Health Inequality and Health Types

Margherita Borella, Francisco Bullano, Mariacristina De Nardi,
Benjamin Krueger, and Elena Manresa*

April 30, 2025

Abstract

While health affects many economic outcomes, its dynamics are still poorly understood. We use k-means clustering, a machine learning technique, and data from the Health and Retirement Study to identify health types during middle and old age. We identify five health types: the vigorous resilient, the fair-health resilient, the fair-health vulnerable, the frail resilient, and the frail vulnerable. They are characterized by different starting health and health and mortality trajectories. Our five health types account for 84% of the variation in health trajectories and are not explained by observable characteristics, such as age, marital status, education, gender, race, health-related behaviors, and health insurance status, but rather, by one's past health dynamics. We also show that health types are important drivers of health and mortality heterogeneity and dynamics. Our results underscore the importance of better understanding health type formation and of modeling it appropriately to properly evaluate the effects of health on people's decisions and the implications of policy reforms.

*Margherita Borella: University of Torino and CeRP-Collegio Carlo Alberto, Italy. Francisco Bullano: University of Minnesota, Mariacristina De Nardi: University of Minnesota, Federal Reserve Bank of Minneapolis, CEPR, and NBER. Benjamin Krueger: Northwestern University. Elena Manresa: New York University. We are grateful to Manuel Arellano, Marco Bassetto, Pamela Giustinelli, John B. Jones, Rory McGee, Vincent Mor, Michela Tincani, Martin Garcia Vazquez, Fan Wang, and Fang Yang for useful comments and suggestions. We also thank Johanna Torres Chain for her contributions to the data cleaning and construction of the cause of death data using the exit interviews file of the HRS. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, the CEPR, any agency of the federal government, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

1 Introduction

While health affects many economic outcomes (including future health, disability, labor supply, earnings, retirement, nursing home entry, savings, and death), its dynamics are still poorly understood. This paper aims at better understanding health dynamics during middle and old age using data from the Health and Retirement Study (HRS), which is a U.S. panel data set covering people aged 51 and older. We use these data to document how health and mortality evolve, how unequal is their evolution, and how to better model the dynamics of health and mortality. We show that people belong to different health types, that these health types have an important influence on one's future health and mortality dynamics, and that ignoring the existence of these health types leads to an important mis-characterization of future health dynamics and mortality.

To achieve these goals, we organize our analysis around five questions. The first question is whether, during middle and old age, there are "health types", that is, whether people have heterogeneous health trajectories. The second question is what those health types look like. The third question is whether, by the time one reaches middle age, health types can mostly be captured by observables, or whether health types mainly reflect unobserved heterogeneity. The fourth question is how important are health types in driving health dynamics and what we miss by ignoring them, specifically when modeling health as function of previous period's health and other commonly used observables. Our fifth question is whether we can capture health dynamics by modeling them as just a function of health, health types, and age, and abstract from other commonly used observables such as gender, marital status, and education.

To start our analysis, we adopt a parsimonious measure of health that has been proposed by the gerontology literature. The frailty index (or "frailty") uses numerous responses about specific impairments (ability to get in and out of bed, feed oneself, use the phone, read a map,...) and conditions (diabetes, cancer, obesity, high blood pressure...) and computes, for each person over time, the proportion of "deficits" as a fraction of the potential ones

considered. It is thus a number between zero and one that varies for each person over time, depending on how that person's underlying deficits evolve. Frailty has been shown to be a good predictor of future health and mortality, but also of many other economic outcomes, including by race, gender, and ethnicity (See for instance Hosseini, Kopecky, and Zhao (2022) and Russo, McGee, De Nardi, Borella, and Abram (2024)).

The HRS collects data every other year. We select people age 52-53 and we follow them until 2018, or until they die, whichever comes first. In this age group, death is primarily a manifestation of health (the HRS "exit interviews," show that only 4% of those in our sample die for causes "unrelated to health"). For this reason, when people die, we attribute them a frailty of one (its maximum possible value) from that time on, and keep them in our health trajectories.

We extract health types by using people's realized health trajectories and k-means clustering. We do so by dividing our sample period in two parts: a clustering period that goes from age 52 to 60, and a validation period which goes from age 62 to 74. That is, we extract health types using k-means during our clustering period, and we then validate our health types during the remaining future period by testing whether the health types that we extract before age 60 have predictive power for the health and mortality dynamics for those alive after age 60. Not using all available history for each person allows us to avoid over-fitting when we test for the predictive power of our health types.

Clustering assigns observations (in our case a health trajectory between age 52 and 60) to groups (clusters) or health types so that the trajectories of health within a given health type are as close as possible, while those across health types are as different as possible. This method provides the important advantages of providing a direct and intuitive mapping between health types and people, and of being non-parametric, and thus not requiring making any functional form assumptions, including about the sources of heterogeneity being observable and unobservable. Within many possible clustering methods, we adopt k-means clustering, which is the only clustering method for which the statistical properties of identify-

ing unobserved heterogeneity from discrete classification have been determined (Bonhomme, Lamadon, and Manresa (2022)).

K-means clustering requires specifying the number of clusters, or health types, K. We select this number by using two sets of criteria. First, we study whether increasing K improves our forecasts of frailty and mortality during our clustering period. Second, we adopt typical machine learning criteria. Both set of criteria are consistent with five health types. People of each health type tend to have similar initial health and health trajectories during our clustering period, and our between types variation accounts for 84% of the total variation across health types trajectories between the ages of 52 and 60.

We also evaluate our health types by evaluating its predictive power (out of sample) in future health and mortality outcomes after age 60, and hence after our clustering period ends, and show that our health types out-predict other potentially relevant observables. More specifically, in a regression of future frailty on demographic characteristics, adding health types to the regression increases the R^2 from 0.120 to 0.571. In a similar regression of future frailty on demographic characteristics and frailty and self-reported health at 52, adding health types increases the R^2 from 0.510 to 0.591. Health types are also statistically significant in a logit regression of mortality on demographics, first period frailty, self-reported health, and health types.

Taking together the extent to which health types explain variation of health and mortality during and after our clustering period, our answer to Question 1 is thus that we find strong evidence for health types, and in particular for five health types.

From this, we move onto our second question. What are these health types like and how do they behave during middle and older ages? We find that individuals belonging to our five health types have very heterogeneous health dynamics, both during and after our clustering period. Looking at age 52 data only, one might conclude that there are only three health types, which start out as "vigorous", in "fair-health", and "frail" and that they have very heterogeneous starting health (2, 6, and 14 health deficits respectively). But looking

over a longer horizon reveals that while those starting out vigorous are also resilient, that is, they face slow health deterioration and are unlikely to die early, those in fair and frail health then bifurcate into two groups each, the resilient ones who are unlikely to experience fast health deterioration and die early, and the vulnerable ones, who instead experience fast health deterioration and early death. A related important observation is that the fraction of people dying for non "health-related" causes is small in all health types. Thus, those who die fast during our sample period are not those having accidents, but rather, die due to underlying health-related issues.

Therefore, our answer to Question 2 is that our health types can be characterized as the "vigorous resilient" which comprises 57% of our sample, the "fair-health resilient," which accounts for 27% of our sample, the "fair-health vulnerable," which are 3 % of our sample, the "frail-resilient," which makes up for 10% , and the "frail-vulnerable," which are 3% . Thus, 6% of our individuals experience very fast health deterioration during middle and older ages. This fraction is similar to the 8% of the most vulnerable of three health types estimated by De Nardi, Pashchenko, and Porapakkarm (2023), that studies the whole life cycle and a different statistical model of types.

Next, we turn to our third question, that is to what extent are health types explained by a rich set of observables, and to what extent they reflect, instead, unobserved heterogeneity which can be identified using individual health trajectories. This is an important question because many models of health and mortality, including those used in many structural models, allow for rich heterogeneity in marital status, gender, education or permanent income, and so on, and find that health trajectories are, to some extent, explained by these observables. But is this the most parsimonious and effective way to model health and mortality?

To address this question we start by simply characterizing the distribution of many key observables by health types. This analysis does reveal several interesting findings. Women are less likely to start at age 52 in the healthiest group, but are also less likely to deteriorate fast and die early. Black people are less likely to be healthy but do not necessarily

deteriorate faster. People with higher education are more likely to belong to the "vigorous resilient groups," at age 52, but conditional on starting out from one of the other health types, they are not less likely to deteriorate faster. People in couples are more likely to be of types 1 and 2, but are similarly represented in types 3, 4, and 5. There is a clear gradient in individual income, with individual income decreasing in frailty type, but not in household income. Having ever smoked is increasing in frailty type and is more prevalent for fast deteriorating health types. Private insurance is decreasing in frailty type, while public insurance is increasing.

While this partial analysis reveals some interesting correlations, we also perform a more systematic exercise to understand the relationship between health types and observables. That is, we run a multinomial logistic regression of health types on one's initial health and a rich set of observables that include demographics, health behaviors, and health insurance status. This part of the analysis answers Question 3: our rich set of observables has poor explanatory power for health types (the pseudo-R^2 is 0.131). Initial frailty alone has a much higher explanatory power (its pseudo-R^2 is 0.430). All of the observables that we consider and initial frailty only generate a marginal improvement in the explanatory power of health types (its pseudo-R^2 is 0.448, compared to 0.430). Thus, we find that, at least once one reaches middle age, health types mainly reflect unobserved heterogeneity and that just modeling health types is a parsimonious way to capture health heterogeneity.

Question 4, in turn, asks how important health types are and what happens if we ignore them while instead taking into account other commonly used variables such as gender, marital status, education, and so on. This is important because many structural models of health that incorporate health risk ignore health types but allow for these other sources of heterogeneity. How much heterogeneity in health and mortality is this approach missing? To address this point, we take our previously computed health types as given and use them as an additional explanatory variable to study the evolution of a different measure: Self-reported health status (SRHS) and death starting from age 52 until death or the end of our

observation period.[1] This measure has, at this stage, some important advantages. In fact, it parsimoniously summarizes an individual's perception of their own health, it is not mechanically related to our frailty measure, and it has also been shown to be a good predictor of many economic outcomes, including by race, gender, and ethnicity (See for instance Hosseini, Kopecky, and Zhao (2022) and Russo, McGee, De Nardi, Borella, and Abram (2024)).

More specifically, we estimate a multinomial logit for SRHS and death as a function of age, previous SRHS, marital status, and education, all interacted with gender. To this, we add information on health types by allowing for a type-specific intercept. The first result from this part of the analysis is that health types are important drivers of health and mortality dynamics, even when we use a state-of-the-art first order Markov formulation. The second result is that ignoring health types misses both the timing and heterogeneity in mortality and the evolution of health conditional on being alive by health types, both during and after our clustering period. Importantly, the model without health types predicts much more mean reversion in health status by health type than is actually in the data.

Finally, we turn to our fifth question, that is, can we abstract from commonly used observables to explain health and death dynamics once we have health types, previous health, and age as explanatory variables? We answer this question by comparing two multinomial logit models of SRHS and mortality. One includes gender, education, and marital status, in addition to previous SRHS. The other one only includes health types, age, and previous SRHS. This part of the analysis reveals that the second parsimonious model outperforms the first one (with a pseudo-R^2 of 0.257 and 0.285, respectively), thus indicating that using a much more parsimonious model of health and mortality with health types actually outperforms the model with a rich set of observables. This is important because it might indicate that we should want to adopt this parsimonious specification, including in structural models, and possibly save on explicitly modeling many state variables such as gender, education, and being in a couple.

[1] Self-reported health status results from a question in which people are asked to self-report their health as "excellent," "very good," "good," "fair," and "poor."

Our findings provide some important lessons and open up interesting and important avenues for research. First, that health types are important to better understand health inequality and, thus, potentially, the extent to which health inequality drives economic inequality. Second, that ignoring health types can generate misleading policy implications both in terms of observed outcomes and welfare. For instance, the vulnerable adult types who die fast would not draw much benefit from increases in Social Security benefits, given their much shorter life spans.

Important questions that we leave unaddressed include the following. How long of a history do we need to identify health types? This is important not only from an operational stand point, but also because it might be related to how long a person needs to learn their health type, based on their family background, and their own realized health history. Related, by when do people learn their health type? To what extent do observed decisions such as retirement, savings, and labor supply depend on health types? Then, what would health types look earlier in life? And how do they relate to other key economic outcomes that we care about, including education, marriage, fertility decisions, disability, length of working life and retirement, and medical expenses? Finally, when and how are health types formed? Much more work is needed to address these important issues.

Related literature. Our paper relates to three main branches of the literature: the analysis of health heterogeneity and health inequality, the study of the effects of health risk on various economic outcomes, and the research on health formation.

The extensive empirical literature on health inequality examines the relationship between health outcomes and various economic factors, such as socioeconomic status, educational attainment, marital status and gender, and health costs. Case and Deaton (2005) use self-reported health status. Hosseini, Kopecky, and Zhao (2022) adopt the frailty index, a measure proposed by the gerontology literature (Goggins et al. (2005); Mitnitski et al. (2002, 2001, 2005); Searle et al. (2008)). Heiss (2011) studies models for health dynamics for self-reported health status and mortality and suggest a joint model with an autocorrelated

latent health component. Keeney et al. (2019) use observed health outcomes and medical utilization to study the composition of the most costly Medicare beneficiaries (and hence for people 65 and older). In this population they identify 5 phenotypes using latent class analysis. Bueren, Amengual, and Josep (2024) use observed health outcomes and lifestyle choices to identify latent types of lifestyle using Gaussian mixture models. They identify 2 lifestyle type and document significant difference in health outcomes driven by lifestyle behavior. Our contribution to this literature is threefold: First, identify health types using k-means clustering (Bonhomme, Lamadon, and Manresa, 2022), a tool previously used in the labor and earnings inequality literature. Importantly, unlike previous latent class analyses, our approach does not rely on parametric assumptions for extracting types. Second, we show that health can be effectively represented by a small number of health types or health trajectories. Third, we document that, in middle age, the effects of health types on health dynamics are much larger than those of education, gender, and income.

Our paper also contributes to the structural literature examining the effects of health risk on various economic outcomes, which typically assumes that health is exogenous and follows a first-order Markov process. Older studies using self-reported health as a health measure include French (2005), French and Jones (2011), French and Jones (2017), De Nardi, French, and Jones (2010), and Kopecky and Koreshkova (2014). More recent studies use frailty, for instance Hosseini, Kopecky, and Zhao (2021), Nygaard (2022), Hosseini, Kopecky, and Zhao (2022), Russo (2022) and Russo, McGee, De Nardi, Borella, and Abram (2024). Our contribution to this literature is to show that ignoring health-type heterogeneity when estimating the evolution of health fails to capture important heterogeneity in health outcomes. Specifically, these models generate insufficient persistence of health states and inadequate heterogeneity in mortality risk, both of which are crucial for forming expectations about future health and thus significantly impact economic decisions. Our findings align with those of De Nardi, Pashchenko, and Porapakkarm (2023),Capatina and Keane (2023) and Bueren, Amengual, and Josep (2024), that show that incorporating health types is essential

for understanding the dynamics of health, as well as its inequality over the life cycle.

Finally, our paper provides facts relevant for health formation. Among previous work in this context, Bolt (2021) shows that about one-third of health inequality can be explained by health investments over the life cycle and that investments during early childhood have a greater impact than those made during adulthood. Ozkan (2024) documents that high-income families invest more in preventive care for their children, leading to lower curative medical expenditures during adulthood and higher life expectancy. Mahler and Yum (2024) and Cole, Kim, and Krueger (2019) examine the non-monetary aspects of health investment, emphasizing how individuals can influence their health through "health effort" over the life cycle.

Because our approach is non-parametric, any attempt to endogenously model health dynamics should generate patterns that are consistent with those that we document. Moreover, it provides insights on how health investments relate to health types in in middle and old age.

2 Data and frailty

We use data from the University of Michigan Health and Retirement Study (HRS), a longitudinal panel study designed to survey individuals aged 51 and older residing in the United States, as well as their spouses. Initiated in 1992, it has been taking place every 2 years, and also collects data on many health measures. It is well known for its good sample size and rigorous follow-up procedures, which result in a low attrition rate.

