Endogenous Health Groups and Heterogeneous Dynamics of the Elderly*

Dante Amengual

CEMFI

<amengual@cemfi.es>

Jesús Bueren

European University Institute

<jesus.bueren@eui.eu>

Julio A. Crego

Tilburg University

<jacrego@uvt.nl>

*We thank Nezih Guner, Barton Hamilton, Pedro Mira, Josep Pijoan-Mas, and Silvia Tiezzi for their comments and suggestions. Annika Bacher provided excellent research assistance. We have also benefited from the feedback provided by seminar participants at CEMFI, as well as audiences at the VII Workshop in Time Series Econometrics (Zaragoza), the 26th European Workshop on Econometrics and Health Economics (Prague), Annual Health Econometric Workshop 2017 (Washington University in St. Louis), Universidad Carlos III de Madrid, and SED meetings 2019 (St. Louis). Financial support from the Spanish Ministry of Economy & Competitiveness through grant 2014-59262 (Amengual) and Santander Research chair at CEMFI (Crego) is gratefully acknowledged.

Abstract

We propose a methodology to classify individuals into few but meaningful health groups by estimating a panel Markov switching model that exploits rich information from panel household surveys. Using the HRS, we identify four persistent health groups, depending on individual's physical and mental disabilities. Our classification outperforms existing health measures at explaining entry in nursing homes, home health care, out-of-pocket medical expenses, and mortality for individuals in the HRS, ELSA, and SHARE. Through a workhorse model of savings, we recover an asset cost of bad health that is twice as big as when using self-reported health.

Keywords: Latent groups, Frailty, Long-Term Care, Medical expenses.

JEL: C23, C38, E21, I14.

1 Introduction

Household surveys of the elderly (HRS, SHARE, or ELSA, among others) contain a wide array of variables with different aspects of individuals' health status. Despite the richness of the data, researchers often need to rely on a single measure that summarizes most of the information about health. For example, structural economists studying the welfare costs of poor health need to incorporate health as a key state variable driving survival, savings, and insurance choices while keeping the state-space manageable. Researchers are thus constrained to undertake an ad-hoc decision over which health variable to use hampering the comparison across studies.

In this paper, we generate a parsimonious health measure by estimating the health status of individuals using a novel dynamic latent variable model that exploits panel-data information on objective health measures: twelve dummy variables on reported difficulty with Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs). The econometric model assumes that individuals belong to one out of a prespecified finite number of latent health groups. In the cross-section, individuals belonging to different health groups differ in the probability of reporting difficulties with each I-ADLs.[1] In the time series, individuals belonging to different groups differ in health transition and survival probabilities. The estimated model parameters define health groups and provide information on how to classify individuals over time in these groups.

Using the Health and Retirement Study, we find that all I-ADLs can be suitably represented by four health groups, which divide individuals into *healthy*, *physically frail*, *mentally frail*, and *impaired*. The last group of individuals have both types of limitations, physical and cognitive, while the *healthy* have no or light difficulties with I-ADLs.[2] In turn, the *physically frail* have limited mobility, while the *mentally frail* have difficulties with more cognitive tasks such as managing money. Importantly, and in line with recent medical literature (e.g. Morris et al., 2013), not all the I-ADLs are equally informative for classifying individuals in health groups. For example, if a person has difficulties with getting in or out of bed, she belongs to the *physically frail* group with a probability higher than one third but to the *mentally frail* with a probability lower than 5%. In contrast, an individual incapable of taking medications is

[1] We denote I-ADLs as the set of both ADLs and IADLs; likewise, I-ADL refers to one of these variables.

[2] Throughout the paper, we use italics to refer to our states, hence a *healthy* individual is a member of the group we label *healthy*.

much more likely to belong to the *mentally* rather than the *physically frail* group.

At the age of 60, around 80% of individuals are *healthy*, 10% are *physically frail* and the remaining 10% includes almost as many *impaired* as *mentally frail*. As individual age, we observe the same stylized facts previously documented in the literature of aging (see Manton and Soldo, 1985, among others): health deteriorates with age, individuals in worse health have larger chances of dying, and females live longer than males. Furthermore, in line with Brown (2002), Meara et al. (2008), and Pijoan-Mas and Ríos-Rull (2014) we find a large educational gradient in life expectancy.

Once health groups and transition probabilities are estimated, the algorithm allocates individuals to health groups by exploiting the time-series information on survival and I-ADLs. More precisely, through the Hamilton filter, filtered probabilities determine the chances that an individual belongs to each health group given past and current history of I-ADLs. We then use these probabilities to compare access to medical and care services across health groups. On average, *impaired* individuals spend \$12,411 per year in out-of-pocket (OOP) medical spending while *healthy* ones spend \$3,330. Likewise, *mentally frail* individuals spend \$1,194 more than *physically frail* ones, who employ \$4,524. The use of long-term care (LTC) services also largely differs across groups. While 7.3% of the *mentally frail* individuals live in a nursing home at the time of the interview, only 1.5% of the *physically frail* do. This disparity widens between members of the *healthy* group, who barely live in nursing homes at all, and those of the *impaired*, out of which 31.8% reside in these facilities. A similar pattern arises if we compare received professional care between these two extreme groups. Nonetheless, *mentally* and *physically frail* individuals need a medical-trained person to look after them at home with a similar probability.

To assess the predictive power of our classification, we contrast it with other commonly used health classifications in the literature, namely, five different levels of self-reported health, whether the individual reports difficulty with any ADL, and the division of a frailty index into five equally sized groups.[3] To do so, we consider three health-related spending variables: OOP medical expenditures as well as indicators of residing in a nursing home and receiving care, which the macro literature has identified as crucial drivers of savings (De Nardi et al. 2010;

[3] De Nardi et al. (2010), Kopecky and Koreshkova (2014), Pijoan-Mas and Ríos-Rull (2014), Dobrescu (2015), and De Nardi et al. (2016) rely on self-reported health, Bohacek et al. (2015) on ADLs, and Braun et al. (2017) on a frailty index.

Barczyk and Kredler 2018; Ameriks et al. 2020). Our classification generates more differentiated groups, which leads to higher explanatory and predictive power of these variables. For instance, conditional on age, education, and gender, self-reported health can explain 1.9 percentage points of the variance of OOP medical expenses and 7.1 of the variance of residing in a nursing home. Our measure explains 4.0 and 27.3, respectively. Besides health measures, we analyze the ability of the different classifications to predict mortality and find that our measure dominates the alternative ones.

Additionally, we explore the consequences of using our measure in a life-cycle model by solving the model proposed by De Nardi et al. (2010). Although the unconditional dissaving pattern of the elderly remains unchanged, dissaving conditional on health differs substantially. De Nardi et al. (2010) predict a similar dissaving pattern by healthy and unhealthy individuals. In contrast, our estimation indicates that unhealthy individuals are responsible for most of the dissaving during retirement. This result arises from a tighter correlation between health and medical expenses and larger differences in survival conditional on health inherent in our classification.

In order to produce the proposed health classification, estimation of the econometric model is necessary. Therefore, to facilitate future usage of our proposed classification, we publicly provide the probabilities of belonging to each estimated health group for the most commonly used retirement surveys: HRS, ELSA, and SHARE on Bueren's webpage for different numbers of health groups.

Literature– Our paper complements the literature analyzing the effect of health on economic decisions which relies on dynamic structural models to quantify the relative importance of alternative mechanisms and their implications for policymaking. As mentioned earlier, due to the curse of dimensionality, researchers undertake an ad-hoc decision over which of all the possible health variables from the available surveys to use as a state variable. For instance, De Nardi et al. (2010) and O'Donnell et al. (2015) use self-reported health which provides the best mortality forecast as argued by Idler and Benyamini (1997); hence, it is ideal to assess the risk of living longer. However, this measure does not capture long-term care needs, which constitute an important component of medical expenses; thus, Ameriks et al. (2016) and Ko (2016) rely on individuals' difficulties with mobility or cognition. Likewise, the subjective nature of self-reported health makes it unfeasible for some applications such as analyzing health and mortality insurance choices. For this reason, Koijen et al. (2016) rely on medical expenses and morbidities to construct a measure of health while Braun et al. (2017) summarize 40 variables into a continuous frailty index, which is then divided into quintiles to obtain a discrete measure. Finally, facing the lack of an optimal measure, some researchers directly use individual choices such as receiving home care or residing in a nursing home as the health state (see for instance Barczyk and Kredler (2018) or Kopecky and Koreshkova (2014)).

As the aforementioned papers, we treat health transition as exogenous to health investments although we allow it to depend on the education level. An alternative approach considers investments in health capital, which in turn decreases the likelihood of future worse health (see for example Yogo (2016) or Ozkan (2014)). This setup would force us to estimate health groups and transition within a structural model with health investments rendering the exercise much more complex.

Moreover, this paper relates to an extensive literature that proposes econometric methods to analyze different issues in health economics (see Jones, 2000, for a survey). Closely related to our paper is Deb and Trivedi (1997) who show that a finite mixture of negative binomials, characterizing "healthy" and "ill" individuals, explains counts of medical care utilization by the elderly in the U.S. better than previously proposed specifications. They, however, do not classify individuals into the aforementioned categories and disregard health dynamics. In contrast, Contoyannis et al. (2004) stress the importance of health persistence using a dynamic panel ordered probit model for self-reported health.

Additionally, we contribute to a growing literature that summarizes health variables into a single index that explains most of the variation in health-related variables (see Searle et al., 2008). Regarding the HRS, Yang and Lee (2009) compute a frailty index based on chronic conditions, ADLs, IADLs, depressing symptoms, self-reported health, and obesity. Nonetheless, its continuous nature prevents researchers to include it in structural models. One exception is Bound et al. (2010) who considers health as a continuous latent variable and include it into a structural model to analyze retirement. To be able to solve the model, they assume that individuals are completely unable to self-insure against medical expenses.

On the econometric modeling side, our paper closely relates to Todd and Zhang (2020) who also consider discrete latent states across which individuals can move over time. They use the latent states to represent types of individuals according to personality traits and unobserved educational and professional characteristics in order to understand educational and occupational choices. Given the context, our latent model needs to accommodate a more flexible specification of transition probabilities across states. In their setting, the probability of switching to a different state is independent of the state from which you transition. In contrast, in our context, this assumption would imply that the probability of recovering from extreme bad health (*impaired*) is the same as the probability of recovering from intermediate health (*physically or mentally frail*) which is not supported by the data.

Finally, our paper contributes to the recent panel data literature on group clustering such as Wang et al. (2018) who allow for heterogeneous-across but homogeneous-within groups slope coefficients. While our set-up also features unobserved individual health status classified within groups, we explicitly identify their changes across time. Similarly, Bonhomme and Manresa (2015) restrict individuals to belong to the same group forever but allow the group characteristics to change over time. In our context, this feature forces every individual in a group to have the same dynamics. Using the physically impaired as an example, their model would imply that either everyone that is physically impaired remains physically impaired, or everyone recovers, or everyone gets worse. While this characteristic makes a lot of sense in their application, it is unrealistic in our setup. With a different objective, Cunha et al. (2010) propose a structural model for skill formation. Our paper relates to theirs in the non-lineal filtering of unobserved variables, though ours are discrete in nature, and, since our model is not structural, there is no need for identifying unobserved optimal decisions.

