Endogenous Health Groups and Heterogeneous Dynamics of the Elderly*

Dante Amengual

CEMFI

<amengual@cemfi.es>

Jesús Bueren

EUI

<jesus.bueren@eui.eu>

Julio A. Crego

Tilburg University

<jacrego@uvt.nl>

*We would like to thank Nezih Guner, Barton Hamilton, Pedro Mira, Josep Pijoan-Mas, and Silvia Tiezzi for their comments and suggestions. 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), and Universidad Carlos III de Madrid. 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 novel methodology to classify individuals into groups of health and characterize their transition across these groups as they age. We use MCMC techniques to estimate a panel Markov switching model that exploits information from both the cross-sectional and time series dimensions. Using the Health and Retirement Study, we identify four clearly differentiated and persistent health groups, depending on individual's physical and mental disabilities, with heterogeneous transitions across gender and education. Our classification outperforms existing measures of health used in the literature at explaining entry in nursing homes, home health care, out-of-pocket medical expenses and mortality.

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

JEL: C23, C38, E21, I14.

1 Introduction

Life-cycle models with heterogeneous agents are becoming increasingly popular among scholars as a tool for designing optimal policies related to Social Security, inequality, insurance markets or health care among others. In order to use these models as accurate laboratory economies, it is therefore crucial to appropriately capture health and earnings risk in order to understand individuals' decisions. As a result, macro models are benefiting from a recent and influential strand of the empirical literature estimating richer earning dynamics (see Guvenen et al. 2015; Arellano et al. 2017) and analysing their macroeconomic implications (De Nardi et al. 2018; Gorea and Midrigan 2017, among others). Different from earnings, health is a multidimensional attribute hard to measure and summarize.

This paper proposes a dynamic latent variable model for jointly estimating a parsimonious health classification and the associated process for health transitions. Health dynamics are allowed to differ across gender and education types to capture heterogeneity in health risks across the population. The methodology exploits both the cross-sectional and the time-series dimension of panel data sets based on detailed surveys, which contain a wide array of variables about different aspects of elderly's health. If we restrict to the cross-sectional dimension, our method mimics the latent class model (Lazarsfeld, 1950); that is, it allocates individuals to latent groups to fit the joint distribution of all the observed health variables considered. On the other side, along the time-series dimension, our strategy emulates Hamilton (1989)'s model inasmuch as it infers the health status of an individual at a point in time using her whole time-series through the auto-correlation structure. Altogether, we assign each individual at each point in time to a given group using her health information and that of every individual in past, current and future periods. We thus reduce the dimensionality of the data to a discrete variable which corresponds to the endogenous groups.

We apply the methodology to allocate individuals in the Health and Retirement Study (HRS) into four groups according to their difficulty with Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs). Additionally, we characterize health dynamics and survival as a hidden Markov chain which incorporates heterogeneity across age, gender, and education. Precisely, we model transitions across health groups as logistic functions of the aforementioned attributes whose parameters change depending on the current health status.

Our modelling approach presents three desirable features. First, it considers the classification of individuals and groups' dynamics jointly. This way the health classification is not based solely on the information of the current period but on all the observations including death events. Moreover, potential misreporting is smoothed out by the algorithm which reduces possible biases affecting groups' dynamics. Secondly, even though the resulting health measure is discrete, we also obtain as a by-product the probability of belonging to each group conditional on the whole sample, which enables to weight observations according to their representativeness of each group using a continuous measure. Third, the latent nature of our groups allows classifying an individual's health even in the case of missing information as long as we have past or future information.

The empirical strategy requires the estimation of thousands of hidden Markov chains, one per individual, together with hundreds of parameters. For that reason, we resort to Markov chain Monte Carlo methods. In particular, we rely on a Metropolis-within-Gibbs algorithm which involves two main blocks. First, given the health group of each individual, it is straightforward to sample the parameters driving the I-ADLs binary processes through a Metropolis step; and likewise, the parameters ruling the dynamics.[1] Then, conditional on these parameters we obtain, for each individual, a realization of the latent health group using Kim (1994)'s smoother algorithm. To save the computational burden to future researchers, the probabilities for each individual and time are available at the authors' website. Further, based on our results we suggest an estimation-free classification that improves currently used ones, although it performs worse than the endogenous one.[2]

Four groups which divide individuals into *physically frail*, *mentally frail*, *impaired*, and *healthy* represent health suitably. The *impaired* have both types of limitations, physical and cognitive, while the *healthy* have no or light difficulties with I-ADLs.[3] 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 gerontology literature (e.g. Morris et al., 2013), not all the I-ADLs are equally informative for classifying individuals in health

[1] Along the paper we use I-ADLs to denote the set of both ADLs and IADLs; likewise, I-ADL refers to one of these variables.

[2] Bueren's webpage provides the probabilities of each individual in the HRS identified by *hhidpn* and *wave*, and the parameters of the model.

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

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

Groups' dynamics features stylized facts previously documented in the literature of aging (Manton and Soldo, 1985): older individuals have relatively worse health, 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) and Meara et al. (2008), we find a large educational gradient in life expectancy. Nonetheless, despite living longer, educated individuals spend, on average, less time *impaired*, consistent with Pijoan-Mas and Ríos-Rull (2014). Even though any health classification reveals the protective effect of education, they lead to very different magnitudes. Precisely, while high-school graduates live on average around 30% less time in our unhealthiest group and 40% more in the healthiest one, these gradients equal 55% and 140% if we rely on self-reported health.

Aside from education, current health status constitutes an important source of heterogeneity because of the groups' persistence. For instance, a 75-years-old *impaired* respondent has a probability of remaining *impaired* of 60%; thus she faces a health risk different from a *healthy* respondent who stays *healthy* with 80% probability. This feature is consistent with our groups being closely linked to long-term care (LTC) needs and it is less pronounced in the case of self-reported heath.

We then compare access to medical and care services across health groups based on the estimated probabilities. On average, *impaired* (*healthy*) individuals spend around $10,043 ($2,310) per year in out-of-pocket medical spending. Likewise, *mentally frail* individuals spend $1,343 more than *physically frail* ones, who employ $3,565. The use of LTC services also presents large differences across groups. While 9% of the individuals *mentally frail* live in a nursing home at the time of the interview, only 1.6% of the *physically frail* do so. This disparity widens between members of the *healthy* group, who avoid the nursing home almost surely, and those of the *impaired*, out of which 33.6% reside in these facilities. A similar pattern arises if we compare the received professional care of these two extreme groups. Nonetheless, *mentally* and *physically frail* individuals need a medical-trained person to look after them at home with the same probability.