Because key variables such as difficulties with daily living first appear in the 1996 survey, we use data from 1996 to 2018 (and hence for 12 interviews). As interviews take place every two years, we group individuals ages in two-year bins.

We select people starting out in our data at age 52-53 and who are 60-61 or older by 2018. From now on, we index these age groups by the lowest age in the group (52 refers to

52-53). Because we use the first five interviews for each person to identify health types and the remaining interviews to validate them, the choice of our sampling period ensures that we have enough interviews for each respondents to perform our identification and validation analysis. Our sample includes 4663 individuals. Appendix A contain detailed information about our sample selection.

2.1 How do we measure health?

Health is an unobserved and potentially multidimensional object. Measuring it presents a challenge on its own. We follow the gerontology literature (Mitnitski, Mogilner, and Rockwood (2001)) and measure health by the frailty index (or frailty), which tracks health deterioration by taking into account that, as people age, they accumulate more *health deficits*, such as difficulty with activities of daily living, functional limitations, and medical diagnoses.

Frailty is defined as the fraction of deficits present for an individual at a certain age over the total number of deficits considered. We follow Russo, McGee, De Nardi, Borella, and Abram (2024) and construct our frailty measure using their proposed 35 underlining deficits. For exposition purposes, Table 1 groups these health deficits into difficulties with Activities of Daily Living (ADLs), difficulties with instrumental activities of daily living (IADLs), other functional limitations, diagnoses by medical professionals, indicators of health care utilization and addictive diseases.[2] These health deficits are recorded as being either present, and thus equal to 1, or not.

Frailty thus weights all deficits equally. One might wonder whether this is a good choice because some deficits might indicate a more severe health impairment than another. Yet, previous literature that adopts principal components to construct a deficit-specific weighted health measure, shows that this measure does not outperform frailty (Hosseini, Kopecky, and Zhao (2022); Russo, McGee, De Nardi, Borella, and Abram (2024)) in terms of predictive

[2]We follow the medical literature in classifying obesity and smoking as diseases. The American Medical Association recognized obesity as a chronic disease in 2013. Many papers in medical literature, including for instance Bernstein and Toll (2019), also consider smoking to be a chronic disease

ability for economic outcomes. While this might seem surprising, there is a substantial degree of comorbidity among deficits, which we also document. That is, people in poor health are unlikely to have just one very severe deficit. Instead, they quickly develop several of them. Moreover, when we analyze the determinants of health types, we find that one's health type is primarily driven by one's level of frailty, rather than by one's frailty composition.

Table 1: Health deficits. Each deficit takes the value of 0 if the respondent reports not having it, or of 1 if the respondent reports having it.

Deficit	Deficit
ADLs	Difficulty lifting a weight heavier than 10 lbs
Difficulty bathing	Difficulty lifting arms over the shoulders
Difficulty dressing	Difficulty picking up a dime
Difficulty eating	Difficulty pulling/pushing large objects
Difficulty getting in/out of bed	Difficulty sitting for two hours
Difficulty using the toilet	
Difficulty walking across a room	***Diagnoses***
Difficulty walking one block	Diagnosed with high blood pressure
Difficulty walking several block	Diagnosed with diabetes
	Diagnosed with cancer
IADLs	Diagnosed with lung disease
Difficulty grocery shopping	Diagnosed with heart condition
Difficulty making phone calls	Diagnosed with a stroke
Difficulty managing money	Diagnosed with psychological or psychiatric problems
Difficulty preparing a hot meal	Diagnosed with arthritis
Difficulty taking medication	
Difficulty using a map	***Healthcare Utilization***
	Has stayed in the hospital in the previous two years
Other Functional Limitations	Has stayed in a nursing home in the previous two years
Difficulty climbing a one flight of stairs	
Difficulty climbing several flights of stairs	***Addictive Diseases***
Difficulty getting up from a chair	Has BMI larger than 30
Difficulty kneeling or crouching	Has ever smoked cigarettes

2.2 Death as an expression of one's health

Because we want to understand health and mortality dynamics, and because death is a manifestation of one's health and health type, we assign a frailty index of one to those who die, where one is the maximum value that frailty can take if a person has all possible possible deficits. To investigate the implications of this assumption we explore the frailty distribution among living individuals, the distribution of frailty changes across two successive periods depending on how far one is from death, and the causes of people's death.

Figure 1: Distribution of changes in frailty while alive, in the period preceding death, and in the period right before death

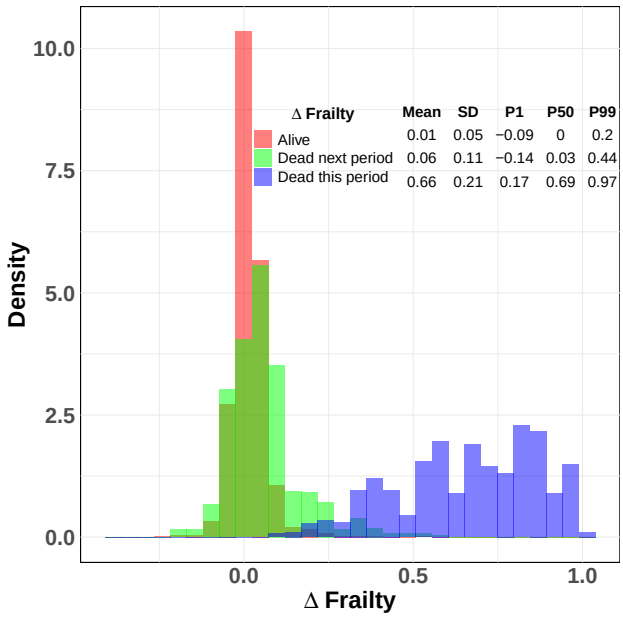


Table 10 in Appendix B illustrates that, among the living, the highest frailty level observed is 0.91 and that 98% of observations have frailty levels below 0.6. Thus, people do not attain a frailty score of 1 while alive.

Figure 1 shows that the average change in frailty over two consecutive waves (a two-year period) for living individuals is 0.01 (equivalent to 0.45 health deficits). In contrast, the average change in frailty during the two-year period immediately preceding death is 0.06 (1.99 health deficits), indicating a faster health deterioration as individuals approach death. For individuals who and thus transition to a frailty score of 1, the average change in frailty is 0.66 (23.1 health deficits). This highlights that the health decline associated with death is significantly larger than the changes observed during life.

Finally, we investigate the causes of death for those who die during our sampling period. Table 11 in appendix B shows that 5.3% and 12.9% die during our clustering period and by 2018, respectively. Among those with known causes of death, 7.2% and 4.4%, respectively, succumbed to non-health-related factors. This suggests that the majority of deaths were attributable to health-related issues. Furthermore, we observe similar health-related causes

of death for those who died prematurely and those who die later on.

3 Uncovering health types

To identify health types, we apply k-means clustering to health trajectories between ages 52 and 60. Here, h_i represents the health trajectory for individual i, yielding a vector with five realizations of frailty, f_{it}.

$$h_i = [f_{i,52}, f_{i,54}, f_{i,56}, f_{i,58}, f_{i,60}] \tag{1}$$

For a given number of health types or clusters ($\bar{k}$), the k-means algorithm generates representative health trajectories and assigns each individual to a specific cluster. The k-means method requires specifying the number of clusters (K), which we determine using two approaches. First, we apply standard machine learning methods, including the elbow method (Thorndike, 1953) and silhouette analysis (Rousseeuw, 1987). Second, we adopt an economic criterion, selecting K such that adding more health types does not significantly improve the predictive power of regression models for mortality and future frailty during the clustering period.

We define the predictive power of health types for $\bar{k}$ health types as $P(\bar{k}) = 1 - \frac{a(\bar{k})}{a(1)}$, where $a(\bar{k})$ is the mean absolute error of the regression model with $\bar{k}$ health types. To estimate $P(\bar{k})$, and to avoid over-fitting, we use 10-fold cross-validation. Finally, we select K so that there is no significant increase in the predictive power of health types when we add an additional health type $(K+1)$. Following this procedure we choose 5 health types. They explain 84% of the variation in health trajectories during the clustering period. Also, 5 health types are within the range suggested by traditional ML methods, which indicate 2 to 5 clusters. This finding is also robust to different specification in our models for frailty and mortality. Appendix C contains detailed information about our procedure to choose K.

The use of k-means to approximate individual heterogeneity with a finite number of

groups is justified when using multiple measurements per individual (Bonhomme, Lamadon, and Manresa (2022)). We use the histories of individuals as different measures. This allows us to reduce miss-classification error and to reliably approximate heterogeneity.[3] Moreover, our goal is not to identify group fixed effects and make a decomposition exercise. Instead, we aim at assigning each individual to a health type that represents a specific health trajectory. Unlike traditional methods for identifying latent classes, we do not impose any functional form assumptions during the type extraction stage.

While k-means is one of many dimension-reduction methods in the statistical literature, in addition to its formal justification to approximate individual heterogeneity, we think k-means provides an advantage in terms of ease of interpretation and simplicity, relative to other methods, such as PCA, random forest, or Neural Networks. Relative to other more classical methods, such as latent types analysis, or random effects, k-means can be seen as the limiting problem when the group-specific probabilities are individual specific (e.g. Bonhomme and Manresa (2015)).

3.1 Validating health types

To what extent do health types capture persistent and unobserved differences in health outcomes? What do health types capture, and what are they useful for?

In this section we evaluate the extent to which health types are able to forecast future health and mortality beyond age 60, that is, after the clustering period. Our main premise is that if we are identifying persistent and unobserved differences in health outcomes, health types should have explanatory power after age 60, when the clustering period ends. It is important to note that this is a very different exercise from the one we use to help choose the number of types, K. Here, we are trying to predict the future. In contrast, in the determination of K we are trying to predict within the clustering period for the sample of individuals that we do not not use for clustering.

[3]Other contexts in which k-means has been used in economics is to capture heterogeneity across firms using the empirical distributions of log-wages in each firm (e.g. Bonhomme, Lamadon, and Manresa (2019)).

To do so, for those individuals alive by the end of the clustering period, we compare the forecasting performance of models for health and mortality with and without health types. We have 3,340 individuals in our sample that satisfy these requirements (see Appendix A for details). We use a linear regression model for frailty and a logistic regression model for mortality and measure the predictive power of those models by the adjusted R^2 and the McFadden pseudo-R^2 (hereafter referred to as pseudo-R^2), respectively. Appendix D contains detailed information about our model specification, sample, and estimates.

Table 2: Forecast performance of health types in future health and mortality

	Future frailty				Future mortality			
Controls	x	x	x	x	x	x	x	x
Initial health			x	x			x	x
Health types		x		x		x		x
R^2	0.119	0.571	0.510	0.591				
Pseudo-R^2					0.145	0.206	0.183	0.210

Notes: See Tables 12 and 13 in Appendix D for detailed information.

Table 2 shows that health types significantly improve the predictive power of future health and mortality. Specifically, adding health types in a regression of future frailty on demographic characteristics, the R^2 increases from 0.119 to 0.571 (Columns 1 and 2). In a similar regression of future frailty on demographic characteristics and frailty and self-reported health at age 52, adding health types increases the R^2 from 0.510 to 0.591 (Columns 3 and 4). When analyzing mortality we find similar patterns. Adding health types to our logistic regression of mortality on demographic characteristics increases the pseudo-R^2 from 0.145 to 0.206 (Columns 5 and 6). The improvement in predictive power is also present when we consider a model that incorporates frailty and self-reported health at age 52, and the pseudo-R^2 increases from 0.183 to 0.210 (Columns 7 and 8).

Our analysis thus suggests that: i) demographic characteristics have relatively low predictive power for future health and mortality; ii) health types considerably increase this predictive power; and iii) the predictive power of health types is still significant and sizable

after controlling for frailty and self-reported health at age 52.

A natural question is to think whether health types are a trait or a state. Disentangling high persistence from heterogeneity is well known to be a difficult question in longitudinal dataset. Hence, our goal in this exercise is to use types to uncover additional ways to predict future outcomes, without taking a stand on whether they are pure heterogeneity or the accumulation of highly-persistent shocks. We leave this very important question for future research.

4 What do those health types look like?

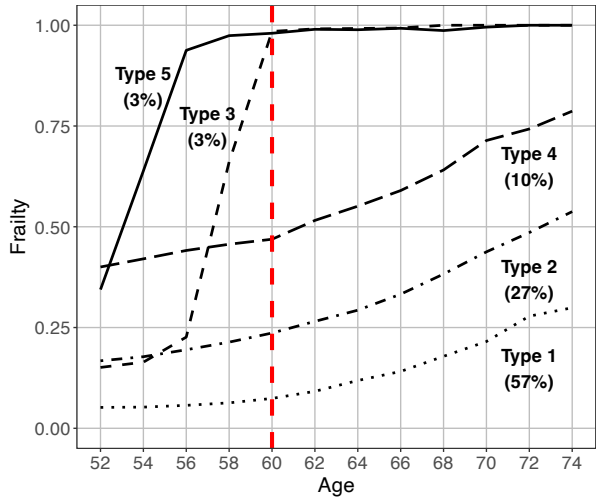
Because health types are established by clustering the data, each individual's health type is defined by the proximity of their realized health path between ages 52 and 60 to the average health path within that cluster. We order the five health types defined by the clustering according to the average frailty between ages 52 and 60 of the centroid of each cluster. Type 5 has the highest average frailty, and Type 1 has the lowest.

Health type 1 includes 57% of the sample and has an average frailty of 0.06 (and 2.1 health deficits) during our clustering period. Health type 2 comprises 27% of the sample and has an average frailty of 0.2 (and 6.9 health deficits). Health types 3 and 4 constitute 3% and 10% of the sample, respectively, and have average frailties of 0.44 (and 15.3 health deficits). Lastly, health type 5 accounts for 3% of the sample and has an average frailty of 0.77 (and 27.1 health deficits).

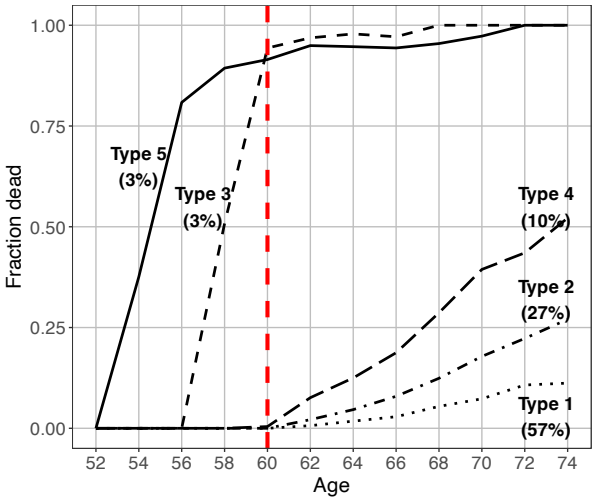
4.1 Health dynamics by health type

Figure 2 illustrates three key health outcomes by health type: (a) mean frailty, (b) fraction of people alive, and (c) mean frailty conditional on survival. They reveal large heterogeneity in health trajectories by health type, both during and after the clustering period.

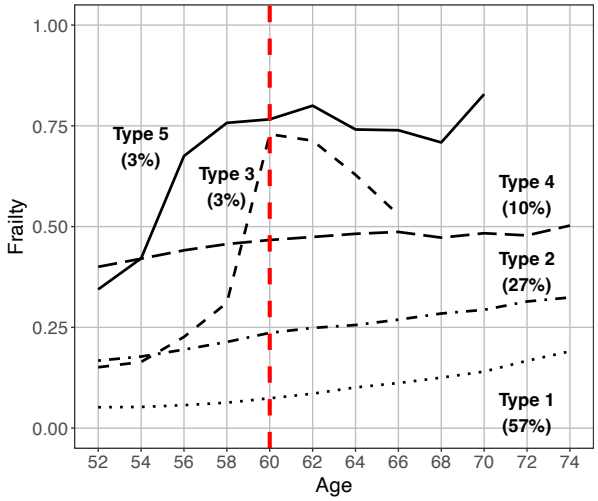
Figure 2: Health dynamics by health type and age



(a) Mean frailty for everyone



(b) Fraction dead



(c) Mean frailty of the survivors

Notes: Red dashed line: end of clustering period

Health at age 52 is very unequally distributed by health type. On average, type 1 has a frailty of 0.05 (1.8 deficits), while types 2 and 3 have frailty levels of 0.17 and 0.15, respectively (5.9 and 5.3 deficits). Health types 4 and 5 start at age 52 with frailty levels of 0.40 and 0.34, respectively (14 and 12 deficits). However, the speed of health deterioration varies greatly among health types.

