinforces this mechanism. Consequently, individuals with low costs of adopting a *protective* lifestyle are overrepresented among the more educated.

We feed the model with earnings, health, and medical expense dynamics estimated from the data, and we calibrate the adult-life stage to match the wealth distribution by age, education, and lifestyle. The resulting value of adopting a *protective* rather than a *detrimental* lifestyle is large and increases with education, which gives a measure of the complementarity between the two investments. We then calibrate the early-life stage to match the joint distribution of education and lifestyles for the 1930s and 1970s cohorts, which mainly differ in their education-specific life-cycle earnings profiles.

Next, we quantify the relative importance of income and selection in driving the education gradient in lifestyles for the 1970s cohort. We find that the income advantage of the more educated

accounts for 41% of the education gradient in lifestyles and 1.3 years of the education gradient in life expectancy. Selection of individuals with lower costs of adopting a *protective* lifestyle into higher education—arising from both complementarity and correlated cost heterogeneity—accounts for 77% of the lifestyle gradient and 2.6 years of the life expectancy gradient. These contributions are not additive because income and selection interact.

We next use the model to compare policy interventions operating on each of the two investment margins. First, we look at interventions that reduce the utility cost of adopting a healthy lifestyle. This could be exposure to a healthier menu in primary school, an increase in sports activities in childhood, or a reduction in smoking among parents. We find that a one-half-standard-deviation reduction in the average cost of adopting a *protective* lifestyle increases its aggregate prevalence from 73% to 86% and increases average life expectancy by 1.1 years. The policy has asymmetric effects across population groups, as the improvement in lifestyle is concentrated among the less educated before the reform: this reduces the life expectancy gap between *ex ante* college graduates and *ex ante* high school dropouts by 1.2 years. Furthermore, because education and lifestyles are jointly chosen, the intervention also raises the educational attainment of many of the individuals who improve their lifestyle: 27.6% of the high school dropouts who improve their lifestyle also complete high school, and 7.6% of the high school graduates who improve their lifestyle also complete college. This changes the selection of individuals into different education groups, which attenuates the improvement in the observed educational gradient of life expectancy to 1.0 years (instead of 1.2). Second, we look at interventions that reduce the utility cost of studying. Reducing the costs of education by the same standardized amount reduces the share of high school dropouts from 10% to 4% and raises the college share from 36% to 55%. At the same time, due to the complementarity in early life investments, the policy slightly increases the share of *protective* individuals, which, in addition to the educational improvement, raises average life expectancy by 0.8 years. As before, the individual lifestyle and education gains are higher at the lower end of the distribution, which reduces the education gradient of life expectancy when measured across the benchmark education groups. However, due to selection, the policy widens the observed (or *ex post*) educational gradient of life expectancy.

Finally, we study how rising labor-earnings inequality across education groups has shaped the increase in the education gradient in life expectancy between the 1930s and 1970s cohorts. The direct effect of changing income prospects increases the education gradient in life expectancy by 0.4 years. Changes in the composition of education groups generate an additional increase of 0.8 years, as the smaller group of high school dropouts in the 1970s cohort contains a larger fraction of individuals for whom adopting a *protective* lifestyle is costly. Overall, the calibrated model generates a 1.2-year increase in the education gradient in life expectancy, compared with an increase of 1.0 year in the data. Thus, endogenous selection accounts for approximately two-thirds of the model-generated increase, while the direct effect of changing income prospects accounts for the remaining one-third.

1.1 Related literature

Our estimation of latent types in health behavior relates to the recent literature estimating unobserved fixed effects on health dynamics. Hosseini et al. (2022) and De Nardi et al. (2025) were the first to show that unobserved fixed effects are needed to explain long-run health dynamics. Contemporaneous work by Borella et al. (2025) estimates unobserved fixed effects by k-means clustering of health trajectories. In contrast, we classify individuals into types using both health behavior data and observed health trajectories. By incorporating health behavior data, our approach offers a clearer interpretation of these fixed effects, helping to separate chance events from more persistent health traits. Our econometric framework allows us to deal with the selection of survivors and, hence, exploit longer trajectories of health data for clustering, which in turn helps disentangle persistent shocks from fixed effects. Finally, our economic model incorporates type formation, which allows us to study the correlation between the unobserved fixed effects and socioeconomic 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), Bueren (2023), or Nakajima and Telyukova (2025)— and welfare —see Capatina (2015), Braun et al. (2019), De Nardi et al. (2025), or Hosseini et al. (2026). 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 lifestyles by individuals making different education choices.

The paper is also related to previous work featuring endogenous health dynamics. Fonseca et al. (2021) considers monetary health investments, but struggles to find a strong causal effect of medical spending on health outcomes. Ozkan (2025) does find a stronger role for medical spending, but mostly due to the preventive health care component, which aligns well with our notion of lifestyle. Capatina and Keane (2026) finds that expanding public health insurance improves health and longevity, but the empirical evidence on the effect of medical spending on health outcomes in the US is mixed.[1] 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 (2024). This justifies our focus on lifestyle choices as a potential driver of the educational gradient of health outcomes. Health behavior choices have been previously studied by Cole et al. (2019), Galama and van Kippersluis (2019), Margaris and Wallenius (2023), Mahler and Yum (2024), and Bairoliya et al. (2026). These models allow agents to adjust their behavior over the life cycle, but education choices and the costs of good health behavior in each education group are exogenously given. Instead, our model allows for the sorting of individuals with different costs of health behavior into different education

[1] Quasi-experimental evidence extending the use of Medicaid to low-income households, like Aron-Dine et al. (2013), Finkelstein et al. (2012), and Baicker et al. (2013), finds very small effects. Instead, Card et al. (2009) find that access to Medicare at the 65-year-of-age discontinuity leads to better health outcomes for emergency room patients.

groups and leverages on the persistence of health behavior —often established early in life— to model health and education investments jointly. This allows us to study type formation and assess the extent to which selection drives differences in health outcomes across education groups. This makes our approach more similar to Hai and Heckman (2026), who study the interaction between smoking, preferences, and education.

Our results on the long-run changes in health inequalities connect 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. This is consistent with Novosad et al. (2022), who find that the worsening of the health outcomes of non-college-educated males is partly due to selection.

2 Data on health behavior

We use data from male individuals from the PSID (1999 to 2019) and the HRS (1996 to 2018). Our data on health behavior collect binary information on whether individuals engage in heavy drinking (reporting more than two alcoholic beverages every day), whether they smoke, whether they have taken a preventive cancer test (prostate cancer screening), 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). We sacrifice some information by restricting to binary variables, but this facilitates the estimation of our complex econometric model. 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, except for drinking, there is a clear education gradient on all other 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 55 years of age, the 4-year autocorrelations of heavy drinking or obtaining a flu shot are 0.44 and 0.53, respectively (see Column 6 in Table 1). This weak persistence is partly the result of a life-cycle pattern, as the incidence of health related behaviors changes with age. For instance, the incidence

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

	Mean					AC	
	HSD	HSG	CG	50-55	70-75	50-55	70-75
Cancer test	0.59	0.69	0.77	0.57	0.78	0.35	0.36
Cholesterol	0.72	0.79	0.86	0.71	0.86	0.37	0.31
Flu shot	0.55	0.58	0.66	0.40	0.73	0.53	0.56
Drinking	0.03	0.04	0.03	0.04	0.04	0.44	0.47
Smoking	0.21	0.17	0.07	0.26	0.11	0.75	0.64
Exercise	0.28	0.41	0.58	0.45	0.44	0.43	0.38

Notes: Data from the HRS. HSD: high school dropout; HSG: high school graduate; CG: college graduate; 50-55: sub-sample of male individuals aged 50 to 55; 70-75: sub-sample of male individuals aged 70 to 75. 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	0.07	-0.02	-0.03	-0.06	-0.01
Smoking	0.11	1.00	-0.12	-0.12	-0.09	-0.09
Cancer test	-0.04	-0.14	1.00	0.38	0.23	0.11
Cholesterol	-0.04	-0.17	0.49	1.00	0.25	0.08
Flu shot	-0.04	-0.07	0.18	0.25	1.00	0.05
Exercise	-0.03	-0.13	0.10	0.03	0.04	1.00

Notes: Data from the HRS. Lower diagonal: individuals aged 50 to 55. Upper diagonal: individuals aged 70 to 75.

of smoking falls from 0.26 of the population aged 50 to 55 to 0.11 of the population aged 70 to 75, while the incidence of flu shots rises from 0.40 to 0.73 (see Columns 4 and 5 in Table 1). And third, these different measures of health behavior are correlated across individuals as one would expect, although the correlations are relatively small (see Table 2). For instance, the correlations across individuals aged 50 to 55 of cholesterol tests with flu shots or cancer tests are fairly large, 0.25 and 0.49, respectively, while the correlations of cholesterol tests with drinking, smoking, and exercise habits are much smaller, -0.04, -0.17, and 0.03, respectively.[2]

The positive but low correlation of different good health behaviors over time and across individuals suggests the need for a latent 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). Ultimately, 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. We present such a factor model in the next Section.

[2]The correlation between healthy behavior and education has been documented before by Lantz et al. (1998) and Cutler and Lleras-Muney (2010), while the imperfect correlation of health behaviors in the cross-section has been highlighted by Cutler and Glaeser (2005) and Bairoliya et al. (2026).

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 $t = 0, \ldots, T$ periods (with T potentially varying across individuals). For each individual and period, we observe standard demographic variables: cohort of birth $\mathrm{c}_i \in \{c_{10}, c_{30}, c_{50}, c_{70}, c_{90}\}$ (individuals born between 1900 and 1920, 1920 and 1940, *etc.*), education $\mathrm{e}_i \in \{\text{HSD}, \text{HSG}, \text{CG}\}$ (high school dropout, high school graduate, college graduate), and age at first interview $a_{i0} \in \mathbb{N}_+$, plus a wide array of health-related variables, which we classify into two groups: health outcomes and health behavior. Finally, for each period we also observe the individual's insurance status, denoted $\text{ins}_{it} \in \{0, 1\}$. 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 obviously an absorbing state. We build this variable using the 5-category self-rated health variable (where the best three 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 information on N_z different binary variables $z_{m,it} \in \{0, 1\}$ describing whether individual i in period t takes some particular action m. These actions are the health behavior variables described in Section 2.[3]

We next assume that both the observed health behavior $\mathbf{z}_{it}$ and health outcome h_{it} depend on some unobserved time-invariant latent variable $\mathrm{y}_i \in Y \equiv \{y_1, y_2, \ldots, y_{N_y}\}$. We interpret the latent variable as the individual lifestyle, which captures the idea that individuals differ in their propensity to undertake actions that are good for their health (see Section 5.1 for a discussion on the assumption of lifestyles being fixed).

We model the joint sequence $\left(\mathbf{z}_i^T, \mathbf{h}_i^T\right)$ of health behaviors and health outcomes for each individual i, conditional on the vector $x_i \equiv (\mathrm{c}_i, \mathrm{e}_i, a_{i0}, h_{i0})$ of demographic characteristics —cohort c_i, education e_i, initial age a_{i0}, and initial health h_{i0}— and on the sequence of insurance status $\mathbf{ins}_i^T$, as a mixture of distributions conditional on unobserved types y_i. In particular, we can write the likelihood of the data as:

$$p\left(\mathbf{z}_i^T, \mathbf{h}_i^T \,|\, \mathbf{ins}_i^T, x_i, \boldsymbol{\theta}\right) = \sum_{y=1}^{N_y} p\left(\mathbf{z}_i^T \,|\, \mathbf{h}_i^T, \mathbf{ins}_i^T, x_i, y; \boldsymbol{\theta}_y\right) \times p\left(\mathbf{h}_i^T \,|\, \mathbf{ins}_i^T, x_i, y; \boldsymbol{\theta}_y\right) \times p\left(y \,|\, \mathbf{ins}_i^T, x_i; \boldsymbol{\theta}\right) \quad (1)$$

where $\boldsymbol{\theta}$ is the vector of parameters, $\boldsymbol{\theta}_y$ is the subset of parameters relevant to type y, and where the joint likelihood is factorized by the product of the conditional and marginal distributions. The three elements on the right-hand side of equation (1) are the probability of observing a sequence of health behaviors, the probability of observing a sequence of health outcomes, and the initial distribution of types. We explain them in more detail in Section 3.1, 3.2, and 3.3, respectively. Importantly, estimation of equation (1) implies that both observed health behavior and observed

[3] 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 preventive health behavior (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.

health transitions help inform the classification of individuals into types.[4]

3.1 Health behavior

To model the probability $p\left(\mathbf{z}_i^T \,|\, \mathbf{h}_i^T, \mathbf{ins}_i^T, x_i, y; \boldsymbol{\theta}_y\right)$ of observing a sequence $\mathbf{z}_i^T$ of health behaviors we make the following assumptions.

First, we assume that the probability of individual i in period t reporting the m^{th} behavior ($z_{m,it} = 1$) depends on the health behavior type y_i but also on age a_{it}, current health status h_{it}, and current health insurance status ins_{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. For behaviors that involve medical utilization —flu shots, cholesterol tests, and cancer screenings— an individual's current insurance status ins_{it} accounts for the possibility that some individuals have a lower propensity to undertake these actions because they are not covered.

Second, conditional on these variables, we impose that health behavior does not depend on cohort c_i and education e_i. We do so because we want the type definitions to be consistent across all demographic groups. That is, a type is defined by the propensity to engage in a series of behaviors conditional on age, health, and insurance status, hence letting the variation in behavior across demographic groups arise from the different distribution of types y.

And third, we assume that conditional on y_i, h_{it}, a_{it}, and ins_{it}, health behaviors are independent between them and over time, so that we can model the probability of observing a history of health behaviors as follows:

$$p\left(\mathbf{z}_i^T \,|\, \mathbf{h}_i^T, \mathbf{ins}_i^T, x_i, y; \boldsymbol{\theta}_y\right) = \prod_{t=0}^{T} \prod_{m=1}^{N_z} p(z_{m,it} | h_{it}, \mathrm{ins}_{it}^{(m)}, a_{it}, y; \boldsymbol{\theta}_y) \tag{2}$$

where $\mathrm{ins}_{it}^{(m)} = \mathrm{ins}_{it}$ for medical-utilization behaviors and is omitted for smoking, drinking, and exercise. We model each element $p(z_{m,it} | h_{it}, \mathrm{ins}_{it}^{(m)}, a_{it}, y; \boldsymbol{\theta}_y)$ as a very flexible probit, see Appendix A.1 for details.

Our substantive restriction is that, conditional on lifestyle type y, current health h_{it}, age a_{it}, and insurance status ins_{it}, we assume that behavior $z_{m,it}$ is independent of health realizations in other periods (past and future). Consequently, once lifestyle type and the current state are held fixed, realized behavior is unrelated to subsequent health transitions. Observed behaviors nevertheless predict future health because they provide information about the individual's permanent

[4] As an alternative, one might think of only using health behavior data to classify individuals into types, and estimate health dynamics conditional on type in a second stage. We do not do so for two reasons. First, the joint use of health behavior and health dynamics data allows us to estimate the third term in equation (1), which controls the selection of better types into survival. That is, as we show in Section 3.3, this term absorbs the different frequencies of good and bad types over age due to differential mortality. Second, we want the classification of individuals based on behavior to be meaningful in terms of health dynamics; that is, we want to avoid grouping behaviors with very different impacts on health dynamics.

lifestyle type, which affects health dynamics, as discussed in Section 3.2 below.

3.2 Health dynamics

To model the probability $p\left(\mathbf{h}_i^T \,|\, \mathbf{ins}_i^T, x_i, y; \boldsymbol{\theta}_y\right)$ of health outcomes, we assume that health dynamics for individual i depends on education e_i, health behavior type y_i, and age a_{it}, but not on cohort c_i. The dependence of health dynamics on education captures differences in health outcomes that are not associated with differences in health behavior types across education groups. The absence of cohort c_i from the set of conditioning variables is an identification assumption: since we do not observe full lifespans for individuals in different cohorts, 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 the estimation of models of health dynamics with survey data, see, for instance, Pijoan-Mas and Ríos-Rull (2014). Hence, our assumption is that the health dynamics of individuals of a given education are, conditional on type, identical across cohorts. Nevertheless, it is important to note that this allows health dynamics by education to differ across cohorts due to differences in type composition.

Finally, we assume that, conditional on education e_i, health behavior type $\mathrm{y}_i = y$, and age a_{it}, the evolution of health outcomes is Markovian, that is, it only depends on one lag of health outcomes:

$$p\left(\mathbf{h}_i^T \,|\, \mathbf{ins}_i^T, x_i, y; \boldsymbol{\theta}_y\right) = \prod_{t=0}^{T-1} p(h_{it+1}|\mathrm{e}_i, a_{it}, h_{i,t}, y; \boldsymbol{\theta}_y) \tag{3}$$

The specific functional form for $p(h_{it+1}|\mathrm{e}_i, a_{it}, h_{i,t}, y; \boldsymbol{\theta}_y)$ is given in Appendix A.2.

Given the assumptions here and in Section 3.1 above, it is clear that the model estimates the relationship between persistent lifestyle patterns and long-run health trajectories rather than the causal effect of any single observed behavior on future health. This conditional-independence restriction is the main simplifying assumption underlying our dimension-reduction exercise.[5]

3.3 Distribution of health types

The final element in the likelihood function is the fraction of individuals of each type y in each demographic group, that is, the mixing weights. In particular, we define $p\left(y|\, \mathbf{ins}_i^T, x_i; \boldsymbol{\theta}\right)$ as the probability that individual i born into cohort c_i, education e_i, first observed at age a_{i0} and health h_{i0} is of type y. In stationary mixture models, it is common to add this term to the likelihood function and estimate it non-parametrically (Dempster et al., 1977). However, in our model with

[5]Estimating the causal effect of each behavior on health transitions is difficult because health behaviors are imperfectly correlated across individuals and over time and because their consequences may emerge only after many years. Our approach therefore focuses on the long-run health trajectories associated with holistic lifestyle patterns rather than on the effect of any behavior in isolation.

health dynamics that are heterogeneous by lifestyle types y, the mixing weights will be different across initial age a_{i0} and health h_{i0}, as well as across $(\mathrm{c}_i, \mathrm{e}_i)$ groups. To avoid estimating different mixing weights for each potential different entry age a_{i0}, we exploit the model of health dynamics in equation (3) to obtain a recursive expression describing the distribution of types conditional on observables at each age. This allows us to estimate only the initial distribution of types at an arbitrary initial age, which we choose to be 26: that is, we only need to estimate the fraction $p\left(\mathrm{y}|\mathrm{c}, \mathrm{e}, 26, h; \boldsymbol{\theta}_y\right)$ of individuals of age 26, cohort c, and education e, in health state h, who are of type y. The specific functional form for this probability, as well as the formal description of the above argument, are given in Appendix A.3.

3.4 Estimation

We need to estimate a large vector of parameters $\boldsymbol{\theta}$ plus the probability of each individual of belonging to each latent type. We resort to Bayesian methods, which are particularly well-suited for models with latent structures. Specifically, we implement a Gibbs sampler to break down the high-dimensional likelihood into smaller, simpler conditional distributions, which significantly reduces the computational complexity of the estimation process, see Appendix A.4 for details.

4 Results from the econometric model

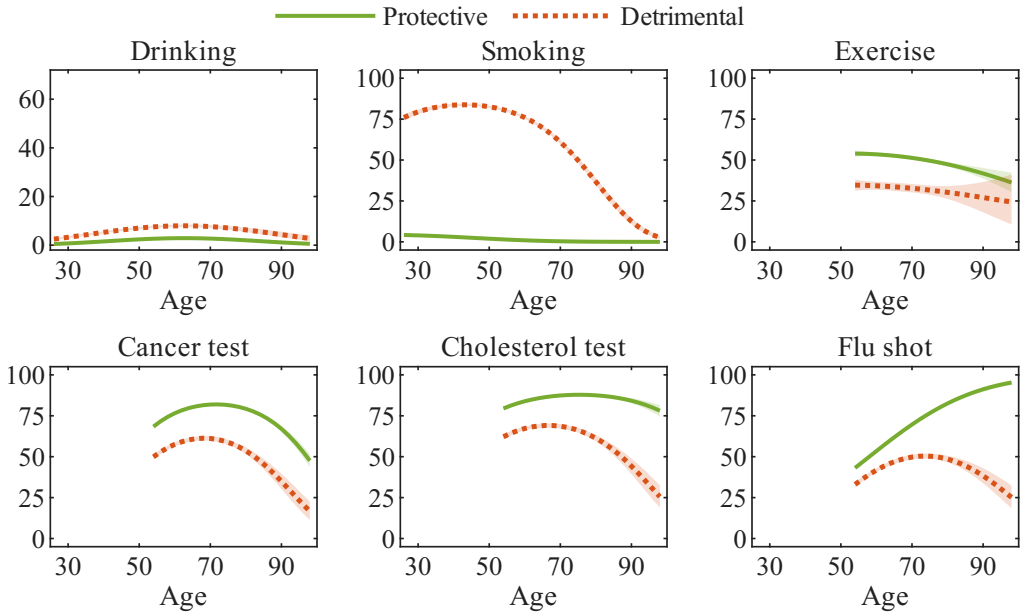
We estimate the model using our panel data discussed in Section 2. The results presented below are for males only, which includes 17,160 individuals (interviewed an average of 5.2 times each) in the HRS and 9,503 individuals (interviewed an average of 6.7 times) in the PSID. In favor of parsimony, we choose $N_y = 2$ for our main estimation and discuss the cases with $N_y > 2$ in Appendix B.

4.1 Health behavior

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 $h_{it} = h_g$.[6] Individuals in the group that we label *protective* (solid green line), have a higher likelihood of reporting health-enhancing behaviors (cancer test, cholesterol test, flu shot, and exercise) and a lower probability of reporting harmful health behaviors (smoking and drinking). For individuals in the other group, which we label *detrimental* (red dashed line), the probability of reporting all the health-enhancing behaviors 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 around age 50 but, as they age, individuals classified as *protective* increase their probability of getting a flu shot while individuals classified as *detrimental* do not.

[6] Figure F.2 and F.3 in Appendix F show how behavior changes with health and insurance status.

FIGURE 1: Health behaviors and lifestyles



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for insured male individuals in good health. The shaded areas represent the 95% credible intervals.

4.2 Health dynamics

Next, we examine the relationship between lifestyles and health dynamics. We summarize this information by looking at the gradient in age-50 life expectancy across types.[7] In Table 3 (second column) we show that, on average, *protective* types live for 9.2 more years than *detrimental* types. The effect of lifestyle on health dynamics is also large within each education category: *protective* types live 8.4 more years than *detrimental* types among males without a high school degree, 7.9 more years among males with a high school degree, and 8.9 more years among males with a college degree (see columns 4, 6, and 8).

4.3 Distribution of health types

Our estimation assigns a different fraction $p\left(\mathrm{y}|\mathrm{c},\mathrm{e},26,h;\boldsymbol{\theta}_y\right)$ of individuals to each type y depending on cohort c and education e. For the 1970s cohort, 73.5% of males are classified as *protective* and 26.5% as *detrimental* (see first column in Table 3). Importantly, there is a large educational gradient of lifestyles: the share of *protective* individuals grows from 50.4% to 89.5% as we move from high school dropout to college graduate (see columns 3, 5, and 7), which reflects a strong correlation between education and lifestyle types. In a sense, this correlation is not too surprising: it is well known that the incidence of smoking declines with education, and Table 1 before shows how in our data the share of individuals taking cancer tests, cholesterol tests, and flu shots also increases

[7] For more details, Figure F.1 in the Appendix shows the actual health transitions and survival functions by lifestyle.

TABLE 3: Life expectancy 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	28.1	100.0	23.6	100.0	26.5	100.0	31.5	7.8	4.5	3.3
PRO	73.5	30.5	50.4	27.8	66.6	29.2	89.5	32.4			
DET	26.5	21.3	49.6	19.4	33.4	21.3	10.5	23.5			
Δ_y	47.1	9.2	0.8	8.4	33.2	7.9	79.0	8.9			

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 education gradient in LE when the distribution of behavior types within high school dropouts is the same as for college-educated individuals. Column (b) is the difference between the actual education gradient in LE and the counterfactual in column (a), that is, it corresponds to the gradient explained by differences in lifestyles across education groups for given health dynamics.

with education. It is therefore natural that our classification of individuals into types according to the observed health-related behavior retains the education gradient of these latter variables. Nevertheless, 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.