Finally, we contrast our estimated health groups with other commonly used health classifications, 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.[4] To do so, we consider three main variables associated with health-related spending, particularly, out-of-pocket medical expenditures, and 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; Barczyk and Kredler 2018; Ameriks et al. 2015). Our four groups classification generates more differentiated groups; furthermore, it explains about three times more variance than self-reported health and twice as much as the use of an ADL indicator. These results resemble an out-of-sample exercise since these variables do not enter the classification model. Additionally, we analyze the ability of the different classification to predict mortality and find that our four groups dominate the alternatives.

Our paper complements the literature analyzing the effect of health on economic decisions. This literature relies on dynamic structural models to quantify the importance of mechanisms or to derive implications for policymaking. 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. Van der Klaauw and Wolpin (2008) and French and Jones (2011) divide individuals into two groups of self-reported health to analyze how health affects the retirement decision. De Nardi et al. (2010) use the same strategy to quantify the effect of health-related expenses on the savings decision of the elderly. With a similar objective, Ameriks et al. (2015) and Barczyk and Kredler (2018) classify individuals as unhealthy if they report a difficulty with ADLs or require care, respectively. Alternatively, Braun et al. (2017) splits a frailty index into five quintiles to introduce health in their insurance demand model.

This paper relates to an extensive literature which 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. Moreover, they disregard health dynamics which is of first-order relevance: Contoyannis et al. (2004) stress the importance of

[4] 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.

health persistence using a dynamic panel ordered probit model for self-reported health.

We also contribute to a growing literature that summarizes health variables into a single index that explains most of the variation related to health (see Searle et al., 2008). Regarding 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; though, they assume that individuals are completely unable to self-insure against medical expenses.

The rest of the paper is structured as follows. We briefly describe the HRS data in Section 2. Then, the econometric model and the estimation strategy are presented in Section 3. 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 data comes from the RAND HRS dataset which comprises a cleaned version of the Health and Retirement Study conducted by the University of Michigan.[5] It contains subjective and objective indicators of health, as well as demographic and economic characteristics, of a representative panel of US households surveyed biannually from 1992 to 2014. In addition, the HRS exit interview records the death of the individual and includes the 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 from 1996 until 2014, which includes ten waves. 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 six waves (12 years). Figure 1 shows that the composition of the sample reflects the survival

[5] 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.

probabilities. 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.

[Figure 1 about here.]

The HRS provides dozens of health-related variables, but we restrict to individual's ability to perform Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) to infer the 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 with cognition as the possibility of using a phone or controlling her medication. Accordingly, 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 utilize twelve binary variables, denoted as I-ADLs, which include six ADLs and six IADLs that describe whether individuals have any difficulty to perform 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 realizing each of them. The most common ADL is not being capable of dressing (12%) whereas eating is the ADL that present fewer difficulties (5%). Likewise, 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, 30% of respondents battle with at least one I-ADL. These probabilities, nevertheless, change substantially across demographic groups and age as Figure 2 shows. When they are 60 years old, more than 40% of the individuals who drop out high school already report difficulties with at least one I-ADL. On the other hand, only one high-school graduate out of five struggle with daily activities. Regarding gender, these proportions are also heterogeneous since 22% of females present some type of difficulty compared to 19% in the case of males. The differences across gender shrink as people age; while at the same time, the share of them facing troubles with an ADL or IADL increases for all groups systematically.

[Table 1 about here.]

[Figure 2 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, in the last five columns of Table 1 we 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 present problems with I-ADLs, nonetheless, the importance of each activity differs. In particular, individuals reporting poor health are not able to walk, dress or bath with probabilities around 40%, while for the remaining three ADLs the corresponding figures barely surpass 30%. Similarly, difficulties with IADLs are also diverse within the worst self-reported health groups since 50% of individuals endeavor to shop but only 20% encounter complications to take their medications.

3 Econometric model

We have an unbalanced panel of individuals $i = 1, \ldots, N$ followed for $t_i = 1, \ldots, T_i$ periods which correspond from age a_1^i to age $a_{T_i}^i$ where $a \in (\underline{a}, \overline{a})$. For each individual, we observe K dummy variables corresponding to each I-ADL across time $(x_{1,i,t}, x_{2,i,t}, \ldots, x_{K,i,t})$, 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. Although we take missing observations into account under the assumption that they occur completely at random, 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. Conditioning on education, e; age, a; and gender, s; the current health cluster of individual i

is independent of previous health clusters except for the most recent one (Markov first-order property). Besides transiting across health groups, individuals may also die which is represented by an observable and absorbing state labeled as D.

Specifically, we consider that individual i at time t belongs to a health group $h_{i,t}$ out of H possible ones. Given her group is 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 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. 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

$$p_{g,c}(a_{it}, s_{it}, e_{it}) = \frac{\exp[f_{g,c}(a_{it}, s_{it}, e_{it})]}{1 + \sum_{c \in \mathcal{H}} \exp[f_{g,c}(a_{it}, s_{it}, e_{it})]} \tag{2}$$

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

$$p_{g,D}(a_{it}, s_{it}, e_{it}) = \frac{1}{1 + \sum_{c \in \mathcal{H}} \exp[f_{g,c}(a_{it}, s_{it}, e_{it})]}.$$

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. To do so, we use a Gibbs sampling procedure to estimate the models for different choices of the number of health groups H. 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 conditional on all the other parameters, 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 all health groups, and μ and β as the vectors stacking the parameters of the I-ADLs process and the transition probabilities, respectively. 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. $p(\mathbf{h}_i^{(m)}|\beta^{(m)}, \mu^{(m)}, \mathbf{X})$: sampling the latent health indicator for each $i = 1, ..., N$ using the Kim (1994)'s smoother.

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

3. $p(\beta^{(m)}|\mu^{(m-1)}, \mathbf{H}^{(m-1)}, \mathbf{X})$: sampling the transition 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 every 100 draws is retained.

3.1.1 Sampling the states: Kim's Smoother

To sample the states, we apply the methodology developed by Kim (1994):

1. Using 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:
$$p(h_{i,t} = g|\beta, \mu, \mathbf{X}, h_{i,t+1} = c) = \frac{p(h_{i,t+1} = c|\beta, h_{i,t} = g) \cdot p(\mathbf{x}_{i,t}|\mu, h_{i,t} = g)}{\sum_{g\in\mathcal{H}} p(h_{i,t+1} = c|\beta, h_{i,t} = g) \cdot p(\mathbf{x}_{i,t}|\mu, h_{i,t} = g)} \quad \forall g, c \in \mathcal{H}$$

As a result, 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 also incorporates information about the individuals' death wave, as well as her age, gender, and education.

To form a complete likelihood, we need to know the unconditional distribution of $h_{i,1}$ for each i, $p(h_{i,1}|\beta)$. Since the model is non-stationary due to its dependence on age, we cannot compute the unconditional distribution without further assumptions. In particular, we consider the unconditional distribution at the age of 60 coincides with the stationary distribution given by the parameters of the first transition (from 60 to 62).