Health type 1 displays initial low frailty and experiences a slow health decline. Conversely, while health types 2 and 3 start with similar frailty levels at age 52, their health trajectories

diverge notably. Type 2 exhibits slow health deterioration, while type 3 undergoes a more rapid decline with higher mortality rates. Similarly, types 4 and 5 begin with comparable frailty levels, but their health evolution differs significantly. While type 4 maintains relatively high frailty levels, type 5 demonstrates rapid health deterioration and higher mortality.

These results underscore the importance of our methodology. First, clustering based on trajectories rather than other measures, such as mean frailty, offers significant advantages. For example, types 3 and 4 exhibit similar average frailty (0.44), yet their trajectories differ significantly. Second, incorporating the time dimension is crucial. While health types 2 and 3, and 4 and 5 appear similar at age 52, their health trajectories prove to be very different.

Health outcomes after the clustering period are very heterogeneous. While health types 3 and 5 suffer from severe health deterioration during the clustering period and most of them die by the age of 60, types 1, 2, and 4 also face very different frailty dynamics and mortality patterns afterward. Particularly, individuals in type 4 die faster than those in health types 1 and 2, with type 2 also experiencing higher mortality than type 1.

A legitimate concern is whether health dynamics of health types 3 and 5 are mostly driven by people who die by age 60. Panel (c) depicts the mean frailty by health types for those who are alive, evidencing the acute health deterioration of types 3 and 5, even among survivors, while health types 1, 2, and 4 show a steady and slow deterioration.

Figure 3 displays the histories of each individual's frailty by health type and age during the clustering period. Darker trajectories refer a higher level of frailty and black trajectories represent death. These graphs highlight both the differences in initial frailty at age 52 and the degree of individual's level health deterioration. Type 1 starts out with low frailty and remains quite healthy during the clustering period. Types 2 and 4 start out with higher frailty (more so for type 4 than for type 2) and also see their frailty increase slowly. Types 3 and 5 have more heterogeneous frailty at age 52 than those of types 1, 2, and 4, and experience faster and larger health deterioration, with trajectories moving from light grey to darker grey, culminating in the majority of these individuals dying by age 60. The heterogeneity in these

health trajectories can be due to several factors, some of which might not be measurable using our data. For instance, individuals may differ in genetic endowments, family background, pollution, access to high-quality health care, eating and exercise habits, and so on.

Figures 12 and 13 in Appendix E refer to versions of the previous graphs in which we choose 4 and 6 health types, respectively, instead of 5. These figures highlight there is remarkable consistency in the behavior of health types when the number of types increases or decreases around our chosen number of types. While most health types display similar dynamics to those in our base case, a second vulnerable (that is with high health deterioration) type emerges when moving from 4 to 5 types, and an additional resilient (that is with slow health deterioration) type, yet with poorer health, appears when going from 5 to 6 health types. This consistency is not something that the k-means method imposes as we change the number of types, yet it emerges from the data very clearly.

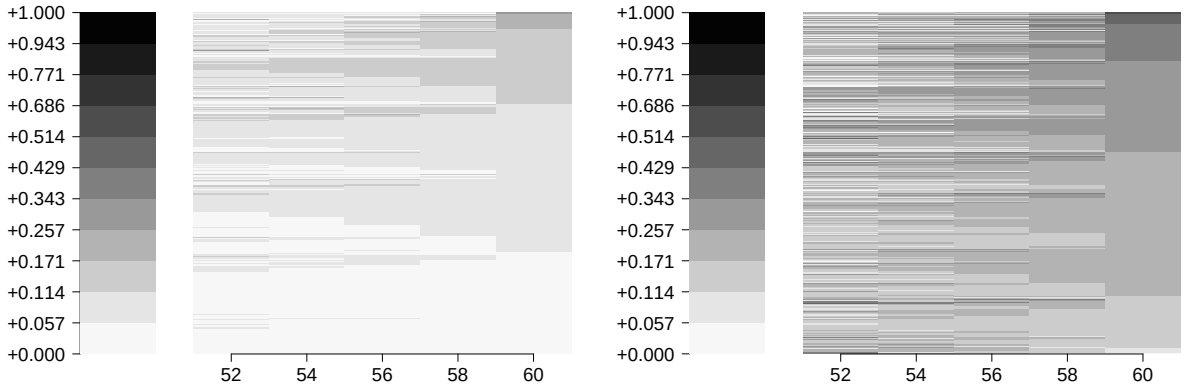
Figures 14 and 15 in Appendix F report the analogous pictures of Figure 2 when we increase the length of our clustering period to age 62 and 64, respectively. They show that both the size of our types and their behaviour before and after the clustering period are remarkably similar.

Finally, Table 3 presents the causes of death among those in our sample.[4] We find that cancer and heart conditions are primary causes, accounting for 61% of deaths, and that 35% can be attributed to other health conditions. This aligns with the National Center for Health Statistics report, which identifies heart disease and cancer as the top two leading causes of death in the United States in 2018 (Xu, Murphy, Kochanek, and Arias (2020)). Only 4% of our sample died from non-health-related reasons. Additionally, approximately half of the deceased individuals (48%) were expected deaths (as reported by the respondent).

Breaking down the data by health types reveals relatively homogeneous distributions of causes of death across them. For example, death due to cancer-related issues accounts for

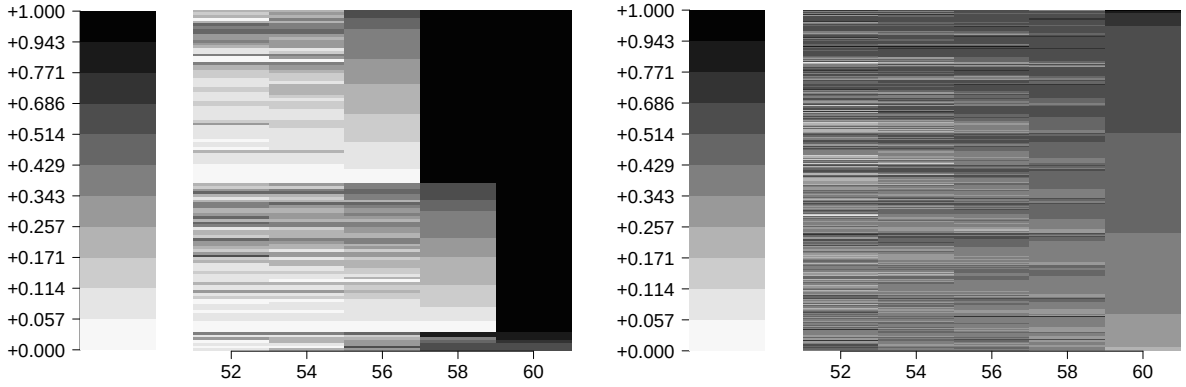
[4]We use the HRS exit interview files. When a participant dies, a family member is interviewed about the participant's death and also asked about the cause of death and if death was expected/unexpected around the time it occurred. About 90% of the people who die has this information available. The response rate for the cause of death is similar across health types.

Figure 3: Individual frailty dynamics by health type and age during clustering period



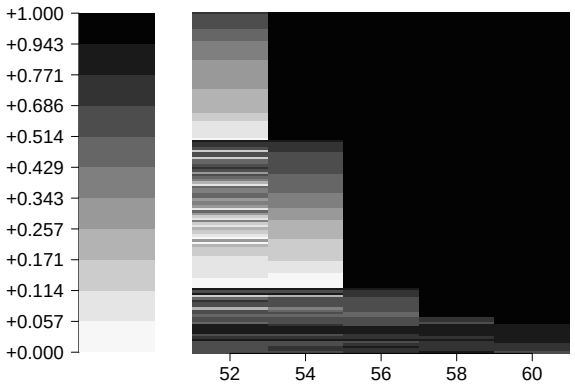
(a) Health type 1

(b) Health Type 2



(c) Health type 3

(d) Health type 4



(e) Health type 5

Notes: Each row in the graphs represents an individual's health trajectory during the clustering period. Darker trajectories indicate higher levels of frailty, with black representing death.

41% of deaths in health types 3 and 29% in health types 5 (most of them died during the clustering period), slightly above and below the overall 35%. Similar patterns emerge for causes of death related to heart issues and whether death was expected or not.

Table 3: Cause of death by health types and all sample

	Death cause				Death expected?		Death	
	Cancer	Heart	Other health-related	Non-health related	Expected	Unexpected	during clustering period	by 2018
Type 1	0.50	0.24	0.24	0.03	0.60	0.40	0.00	0.05
Type 2	0.35	0.30	0.33	0.03	0.49	0.51	0.00	0.10
Type 3	0.41	0.19	0.32	0.08	0.47	0.53	0.94	0.97
Type 4	0.16	0.27	0.55	0.01	0.38	0.62	0.00	0.19
Type 5	0.29	0.30	0.36	0.06	0.44	0.56	0.91	0.96
Overall	0.35	0.26	0.35	0.04	0.48	0.52	0.053	0.129

4.2 Health types and observable characteristics

Research in health dynamics has traditionally focused on analyzing health outcomes conditional on observable characteristics. We deviate from this approach and identify typical health trajectories. This section examines to what extent our health types correlate with commonly used observable characteristics. Table 4 presents these relationships.

We first observe some gender disparities. Women have a lower likelihood of being in good health; but their health tends to deteriorate at a slower pace compared to that of men. Specifically, women are over-represented in health types 2 and 4, while men show a relatively larger presence in health types 1, 3, and 5. These patterns are consistent with the gender-specific health dynamics previously documented in the literature (Case and Deaton (2005); De Nardi, French, Jones, and McGee (2024); Hosseini, Kopecky, and Zhao (2022))

Race is another factor related to health outcomes. Black individuals are less likely to belong to the healthiest type and are relatively more likely to belong to the rapidly deteriorating health groups, such as health types 3 and 5.

As previous work, we find an educational attainment gradient by health types. Individuals with more education are more likely to belong to the healthiest health type. However,

this gradient is weaker for the remaining health types. For example, health types 2 and 3 exhibit similar educational achievement but have distinct health trajectories.

Furthermore, our analysis reveals that coupled individuals are more likely to belong to the healthiest types (1 and 2) and experience lower health deterioration relative to single individuals. These patterns are consistent with previous studies analyzing heterogeneity of health dynamics and marital status (Borella, De Nardi, and Yang (2023); Braun, Kopecky, and Koreshkova (2017); De Nardi, French, Jones, and McGee (2024))

In line with many previous studies, we identify a strong income gradient among health types, where individuals with higher individual incomes are more likely to belong to the healthiest types. Notably, this gradient is weaker when analyzing household income.

We also investigate the prevalence of health-related behaviors among health types. We find that the proportion of individuals who have ever smoked increases in frailty type and is particularly prevalent for health types 3 and 5, which are associated with faster health deterioration. Notably, the fraction of individuals reporting vigorous physical activity at age 52 decreases in frailty type, but is similar for slow and fast health deterioration types. While both behaviors are considered forms of health investment, their differing impacts on mortality challenge the state-of-art health investment model that assumes a uniform measure of "health effort". (Cole, Kim, and Krueger (2019); Mahler and Yum (2024)

Finally, we examine health insurance status at age 52 and find that while private insurance decreases with frailty type, public insurance increases.

4.2.1 Predicting health types

We now turn to studying to what extent heterogeneity in observable characteristics at age 52 explains who belongs to each health type. We do so by estimating the probability of belonging to a health type conditional on observable characteristics at age 52 using a multinomial logit model. The explanatory variables include demographic variables (education, race, gender, HRS cohort, marital status, and household total income), health-related be-

Table 4: Health type and observable characteristics

	All sample	Type 1	Type 2	Type 3	Type 4	Type 5
Fraction of people	1	0.57	0.27	0.03	0.10	0.03
Health outcomes during clustering period						
Average frailty	0.17	0.06	0.20	0.44	0.44	0.77
Average health deficit	5.85	2.09	6.93	15.31	15.31	27.12
Fraction dead by 60	0.05	0	0	0.94	0	0.91
Health at 52						
Average frailty	0.13	0.05	0.17	0.15	0.40	0.34
Average health deficit	4.50	1.80	5.90	5.30	14	12
Average SRHS	2.64	2.12	3.01	3.18	4.01	3.89
Std. Dev. of frailty	0.14	0.04	0.08	0.13	0.13	0.23
Demographics						
Fraction women	0.63	0.59	0.69	0.57	0.74	0.54
Fraction black people	0.17	0.13	0.20	0.27	0.28	0.28
Mean years of education	13.01	13.60	12.42	12.60	11.56	12.36
Fraction partnered at 52	0.78	0.82	0.77	0.67	0.63	0.65
Mean individual income at 52	30,580	38,928	24,200	17,395	10,478	10,304
Mean household income at 52	56,102	69,818	45,340	34,185	22,368	27,852
Health behaviours						
Fraction ever smoked	0.56	0.49	0.64	0.72	0.67	0.77
Fraction vigorous activity at 52	0.50	0.61	0.44	0.46	0.21	0.23
Health insurance status						
Health insurance status						
Private health insurance at 52	0.76	0.85	0.74	0.60	0.42	0.43
Public health insurance at 52	0.13	0.04	0.13	0.20	0.45	0.48
Medicaid	0.06	0.01	0.06	0.08	0.23	0.27
Medicare	0.05	0.01	0.06	0.12	0.25	0.26
Uninsured at 52	0.14	0.12	0.17	0.22	0.21	0.16

haviors (ever-smoked and vigorous activity indicators), health insurance status (public and private health insurance indicators), and frailty at age 52. Table 5 reports the pseudo-R^2 of those models, which we take as our metric for predictive power.

Column 1 refers to a model comprising all observable characteristics but excludes initial frailty at age 52. It has low predictive power (a pseudo-R^2 of 0.131). This suggests that although some relationships emerge in the data between health types and the observables that we consider, these variables only explain a small fraction of the variation in health types.

Column 2 reports results for a model that only includes initial frailty at age 52 and shows that it has higher predictive power, with a pseudo-R^2 of 0.430. Column 3 includes all observable characteristics at age 52, including initial frailty, and shows that observable characteristics have negligible explanatory power for health types once we include initial health (Column 3 and 2). These results support our clustering empirical strategy of clustering only based on health histories and not other observables, as other observables only turn out to be weakly related to health types.

Table 5: Multinomial logistic regression of health type on observable characteristics

	Health types		
	(1)	(2)	(3)
Initial frailty		x	x
Demographics	x		x
Health behaviours	x		x
Health insurance	x		x
Pseudo-R^2	0.131	0.43	0.448

Notes: See Appendix G for further details

We also investigate whether our health types, which are constructed from a health measure that weights each health deficit equally, are related to the composition of deficit prevalence at age 52. We do so by constructing frailty indexes for each deficit subgroup listed in Table 1, estimating our multinomial logit model for health types including these six frailty subcategories, and comparing its predictive performance with those from the multinomial logit that only includes initial frailty at age 52. We find that using the frailty sub-indexes has little effect on the predictive power for health types compared with the models that use only frailty at 52. These findings suggest a high level of comorbidity among deficits: while the composition of frailty varies across types, the equal-weighted index also varies, leaving little role for additional predictive capability. We discuss this more in Appendix G.

5 What health dynamics do we miss if we ignore health types?

Health types are commonly overlooked when estimating health dynamics. What do we miss by doing so?

To answer this question, we turn to what is commonly done in the extensive literature of structural papers modeling health dynamics. That is, we model health as a variable that can take a small number of states and is a Markov process of order one that also depends on a number of observable characteristics, in addition to health during the previous period. Then, we compare what happens in these models in their base case and if we include health types as an additional explanatory variable.