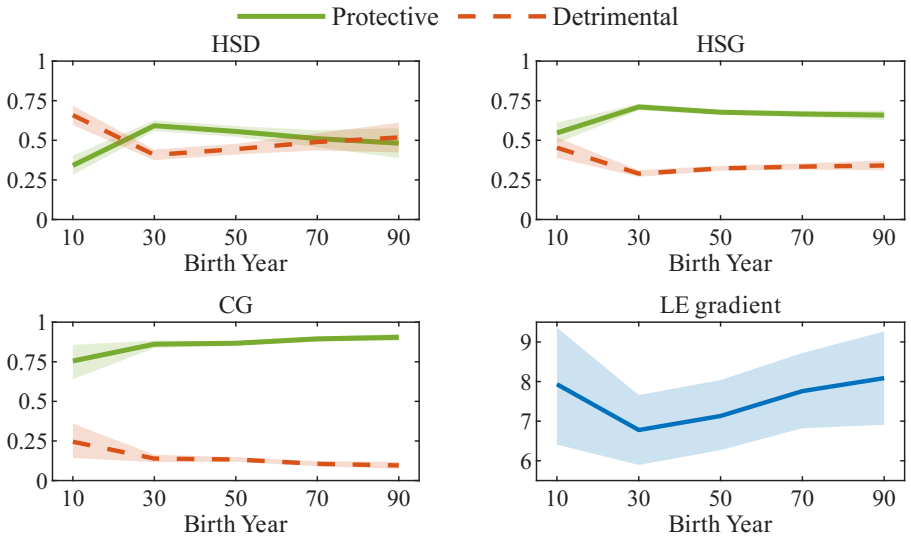
4.4 Evolution over time

Next, we look at the evolution over time of the joint distribution of health types and education. In Panels (a) to (c) of Figure 2 we report the distribution of types by education group from the 1910s to the 1990s cohorts. Within the high school dropouts, the *detrimental* types increased monotonically from 40% in the 1930s cohort to 52% in the 1990s cohort. This implies a severe deterioration in the lifestyle of individuals in the least educated group, which reverses the improvements in the type distribution between the 1910s and 1930s 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 between the 1930s and the 1990s cohort. As we will see in the next Section, this will generate an increasing life expectancy gap across education groups.

4.5 Decompositions

Combining all previous results, we can show that the different type composition across education groups explains 42% of the educational gradient of life expectancy, which is 7.8 years for the 1970s cohort. In particular, we compute counterfactual health dynamics for males without a high school degree for the case in which they have a distribution of lifestyles as the males with a college degree. In this counterfactual case, we find that their life expectancy would rise by 3.3 years, reducing the education gradient of life expectancy from 7.8 to 4.5 years. That is, the different health dynamics across education groups for a fixed distribution of lifestyle explains 4.5 years of the education

FIGURE 2: Distribution of types by education and cohort at age 50



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. The shaded area represents the 95% credible intervals.

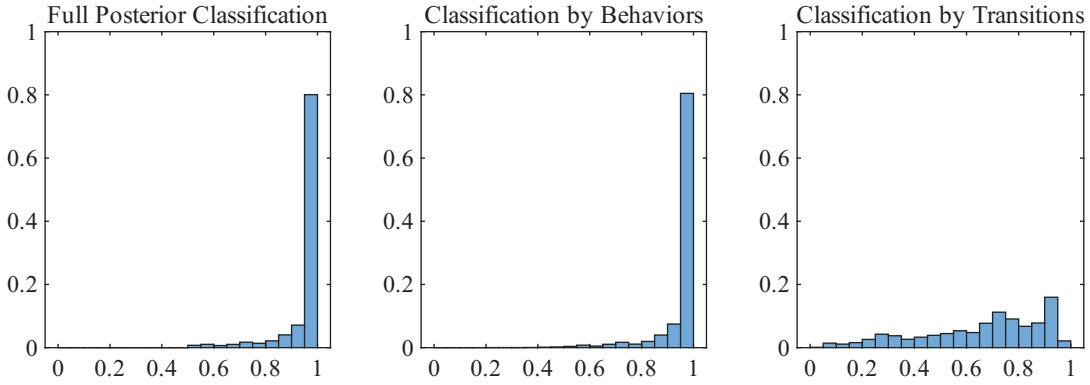
gradient in life expectancy. The difference, 3.3 years of the observed gradient, is explained by the different distribution of lifestyles across education groups (see the last two columns in Table 3).

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 cohorts have different life expectancies because of different compositions 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 findings show a growing education gradient in life expectancy: from 6.8 years in the 1930s cohort to 8.1 years in the 1990s cohort, which follows a 1.1-year decline in the gradient between the 1910s and the 1930s cohorts.

4.6 Identification of types

To understand how individuals are assigned to each latent lifestyle type, we decompose the role played by health behaviors and health transitions in the posterior classification. For each individual, we first compute the full posterior $p\big(y \mid \mathbf{z}_i^T, \mathbf{h}_i^T, \mathbf{ins}_i^T, x_i\big)$, where $x_i \equiv (\mathrm{c}_i, \mathrm{e}_i, a_{i0}, h_{i0})$ collects all individual covariates (cohort, education, initial age, and initial health), and we select the type $y \in \{\text{DET}, \text{PRO}\}$ with the highest posterior probability. The left panel of Figure 3 shows the distribution of this maximum posterior probability across individuals. The model classifies individuals very clearly: 78% of the sample is assigned to one of the two types with posterior probability above 0.95.

FIGURE 3: Posterior classification: health behaviors versus health transitions



Notes: The three panels display the distribution of the estimated posterior probabilities for the latent lifestyle type of each individual. The left panel shows the distribution of the maximum posterior probability $\max_y p(y \mid \mathbf{z}_i^T, \mathbf{h}_i^T, x_i)$ across individuals. The middle panel shows $p(y \mid \mathbf{z}_i^T, x_i)$, isolating the contribution of observed health behaviors. The right panel shows $p(y \mid \mathbf{h}_i^T, x_i)$, isolating the contribution of health transitions.

We then separate the contribution of behaviors and transitions by computing $p\big(y \mid \mathbf{z}_i^T, x_i\big)$ and $p\big(y \mid \mathbf{h}_i^T, x_i\big)$. The middle panel of Figure 3 plots the distribution of $p\big(y \mid \mathbf{z}_i^T, x_i\big)$ for the same group assignment chosen using the full posterior. This distribution is nearly identical to the full posterior, indicating that health behaviors almost entirely drive the classification of individuals into latent types. The right panel reports the distribution of $p\big(y \mid \mathbf{h}_i^T, x_i\big)$ for the same individuals. This distribution is much flatter: the model is largely unable to classify individuals using only their health transitions. Indeed, for roughly 25% of the sample, the type that maximizes $p\big(y \mid \mathbf{h}_i^T, x_i\big)$ is the opposite of the type identified using the full posterior. In short, while the likelihood incorporates both behaviors and transition dynamics, the grouping structure is fundamentally determined by differences in observed health behaviors, not by differences in health transitions.

5 An economic model

Our economic model considers two distinct life stages. The *early life* stage (Section 5.1) 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 y $\in \{\text{DET}, \text{PRO}\}$. This problem is a stand-in for choices and investments made by parents in early childhood or young adults before entering the labor market. This stage serves to account for the observed correlation between education and lifestyles. The *adult life* stage (Section 5.2) is a dynamic life-cycle consumption-saving problem under uncertainty in both labor market and health outcomes where individuals differ in fixed characteristics (education and lifestyle). This stage provides the endogenous value of starting life in each type, which is used in the *early life* stage.

5.1 Stage 1: early life

Let $V_0^{c,e,y}$ be the value of starting adult life with education e and lifestyle y for an individual in cohort c, as determined by the adult-life problem described in Section 5.2. Before entering the labor market, young individuals choose education and lifestyle jointly by solving

$$\max_{e,y} \left\{ V_0^{c,e,y} - \tau_e - \tau_y \right\}. \tag{4}$$

The terms τ_e and τ_y denote the utility costs of choosing education e and lifestyle y, respectively. These costs are heterogeneous and possibly correlated. We normalize the cost of a *detrimental* lifestyle and the cost of not finishing high school to zero, $\tau_{DET} = 0$ and $\tau_{HSD} = 0$. The costs of adopting a *protective* lifestyle and of completing higher levels of education, τ_{PRO}, τ_{HSG}, and τ_{CG}, are drawn from a joint distribution.

We want to comment on three assumptions embedded in equation (4). First, education and lifestyle choices are taken jointly before adult life and are fixed thereafter.[8] The main reason for this assumption is the empirical evidence on the early-age adoption of health-related habits and on the persistence of lifestyles once adopted.[9] Since our objective is to understand differences in lifestyles across education groups and cohorts, the initial formation of persistent lifestyles is the relevant margin. Second, we model early life investment as happening in a single period. This is different from dynamic settings like Cunha and Heckman (2007) or Conti and Heckman (2010). We make this assumption for simplicity, as well as for lack of data to discipline the potential early-life dynamic complementarities between the two different investments. We note, however, that the distributions of the costs of education and lifestyle investments are allowed to be correlated in our setting. In particular, individuals who face a low cost of adopting a *protective* lifestyle may also face a low cost of acquiring education. This positive correlation is the natural case if family background, parental investments, neighborhood conditions, or early-life environments jointly shape education and health-related investments. But it can also capture in reduced form the dynamic complementarities between education and lifestyle investments in early life. And third, we abstract from heterogeneous discount factors. Such heterogeneity could also generate correlated education and lifestyle choices, but our evidence on wealth accumulation does not support the view that more patient individuals largely self-select into both higher education and healthier lifestyles.[10]

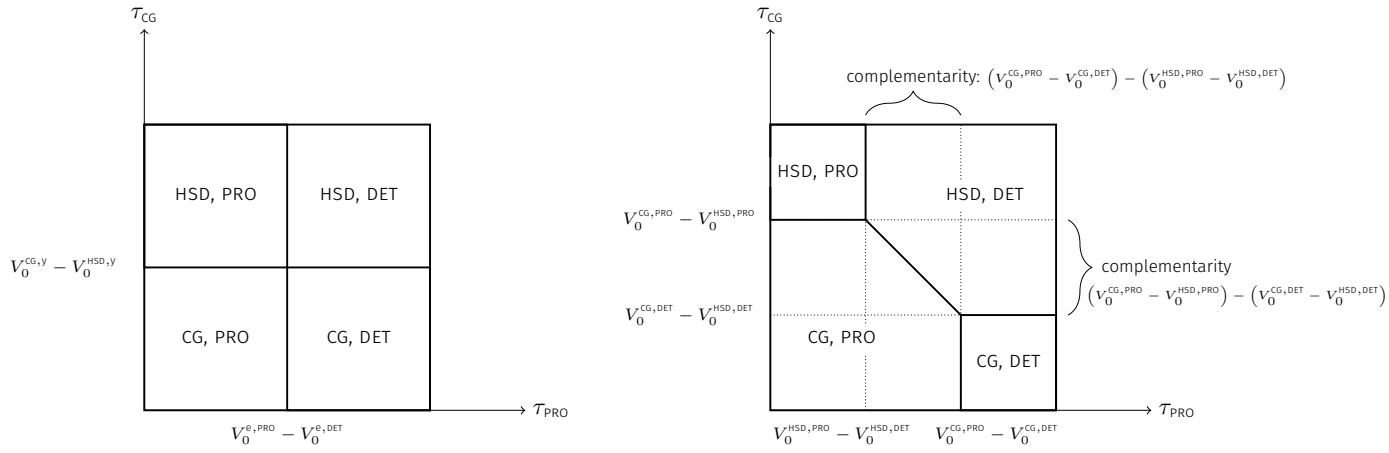
Separable choices. To understand the mechanics of the model, first consider the case in which education and lifestyle choices are separable. This occurs when the value of adopting a *protective*

[8] Recall that, according to our econometric specification in Section 3, a permanent latent lifestyle y allows observed behaviors, such as smoking, exercise, or preventive tests, to vary over the life cycle and across health states.

[9] See Farrell and Fuchs (1982) and Hai and Heckman (2026) for evidence on the early adoption of smoking and its correlation with later education decisions. Conner and Norman (2017) and references therein document the limited effects of interventions aimed at changing lifestyles during adulthood. Miller et al. (2026) and van Kippersluis and Galama (2014) find no evidence of improvements in health-related behaviors after plausibly exogenous income or wealth increases in adulthood.

[10] Wealth accumulation in the data is larger for the more educated and for the *protective*, but the model with homogeneous discount factors accounts for these patterns well; see Section 6.1.2.

FIGURE 4: Early life choices of education and lifestyle



(A) Separable choices (B) Complementarity

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

lifestyle is the same for all education groups:

$$V_0^{CG,PRO} - V_0^{CG,DET} = V_0^{HSG,PRO} - V_0^{HSG,DET} = V_0^{HSD,PRO} - V_0^{HSD,DET}.$$

In this case, there is a threshold $\tau^*_{PRO} \equiv V_0^{e,PRO} - V_0^{e,DET}$ that does not depend on the chosen education $e \in \{HSD, HSG, CG\}$ such that individuals choose a *protective* lifestyle if $\tau_{PRO} < \tau^*_{PRO}$ and a *detrimental* lifestyle with the reverse inequality. In terms of primitives, lifestyle choices depend on τ_{PRO} but not on τ_{CG} or τ_{HSG}. Symmetrically, in the separable case the value of education does not depend on the chosen lifestyle, and there are thresholds τ^*_{CG} and τ^*_{HSG} driving the choice of education that do not depend on the chosen lifestyle. In terms of primitives, education choices depend on τ_{CG} and τ_{HSG} but not on τ_{PRO}. Panel (A) of Figure 4 illustrates this case for the simpler environment with two education categories.

Complementarity. In contrast, education and lifestyle choices are complementary whenever the value of a *protective* lifestyle is increasing in education:

$$V_0^{CG,PRO} - V_0^{CG,DET} > V_0^{HSG,PRO} - V_0^{HSG,DET} > V_0^{HSD,PRO} - V_0^{HSD,DET}.$$

This is the relevant case in our calibrated model. When the value of a *protective* lifestyle rises with education —symmetrically, the value of education is larger with a *protective* lifestyle— the two choices reinforce each other. In this situation, the threshold costs $\tau^*_{PRO}(e)$ driving the lifestyle choices increase with education e, and the threshold costs $\tau^*_{CG}(y)$ and $\tau^*_{HSG}(y)$ driving the education decision increase with the lifestyle y. In terms of primitives, lifestyle choices depend on both sets of

costs, and so do education choices. For instance, individuals with, say, intermediate education costs may choose higher education if their cost τ_{PRO} of adopting a *protective* lifestyle is low, but choose lower education if τ_{PRO} is high. Panel (B) of Figure 4 illustrates this mechanism and Appendix C describes it in detail. The complementarity between education and lifestyle choices is consistent with empirical evidence showing that increases in education improve longevity (e.g. Lleras-Muney, 2005) and that increases in life expectancy raise educational attainment (e.g. Jayachandran and Lleras-Muney, 2009; Oster et al., 2013).

Selection. The model contains two sources of selection. The first is primitive selection through the joint distribution of costs: if education costs and lifestyle costs are positively correlated, individuals with low costs of adopting a *protective* lifestyle also tend to have low costs of education. This force directly increases the concentration of *protective* lifestyles among the more educated, which generates a positive education gradient in lifestyles. The second is endogenous selection generated by complementarities in adult-life values. Even if the primitive costs were independent, complementarity would induce individuals with lower τ_{PRO} to select into higher education, because the return to a higher education is larger for individuals with a *protective* lifestyle. In particular, for the case with two education categories and independent costs, complementarity implies that the distribution of τ_{PRO} conditional on choosing HSD first-order stochastically dominates the distribution of τ_{PRO} conditional on choosing CG. That is,

$$\Pr\left[\tau_{\text{PRO}} \leq \tau \mid \text{HSD}\right] \leq \Pr\left[\tau_{\text{PRO}} \leq \tau \mid \text{CG}\right] \quad \forall \tau,$$

with strict inequality for some τ; see Proposition 1 in Appendix C. The endogenous selection also generates a positive education gradient in lifestyles.

5.2 Stage 2: adult life

In this Section, we model adult life, which generates the values $V_0^{\text{c,e,y}}$ used in the early life stage.

Demographics and preferences. The model period corresponds to two years. Individuals live for at most T periods, but survival is stochastic every period. People are exposed to health shocks, medical expenditure shocks and, up to period $R-1$, 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:

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

where σ is the risk-aversion parameter, $\bar{n}_t$ 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:

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

where b_0 drives the strength of the bequest motive and b_1 the degree of non-homotheticity.

Health dynamics. Following the empirical model in Section 3, health h_t can be either good (h_g) or bad (h_b) and, conditional on survival, it evolves according to the age-dependent first-order Markov chain $\Gamma_t^{\mathrm{e,y}}\left(h'|h\right)$, which depends on age t, education e, and lifestyle y. The survival probability $s_t^{\mathrm{e,y}}(h_t)$ also depends on age t, health h_t, education e, and lifestyle y.

Medical expenses and insurance Each period, after making their consumption and saving decisions, individuals draw medical expenditure and insurance shocks. Medical expenditures are given by $m_t^e(h_t, \zeta_t)$, where the i.i.d. shock ζ_t determines the level of expenditures conditional on age, education, and health. Working-age individuals receive private insurance with probability $\mathrm{Prob}_t^e(\ell)$, which depends on age, education, and employment status. Conditional on being insured, $q_t^e(h_t, \zeta_t)$ denotes the fraction of medical expenditures paid by the insurer. Thus, uninsured individuals pay the full amount, whereas insured individuals pay the fraction $1 - q_t^e(h_t, \zeta_t)$. Employed individuals with private insurance pay no premium because we assume that their employer covers its full cost. Unemployed individuals with private insurance pay the premium $p_t^e(h_t)$. All retired individuals are covered by Medicare, which pays 50% of their medical expenditures. The annual Medicare Part B premium is \$1,055. Appendix D describes the calibration of these objects.

Labor earnings. During working life, individuals face exogenous labor earnings risk, which we model in two components. First, employment is determined by a probit process. Let

$$\ell_t^* = X_t^{\mathrm{c,e}}(h_t)'\beta_e + u_t, \qquad u_t \sim N(0,1),$$

where $X_t^{\mathrm{c,e}}(h_t)$ is a vector of variables with cohort, education, age, and health. Individuals work when $\ell_t^* > 0$, so that

$$\Pr\left(\ell_t = 1 \mid \mathrm{c,e}, t, h_t\right) = \Phi\left(X_t^{\mathrm{c,e}}(h_t)'\beta_e\right).$$

This component captures differences in employment across education groups and the higher non-employment probability of individuals in bad health, which is one of the channels through which health affects labor income.

Second, conditional on working, individuals receive labor earnings

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

where $\mu_t^{\mathrm{c,e}}(h_t)$ is a deterministic component of log earnings that depends on cohort, education, age, and health. The terms η_t and ϵ_t are persistent and transitory earnings shocks, respectively. The

persistent component follows the stationary AR(1) process

$$\eta_{t+1} = \rho_e \eta_t + \nu_{t+1}, \qquad \nu_{t+1} \sim N\left(0, \sigma^2_{\nu,e}\right),$$

where one model period corresponds to two calendar years. Its initial value is drawn from the stationary distribution, $\eta_0 \sim N\left(0, \sigma^2_{\eta,e}\right)$ with $\sigma^2_{\eta,e} = \sigma^2_{\nu,e} / \left(1 - \rho^2_e\right)$. The transitory component is Gaussian white noise, $\epsilon_t \sim N\left(0, \sigma^2_{\epsilon,e}\right)$.

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}), with the latter only affecting earnings below w_{ss}. Following Heathcote et al. (2020), we define a progressive labor income tax function $T(w) = w - a_{\tau 0} w^{1-a_{\tau 1}}$. We represent several existing means-tested programs in a stylized way through a public safety-net program guaranteeing 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 as follows. First, we divide individuals into groups based on their labor force participation just before retirement, their last draw of the persistent productivity shock, and their education and lifestyle. Then, for each group, we compute average earnings over the 17 model periods (34 years) with the highest earnings, and we apply the Social Security benefits formula to these average earnings.[11]

5.2.1 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, the persistent component of productivity η_t, 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), medical expenses, and insurance coverage.

[11] Of course, agents in the model are aware of how these variables and shocks affect their pension income.

The optimization problem for working-age individuals ($t < R$) is:

$$V_t^{\mathrm{c,e,y}}(x,h,\eta) = \max_{c,k'} \Bigg\{ u(c) + \beta \Bigg[s_t^{\mathrm{e,y}}(h) \sum_{h'} \Gamma_t^{\mathrm{e,y}}(h' \mid h) \mathbb{E}_{\ell',\eta',\epsilon',\zeta',\iota'} \left[V_{t+1}^{\mathrm{c,e,y}}(x',h',\eta') \right] + \left(1 - s_t^{\mathrm{e,y}}(h)\right) v(k') \Bigg] \Bigg\}, \tag{6}$$

$$\text{s.t.} \quad c + k' = x,$$

$$x' = \max \left\{ w_{t+1}^{\mathrm{c,e}}(h',\eta',\epsilon')\ell' - \mathrm{Tax}_{t+1} + (1+r)k' - d_{t+1}^{\mathrm{c,e}} - \widetilde{m}_{t+1}^{e}(h',\zeta',\iota',\ell'),\ \underline{x} \right\},$$

$$\widetilde{m}_{t+1}^{e}(h',\zeta',\iota',\ell') = \left[1 - \iota' q_{t+1}^{e}(h',\zeta')\right] m_{t+1}^{e}(h',\zeta') + \iota'(1-\ell') p_{t+1}^{e}(h'),$$

$$\mathrm{Tax}_{t+1} = T\left(w_{t+1}^{\mathrm{c,e}}(h',\eta',\epsilon')\ell'\right) + \tau_{MCR} w_{t+1}^{\mathrm{c,e}}(h',\eta',\epsilon')\ell' + \tau_{ss} \min\left\{ w_{t+1}^{\mathrm{c,e}}(h',\eta',\epsilon')\ell', w_{ss} \right\}.$$

Here, primes denote next-period realizations. The expectation is taken over next-period employment ℓ', persistent productivity η', transitory productivity ϵ', the medical expenditure shock ζ', and private insurance status ι'. The indicator ι' equals one if the individual has private insurance in period $t+1$ (with probability $\mathrm{Prob}_{t+1}^{e}(\ell')$) and zero otherwise. The object $\widetilde{m}_{t+1}^{e}(h',\zeta',\iota',\ell')$ denotes the individual's total out-of-pocket medical payment, including any private insurance premium. Uninsured individuals pay the full medical expenditure $m_{t+1}^{e}(h',\zeta')$. Privately insured individuals pay the fraction $1 - q_{t+1}^{e}(h',\zeta')$, where $q_{t+1}^{e}(h',\zeta')$ is the fraction paid by the insurer. Privately insured individuals who are unemployed also pay the premium $p_{t+1}^{e}(h')$.

During the first five model periods of adult life (10 calendar years), college graduates repay the pecuniary cost of college education. Let $d_t^{\mathrm{c,e}}$ denote the repayment due in period t. We set $d_t^{\mathrm{c,e}} = 0$ for $\mathrm{e} \neq \mathrm{CG}$ or $t > 5$. For college graduates and $t \leq 5$, $d_t^{\mathrm{c,CG}}$ is the annuity payment associated with four years of college costs for cohort c, repaid over the first ten years of adult life.

In this problem, wealth accumulation will be larger for the more educated and for the *protective* lifestyle for two different reasons. First, both groups earn more labor income net of medical expenses (either directly or through the lower incidence of bad health). Second, both groups feature higher survival, which translates into larger effective discount factors and larger willingness to save.

6 Calibration

The parameterization of the model proceeds in two parts. First, we calibrate or estimate the parameters of the adult life-cycle model (Section 6.1). Second, we estimate the parameters governing heterogeneity in education and lifestyle costs in the early-life stage (Section 6.2).