The rest of the paper is structured as follows. We briefly describe the HRS data in Section 2. Section 3 presents the econometric model and the estimation strategy. Next, we present the main results in Section 4 and we compare our proposed classification with alternative ones in Section 5. Finally, Section 6 concludes.

2 HRS and I-ADLs

Our main dataset is the RAND HRS Longitudinal File which is a clean, easy-to-use data product containing information from Core and Exit Interviews of the Health and Retirement Study, conducted by the University of Michigan.[4] It contains subjective and objective indicators of health, as well as demographic and economic characteristics, of a representative panel

[4]Version P. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (August 2016). The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

of US households surveyed biannually from 1992 to 2014. In addition, the HRS exit interview records the death of the individual and includes answers from a proxy informant. The completeness of this data source has led to its omnipresence in the recent literature.

Since not all the variables used in the estimation are available for early waves, we restrict the sample to ten waves, from 1996 until 2014. Moreover, to focus on health needs, we select individuals over 60 years old. The final sample, after excluding individuals whose education, gender, or age are missing (<0.1% of observations), consists of 159,025 interviews (including exit waves), which corresponds to 27,369 individuals followed on average for six waves (12 years). The composition of the sample reflects changes in survival probabilities. Not surprisingly, while the median age is 72 years, the share of individuals is decreasing in age as they die. Likewise, females account for 58% of the sample as their life expectancy is higher than the males' one. In terms of education, 72% of individuals completed high school which constitutes 74% of the sample due to its superior life expectancy.

The HRS provides dozens of health-related variables, but we restrict our focus to individual's ability to perform ADLs and IADLs to infer their health status. ADLs were proposed by Katz et al. (1963) as a measure of how independent a patient is, and consequently, they include very basic activities such as if they can walk or dress. IADLs, in contrast, consist of activities more closely related to cognition as the ability to use a phone or controlling her medication. These variables relate to the need for LTC which is the dimension of health we aim to identify. Although our model could incorporate more information, reducing the set of variables eases the interpretation of the groups. Besides, by excluding other variables, we can use them to compare the performance of our classification against other alternatives.

Precisely, we employ twelve binary variables, denoted as I-ADLs, which include six ADLs and six IADLs that describe whether individuals have any difficulty performing these types of basic tasks. We extract this information from the HRS questionnaire to which respondents select one out of six possible answers: *Yes* and *Can't Do* that we label as 1, *No* to which we assign a value of 0, and *Don't Do*, *Don't Know*, and *Refuse to answer*, which are recorded as missing. Table 1 defines the activities included in the HRS and provides the proportion of observations in which an individual declares to have difficulties performing each of them. The most common ADL is not being capable of dressing (12%) whereas eating is the ADL that fewer individuals report having difficulties with (5%); at the same time, the frequency of IADLs differs across activities from 5% of respondents who claim to face problems when taking medications to 15% that struggle reading a map. Table 1 also indicates that 21% of individuals report difficulties with at least one ADL; meanwhile, 23% of them encounter problems when they carry out one or more IADLs. Altogether, almost 30% of respondents report difficulties with at least one I-ADL.

[Table 1 about here.]

The HRS also includes a question to qualify respondent's self-reported health (SRH). Since another strand of the literature hinges on subjective measures of health to classify individuals, the last five columns of Table 1 compare this measure with the answers related to ADLs and IADLs. Not surprisingly, we observe that as people report worse health, they are more likely to have problems with I-ADLs. Nonetheless, the importance of each activity differs. In particular, individuals reporting poor health are not able to walk, dress, or bathe with around 40% probability, while for the remaining three ADLs the corresponding figures barely surpass 30%. Similarly, difficulties with IADLs are diverse within the worst self-reported health groups: 50% of individuals report difficulties to shop for groceries whereas only 20% encounter complications to take their medications.

3 Econometric model

We have an unbalanced panel of individuals $i = 1, \dots, N$ followed for $t_i = 0, \dots, T_i$ periods which correspond to ages $a_0^i, \dots, a_{T_i}^i$. For each individual, we observe K dummy variables corresponding to each I-ADL across time $(x_{1,i,t}, x_{2,i,t}, \dots, x_{K,i,t})$, provided the individual is alive and interviewed. All or some of the variables for a given individual who is alive can also be missing for some period t_i. We take missing observations into account under the assumption that they occur completely at random, but we abstract from them in the model description to simplify the exposition.

We assume that the main source of heterogeneity in the population is represented by a finite number of possible health groups or clusters which are not observed by the researcher. Individuals belong to each cluster with a probability that depends on education, e; age, a; gender, s; and the current health cluster but it is independent of the individual's previous health (Markov first-order property). Besides transiting across health groups, individuals may also die, which we represent as an observable and absorbing state, D.

Specifically, we consider that individual i at time t belongs to a health group $h_{i,t}$ out of H possible groups. Given her group, g, the probability of facing difficulties with the k'th I-ADL, say $x_{i,k,t} = 1$, is $\mu_{k,g}$. Under the assumption that I-ADLs are independently distributed conditional on the health status, the joint distribution of $\mathbf{x}_{i,t} = (x_{1,i,t}, x_{2,i,t}, \ldots, x_{K,i,t})'$ is then characterized by

$$p(\mathbf{x}_{i,t}|\mu_g, h_{i,t} = g) = \prod_{k=1}^{K} \mu_{k,g}^{x_{k,i,t}} (1 - \mu_{k,g})^{1-x_{k,i,t}}, \tag{1}$$

where $\mu_g = (\mu_{1,g}, \mu_{2,g}, \ldots, \mu_{K,g})'$. Therefore, individuals within the same health group have the same probabilities of experiencing problems with an I-ADL whereas these probabilities might vary if individuals do not belong to the same group. Similarly, the same individual might face a different likelihood regarding I-ADLs if she changes groups during her life.

In favor of parsimony, we model health outcomes as independent across time and individuals *conditional* on the health group. In the case of I-ADLs, it seems plausible that their persistent component is only due to health, nonetheless, the model can accommodate other types of persistence if the researcher wants to extend the set of conditioning variables. To do so, we take into account health dynamics by explicitly modeling the transition probabilities across groups. In particular, an individual i at time t who belongs to group g transits to group c with probability

$$Pr[h_{i,t+1} = c|a_{it}, s_i, e_i, h_{i,t} = g] = \frac{\exp[f_{g,c}(a_{it}, s_i, e_i)]}{1 + \sum_{c \in \mathcal{H}} \exp[f_{g,c}(a_{it}, s_i, e_i)]} \tag{2}$$

where $\mathcal{H}$ is the set that contains the H health groups. The remaining possible event is that the individual dies, which is an observable state that occurs with probability

$$Pr[h_{i,t+1} = D|a_{it}, s_i, e_i, h_{i,t} = g] = \frac{1}{1 + \sum_{c \in \mathcal{H}} \exp[f_{g,c}(a_{it}, s_i, e_i)]}.$$

This specification allows health groups to own distinct dynamics as parameters differ according to the current health group. Moreover, to capture within-group heterogeneity, transition probabilities can depend on age, gender and education level through the function $f_{g,c}(a, s, e)$ whose parametric specification is given by

$$f_{g,c}(a, s, e) = \beta_{1,g,c} + \beta_{2,g,c}a + \beta_{3,g,c}s + \beta_{4,g,c}e + \beta_{5,g,c}(a \times s) + \beta_{6,g,c}(a \times e).$$

3.1 Posterior simulation

We aim to recover the posterior of all the parameters and the latent variables that classify the health group to which each individual belongs at each point in time. We first need to set the number of health groups H. Then, we use a Gibbs sampling procedure to estimate the econometric model. In essence, this amounts to reducing a complex problem, that is, sampling from the joint posterior distribution of both parameters and state variables, into a sequence of tractable ones, i.e., sampling from conditional distributions for a subset of the parameters or latent variables conditional on all the others, for which the literature already provides a solution.

We define $\mathbf{H} = \{\mathbf{h}_i\}_{i=1}^{N}$, where $\mathbf{h}_i = \{h_{i,t}\}_{t=1}^{T_i}$, as the collection of health groups after the first time period, and $\mathbf{H}_0 = \{h_{i,0}\}_{i=1}^{N}$ the collection of health states in the first time period. Further, we denote the vectors stacking the parameters of the I-ADLs process and the transition probabilities as μ and β. In addition, we include in $\mathbf{X}$ the data we observe; that is, age, gender, education, if the individual is death or alive, and her situation in terms of ADLs and IADLs. The Metropolis-within-Gibbs algorithm involves sampling sequentially from several blocks. Specifically, iteration m involves:

1. $p(h_{i,0}^{(m)}|\beta^{(m-1)}, \mu^{(m-1)}, \mathbf{X})$: sampling the initial condition using Hamilton (1989)'s smoother.

2. $p(\mathbf{h}_i^{(m)}|\beta^{(m-1)}, \mu^{(m-1)}, \mathbf{X}, \mathbf{H}_0^{(m-1)})$: sampling the latent health indicator for each $i = 1, ..., N$ and $t > 0$ using the Kim (1994)'s smoother.

3. $p(\beta^{(m)}|\mu^{(m-1)}, \mathbf{H}^{(m)}, \mathbf{H}_0^{(m)}, \mathbf{X})$: sampling the transition parameters (Metropolis).

4. $p(\mu^{(m)}|\beta^{(m)}, \mathbf{H}^{(m)}, \mathbf{H}_0^{(m)}, \mathbf{X})$: sampling the Bernoulli mixture parameters (Metropolis).

The empirical results shown in the next sections are based on 40,000 draws. The first 2,000,000 draws are disregarded as burn-in and, of the remaining 4,000,000, one in every 100 draws is retained.

3.1.1 Sampling the states: Kim's Smoother

We obtain $p(h_{i,0}|\beta, \mu, \mathbf{X})$ for each individual from Hamilton's smoother. Then, to sample the health states, we apply the methodology developed by Kim (1994):

1. Using $p(h_{i,0}|\beta, \mu, \mathbf{X})$ and the filter proposed in Hamilton (1989) we obtain $p(h_{i,T} = g|\beta, \mu, \mathbf{X})$ for all $g \in \mathcal{H}$.

2. We sample $h_{i,T}$ from $p(h_{i,T}|\beta, \mu, \mathbf{X})$.

3. Similarly, we sample $h_{i,t}$ conditional on β, μ, $\mathbf{X}$ and $h_{i,t+1}$, using the following result: $p(h_{i,t} = g|\beta, \mu, \mathbf{X}, h_{i,t+1} = c) = \dfrac{p(h_{i,t+1} = c|\beta, h_{i,t} = g) \cdot p(\mathbf{x}_{i,t}|\mu, h_{i,t} = g)}{\sum_{g \in \mathcal{H}} p(h_{i,t+1} = c|\beta, h_{i,t} = g) \cdot p(\mathbf{x}_{i,t}|\mu, h_{i,t} = g)}$ for all $g, c \in \mathcal{H}$

As a result, in the econometric model, each individual has a different probability of belonging to a given group depending on her past, current, and future answers regarding I-ADLs. Moreover, this probability incorporates information about the individuals' death wave, as well as her age, gender, and education. As such, this probability contains information available to the econometrician but surely not available to the individual.

3.1.2 Sampling the transition probabilities and the Bernoulli parameters

In this step, we sample from the posterior of the parameters of the Bernoulli distributions and the ones governing the health dynamics (μ, β) conditional on the health groups, $\mathbf{H}$, and the data, $\mathbf{X}$.