More specifically, we take our previously computed health types based on frailty as given, and use them to study the evolution of a different measure: Self-reported health status (SRHS) and death starting from age 52 and until death or the end of our observation period.

Self-reported health status results from a question in which people are asked to self-report their health as "excellent," "very good," "good," "fair," and "poor." This measure has, at this stage, some important advantages for us. It parsimoniously summarizes an individual's perception about their own health, it is not mechanically related to our frailty measure, and it has also been shown to be a good predictor of many economic outcomes. Moreover, the vast majority of structural models using health status adopts this health measure, which allows us to make our exercise relevant for these studies.[5]

Let $h_{i,t}$ represent the health state of individual i at age t. The variable $h_{i,t}$ takes values in the set $H = \{0, 1, 2, 3, 4, 5\}$, where 0 corresponds to "dead," and 1 to 5 map to self-reported health statuses: "poor," "fair," "good," "very good," and "excellent," respectively. Let g denote a generic health state, such that $g \in H$. We estimate the two-year health/mortality

[5] As Russo, McGee, De Nardi, Borella, and Abram (2024), we could instead set frailty thresholds to map frailty to discrete states. However, efficient discretization implies taking a stand about the underlying process for frailty, which we can avoid by using SRHS.

transition probabilities by fitting the observed transitions to a multinomial logit model. We allow the transition probabilities to depend on one's current health, age, marital status, education, and gender, and take into account that death is an absorbing state (i.e $h_{i,t} = 0 \rightarrow h_{i,t+2} = 0$).

Our assumptions gives the following expression for each probability

$$Pr(h_{i,t+2} = g \mid X_{it}) = \frac{e^{X_{it}\beta_g}}{\sum_{n=1}^{5} e^{X_{it}\beta_n}} \tag{2}$$

Where X_{it} includes age, age squared, current self-reported health status dummies, couple dummies, and education dummies, all interacted with a gender dummy, which translates into gender-specific transition probabilities. This specification parallels state-of-the-art health dynamics models in the literature (See for instance De Nardi, French, Jones, and McGee (2024)) that use the same data set. Our sample includes all individuals to whom we assign a health type from age 52 until death or 2018.

To incorporate health type heterogeneity, we estimate an augmented version of the previous model that accounts for health transitions based on an individual's health type. Specifically, we extend the baseline model by including a vector of health type indicators, $\Theta_i = [1_{\eta=1}, 1_{\eta=2}, 1_{\eta=3}, 1_{\eta=4}, 1_{\eta=5}]$, in the multinomial logistic specification. The model specification is shown in Equation 3.

$$Pr(h_{i,t+2} = g \mid X_{it}, \Theta_i) = \frac{e^{X_{it}\beta_g + \Theta_i\gamma_g}}{\sum_{n=1}^{5} e^{X_{it}\beta_n + \Theta_i\gamma_n}} \tag{3}$$

Table 6 summarizes the main estimation results. The log-likelihood of the model with health types is greater than that of the model without health types. A formal log-likelihood ratio test indicates that these differences are statistically significant, suggesting that health types are important determinants of health dynamics, even when we include past period health and a rich set of observable characteristics. We also compare these models using the Akaike information criterion and the Bayesian information criterion. Models with lower

Table 6: Multinomial logistic regression for health dynamics

	Future SRHS	
	Without health types	Including health types
Log-likelihood	-33454	-31932
Log-likelihood ratio test p-value	-	0.00
Akaike information criterion	67108	64105
Bayesian information criterion	67938	65100
Pseudo-R^2	0.257	0.291

Notes: See the text for a detailed description of the models.

values for these criteria are preferred. The results indicate that both criteria select the model with health types as the better model. Finally, our results show that including health types significantly increases the pseudo-R^2 from 0.257 to 0.291.

But to what extent does ignoring health types lead to misrepresenting the observed heterogeneity in health trajectories? To answer this question, we simulate the health and mortality path from the multinomial model of health dynamics with and without health types and using the initial distribution of health and its covariates at age 52.[6] To understand the implications of ignoring health types we focus on two important metrics from the standpoint of many structural models: the fraction of individuals alive by age and the fraction of individuals in good health (where "good" now includes good, very good, or excellent) conditional on being alive by age. We also compare the models' implied simulation with their counterparts in the observed data.

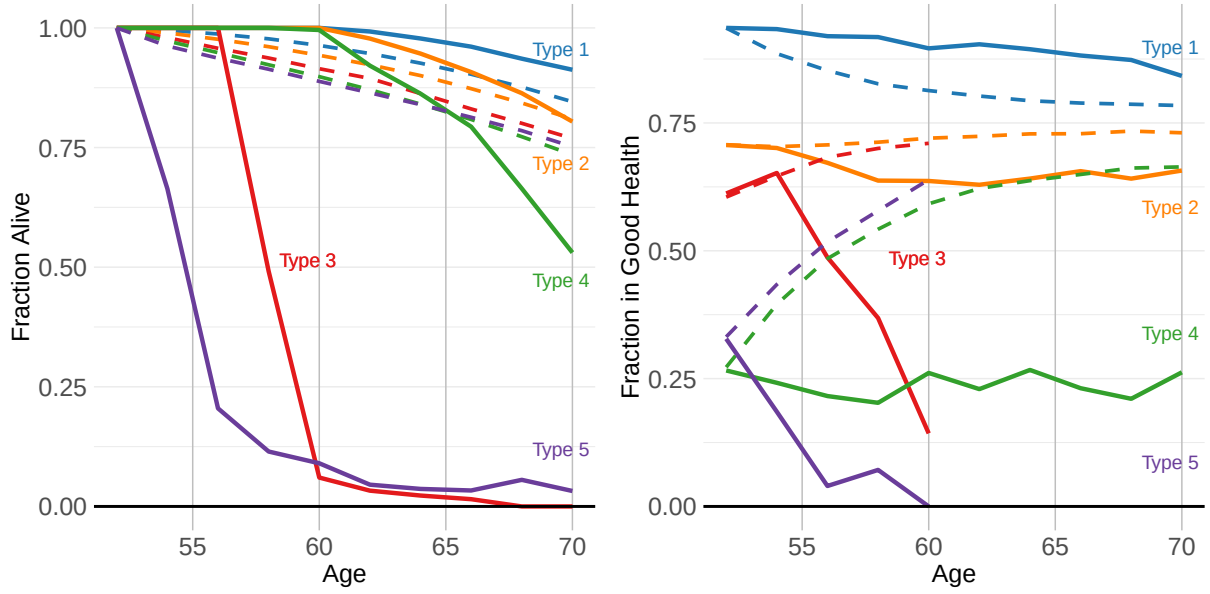
Figure 4 and 5 report the results of the models without and with health types, respectively. The left panel shows the fraction of individuals alive by health type and age, and the right panel shows the fraction of individuals in good health, conditional on being alive. Solid lines refer to the data and dashed line to models' simulations.

The left panel of Figure 4 shows that the model without health types misses the timing and heterogeneity in mortality. That is, it fails to account for the significantly higher mortality rates experienced by health types 3 and 5, and underestimates the magnitude in

[6]We sample the entire sequence of marital status for each individual.

the difference in mortality of types 1, 2, and 4. For example, by age 70, the model implies a mortality gap of about 11 percentage points (pp) between those of types 1 and 4, which is significantly lower than the 38 percentage points (pp) observed gap. Similarly, the model underestimates the mortality gap between health types 1 and 2, predicting a 4 pp difference, while the data indicate that it is nearly 11 pp.

Figure 4: Health dynamics by health type and age. Model without health types



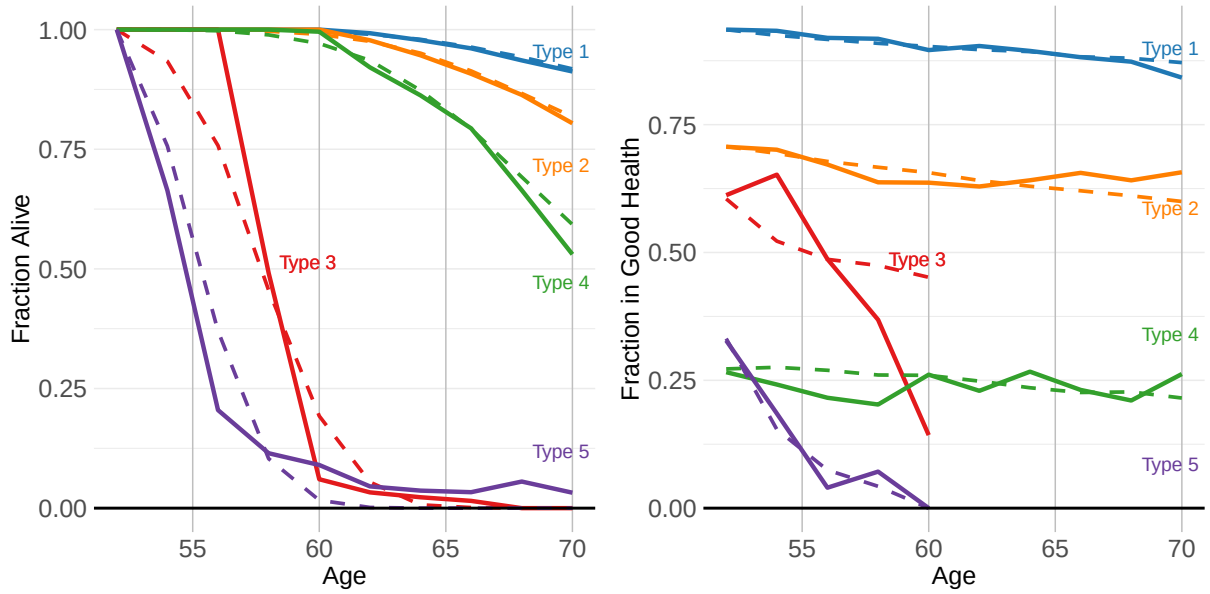
(a) Fraction alive

(b) Fraction in good health

Notes: Solid lines represent the data, while dashed lines depict the model's simulations. Lines in blue, orange, red, green, and purple correspond to health types 1, 2, 3, 4, and 5, respectively. The fraction in good health for types 3 and 5 is plotted only up to age 60 due to reduced cell size. Model simulations are adjusted accordingly.

The right panel of Figure 4 shows that the model without health types also misses the persistence of good health among the living. Specifically, conditional on being alive, it predicts fast mean reversion, forecasting improved health for individuals who initially report poor health. It projects fast recoveries for health types 4 and 5, while in the data these individuals remain in relatively poor health. Conversely, for individuals starting with relatively good health (health type 1), the model forecasts a faster decline than observed in the data. As a result, the model cannot explain the observed gap between the fractions of people reporting good health by health type and age observed in the data. For example, the

Figure 5: Health dynamics by health type and age. Model with health types



(a) Fraction alive

(b) Fraction in good health

Notes: Solid lines represent the data, while dashed lines depict the model's simulations. Lines in blue, orange, red, green, and purple correspond to health types 1, 2, 3, 4, and 5, respectively. The fraction in good health for types 3 and 5 is plotted only up to age 60 due to reduced cell size. Model simulations are adjusted accordingly.

model predicts that the share of people reporting good health in health type 1 will decline from 0.94 at age 52 to 0.78 by age 70, significantly lower than the observed 0.84. Similarly, for those in health type 2, it predicts the fraction reporting good health will increase from 0.71 at age 52 to 0.73 at age 70, significantly higher to the observed 0.66. Overall, the model predicts that the initial 23 percentage points (pp) gap between health type 1 and 2 (0.94-0.71) will reduce to 5 pp by age 70. However, the observed gap at age 70 remains at 19 pp (0.84- 0.66). Large discrepancies between data and model implications are also observed among other health types as well. We do not plot health types 3 and 5 beyond age 60 due to the small number of individuals remaining alive, which would result in noisy estimates.

Figure 5 refers to the model with health types and shows that it generates patterns of mortality and good health by health type that are more consistent with the data. That is, it captures the elevated mortality rates experienced by health types 3 and 5 and the mortality disparities among health types 1, 2, and 4, both before and after age 60. Furthermore, the

model's simulations closely match the fraction of individuals in good health by health types in the data. This alignment persists both during and after the clustering period. Particularly, this version of the model does not predict mean reversion in the fraction of people in good health but, rather, captures well the persistence of these individuals' health states.

These results showcase the importance of health types for capturing the rich heterogeneity in health dynamics and their persistence that we observe in the data. Many economic outcomes, such as labor supply and savings, are based on future health and mortality expectations. Individual choices may vary a lot depending on whether their projected future health state evolves, as illustrated in Figure 4 or Figure 5.

5.1 Modeling health and mortality parsimoniously

The literature modeling health as an exogenous process typically allows health to be a function of several important variables, which include age, gender, marital status, and educational attainment (or permanent income). While estimating processes that take into account this observed heterogeneity may sound appealing, this richness translates into a rapid increase in the number of state variables. This, in turn, increases the computational burden of solving these structural models.

We have so far shown that health types have important implications for health dynamics. This raises the question of whether we can just focus on health types and avoid to explicitly modeling other observable characteristics that the previous literature has spent computational power on so far.

To answer this question, we evaluate how well a parsimonious model of health dynamics that only includes age and health types performs against a model that include the observed heterogeneity that the structural literature has used so far. Specifically, we compare the baseline model without health types (Equation 2) and a simplified version of the model with health types (Equation 3), where X_{it} only includes age, age squared, and current self-reported health status dummies.

Table 7 shows that the log-likelihood of the parsimonious model with health types is greater than the one with demographics but without health types, and that those differences are statistically significant. The Akaike information criterion and the Bayesian information criterion also suggest that the simple model with health types dominates the full model that ignores them. Finally, the pseudo-R^2 of the simpler model with just age and health types is also larger (0.285 against 0.257) that of the model with the rich set of observables.

These results suggest that a parsimonious model that includes health types captures the heterogeneity in health outcomes well, and covariates such as gender, education, and marital status don't need to be included. It is worth noticing that this does not imply that those variables do not affect health dynamics earlier on in life; but rather, that their earlier influence is already reflected in health types once people reach middle age.

Table 7: Multinomial logistic regression for health dynamics

	Future SRHS	
	Without health types	Including health types
Log-likelihood	-33454	-32219
Log-likelihood ratio test p-value	-	0.00
Akaike information criterion	67108	64547
Bayesian information criterion	67938	65003
Pseudo-R^2	0.257	0.285

Notes: See the text for a detailed description of the models.

6 Conclusions and directions for future research

We use HRS data and k-means clustering to identify health types in middle age. We identify five health types: the "vigorous resilient" (57% of our sample), the "fair-health resilient," (27%), the "fair-health vulnerable," (3%), the "frail-resilient," (10%), and the "frail-vulnerable," (3%). Thus, 6% of our individuals experience very fast health deterioration during middle and older ages. We also show that observable characteristics earlier in middle age have little explanatory power for health types and that using health trajectories is

key to identifying health types. Our findings offer valuable insights and highlight promising avenues for future research. First, health types are crucial for understanding health inequality and its potential role in driving economic inequality. Second, ignoring health types can lead to misleading policy implications regarding both observed outcomes and welfare.

Several important questions we leave unadressed. How much historical information is needed to identify health types? This is relevant both operationally and for understanding how long it takes individuals to learn their health type based on family background and personal health history. At what point do individuals become aware of their health type? To what extent do decisions related to retirement, savings, and labor supply depend on health types? Another open question is how health types manifest earlier in life and how they connect to key economic outcomes such as education, marriage, fertility, disability, working life duration, retirement, and medical expenses. Finally, when and how are health types formed? Addressing these critical issues requires substantial further research.

Finally, we identify health types in adulthood and model them as exogenous ex-ante unobserved heterogeneity. However, a health type formation period exists during which individuals invest in their health, possibly well before adulthood. Understanding the factors contributing to health type formation and its timing offers significant potential for future research.

References

Bernstein, Steven L and Benjamin A Toll (2019), "Ask about smoking, not quitting: a chronic disease approach to assessing and treating tobacco use." *Addiction Science & Clinical Practice*, 14, 29.

Bolt, Uta (2021), "What is the source of the health gradient? the case of obesity."