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 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. 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 set of parameters 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.[6] 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.

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

3.2 Obtaining moments

In most applications, as in the following sections, researchers aim to compute several sample moments conditional on a given health level. Our model, however, results in a probability of being in each group even if one fixes the parameters. While we can impute individuals to their most likely groups, using these probabilities to weight observations enhances our measure without losing the discrete nature of the variable.

For instance, assume the researcher wants to obtain the expectation of several outputs, say $\mathbb{E}_M[A(\mathbf{x}_{i,t})|\mathbf{X}]$, where M denotes the specific structural economic model in hand and A are the quantities of interest. In our context, the dimension of the state space is greater than 2^K, thus she must discretize $\mathbf{X}$ into $\tilde{X} = \{\tilde{x}_1, ..., \tilde{x}_b\}$ and then the final result equals $\mathbb{E}_M[A(\tilde{x}_{i,t})|\tilde{X}]$. First, our procedure provides a natural way of obtaining $\tilde{X}$. Second, the proposed methodology also determines the probabilities of each $\tilde{x}$ given the sample such that we can obtain

$$\sum_{\tilde{x}\in\tilde{X}} \mathbb{E}_M[A(\tilde{x}_{i,t})|\tilde{X}] \cdot P(\tilde{X}|\mathbf{X}).$$

Thus, even though we can only compute A at some points, we can weight each observation by its representativeness of each group.

4 Health groups

We first describe how the algorithm classifies individuals into groups and then how health evolves as individuals age taking into account differences in education and gender. To define the groups, the model identifies those that explain the joint distribution of difficulties with I-ADLs the best, taking into account the dynamics. In this context, the only parameter that is not endogenous is the total number of clusters, whose value we vary from two to five to discern what is the contribution of each successive cluster.

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

4.1 Endogenous classification

Figure 3 reports 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 is, the more likely is that an individual in that specific group struggles with the corresponding I-ADL.

[Figure 3 about here.]

If 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 owns a higher likelihood of facing problems with every I-ADLs. We label the former group as *healthy* (circumferences) and the latter as *impaired* (triangles). We also find large differences in the probabilities across I-ADLs within the *impaired* group which suggests that activities differ in their importance for categorizing individuals. For example, as regards the *impaired* group these probabilities range from 31% in the case of eating to 77% in the case of shopping.

The upper right panel of Figure 3 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 with again the highest probabilities of struggling with all 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 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 remaining of the paper, we focus on the case of four groups.

While Figure 3 characterizes individual's 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 present 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.

To overcome this issue, Figure 4 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 4 about here.]

Following the same example, if a person has difficulties to eat, she belongs to the *impaired* group with probability 90%, 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; thus, MAP is uninformative. The pattern of these two I-ADLs remains unchanged when $H = 3$ and $H = 4$; MAP is never informative while EAT is the best indicator to classify individuals into the *impaired* group. This evidence is in line with previous evidence in the medical literature (see Morris et al., 2013, and references therein) which argues that difficulties with eating are the best predictor of full dependence.

Figure 4 characterizes the importance of each I-ADL separately for descriptive purposes; however, the joint structure of these variables also contributes significantly to identification. 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; actually, they face problems with one I-ADL almost surely. The other side of the coin is the *healthy* group which probability of reporting troubles with ADLs varies between around 4% 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 80% of respondents struggle with ADLs and 61% with I-ADLs, the latter faces more problems with IADLs (100%) and less with ADLs (55%).

[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 they 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. Next, in terms of education, high school graduates are overrepresented in the *healthy* group which is 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 pattern of heterogeneity of health dynamics across gender and education groups.

4.2 Heterogeneous health dynamics

The distribution of elderly into health groups changes with age, gender and education. Figure 5 plots the probability of being in each group through age. The left panels correspond to 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 is *healthy* at early ages but starting at age 90, *impaired* becomes the predominant group. Further, the *physically* and *mentally frail* have very different dynamics. The former is stable throughout life while the latter increases steeply as elderly age. These patterns are very similar across education and gender, although the initial composition of individuals varies with demographic characteristics.

[Figure 5 about here.]

Since, in the estimation, mortality and health deterioration is allowed to vary by education group, we find that dropouts and high-school graduates encounter a very distinct health risks. Table 3 shows the expected time an individual at age 60 lives in each health group. Even if the more educated elderly 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 stay 60% more time (or 0.3 extra years) in the *impaired* state. This empirical fact generates extra motives for precautionary savings for low-income earners. Nonetheless, the superior life expectancy of high-school graduates also increments their need for savings due to uncertain life span.

[Table 3 about here.]

Individuals' incentives might also change across health groups since their expected health path might differ. Figure 6 displays the transition probabilities according to age and current health status. For example, a *healthy* elderly owns a very low probability to become *impaired*, thus a low health risk, everything else equal. In contrast, once an individual enters the *impaired* group, she is very likely to stay in that group; hence, her expected future medical spending is very high. In general, groups are very persistent and health is more likely to worsen than to improve, in line with our interpretation of the endogenous groups as different levels of LTC needs. Although *mentally frail* and *impaired* individuals do not recover, their large mortality rates limit the time spent in high levels of need.

[Figure 6 about here.]

5 Comparison with alternative indices

The need for a discrete measure of health has led researchers to use ad-hoc classifications. In this section, we compare our endogenous classification with the main three alternatives: self-reported health, if the individual struggles with an ADL, and the quintiles of a frailty index. In addition, we also consider 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 endogenous classification. To perform the comparison, we focus on mortality and three variables related to the financial risk due to health: OOP medical expenditures, and indicators of receiving home-care and residing in a nursing home. 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 an individuals self-assessment on 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 more about their health than researchers can ever measure. On the other hand, respondents might misjudge their health condition, incorporate other information as mood or consider different benchmarks of being *good*. Previous literature has analyzed the net effect of these two channels and establishes that the disadvantages often offset any benefit. For instance, Crossley and Kennedy (2002) directly checks the reliability of self-assessment and finds that 28% of individuals change their answer from the beginning to the end of the survey. Moreover, this measurement error correlates with important socioeconomic variables; hence, it raises concerns about the validity of self-reported health (see Currie and Madrian, 1999, for a survey).

Nevertheless, the first panel of Table 4 confirms that self-reported health has information about the financial risks. Those 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. The difference between the five groups varies though. In particular, answering *excellent*, *very good*, and *good* relates to almost the same risk, whilst *fair* and *poor* correspond to much more spending. Previous literature, thus, merges the three healthiest and the two worst groups. We denote this latter classification as *self-reported health (2 groups)*.