6.1 Stage 2: Adult life

We calibrate the adult life stage with the 1950s cohort. We fix the annual interest rate at 4%, retirement age at 66, and maximum lifespan to 98. The survival and health-transition processes, $s_t^{\mathrm{e,y}}(h)$ and $\Gamma_t^{\mathrm{e,y}}(h' \mid h)$, are taken from Section 3.2. Below, we describe the estimation of the income process (Section 6.1.1), the estimates of the vector of parameters $\{\beta, \sigma, \underline{x}, b_0, b_1\}$ by simulated method of moments (Section 6.1.2), and the calibration of $\bar{b}$ to match the value of a statistical life (6.1.3). The tax and Social Security parameters are described in Appendix D. All monetary values are in 2018 dollars.

6.1.1 Income process

We use PSID data covering 1968–2019 to estimate the labor earnings process $w_t^{\mathrm{c,e}}(h, \eta, \epsilon)$ and employment probabilities $\Pr(\ell_t = 1 \mid \mathrm{c,e}, t, h_t)$. The sample consists of working-age men before retirement. We define the employment state as annual labor earnings above \$3,770.[12] Conditional on employment, we estimate the deterministic component of log earnings separately by education using an individual fixed-effects model. Age and health effects are hence identified from within-person variation, while cohort differences are recovered in a second step. We estimate the persistent and transitory components of residual earnings by GMM. Finally, we estimate employment probabilities separately by education using Mundlak correlated-random-effects probit models. Appendix D.1 provides the estimation details.

Figure 5 illustrates the education and health gradients in labor-market outcomes. The left panel reports mean log labor earnings conditional on work, and the right panel reports labor force participation, separately by education and health status. Bad health is associated with lower earnings and lower employment at all ages and education categories. These patterns imply that poor health generates large labor-market losses, consistent with recent findings (e.g. Hosseini et al., 2026). However, these losses are not particularly different across education groups.

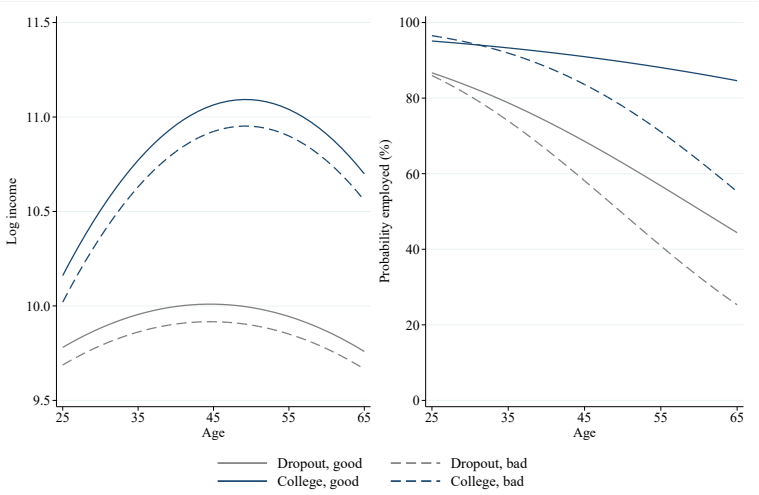
Finally, the estimated parameters of the stochastic component of the income process show that the variances of persistent and transitory earnings shocks differ across education groups; see Table F.1 in Appendix F.

For college graduates, the adult budget constraint includes the repayment of college costs during the first ten years of working life, as described in Section 5.2. We calibrate these costs using NCES data on average undergraduate tuition, fees, room, and board. For each cohort, we take the annual college cost, multiply it by four years of college, and convert it into a ten-year annuity using the model interest rate.[13]

[12] This threshold corresponds to annual earnings from 520 hours at the 2018 federal minimum wage of \$7.25.

[13] The annual costs are \$10,500, \$10,900, and \$15,900 for the 1930s, 1950s, and 1970s cohorts, respectively. Source: NCES Digest of Education Statistics, Table 330.10.

FIGURE 5: Labor-market profiles by health and education



Notes: The left panel reports mean log labor earnings conditional on work and the right panel reports labor force participation. Profiles are shown for high school dropouts (HSD) and college graduates (CG), separately by health status. Profiles correspond to the 1950s cohort.

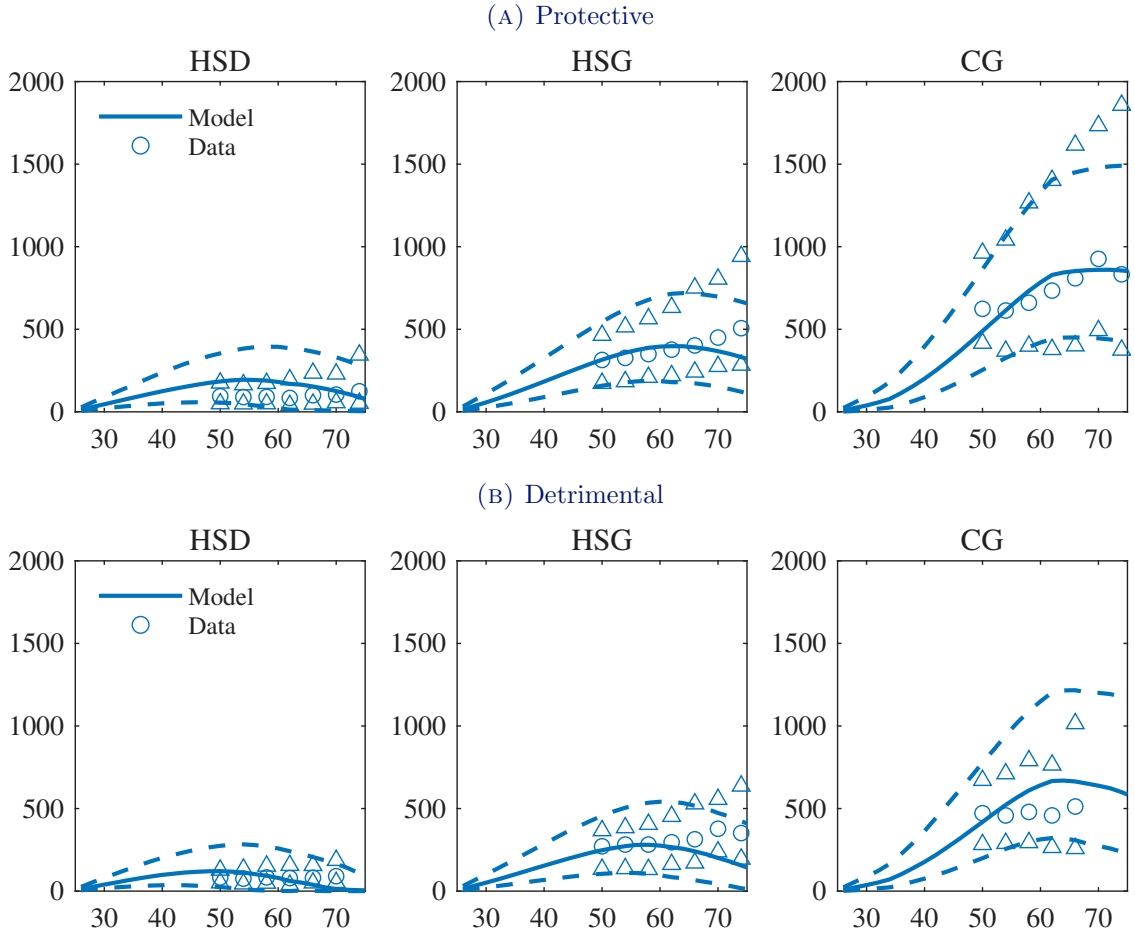
6.1.2 Matching wealth trajectories

Conditional on the externally calibrated inputs discussed above, we estimate the remaining adult-life parameters, $\{\beta, \sigma, \underline{x}, b_0, b_1\}$, using the simulated method of moments. We minimize the sum of squared differences between the 25th, 50th, and 75th percentiles of assets in the data and their corresponding model-implied values, by education e, lifestyle y, and age t, using individuals born in the 1950s.

We construct the empirical wealth moments by assigning each individual to the lifestyle type with the highest posterior probability, as described in Section 4.6. This assignment is relatively unambiguous within the estimated model: 78% of individuals have a maximum posterior probability above 0.95. We then compute the 25th, 50th, and 75th percentiles of wealth by age, education, and assigned lifestyle type. Figure 6 presents the resulting moments together with their model counterparts. In particular, the solid blue lines represent median wealth, while the dashed blue lines represent the 25th and 75th percentiles. As is well known, wealth accumulation is positively correlated with education. In addition, we highlight two results. First, conditional on education, wealth accumulation is stronger for the *protective* type. Second, the difference in wealth accumulation across types is especially apparent among college-educated individuals and smaller within the other two education categories. Overall, our calibrated model reproduces these facts well.

As an untargeted validation, Figure 7 compares wealth trajectories by health status in the data and the model. These moments are not included in the calibration objective. Wealth profiles display substantial heterogeneity by health. The model reproduces these differences well, providing additional evidence on its ability to capture the relationship between health and wealth accumulation.

FIGURE 6: Model fit: wealth trajectories by lifestyle



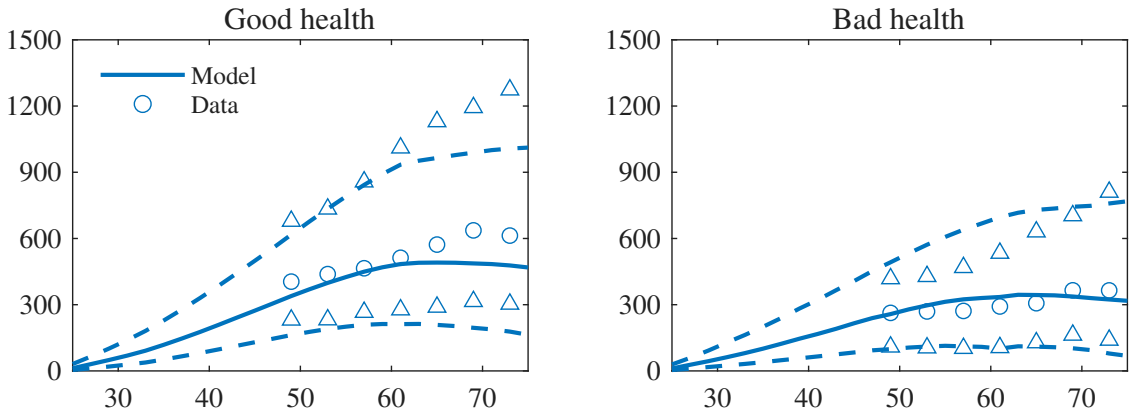
Notes: Each panel plots wealth over the life cycle for one lifestyle type, separately by education group (HSD, HSG, and CG). In each subplot, the solid line shows the model-implied median wealth, while the two dashed lines show the model-implied 25th and 75th percentiles. The markers report the corresponding moments in the data after assigning individuals to their most likely lifestyle type: circles denote the median and triangles denote the 25th and 75th percentiles.

Finally, Panel A in Table 4 reports our estimated parameters (see "Benchmark" column). The estimated discount factor is 0.98, and the estimated coefficient of relative risk aversion is 1.33. The bequest parameters b_0 and b_1 are 33.93 and 895.81, respectively. These values imply that in the period before certain death, the bequest motives become active at \$63,000, and the marginal propensity to bequeath is 93 cents out of every additional dollar. The biennial income floor is estimated at \$5,020. These parameters fall in the range of parameters estimated by the previous literature, see discussion in Ameriks et al. (2020).

6.1.3 Value of life

To identify the parameter $\bar{b}$, we use a measure of the value of a statistical life (VSL). The use of VSL to calibrate the marginal rate of substitution between life years and consumption in macro models

FIGURE 7: Model fit: wealth gradient by health



Notes: Each panel plots wealth over the life cycle across health status. In each subplot, the solid line shows the model-implied median wealth, while the two dashed lines show the model-implied 25th and 75th percentiles. The markers report the corresponding moments in the data: circles denote the median and triangles denote the 25th and 75th percentiles. The wealth moments by health status are not targeted in the calibration.

follows Rosen (1988) and is today relatively common.[14] 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. 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 40 years of age, see Appendix D.4 for details. We target a VSL of $2,000,000 and in Appendix E we discuss the case with a VSL of $4,000,000.

6.1.4 Taking stock

With this calibration, the model delivers a large value of adopting a *protective* lifestyle, $V_0^{e,PRO} - V_0^{e,DET}$, which increases with education e. We express this welfare gain as a consumption equivalent. At the start of working life, a high school dropout adopting a *protective* lifestyle could give up 51% of this consumption stream and still be as well off as a high school dropout with a *detrimental* lifestyle. The corresponding consumption equivalents increase with education: 56% for high school graduates and 62% for college graduates. Therefore, the calibrated model delivers a strong complementarity between lifestyle and education investments. We note that this complementarity has been informed by the adult life stage of the model, and that it does not use any information on the observed correlation between education and lifestyles.

[14] See for instance Murphy and Topel (2006), Hall and Jones (2007), Jones and Klenow (2016), or De Nardi et al. (2025) among others.

TABLE 4: Calibrated parameters

Parameter	Description	Value
PANEL A: ADULT LIFE		
β	Discount factor	0.98
σ	Relative risk aversion	1.33
$\underline{x}$	Income floor	5.02
b_0	Bequest motive: marginal utility	33.93
b_1	Bequest motive: non-homotheticity	895.81
$\bar{b}$	Value of life	3.20
PANEL B: EARLY LIFE		
μ_{HSG}	Average cost of HSG education	4.01
μ_{CG}	Average cost of CG education	15.06
μ_{PRO}	Average cost of PRO lifestyle	5.39
σ_{HSG}	Sd. cost of HSG education	2.56
σ_{CG}	Sd. cost of CG education	11.34
σ_{PRO}	Sd. cost of PRO lifestyle	5.78
ρ	Correlation edu,PRO	0.20

Notes. Panel A: internally calibrated parameters for the adult life stage. Panel B: calibrated parameters for the distributions of the utility costs τ_{PRO}, τ_{CG}, and τ_{HSG} in the early life stage of the model.

6.2 Stage 1: Early life

We parameterize the cost of adopting a *protective* lifestyle and the costs of completing high school and college education as

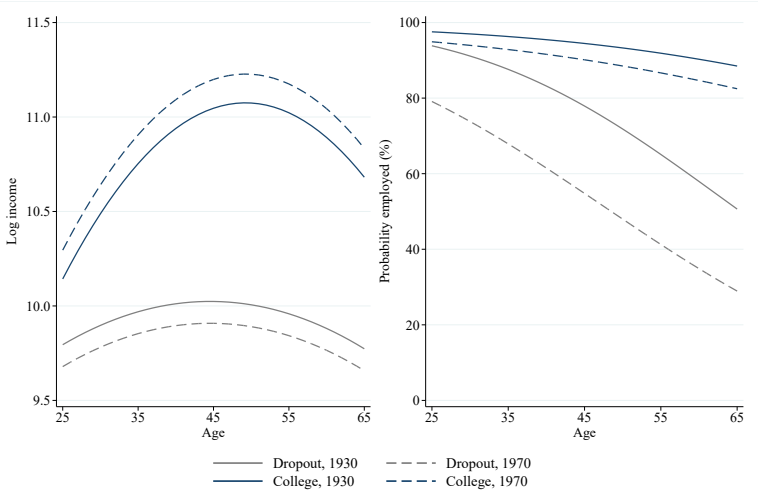
$$\tau_{\text{PRO}} = \mu_{\text{PRO}} + \sigma_{\text{PRO}}\epsilon_{\text{PRO}}, \qquad \tau_{\text{HSG}} = \mu_{\text{HSG}} + \sigma_{\text{HSG}}\epsilon_e, \qquad \tau_{\text{CG}} = \mu_{\text{CG}} + \sigma_{\text{CG}}\epsilon_e.$$

The terms μ_{PRO}, μ_{HSG}, and μ_{CG} capture the average of these costs in the population, and σ_{PRO}, σ_{HSG}, and σ_{CG} capture their standard deviation. We model the idiosyncratic components, ϵ_e and ϵ_{PRO}, to be jointly and normally distributed across the population, with zero mean, unit variance, and correlation ρ.

We calibrate these parameters to match the joint distribution of education and lifestyles for two cohorts, the 1930s and the 1970s. The value functions $V_0^{\text{c,e,y}}$ driving early-life choices will vary by cohort c because we fit the adult stage of the model with the corresponding cohort-specific life-cycle paths of labor earnings and employment rates.[15] Figure 8 illustrates the cohort variation in labor-market prospects that enters the model. The left panel reports mean log labor earnings conditional on work, while the right panel reports labor force participation profiles. We show the profiles for high school dropouts and college graduates in the 1930s and 1970s cohorts. Two patterns are apparent. First, labor force participation deteriorates substantially for high school dropouts in the 1970s cohort relative to the 1930s cohort. Second, earnings conditional on labor

[15] We also fit the model with a higher college cost for the 1970s cohort, see footnote 13. However, the college cost change contributes negligibly compared to the changes in lifetime incomes.

FIGURE 8: Labor-market profiles by cohort and education



Notes: The left panel reports mean log labor earnings conditional on work and the right panel reports labor force participation. Profiles are shown for high school dropouts (HSD) and college graduates (CG) in the 1930s and 1970s cohorts. Profiles are for individuals in good health.

market participation evolve in opposite directions across education groups: they fall for high school dropouts and rise for college graduates. All in all, lifetime income increased by 8% for college graduates and declined by 23% for high school dropouts.

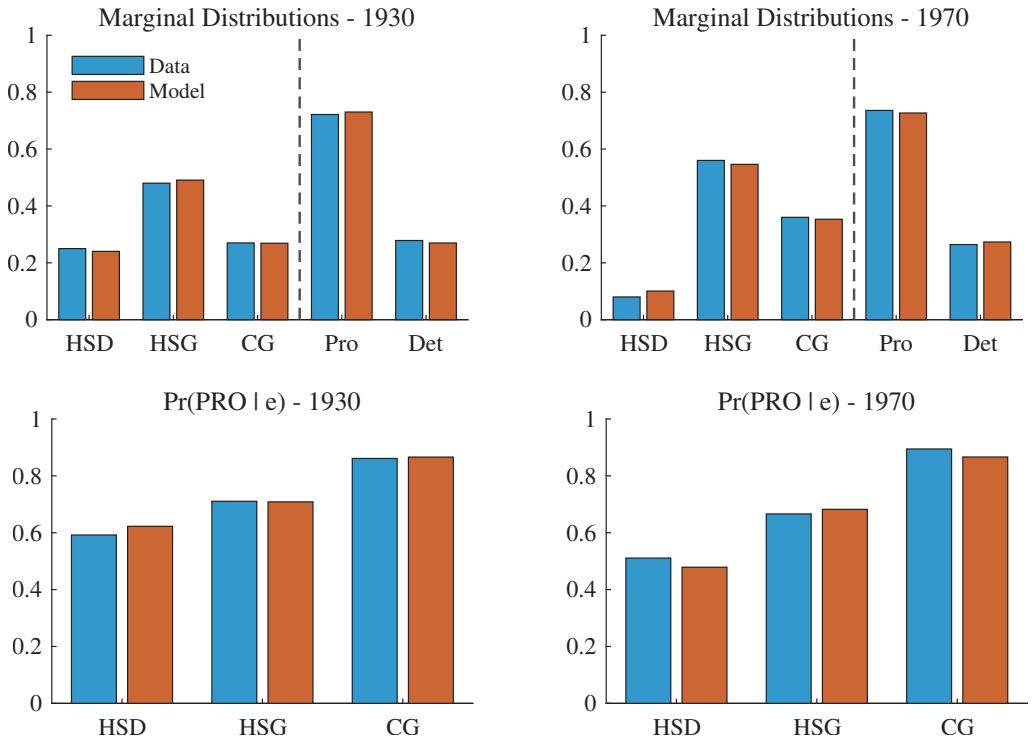
6.2.1 Identification and results

We have 7 parameters to estimate: $(\mu_{\text{CG}}, \mu_{\text{HSG}}, \mu_{\text{PRO}}, \sigma_{\text{CG}}, \sigma_{\text{HSG}}, \sigma_{\text{PRO}}, \rho)$. Because we require the model to match the joint distribution of (e, y) for two cohorts, there are 10 moment conditions, $3\times2-1$ targets per cohort. The marginal distributions of education and lifestyles in the 1930s cohort mainly discipline the average cost parameters $(\mu_{\text{CG}}, \mu_{\text{HSG}}, \mu_{\text{PRO}})$. Changes in these distributions between the 1930s and 1970s cohorts discipline the dispersion parameters $(\sigma_{\text{CG}}, \sigma_{\text{HSG}}, \sigma_{\text{PRO}})$, because they determine how choices respond to changes in labor earnings and employment prospects.[16] The level and changes of the joint distribution of education and lifestyles are informative about ρ, since this parameter directly affects how strongly low costs of *protective* behavior are associated with low education costs.

A positive ρ works in the same direction as the complementarity mechanism described in Section 5.1. By itself, positive correlation in costs generates more selection of individuals with low τ_{PRO} into higher education. Hence, allowing for $\rho > 0$ helps the model reproduce the observed education gradient in lifestyles. This is useful for interpreting the decomposition exercises below: part of the observed sorting is captured by primitive correlated heterogeneity, while the remaining sorting is generated endogenously by the adult-life complementarity between education and *protective*

[16]This logic 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 9: 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.

lifestyles.

The upper panels of Figure 9 show the marginal distributions of education and lifestyles, while the lower panel reports lifestyle distributions conditional on education. For the 1930s cohort, the share with a *protective* lifestyle among high school dropouts is 59% in the data and 62% in the model. Among college graduates, the corresponding shares are 86% in the data and 86.5% in the model.

Between the 1930s and 1970s cohorts, the share with a *protective* lifestyle among high school dropouts declines by 8 percentage points in the data and by 15 percentage points in the model. Among college graduates, this share increases by 3 percentage points in the data but remains approximately constant in the model. Consequently, the education gradient in lifestyles increases by 11 percentage points in the data and by 15 percentage points in the model. The corresponding increase in the life expectancy gradient is 1.0 year in the data and 1.2 years in the model.

6.2.2 Early-life determinants of education and lifestyle choices

Our calibration recovers the joint distribution of early-life education and lifestyle unobserved utility costs. We now give an empirical interpretation to this distribution by examining how education and estimated lifestyles are related to genetic predispositions and childhood family characteristics. We

TABLE 5: Early-life characteristics by education and lifestyle

	Lifestyles		Education		
	DET	PRO	HSD	HSG	CG
Panel A: Polygenic scores (SD)					
Smoking initiation	0.196	-0.102	0.092	-0.008	-0.170
Cigarettes per day	0.132	-0.075	0.138	-0.002	-0.182
Drinks per week	0.033	-0.021	-0.046	0.006	-0.020
Educational attainment	-0.258	0.076	-0.353	-0.090	0.395
# obs	1372	4985	971	3624	1762
Panel B: Family background					
Parents/guardians smoked (%)	75.6	68.1	66.7	71.4	67.8
# obs	1292	4796	883	3483	1722
Mother: 12+ years of schooling (%)	11.8	16.3	3.8	10.1	30.6
# obs	1198	4578	746	3307	1723
Father: 12+ years of schooling (%)	12.4	18.9	1.9	12.3	33.7
# obs	1086	4402	670	3127	1691

Notes: The sample consists of 6,357 HRS men with all four polygenic scores and an estimated lifestyle assignment. Panel A reports scores in the supplied within-ancestry standard-deviation units. Panel B reports percentages. Parental smoking refers to parents or guardians smoking during the respondent's childhood. Parental schooling indicators equal one for more than 12 years and zero for 12 years or less. N reports the corresponding number of observations.

interpret these observable characteristics as proxies for the unobserved costs of acquiring education, τ_{e}, and adopting a *protective* lifestyle, τ_{PRO}, but we recognize that other factors, such as schools or neighborhoods, may play an equally important role.