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2019), "A distributional framework for matched employer employee data." *Econometrica*, 87, 699–739.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2022), "Discretizing unobserved heterogeneity." *Econometrica*, 90, 625–643.

Bonhomme, Stéphane and Elena Manresa (2015), "Grouped patterns of heterogeneity in panel data." *Econometrica*, 83, 1147–1184.

Borella, Margherita, Mariacristina De Nardi, and Fang Yang (2023), "Are marriage-related taxes and social security benefits holding back female labour supply?" *The Review of Economic Studies*, 90, 102–131.

Braun, R Anton, Karen A Kopecky, and Tatyana Koreshkova (2017), "Old, sick, alone, and poor: A welfare analysis of old-age social insurance programmes." *The Review of Economic Studies*, 84, 580–612.

Bueren, Jesus, Dante Amengual, and Pijoan-Mas Josep (2024), "Education, healthy habits, and inequality."

Capatina, Elena and Michael Keane (2023), "Health shocks, health insurance, human capital, and the dynamics of earnings and health." Working Paper 080, Federal Reserve Bank of Minneapolis.

Case, Anne and Angus S Deaton (2005), "Broken down by work and sex: How our health declines." In *Analyses in the Economics of Aging*, 185–212, University of Chicago Press.

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

De Nardi, Mariacristina, Eric French, and John B Jones (2010), "Why do the elderly save?" *Journal of Political Economy*, 118, 39–75.

De Nardi, Mariacristina, Eric French, John B Jones, and Rory McGee (2024), "Why do couples and singles save after retirement? household heterogeneity and its aggregate implications." NBER Working Paper No. 28828, Revised July 2023. Accepted for publication in the Journal of Political Economy, 2024.

De Nardi, Mariacristina, Svetlana Pashchenko, and Ponpoje Porapakkarm (2023), "The lifetime costs of bad health." Working Paper Series No. 23963, Revised March 2022. Accepted for publication in The Review of Economic Studies, 2023.

French, Eric (2005), "The effects of health, wealth, and wages on labour supply and retirement behaviour." *The Review of Economic Studies*, 2, 395–427.

French, Eric and John B Jones (2011), "The effects of health insurance and self-insurance on retirement behavior." *Econometrica*, 79, 693–732.

French, Eric and John B Jones (2017), "Health, health insurance, and retirement: a survey." *Annual Review of Economics*, 9, 383–409.

Goggins, William B, Jean Woo, Aprille Sham, and Suzanne C Ho (2005), "Frailty index as a measure of biological age in a chinese population." *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 60, 1046–1051.

Health and Retirement Study (HRS) (2018a), "RAND Fat Files, public use database." Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant numbers NIA U01AG009740 and NIA R01AG073289). Ann Arbor,

MI, (1994-2018). https://hrsdata.isr.umich.edu/data-products/rand (accessed January 14, 2025).

Health and Retirement Study (HRS) (2018b), "RAND HRS Longitudinal File 2018 (V2), public use dataset." Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant numbers NIA U01AG009740 and NIA R01AG073289). Ann Arbor, MI, (2018). https://hrsdata.isr.umich.edu/data-products/rand-hrs-archived-data-products (accessed January 14, 2025) .

Health and Retirement Study (HRS) (2020), "RAND HRS Exit/Post-Exit Interview and Finder Files 2020 (V1)), public use database." Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant numbers NIA U01AG009740 and NIA R01AG073289). Ann Arbor, MI, (2020). https://hrsdata.isr.umich.edu/data-products/rand-hrs-exitpost-exit-interview-and-finder-files-2020 (accessed January 14, 2025).

Health and Retirement Study (HRS) (2022), "Cross-Wave Tracker file, public use database." Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant numbers NIA U01AG009740 and NIA R01AG073289). Ann Arbor, MI, (2022). https://hrsdata.isr.umich.edu/data-products/cross-wave-tracker-file (accessed January 14, 2025).

Heiss, Florian (2011), "Dynamics of self-rated health and selective mortality." *Empirical economics*, 40, 119–140.

Hosseini, Roozbeh, Karen A Kopecky, and Kai Zhao (2021), "How important is health inequality for lifetime earnings inequality?" *Federal Reserve Bank of Atlanta*.

Hosseini, Roozbeh, Karen A Kopecky, and Kai Zhao (2022), "The evolution of health over the life cycle." *Review of Economic Dynamics*, 45, 237–263.

Keeney, Tamra, Emmanuelle Belanger, Rich Jones, Nina Joyce, David Meyers, and Vincent Mor (2019), "High-need phenotypes in medicare beneficiaries: Drivers of variation in utilization and outcomes." *Journal of the American Geriatrics Society*, 68.

Kopecky, Karen A and Tatyana Koreshkova (2014), "The impact of medical and nursing home expenses on savings." *American Economic Journal: Macroeconomics*, 6, 29–72.

Mahler, Lukas and Minchul Yum (2024), "Lifestyle behaviors and wealth-health gaps in germany." *Econometrica*, 92, 1697–1733.

McFadden, Daniel (1975), *Urban Travel Demand: A Behavioral Analysis*, 122–125. North-Holland Publishing Co. Reprinted 1996.

Mitnitski, Arnold, Alexander Mogilner, Chris MacKnight, and Kenneth Rockwood (2002), "The accumulation of deficits with age and the possible invariants of aging." *The Scientific World*, 2, 1816–1822.

Mitnitski, Arnold, Alexander Mogilner, and Kenneth Rockwood (2001), "Accumulation of deficits as a proxy measure of aging." *The Scientific World*, 1, 323–336.

Mitnitski, Arnold, Xiaowei Song, Ingmar Skoog, GA Broe, Jafna Cox, Eva Grunfeld, and Kenneth Rockwood (2005), "Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality." *Journal of American Geriatrics Society*, 53, 2184–2189.

Nygaard, Vegard M (2022), "Causes and consequences of life expectancy inequality."

Ozkan, Serdar (2024), "Income differences and health disparities: Roles of preventive vs. curative medicine." Technical report, Federal Reserve Bank of St. Louis.

Rousseeuw, Peter J. (1987), "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis." *Journal of Computational and Applied Mathematics*, 20, 53–65.

Russo, Nicolo (2022), "Health-depdenent preferences, consumption, and insurance.", URL https://nicolorusso.github.io/jmp_current.pdf. Mimeo.

Russo, Nicolò, Rory McGee, Mariacristina De Nardi, Margherita Borella, and Ross Abram (2024), "Health inequality and economic disparities by race, ethnicity, and gender." Working Paper 32971, National Bureau of Economic Research, URL http://www.nber.org/papers/w32971.

Searle, Samuel, Arnold Mitnitski, Evelyne Gahbauer, Thomas Gill, and Kenneth Rockwood (2008), "A standard procedure for creating a frailty index." *BMC Geriatrics*, 8, 1.

Thorndike, Robert (1953), "Who belongs in the family?" *Psychometrika*.

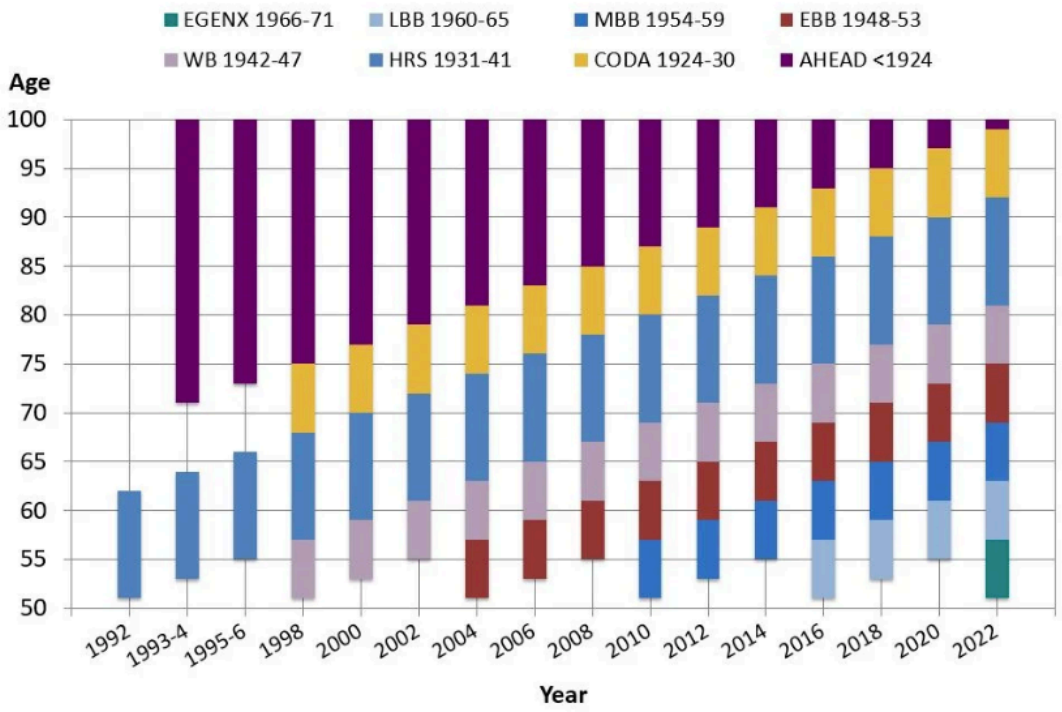
Xu, JQ, SL Murphy, KD Kochanek, and E Arias (2020), "Mortality in the united states, 2018." *NCHS Data Brief, no 355*.

Appendix

A Data, the HRS

The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The HRS data includes eight cohorts which enter the sample in various years. Figure 6 showcases the HRS sampling structure.

Figure 6: HRS sampling scheme



Notes: This graphic illustrates the longitudinal cohort design of the HRS. The initial 1992 HRS cohort included individuals born 1931–1941 (aged 51–61) and their spouses of any age, followed biennially since 1992. Additional cohorts were added over time: AHEAD in 1993 (born before 1924, aged 70+), CODA (1924–1930) and War Babies (1942–1947) in 1998, Early Baby Boomers (1948–1953) in 2004, Mid Baby Boomers (1954–1959) in 2010, Late Baby Boomers (1960–1965) in 2016, and Early Generation X (1966–1971) in 2022. HRS replenishes the sample every six years, always including both members of couples.
Source: *HRS Survey Design and Methodology: Longitudinal Cohort Sample Design.* Available at https://hrs.isr.umich.edu/documentation/survey-design. Accessed: April 10, 2024.

We use the RAND HRS files and the exit files and restrict our analysis to the years 1996 to 2018. The exit files contain information collected in the wave following an individual's

Table 8: Number of individuals each step approaching the sample used for clustering.

Step	Number of individuals
Full sample	46852
Alive and observed at age 52-53	20395
Observed at age 60-61 or older by 2018	12390
Constructible frailty index during clustering	4663

death. Because our data is biennial, we group ages into two-year bins, which means that, for example, people age 52 and 53 are in the same bin. This solves the problem that, due to the biennial nature of the data, we never observe some individuals at even ages and others at odd ages.

There are 46,852 individuals in the main data set. We impose the following screens to ensure that we have enough interviews for each respondent to identity their health type during our clustering period and to validate it after our clustering periods ends.

We start by imposing that, for an individual to be included in our sample, he or she is alive and observed at age 52-53. Hence, we exclude those who enter the HRS study at age 54 or older and include those who enter the sample at younger ages but that we observe at 52-53. This yields a total of 20,395 individuals. Additionally, we require that individuals are 60-61 or older by 2018. Individuals who die before age 60-61, are included as long as their potential age (e.g age if alive) satisfies our criteria. Our restrictions shrink the sample to 12,390 individuals. We also restrict our sample to those for whom we can construct our health measure for all observations during the clustering periods (ages 52-53 to 60-61). This leaves us with 4,663 individuals. Table 8 reports the number of individuals at each stage of our sample selection.

To validate our health types in terms of mortality and frailty predictions after age 60 (see Appendix D), we use the 4,663 individuals to whom we can assign a health type and perform further screenings. First, we need those who are alive by the end of the clustering periods, which includes 4,415 individuals. Second, to evaluate predictive power of our health types on mortality after age 60, we require individuals who have at least one health realization

(including death) after age 60-61, resulting in 3,346 individuals. Third, since our regressions models include several co-variates (e.g., marital status, race, education, among others), we can only use individuals for whom this information is non-missing, resulting in 3,340 individuals and 12,890 observations. Finally, to evaluate the predictive power of our health types on frailty after age 60, we can only use information for individuals who are alive. This last selection yields 3,279 individuals and 11,964 observations. Table 9 summarizes this selection process.

Table 9: Number of individuals each step approaching the sample used for predictive exercise

Step	Number of individuals	N of observations
Assigned a health type	4663	
Alive at 60-61	4415	
At least one health realization after age 60-61	3346	
Non-missing information	3340	12890
Non-missing information and alive	3279	11964

B Frailty distribution and cause of death

Table 10 report the number of deficits and the distribution of frailty while alive for all observations in our sample. This sample includes the 4,663 individuals (from age 52 until they die or 2018) and 34,734 observations.

About 6 percent of cases have a frailty of 0 (i.e., no health deficit) and the median frailty is 0.11, which correspond to 4 health deficits. Moreover, in 98% of the cases alive individuals experience a frailty that is lower than 0.6, which corresponds to 21 health deficits.

Table 11 reports the major illness that lead to death as reported by the HRS exit interview respondent. We group them into 4 categories[7] and show their distribution conditional on

[7]There are 13 categories in the HRS Exit interview files: 1) Cancers and tumors; skin conditions, 2) Musculoskeletal system and connective tissue, 3) Heart, circulatory and blood conditions, 4) Allergies; hay fever; sinusitis; tonsillitis, 5) Endocrine, metabolic and nutritional conditions, 6) Digestive system (stomach, liver, gallbladder, kidney, bladder) 7) Neurological and sensory conditions, 8) Reproductive system and prostate conditions 9) Emotional and psychological conditions, 10) Miscellaneous, 11) Other symptoms, 12) Not a health condition, and 13) Other health condition.

dying either during the clustering period or by 2018. Of these 4,663 individuals, 600 die by 2018. We have a known cause of death for 526 of them (88%). The data show that most deaths are due to health-related causes and that cancer and heart conditions are leading causes of death.

Table 10: Number of deficit and frailty distribution in our sample

Number of Deficits	Average Frailty	Freq.	Percent.	Cumul Percent.
0	0.00	2026	5.83	5.83
1	0.03	4771	13.74	19.57
2	0.06	4953	14.26	33.83
3	0.09	4050	11.66	45.49
4	0.11	3462	9.97	55.46
5	0.14	2792	8.04	63.49
6	0.17	2088	6.01	69.51
7	0.20	1691	4.87	74.37
8	0.23	1321	3.80	78.18
9	0.26	1265	3.64	81.82
10	0.29	1020	2.94	84.76
11	0.31	839	2.42	87.17
12	0.34	635	1.83	89.00
13	0.37	664	1.91	90.91
14	0.40	569	1.64	92.55
15	0.43	481	1.38	93.93
16	0.46	418	1.20	95.14
17	0.49	328	0.94	96.08
18	0.51	254	0.73	96.81
19	0.54	209	0.60	97.41
20	0.57	183	0.53	97.94
21	0.60	180	0.52	98.46
22	0.63	147	0.42	98.88
23	0.66	120	0.35	99.23
24	0.69	69	0.20	99.43
25	0.71	55	0.16	99.59
26	0.74	34	0.10	99.68
27	0.77	38	0.11	99.79
28	0.80	33	0.10	99.89
29	0.83	17	0.05	99.94
30	0.86	15	0.04	99.98
31	0.89	6	0.02	100.00
32	0.91	1	0.00	100.00

Table 11: Death cause

	Fraction dead	Cancer / tumors	Heart, circ. and blood conditions	Other health related	Non-health related
Dead during clustering period	5.3%	34.1%	25.6%	33.2%	7.2%
Dead by 2018	12.9%	34.8%	26.0%	34.8%	4.4%

Notes: Other health related include musculoskeletal system and connective tissue, allergies; hay fever; sinusitis; tonsillitis, endocrine, metabolic and nutritional conditions, digestive system (stomach, liver, gallbladder, kidney, bladder), neurological and sensory conditions, reproductive system and prostate conditions, emotional and psychological conditions, miscellaneous, other symptoms, and other health conditions.