Grouping individuals according to if they have 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. Actually, elderly who struggle to eat are usually more dependent than those who are unable to dress themselves (Williams et al., 1994).

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 them to allocate 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 present the same features as those who have an ADL in the Yes/No classification.

Finally, classifying individuals regarding whether they struggle with at least one ADL, IADL, both or none, which we denote as 4-I-ADL, can be understood as a simple approximation to our four groups. In contrast to the frailty index by Braun et al. (2017) who effectively separates individuals without problems with any ADL in four groups, this method divides respondents who recognize problems to perform an ADL into three groups. Since these individuals are more heterogeneous, the resulting groups become more differentiated in all the variables considered.

Even if the four aforementioned alternative classifications are highly correlated with the health outcomes that we use, our estimated groups seem to be more differentiated across them. For instance, using our methodology, the average difference in terms of OOP between *healthy* and *impaired* elderly is \$7,751. According to self-reported health, however, an individual belonging to the worst group only expends \$3,333 more than one in the best group. Similarly, the fact that you report an ADL implies that your average OOP medical spending is \$2,648 higher; meanwhile, being a part of 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 hardly surpasses \$4,000. As for the intermediate groups, they are again less distinct in the case of the alternative classifications as their increment in spending is below \$1,000 except from the two worst groups, compared to \$1,281 which is the minimum difference between our groups.

Regarding the probability of residing in a nursing home, a similar pattern arises and the difference between the best and worst of our health groups at least duplicates the same difference using the alternative methods. The same holds true for home care when we look at self-reported health or struggling with at least one ADL but, 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 40% probability whereas only 3.4 out of 100 *healthy* ones do not survive to the next wave. Instead, the difference between the healthiest and unhealthiest groups does not reach 25 percentage points with alternative classifications. Though relevant, this result might follow from the inclusion of death in the classification algorithm.

5.1 A horse race

Most of the time, the researcher's concern might not be to classify individuals into distant groups but to create a categorical index that captures most of the variation coming from health. To assess the performance of the 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.[7] In the case of our classification, we use two alternative approaches. First, we substitute $\mathbf{d}_{i,t}$ by a vector containing the probability of individual i at time t of belonging to each cluster (we label it Probs). Secondly, we assign each individual to her most likely state (which we label as Mode).

[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 an ADL. If we also include IADLs, the fit improves by 1 percentage point. Altogether, our classification explains 4.4% of the medical spending variance which exceeds every alternative. The same conclusion arises from considering the spending reported in the following wave.

Nursing home residency, by virtue of being binary, contains a lower measurement error; nevertheless, the same ranking persists. Any measure that includes ADLs beats self-reported health by at least 5 percentage points, which doubles if we consider our unsophisticated method. Further, weighting each I-ADL, our health groups enhance the naive 4-I-ADL by 65% because it identifies the extreme dependent individuals better. In sum, our proposed classification explains almost 4 times more variance than self-reported health and 2.5 times more than *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

[7]The exclusion of the covariates does not modify our results. It just changes the level of the R^2 for all classifications

measure, although remains to be the optimal, barely improves classifications based on ADLs. Nonetheless, it explains 50% more variance than self-reported health.

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 exceeds by 0.7 percentage points that of self-reported health which indicates that part of the improvement on the mortality prediction relies on the incorporation of I-ADLs, and not on the use of death in the classification algorithm.

5.2 Dynamics: self-reported health versus endogenous classification

The comparison regarding groups' dynamics generates new 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 groups 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 groups of self-reported health in the comparison.[8]

There are two main risks associated with health transitions which increase the incentives to save. The first one is survival risk. Individuals optimally want to consume everything but the bequest they desire to leave before their death day. In reality, however, this day is not known, hence they have to save in case they live more than expected. The second risk relates to the direct costs of health. Under the fear of entering into a health status with high medical costs, individuals increase their savings.

Figure 7 reports the median probability of dying. The left panel corresponds to the healthiest groups, whereas the right panel presents the results for the most unhealthy ones. Up to the age of 80, individuals who report an *excellent* health, as well as those classified as *healthy* own very small probabilities of dying. After this age, elderly with a low survival probability still assess their health as *excellent*. On the other hand, 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

[8]One groups includes *excellent*, *very good*, and *good*; while the other comprises *fair* and *poor*.

status; thus, respondents of age 65 and 90 have a different benchmark. Furthermore, while the difference between the 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. In addition, *impaired* individuals feature a higher death probability than those who assess themselves as in *poor* health at any age.

[Figure 7 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 tomorrows health and survival probabilities thus affecting individuals' saving behavior. Additionally, the persistence of each classification sheds some light on the type of health process. In particular, we aim to create an indicator of LTC needs which is by definition persistent in contrast to others such as the flu or a sprained ankle. Figure 8 depicts the probability of remaining in the same group conditional on the group you are at a given age. We find that individuals that report excellent in one wave have less than 40% of probability to provide the same answer in the following wave, whereas respondents classified as *healthy* are extremely likely to remain in that state. This fact indicates that some non-persistent factors might drive self-reported health. If we focus on individuals in bad health, our classification displays a larger persistence as individuals age which is in line with the idea that as you become older the harder it is to recover. In contrast for fair and/or poor self-reported health, individuals are more likely to report improvements in their health status as people age which points towards changes in their health benchmark.

[Figure 8 about here.]

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.[9] To shed some light on the former, Figure 9 plots the additional percentage of time than 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, using self-reported, these differences at least double. Given that our classification was able to explain a

[9]Using a fully-fledged structural model, Bueren (2018) shows that the importance of medical expenses as a saving motive might have been overstated due to the use of self-reported health.

larger fraction of the variance of different health outcomes, these results suggest that self-reported health contains a measurement error correlated with education. More precisely, low educated individuals tend to report worse health status or high-school graduates overestimate their wellness, or both.

[Figure 9 about here.]

6 Conclusion

As retirees age, they face large risks of requiring persistent and expensive care. The macroeconomic literature underlines the importance of this uncertainty to explain the dissaving pattern of the elderly and the labor supply decisions of the individuals close to retirement. They 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 the health information available in panel surveys. In addition, 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.

Individuals LTC needs 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 ADLs and IADLs. Moreover, and in line with the previous literature, 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 measures. Our comparison exercises show 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, out of pocket medical expenses, and mortality.

References

Ameriks, J., Briggs, J. S., Caplin, A., Shapiro, M. D., and Tonetti, C. (2015). Long-term care utility and late in life saving. Technical report, National Bureau of Economic Research.

Arellano, M., Blundell, R., and Bonhomme, S. (2017). Earnings and consumption dynamics: A nonlinear panel data framework. *Econometrica*, 85(3):693–734.

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.

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.

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.

Bueren, J. (2018). Long-term care needs: Implication for savings, welfare and public policy. *Wroking Paper*.