We use supplemental data from the HRS on polygenic scores and childhood family background. Polygenic scores aggregate information across genetic variants, weighted by their associations with a particular trait in genome-wide association studies. We consider scores for educational attainment, smoking initiation, cigarettes per day, and drinks per week, based on the studies of Lee et al. (2018) and Liu et al. (2019). Higher scores indicate a greater genetic propensity toward the corresponding trait. The supplied scores are standardized to have zero mean and unit standard deviation within each ancestry group in the HRS genetic sample. For this exercise, we assign each individual to the lifestyle type with the highest posterior probability. We also observe whether respondents' parents or guardians smoked during childhood and the schooling of their mothers and fathers. For each parent, we construct an indicator for having completed more than 12 years of schooling.

Table 5 reports the average polygenic scores and family background characteristics across lifestyle types and education categories.[17] The averages differ across groups in the expected direction: individuals with a *detrimental* lifestyle have higher polygenic scores for smoking initiation, cigarettes per day, and drinks per week than those with a *protective* lifestyle, by approximately 0.30, 0.21, and 0.05 standard deviations, respectively. The educational-attainment score increases with

[17]Polygenic scores are noisy measures of genetic propensity, so the differences in Panel A understate the underlying associations, see van Kippersluis et al. (2023). One should therefore read more the sign rather than the magnitudes of the differences between the columns of Panel A.

education, with a difference of approximately 0.75 standard deviations between college graduates and high school dropouts. Family background characteristics follow the same pattern. Individuals with a *detrimental* lifestyle are more likely to report that their parents or guardians smoked during childhood (75.6% versus 68.1% for the *protective* type), and more educated individuals are more likely to have mothers and fathers with more than 12 years of schooling. For example, 30.6% of college graduates have mothers with more than 12 years of schooling, compared with 3.8% of high school dropouts. These findings are consistent with the model's assumption that the costs of acquiring education and adopting a *protective* lifestyle are heterogeneous across the population and matter for these choices.

More interestingly, the table also shows that characteristics associated with one investment differ across choices on the other margin. Individuals with a *protective* lifestyle have an educational-attainment score that is approximately 0.33 standard deviations higher than that of the *detrimental* type. They are also more likely to have mothers and fathers with more than 12 years of schooling (16.3% versus 11.8% for mothers and 18.9% versus 12.4% for fathers). Conversely, college graduates have lower smoking-initiation and cigarettes-per-day scores than high school dropouts, by approximately 0.26 and 0.32 standard deviations, respectively.[18] Thus, proxies for lower education costs are also associated with a *protective* lifestyle, while proxies for lower costs of adopting a *protective* lifestyle are also associated with higher education. These patterns are consistent both with complementarity between education and lifestyle investments and with a positive correlation between the early life characteristics in the first-stage model.

7 Counterfactual results

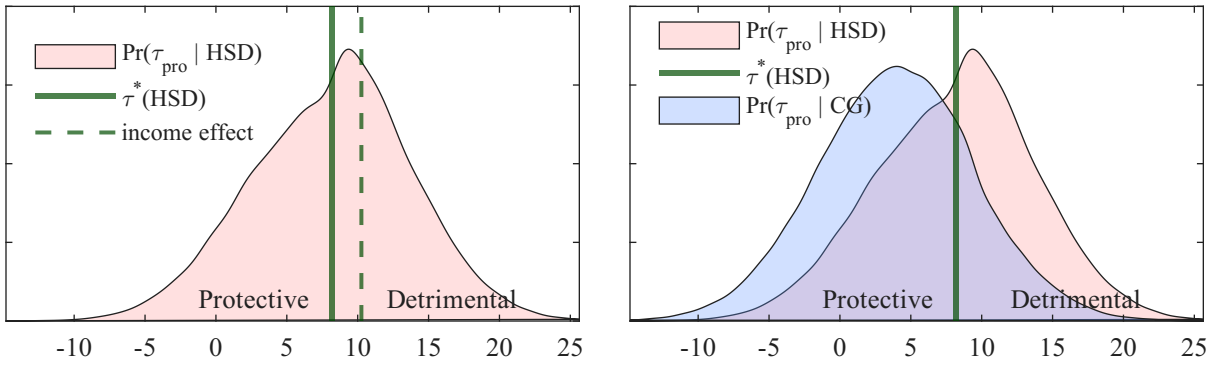
In this Section, we aim to answer three questions: why is there an education gradient of lifestyles (Section 7.1), how early life interventions could limit inequality in health outcomes (Section 7.2), and what has been the effect of the rising education premium of labor earnings on the increase in education gradient of life expectancy (Section 7.3).

7.1 The education gradient of lifestyles

Our calibrated model generates a positive association between education and lifestyle investments through two related mechanisms. First, the higher expected income associated with education encourages healthier behavior because greater consumption possibilities increase the value of life and, consequently, the return to adopting a *protective* lifestyle. This mechanism makes education and lifestyle investments complementary and is standard in models of health investment through health behavior; see Becker (2007) and the recent contributions of Cole et al. (2019), Mahler and Yum (2024), and Margaris and Wallenius (2023). Second, this complementarity —together with

[18] Only the smoking of parent/guardian does not adhere to this pattern.

FIGURE 10: Decomposition of lifestyle choices conditional on education (1970s cohort)



Notes: In Panel (a), the shaded red area shows the density of τ_{PRO} among individuals who choose HSD. The solid green vertical line reports the benchmark threshold $\tau^*(\text{HSD})$ separating individuals who choose *protective* and *detrimental* lifestyles. The dashed green vertical line reports the corresponding threshold when high school dropouts face the income prospects of college graduates. Panel (b) compares the conditional distributions of τ_{PRO} among high school dropouts and college graduates.

the calibrated positive correlation between the unobserved utility costs of education and lifestyle—generates self-selection of individuals facing lower costs of adopting a *protective* lifestyle into higher education. Consequently, individuals with low τ_{PRO} are disproportionately represented among the more educated, further strengthening the education gradient in lifestyles.

To gauge the importance of each mechanism, we show graphically the lifestyle choice conditional on education for the calibrated model. The shaded red area in Panel (a) of Figure 10 plots $\Pr(\tau_{\text{PRO}}|\text{HSD})$, the distribution of *protective* lifestyle costs τ_{PRO} for individuals who choose to drop out of high school in the 1970s cohort. The vertical solid green line represents $\tau^*(\text{HSD}) \equiv V_0^{70,\text{HSD,PRO}} - V_0^{70,\text{HSD,DET}}$, the value of choosing a *protective* lifestyle over a *detrimental* lifestyle for individuals choosing to drop out of high school. Individuals with $\tau_{\text{PRO}} < \tau^*(\text{HSD})$ opt for *protective* behavior, while the rest choose *detrimental* behavior. The integral of $\Pr(\tau_{\text{PRO}}|\text{HSD})$ between minus infinity and $\tau^*(\text{HSD})$ gives the fraction of dropouts adopting a *protective* lifestyle.

We start by conducting a decomposition exercise that holds individuals' education choices fixed, and examines how lifestyle choices respond to changes in income prospects and changes in the distribution of lifestyle costs. In a first scenario, we compute the choices of high school dropouts if they were to face the income prospects of college graduates. Due to the higher consumption possibilities, the increase in expected earnings leads to higher value functions for both lifestyle options, which we denote as $\widetilde{V}_0^{70,\text{HSD,PRO}}$ and $\widetilde{V}_0^{70,\text{HSD,DET}}$. We also find that the gain is higher for y = PRO, as life expectancy is higher for this type and the higher consumption flow is enjoyed for more years. Thus, we have that $\tilde{\tau}^*(\text{HSD}) \equiv \widetilde{V}_0^{70,\text{HSD,PRO}} - \widetilde{V}_0^{70,\text{HSD,DET}} > \tau^*(\text{HSD})$, which shifts to the right the threshold value that prompts individuals to adopt a *protective* lifestyle, see the dashed green vertical line in Figure 10. In this case, the proportion of high school dropouts choosing a *protective* lifestyle increases from 47% to 63% and the education gradient in life expectancy falls by 1.3 years (from 7.8 to 6.5 years), a reduction of approximately 17%; see the second and fourth rows of Table 6.

TABLE 6: Decomposition: 1970s cohort

	Pr(e)		Pr(PRO|e)			LE(e)		
	CG	HSD	CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.36	0.08	0.89	0.51	0.38	31.5	23.7	7.8
Model	0.36	0.10	0.87	0.47	0.39	31.2	23.4	7.8
Same lifestyle			0.87	0.87	0.00	31.2	26.7	4.5
Income			0.87	0.63	0.23	31.2	24.7	6.5
Selection			0.87	0.78	0.09	31.2	25.9	5.2
Selection + income			0.87	0.88	-0.01	31.2	26.8	4.4
No correlated costs	0.35	0.10	0.81	0.61	0.20	30.7	24.5	6.1

Notes: The first two rows report results from the data and the calibrated model. The third row equalizes the conditional lifestyle distribution across education groups. The fourth row gives high school dropouts the income prospects of college graduates. The fifth row gives high school dropouts the distribution of τ_{PRO} faced by college graduates. The sixth row combines both and the seventh row imposes $\rho = 0$. See Section 7.1 for details.

In a second scenario, we compute the behavior of high school dropouts if the distribution of τ_{PRO} was the same as for the college graduates. Panel (b) in Figure 10 displays the distributions of the cost of *protective* behavior for both high school dropouts (red shaded area) and college graduates (blue shaded area), with the latter shifted to the left as predicted by Proposition 1. As a consequence, the mass of individuals below the threshold $\tau^*(\text{HSD})$ is larger under the distribution of τ_{PRO} faced by college graduates. In this counterfactual, the proportion of high school dropouts choosing a *protective* lifestyle increases from 47% to 78% and the education gradient in life expectancy falls by 2.6 years (from 7.8 to 5.2 years), a reduction of approximately 33%; see the fifth row of Table 6.

Finally, when we put the two scenarios together (income prospects and selection), the lifestyle choices of the high school dropout are almost equal to those of the college graduates, and the education gradient in life expectancy drops by 3.4 years (from 7.8 to 4.4 years), see the second and sixth rows in Table 6. The combined effect is smaller than the sum of each channel (1.3 and 2.6 years, respectively) because of the negative interaction, which can be seen in Figure 10. The region of the distribution of τ_{PRO} that switches from *detrimental* to *protective* lifestyle when increasing income becomes less populated when giving high school dropouts the distribution of τ_{PRO} of college graduates.[19]

We close this section by assessing the role of primitive correlated heterogeneity by setting the correlation between education and lifestyle costs to zero while keeping all other calibrated parameters and adult-life values fixed. Absent the primitive correlation, the share of individuals choosing a *protective* lifestyle would fall from 87% to 81% among college graduates and would rise from 47%

[19]We note that giving high school dropouts the income prospects and τ_{PRO} distribution of college graduates (row 6) does not completely equalize the lifestyle choices across education groups (row 3) because some differences across education groups remain; in particular, the earnings losses due to bad health and the health dynamics conditional on lifestyle are slightly different across education groups.

TABLE 7: Policy Counterfactuals

	Pr(e)		Pr(PRO)	Pr(PRO|e)			LE(e)			
	CG	HSD		CG	HSD	CG-HSD	All	CG	HSD	CG-HSD
Benchmark	0.36	0.10	0.73	0.87	0.47	0.39	28.0	31.2	23.4	7.8
Lower μ_{PRO}	0.36	0.09	0.86	0.95	0.67	0.27	29.1	31.9	25.1	6.8
Diff	0.00	-0.01	0.13	0.08	0.20	-0.12	1.1	0.7	1.7	-1.0
Lower μ_e	0.55	0.04	0.74	0.84	0.42	0.42	28.8	31.0	22.9	8.1
Diff	0.19	-0.06	0.01	-0.03	-0.05	0.03	0.8	-0.2	-0.5	0.3

Notes: The first row reports the benchmark model. The second row reduces the cost of *protective* behavior by one-half of its standard deviation. The fourth row reduces the costs of completing high school and college by one-half of their respective standard deviations. Both education and lifestyle choices are allowed to adjust in the two policy experiments. Difference rows are relative to the benchmark.

to 61% among high school dropouts. The education gradient in lifestyle therefore would fall from 39 to 20 percentage points, while the education gradient in life expectancy would fall from 7.8 to 6.1 years. Educational attainment changes little. This exercise shows that primitive correlated heterogeneity contributes substantially to the observed sorting, while a sizable gradient remains through differences in economic prospects and endogenous selection generated by complementarity.

The literature on the early-age formation of the traits that jointly govern schooling and health has also reached our conclusion about the importance of selection. For instance, Conti et al. (2010) find that selection on factors determined early in life accounts for more than half of the education differential in poor health, depression, and obesity; see also Conti and Heckman (2010) and Heckman et al. (2018). These papers identify selection by measuring early traits directly and conditioning on them. We take the complementary route: we do not observe the traits, but we recover the joint cost distribution that rationalizes observed education and lifestyle choices. In Section 6.2.2 we showed that this distribution lines up with traits measured independently in the data. Our exercise in this Section allows us to separate the importance of correlated early life traits ($\rho > 0$) from endogenous selection through the complementarity of both investments.

7.2 Policy Implications

Our model features dispersion in utility costs of education and lifestyle as a key determinant of education and lifestyle choices. Therefore, policies aimed at reducing these costs may encourage higher education attainment, better lifestyles, and higher life expectancy. Furthermore, because the two investments are complementary, a policy directed at either early-life investment can affect the outcome of both of them.

We next use the model to simulate a policy aimed at reducing the average cost of a *protective* lifestyle. This is more a proof of concept than an actual policy; but one may think this policy could capture exposure to a healthier menu in primary school, an increase in sports activities in childhood, or a reduction in smoking among parents. In our implementation, we reduce μ_{PRO} by one-half of the

TABLE 8: Individual responses to the policy counterfactuals

Benchmark education	Responses (%)		Life expectancy (years)		
	Switch to PRO	Move to higher education	Benchmark	Policy	Change
Panel A: Lower cost of protective behavior					
HSD	22.1	6.1	23.4	25.3	1.9
HSG	15.7	1.2	26.7	28.0	1.3
CG	8.0	0.0	31.2	31.9	0.7
CG-HSD	-14.2	-6.1	7.8	6.6	-1.2
Panel B: Lower cost of education					
HSD	2.1	61.9	23.4	24.5	1.1
HSG	2.6	35.6	26.7	28.0	1.3
CG	0.0	0.0	31.2	31.2	0.0
CG-HSD	-2.1	-61.9	7.8	6.7	-1.1

Notes: Panel A lowers the cost of adopting a *protective* lifestyle by one-half of its standard deviation. Panel B lowers the costs of completing high school and college by one-half of their respective standard deviations. "Switch to PRO" reports the percentage of individuals in each benchmark education group who switch from a *detrimental* to a *protective* lifestyle. "Move to higher education" reports the percentage who move to a higher education category. Life expectancy is reported for individuals grouped by their benchmark education.

standard deviation of the distribution of τ_{PRO}. We find that the aggregate fraction of individuals with a *protective* lifestyle rises from 73% to 86%, see rows 1 and 2 in Table 7. Complementarity generates an additional response on the education margin, with the share of high school dropouts falling by 1 percentage point and the share of college graduates rising slightly, despite no changes in the costs of education.[20] Overall, due to these two changes, life expectancy rises by 1.1 years (from 28.0 to 29.1 years).

Interestingly, the health gains from this intervention are largest among the least advantaged. Among benchmark high school dropouts (individuals who dropped out of high school without the intervention), 22.1% switch from a *detrimental* to a *protective* lifestyle, compared with 15.7% of benchmark high school graduates and 8.0% of benchmark college graduates, see Panel A in Table 8. As a result, the *ex ante* education gradient in life expectancy narrows from 7.8 to 6.6.[21] At the same time, the increase in education triggered by a reduction in the cost of a *protective* lifestyle is also asymmetric in the population: 6.1% of benchmark high school dropouts and 1.2% of benchmark high school graduates move to a higher education category. This worsens the distribution of τ_{PRO} in all education groups, but more so for high school dropouts, and as a result, the *ex post* educational gradient of life expectancy is mitigated by selection, falling from 7.8 to 6.8 instead of 6.6 years, see again rows 1 and 2 in Table 7.

[20] The rise in education achievement following a reduction in the cost of undertaking healthy lifestyles is consistent with some recent empirical evidence. Cappelen et al. (2026) shows how randomized access to a free gym card increases academic outcomes of university students in the US, while Bagues and Villa (2025) finds that increasing the minimum legal drinking age in Spain raises education outcomes of teenagers.

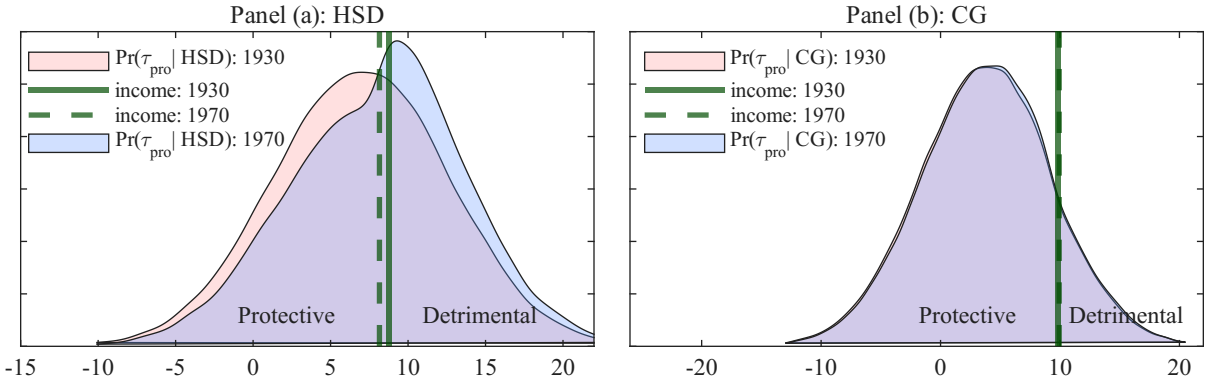
[21] We call the *ex ante* education gradient in life expectancy the difference in life expectancy with the policy intervention based on the education groups without the intervention.

As a comparison, we next reduce the costs of completing high school and college education by one-half of their respective standard deviations. The direct education response is large: college attainment rises from 36% to 55%, while the high school dropout share falls from 10% to 4%. Complementarity also produces a response on the margin not directly targeted by the policy, with the fraction of *protective* individuals increasing by 1 percentage point. As a result, average life expectancy increases by 0.8 years (from 28.0 to 28.8), see rows 1 and 4 in Table 7. This result is consistent with the empirical evidence showing a causal effect of education on health outcomes (e.g. Lleras-Muney, 2005). In our model, this happens for two reasons. First, because more educated individuals experience better health dynamics conditional on lifestyle (this is the estimated direct education effect in $p\left(\mathbf{h}_i^T \mid \mathbf{ins}_i^T, x_i, y; \boldsymbol{\theta}_y\right)$, see Section 3.2). Second, because some of the individuals increasing education also choose better lifestyles, which additionally increases their life expectancy.

The distributional effects of this policy are also very interesting. The increase in education is very large among benchmark high school dropouts (62% of them increase their educational attainment), which, through complementarity, also raises the share of *protective* individuals within this group by 2.1 percentage points. Overall, this leads to an increase in life expectancy of benchmark high school dropouts of 1.1 years (from 23.4 to 24.5 years), see Panel B in Table 8. In contrast, by definition there is no educational improvement among benchmark college graduates, and hence there is no change in the lifestyle composition of this group and no change in their life expectancy either. As a result, there is a reduction in the *ex ante* education gradient of life expectancy of 1.1 years. However, a different picture emerges when looking at the attained (or *ex post*) education gradient of life expectancy. The share of *protective* individuals falls from 47% to 42% among high school dropouts and from 87% to 84% among college graduates. This pattern reflects selection induced by the policy. Individuals who are induced to acquire more education are less likely to have a *protective* lifestyle than incumbent members of their destination education group, while the smaller group that remains among high school dropouts becomes more negatively selected. As a result, the *ex post* education gradient in lifestyle rises by 3 percentage points (from 39% to 42%), and the *ex post* education gradient in life expectancy rises by 0.3 years (from 7.8 to 8.1 years), see again rows 1 and 4 in Table 7. Therefore, selection reverses the reduction in the *ex ante* education gradient of life expectancy generated by the education policy, and hence, it masks a reduction in health inequalities.

The comparison delivers two policy lessons. First, complementarity makes policies more powerful than their direct target suggests: lifestyle interventions can promote education, and education interventions can improve health. Second, whether these spillovers reduce measured health inequality depends on the selection they generate. Lowering the cost of a *protective* lifestyle disproportionately benefits the initially least educated and narrows the education gradient in life expectancy. Lowering education costs also improves individual health outcomes, but the resulting reallocation across education groups widens the observed gradient. Evaluating these policies therefore requires distinguishing changes in individuals' outcomes from changes in inequality across groups whose composition is itself affected by the policy.

FIGURE 11: Cost of *protective* lifestyle across cohorts



Notes: In Panel (a), the red shaded area describes the density function of τ_{PRO} for individuals in the 1930s cohort who chose HSD education. The solid green vertical line reports the threshold $\tau^*(\text{HSD})$ that separates HSD individuals into *detrimental* and *protective* in the 1930s cohort. The dashed green vertical line reports this same threshold if the income prospects of HSD individuals were as in the 1970s cohort. The dashed shaded blue area describes the density function of τ_{PRO} for individuals who choose HSD education in the 1970s cohort. Panel (b) reports the same information for CG individuals.

7.3 Changes over time

The econometric model estimated in Section 4 shows that the life expectancy gap between college graduates and high school dropouts increased by 1.0 year between the 1930s and 1970s cohorts. This increase is associated with an important decline in the prevalence of *protective* lifestyles among high school dropouts and a small increase among college graduates. The calibrated model generates a slightly larger increase of 1.2 years. In this Section, we quantify how much of the increase in the education gradient of life expectancy is due to the direct effect of diverging income prospects across education groups and how much is due to the induced changes in the composition of individuals in each education group.

To quantify the direct effect of income changes, we fix the distributions of lifestyle costs conditional on education to the ones for the 1930s cohort. We then analyze how investments in lifestyle would have changed solely due to changes in income prospects across education groups. Recall that in our calibration, expected lifetime earnings increase by 8% for college graduates and decrease by 23% for high school dropouts.

In Figure 11, the shaded red areas represent the distributions $\Pr(\tau_{\text{PRO}}|\text{HSD})$ and $\Pr(\tau_{\text{PRO}}|\text{CG})$ of the cost of *protective* lifestyle τ_{PRO} faced by individuals in the 1930s cohort for high school dropouts (Panel a) and college graduates (Panel b), respectively. These two distributions are different due to selection, see Proposition 1. The solid green vertical lines indicate the threshold values $\tau^*(\text{HSD})$ and $\tau^*(\text{CG})$ that separate individuals of that cohort into *protective* and *detrimental* lifestyles conditional on each education choice. The changes in income prospects are represented by the dashed green vertical lines. For high school dropouts, the threshold value $\tau^*(\text{HSD})$ decreases across cohorts due to the decline in expected labor earnings, and consequently, a smaller fraction of individuals dropping out of high school invest in their health. In particular, the share of high school dropouts choosing a

protective lifestyle declines by 4 percentage points, compared with a decline of 15 percentage points in the complete model, see the second and third rows in Table 9. As a result, the life expectancy of high school dropouts falls by 0.3 years, compared with 1.2 years in the model. Instead, the threshold value $\tau^*(\text{CG})$ for college graduates increases due to the increase in expected lifetime income, although the actual increase is rather small. As a consequence, the fraction of individuals with a *protective* lifestyle among college graduates only increases by 1 percentage point. Overall, the direct effect of changing income prospects increases the education gradient in life expectancy by 0.4 years. This represents approximately one-third of the 1.2-year increase generated by the model.