C Choosing the number of clusters or health types

We select our number of clusters, or health types, based on an economic criterion and traditional machine learning criteria. The economic criterion is that adding health types no longer increases their predictive power for future frailty and mortality during the clustering period.

To evaluate the predictive power of health types on frailty, we use the following linear regression model:

$$f_{i,t} = X_{it}\beta + \Theta_i(\bar{k})\gamma + \epsilon_{i,t} \tag{4}$$

where X_{it} includes age, age squared, age cubed, education dummies, race dummies, gender dummies, HRS cohort dummies, and marital status dummies. $\Theta_i(\bar{k})$ is a vector of health type indicators when $\bar{k}$ clusters are considered:

$$\Theta_i(\bar{k}) = \left[1_{i,\eta=1}, 1_{i,\eta=2}, \ldots, 1_{i,\eta=\bar{k}}\right]$$

To evaluate the predictive power of health types on mortality, we use a logit model for the probability of dying at age t. Let $D_{i,t}$ be a binary variable equal to 1 if individual i is dead at age t, and 0 otherwise. The logit model is specified as:

$$Pr(D_{i,t} = 1|X_{it}, \Theta_i(\bar{k})) = \frac{e^{X_{it}\beta+\Theta_i(\bar{k})\gamma}}{1 + e^{X_{it}\beta+\Theta_i(\bar{k})\gamma}} \tag{5}$$

The predictive power of a model with $\bar{k}$-clusters is defined as

$$P(\bar{k}) = 1 - \frac{a(\bar{k})}{a(1)},$$

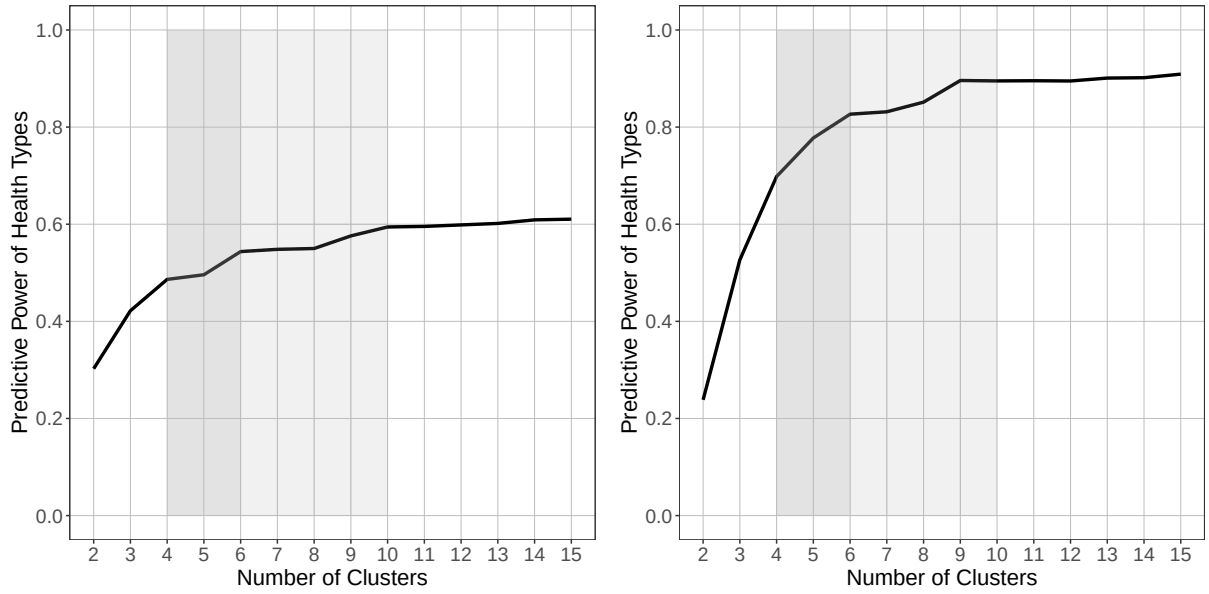
where $a(\bar{k})$ is the mean absolute error for a model with $\bar{k}$-health types, and $a(1)$ is the mean absolute error for a model without health types.

We perform both regressions and predictions during the age range used to cluster health types. To estimate $P(\bar{k})$, we use 10-fold cross-validation as follows. We split the 4,663 individuals into 10 roughly equal-sized subsamples. For each subsample, we split the data into that subsample and its complement and name them the test and training sample, respectively. We run k-means clustering with $\bar{k}$ clusters on the training sample, recovering individuals' health type. We then run regression 4 and 5 on the training sample, and store the estimated coefficients. Using the centroids established by running k-means on the training sample, we assign each individual in the test sample to the cluster with the nearest centroid, and use the stored coefficients to predict frailty and mortality in the test sample. Finally we calculate $\hat{P}_n(k) = 1 - \frac{\hat{a}_n(k)}{\hat{a}_n(1)}$, where $\hat{a}_n(1)$ and $\hat{a}_n(k)$ are calculated using the predictions on the n test sample. That is, for each test sample, we construct an estimate of the predictive power of $\bar{k}$-health types for frailty and mortality. Finally, we estimate the predictive power of $\bar{k}$-health types by taking the average across them, that is: $\hat{P}(\bar{k}) = \frac{\sum_{n=1}^{10} \hat{P}_n(\bar{k})}{10}$.

Figure 7 displays the predictive power of health types as a function of the number of health types, $\bar{k}$. These results suggest that, for frailty, the gain from moving from 3 to 4 clusters is substantial, and that the gain of having more than 6 clusters is very small. Because a clear elbow is present at 5 clusters for predictions of mortality, we consider the gain associated with moving from 4 to 5 clusters to be substantial, and proceed with 5 clusters.

We also explore model specifications in which we include $f_{i,52}$ and $s_{i,52}$, that is one's

Figure 7: The predictive power of health types as a function of the number of health types



(a) Frailty (b) Mortality

Notes: Dark shaded areas indicate zones with the highest increase in predictive power, while light shaded areas represent zones with smaller increases in predictive power.

frailty and self-reported health[8] at age 52, as follows

$$f_{i,t} = X_{it}\beta + \Theta_i(\bar{k})\gamma + \beta_s s_{i,52} + \beta_f f_{i,52} + \epsilon_{i,t}. \tag{6}$$

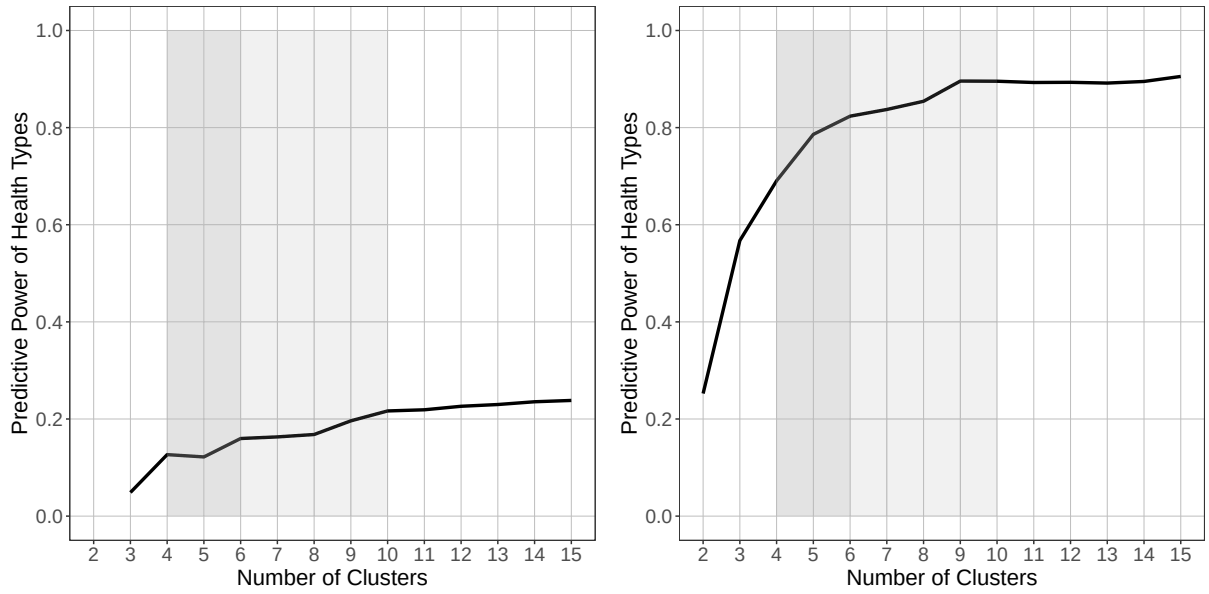
$$Pr(D_{i,t} = 1 | X_{it}, \Theta_i(\bar{k}), s_{i,52}, f_{i,52}) = \frac{e^{X_{it}\beta + \Theta_i(\bar{k})\gamma + \beta_s s_{i,52} + \beta_f f_{i,52}}}{1 + e^{X_{it}\beta + \Theta_i(\bar{k})\gamma + \beta_s s_{i,52} + \beta_f f_{i,52}}} \tag{7}$$

These results (see Figure 8) suggest that health types have significant predictive power for future frailty and mortality, even after controlling for frailty and self-reported health at age 52.

Finally, we adopt two conventional machine learning criteria to evaluate how many clusters we should use. They are the silhouette and elbow methods. The silhouette of a cluster (see Rousseeuw (1987)) is a measure that increases with the average distance between clus-

[8]HRS respondents are asked to rate their health as excellent, very good, good, fair, and poor health, which are used to compute the SRHS by assigning numerical values (from 5 to 1) to each of the health states. The SRHS has been shown to be associated with key health and economic outcomes, including labor earnings and mortality (De Nardi et al. (2023)).

Figure 8: The predictive power of health types as a function of the number of health types, when including one's frailty and self-reported health at age 52



(a) Frailty (b) Mortality

Notes: Dark shaded areas indicate zones with the highest increase in predictive power, while light shaded areas represent zones with smaller increases in predictive power.

ters and decreases with variance within clusters. More formally, given some point i, letting $i \in C_I$ for some cluster C_I, define:

$$a(i) = \frac{1}{|C_I| - 1} \sum_{j \in C_I, j \neq i} d(i,j)$$
$$b(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{j \in C_J} d(i,j)$$

Where $|\cdot|$ gives set size and d is the Euclidean distance, so that $a(i)$ is the mean distance between i and other points within the same cluster and $b(i)$ is the mean distance between i and the points in the nearest cluster. Then the silhouette at point i is given by:

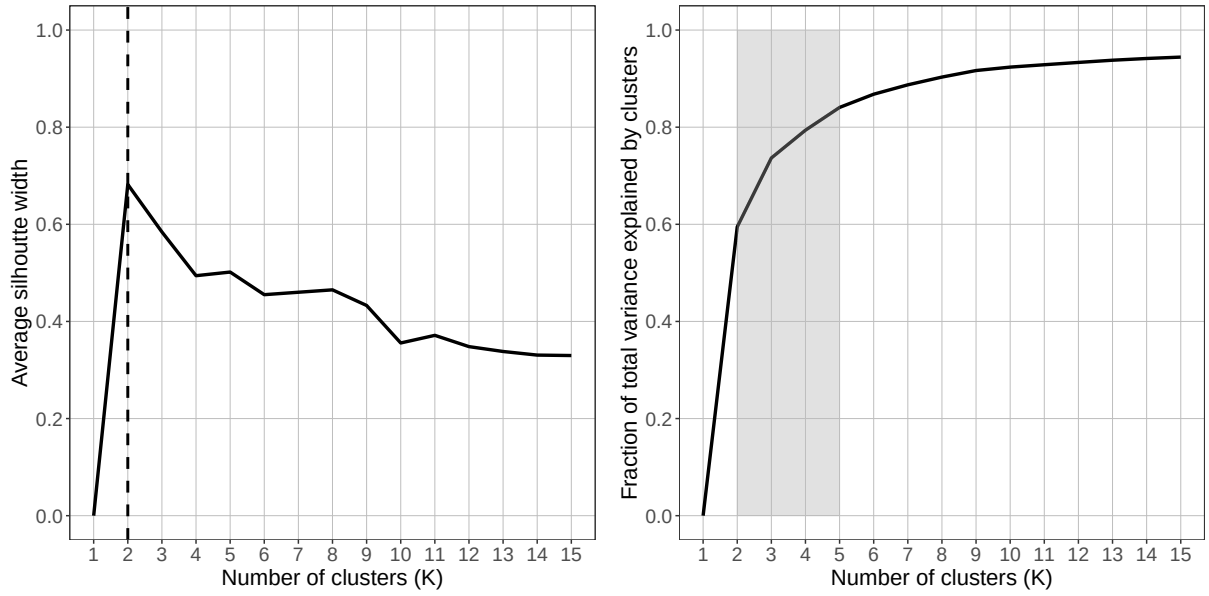
$$s(i) = \begin{cases} 0 & |C_I| = 1 \\ \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} & \text{otherwise} \end{cases}$$

The silhouette criterion states that we should select the number of clusters that maximizes

the average silhouette of the clustering.

The elbow method (see Thorndike (1953)) involves plotting within cluster variance against the number of clusters, and selecting a number of clusters such that within-cluster variance decreases very little past that point. This will appear as an "elbow" in the graph. An inversion of this plot, which will give the same optimal number of clusters, uses the proportion of variance explained by the clusters rather than within-cluster variance.

Figure 9: Clustering frailty trajectories: Average silhouette against the number of clusters on the left-head-side and proportion of total variance explained by clusters against the number of clusters on the right-hand-side.



(a) Silhouette (b) Elbow

Notes: Silhouette analysis: Black dashed line represents the maximum average silhouette at 2 health types. Elbow analysis: Dark shaded areas indicate zones where an "elbow" appears, suggesting a potential number of health types.

The graph on the left-hand-side of Figure 9 displays the silhouette of clustering frailty trajectories by the number of clusters, while its left-hand-side displays the results for the Elbow method. The silhouette method suggests two clusters and the elbow method suggests a range of 2 to 5 clusters. Therefore, selecting 5 clusters is consistent with the recommendations of these criteria.

D Validating health types: Out-of-sample forecasting exercise

In this section, we explain how we conduct our out-of-sample forecast exercise for frailty and mortality after age 60. To evaluate the predictive power of health types for future frailty, we only include observations for those who are still alive at age 60 and for the periods during which individuals are alive and have non-missing information. We run the following specifications

$$f_{i,t} = X_{it}\beta + \Theta_i(5)\gamma + \epsilon_{i,t}, \tag{8}$$

$$f_{i,t} = X_{it}\beta + \Theta_i(5)\gamma + \beta_s s_{i,52} + \beta_f f_{i,52} + \epsilon_{i,t}. \tag{9}$$

Where X_{it} includes, age, age squared, third degree polynomial in age, education dummies, race dummies, gender dummies, HRS cohort dummies, and marital status dummies. The term $\Theta_i(5)$ represents a vector of health type indicators derived from our clustering exercise, based on five health types. Additionally $f_{i,52}$ and $s_{i,52}$ represent one's frailty and self-reported health score at age 52, respectively.

To evaluate the predictive power of health types for future mortality, we include in our sample observations for those who are still alive at age 60, we keep those who die after that age (and until 2018), and we flag them as dead until 2018.[9] This procedure provides a better sense of the effects of time-invariant factors like health types and first period frailty on how long one lives. We model the probability of dying at age t as the following logistic regressions:

$$Pr(D_{i,t} = 1|X_{it}, \Theta_i(5)) = \frac{e^{X_{it}\beta + \Theta_i(5)\gamma}}{1 + e^{X_{it}\beta + \Theta_i(5)\gamma}}, \tag{10}$$

$$Pr(D_{i,t} = 1|X_{it}, \Theta_i(5), s_{i,52}, f_{i,52}) = \frac{e^{X_{it}\beta\Theta_i(5)\gamma + \beta_s s_{i,52} + \beta_f f_{i,52}}}{1 + e^{X_{it}\beta\Theta_i(5)\gamma + \beta_s s_{i,52} + \beta_f f_{i,52}}}. \tag{11}$$

Table 12 reports the results for these regressions. The first column refers to Equation

[9] We also fill in X_{it} using the last information available for the individual in case of death.