Regarding priors, we consider a uniform distribution on $[0, 1]$ for the elements of μ and a diffuse Gaussian prior centered at $\mathbf{0}$ and covariance matrix $100 \cdot \mathbf{I}$ for β, where $\mathbf{I}$ denotes the identity matrix. Hence, the posterior of the parameters governing the health dynamics and the one driving the Bernoulli distributions are independent conditional on the latent health group. Precisely, their posterior distributions are given by

$$p(\mu|\mathbf{X}, \mathbf{H}) = \prod_{i=1}^{N} \prod_{t=1}^{T_i} p(\mathbf{x}_{i,t}|h_{i,t}, \mu) \cdot p(\mu)$$

and

$$p(\beta|\mathbf{X}, \mathbf{H}) = \prod_{i=1}^{N} \prod_{t=2}^{T_i} p(h_{i,t}|\beta, h_{i,t-1}) \cdot p(h_{i,1}|\beta) \cdot p(\beta).$$

3.1.3 Starting the algorithm

To obtain the starting parameter values μ^0 and β^0 for the algorithm, we sample from an approximate model in two steps. First, we obtain μ^0 as the mode of the posterior described in equation (1) under the assumption that $h_{i,t}$ are independent across both dimensions.[5] Second, we use the same model to simulate $h_{i,t}$ from the posterior probability $p(h_{i,t}|\mu, \mathbf{x}_{i,t})$. Given a sample of health groups, we get the mode of the posterior of β, β^0, under the assumption that groups follow the same multinomial logit specification as in the baseline model.

[5]This model is also known as latent class analysis (Lazarsfeld, 1950; McLachlan and Peel, 2004).

3.1.4 Selecting the number of groups

Two main aspects play a role when selecting the number of groups. On the one hand, more groups might lead to a better model fit. Indeed, Bayes's odd ratios indicate that a higher number of groups would be the best choice. On the other hand, each additional group creates an extra burden for structural models, which rarely use more than four. Therefore, the optimal number of groups depends on the application. We focus on four groups as they can be implemented in many models, and provide a significant improvement against two and three groups. Nonetheless, we also provide results for the two-groups case for more complex structural models. Considering more than four groups would improve, mildly, some results; hence, the paper provides a conservative lower bound of our model performance.[6]

4 Results

In this section we first describe the estimated health groups, then we explain how health evolves with age taking into account differences in education and gender. Finally, we show how we can use our econometric model to produce a new health classification.

In what follows, we report the median of the posterior distribution of the parameters -or relevant functions of them.

4.1 Health Groups

Figure 1 displays the probability of reporting difficulties with each I-ADL conditional on being in each cluster, that is $\mu_{k,g}$ in equation (1). Each panel corresponds to a different number of clusters H. Meanwhile, each marker symbol represents a cluster and each tick in the horizontal axis refers to an ADL (the first six) or an IADL (the remaining ones). The higher the marker, the more likely the individual in that specific group struggles with the corresponding I-ADL.

[Figure 1 about here.]

[6] Nonetheless, we acknowledge that restricting to four groups prevents the posterior distribution to achieve consistency à la Barron et al. (1999).

When we set $H = 2$, the algorithm divides individuals into one group whose probability of declaring problems with an I-ADL is close to 0 for every I-ADL and another one which has a relatively high likelihood of facing problems with every I-ADL. We label the former group as *healthy* (circumferences) and the latter as *impaired* (triangles).[7] We find large differences in the probabilities of reporting difficulties across I-ADLs within the *impaired* group which suggests that activities differ in their importance for categorizing individuals. For example, in the *impaired* group, individuals have a 31% chance of reporting difficulties eating whereas their likelihood of reporting difficulties with shopping is 77%.

The upper right panel of Figure 1 presents the same graph but with $H = 3$. There is still one group with almost zero probability to face difficulties with any I-ADL and another one that faces a relatively high probability of struggling with each I-ADL. Nevertheless, the probabilities of this group are slightly higher than when we consider only two groups as some individuals previously classified as *impaired* belong to the new group whose probabilities of reporting difficulties with I-ADLs lie between the other two.

When we allow for four groups, the *impaired* and the *healthy* groups become more distant. In addition, the middle group splits into two very different ones. One group with moderate probabilities to suffer difficulties with an ADL but low probabilities to have problems with IADLs, reflecting that those individuals are *physically frail*; and another one which consists of *mentally frail* elderly in the sense that they are mostly dependent in terms of IADLs but not as much in terms of ADLs.

Lastly, we consider $H = 5$ in the lower right panel. In that case, the previous groups remain almost unchanged and the new group that emerges is extremely similar to the *healthy* one, with the exception that individuals struggle reading a map. As one adds more groups, their connection to health is even weaker; therefore, in the remainder of the paper, we focus on the case of four groups.

While Figure 1 characterizes individuals' health in each cluster, it is silent about the meaningfulness of each I-ADLs for classifying individuals. For instance, in the case of $H = 2$, the elderly in the *impaired* group have a much higher probability of facing difficulties reading a map than eating. This comparison, however, disregards that unconditionally only 5% of individuals struggle to eat but 16% are not able to read a map.

[7]As in most latent variable models, the labeling of the groups is somewhat arbitrary. Nonetheless, it eases the presentation without altering our main findings.

To overcome this issue, Figure 2 plots the probability of belonging to group g given that the individual faces difficulties with I-ADL k, that is,

$$\Pr(h = g|x_k = 1) = \Pr(x_k = 1|h = g)\frac{\Pr(h = g)}{\Pr(x_k = 1)};$$

where the relative size of the bars indicates which I-ADL is more informative.

[Figure 2 about here.]

Following the same example, if a person has difficulties to eat, she belongs to the *impaired* group with 90% probability, according to the upper left panel. Meanwhile, individuals incapable of reading a map have almost the same likelihood to be part of the *impaired* or *healthy* group, suggesting that MAP is uninformative. The pattern of these two I-ADLs remains unchanged when $H = 3$ and $H = 4$; MAP is again uninformative while EAT is the best indicator to classify individuals into the *impaired* group. This finding is in line with previous evidence in the medical literature (see Morris et al., 2013, and references therein) which argues that difficulties with eating best predict full dependence.

While Figure 2 characterizes the importance of each I-ADL separately for descriptive purposes, their joint structure significantly contributes to identification as well. To see this, in the third and fourth columns in Table 2 we provide the proportion of respondents who report difficulties with at least one ADL or IADL. Consistent with the previous discussion, individuals in the *impaired* group are the ones more likely to present difficulties with an I-ADL; in fact, they face problems with at least one I-ADL with almost certainty. In contrast, a healthy individual's probability of reporting troubles with ADLs varies between 1% and 9% depending on the number of groups. In the third panel (four groups), the distinction between *physically frail* and the *mentally frail* becomes salient. While in the former 87% of respondents struggle with ADLs, the latter faces more problems with IADLs (99.4%) and less with ADLs (51.3%).

[Table 2 about here.]

Groups are not only different in terms of I-ADLs but also in terms of demographics. For instance, if our classification correctly identifies the health status of individuals we expect members of the *impaired* group to be older than those of the other groups. In that regard, Table 2 shows impaired individuals are indeed on average nine years older than the ones in the *healthy* cluster and six years older than those *physically frail*. Additionally, the difference between *mentally frail* and *impaired* is smaller which is consistent with mental conditions caused by aging. In terms of education, high school graduates are overrepresented in the *healthy* group in line with previous literature on health inequality such as Mackenbach et al. (2008). Another interesting pattern is that worse health groups contain a significantly higher proportion of women. These differences lead us to study heterogeneity in health dynamics across gender and education groups.

4.2 Heterogeneous health dynamics

Indeed, the distribution of elderly across health groups changes with age, gender, and education as can be seen in Figure 3, which plots the probability of belonging to each group through age. The left panels correspond to high-school dropouts whereas the right ones present the results for high-school graduates; meanwhile, the upper graphs refer to males and the lower ones to females. The most common health status at early ages is *healthy* but starting around the age of 90, *impaired* becomes the predominant group. Further, the *physically* and *mentally frail* have very different dynamics: while the former is stable throughout life, the latter increases steeply as individuals age. Although these patterns are very similar across education and gender, the initial composition of individuals varies with demographic characteristics.

[Figure 3 about here.]

We should notice that different transitions translate into different risks, which we summarize as expected time an individual at age 60 lives in each health group in Table 3. Even if more educated individuals live longer, they spend fewer years as *impaired* and *frail*, which suggests that richer individuals face lower health risks. For instance, in the case of males, dropouts live 80% more time (or 0.4 extra years) in the *impaired* state. Additionally, the difference between males and females indicates that females expect a longer life but during those extra years, they do not expect to be healthy.

[Table 3 about here.]

4.3 Health Classification

The estimated econometric model exploits past, present, and future information on I-ADLs and survival to construct a parsimonious model for health groups and health transitions. Therefore, the econometrician has information that does not belong to the individual's information set. Only in the case in which individuals had perfect foresight the information set of both coincides. To avoid making the perfect foresight assumption, from our estimates, we construct a health classification that assumes that individuals are aware of the underlying health process (transitions and probabilities of belonging to health groups for each I-ADL) but only know their past and present information on survival and I-ADLs.

To do so, we use the mean of the posterior of μ and β ($\bar{\mu}, \bar{\beta}$) to recursively update each individual's probability of belonging to each health group. At each age t, we obtain $p(h_{i,t}|X_i^{t-1}, \bar{\mu}, \bar{\beta})$ which represents the probability of belonging to each health group using I-ADLs information up to time $t-1$. Then, we update this probability with the I-ADL information at time t to obtain the probability of each state conditional on present and past information, $p(h_{i,t}|X_i^t; \bar{\mu}, \bar{\beta})$:

$$p(h_{i,t}^{(g)}|X_i^t, \bar{\mu}, \bar{\beta}) \propto p(\mathbf{x}_{i,t}|X_i^{t-1}, h_{i,t}^{(g)}, \bar{\mu}, \bar{\beta}) \cdot p(h_{i,t}^{(g)}|X_i^{t-1}, \bar{\mu}, \bar{\beta}) = p(X_{i,t}|h_{i,t}^{(g)}, \bar{\mu}) \cdot p(h_{i,t}^{(g)}|X^{t-1}, \bar{\mu}, \bar{\beta}),$$

where $p(X_t|h_{i,t}^{(g)}, \bar{\mu})$ is the likelihood of current I-ADLs conditional on being in state g given by (1). Then, we obtain forecasts of health groups given information up to t recursively, $p(h_{i,t+1}^{(g)}|X^t, \bar{\mu}, \bar{\beta})$, starting with $p(h_{i,0}|\bar{\mu}, \bar{\beta})$, and using the transition probabilities $p(h_{t+1}|h_t, \bar{\beta})$:

$$p(h_{i,t+1}^{(g)}|X_i^t, \bar{\mu}, \bar{\beta}) = \sum_{h_{i,t}^{(c)} \in \mathcal{H}} p(h_{t+1}^{(g)}|h_t^{(c)}, \bar{\beta}) \cdot p(h_{i,t}^{(c)}|X_i^t, \bar{\mu}, \bar{\beta}) \qquad t \geq 0,$$

To obtain $p(h_{i,0}|\bar{\mu}, \bar{\beta})$, we first assume that, at the age of 60, individuals' probability of belonging to each health group equals the average probability at that age for their gender and education level: $p(h_{i,60-a_0^i}|\bar{\mu}, \bar{\beta}) \equiv \frac{1}{N_{60}} \sum_{i=1}^{N_{60}} p(h_{i,60-a_0^i}|\mathbf{X}, a = 60, \bar{\mu}, \bar{\beta})$, where the average is taken across the N_{60} individuals that we observe at age 60. For those that we observe for the first time when they are older than 60, we update this probability until $t = 0$ using the transitions:

$$p(h_{i,t+1}^{(g)}|\bar{\mu}, \bar{\beta}) = \sum_{h_{i,t}^{(c)} \in \mathcal{H}} p(h_{i,t+1}^{(g)}|h_{i,t}^{(c)}, \bar{\beta}) \cdot p(h_{i,t}^{(c)}|\bar{\mu}, \bar{\beta}) \qquad 60 - a_0^i \leq t < 0$$

Intuitively, to ensure that we do not use information outside the information set of each individual, we assume they are homogeneous at the age of 60 and transition across states as the average individual until we observe their health data for the first time.