Next, selection into each education category changes with income changes for two reasons. First, the increase in the education wage premium shrinks the group of high school dropouts and increases the size of the other two. This leads to a worsening of the distribution of τ_{PRO} in all education groups, as the marginal individuals who chose HSD in the 1930s cohort chose higher education in the 1970s. Second, the increase in income inequality across education groups increases the complementarity between education and lifestyles, which leads to a worsening of the distribution of τ_{PRO} among individuals choosing HSD and an improvement among individuals choosing CG. The overall effect on the conditional distributions of τ_{PRO} can be observed in Figure 11, where the shaded blue areas in each panel represent the distributions $\Pr(\tau_{\text{PRO}}|\text{HSD})$ and $\Pr(\tau_{\text{PRO}}|\text{CG})$ for the 1970s cohort. The distribution of τ_{PRO} worsened for high school dropouts (it shifted to the right) as both effects are pushing in the same direction, while it did not change significantly for the college graduates, as each effect pushes in opposite directions. As reported in Table 9, changes in selection reduce the share of high school dropouts choosing a *protective* lifestyle by 10 percentage points, compared with a decline of 15 percentage points in the complete model. Their life expectancy falls by 0.9 years, compared with 1.2 years in the model. With essentially no change among college graduates, selection increases the education gradient in life expectancy by 0.8 years. This represents approximately two-thirds of the increase generated by the model.

Therefore, we conclude that the increase in the gap in earnings prospects between different education groups can explain the increase in the educational gradient in life expectancy across cohorts in the US. However, only about 1/3 is due to the direct effect of income prospects on lifestyle choices. The other 2/3 come from the worsening of the pool of high school dropouts due to the changes in education choices induced by the wage changes.

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*. More educated individuals choose *protective* behavior more frequently, and differences in lifestyles explain 42% of

TABLE 9: Decomposition: changes across cohorts

	Δ_cPr(e)		Δ_cPr(PRO|e)			Δ_cLE(e)		
	CG	HSD	CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.09	-0.17	0.03	-0.08	0.11	0.3	-0.7	1.0
Model	0.09	-0.13	0.00	-0.15	0.15	0.0	-1.2	1.2
Income			0.01	-0.04	0.04	0.1	-0.3	0.4
Selection			-0.00	-0.10	0.10	-0.0	-0.9	0.8

Notes: The first row reports statistics from the data, comparing the cohorts of 1930s and 1970s. The second row reports the results from the calibrated model. The 3rd row performs the counterfactual exercise where only income prospects change between the 1930s and 1970; the 4th row performs the counterfactual exercise where only the distribution of τ_{PRO} conditional on education changes between the 1930s and 1970s. See Section 7.3 for details.

the education gradient in life expectancy. We also find an increasing education gradient in life expectancy between the 1930s and 1970s cohorts, which is mostly driven by worsening lifestyles among the less educated.

To understand the relationship between education and lifestyles, we develop a heterogeneous-agents model with two stages. In the first stage, individuals jointly choose their education and lifestyle, subject to heterogeneous costs of both investments. In the second stage, they solve a life-cycle consumption and saving problem subject to income and health risks. Higher expected income among the more educated increases the value of life and therefore the return to adopting a *protective* lifestyle, making education and lifestyle investments complementary. This complementarity, together with the positive correlation between the unobserved costs of the two investments, generates endogenous selection: individuals with lower costs of adopting a *protective* lifestyle are disproportionately represented among the more educated. Our counterfactual exercises show that both income differences and this selection mechanism contribute substantially to the education gradient in lifestyles and life expectancy.

The policy experiments illustrate a related implication of the complementarity and selection mechanisms. Reducing the cost of adopting a *protective* lifestyle generates large health gains, narrows the education gradient in life expectancy, and induces some individuals to acquire more education. By contrast, an equally sized standardized reduction in education costs substantially expands educational attainment but widens the measured health gradient across education groups by changing their composition. Policies affecting lifestyle formation early in life may therefore produce educational gains in addition to their direct health benefits (Bagues and Villa, 2025; Cappelen et al., 2026). More generally, changes in health gaps across education categories need not measure the individual health effects of education policies. Because lifestyles are permanent in the model, these results should be interpreted as mechanisms rather than as comparisons of particular interventions.

Finally, between the 1930s and 1970s cohorts, the model generates a 1.2-year increase in the education gradient in life expectancy, compared with an increase of 1.0 year in the data. Changes in

income prospects directly generate 0.4 year of the model-implied increase, while changes in selection across education groups generate the remaining 0.8 year.

While the current model closely reproduces the education gradient in life expectancy, it treats the correlation ρ between the unobserved costs of education and a *protective* lifestyle as a reduced-form primitive. The human-capital literature on early-life capability formation provides the natural microfoundation (Cunha and Heckman, 2007; Conti and Heckman, 2010) and Section 6.2.2 shows that its observable correlates behave as that account predicts. Embedding an explicit technology of capability formation in the first stage, so that ρ is an outcome rather than a parameter, is a natural next step. The model also assumes that individuals understand the long-run health consequences of their lifestyle choices. Future research could explicitly model the forces generating the correlation between education and lifestyle costs and allow health knowledge to vary across cohorts and education groups, as suggested by Cutler and Lleras-Muney (2010).

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 pages 4 and 23.)

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.)

Bagues, M. and C. Villa (2025): "Minimum Legal Drinking Age and Educational Outcomes," *Journal of Health Economics*, 104, 103078. (Cited on pages 33 and 37.)

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 (2026): "Exercise or Extra Fries? Behavioral Drivers of Obesity Over the Life Cycle," *Journal of the European Economic Association*, forthcoming. (Cited on pages 4 and 6.)

Becker, G. S. (2007): "Health as Human Capital: Synthesis and Extensions," *Oxford Economic Papers*, 59, 379–410. (Cited on pages 2 and 29.)

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 (2025): "Health Inequality and Health Types," *The Econometrics Journal*, 28, 341–384. (Cited on pages 4 and A.10.)

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.)

Capatina, E. and M. Keane (2026): "Health Shocks, Health Insurance, Human Capital, and the Dynamics of Earnings and Health," *Journal of Political Economy*, 134, 1714–1774. (Cited on page 4.)

Cappelen, A. W., G. Charness, M. Ekstrom, U. Gneezy, and B. Tungodden (2026): "Exercise Improves Academic Performance," *Journal of Political Economy*, 134, 397–434. (Cited on pages 33 and 37.)

Card, D., C. Dobkin, and N. Maestas (2009): "Does Medicare Save Lives?" *Quarterly Journal of Economics*, 124, 597–636. (Cited on page 4.)

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 5.)

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," *Review of Economic Studies*, 86, 1123–1169. (Cited on pages 4 and 29.)

Conner, M. and P. Norman (2017): "Health Behaviour: Current Issues and Challenges," *Psychology & Health*, 32, 895–906. (Cited on page 15.)

Conti, G. and J. J. Heckman (2010): "Understanding the Early Origins of the Education-Health Gradient," *Perspectives on Psychological Science*, 5, 585–605. (Cited on pages 15, 32, and 38.)

Conti, G., J. J. Heckman, and S. Urzua (2010): "The Education-Health Gradient," *American Economic Review*, 100, 234–238. (Cited on page 32.)

Cunha, F. and J. J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47. (Cited on pages 15 and 38.)

Cutler, D. M. and E. Glaeser (2005): "What Explains Differences in Smoking, Drinking, and Other Health Related Behaviors?" *American Economic Review*, 95, 238–242. (Cited on page 6.)

Cutler, D. M. and A. Lleras-Muney (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29, 1–28. (Cited on pages 1, 6, and 38.)

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 (2025): "The Lifetime Costs of Bad Health," *Review of Economic Studies*, 92, 1987–2026. (Cited on pages 4, 24, A.18, and A.19.)

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society: series B (methodological)*, 39, 1–22. (Cited on page 9.)

Farrell, P. and V. Fuchs (1982): "Schooling and Health: the Cigarette Connection," *Journal of Health Economics*, 1, 217–230. (Cited on page 15.)

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.)

Galama, T. J. and H. van Kippersluis (2019): "A Theory of Socio-economic Disparities in Health over the Life Cycle," *Economic Journal*, 129, 338–374. (Cited on page 4.)

Hai, R. and J. J. Heckman (2026): "The Causal Effects of Youth Cigarette Addiction and Education," *Journal of Political Economy*, 134, 1347–1400. (Cited on pages 5 and 15.)

Hall, R. and C. Jones (2007): "The Value of Life and the Rise in Health Spending," *Quarterly Journal of Economics*, 122, 39–72. (Cited on page 24.)

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 26.)

Heathcote, J., K. Storesletten, and G. L. Violante (2020): "Optimal Progressivity with Age-Dependent Taxation," *Journal of Public Economics*, 189, 104074. (Cited on page 19.)

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018): "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking," *Journal of Political Economy*, 126, S197–S246. (Cited on page 32.)

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 A.20.)

Hosseini, R., K. A. Kopecky, and K. Zhao (2022): "The Evolution of Health over the Life Cycle," *Review of Economic Dynamics*, 45, 237–263. (Cited on page 4.)

——— (2026): "How Important is Health Inequality for Lifetime Earnings Inequality?" *Review of Economic Studies*, 93, 556–599. (Cited on pages 4 and 21.)

Jayachandran, S. and A. Lleras-Muney (2009): "Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines," *Quarterly Journal of Economics*, 124, 349–397. (Cited on page 17.)

Jones, C. and P. Klenow (2016): "Beyond GDP? Welfare across Countries and Time," *American Economic Review*, 106, 2426–2457. (Cited on page 24.)

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 pages 1 and 6.)

Lee, J. J., R. Wedow, A. Okbay, and E. Kong (2018): "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals," *Nature Genetics*, 50, 1112–1121. (Cited on page 28.)

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.)

Liu, M., Y. Jiang, R. Wedow, and Y. Li (2019): "Association Studies of Up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use," *Nature Genetics*, 51, 237–244. (Cited on page 28.)

Lleras-Muney, A. (2005): "The Relationship Between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72, 189–221. (Cited on pages 17 and 34.)

Mahler, L. and M. Yum (2024): "Lifestyle Behaviors and Wealth-Health Gaps in Germany," *Econometrica*, 92, 1697–1733. (Cited on pages 4 and 29.)

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 pages 4 and 29.)

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.)

Miller, S., E. Rhodes, A. W. Bartik, D. E. Broockman, P. K. Krause, and E. Vivalt (2026): "Does Income Affect Health? Evidence from a Randomized Controlled Trial of a Guaranteed Income," NBER Working Paper 32711. (Cited on page 15.)

Murphy, K. M. and R. H. Topel (2006): "The Value of Health and Longevity," *Journal of Political Economy*, 114, 871–904. (Cited on page 24.)

Nakajima, M. and I. A. Telyukova (2025): "Medical Expenses and Saving in Retirement: The Case of U.S. and Sweden," *American Economic Journal: Macroeconomics*, 17, 161–202. (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 5.)

Oster, E., I. Shoulson, and E. R. Dorsey (2013): "Limited Life Expectancy, Human Capital and Health Investments," *American Economic Review*, 103, 1977–2002. (Cited on page 17.)

Ozkan, S. (2025): "Income Differences and Health Disparities: Roles of Preventive vs. Curative Medicine," *Journal of Monetary Economics*, 150, 103698. (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 9.)

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.)

Rosen, S. (1988): "The Value of Changes in Life Expectancy," *Journal of Risk and Uncertainty*, 1, 285–304. (Cited on page 24.)

Rousseeuw, P. J. (1987): "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, 20, 53–65. (Cited on page A.10.)

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.)

Thorndike, R. L. (1953): "Who Belongs in the Family?" *Psychometrika*, 18, 267–276. (Cited on page A.10.)

van Kippersluis, H., P. Biroli, R. D. Pereira, T. J. Galama, S. von Hinke, S. F. W. Meddens, D. Muslimova, E. A. W. Slob, R. de Vlaming, and C. A. Rietveld (2023): "Overcoming Attenuation Bias in Regressions Using Polygenic Indices," *Nature Communications*, 14, 4473. (Cited on page 28.)

van Kippersluis, H. and T. J. Galama (2014): "Wealth and Health Behavior: Testing the Concept of a Health Cost," *European Economic Review*, 72, 197–220. (Cited on page 15.)

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

Table of Contents

A Estimation details of the health model **A.2**

A.1 Health Behaviors A.2

A.2 Health Dynamics A.2

A.3 Mixture Weights A.3

A.4 Gibbs sampler A.4

B Alternative estimations **A.7**

B.1 More than 2 types A.7

B.2 Choosing the number of types A.10

B.3 Lifestyle types by smoking A.11

C Theoretical results of the first-stage model **A.13**

D First step estimation details **A.17**

D.1 Income process A.17

D.2 Medical shocks and insurance A.18

D.3 Tax system A.19

D.4 Value of statistical life A.20

E Robustness exercises **A.24**

F Extra figures and tables **A.30**

Appendix A: Estimation details of the health model

In this appendix, we explain the details of the estimation of the health model in Section 3. Sections A.1 to A.3 give details on how we model the three different elements of the likelihood function in equation (1), while Section A.4 explains the estimation procedure.

In terms of notation, we break the parameter space into three different types of parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$, where $\boldsymbol{\beta}$ denotes the parameters associated with health transitions, $\boldsymbol{\gamma}$ denotes the parameters associated with health behaviors, and $\boldsymbol{\delta}$ denotes the parameters associated with the probability of having lifestyle y at the initial age.

A.1 Health Behaviors

Given lifestyle $\mathrm{y}_i = y$, age a_{it}, insurance status ins_{it}, and health h_{it}, we model the probability of an individual i at period t displaying health behavior m ($z_{m,it} = 1$) as the probability that $z^{\star}_{m,it} > 0$, where

$$z^{\star}_{m,it} \equiv \gamma_{1,m,y} + \gamma_{2,m,y} a_{it} + \gamma_{3,m,y} a_{it}^2 + \gamma_{4,m,y} I(h_{it} = h_b) + \gamma_{5,m,y} \times \mathrm{ins}_{it}^{(m)} + \epsilon_{m,it}, \quad \epsilon_{m,it} \sim N(0,1)$$

Hence,

$$\begin{aligned} p(z^{*}_{m,it} > 0 | h_{it}, \mathrm{ins}_{it}^{(m)}, a_{it}, y) &= \Phi\Big(\gamma_{1,m,y} + \gamma_{2,m,y} a_{it} + \gamma_{3,m,y} a_{it}^2 + \gamma_{4,m,y} I(h_{it} = h_b) + \gamma_{5,m,y} \times \mathrm{ins}_{it}^{(m)}\Big) \\ &\equiv \alpha(h_{it}, \mathrm{ins}_{it}^{(m)}, a_{it}, y; \boldsymbol{\gamma_{y,m}}) \end{aligned}$$

where Φ is the cdf of the normal distribution and $\boldsymbol{\gamma_{y,m}} \subset \boldsymbol{\gamma}$ is the relevant subset of parameters. Then, the term $p(z_{m,it} | h_{it}, ins_{it}, a_{it}, y; \boldsymbol{\theta}_y)$ in equation (2) is given by,

$$p(z_{m,it} | h_{it}, ins_{it}, a_{it}, y; \boldsymbol{\gamma_{y,m}}) = \alpha(h_{it}, ins_{it}, a_{it}, y; \boldsymbol{\gamma_{y,m}})^{z_{m,it}} \Big(1 - \alpha(h_{it}, ins_{it}, a_{it}, y; \boldsymbol{\gamma_{y,m}})\Big)^{1 - z_{m,it}}$$

We only allow the insurance status to affect the probability of health behaviors associated to access to medical care (Flu shot, Cholesterol test and Cancer test) setting $\gamma_{5,m,y} = 0$ for drinking, smoking and exercise.

A.2 Health Dynamics

We model the term $p(h_{i,t+1} | \mathrm{e}_i, a_{i,t}, h_{i,t}, y; \boldsymbol{\theta}_y)$ in equation (3) reflecting health dynamics of an individual i at period t as a nested probit model. In the first nest, individuals are exposed to a survival/mortality shock. Then, conditional on surviving, individuals suffer a good/bad health shock. We thus partition the set of parameters of health transitions $\boldsymbol{\beta} = \{\boldsymbol{\beta}^s, \boldsymbol{\beta}^h\}$ where $\boldsymbol{\beta}^s$ denotes parameters driving survival probabilities and $\boldsymbol{\beta}^h$ drive health transition probabilities given survival.

In the first nest, an individual i with education $\mathrm{e}_i = e$, lifestyle type $\mathrm{y}_i = y$, health $h_{it} = h$, and age a_{it} at time t will survive ($h_{i,t+1} \neq h_d$) if $h^*_{s,i,t+1} > 0$:

$$h^\star_{s,i,t+1} = \beta^s_{1,h,e,y} + \beta^s_{2,h,e,y}\, a_{it} + \epsilon_{s,it}, \quad \epsilon_{s,it} \sim N(0,1)$$

In the second nest, conditional on survival, an individual can transition into good or bad health. An individual will transition to good health ($h_{i,t+1} = h_g$) if $h^*_{g,i,t+1} > 0$:

$$h^\star_{g,i,t+1} = \beta^h_{1,h,e,y} + \beta^h_{2,h,e,y}\, a_{it} + \epsilon_{g,it}, \quad \epsilon_{g,it} \sim N(0,1)$$

Therefore we can write each probability $p(h_{i,t+1}|\mathrm{e}_i, a_{i,t}, h_{i,t}, y; \boldsymbol{\theta}_y)$ in equation (3) as:

$$\begin{aligned} p(h_{i,t+1}|\mathrm{e}_i, a_{i,t}, h_{i,t}, y; \boldsymbol{\theta}_y) \;=\; & \Phi\left(f(\mathrm{e}_i, a_{it}, h_{it}, y; \boldsymbol{\beta}^h)\right)^{I(h_{i,t+1}=h_g)} \\ \times \; & \left[1 - \Phi\left(f(\mathrm{e}_i, a_{it}, h_{it}, y; \boldsymbol{\beta}^h)\right)\right]^{I(h_{i,t+1}=h_b)} \\ \times \; & \Phi\Big(f(\mathrm{e}_i, a_{it}, h_{it}, y; \boldsymbol{\beta}^s)\Big)^{I(h_{i,t+1}\neq h_d)} \\ \times \; & \left[1 - \Phi\Big(f(\mathrm{e}_i, a_{it}, h_{it}, y; \boldsymbol{\beta}^s)\Big)\right]^{I(h_{i,t+1}=h_d)} \end{aligned}$$

where $f(\mathrm{e}, a, h, y; \boldsymbol{\beta}) = \beta_{1,h,e,y} + \beta_{2,h,e,y}\, a$.

A.3 Mixture Weights

As discussed in Section 3.3, the mixture weights $p\left(\mathrm{y}|\mathrm{c}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right)$ differ across entry age a_{i0} given the differential health transitions and mortality conditional on type y. The way they do is determined by the model of health dynamics in equation (3). In particular, let's write,

$$p\left(\mathrm{y}|\mathrm{c}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right) = \frac{p\left(\mathrm{y}, h_{i0}|\mathrm{c}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right)}{\sum_{y=1}^{N_y} p\left(y, h_{i0}|\mathrm{c}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right)} \tag{A.1}$$

The joint probability $p\left(\mathrm{y}, h_{i0}|\mathrm{c}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right)$ can be decomposed as,

$$\begin{aligned} p\left(\mathrm{y}, h_{i0} \mid \mathrm{c}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right) &= \sum_{h_{i(-1)} \in H} p\left(\mathrm{y}, h_{i0} \mid \mathrm{c}_i, \mathrm{e}_i, a_{i0}, h_{i(-1)}; \boldsymbol{\beta}, \boldsymbol{\delta}\right) p\left(h_{i(-1)} \mid \mathrm{c}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right) \\ &= \sum_{h_{i(-1)} \in H} p\left(h_{i0} \mid h_{i(-1)}, \mathrm{c}_i, \mathrm{e}_i, \mathrm{y}, a_{i0}; \boldsymbol{\beta}\right) p\left(\mathrm{y}, h_{i(-1)} \mid \mathrm{c}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right), \end{aligned} \tag{A.2}$$

where we integrate over all possible health states in the period before the individual enters our sample ($h_{i(-1)}$). The first term in the right-hand side of equation (A.2) describes the health dynamics, and it has been discussed in Section 3.2, see equation (3). The second term on the right-hand side is the same as the left-hand side, just one period before. Hence, we can use equation (A.2) recursively

up to some common initial age, which we choose to be age 26. This gives us $p\left(\mathrm{y}, h_{i,26}|\mathrm{c}_i, \mathrm{e}_i, 26; \boldsymbol{\delta}\right)$, the joint probability of an individual i in cohort c_i and education e_i of being of health $h_{i,26}$ and type y at age 26. We decompose this probability in two pieces:

$$p\left(\mathrm{y}, h_{i,26}|\mathrm{c}_i, \mathrm{e}_i, 26; \boldsymbol{\delta}\right) = p\left(\mathrm{y}|\mathrm{c}_i, \mathrm{e}_i, 26, h_{i,26}; \boldsymbol{\delta}\right) p\left(h_{i,26}|\mathrm{c}_i, \mathrm{e}_i, 26\right)$$

The second term in the right-hand side describes the share of individuals of age 26 of given cohort c_i and education e_i that have health $h_{i,26}$. This can be measured directly in the PSID for many but not all cohorts. We thus assume that conditional on education, the probability of good and bad health at age 26 does not change across cohorts, and we set it equal to the one that we observe in the PSID.[22] The first term on the right-hand side describes the fraction of individuals of age 26, cohort c_i, education e_i, and health $h_{i,26}$ who are of type y_i. This is the probability that we have to estimate, and we model it through a multinomial probit.

In order to compute $p(\mathrm{y}|\mathrm{c}_i, \mathrm{e}_i, 26, h_{i,26}; \boldsymbol{\delta}, \boldsymbol{\beta})$ we need to specify the probability of the lifestyle at the initial age. For this purpose we define:

$$\begin{aligned} y^*_{1,i} &= \delta^1_{0,e} + \delta^1_{1,e}\mathbb{1}_{h=h_g} + \sum_c \delta^1_{2,e,c}\mathbb{1}_{\mathrm{c}_i=c} + \epsilon_{1,i} \\ &\vdots \\ y^*_{N_y,i} &= \delta^{N_y}_{0,e} + \delta^{N_y}_{1,e}\mathbb{1}_{h=h_g} + \sum_c \delta^{N_y}_{2,e,c}\mathbb{1}_{\mathrm{c}_i=c} + \epsilon_{N_y,i} \end{aligned}$$

where $\epsilon_{1,i}, ..., \epsilon_{N_y,i}$ are independently normally distributed across lifestyles with mean 0 and variance 1.

$$y_i = \begin{cases} 1 \text{ if } y^*_{1,i} > y^*_{2,i}, \ldots, y^*_{N_y,i} \\ 2 \text{ if } y^*_{2,i} > y^*_{1,i}, \ldots, y^*_{N_y,i} \\ \vdots \\ N_y \text{ otherwise} \end{cases}$$

A.4 Gibbs sampler

Estimating the mixture model poses a challenge because the health behavior type, y, which classifies individuals into different lifestyle categories, is unobserved. This means that the likelihood function must account for all possible assignments of y, making direct estimation computationally complex. A particularly effective solution in a Bayesian framework is *data augmentation*, which introduces y as an auxiliary variable and allows us to estimate the model using Markov Chain Monte Carlo (MCMC) methods.