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.

Crossley, T. F. and Kennedy, S. (2002). The reliability of self-assessed health status. *Journal of Health Economics*, 21(4):643 – 658.

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

De Nardi, M., Fella, G., and Pardo, G. P. (2018). Nonlinear household earnings dynamics, self-insurance, and welfare. (24326).

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.

French, E. and Jones, J. B. (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, 79(3):693–732.

Gorea, D. and Midrigan, V. (2017). Liquidity constraints in the u.s. housing market. Working Paper 23345, National Bureau of Economic Research.

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2015). What do data on millions of u.s. workers reveal about life-cycle earnings risk? (20913).

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.

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.

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.

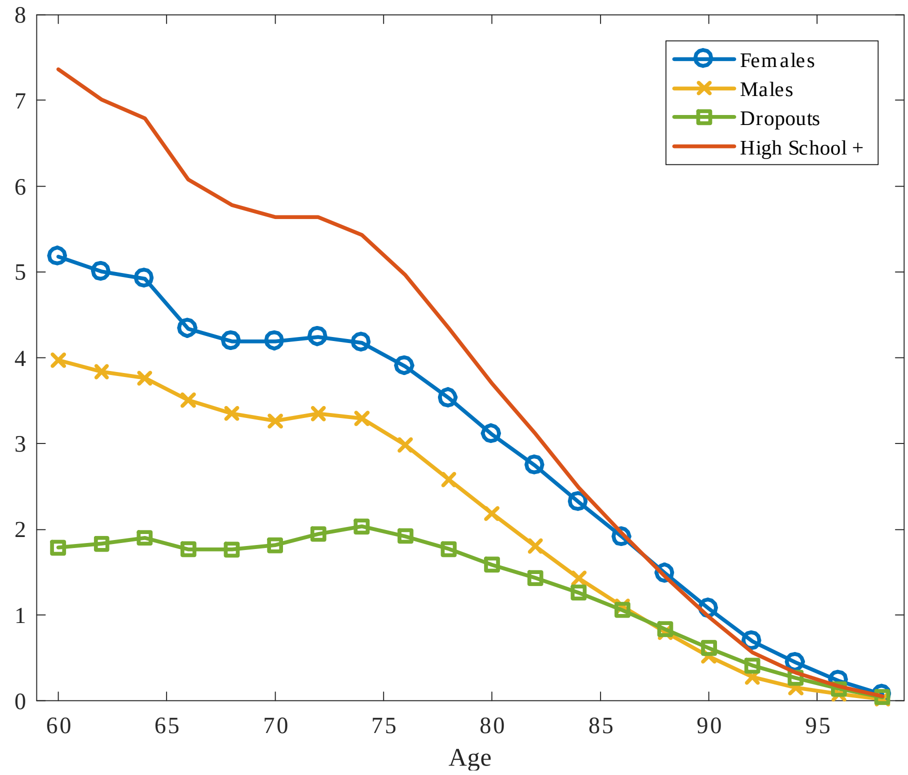
Van der Klaauw, W. and Wolpin, K. I. (2008). Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, 145(1):21 – 42. The use of econometrics in informing public policy makers.

Williams, B. C., Fries, B. E., Foley, W. J., Schneider, D., and Gavazzi, M. (1994). Activities of daily living and costs in nursing homes. *Health Care Financing Review*, 15(4):117.

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.