Upon updating until the last period, we obtain $p(h_{i,t}|X_i^t, \bar{\mu}, \bar{\beta})$ for each individual and time period. Therefore, the weighting of that observation when we estimate moments conditional on being in group g corresponds to the probability for health group g. However, in our context, the average probability of the most likely group is 0.98 ($H = 2$), 0.95 ($H = 3$), and 0.94 ($H = 4$); hence, these weights play a relatively minor role.

5 Comparison with alternative indices

We next compare our proposed classification with the main three alternatives used in the literature: self-reported health, if the individual struggles with one ADL, and the quintiles of a frailty index. In addition, we propose an alternative classification: the Cartesian product of whether the individuals report difficulty with i) at least one ADL and ii) IADLs (excluding MAP) as an unsophisticated proxy of our classification. This classification does not require any estimation and can be implemented with many health surveys (e.g. SHARE or ELSA); nonetheless, its similarity to our measure suggests that it might perform better than current classifications.

To perform the comparison, we focus on mortality and three variables related to the financial risk due to health: OOP medical expenditures as well as indicators of receiving home care and residing in a nursing home. As mentioned earlier, OOP medical spending is a direct measure of the economic consequences of health. It includes the costs -in constant 2000 US dollars- of hospital and nursing home stays, doctor visits, dental treatments, outpatient surgery, prescription drugs, home health care, and special facilities. Received home care equals 1 if a medically-trained person has come to the respondent's home to help her, and nursing home resident takes value 1 for those individuals who live in a nursing home at the time of the interview.

[Table 4 about here.]

The health classification most widely used in the literature relies on individuals' self-assessment of their health status which can take 5 different values between excellent and poor. The self-reporting nature of the answer induces two opposing effects. On the one hand, individuals might know features of their health unknown to everyone else. On the other hand, respondents might misjudge their health condition, bias their answers according to their mood, or consider different benchmarks of being good. As documented by previous literature, the net effect of these two channels and establishes that the disadvantages often offset any benefit (see Currie and Madrian, 1999, for a survey). Nonetheless, the first panel of Table 4 confirms that self-reported health has information predictive for health-related financial risk. Respondents reporting worse health spend more on medical consumption and care, and are more likely to reside in a nursing home than those who claim to be healthy. However, the difference between the five groups varies. In particular, answering excellent, very good, and good relates to almost the same risk, whilst fair and poor correspond to significantly more spending. Thus, it is common practice merging the three healthiest and the two worst groups, which we denote as *self-reported health (2 groups)*.

Besides, grouping individuals according to whether they report difficulty with an ADL or not is similar to our approach, specifically to identify the *healthy* respondents; hence the proportions of *healthy* and No-ADL almost coincide. This classification, however, considers every ADL equally important and disregards the number of ADLs, as well as difficulties with IADLs.

Recently, Braun et al. (2017) construct a frailty index based on Searle et al. (2008) by merging information on I-ADLs, chronic conditions, cognitive impairment, and information about smoking and alcohol consumption to create a frailty index. Although the inclusion of more information improves the measure of health and allows to create more groups, the relevance of each variable is still assumed to be the same. Additionally, the resulting index is continuous which forces to discretize it, and they do so by allocating individuals into five equally sized groups according to the quintiles of the index. As a result, the healthiest groups are very similar among themselves and the worst group presents the same features as those who have an ADL in the Yes/No classification.

Finally, we propose a simple approximation to our four groups which we denote as 4-I-ADL: classifying individuals regarding whether they struggle with at least one ADL but no IADL, at least one IADL but no ADL, difficulties with at least one ADL and one IADL or no reported difficulties. In contrast to the frailty index by Braun et al. (2017) who effectively separates individuals without problems with any ADL into four groups, this method divides respondents who recognize problems to perform at least one ADL into three groups.

Although the four aforementioned alternative classifications are highly correlated with the health outcomes that we use, our estimated groups seem to be more differentiated as can be seen in Table 4. For instance, using our methodology, the average difference in terms of medical spending between *healthy* and *impaired* elderly is \$9,081. According to self-reported health, however, an individual belonging to the worst group spends on average \$3,333 more than an individual in the healthiest group. Similarly, the fact that you report an ADL implies that your average OOP medical spending is \$2,648 higher; meanwhile, belonging to the worst, rather than the best, frailty quintile costs \$3,305. Not surprisingly, 4-I-ADL is the closest to our classification but the distance between the best and worst groups is just a little above \$4,000. Similarly, the intermediate groups are less distinct in terms of medical spending when using alternative classifications. For example, the increment in spending is always below \$1,000 which corresponds to the minimum observed difference between our groups.

Regarding the probability of residing in a nursing home, a similar pattern arises. The difference in the probability of residing in a nursing home between the best and worst of our health groups is at least twice as large as when using the alternative measures. The same holds true for home care when we look at self-reported health or struggling with at least one ADL although, in this case, our four groups outranks 4-I-ADL just mildly.

In line with the previous discussion, our classification also identifies future death events more accurately. In particular, an *impaired* individual dies with 33% probability within the next two years whereas only 3.6% of *healthy* ones do not survive until the next wave. In contrast, the difference between the healthiest and unhealthiest groups does not exceed 25 percentage points with any alternative classifications.

5.1 A horse race

Most of the time, the researcher's objective is not be to classify individuals into distant groups but to create a categorical index that captures most of the variation coming from health conditional on age, gender, and education. To assess the performance of the alternative grouping methods in that context, Table 5 displays the R^2 of the following regression:

$$y_{i,t} = c + \mathbf{d}'_{i,t}\beta + \mathbf{z}'_{i,t}\gamma + (\mathbf{d} \otimes \mathbf{z})'\,\theta + age_{i,t}\left(\mathbf{d}'_{i,t}\beta_1 + \mathbf{z}'_{i,t}\gamma_1\right) + \varepsilon_{i,t}$$

where $y_{i,t}$ is the variable used as a reference, $\mathbf{z}_{i,t}$ includes gender and education, and $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs. As a comparison, we also add the R^2 of the regression including, in the $\mathbf{d}_{i,t}$ vector, dummies for each I-ADL.

[Table 5 about here.]

Even though self-reported health only explains 1.9% of the variation of out-of-pocket medical spending, it doubles the variance explained solely by age, education, gender, and their interactions. Similarly, we can explain up to 2.2% by dividing individuals according to whether they report problems with at least one ADL. Including IADLs improves the fit by one percentage point. Altogether, our classification explains 3.8% of the medical spending variance when using the modal health group and 4.0% when using the set of probabilities of belonging to each health group, which exceeds every alternative. The same conclusion arises from considering the OOP spending reported in the following wave. Interestingly, our classification explains just 0.2% less than including all the full set of dummies for ADLs and IADLs in the regression.

The same ranking persists when we consider nursing home residency. Any measure that includes ADLs beats self-reported health by at least five percentage points, which doubles if we consider the 4-I-ADL classification. Further, weighting each I-ADL, our health groups enhance the 4-I-ADL classification by 62% because it identifies the extreme dependent individuals better. Likewise, the 4-I-ADL classification performs poorly compared to the one that includes each I-ADL separately because of the importance of each I-ADL to predict nursing home residency. Our proposed classification explains almost four times more variance than self-reported health and twice as much as*ADL:Yes/No*.

In contrast to nursing home residents, most elderly who need home care preserve a high degree of independence. As a consequence, the weighting of I-ADLs loses importance, and our measure, although it still explains most of the variation in the outcome variable hardly improves a classification based on ADLs. Nonetheless, it explains 50% more variance than self-reported health and reaches the performance of including all the I-ADLs.

Finally, regarding mortality, we have constructed a division that performs better than self-reported health. This contribution is relevant because most of the literature (see Idler and Benyamini, 1997, for a survey) shows that subjective measures of health usually predict mortality beyond objective indicators. Notably, the R^2 using 4-I-ADL is 0.7 percentage points larger than that of self-reported health which indicates that part of the improvement on the mortality prediction relies on the incorporation of I-ADLs, while the remaining improvement is due to the dynamics and the weighting of each I-ADL.

The Case of Europe– In order to assess how our proposed classification would perform in different regions, we re-estimate our econometric model using data from the English Longitudinal Study of Aging (ELSA) for the U.K. and from the Survey of Health, Ageing and Retirement in Europe (SHARE) for 18 European countries. Given the small sample size for some countries in SHARE, we pool countries into five different regions: Continental Europe (Austria, Germany, Netherlands, Switzerland, Belgium, and Luxembourg); Nordic countries (Sweden and Denmark); Mediterranean (Spain, Italy, France, Israel, and also Portugal); Eastern Europe (Czech Republic, Poland, Hungary, Slovenia, and Estonia). In order to ease the comparison across regions, we fix the probability of reporting I-ADLs by health group to the ones estimated for the U.S. Nonetheless, we allow variation in health transitions to capture heterogeneous dynamics across regions. In particular, we allow the intercept of function $f_{g,c}(a,s,e)$ in equation (2) to vary across regions r following:

$$f_{g,c}(a,s,e,r) = \beta_{1,g,c,r} + \beta_{2,g,c}a + \beta_{3,g,c}s + \beta_{4,g,c}e + \beta_{5,g,c}(a \times s) + \beta_{6,g,c}(a \times e).$$

Table 6 mimics Table 5 using European data. We observe that even though the health groups are defined using the estimates from the U.S. data, the proposed classification outperforms other health measures used in the literature. Indeed, the relative differences between measures remain stable despite the difference in explanatory power of each measure, suggesting that fixing the parameters does not have extreme consequences. There are two caveats when comparing Table 5 and Table 6. First, by design, attrition in ELSA and SHARE is non-random, and those individuals who move to a nursing home are more likely to exit the sample (Banks et al., 2011). This feature explains the low explanatory power of any health classification for nursing home residency. Second, the time spell between two consecutive interviews in share varies between two and six years. Hence, with the exception of mortality for which we know the date, any attempt to compute the explanatory power in the next wave will be affected by the heterogeneity in time between consecutive surveys across countries.

[Table 6 about here.]

5.2 Dynamics: self-reported health versus endogenous classification

Comparing groups' dynamics generates additional insights about the differences between grouping methods. To obtain smooth dynamics, we assume that the transition probabilities of self-reported health follow a logistic specification as described by equation (2). Furthermore, to ease the comparison we focus on the best and worse group of each method, that is, we compare *healthy* according to our method with excellent as reported by individuals and *impaired* with poor. For completeness, we also include the two-group classification based on self-reported health in the comparison.