8. Our excluded education category is not degree. Thus, higher education is associated with lower frailty compared to those with no degree, and those with a bachelor's or master's degree have the lowest frailty. Black people have, on average, a frailty index 0.04 higher than white people, corresponding to 1-2 additional health deficits. Women have a slightly higher frailty index than men. Married individuals have a frailty index 0.046 lower than single individuals, corresponding to 1-2 fewer health deficits. The R^2 on this restricted regression is 0.120, and thus suggests that demographics explain little of the variation in future frailty.

The second column of Table 12 reports the results for Equation 8. All health types have significantly higher frailty after age 60 than type 1, on average. The inclusion of health types increases the R^2 of the regression from 0.120 to 0.571, which implies that health types explain the majority of variation in health not explained by demographics.

The third column of Table 12 reports the results for Equation 9 when the coefficients of health types are restricted to be zero. Future frailty increases in both initial frailty and self-reported health. The R^2 of the regression is 0.510, lower than the R^2 associated with the regression including health types. This is noteworthy because frailty is nearly continuous, while health types are discrete and take on only 5 values.

The final column of Table 12 shows the estimated coefficients for Equation 9. All health types have significantly higher frailty than Type 1, and higher first period frailty and self-reported health are associated with higher future frailty. The R^2 of this regression is 0.591.

Thus, health types explain a large amount of variation in future frailty. Health type, frailty, and self-reported health in the first period all have an independent influence on future frailty. Both health types and first period frailty and self-reported health explain much more variation in future frailty than demographics. Health types explain more variation in future frailty than initial frailty and self-reported health.

Table 13 displays the estimates for the logistic regression models for mortality. We use the McFadden pseudo-R^2 as measure of goodness of fit (see McFadden (1975)).

The first column of Table 13 reports the results for regression of mortality on demograph-

Table 12: Regressions of future health on demographics, health type, and initial frailty and self-reported health.

	Dependent variable:			
	frailty_bl			
	(1)	(2)	(3)	(4)
age	0.113	−0.070	−0.067	−0.096
	(0.389)	(0.271)	(0.291)	(0.266)
age^2/100	−0.172	0.089	0.095	0.130
	(0.576)	(0.402)	(0.431)	(0.393)
age^3/10000	0.090	−0.032	−0.040	−0.053
	(0.284)	(0.198)	(0.212)	(0.194)
GED	0.007	−0.009*	0.004	−0.004
	(0.008)	(0.005)	(0.006)	(0.005)
HS	−0.050***	−0.007**	0.003	0.002
	(0.004)	(0.003)	(0.003)	(0.003)
HS/GED	−0.063***	−0.009***	−0.003	0.001
	(0.005)	(0.003)	(0.004)	(0.003)
Associate's	−0.056***	−0.001	−0.001	0.006
	(0.007)	(0.005)	(0.005)	(0.005)
Bachelor's	−0.107***	−0.028***	−0.021***	−0.014***
	(0.005)	(0.004)	(0.004)	(0.004)
Master's	−0.116***	−0.026***	−0.020***	−0.011***
	(0.006)	(0.004)	(0.004)	(0.004)
Doctorate	−0.112***	−0.037***	−0.026***	−0.021***
	(0.009)	(0.006)	(0.007)	(0.006)
Black	0.040***	0.024***	0.017***	0.020***
	(0.004)	(0.003)	(0.003)	(0.003)
Other Non-White	0.015***	0.011***	0.007	0.007*
	(0.006)	(0.004)	(0.004)	(0.004)
Woman	0.013***	−0.002	0.003	−0.001
	(0.003)	(0.002)	(0.002)	(0.002)
Cohort 5	0.004	−0.002	−0.002	−0.003
	(0.009)	(0.006)	(0.007)	(0.006)
Married	−0.046***	−0.010***	−0.012***	−0.008***
	(0.003)	(0.002)	(0.002)	(0.002)
Type 2		0.150***		0.111***
		(0.002)		(0.003)
Type 3		0.536***		0.522***
		(0.041)		(0.040)
Type 4		0.358***		0.250***
		(0.004)		(0.006)
Type 5		0.645***		0.432***
		(0.027)		(0.028)
f_0			0.794***	0.273***
			(0.011)	(0.015)
s_0			0.016***	0.011***
			(0.001)	(0.001)
Constant	−2.325	1.783	1.521	2.273
	(8.740)	(6.103)	(6.539)	(5.974)
Observations	11,964	11,964	11,875	11,875
R^2	0.120	0.571	0.510	0.591
Adjusted R^2	0.119	0.571	0.510	0.591
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 13: Regressions of future mortality on demographics, health type, and initial frailty and self-reported health.

	Dependent variable:			
	Dead			
	(1)	(2)	(3)	(4)
age	30.090***	29.073***	31.186***	29.905***
	(10.740)	(11.068)	(10.929)	(11.103)
age^2/100	−42.449***	−40.918**	−44.006***	−42.115***
	(15.603)	(16.080)	(15.878)	(16.131)
age^3/10000	20.088***	19.339**	20.836***	19.914**
	(7.545)	(7.777)	(7.679)	(7.801)
GED	0.649***	0.549***	0.739***	0.631***
	(0.155)	(0.163)	(0.161)	(0.165)
HS	−0.171	0.112	0.175	0.162
	(0.104)	(0.110)	(0.109)	(0.111)
HS/GED	−0.163	0.133	0.200*	0.197
	(0.112)	(0.119)	(0.119)	(0.122)
Associate's	0.103	0.445***	0.476***	0.530***
	(0.163)	(0.170)	(0.169)	(0.171)
Bachelor's	−0.767***	−0.259*	−0.211	−0.152
	(0.150)	(0.157)	(0.157)	(0.159)
Master's	−1.335***	−0.715***	−0.728***	−0.603**
	(0.226)	(0.232)	(0.233)	(0.234)
Doctorate	−1.501***	−1.288***	−1.098***	−1.239***
	(0.397)	(0.437)	(0.410)	(0.440)
Black	0.403***	0.265***	0.290***	0.246**
	(0.094)	(0.099)	(0.096)	(0.099)
Other Non-White	0.238	0.136	0.159	0.104
	(0.153)	(0.160)	(0.157)	(0.160)
Woman	−0.437***	−0.600***	−0.523***	−0.587***
	(0.078)	(0.081)	(0.080)	(0.082)
Cohort 5	0.851**	0.795**	0.781**	0.755**
	(0.360)	(0.370)	(0.365)	(0.372)
Married	−0.573***	−0.353***	−0.395***	−0.359***
	(0.076)	(0.080)	(0.079)	(0.080)
Type 2		0.914***		0.886***
		(0.086)		(0.100)
Type 3		4.663***		4.553***
		(0.535)		(0.540)
Type 4		1.897***		1.890***
		(0.110)		(0.185)
Type 5		3.960***		4.201***
		(0.449)		(0.531)
f_0			2.736***	−1.041**
			(0.320)	(0.493)
s_0			0.254***	0.205***
			(0.043)	(0.045)
Constant	−716.041***	−694.893***	−743.405***	−714.596***
	(246.096)	(253.578)	(250.401)	(254.391)
Pseudo-R^2	0.145	0.206	0.183	0.21
Observations	12,890	12,890	12,797	12,797
Log Likelihood	−2,848.792	−2,643.110	−2,707.151	−2,618.481
Note:				*p<0.1; **p<0.05; ***p<0.01

ics only. People with a GED are more likely to die than people with no degree. People with bachelors, masters, and doctoral degrees are the least likely to die in a given period. Black people are more likely than white people to die earlier, and women are less likely than men to die earlier. The pseudo-R^2 of this regression is 0.145.

The second column of Table 13 refers to Equation 10. All health types are significantly more likely to die than type 1, and type 4 is the most likely to die. The pseudo-R^2 of this regression increases to 0.206.

The third column of Table 13 displays results for Equation 11 and restricts the coefficient of health types to be zero. An increase in either first period frailty or self-reported health significantly increases the likelihood of dying in a given period. The pseudo-R^2 of this regression is 0.183. Hence, health types explain more mortality variation than first period frailty and self-reported health.

The final column of Table 13 pertains to Equation 11. All health types are associated with a higher likelihood of dying than individuals of type 1. There is a significant negative relationship between first period frailty and mortality, but individuals with higher first period self-reported health are more likely to die earlier. The pseudo-R^2 of this regression is 0.210, which is not much higher than that for the regression including only health types.

Hence, health types explain more variation in mortality than first period frailty and self-reported health.

More generally, demographic characteristics explain very little of the variation in both future health and mortality. Both health types and frailty and self-reported health at age 52 have more explanatory power for both future health and mortality. Health types explain more variation in both frailty and mortality than first period frailty and self-reported health.

D.1 Sensitivity to the number of health types

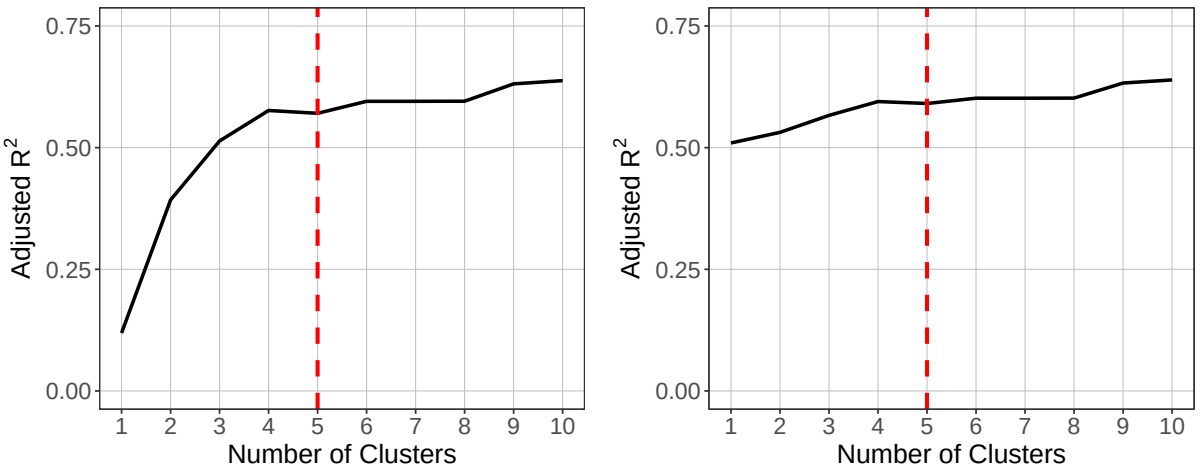
We have first shown that five health types help predict health dynamics during our clustering period. We have then validated our health types by showing that they have

predictive power for frailty and mortality after our clustering period.

As an additional way to validate the choice of our number of types, we now turn to examining how the results from the latter exercise change if we use a different number of health types. Figure 10 depicts the R^2 of the regressions in Equations 8 (left) and 9 (right), when the number of health types $\bar{k}$, ranges from 1 (no heterogeneity in health types) to 10. The red vertical dotted line indicates our benchmark number of health types. We observe a clear increase in predictive power when moving from no health type heterogeneity to five health types, with the predictive power remaining quite flat beyond five health types. This strongly supports five health types as the maximum number needed. However, we also find that four health types perform as well as five in the out-of-sample exercise. This result holds for models including initial health information.

Figure 11 shows the pseudo-R^2 for the logistic regression of the probability of dying, represented by Equations 10 (left) and 11 (right). The patterns are similar. The predictive power of health types increases from no health type heterogeneity to five health types, then remains flat. As in the previous exercise, four health types perform as well as five in the out-of-sample exercise.

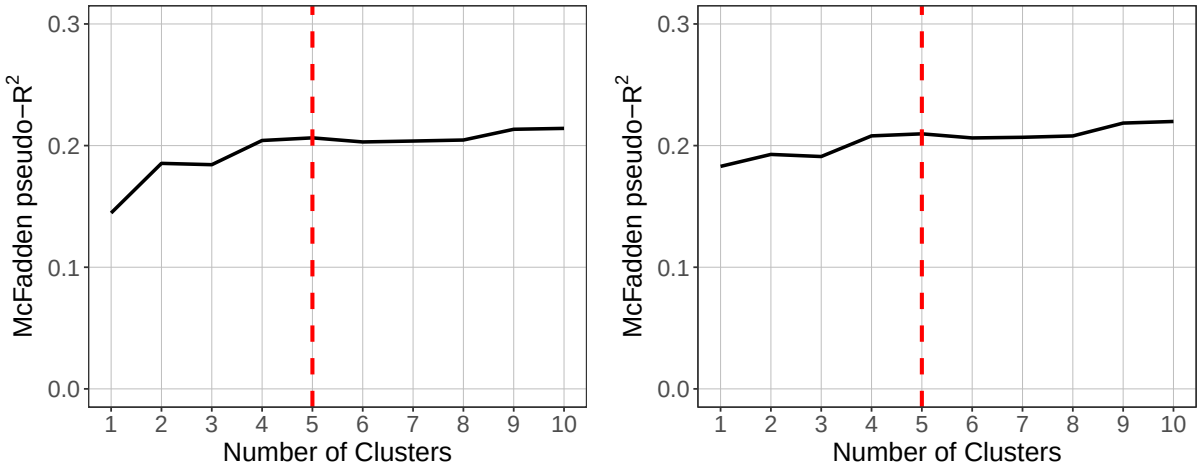
Figure 10: Adjusted R^2 for regressions of frailty after age 60 by number of health type



(a) Demographics (b) Demographics and health at 52

Notes: The dotted line is our benchmark number of health types

Figure 11: McFadden pseudo-R^2 for the probability of dying after age 60 by number of health type



(a) Demographics

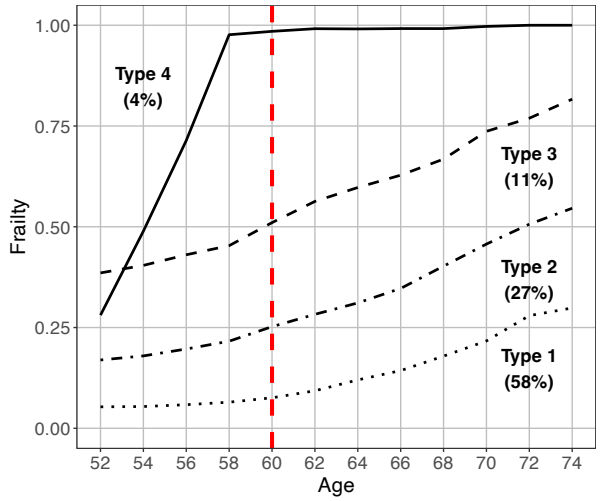
(b) Demographics and health at 52

Notes: The dotted line is our benchmark number of health types

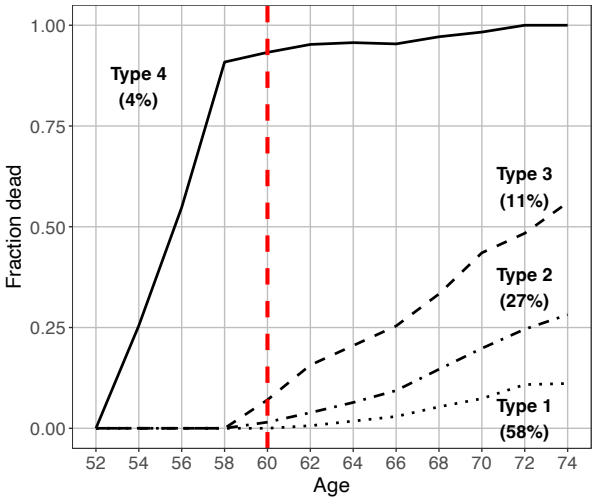
E How do health types look with four and six clusters?

Figures 12 and 13 report the analogous pictures of Figure 2 in the main text when we choose 4 and 6 health types, respectively, instead of 5. They highlight there is remarkable consistency in the behavior of health types when the number of types increases or decreases around our chosen number of types. While most health types display similar dynamics to those in our base case, a second vulnerable (that is with high health deterioration) type emerges when moving from 4 to 5 types, and an additional resilient (that is with slow health deterioration) type, yet with poorer health, appears when going from 5 to 6 health types. This consistency is not something that the k-means method imposes as we change the number of types, yet it emerges from the data very clearly.