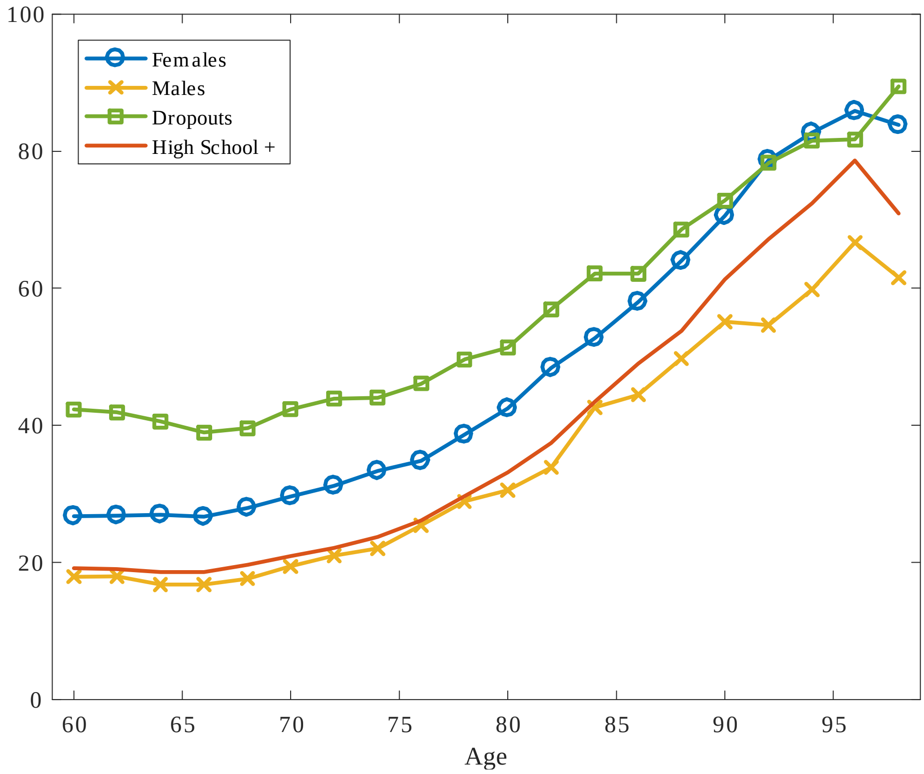
Figures

Figure 1: Share of interviewees by age



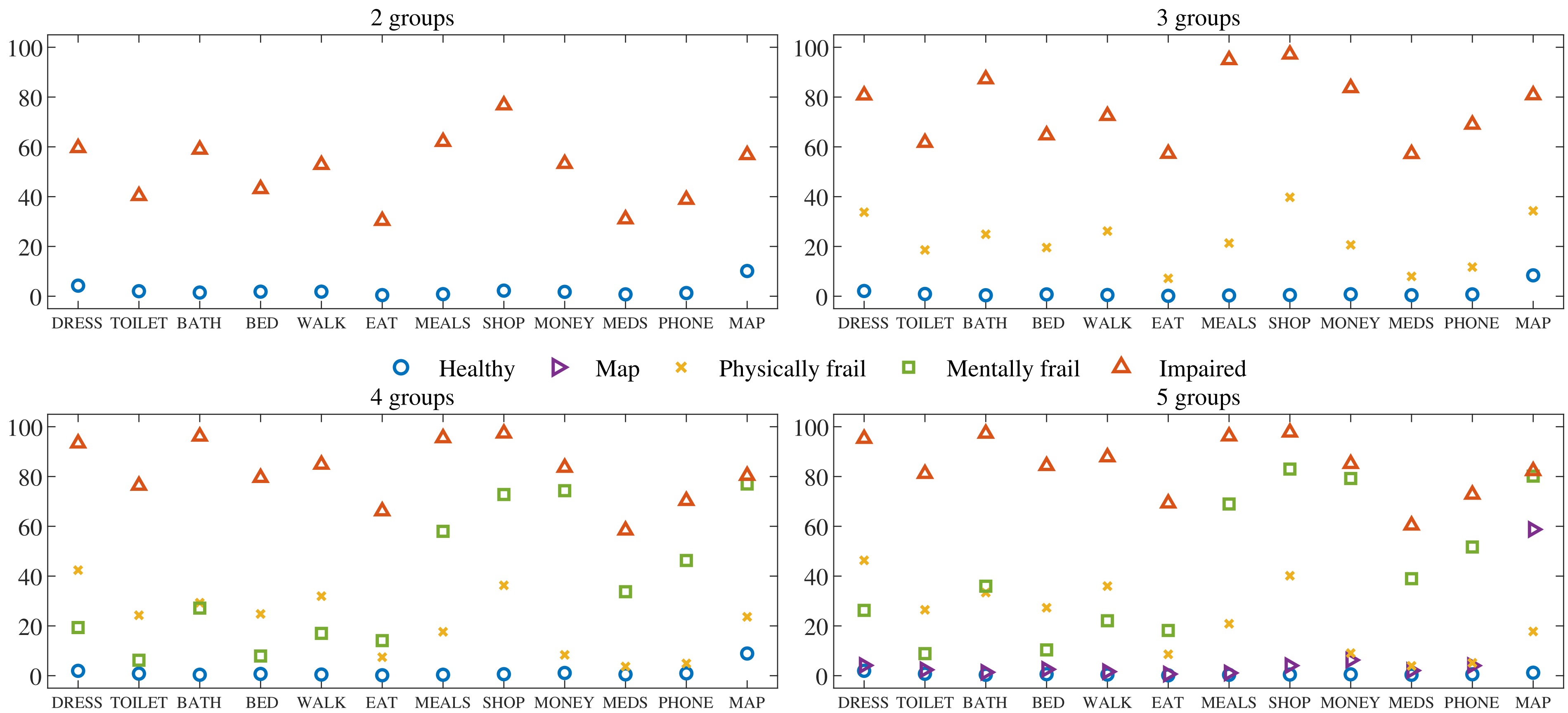
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing ($< 0.1\%$ of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. The y-axis is measured in percentage points and the x-axis in years.

Figure 2: Share of interviewees reporting at least one difficulty with an I-ADL by age



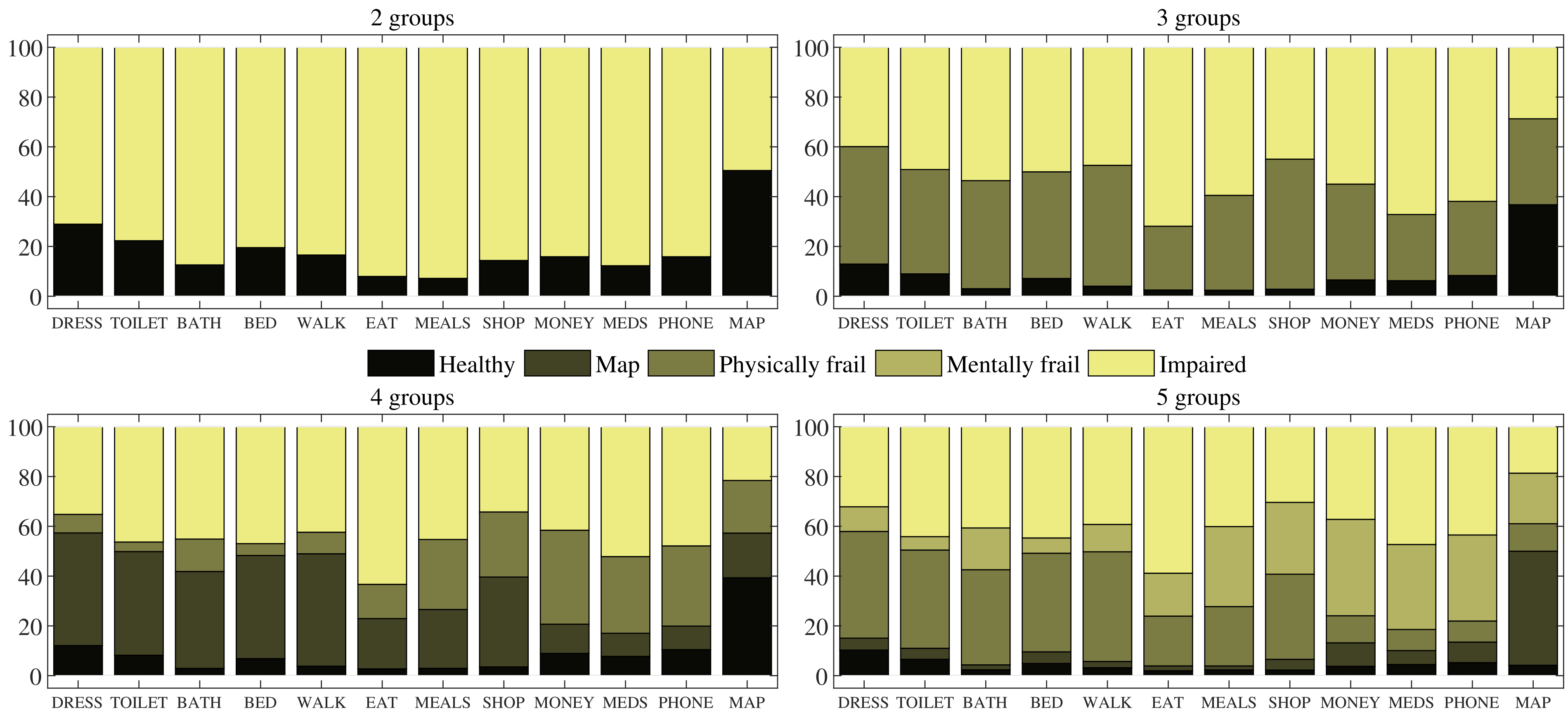
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals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

Figure 3: Probability of reporting a difficulty with a given I-ADL by health group