The upper plots in Figure 4 report the median probability of dying in the HRS for self-reported health and our estimated health groups. The left one corresponds to the healthiest groups in both classifications, whereas the right one presents the results for the most unhealthy ones. Up to the age of 80, individuals who report excellent health, as well as those classified as *healthy* face a very small probability of dying. After this age, the elderly with a low survival probability still assess their health as excellent. On the other hand, variation in age is not as important for the *healthy* group as mortality less than doubles between age 80 and 98. One possible explanation is that individuals compare themselves with relatives and friends of the same age to assess their health status; thus, respondents of age 65 and 90 have a different benchmark. Furthermore, while the difference between mortality rates of *healthy* and *impaired* are sizable, this is not the case for the groups based on self-reported health, which suggests that this method does not predict mortality at older ages. Additionally, *impaired* individuals have a higher death probability than those who assess themselves in poor health regardless of their age.

[Figure 4 about here.]

The second relevant element of health risk is persistence. If the process is not persistent, health today would contain relatively little information on tomorrow's health and survival probabilities thus affecting individuals' saving behavior. In addition, the persistence of each classification sheds some light on the type of health process. In particular, we create an indicator of LTC needs which is by definition persistent in contrast to other health problems such as the flu or a sprained ankle. The lower plots in Figure 4 depict the probability of remaining in the same group conditional on the current health group at any given age. Specifically, we find that individuals who report *excellent* health in one wave have less than 40% probability of providing the same answer in the following wave. In contrast, respondents classified as *healthy* are extremely likely to remain in that state, which indicates that some non-persistent factors might drive self-reported health. If instead, we focus on individuals in bad health, our classification displays a larger persistence as individuals age, in line with the idea that recovery becomes harder for older individuals. In contrast, for fair and/or poor self-reported health, individuals are more likely to report improvements in their health status as they age which points towards changes in their health benchmark.

Lower persistence and a worse ability to predict mortality indicate that self-reported health overestimates the uncertainty faced by individuals. The effect of this bias on individuals' decisions depends on its severity across socio-economic groups and the specific structural model. To shed some light on the former, Figure 5 plots the additional time that a high-school graduate spends in the healthiest state (left-hand panel) and the unhealthiest state (right-hand side) in expectation. While our classification indicates that high school graduates spend around 40% more time in the *healthy* state and 30% less in the *impaired* state, these differences at least double when using self-reported measures of health. Given that our classification explains a larger fraction of the variance of different health outcomes, these results suggest that self-reported health contains a measurement error correlated with education: Low-educated individuals tend to report worse health status or high-school graduates overestimate their well-being or both.

[Figure 5 about here.]

5.3 The asset cost of bad health

In order to show how to implement our health classification in a structural model and its implications, we replicate De Nardi et al. (2010) to quantify the asset cost of being in bad health across classifications. For this purpose, we solve and simulate the baseline model of De Nardi et al. (2010) using two levels of self-reported health (as in the original paper) and three different versions of our proposed classification. First, we consider our two-group classification to compare the effects of our method without increasing the state space. Second, we discretize the probability of being *impaired* when $H = 2$ into four equally spaced points. This measure assesses the empirical relevance of probabilities and sets up a benchmark for a four-group classification. Finally, we implement our four-group classification.

Besides the health classifications, we follow the original paper as closely as possible to ease the comparison. De Nardi et al. (2010) estimate the model on a sample of single retirees using a two-step procedure: in the first step, the authors estimate the processes for survival, health transitions, and medical expenses risk as functions of health; hence we have to re-estimate them for every classification. In the second step, they estimate the preference parameters and the generosity of Medicaid using the method of simulated moments. To ease the comparison, and due to the good aggregate fit, we maintain the same parameters as in the original model. Therefore, differences in policy functions come from differences in the health classification and not in the preference parameters. [8]

To analyze how health affects dissaving, we focus on rich female individuals because they face a larger risk of outliving their assets and we simulate 500,000 female individuals in the top quintile of the income distribution who are endowed with $175,000 (the median assets of the top income quintile in the HRS for individuals between 70 and 75). We fix half of the simulated sample to the good health state forever while the other half stays always in bad health ruling out attrition in order to ease the interpretation of results.

[Figure 6 about here.]

The upper left panel in Figure 6 shows the median asset holdings for individuals in each of the two subsamples when using self-reported health as in De Nardi et al. (2010). We can see that after fifteen years in bad health a representative individual holds around 15% less wealth than an individual who remains healthy. Under any of our proposed classifications, the difference in the dissaving profile across health groups is much more salient. When using our classification with two and four health groups, the representative individual who remains impaired for fifteen years holds 30% and 64% less wealth, respectively. Differences between using probabilities or the modal health state are relatively small because the average individual has a 94% probability of belonging to her modal group.

6 Conclusion

As retirees age, they face large risks of requiring persistent and expensive care. The structural literature analyzing savings in retirement underlines the importance of this uncertainty to explain the dissaving pattern of the elderly and labor supply decisions of individuals. They

[8]The Online Appendix gathers the parameters of the first-step estimation and results from the original paper to assess the accuracy of the replication.

face, however, an important empirical challenge: summarizing the information content of several health variables into a few groups, which is a requirement for quantitative models to be computationally feasible.

This paper develops a methodology to classify individuals into a reduced number of categories, exploiting the richness of health information available in panel surveys. By profiting from the panel dimension of the data we estimate transitions across groups conditioning on current health, age, education, and gender, which are of paramount importance when calibrating macroeconomic models.

We find that individuals' health can be parsimoniously represented with four different groups, namely, *healthy*, *impaired*, *physically* and *mentally frail*. While *healthy* and *impaired* have the usual extreme interpretation, the distinction between *physically* and *mentally frail* arises from the different pattern of respondents struggling with physical and cognitive tasks, respectively. Moreover, and in line with the previous literature, our empirical findings show that health status is highly persistent over time, but with significant differences in the dynamics of health across demographic groups.

We then assess our proposed classification against other commonly used health measures, finding that previous health indices are weakly related to health outcomes and medical utilization rates. In contrast, our health groups explain a significant fraction of the variance in the use of nursing homes, home health care, OOP medical expenses, and mortality. Moreover, we show that using a more accurate health measure changes significantly the saving patterns predicted by life-cycle models. In particular, self-reported health predicts a slower dissaving of individuals in bad health compared to our classification.

Finally, we make publicly available for future research our health classification for two, three, and four groups for future research exploiting the main retirement surveys: HRS, ELSA, and SHARE. Its discreteness and good fit make the classification valid for most applications; hence, it constitutes a good candidate as a unified health measure.

References

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., and Tonetti, C. (2016). The long-term-care insurance puzzle: Modeling and measurement. Technical report, National Bureau of Economic Research.

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., and Tonetti, C. (2020). Long-term-care utility and late-in-life saving. *Journal of Political Economy*, 128(6):2375–2451.

Banks, J., Muriel, A., and Smith, J. P. (2011). Attrition and health in ageing studies: evidence from elsa and hrs. *Longitudinal and life course studies*, 2(2).

Barczyk, D. and Kredler, M. (2018). Evaluating long-term care policy options, taking the family seriously. *The Review of Economic Studies*, 85(2):766–809.

Barron, A., Schervish, M. J., Wasserman, L., et al. (1999). The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics*, 27(2):536–561.

Bohacek, R., Crespo, L., Mira, P., and Pijoan-Mas, J. (2015). The educational gradient in life expectancy in europe: Preliminary evidence from share. *Ageing in Europe-Supporting Policies for an Inclusive Society*, page 321.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bound, J., Stinebrickner, T., and Waidmann, T. (2010). Health, economic resources and the work decisions of older men. *Journal of Econometrics*, 156(1):106–129.

Braun, R. A., Kopecky, K. A., and Koreshkova, T. (2017). Old, frail, and uninsured: Accounting for puzzles in the us long-term care insurance market. Technical report, FRB Atlanta Working Paper No. 2017-3.

Brown, J. (2002). Differential mortality and the value of individual account retirement annuities. In *The distributional aspects of social security and social security reform*, pages 401–446. University of Chicago Press.

Contoyannis, P., Jones, A. M., and Rice, N. (2004). The dynamics of health in the British household panel survey. *Journal of Applied Econometrics*, 19(4):473–503.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Currie, J. and Madrian, B. C. (1999). Health, health insurance and the labor market. *Handbook of Labor Economics*, 3:3309–3416.

De Nardi, M., French, E., and Jones, J. B. (2010). Why do the elderly save? The role of medical expenses. *Journal of Political Economy*, 118(1):39–75.

De Nardi, M., French, E., and Jones, J. B. (2016). Medicaid insurance in old age. *The American Economic Review*, 106(11):3480–3520.

Deb, P. and Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics*, 12(3):313–336.

Dobrescu, L. I. (2015). To love or to pay savings and health care in older age. *Journal of Human Resources*, 50(1):254–299.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.

Idler, E. L. and Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of health and social behavior*, pages 21–37.

Jones, A. M. (2000). Health econometrics. *Handbook of Health Economics*, 1:265 – 344.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., and Jaffe, M. W. (1963). Studies of illness in the aged: the index of adl: a standardized measure of biological and psychosocial function. *Journal of the American Medical Association*, 185(12):914–919.

Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

Ko, A. (2016). An equilibrium analysis of the long-term care insurance market. *Draft, University of Pennsylvania*.

Koijen, R. S., Van Nieuwerburgh, S., and Yogo, M. (2016). Health and mortality delta: Assessing the welfare cost of household insurance choice. *The Journal of Finance*, 71(2):957–1010.

Kopecky, K. A. and Koreshkova, T. (2014). The impact of medical and nursing home expenses on savings. *American Economic Journal: Macroeconomics*, 6(3):29–72.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. *Studies in Social Psychology in World War II Vol. IV: Measurement and Prediction*, pages 362–412.

Mackenbach, J. P., Stirbu, I., Roskam, A.-J. R., Schaap, M. M., Menvielle, G., Leinsalu, M., and Kunst, A. E. (2008). Socioeconomic inequalities in health in 22 european countries. *New England Journal of Medicine*, 358(23):2468–2481.

Manton, K. G. and Soldo, B. J. (1985). Dynamics of health changes in the oldest old: New perspectives and evidence. *The Milbank Memorial Fund Quarterly. Health and Society*, pages 206–285.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Meara, E. R., Richards, S., and Cutler, D. M. (2008). The gap gets bigger: Changes in mortality and life expectancy by education, 1981–2000. *Health Affairs*, 27(2):350–360.

Morris, J. N., Berg, K., Fries, B. E., Steel, K., and Howard, E. P. (2013). Scaling functional status within the interrai suite of assessment instruments. *BMC Geriatrics*, 13(1):128.

O'Donnell, O., Van Doorslaer, E., and Van Ourti, T. (2015). Health and inequality. In *Handbook of income distribution*, volume 2, pages 1419–1533. Elsevier.

Ozkan, S. (2014). Preventive vs. curative medicine: A macroeconomic analysis of health care over the life cycle. *Working Paper*.

Pijoan-Mas, J. and Ríos-Rull, J.-V. (2014). Heterogeneity in expected longevities. *Demography*, 51(6):2075–2102.