The key idea behind data augmentation is to treat y as missing data and iteratively infer it

[22] The probability of being in good health at age 26 varies between 85% for dropout to 98% for college graduates.

alongside the model parameters. In a Bayesian setting, we sample from the posterior distribution of y given the observed health behaviors and health outcomes. This imputed dataset is then used to update the posterior distributions of the model parameters. Specifically, given prior distributions for the parameters, we construct a Gibbs sampler where we alternate between:

i) drawing lifestyle assignments y_i from their conditional posterior distribution, and

ii) updating the parameters given the imputed values of y

We can write the complete data likelihood as:

$$p\left(\mathbf{y}, \mathbf{h}^T, \mathbf{z}^T | \mathbf{ins}^T, \mathbf{c}, \mathbf{e}, \boldsymbol{a}_0, \boldsymbol{h}_0; \boldsymbol{\theta}\right) = \prod_{i=1}^{N} \prod_{y=1}^{N_y} \Big[p\left(\mathbf{z}_i^T | y, a_{i0}, \mathbf{h}_i^T, \mathbf{ins}_i^T; \boldsymbol{\gamma}_y\right) \\ p\left(\mathbf{h}_i^T | \mathrm{e}_i, y, a_{i0}, h_{i0}; \boldsymbol{\beta}_y\right) \\ p\left(y | \mathrm{c}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}_y, \boldsymbol{\delta}_y\right) \Big]^{\mathbb{1}_{\mathrm{y}_i = y}}$$

where y_i denotes the lifestyle assigned to individual i and $\boldsymbol{h}^T, \boldsymbol{z}^T$ and $\mathbf{y}$ denote health status, health behaviors, and assigned lifestyles for every individual in the sample, respectively. We can now write the joint posterior distribution as:

$$p\left(\boldsymbol{\theta} | \mathbf{ins}^T, \mathbf{c}, \mathbf{e}, \mathbf{y}, \boldsymbol{a}_{i0}, \mathbf{h}_0, \mathbf{h}^T, \mathbf{z}^T\right) \propto \prod_{i=1}^{N} \prod_{y=1}^{N_y} \Big[p\left(\mathbf{z}_i^T | a_{i0}, y, \mathbf{ins}_i^T, h_{i0}, \mathbf{h}_i^T; \boldsymbol{\gamma}_y\right) \\ p\left(\mathbf{h}_i^T | \mathrm{e}_i, y, a_{i0}, h_{i0}; \boldsymbol{\beta}_y\right) \\ p\left(y | \mathrm{c}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}_y, \boldsymbol{\delta}_y\right) \Big]^{\mathbb{1}_{\mathrm{y}_i = y}} \; p(\boldsymbol{\theta}).$$

To estimate the parameters of the model, we implement a Gibbs sampler that sequentially updates different sets of parameters by drawing from their respective conditional posterior distributions. The sampler consists of four blocks: (i) *Transition parameters*, which govern the evolution of health status over time and are updated based on the observed health trajectories given the lifestyle; (ii) *Health behavior parameters*, which capture the relationship between lifestyles and observed health behaviors; (iii) *Mixture weights*, which determine the prior probability of assignment to each lifestyle type given individual characteristics; and (iv) *Latent lifestyle*, where we impute the unobserved health behavior type y for each individual based on the current parameter estimates. By iterating over these blocks, the algorithm ensures convergence to the joint posterior distribution of the model parameters and latent variables.

Block 1: transition parameters The first block is given by

$$p\left(\boldsymbol{\beta}|\mathbf{y},\mathbf{c},\mathbf{e},\boldsymbol{a}_0,\boldsymbol{h}_0,\mathbf{h}^T,\mathbf{z}^T;\boldsymbol{\delta},\boldsymbol{\gamma}\right) \propto \prod_{y=1}^{N_y}\prod_{i:\mathrm{y}_i=y} p\left(\mathbf{h}_i^T|\mathrm{e}_i,y,a_{i0},h_{i0};\boldsymbol{\beta}_y\right) p\left(y|\mathrm{c}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y,\boldsymbol{\delta}_y\right) p(\boldsymbol{\beta}_y)$$

Using flat prior in $\boldsymbol{\beta}$, we sample from the posterior distribution using a Metropolis-within-Gibbs algorithm.

Block 2: health behaviors parameters The second block is given by:

$$p\left(\boldsymbol{\gamma}|\mathbf{ins}^T,\mathbf{c},\mathbf{e},\mathbf{y},\boldsymbol{a}_0,\mathbf{h}^T,\mathbf{z}^T;\boldsymbol{\delta},\boldsymbol{\beta}\right) \propto \prod_{y=1}^{N_y}\prod_{i:\mathrm{y}_i=y}\prod_{t=0}^{T} p\left(\mathbf{z}_{it}|\mathrm{ins}_{it},y,a_{i0},h_{it};\boldsymbol{\gamma}_y\right) p(\boldsymbol{\gamma}_y)$$

Using data augmentation for z_m^*, the posterior of γ is normally distributed (probit model).

Block 3: mixture weights parameters The third block is given by:

$$p\left(\boldsymbol{\delta}|\mathbf{c},\mathbf{e},\mathbf{y},\boldsymbol{a}_0,\mathbf{h}^T,\mathbf{z}^T;\boldsymbol{\gamma},\boldsymbol{\beta}\right) \propto \prod_{y=1}^{N_y}\prod_{i:\mathrm{y}_i=y} p\left(y|\mathrm{c}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y,\boldsymbol{\delta}_y\right) p(\boldsymbol{\delta}_y)$$

we sample δ using a Metropolis algorithm with flat priors in $p(\boldsymbol{\delta})$

Block 4: latent lifestyles The posterior distribution of the latent lifestyles with flat priors is given by:

$$\begin{aligned} p\left(\mathrm{y}_i|\mathbf{ins}_i^T,\mathrm{c}_i,\mathrm{e}_i,a_{i0},h_{i0},\mathbf{h}_i^T,\mathbf{z}_i^T;\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\delta}\right) &\propto p\left(\mathbf{z}_i^T|\mathbf{ins}_i^T,\mathrm{y}_i,a_{i0},\mathbf{h}_i^T;\boldsymbol{\gamma}_y\right) \\ &\times p\left(\mathbf{h}_i^T|\mathrm{e}_i,\mathrm{y}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y\right) \\ &\times p\left(\mathrm{y}_i|\mathrm{c}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y,\boldsymbol{\delta}_y\right) \end{aligned}$$

Thus, we can directly sample the latent lifestyle from:

$$p\left(\mathrm{y}_i=y|\mathrm{c}_i,\mathrm{e}_i,a_{i0},h_{i0},\mathbf{ins}_i^T,\mathbf{h}_i^T,\mathbf{z}_i^T;\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\delta}\right) = \frac{p\left(\mathbf{z}_i^T|y,a_{i0},h_{i0},\mathbf{h}_i^T,\mathbf{ins}_i^T;\boldsymbol{\gamma}_y\right) p\left(\mathbf{h}_i^T|\mathrm{e}_i,y,a_{i0},h_{i0};\boldsymbol{\beta}_y\right) p\left(y|\mathrm{c}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y,\boldsymbol{\delta}_y\right)}{\sum_{\tilde{y}=1}^{N_y} p\left(\mathbf{z}_i^T|\tilde{y},a_{i0},h_{i0},\mathbf{h}_i^T,\mathbf{ins}_i^T;\boldsymbol{\gamma}_{\tilde{y}}\right) p\left(\mathbf{h}_i^T|\mathrm{e}_i,\tilde{y},a_{i0},h_{i0};\boldsymbol{\beta}_{\tilde{y}}\right) p\left(\tilde{y}|\mathrm{c}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_{\tilde{y}},\boldsymbol{\delta}_{\tilde{y}}\right)}$$

Appendix B: Alternative estimations

In our main specification, we estimate the econometric model with only two latent types, which we label *protective* and *detrimental*. In this estimation, the different composition of types across education groups explains 42.3% of the education gradient in life expectancy. In this Appendix, we extend the model by estimating specifications with three and four groups, and we also compare them to a simple smoking/non-smoking classification.

B.1 More than 2 types

When estimating the model with three groups, the *protective* group remains essentially unchanged ("group 1"), while the *detrimental* lifestyle is split into two subgroups. These two subgroups are very similar to each other in terms of behavior, including smoking incidence at younger ages. However, as individuals age, those in what we call "group 2" tend to quit smoking, whereas those in "group 3" do not (see Figure B.1). This classification increases the difference in life expectancy between the best and worst lifestyles ("group 1" and "group 3") to 10.2 years (compared with 9.2 years in the two-group classification; see Table B.1). Furthermore, the share of the education gradient in life expectancy explained by differences in lifestyle composition across education groups increases modestly to 44.2%.

Next, we estimate the four-group model. Group 2 displays the main features of the *protective* lifestyle, with very low smoking probabilities and relatively high participation in exercise and preventive health care. Group 1 has a smoking probability that declines rapidly with age and lower participation in preventive care, while Group 3 is distinguished by substantially higher heavy-drinking probabilities, peaking around age 70. Group 4 maintains high smoking probabilities through middle age (see Figure B.2). Although Group 1 participates less in preventive care than Group 3, its remaining life expectancy at age 50 is substantially higher: 29.7 versus 22.9 years. This difference reflects its more favorable estimated health dynamics, including lower probabilities of entering poor health and greater probabilities of recovery. Health-transition information also plays a larger role in classification in the four-group specification, particularly in identifying Group 3. The largest life expectancy gap is between Groups 2 and 4 and amounts to 11.5 years (see Table B.2). The fraction of the education gradient explained by differences in type composition increases to 61.0%.

TABLE B.1: Life expectancies at age 50: males born in 1970s (3 groups)

	All		HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	28.5	100.0	24.1	100.0	27.0	100.0	31.9	7.7	4.4	3.4
Group 1	65.9	30.7	40.5	28.1	58.1	29.3	83.7	32.5			
Group 2	14.2	29.4	18.1	27.0	16.2	28.7	10.2	32.2			
Group 3	19.9	20.5	41.4	18.9	25.6	20.6	6.1	22.8			

Notes: See footnote in Table 3.

TABLE B.2: Life expectancies at age 50: males born in 1970s (4 groups)

	All		HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	28.8	100.0	24.5	100.0	27.2	100.0	32.2	7.7	3.0	4.7
Group 1	13.7	29.7	16.8	27.1	15.4	28.7	10.4	32.7			
Group 2	59.1	32.0	33.6	30.9	50.6	30.8	77.9	33.4			
Group 3	7.5	22.9	7.4	21.2	8.5	22.7	5.8	23.9			
Group 4	19.8	20.5	42.2	18.9	25.4	20.6	5.9	22.7			

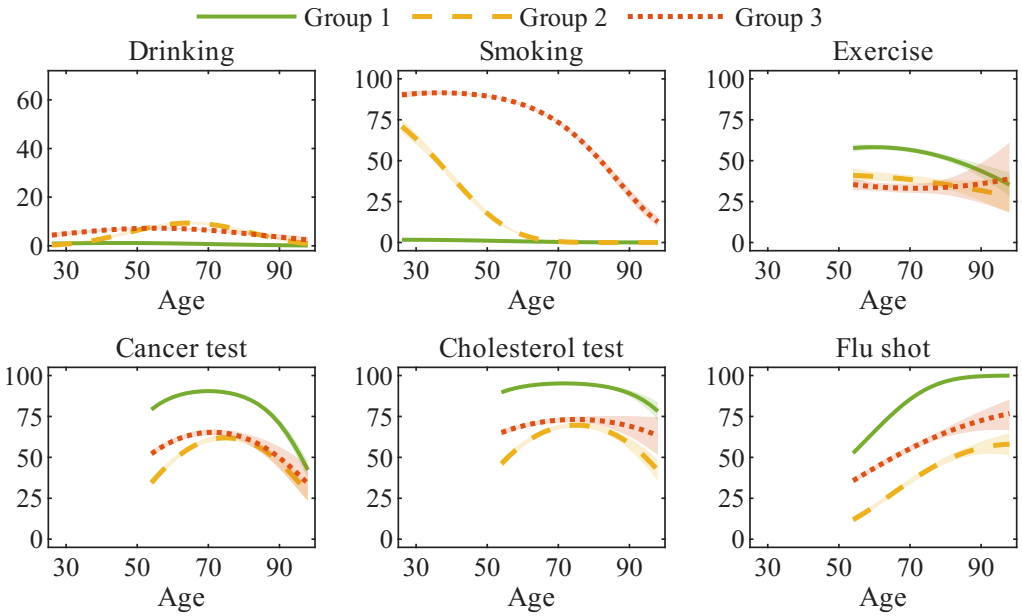
Notes: See footnote in Table 3.

TABLE B.3: Life expectancies at age 50: males born in 1970s (Smokers vs non-smokers)

	All		HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	28.0	100.0	23.9	100.0	26.5	100.0	31.1	7.2	5.1	2.1
Non-Smoking	70.1	30.2	52.1	27.0	63.5	28.8	84.3	32.3			
Smoking	29.9	22.8	47.9	20.5	36.5	22.6	15.7	25.0			
Δ_y	40.2	7.4	4.1	6.5	27.0	6.2	68.7	7.3			

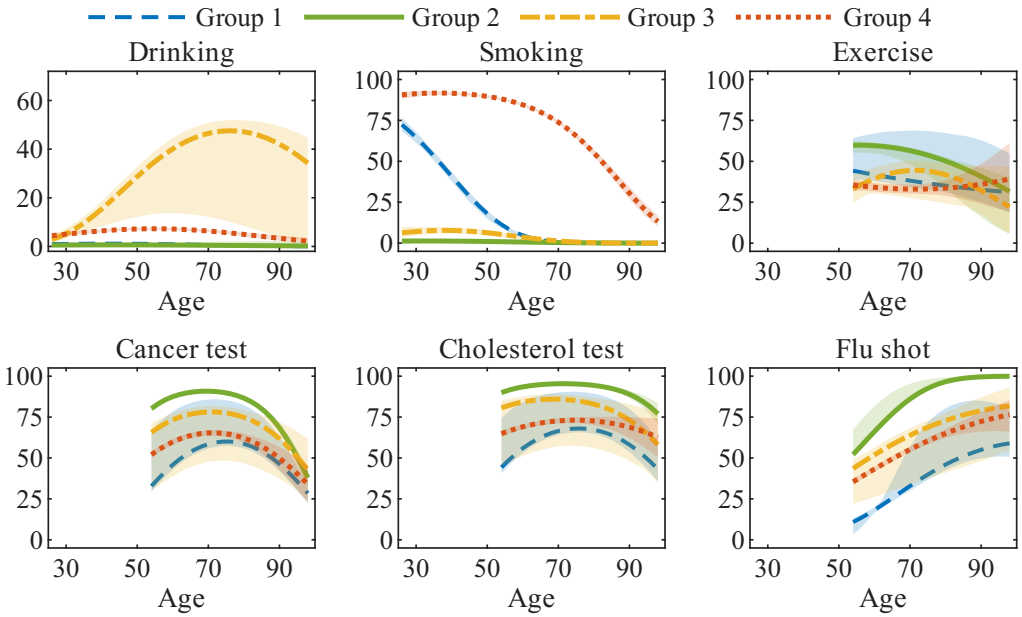
Notes: See footnote in Table 3.

FIGURE B.1: Health behaviors and lifestyles (3 groups)



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health. The shaded areas represent the 95% credible intervals.

FIGURE B.2: Health behaviors and lifestyles (4 groups)



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health. The shaded areas represent the 95% credible intervals.

B.2 Choosing the number of types

To decide on the optimal number of types and following Borella et al. (2025) we use two metrics from the machine learning literature: silhouette (Rousseeuw, 1987) and elbow (Thorndike, 1953). For this purpose, we focus on a sample of 2,223 HRS men observed at model ages 50–60. We select this age range because information on health behaviors is relatively abundant, while attrition due to mortality is still low.

For each individual, we construct six behavior frequencies by averaging the observed binary indicators over this age range. We also compute q_i, the fraction of valid self-reported health observations in which the individual reports poor or fair health. We retain individuals with at least one valid observation for each behavior and for health status, leaving the sample unchanged at 2,223 men. Let f_{ik} denote individual i's observed frequency of behavior k. We measure the distance between individuals i and j as

$$d_{ij} = \frac{1}{7}\left[\sum_{k=1}^{6} |f_{ik} - f_{jk}| + |q_i - q_j|\right].$$

This distance gives equal weight to each of the six behaviors and to health status. For each individual i, we define a_i as his average distance to all other members of his assigned type, excluding himself. We then calculate his average distance to the members of each alternative type separately and define b_i as the smallest of these averages. The silhouette of individual i is

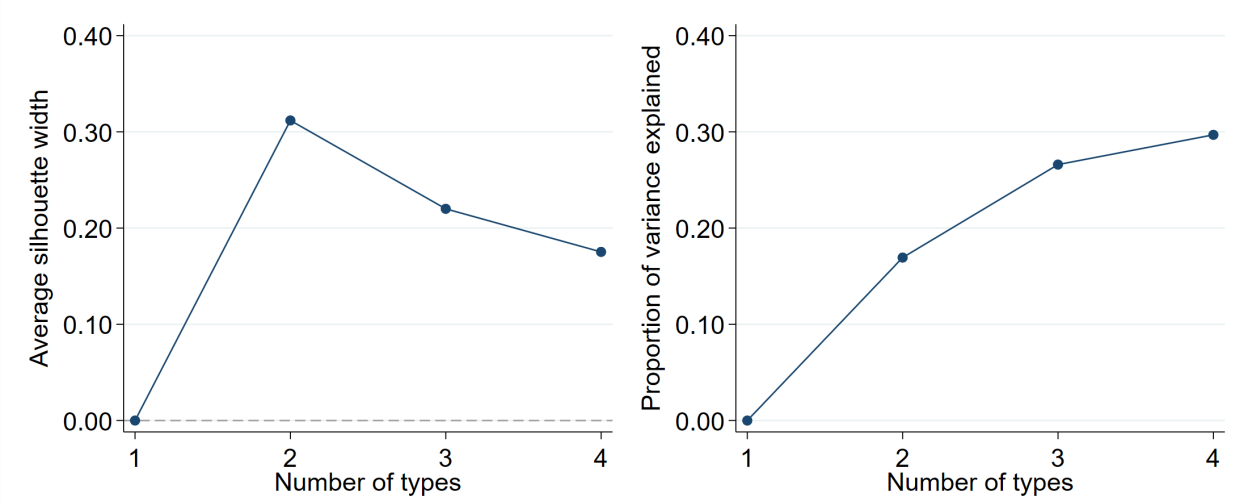
$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}.$$

The statistic ranges from -1 to 1. Values close to 1 indicate that the individual's behavior profile is much closer to members of his assigned type than to members of any alternative type. Values close to zero indicate that his assigned type and the closest alternative type are similarly distant. Negative values indicate that the individual is, on average, closer to an alternative type than to his assigned type. We average s_i across individuals to obtain the average silhouette width for each specification, with higher values indicating better behavioral separation between types.

The left panel of Figure B.3 reports average silhouette width, which compares each individual's distance to members of his own group with his distance to the nearest alternative group. Higher values indicate better separation. Average silhouette width falls from 0.312 with two groups to 0.220 and 0.175 with three and four groups. This would indicate the 2 types as a good choice.

For the elbow diagnostic, we first measure total variation by summing the squared differences between each individual's seven frequencies and their corresponding overall sample means (WSS_1). Then, with K types, we instead calculate the within-type sum of squares, WSS_K, using deviations from the corresponding means within each individual's assigned type. The proportion of variation

FIGURE B.3: Separation and variation explained by latent types



Notes: Common sample of 2,223 HRS men with at least one observed value for each of the six behaviors and a valid health report at model ages 50–60. The seventh frequency is the fraction of health reports in poor or fair health. The left panel uses the mean absolute difference across all seven frequencies as the distance measure. The right panel reports $1 - \text{WSS}_K/\text{WSS}_1$, where WSS_K sums squared deviations from group means across individuals and the seven frequencies.

explained by the types is

$$R^2_K = 1 - \frac{\text{WSS}_K}{\text{WSS}_1},$$

where WSS_1 is the total variation obtained when all individuals belong to a single type. This statistic measures the fraction of total variation accounted for by differences in mean profiles across types. A value of zero indicates that type membership explains none of the variation, while a value of one indicates that individuals within each type have identical profiles. All seven frequencies receive equal weight. The right panel of Figure B.3 shows that this proportion increases from 16.9% with two groups to 26.6% and 29.7% with three and four groups. By construction, R^2_K will always increase with the number of groups K. The idea is to pick a K such that the function becomes clearly concave with K, suggesting small gains from increasing the number of types.

The silhouette criterion therefore favors two groups, whereas the elbow plot may suggest three groups. Given the modest increase in explained variation from adding a third group and the silhouette criterion, we select two lifestyles for the structural model in favor of parsimony.

B.3 Lifestyle types by smoking

Finally, we compare our estimation method to a simple two-group classification based solely on whether an individual reports smoking in any interview. Table B.3 shows that the life expectancy difference between non-smokers and smokers is 7.4 years, which is 1.8 years smaller than the 9.2 gap in our protective/detrimental classification. For the 1970s cohort, differences in smoking-group composition explain 29.2% of the education gradient in life expectancy, compared with 42.3% in

our two-type specification.

Appendix C: Theoretical results of the first-stage model

In this appendix, we obtain two theoretical results for the first-stage model in Section 5.1. The results are proved for a slightly simpler version of the model with only two education categories (without loss of generality, we drop e = HSG by imposing $\tau_{\text{HSG}} = \infty$ for all individuals). First, in Lemma 1 and 2 we characterize the choice of $(\text{e}, \text{y}) \in \{(\text{CG},\text{PRO}), (\text{CG},\text{DET}), (\text{HSD},\text{PRO}), (\text{HSD},\text{DET})\}$ described in Figure 4, and second, in Proposition 1 we characterize the pattern of selection of individuals into the different (e, y) groups. All our discussion below focuses on the cases where the choices of education e and lifestyle y are separable or complementary, see Definition 1. Throughout this appendix, *separability* and *complementarity* are properties of the adult-life values V_0, whereas *independence* refers to the joint distribution of the cost draws $(\tau_{\text{CG}}, \tau_{\text{PRO}})$, that is, to the case $\rho = 0$.

Definition 1. *We say that the choices of education e and lifestyle y are separable whenever* $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} = V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$ *and complementary whenever* $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$.

Individuals in the model are heterogeneous and characterized by a pair $(\tau_{\text{CG}}, \tau_{\text{PRO}}) \in \mathbb{R}^2$. Given the objective function (4), we can characterize the choice of education for an individual with taste shocks $(\tau_{\text{CG}}, \tau_{\text{PRO}})$ with a threshold function $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ making the individual indifferent between going to college or not such that the individual will choose e = CG whenever $\tau_{\text{CG}} < \tau^*_{\text{CG}}(\tau_{\text{PRO}})$ and e = HSD with the reverse inequality. Lemma 1 below characterizes this threshold function, which can be seen in Panel (B) of Figure 4 as the piecewise linear function that separates the (CG, PRO) and (CG, DET) regions from the (HSD, PRO) and (HSD, DET) regions. Symmetrically, we can characterize the choice of lifestyle with a threshold function $\tau^*_{\text{PRO}}(\tau_{\text{CG}})$ such that the individual will choose y = PRO whenever $\tau_{\text{PRO}} < \tau^*_{\text{PRO}}(\tau_{\text{CG}})$ and y = DET with the reverse inequality. Lemma 2 below characterizes this threshold function, which can be seen in Panel (B) of Figure 4 as the piecewise linear function that separates the (CG, PRO) and (HSD, PRO) regions from the (CG, DET) and (HSD, DET) regions. Hence, the crossing of these two threshold functions fully characterizes the joint decision on (e, y).

Lemma 1. *Let* $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ *be the threshold function making an individual with taste shock* τ_{PRO} *indifferent between going to college or not such that if* $\tau_{\text{CG}} < \tau^*_{\text{CG}}(\tau_{\text{PRO}})$ *the individual goes to college. This function is continuous, piecewise linear, and non-increasing in* τ_{PRO}. *Furthermore,*

(a) For the case where education and lifestyle are complementary, the function is defined by

$$\tau^*_{\text{CG}}(\tau_{\text{PRO}}) = \begin{cases} V_0^{\text{CG,PRO}} - V_0^{\text{HSD,PRO}} & \text{if } \tau_{\text{PRO}} < V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}} \\ V_0^{\text{CG,PRO}} - V_0^{\text{HSD,DET}} - \tau_{\text{PRO}} & \text{if } \tau_{\text{PRO}} \in \left[V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}, V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}}\right] \\ V_0^{\text{CG,DET}} - V_0^{\text{HSD,DET}} & \text{if } \tau_{\text{PRO}} > V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} \end{cases}$$

(b) For the case where education and lifestyle are separable, the function is just defined by

$$\tau^*_{\text{CG}}(\tau_{\text{PRO}}) = V_0^{\text{CG,PRO}} - V_0^{\text{HSD,PRO}} = V_0^{\text{CG,DET}} - V_0^{\text{HSD,DET}}$$

Proof. An individual with taste shocks $(\tau_{\text{CG}}, \tau_{\text{PRO}})$ chooses to go to college if

$$\max\{V_0^{\text{CG,PRO}} - \tau_{\text{PRO}}, V_0^{\text{CG,DET}}\} - \tau_{\text{CG}} > \max\{V_0^{\text{HSD,PRO}} - \tau_{\text{PRO}}, V_0^{\text{HSD,DET}}\} \tag{C.1}$$

that is, the education decision depends on both τ_{CG} and τ_{PRO}. Let's define the threshold function $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ making an individual with taste shocks $(\tau_{\text{CG}}, \tau_{\text{PRO}})$ indifferent between going to college or not. To characterize this function, we distinguish three cases according to the value of τ_{PRO}:

Case 1: $\tau_{\text{PRO}} < V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$

This inequality implies the choice of y = PRO whenever e = HSD. Furthermore, as long as $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} \geq V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$ (complementary or separable choices) we also have that $\tau_{\text{PRO}} < V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}}$, which also implies the choice of y = PRO whenever e = CG. Therefore, case 1 guarantees the choice of y = PRO for all education levels (in both the complementary and separable cases). In this case, inequality (C.1) states that the individual will choose e = CG iff $\tau_{\text{CG}} < V_0^{\text{CG,PRO}} - V_0^{\text{HSD,PRO}}$.