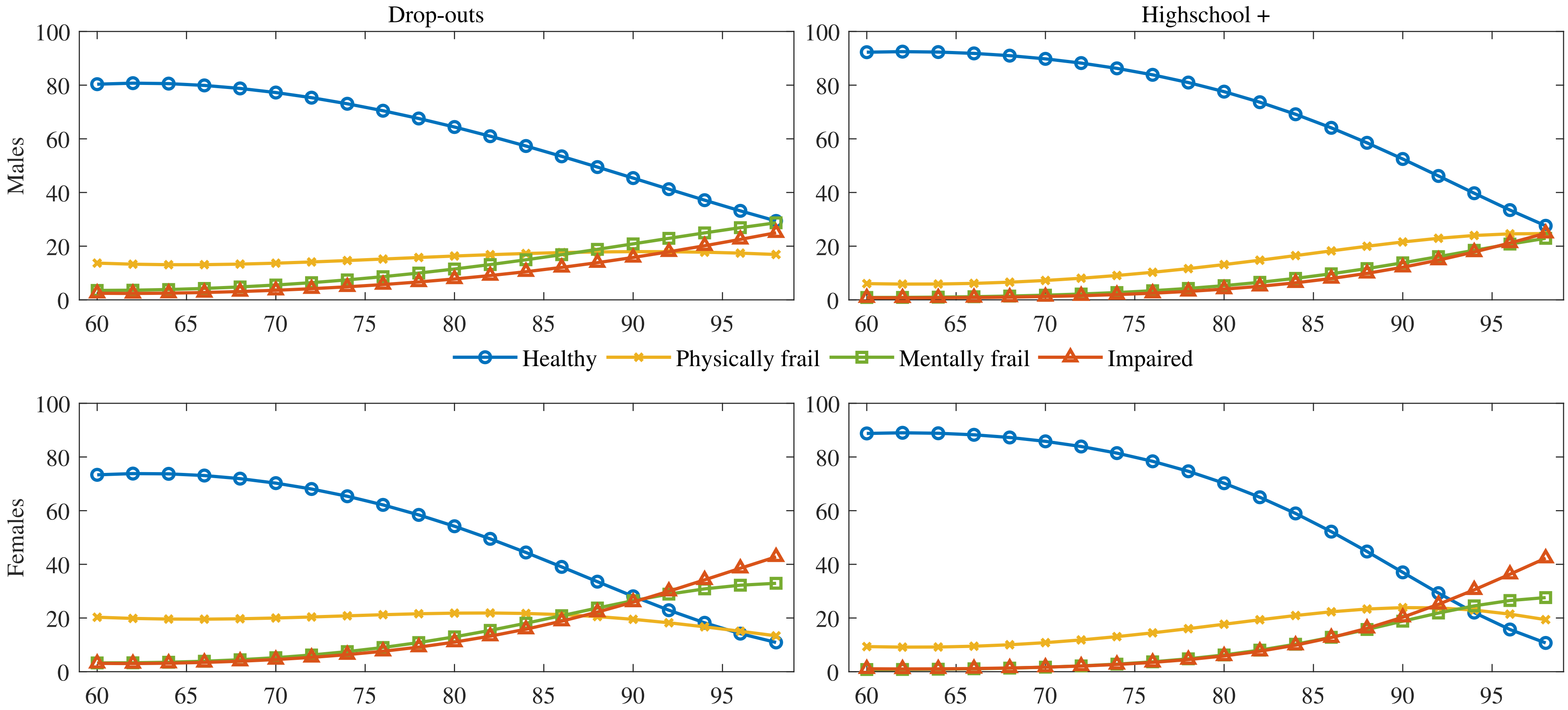
Notes: ADLs: Some difficulty with dressing (DRESS), using the toilet (TOILET), bathing (shower, BATH), getting in or out of bed (BED), to walk across a room (WALK) and eating (EAT). IADLs: Some difficulty with preparing hot meal (MEALS), shopping for groceries (SHOP), managing money (MONEY), taking medications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

Figure 4: Probability of belonging to a given health group by I-ADL



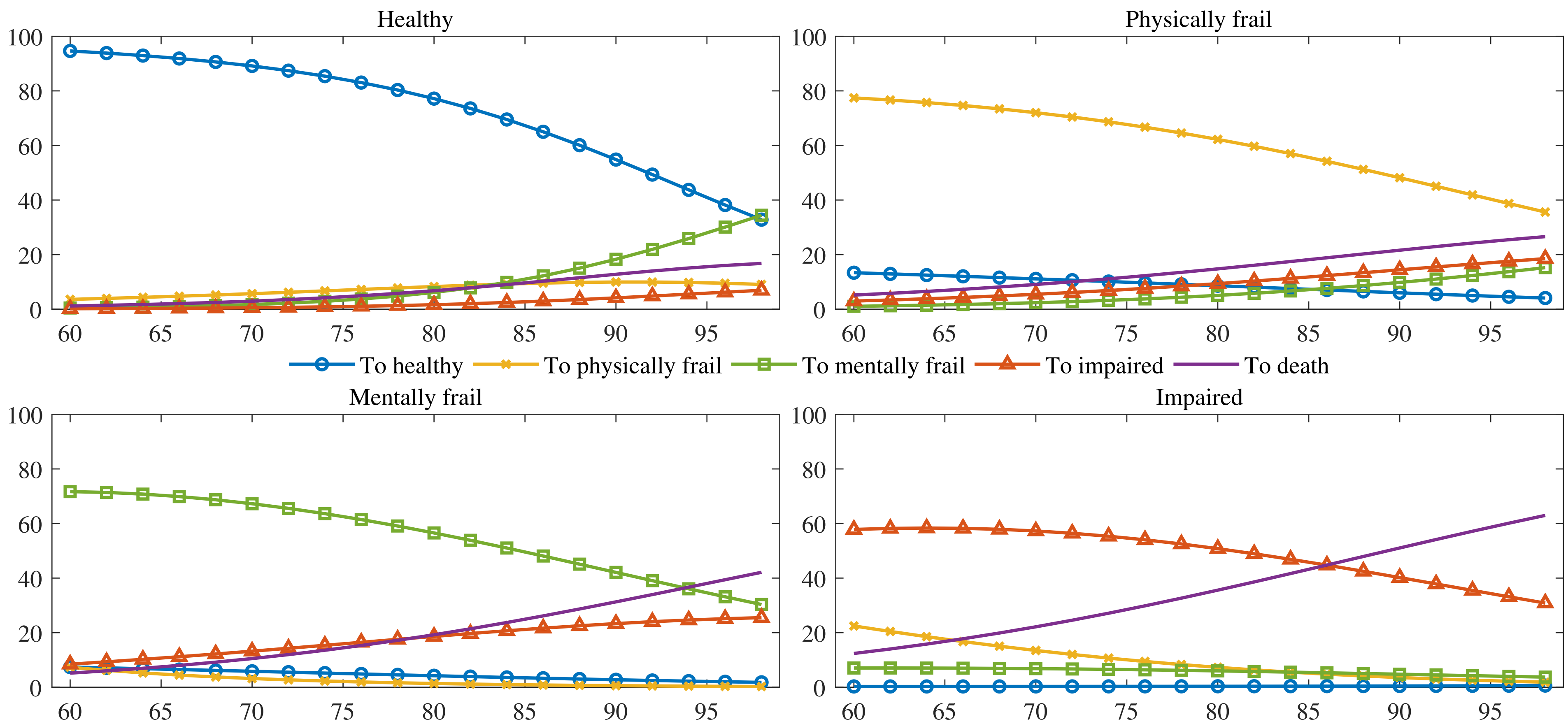
Notes: ADLs: Some difficulty with dressing (DRESS), using the toilet (TOILET), bathing (shower, BATH), getting in or out of bed (BED), to walk across a room (WALK) and eating (EAT). IADLs: Some difficulty with preparing hot meal (MEALS), shopping for groceries (SHOP), managing money (MONEY), taking medications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