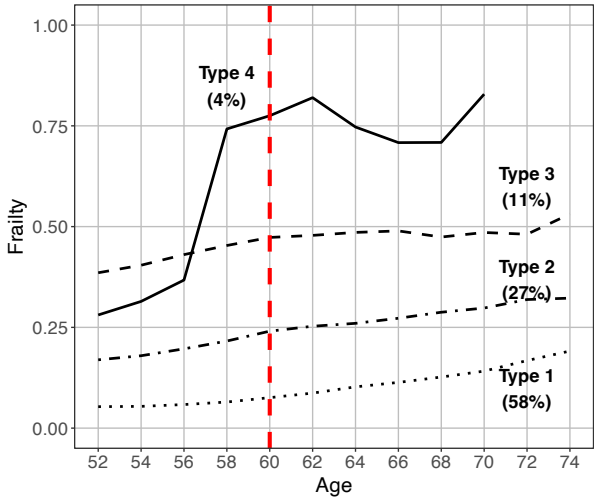
Figure 12: Health dynamics by health type and age for 4 health types



(a) Mean frailty for everyone



(b) Fraction dead



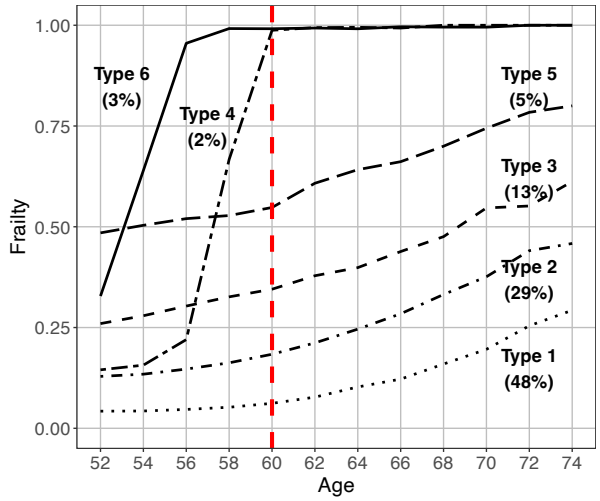
(c) Mean frailty of the survivors

Notes: Red dashed line: end of clustering period

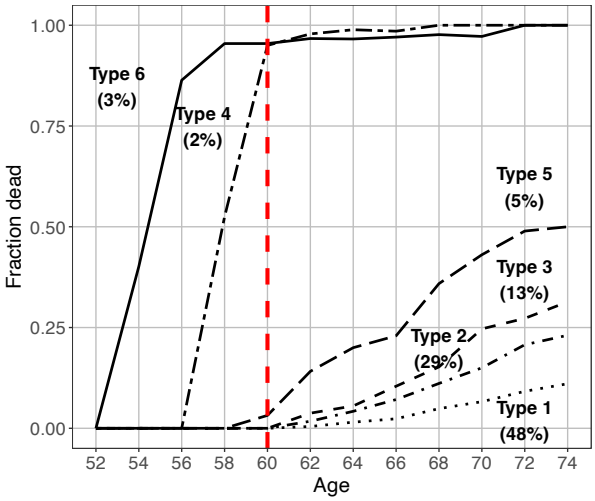
F How do health types look with a longer clustering period?

Figures 14 and 15 report the analogous pictures of Figure 2 in the main text when we increase the length of our clustering period to age 62 and 64, respectively, instead of 60. They show that both the size of our types and their behaviour before and after the clustering

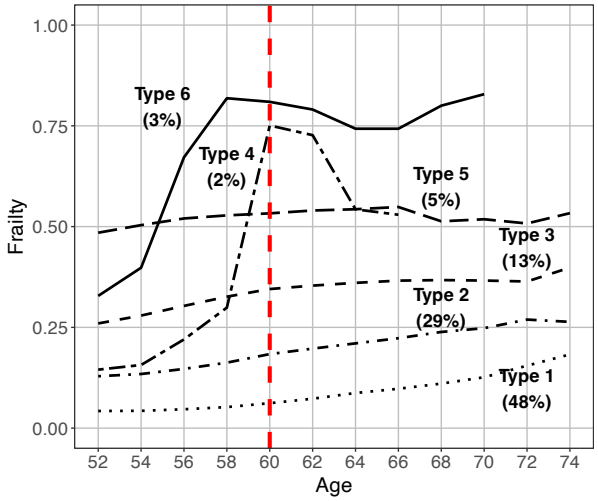
Figure 13: Health dynamics by health type and age for 6 health types



(a) Mean frailty for everyone



(b) Fraction dead

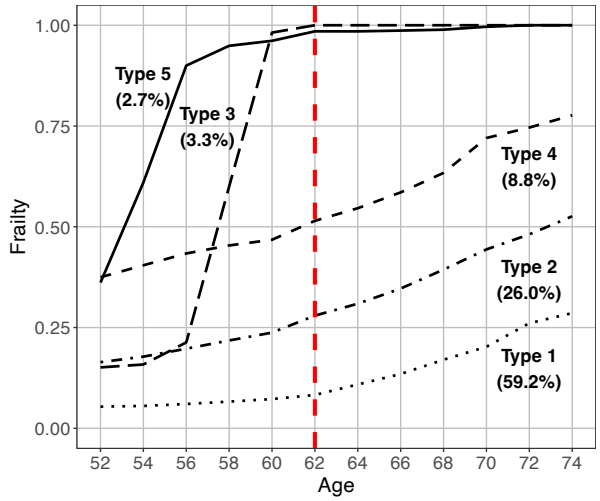


(c) Mean frailty of the survivors

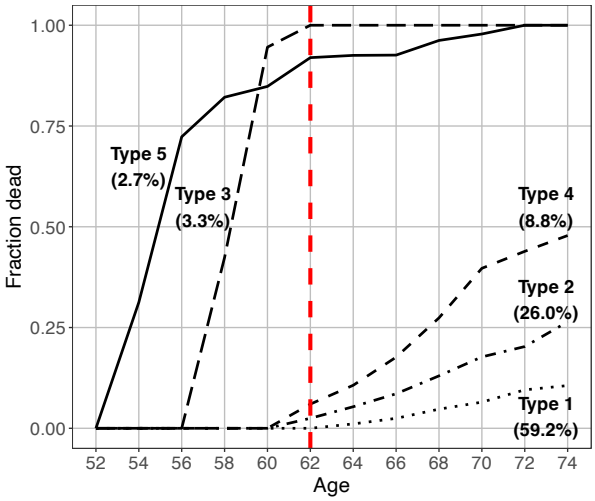
Notes: Red dashed line: end of clustering period

period are remarkably similar.

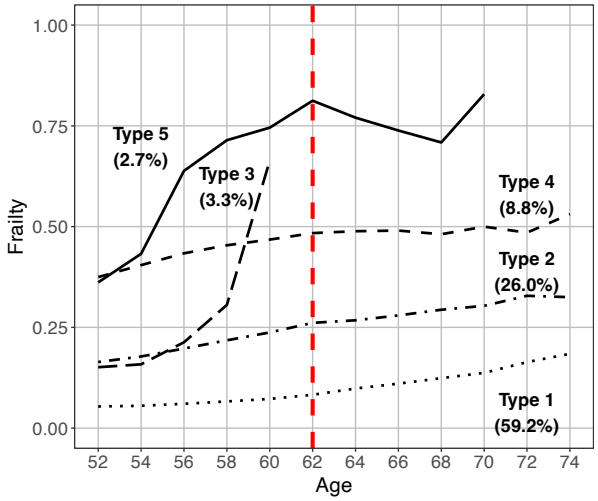
Figure 14: Health dynamics by health type and age for types constructed using 6 periods



(a) Mean frailty for everyone



(b) Fraction dead



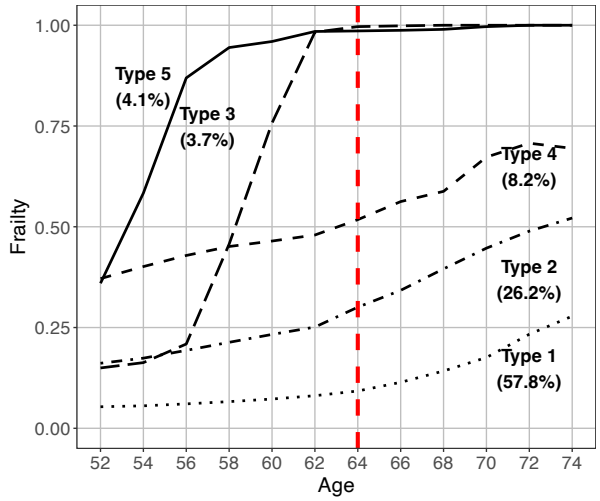
(c) Mean frailty of the survivors

Notes: Red dashed line: end of clustering period

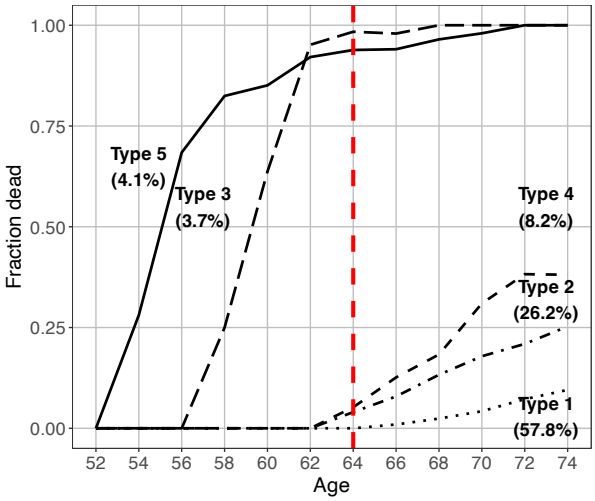
G Health types and observable characteristics

We now turn to evaluating the extent to which observable characteristics at age 52 predict health-type membership. Let η_i represent the health type membership of individual i. The variable η_i takes values from the set $\zeta = \{1, 2, 3, 4, 5\}$, where each value corresponds to one of the five health types (1 to 5). We estimate this relationship using a multinomial logit model, resulting in the following expression for each probability:

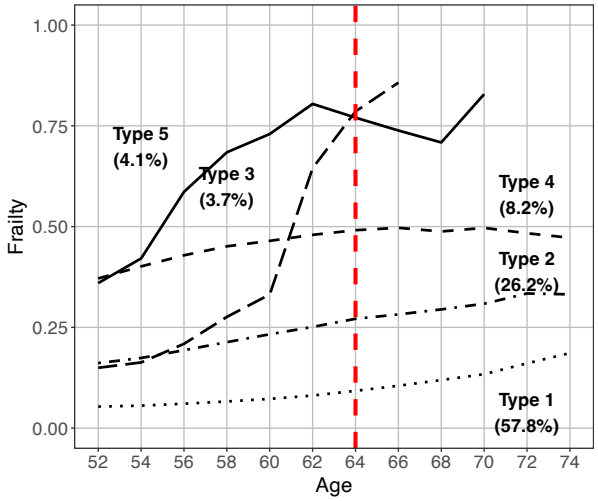
Figure 15: Health dynamics by health type and age for types constructed using 7 periods



(a) Mean frailty for everyone



(b) Fraction dead



(c) Mean frailty of the survivors

Notes: Red dashed line: end of clustering period

$$Pr(\eta_i = j \mid X_{it}) = \frac{e^{X_{it}\beta_j}}{\sum_{n=1}^{5} e^{X_{it}\beta_n}}, \tag{12}$$

where $j \in \zeta$ represents a specific health type, and the vector X_{it} includes *Demographics*: education dummies, race dummies, gender dummies, HRS cohort dummies, marital status dummies, and household total income, *Health behaviors*: Ever Smoked and vigorous activity dummies, and *Health insurance*: Private and public health insurance dummies. When

explicitly mentioned, X_{it} also includes frailty at age 52. The term β_j denotes the vector of coefficients for health type j. The coefficient for health type 1, β_1, is normalized to zero.

Column 1 of Table 14 reports the results when we allow health types to only depend on initial frailty and shows that the pseudo-R^2 from this specification is high. Column 3 of the same table shows that adding a rich set of demographics, health behaviors, and health insurance status to the previous specification only marginally improves the pseudo-R^2. This indicates that observables are not very helpful in explaining health types above and beyond initial health.

Table 14: Multinomial logistic regression of health types on observable characteristics

	Health types			
	(1)	(2)	(3)	(4)
Initial frailty	x		x	
Initial frailty composition		x		x
Demographics			x	x
Health behaviours			x	x
Health insurance			x	x
Pseudo-R^2	0.43	0.45	0.448	0.468

G.1 Frailty composition and health types

What are we missing by weighting all health deficits equally in our frailty index? Table 15 shows the contribution of each deficit category to the total average number of deficits by health type at age 52, where we split deficits into groups according to the classification that we adopt for Table 1. It shows that there is substantial heterogeneity in the relative contribution of deficits across health types. That is, the contribution of *ADLs* is low for health type 1 and increases with frailty type. *IADLs* also depict displays a similar pattern but its increasing prevalence by frailty type is less steep than for *ADLs*. *Health care utilization* and *other functional limitations* have as a relatively constant contribution by health type.

Diagnoses and *addictive* deficits showcase a decreasing relative contribution in frailty types.

Table 15: Deficits group prevalence at age 52: All sample and by health types

	All Sample		Type 1		Type 2		Type 3		Type 4		Type 5	
	Share	Level	Share	Level	Share	Level	Share	Level	Share	Level	Share	Level
ADLs	10.00	0.44	1.00	0.02	6.00	0.34	8.00	0.42	18.00	2.56	19.00	2.35
IADLs	5.00	0.23	3.00	0.05	3.00	0.20	5.00	0.27	7.00	1.03	9.00	1.09
Other functional lim	37.00	1.69	24.00	0.43	41.00	2.41	36.00	1.91	43.00	6.03	36.00	4.31
Diagnoses	25.00	1.12	29.00	0.52	27.00	1.60	27.00	1.45	19.00	2.66	21.00	2.57
Health care utilization	3.00	0.15	4.00	0.06	3.00	0.18	4.00	0.22	3.00	0.42	5.00	0.57
Addictive	20.00	0.92	40.00	0.72	20.00	1.16	19.00	1.02	9.00	1.32	10.00	1.16
Deficits at 52	100.00	4.55	100.00	1.81	100.00	5.87	100.00	5.29	100.00	14.02	100.00	12.05

Notes: The left column (Share) shows the relative contribution of each deficit group to the average total number of deficits. The right column (Level) shows the contribution of each deficit group to the average total number of deficits.

Given this substantial heterogeneity in the composition of frailty by health types, we now turn to examining to what extent one's frailty composition in the initial period affects one's probability of belonging to a given health type. To do so, we include the composition of frailty at age 52, in addition to the other observable characteristics that we have so far analyzed, in our health types classification exercise.

More precisely, we construct frailty indexes for each deficit group listed in Table 1. For example, using the eight deficits in the ADLs group (difficulty bathing, dressing, eating, getting in/out of bed, using the toilet, walking across the room, walking one block, and walking several blocks), we compute the number of reported ADLs and divide them by the total number of possible ADLs, thus obtaining an ADL frailty index. We repeat this process for the remaining five groups of deficits: IADLs, other functional limitations, diagnoses, healthcare utilization, and addictive diseases. This procedure results in six frailty indexes for each individual, thus providing information about individual deficit prevalence for each deficit group.

Comparing Columns 1 and 2, and 3 and 4 in Table 14 reveals that the composition of frailty does not add much explanatory power in terms of explaining which health type one belongs to. The high heterogeneity in frailty composition across health types and its small

effect in explaining health types' membership come from the fact that there is a large degree of co-mobility among deficits. This happens because of two important features of the data. First, all health types are very heterogeneous both in their level of frailty and in their frailty composition. Second, health types 2 and 3, and 4 and 5, are very similar in both their average frailty structure and in the average number of deficits at 52.