Case 2: $\tau_{\text{PRO}} > V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}}$

This inequality implies the choice of y = DET whenever e = CG. Furthermore, as long as $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} \geq V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$ (complementary or separable choices) we have that $\tau_{\text{PRO}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$, which also implies the choice of y = DET whenever e = HSD. Therefore, case 2 guarantees the choice of y = DET for all education levels (in both the complementary and separable cases). In this case, inequality (C.1) states that individual will choose e = CG iff $\tau_{\text{CG}} < V_0^{\text{CG,DET}} - V_0^{\text{HSD,DET}}$.

Case 3: $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > \tau_{\text{PRO}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$

This intermediate case for τ_{PRO} is only possible in the case of complementary choices, as in the separable case the interval has length zero. In this case, inequality (C.1) states that the individual will choose e = CG iff $\tau_{\text{CG}} < V_0^{\text{CG,PRO}} - V_0^{\text{HSD,DET}} - \tau_{\text{PRO}}$. That is, in this third case, the threshold $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ making individuals indifferent between e = CG and e = HSD is strictly decreasing in τ_{PRO}. □

Lemma 2. *Let $\tau^*_{\text{PRO}}(\tau_{\text{CG}})$ be the threshold function making an individual with taste shock τ_{CG} indifferent between protective and detrimental lifestyles such that if $\tau_{\text{PRO}} < \tau^*_{\text{PRO}}(\tau_{\text{CG}})$ the individual chooses a protective lifestyle. This function is continuous, piecewise linear, and non-increasing in τ_{CG}. Furthermore,*

(a) For the case where education and lifestyle are complementary, the function is defined by

$$\tau^*_{\text{PRO}}(\tau_{\text{CG}}) = \begin{cases} V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} & \text{if } \tau_{\text{CG}} < V_0^{\text{CG,DET}} - V_0^{\text{HSD,DET}} \\ V_0^{\text{CG,PRO}} - V_0^{\text{HSD,DET}} - \tau_{\text{CG}} & \text{if } \tau_{\text{CG}} \in \left[V_0^{\text{CG,DET}} - V_0^{\text{HSD,DET}}, V_0^{\text{CG,PRO}} - V_0^{\text{HSD,PRO}}\right] \\ V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}} & \text{if } \tau_{\text{CG}} > V_0^{\text{CG,PRO}} - V_0^{\text{HSD,PRO}} \end{cases}$$

(b) For the case where education and lifestyle are separable, the function is just defined by

$$\tau^*_{\text{PRO}}(\tau_{\text{CG}}) = V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} = V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$$

Proof. The proof is symmetric to Lemma 1 □

Proposition 1. *Assume that the two cost draws are independently distributed, $\tau_{\text{CG}} \perp \tau_{\text{PRO}}$ (that is, $\rho = 0$). Let $F(\tau_{\text{PRO}})$ be the CDF of τ_{PRO}, $f(\tau_{\text{PRO}})$ its associated PDF, and $F(\tau_{\text{PRO}}|e)$ be the CDF of τ_{PRO} conditional on a particular educational choice e. Then, in the case where the education and lifestyle choices are complementary we have that $F(\tau_{\text{PRO}}|\text{HSD})$ first-order stochastically dominates $F(\tau_{\text{PRO}})$, which in turn first-order stochastically dominates $F(\tau_{\text{PRO}}|\text{CG})$. In the separable case, the three distributions are identical.*

Proposition 1 isolates the selection generated by complementarity alone: it holds even when the two cost draws are independent, and the independence assumption is used only in the first step of the proof below, where $\Pr(\text{CG}|\tau_{\text{PRO}})$ is computed from the marginal distribution of τ_{CG}. A positive correlation $\rho > 0$ reinforces the ranking, since it directly concentrates low-τ_{PRO} individuals among those facing low education costs.

Proof. An individual with taste shocks $(\tau_{\text{CG}}, \tau_{\text{PRO}})$ goes to college whenever $\tau_{\text{CG}} < \tau^*_{\text{CG}}(\tau_{\text{PRO}})$. Let us start with the case in which the choices of e and y are complementary. Lemma 1 states that $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ is non-increasing in τ_{PRO}, and strictly decreasing in τ_{PRO} within an interval of positive length. That is, if $\tau^a_{\text{PRO}} > \tau^b_{\text{PRO}}$ then $\tau^*_{\text{CG}}(\tau^a_{\text{PRO}}) \leq \tau^*_{\text{CG}}(\tau^b_{\text{PRO}})$, with strict inequality for some pairs $\left(\tau^a_{\text{PRO}}, \tau^b_{\text{PRO}}\right)$. Therefore $\Pr(\text{CG}|\tau_{\text{PRO}}) \equiv \Pr(\tau_{\text{CG}} \leq \tau^*_{\text{CG}}(\tau_{\text{PRO}}))$ is weakly decreasing with τ_{PRO}, and so is $\Pr(\text{CG}|\tau \leq \tau_{\text{PRO}}) = \int_{-\infty}^{\tau_{\text{PRO}}} \Pr(\text{CG}|\tau) \frac{f(\tau)}{F(\tau_{\text{PRO}})} d\tau$.[23] Using Bayes theorem, we have that

$$F(\tau_{\text{PRO}}|\text{CG}) = \Pr(\tau \leq \tau_{\text{PRO}}|\text{CG}) = \frac{\Pr(\text{CG}|\tau \leq \tau_{\text{PRO}})\Pr(\tau \leq \tau_{\text{PRO}})}{\Pr(\text{CG})} \geq \Pr(\tau \leq \tau_{\text{PRO}}) = F(\tau_{\text{PRO}})$$

where the last inequality comes from the fact that, since $\Pr(\text{CG}|\tau \leq \tau_{\text{PRO}})$ is non-increasing in τ_{PRO} and strictly decreasing for some τ_{PRO}, it must be that

$$\frac{\Pr(\text{CG}|\tau \leq \tau_{\text{PRO}})}{\Pr(\text{CG})} \geq 1$$

[23] To see this last part, note that $\frac{\partial \Pr(\text{CG}|\tau \leq \tau_{\text{PRO}})}{\partial \tau_{\text{PRO}}} = \left(\Pr(\text{CG}|\tau_{\text{PRO}}) - \Pr(\text{CG}|\tau \leq \tau_{\text{PRO}})\right) \frac{f(\tau_{\text{PRO}})}{F(\tau_{\text{PRO}})} \leq 0$, where the first term inside the parentheses is no greater than the second term whenever $\Pr(\text{CG}|\tau_{\text{PRO}})$ is non-increasing in τ_{PRO}.

For high school dropouts we have equivalently:

$$F(\tau_{\text{PRO}}|\text{HSD}) = \Pr(\tau \leq \tau_{\text{PRO}}|\text{HSD}) = \frac{\Pr(\text{HSD}|\tau \leq \tau_{\text{PRO}})\Pr(\tau \leq \tau_{\text{PRO}})}{\Pr(\text{HSD})}$$
$$= \frac{[1 - \Pr(\text{CG}|\tau \leq \tau_{\text{PRO}})]\Pr(\tau \leq \tau_{\text{PRO}})}{1 - \Pr(\text{CG})} \leq \Pr(\tau \leq \tau_{\text{PRO}}) = F(\tau_{\text{PRO}})$$

thus:

$$F(\tau_{\text{PRO}}|\text{HSD}) \leq F(\tau_{\text{PRO}}) \leq F(\tau_{\text{PRO}}|\text{CG}) \quad \forall \tau_{\text{PRO}}$$

with strict inequality for some τ_{PRO}.

Finally, note that whenever we are in the separable case, Lemma 1 states that $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ is just a constant that does not depend on τ_{PRO}. Hence, $\Pr(\text{CG}|\tau_{\text{PRO}})$ is also independent from τ_{PRO} and we have $F(\tau_{\text{PRO}}|\text{HSD}) = F(\tau_{\text{PRO}}) = F(\tau_{\text{PRO}}|\text{CG}) \ \ \forall \tau_{\text{PRO}}$. □

Appendix D: First step estimation details

D.1 Income process

We estimate the labor-market processes separately for each education category using PSID data from 1968 to 2019. The sample consists of men aged 25–62, born between 1920 and 1980. We group birth years into the 1930s cohort (1920–1939), the 1950s cohort (1940–1959), and the 1970s cohort (1960–1979). Health h_{it} is coded as bad when self-reported health is fair or poor. Because health is only available starting in the 1984 wave, we also include an indicator o_{it} for whether health is observed.

Deterministic component of earnings: fixed effects. For individuals with positive labor earnings, let y_{it} denote log earnings in 2018 dollars, a_{it} denote age, and h_{it} indicate bad health. We estimate

$$y_{it} = \alpha_i + \beta_{1e} a_{it} + \beta_{2e} a_{it}^2 + \delta_e h_{it} + \omega_e o_{it} + v_{it}.$$

The individual fixed effects α_i absorb all permanent observed and unobserved individual characteristics. Consequently, the coefficients on age and health are identified from within-person variation separately for different education groups. Birth cohort is time invariant and cannot be identified jointly with the individual effects. We recover cohort levels in a second step. In particular, using one observation per person, we regress the estimated $\widehat{\alpha}_i$ on indicators $\lambda_{1930,e}$, $\lambda_{1950,e}$, and $\lambda_{1970,e}$ for cohort $\times$ education. The deterministic earnings profile supplied to the structural model is therefore

$$\mu_t^{\mathrm{c,e}}(h_t) = \widehat{\beta}_{1e} a_t + \widehat{\beta}_{2e} a_t^2 + \widehat{\delta}_e h_t + \widehat{\omega}_e o_{it} + \lambda_{1930,e} \mathbb{1}\{\mathrm{c} = 1950\} + \lambda_{1950,e} \mathbb{1}\{\mathrm{c} = 1950\} + \lambda_{1970,e} \mathbb{1}\{\mathrm{c} = 1970\}.$$

Employment: correlated random effects. Individual fixed-effects probit would discard people whose employment status never changes and would not identify the time-invariant cohort effects. We instead estimate an education-specific Mundlak correlated-random-effects (CRE) probit. Let $z_{it} = (b_{it}, a_{it}, b_{it}a_{it}, o_{it})'$. The latent employment equation is

$$\ell_{it}^* = \theta_e' X_{it} + \pi_e' \overline{z}_i + c_{i,e} + u_{it}, \qquad \ell_{it} = \mathbb{1}\{\ell_{it}^* > 0\},$$
$$c_{i,e} \sim N(0, \sigma_{c,e}^2), \qquad u_{it} \sim N(0,1),$$

where X_{it} contains a constant, current age, current health, their interaction, the observed-health indicator, and cohort indicators, while $\overline{z}_i$ contains the corresponding person means of all time-varying regressors. Including these means allows the random intercept to be correlated with an individual's observed age and health history. Conditional on $\overline{z}_i$, the model retains both within-person changes and between-person levels and identifies the cohort coefficients. We integrate out

the random intercept when constructing employment probabilities,

$$\Pr(\ell_{it} = 1 \mid X_{it}, \overline{z}_i, e) = \Phi\left(\frac{\theta_e' X_{it} + \pi_e' \overline{z}_i}{\sqrt{1 + \sigma_{c,e}^2}}\right),$$

and average these probabilities over the empirical distribution of $\overline{z}_i$ within each education–cohort group before passing them to the structural model.

Persistent–transitory error decomposition. Let $r_{it} = y_{it} - \widehat{\mu}_{it}^{\mathrm{c,e}}(h_{it})$ be the residual from the deterministic earnings profile. We decompose the residual as

$$\begin{aligned} r_{it} &= \eta_{it} + \epsilon_{it}, \\ \eta_{i,t+2} &= \rho_e \eta_{it} + \nu_{i,t+2}, & \nu_{i,t+2} &\sim N(0, \sigma_{\nu,e}^2), \\ \epsilon_{it} &\sim N(0, \sigma_{\epsilon,e}^2), & \eta_{i0} &\sim N(0, \sigma_{\eta,e}^2), \end{aligned}$$

where the persistent and transitory innovations are mutually independent and serially independent, and η_{it} is initialized at its stationary distribution, so that $\sigma_{\nu,e}^2 = \sigma_{\eta,e}^2(1 - \rho_e^2)$.

We estimate the stochastic component by just-identified GMM, separately by education. Let $\widehat{\gamma}_{j,e}$ denote the sample covariance of r_{it} and $r_{i,t-j}$, using matched observations for $j = 2$ and $j = 4$ calendar years. The three moment conditions are

$$\begin{aligned} \widehat{\gamma}_{0,e} - \left(\sigma_{\eta,e}^2 + \sigma_{\epsilon,e}^2\right) &= 0, \\ \widehat{\gamma}_{2,e} - \rho_e \sigma_{\eta,e}^2 &= 0, \\ \widehat{\gamma}_{4,e} - \rho_e^2 \sigma_{\eta,e}^2 &= 0. \end{aligned}$$

Because the system is exactly identified, its GMM solution has the closed form

$$\begin{aligned} \widehat{\rho}_e &= \frac{\widehat{\gamma}_{4,e}}{\widehat{\gamma}_{2,e}}, & \widehat{\sigma}_{\eta,e}^2 &= \frac{\widehat{\gamma}_{2,e}^2}{\widehat{\gamma}_{4,e}}, \\ \widehat{\sigma}_{\epsilon,e}^2 &= \widehat{\gamma}_{0,e} - \widehat{\sigma}_{\eta,e}^2, & \widehat{\sigma}_{\nu,e}^2 &= \widehat{\sigma}_{\eta,e}^2(1 - \widehat{\rho}_e^2). \end{aligned}$$

Table F.1 reports the resulting estimates.

D.2 Medical shocks and insurance

To estimate medical expenditures, we follow De Nardi et al. (2025). We first convert all medical spending in the MEPS to 2018 dollars. For each individual, we then compute total medical expenditures over a two-year period. Conditional on health status in the first year, we divide individuals at each age into three expenditure groups: the bottom 50%, the 50th–90th percentiles, and the top

10%. We compute mean medical expenditures within each group and represent the three groups as discrete realizations of the medical expenditure shock ζ_t. Following De Nardi et al. (2025), we multiply these means by 1.60 for individuals younger than age 65 and by 1.90 for individuals aged 65 and older. These adjustments ensure that medical spending in our model matches aggregate medical spending in the National Health Expenditure Accounts (NHEA).

For each health status, education group, and realization of the medical expenditure shock, we estimate two-year medical expenditures using an OLS regression on a quadratic polynomial in age. We use the estimated regressions to construct smooth expenditure profiles at two-year age intervals from age 26 to age 98 for every (h, e, ζ) cell.

Figure D.1 presents the resulting medical expenditure profiles for each expenditure group and education level, shown separately by health status.

To measure private insurance coverage, we compute total payments made by private insurers over a two-year period and divide these payments by total two-year medical expenditures. Let $q_t^e(h, \zeta)$ denote the fraction of medical expenditures paid by the insurer, conditional on age, education, health, and the medical expenditure shock. For each education group, health state, and realization of ζ, we estimate this fraction using an OLS regression on a quadratic polynomial in age. We use the estimated regressions to construct smooth coverage profiles at two-year age intervals from age 26 to age 98. Figure D.2 presents the resulting profiles.

To estimate the probability of having private insurance, we regress insurance status on a quadratic polynomial in age, fully interacted with education and employment status. We estimate the probit model using individuals aged 30–60 and use the fitted model to predict insurance probabilities over the working-age range of the structural model. The resulting probability, denoted by $\text{Prob}_t^e(\ell)$, depends on age, education, and employment status. Figure D.3 presents the estimated profiles.

We assume that employed individuals with private insurance receive employer-sponsored coverage and pay no premium. Unemployed individuals with private insurance pay the premium

$$p_t^e(h) = \xi \, \mathbb{E}_\zeta \left[q_t^e(h, \zeta) m_t^e(h, \zeta) \right] + 2\psi^h,$$

where $\xi = 1.07$ is a proportional load and ψ^h is an annual fixed load. Following De Nardi et al. (2025), we set ψ^h equal to \$100 in good health and \$2,100 in bad health per year.

Finally, individuals aged 65 or older are covered by Medicare. Medicare pays 50% of their medical expenditures. The annual Medicare Part B premium is \$1,055.

D.3 Tax system

We follow De Nardi et al. (2025) 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} =\$128,400.

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

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

D.4 Value of statistical life

To obtain the model equivalent of the VSL, we proceed as follows. Using the value function expressed in Section 5.2.1 we can obtain the total differential:

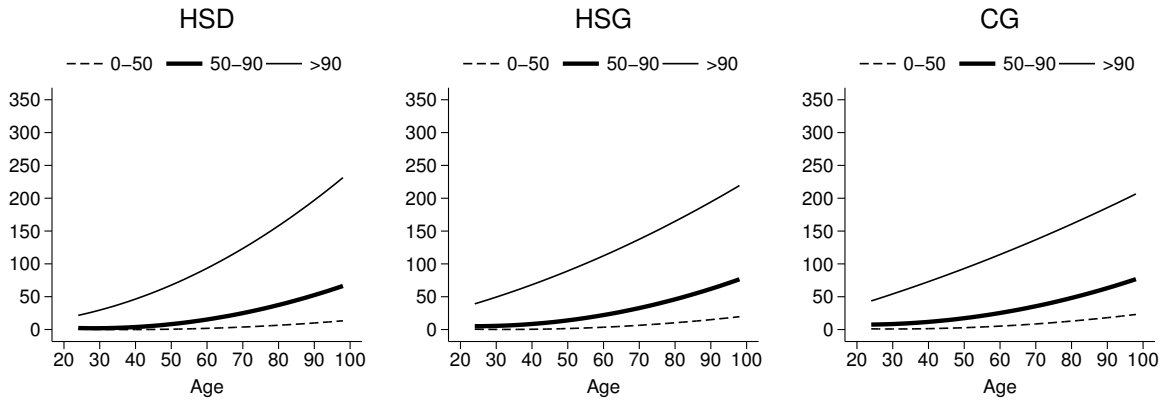
$$\frac{\partial V_t^{\mathrm{c,e,y}}(x,h,\eta)}{\partial x} dx + \frac{\partial V_t^{\mathrm{c,e,y}}(x,h,\eta)}{\partial s_t^{\mathrm{e,y}}(h)} ds_t^{\mathrm{e,y}}(h) = 0$$

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

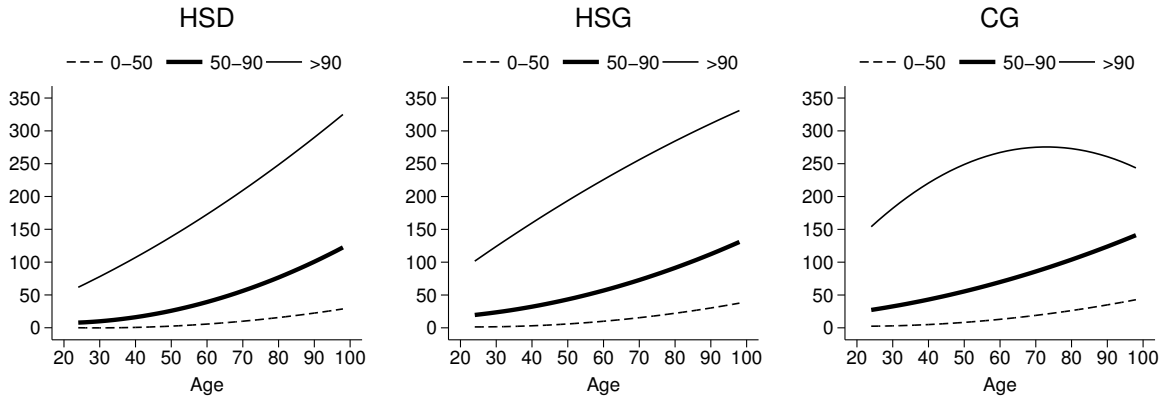
$$-\frac{dx}{ds_t^{\mathrm{e,y}}(h)} = \frac{\partial V_t^{\mathrm{c,e,y}}(x,h,\eta)}{\partial s_t^{\mathrm{e,y}}(h)} \left[\frac{\partial V_t^{\mathrm{c,e,y}}(x,h,\eta)}{\partial x}\right]^{-1}$$

Hence, an individual of type (c,e,y) with state variables (x,h,η) at age t would be willing to give up $0.01 \times dx/ds_t^{\mathrm{e,y}}(h)$ units of cash on hand in exchange for an increase of one percentage point in his survival probability. Putting 100 identical agents together, such an increase avoids one death on average, so their joint willingness to pay of $dx/ds_t^{\mathrm{e,y}}(h)$ units of income is the model counterpart of the value of one statistical life.

FIGURE D.1: Two-year Medical expenses distribution by education and health



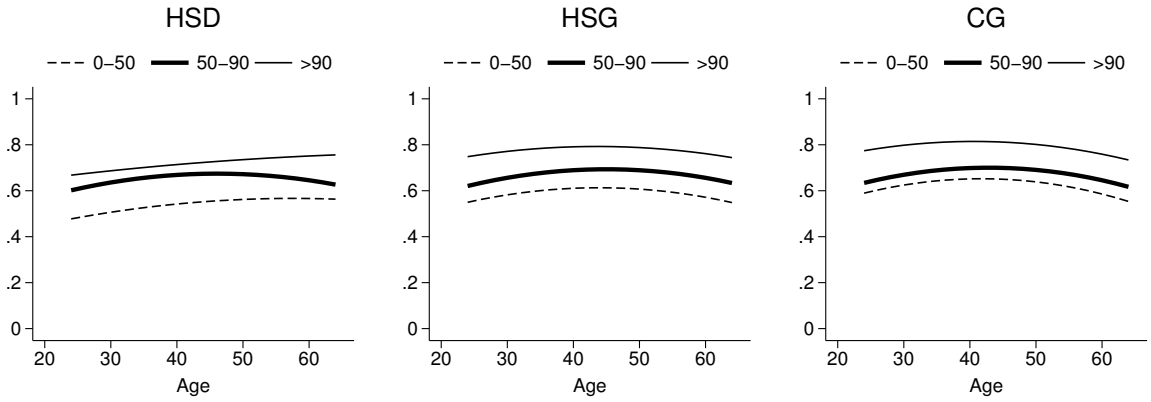
(A) Good health



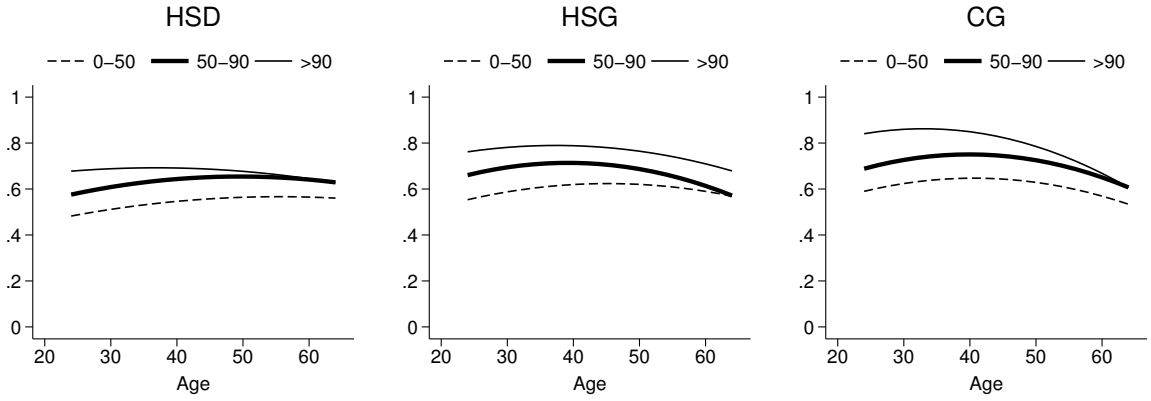
(B) Bad health

Notes: Two-year total medical expenditures are in thousands of constant 2018 USD, derived from the MEPS data years 2010-2019. Estimates generated from quadratic regression models by health status and education level (HSD: high-school dropout; HSG: high-school graduate; CG: college graduate). Dashed lines represent expenditures below the 50th percentile; solid lines represent expenditures between the 50th and 90th percentiles; and thick solid lines represent expenditures above the 90th percentile.