Searle, S. D., Mitnitski, A., Gahbauer, E. A., Gill, T. M., and Rockwood, K. (2008). A standard procedure for creating a frailty index. *BMC Geriatrics*, 8(1):24.

Todd, P. E. and Zhang, W. (2020). A dynamic model of personality, schooling, and occupational choice. *Quantitative Economics*, 11(1):231–275.

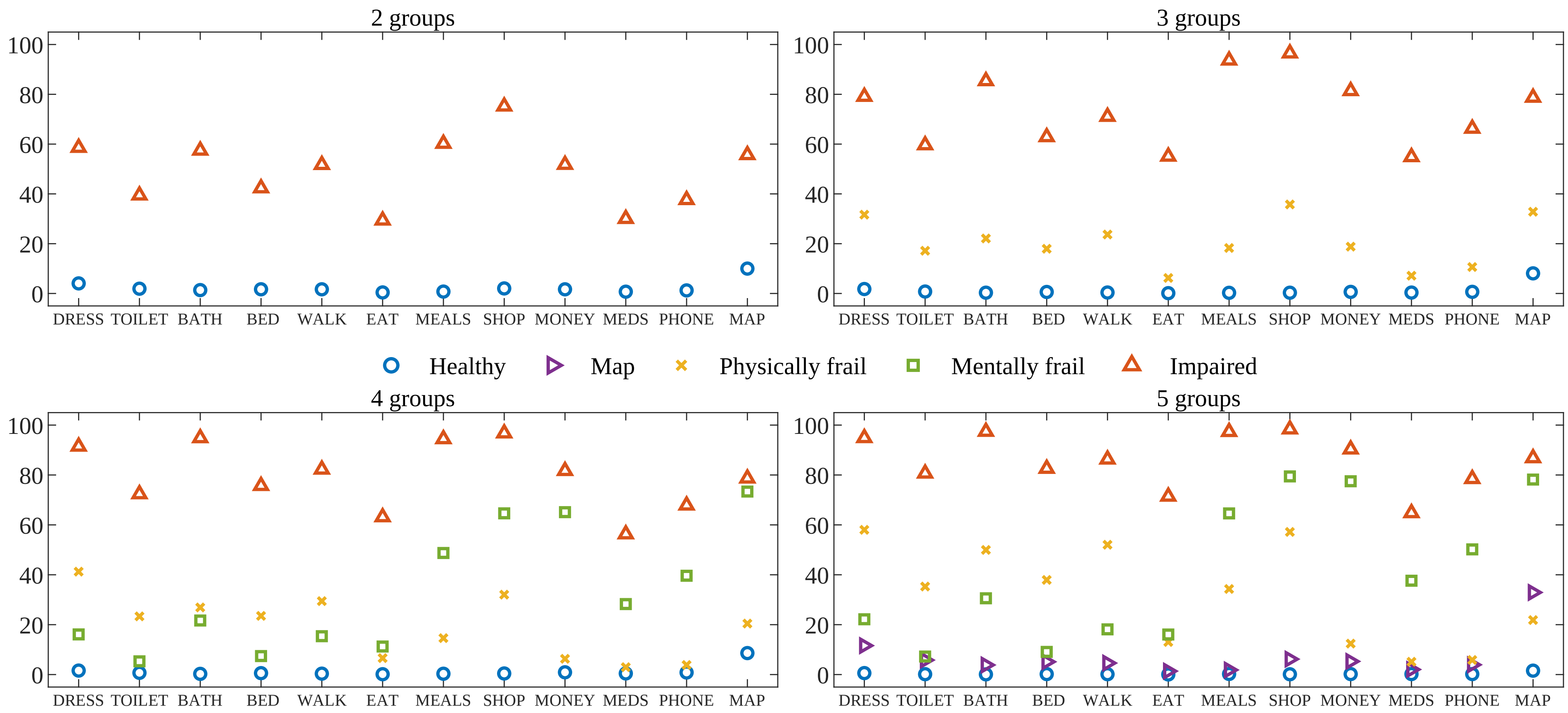
Wang, W., Phillips, P. C., and Su, L. (2018). Homogeneity pursuit in panel data models: Theory and application. *Journal of Applied Econometrics*, 33(6):797–815.

Yang, Y. and Lee, L. C. (2009). Dynamics and heterogeneity in the process of human frailty and aging: evidence from the us older adult population. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, pages 246–255.

Yogo, M. (2016). Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets. *Journal of Monetary Economics*, 80:17–34.

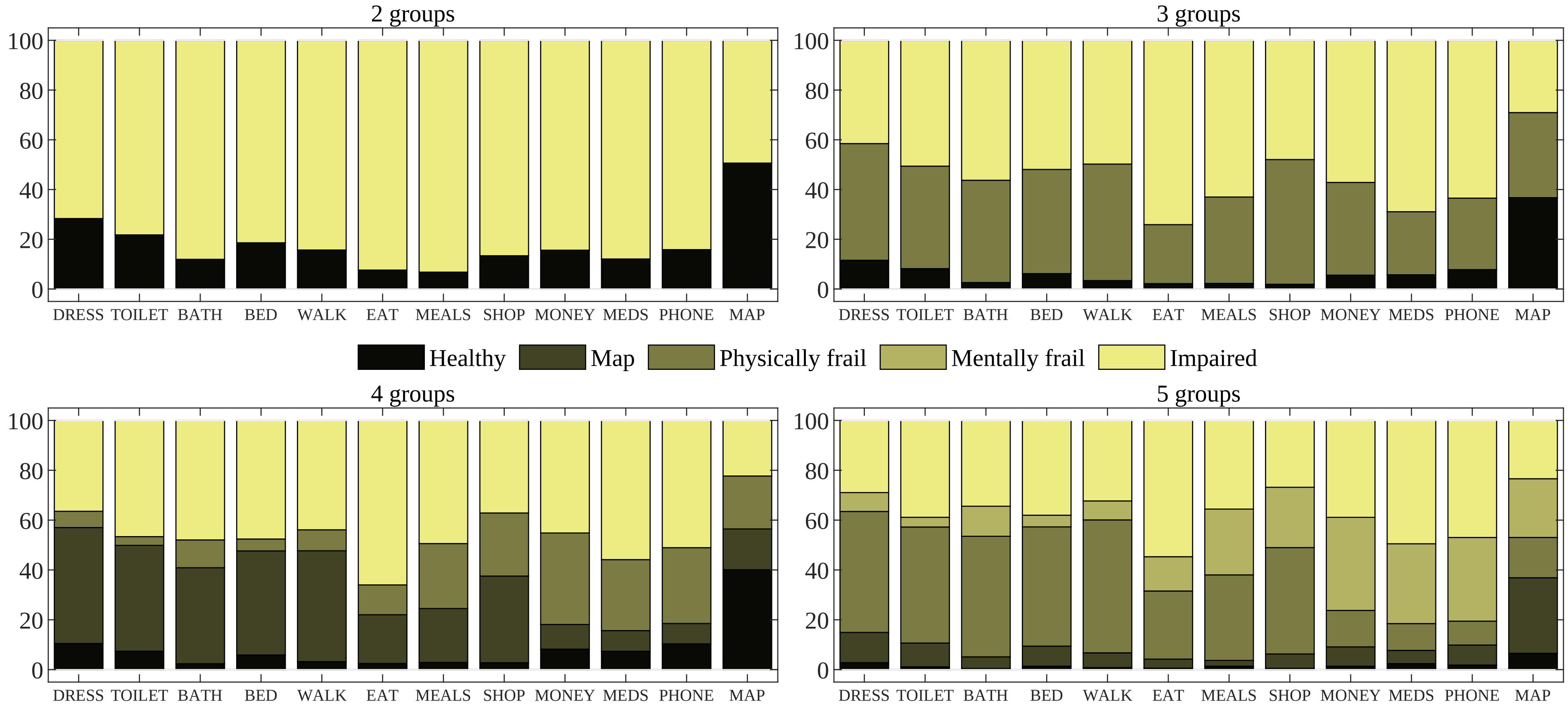
Figures

Figure 1: Probability of reporting a difficulty with a given I-ADL by health group



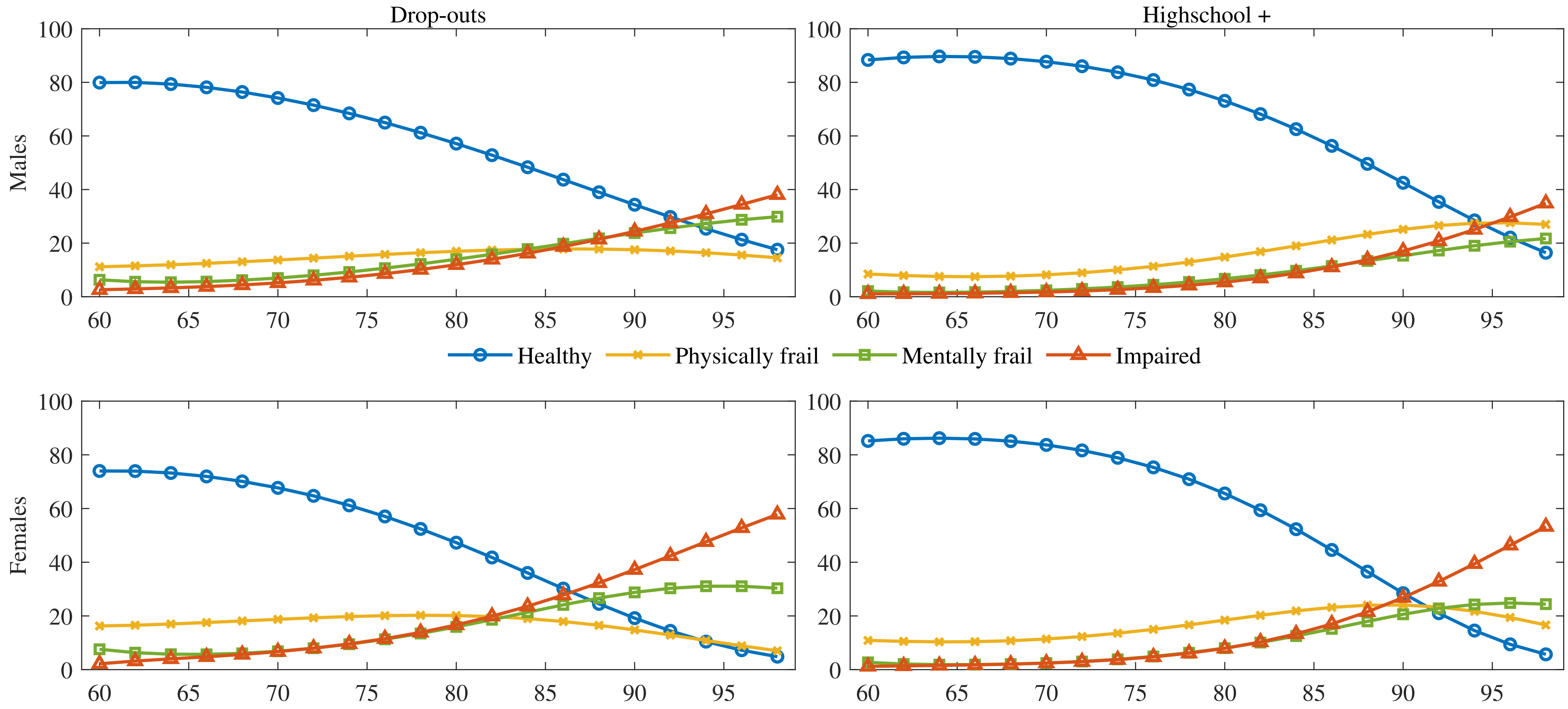
Notes: ADLs: Some difficulty with dressing (DRESS), using the toilet (TOILET), bathing (shower, BATH), getting in or out of bed (BED), to walk across a room (WALK) and eating (EAT). IADLs: Some difficulty with preparing hot meal (MEALS), shopping for groceries (SHOP), managing money (MONEY), taking medications (MEDS), using a phone (PHONE), and using a map (MAP). The units of the y-axis are percentage points.

Figure 2: Probability of belonging to a given health group by I-ADL



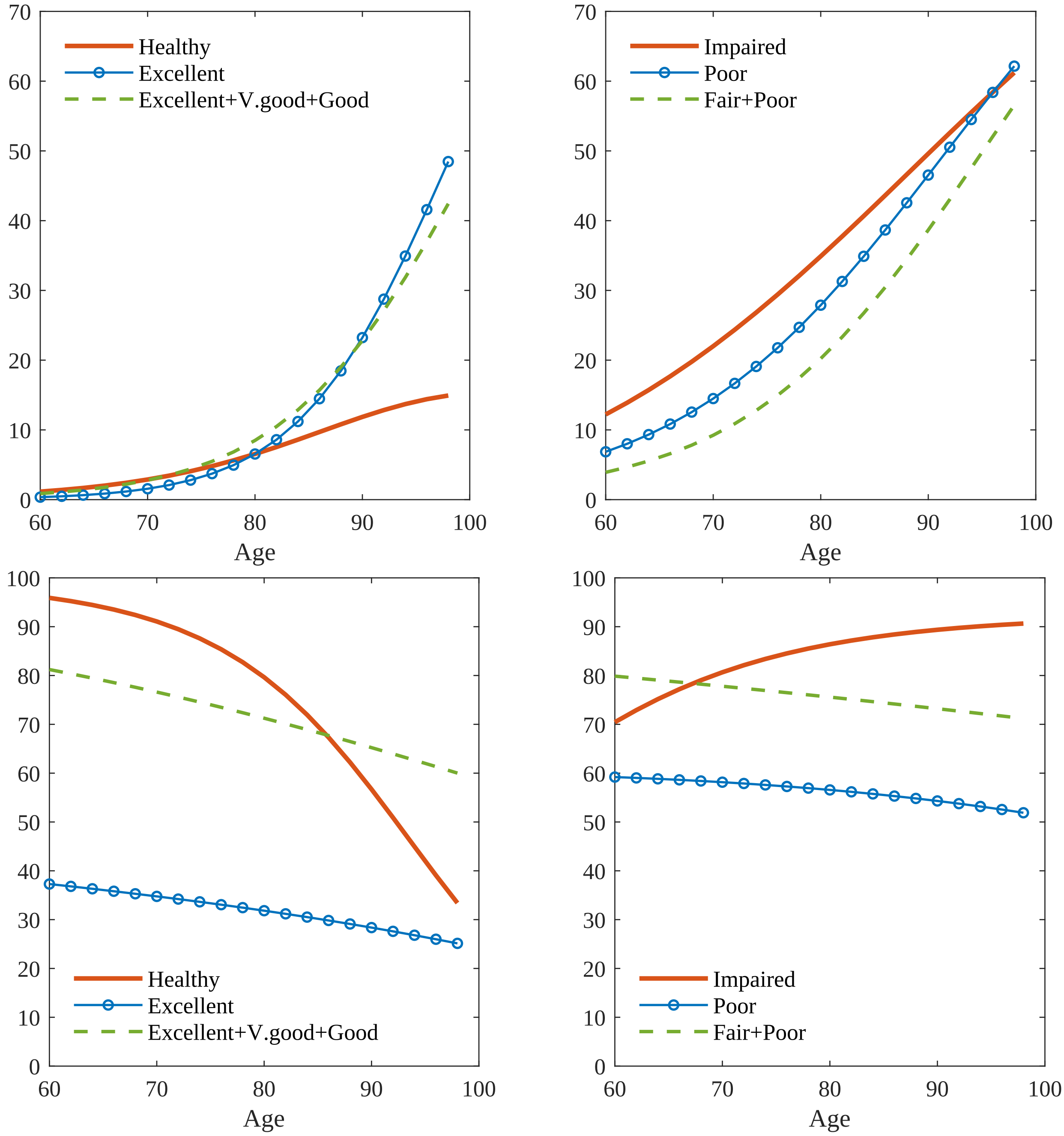
Notes: ADLs: Some difficulty with dressing (DRESS), using the toilet (TOILET), bathing (shower, BATH), getting in or out of bed (BED), to walk across a room (WALK) and eating (EAT). IADLs: Some difficulty with preparing hot meal (MEALS), shopping for groceries (SHOP), managing money (MONEY), taking medications (MEDS), using a phone (PHONE), and using a map (MAP). The units of the y-axis are percentage points.

Figure 3: Share of individuals in each group conditional on being alive by education and gender as individuals age



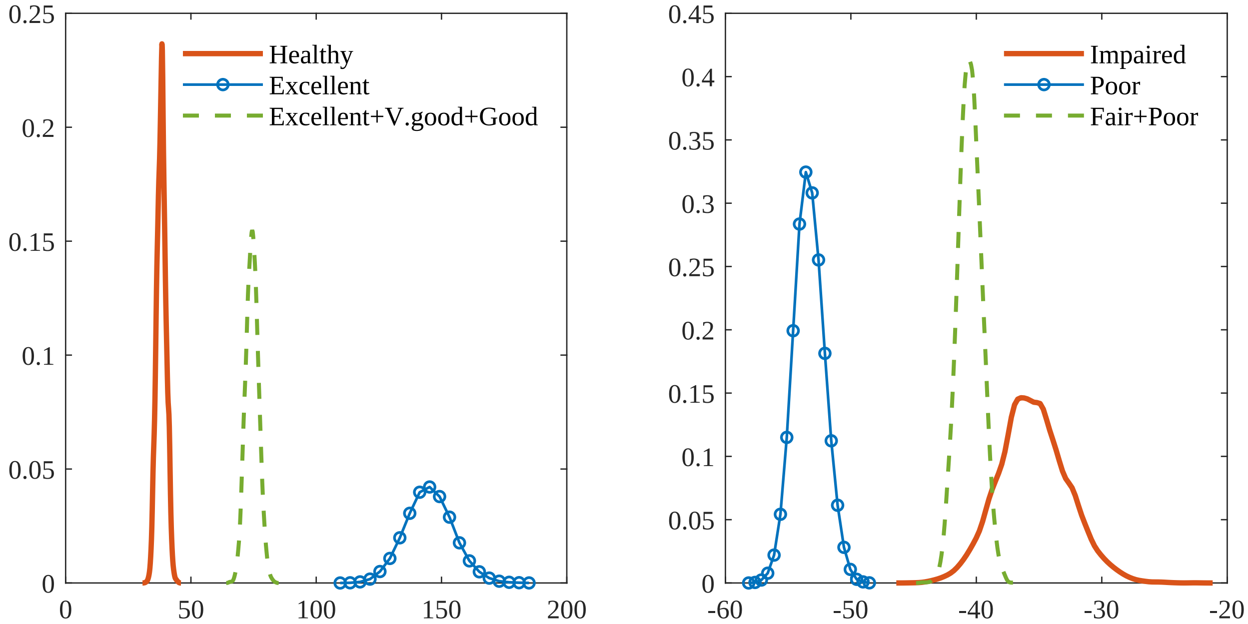
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender, or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. See Section 3 for details about the econometric model and the estimation procedure. The units of the y-axis are percentage points and those of the x-axis are years.

Figure 4: Transition to death and persistence



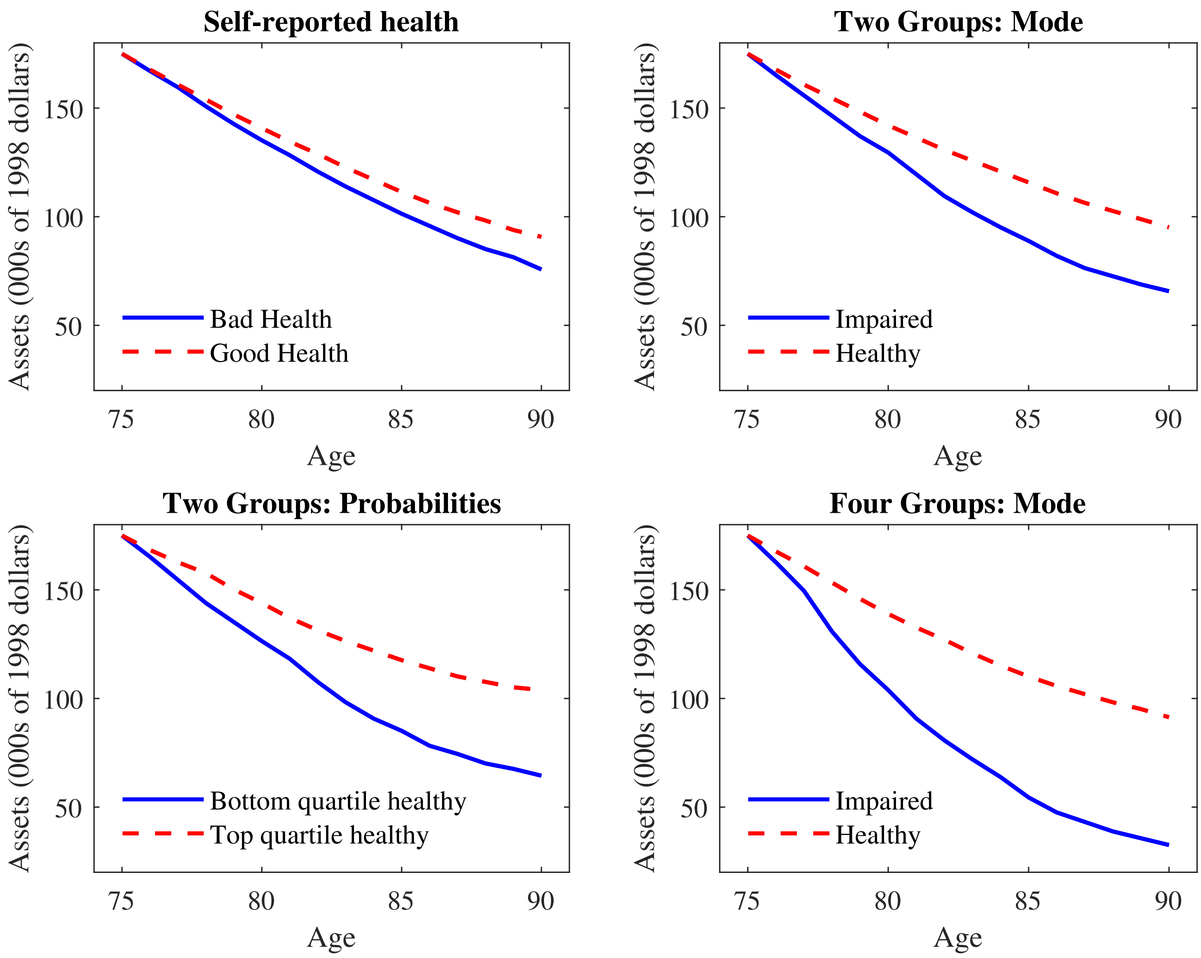
Notes: Upper plots: probability of dying per health group. Lower plots: Probability of maintaining the same health state. RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. The units of the y-axis are percentage points and those of the x-axis are years. This graph corresponds to female dropouts but it is similar if we look at other socio-economic groups.