FIGURE D.2: Private health insurance coverage by medical expenditure group, education, and health



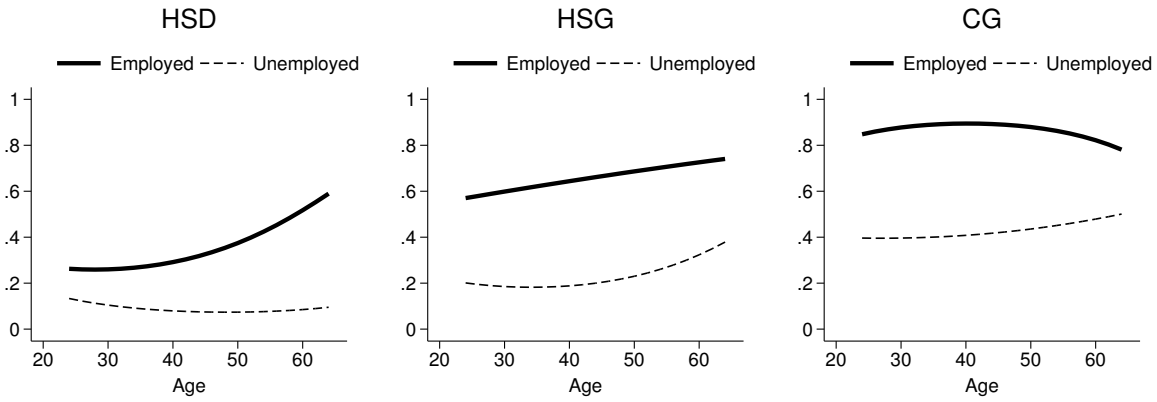
(A) Good health



(B) Bad health

Notes: Fraction of two-year total medical expenditures covered by insurance, derived from the MEPS data years 2010-2019. Estimates generated from quadratic regression models by health status and education level (HSD: high-school dropout; HSG: high-school graduate; CG: college graduate). Dashed lines represent coverage when expenditures below the 50th percentile; solid lines represent coverage when expenditures between the 50th and 90th percentiles; and thick solid lines represent coverage when expenditures above the 90th percentile.

FIGURE D.3: Probability of having private insurance



Notes: Probability of being covered by private insurance, derived from the MEPS data years 2010-2019. Estimates generated from probit regression on quadratic in age, employment status and education level (HSD: high-school dropout; HSG: high-school graduate; CG: college graduate). Dashed lines represent probability of being covered when employed; and thick solid lines represent probability of being covered when unemployed.

Appendix E: Robustness exercises

In this appendix, we consider two different model robustness exercises.

Uncorrelated heterogeneity. In the benchmark model, education and lifestyle costs are positively correlated, introducing a primitive source of selection alongside the endogenous selection generated by complementarity. As a robustness exercise, we impose $\text{corr}(\epsilon_{\text{PRO}}, \epsilon_e) = 0$ and re-estimate the early-life cost parameters (see Figure E.1 for the model fit). Unlike the counterfactual in Section 7.1, this exercise allows the remaining parameters to adjust to the absence of primitive correlation. The estimated dispersion of lifestyle costs falls substantially, from $\sigma_{\text{PRO}} = 5.78$ to 1.97, while their mean rises from $\mu_{\text{PRO}} = 5.39$ to 7.79. The dispersion of education costs also declines; see Tables 4 and E.1. The lower dispersion of lifestyle costs makes choices more responsive to differences in adult-life returns, increasing the quantitative importance of income prospects. For the 1970s cohort, the income counterfactual reduces the life expectancy gradient by 3.2 years, from 7.8 to 4.6 years, compared with a reduction of 1.3 years in the benchmark. The selection counterfactual reduces the gradient by 0.9 year, from 7.8 to 6.9 years, compared with 2.6 years in the benchmark; see Table E.2. Selection remains present because education and lifestyle investments are complementary even when their primitive costs are independent. Across cohorts, however, this specification overstates the deterioration in lifestyles among high school dropouts: their protective share falls by 19 percentage points, compared with 8 percentage points in the data. The life expectancy gradient consequently increases by 1.8 years, compared with 1.2 years in the benchmark and 1.0 year in the data. The income and selection exercises generate increases of 1.1 and 0.6 years, respectively; see Table E.3. Thus, removing primitive correlation and re-estimating the model shifts the explanation toward the direct income channel.

Higher VSL. We also consider a calibration targeting a VSL of \$4 million rather than the benchmark value of \$2 million (see Figure E.2 for the model fit). The higher value attached to survival raises the value of adopting a *protective* lifestyle. Consistent with this, the estimated mean lifestyle cost rises from 5.39 to 14.14. The estimated correlation between education and lifestyle costs also increases, from 0.20 to 0.38; see Tables 4 and E.1. For the 1970s cohort, the income counterfactual reduces the life expectancy gradient by 1.1 years, from 7.9 to 6.8 years, compared with 1.3 years in the benchmark. The selection counterfactual reduces the gradient by 2.9 years, from 7.9 to 5.0 years, compared with 2.6 years in the benchmark. Selection therefore remains the larger channel; see Table E.4. Setting the cost correlation to zero while holding the remaining parameters fixed reduces the lifestyle gradient from 40 to 6 percentage points, indicating a substantial role for primitive correlated heterogeneity in this calibration. Across cohorts, the higher-VSL specification generates a 0.9-year increase in the life expectancy gradient, compared with 1.2 years in the benchmark and 1.0 year in the data. The income and selection exercises generate increases of 0.4 and 0.6 years, respectively; see Table E.5. Although the predicted change in the life expectancy gradient is close to the data, the model understates changes in educational

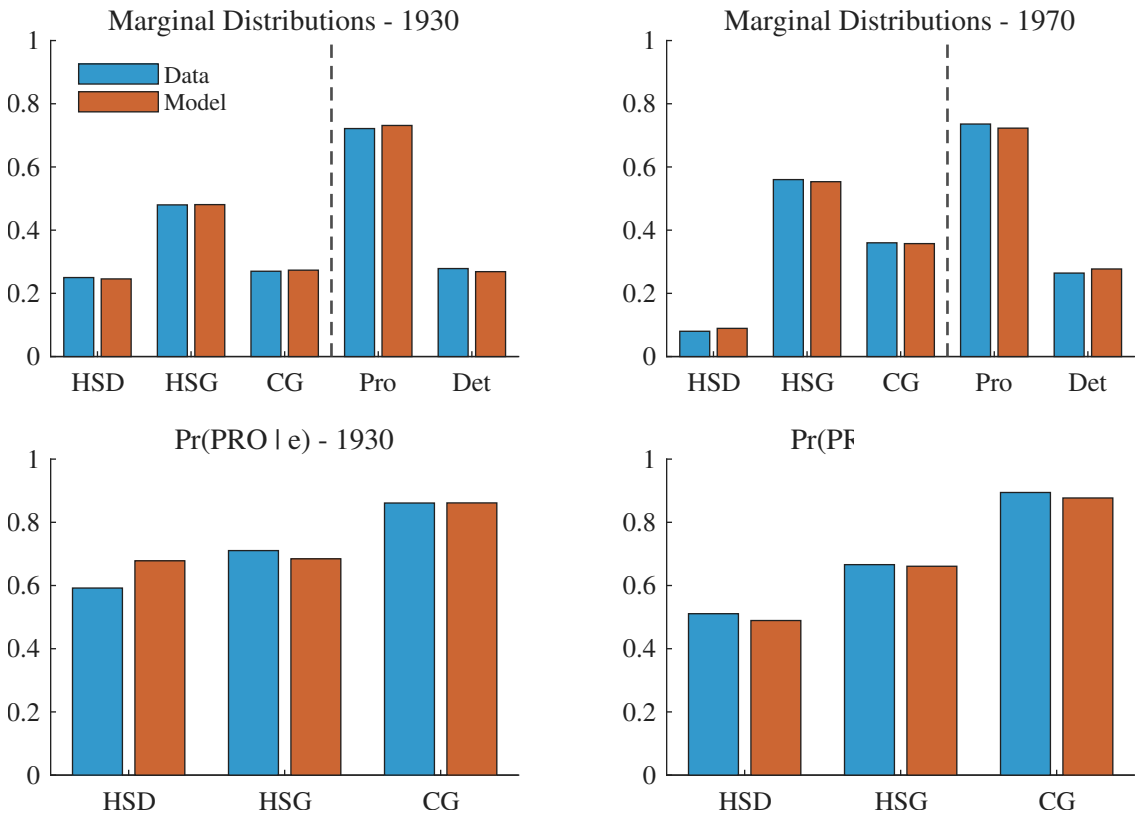
TABLE E.1: Calibrated parameters: robustness checks

Parameter	Description	VSL= 4 M	$\rho = 0$
$\bar{b}$	Value of life	5.55	3.20
μ_{HSG}	Average cost of HSG education	4.03	4.34
μ_{CG}	Average cost of CG education	30.54	15.29
μ_{PRO}	Average cost of PRO lifestyle	14.14	7.79
σ_{HSG}	Sd. cost of HSG education	4.18	2.21
σ_{CG}	Sd. cost of CG education	38.19	10.98
σ_{PRO}	Sd. cost of PRO lifestyle	5.80	1.97
ρ	Correlation edu,PRO	0.38	0.00

Notes. Panel A: internally calibrated parameters for the adult life stage. Panel B: calibrated parameters for the distributions of the utility costs τ_{PRO}, τ_{CG}, and τ_{HSG} in the early life stage of the model. The first column refers to the model parameters for the calibration with a higher VSL. The second column refers to the model parameters for the model with uncorrelated unobserved heterogeneity.

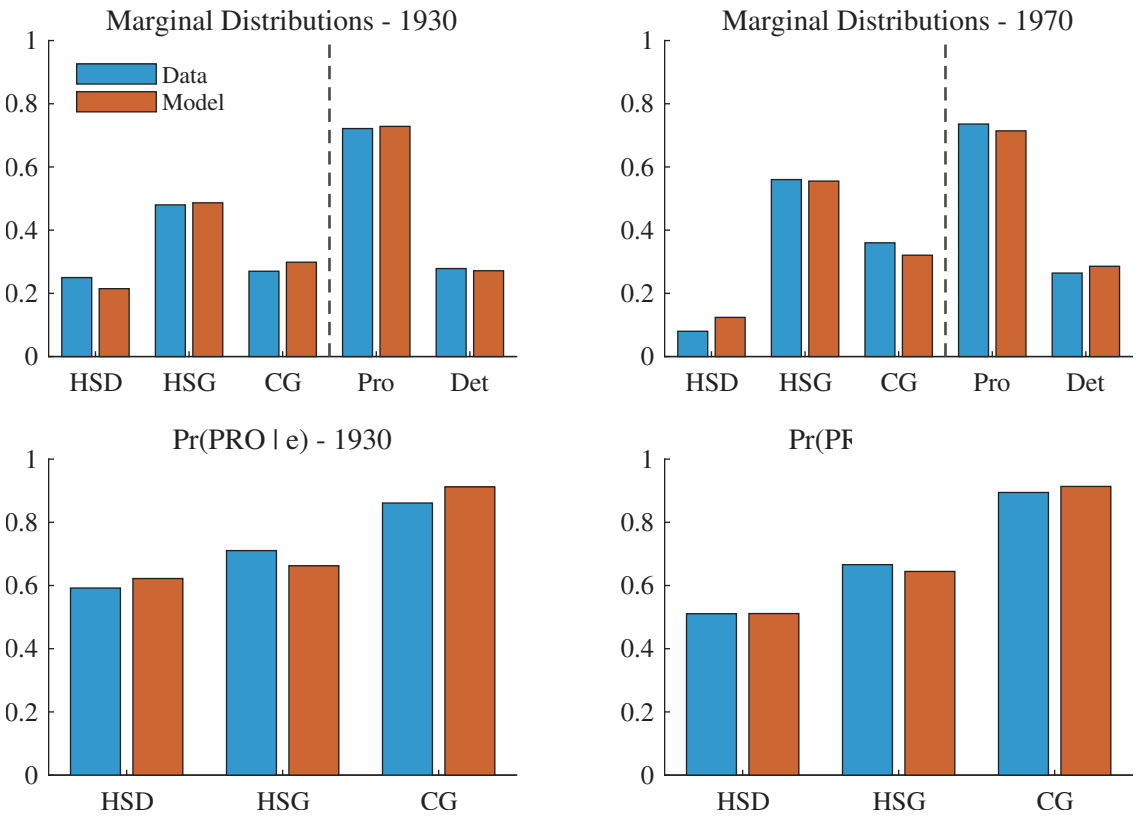
attainment: college attainment increases by 2 percentage points rather than 9, while the dropout share declines by 9 percentage points rather than 17.

FIGURE E.1: Model Fit. First-Stage (uncorrelated unobserved heterogeneity)



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.

FIGURE E.2: Model Fit. First-Stage (larger VSL)



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.

TABLE E.2: Decomposition: 1970s cohort (uncorrelated unobserved heterogeneity)

	Pr(e)		Pr(PRO|e)			LE(e)		
	CG	HSD	CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.36	0.08	0.89	0.51	0.38	31.5	23.7	7.8
Model	0.36	0.09	0.88	0.49	0.39	31.3	23.5	7.8
Same lifestyle			0.88	0.88	0.00	31.3	26.8	4.5
Income			0.88	0.87	0.01	31.3	26.7	4.6
Selection			0.88	0.59	0.29	31.3	24.4	6.9
Selection + income			0.88	0.91	-0.03	31.3	27.0	4.3

Notes: The exercises correspond to the first six rows of Table 6, using the specification re-estimated under zero correlation between education and lifestyle costs.

TABLE E.3: Decomposition: changes across cohorts (no correlated unobserved heterogeneity)

	Δ_cPr(e)		Δ_cPr(PRO|e)			Δ_cLE(e)		
	CG	HSD	CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.09	-0.17	0.03	-0.08	0.11	0.3	-0.7	1.0
Model	0.09	-0.15	0.01	-0.19	0.21	0.1	-1.6	1.8
Income			0.01	-0.11	0.13	0.1	-1.0	1.1
Selection			0.00	-0.07	0.08	0.0	-0.6	0.6

Notes: See footnote in Table 9.

TABLE E.4: Decomposition: 1970s cohort (larger VSL)

	Pr(e)		Pr(PRO|e)			LE(e)		
	CG	HSD	CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.36	0.08	0.89	0.51	0.38	31.5	23.7	7.8
Model	0.32	0.12	0.91	0.51	0.40	31.6	23.7	7.9
Same lifestyle			0.91	0.91	0.00	31.6	27.1	4.5
Income			0.91	0.65	0.26	31.6	24.9	6.8
Selection			0.91	0.86	0.05	31.6	26.7	5.0
Selection + income			0.91	0.93	-0.02	31.6	27.2	4.4
No correlated costs	0.32	0.13	0.80	0.75	0.06	30.6	25.7	5.0

Notes: See footnote in Table 6.

TABLE E.5: Decomposition: changes across cohorts (larger VSL)

	Δ_cPr(e)		Δ_cPr(PRO|e)			Δ_cLE(e)		
	CG	HSD	CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.09	-0.17	0.03	-0.08	0.11	0.3	-0.7	1.0
Model	0.02	-0.09	0.00	-0.11	0.11	0.0	-0.9	0.9
Income			0.00	-0.04	0.04	0.0	-0.3	0.4
Selection			-0.00	-0.07	0.07	-0.0	-0.6	0.6

Notes: See footnote in Table 9.

Appendix F: Extra figures and tables

TABLE F.1: GMM estimates of the stochastic component of labor earnings

	ρ_e	$\sigma^2_{\eta,e}$	$\sigma^2_{\nu,e}$	$\sigma^2_{\epsilon,e}$
HSD	0.905	0.463	0.084	0.333
HSG	0.870	0.403	0.098	0.282
CG	0.853	0.454	0.123	0.217

Notes: Estimates are obtained separately by education group by matching the variance and the two- and four-year autocovariances of residual log earnings. The parameter ρ_e is the two-year persistence coefficient, $\sigma^2_{\eta,e}$ is the stationary variance of the persistent component (and its initial variance), $\sigma^2_{\nu,e}$ is the innovation variance, and $\sigma^2_{\epsilon,e}$ is the variance of the transitory component.

FIGURE F.1: Health transitions across education and lifestyles

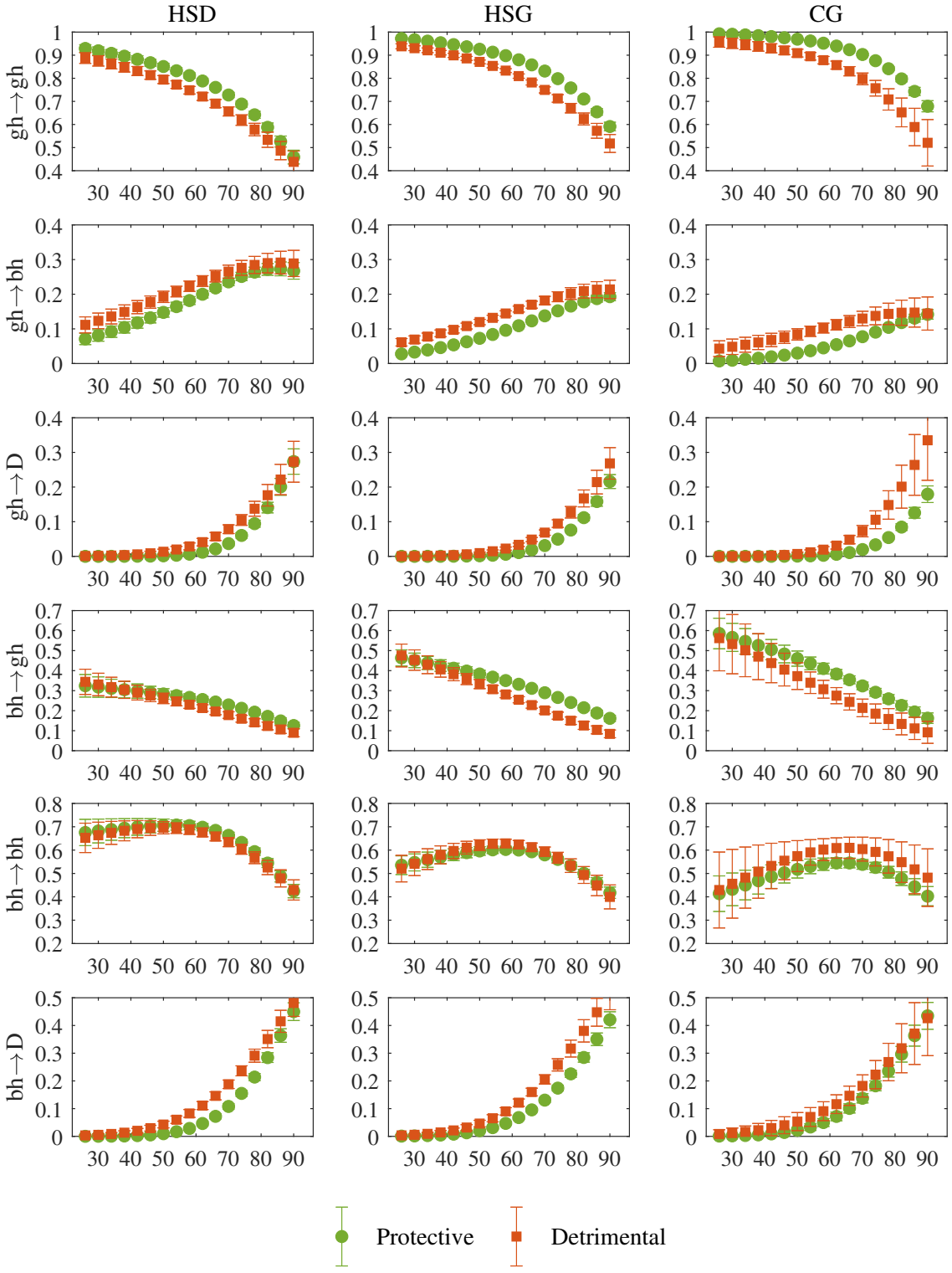
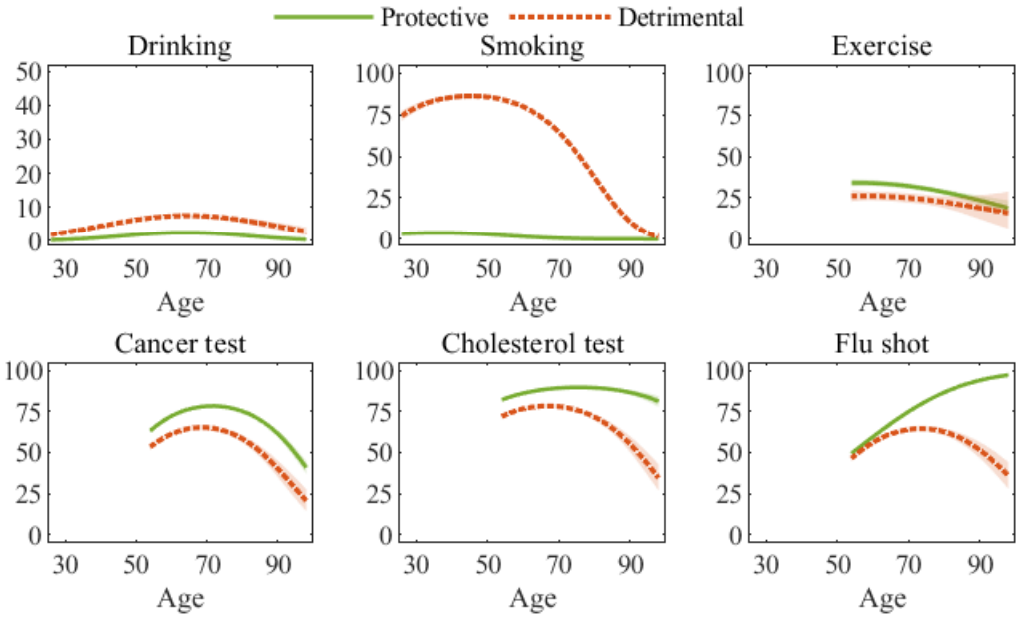
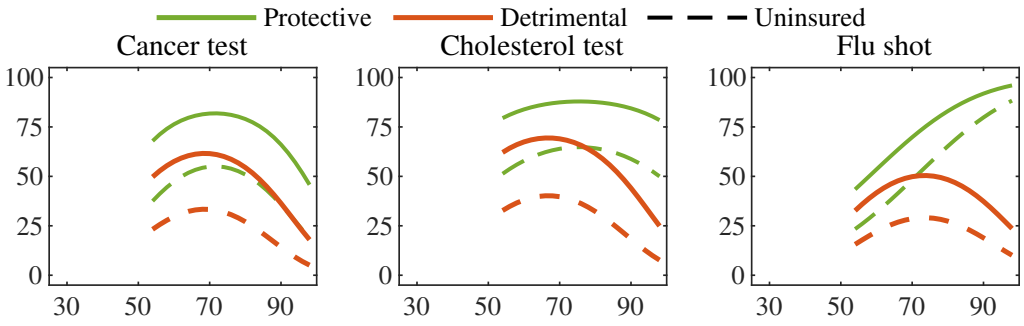


FIGURE F.2: Health behaviors and lifestyles: bad health



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

FIGURE F.3: Medically related health behaviors and lifestyles by insurance status



Notes: Estimation results. Probability of engaging in each health behavior by age, type and insurance status.