Figure 5: Expected educational gradient across health classification



Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender, or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. The x-axis in the left (right) figure corresponds to the proportion of extra years in percentage points that a high-school graduate lives healthy (unhealthy) compared to a similar high-school dropout. The y-axis is the pdf of the posterior distribution.

Figure 6: Comparison of dissaving across health states



Notes: This figure represents the dissaving pattern in the model by De Nardi et al. (2010) of the median female rich individual who stays in the healthy state from the age 75 to age 90 (dashed) and one who has bad health from age 75 to age 90. The upper left-hand plot uses self-reported as the measure of healthy/unhealthy; the upper right-hand plot uses our two-group classification; the lower left-hand plot uses the probability of being healthy in our two group classification discretized into quartiles and we label the top/bottom quartile as healthy/unhealthy the lower right-hand plot uses our four-group classification;

Tables

Table 1: Fraction of individuals reporting difficulties with I-ADLs by self-reported health

Variable	Definition	# Obs	All	Self-reported health Exc.	Very	Good	Fair	Poor
	Activities of daily living (ADLs): Some difficulty...							
DRESS	Dressing	134,980	12.4	2.2	3.5	8.1	20.2	44.1
TOILET	Using the toilet	134,785	7.6	1.0	2.1	4.8	12.1	29.2
BATH	Bathing (shower)	134,949	10.0	1.6	2.3	5.7	16.0	40.3
BED	Getting in or out of bed	134,900	7.9	1.0	1.4	4.3	13.0	33.2
WALK	To walk across a room	134,913	9.4	1.1	1.9	5.2	14.8	39.4
EAT	Eating	134,908	4.9	0.8	1.0	2.5	7.4	21.5
	Instrumental activities of daily living (IADLs): Some difficulty...							
MEALS	Preparing hot meal	127,840	9.6	1.8	2.4	5.6	14.7	39.3
SHOP	Shopping for groceries	130,313	12.8	2.2	3.1	7.7	21.0	50.2
MONEY	Managing money	130,013	9.2	2.5	3.1	6.2	14.1	32.2
MEDS	Taking medications	131,264	5.3	1.2	1.5	3.1	7.9	20.4
PHONE	Using a phone	134,259	6.8	1.6	2.2	4.4	10.2	24.7
MAP	Using a map	117,200	15.7	6.5	8.7	13.6	23.8	39.3
	Some difficulties with...							
ADL	At least one ADL	134,366	21.1	4.0	6.9	15.6	35.6	66.0
IADL	At least one IADL	103,910	23.2	10.8	14.2	24.5	47.0	74.3
I-ADL	At least one I-ADL	103,663	29.6	10.8	16.1	28.2	51.3	78.5

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (< 0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed on average 6 waves (12 years). The column *All* indicate the percentage of observations who have problems with a given I-ADL. The last five columns present the same percentage by group of self-reported health (excellent (Exc.), very good (Very), good, fair and poor).

Table 2: Summary statistics for estimated health clusters

Group	Share	ADL	IADL	Age	Female	Dropout
		Average				
	100	21.3	33.2	72.7	57.4	27.2
		2 groups				
Healthy	85.6	9.7	21.7	71.1	56.0	23.0
Impaired	14.4	89.0	99.9	78.0	68.0	46.4
		3 groups				
Healthy	77.2	4.2	13.7	70.7	55.1	21.4
Physically frail	16.5	70.9	97.0	75.7	65.6	41.3
Impaired	6.3	97.0	99.9	80.6	67.0	49.2
		4 groups				
Healthy	78.4	4.1	14.5	70.8	55.2	21.8
Physically frail	11.8	87.4	98.4	74.6	67.1	36.6
Mentally frail	4.9	51.3	99.4	79.5	62.9	52.7
Impaired	5.0	100.0	99.9	80.5	69.5	48.8
		5 groups				
Healthy	61.1	0.8	2.1	70.1	51.0	17.0
Map	24.3	31.8	70.4	73.8	68.7	39.0
Physically frail	6.8	97.0	99.9	74.6	68.7	38.4
Mentally frail	3.9	62.1	99.7	80.4	64.2	51.3
Impaired	3.8	100.0	99.9	81.3	69.0	49.2

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. Results reported in percentage points. See Section 3 for details about the econometric model and the estimation procedure.

Table 3: Expected forthcoming time in each health group by education and gender at age 60.

Education	*Healthy*	+	*Physically frail*	+	*Mentally frail*	+	*Impaired*	=	Life Expectancy
				Females					
Dropouts	12.7		3.6		1.9		1.7		19.9
	(0.2)		(0.1)		(0.1)		(0.1)		(0.2)
Highschool	17.6		3.1		1.1		1.1		23.0
	(0.1)		(0.1)		(0.0)		(0.0)		(0.1)
				Males					
Dropouts	12.4		2.2		1.3		0.9		16.7
	(0.2)		(0.1)		(0.1)		(0.0)		(0.2)
Highschool	16.7		1.9		0.7		0.5		19.8
	(0.1)		(0.1)		(0.0)		(0.0)		(0.1)

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing ($<0.1\%$ of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. Results reported in years. In parentheses we report the standard deviation of the posterior distribution. See Section 3 for details about the econometric model and the estimation procedure.

Table 4: Long-term care needs by health classification

	Share	OOP med spending	Nurs-h resident	Received h-care	Dead next wave	ADL>0	IADL>0	IADL>0 w/o MAP
Self-reported health								
Excellent	9.2	1,805	0.6	2.5	2.4	3.5	9.4	4.5
Very good	28.0	2,129	0.7	3.9	3.2	6.4	13.1	6.2
Good	32.2	2,764	1.3	7.4	5.5	14.9	22.5	12.8
Fair	21.0	3,594	3.0	14.2	11.2	34.6	44.1	30.1
Poor	9.4	5,138	7.9	28.1	24.6	65.2	71.2	60.1
ADL: Yes/No								
No	79.6	2,357	0.3	5.3	4.7	0.0	16.7	7.8
Yes	20.4	5,005	8.8	25.7	19.0	100.0	69.5	59.2
Frailty Index Quintiles								
Lowest quintile	19.6	1,743	0.1	1.5	1.5	0.1	2.8	0.7
2	19.8	2,062	0.1	3.2	2.8	0.9	8.4	2.4
3	20.8	2,524	0.2	5.6	4.9	4.7	15.5	5.4
4	19.0	3,017	0.7	10.3	8.4	21.9	31.8	16.7
Highest quintile	20.8	5,048	8.9	25.9	20.6	72.5	77.1	64.6
4-I-ADL (i, j): ADL=i & IADL=j, IADL without MAP								
(0,0)	73.4	2,274	0.1	4.5	3.9	0.0	9.0	0.0
(1,0)	8.3	3,023	0.6	13.6	8.7	100.0	17.2	0.0
(0,1)	6.2	3,337	2.6	14.0	13.7	0.0	100.0	100.0
(1,1)	12.1	6,371	14.5	34.7	26.2	100.0	100.0	100.0
2 groups (mode)								
Healthy	86.1	3,122	0.2	5.7	4.1	9.5	14.6	6.6
Impaired	13.9	7,778	13.5	30.6	21.6	87.8	93.2	90.6
4 groups (mode)								
Healthy	77.6	3,330	0.1	4.8	3.6	3.9	11.0	3.3
Physically frail	13.0	4,524	1.5	19.2	10.9	80.0	59.2	50.0
Mentally frail	5.1	5,780	7.3	22.7	18.0	51.3	99.0	96.8
Impaired	4.3	12,411	31.8	41.4	33.1	100.0	99.9	99.9

Notes: Results reported in percentage points, except for OOP medical spending which is reported in 2000 US dollars. See Section 2 for details about the data and Section 3 for details about the econometric model and the estimation procedure.

Table 5: Fraction of explained variance by health classification

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
No health	0.7	0.8	4.3	5.1	3.6	3.5	5.9
SRH (2 groups)	1.5	1.3	6.0	6.2	7.3	6.2	9.3
SRH (5 groups)	1.9	1.5	7.1	6.8	9.0	7.3	11.2
ADL: Yes/No	2.2	1.7	11.4	9.8	9.9	7.5	9.8
Frailty index	2.7	2.3	12.6	11.5	11.6	9.6	12.1
All I-ADL dummies	4.0	2.7	31.5	20.2	13.1	8.7	13.1
4-I-ADL	3.2	2.5	16.2	13.8	12.3	9.1	11.9
2 groups (mode)	2.6	2.2	15.8	13.7	11.3	8.2	11.4
4 groups (mode)	3.8	2.5	26.3	18.2	12.8	9.1	12.8
4 groups (probabilities)	4.0	2.7	27.3	19.0	13.5	9.6	13.2
Observations	118,706	94,544	118,706	94,544	117,408	93,268	102,292

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing ($<$0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. Then, we restrict the sample to those observations that can be classified according to all criteria. Results correspond to the R^2 (in percentage points) of the following regression::

$$y_{i,t} = c + \mathbf{d}'_{i,t}\beta + \mathbf{z}'_{i,t}\gamma + (\mathbf{d} \otimes \mathbf{z})'\,\theta + age_{i,t}\left(\mathbf{d}'_{i,t}\beta_1 + \mathbf{z}'_{i,t}\gamma_1\right) + \varepsilon_{i,t}$$

where $y_{i,t}$ is the variable used as a reference, $\mathbf{z}_{i,t}$ includes gender and education, and $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs.

Table 6: Fraction of explained variance by health classification: SHARE and ELSA

	OOP medical spending	Nursing home resident	Received home care	Mortality in 2 years
No health	0.5	0.4	6.7	4.4
SRH (2 groups)	0.7	0.6	10.2	6.5
SRH (5 groups)	1.0	0.9	11.7	8.3
ADL: Yes/No	1.2	1.3	11.7	6.9
Frailty Index	1.2	1.2	13.1	8.1
All I-ADL dummies	1.9	2.8	16.6	9.5
4-I-ADL	1.5	2.0	15.6	8.9
4 groups (mode)	1.9	2.7	15.6	9.4
4 groups (probabilities)	2.0	2.8	16.2	9.6
Observations	74,196	219,989	61,174	159,121

Notes: SHARE and ELSA Data; sample from 2002 to 2017 (7 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing ($<0.1\%$ of observations). The final sample consists of 71,416 individuals. Then, we restrict the sample to those observations that can be classified according to all criteria. Results correspond to the R^2 (in percentage points) of the following regression:

$$y_{i,t} = c + \mathbf{d}'_{i,t}\beta + \mathbf{z}'_{i,t}\gamma + (\mathbf{d} \otimes \mathbf{z})'\,\theta + age_{i,t}\left(\mathbf{d}'_{i,t}\beta_1 + \mathbf{z}'_{i,t}\gamma_1\right) + \varepsilon_{i,t}$$

where $y_{i,t}$ is the variable used as a reference, $\mathbf{z}_{i,t}$ includes gender and education, and $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs. ELSA does not report data on OOP medical expenditures.