Figure 5: Share of individuals in each group conditional on being alive by education and gender as individuals age



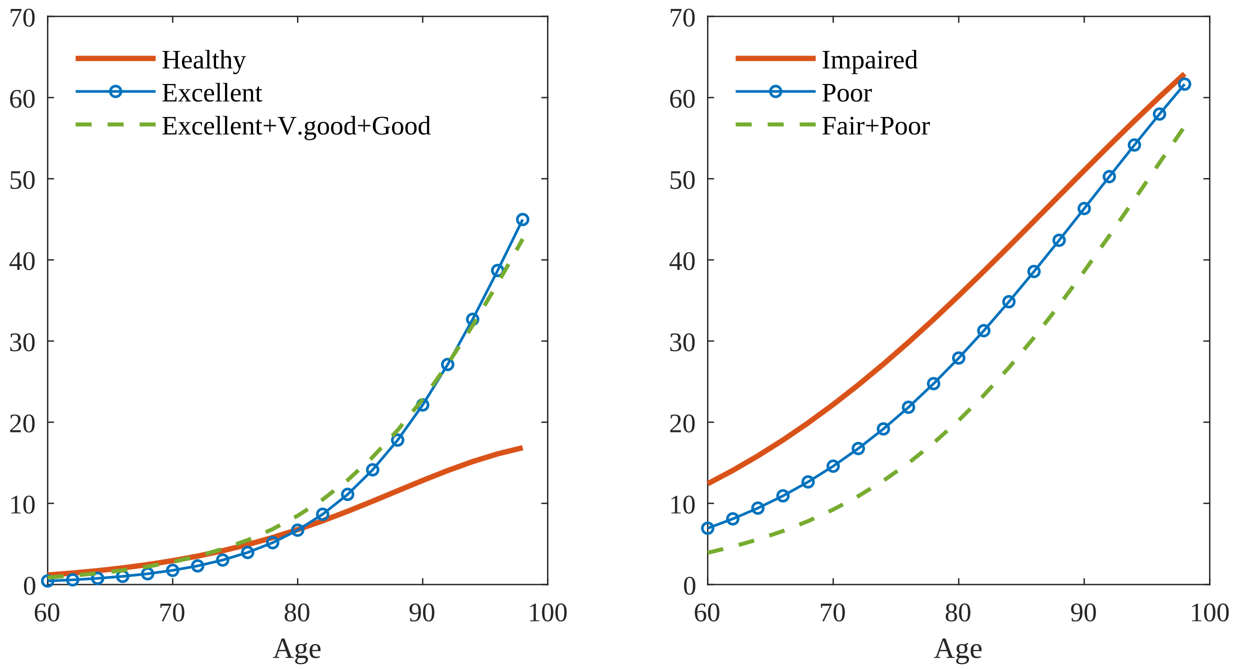
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals 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 6: Transitions by group as individuals age



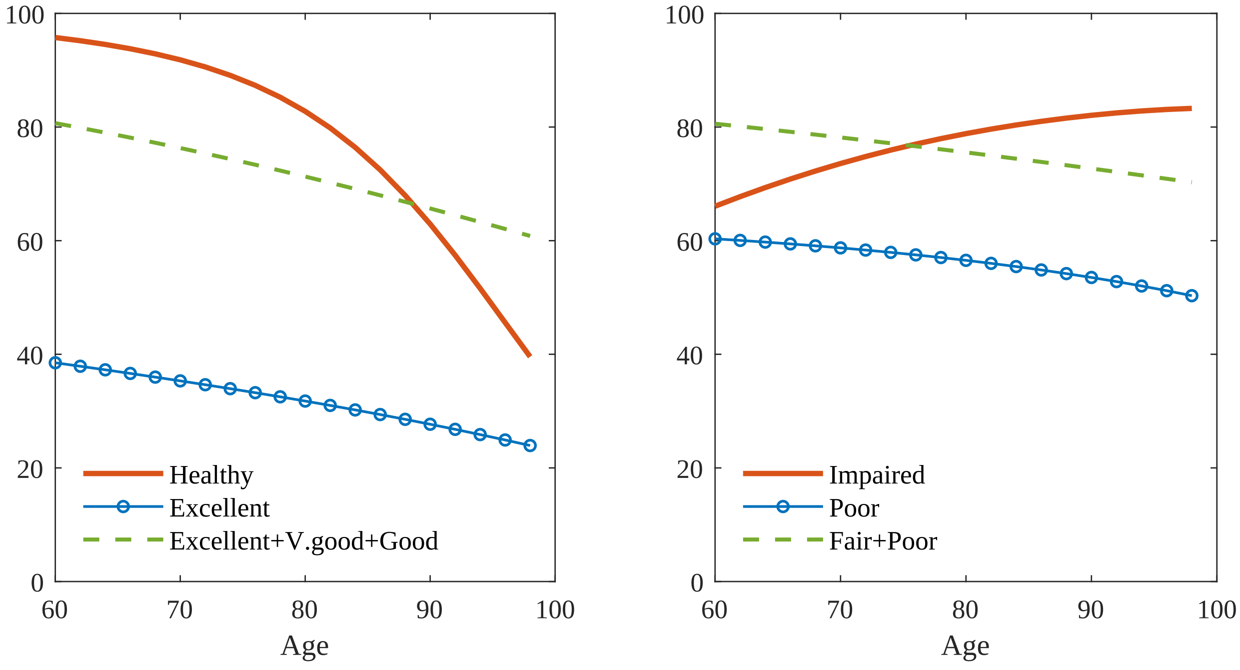
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals 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. This graph corresponds to female dropouts but it is similar if we look at other socio-economic groups (see Supplemental Material)

Figure 7: Transition to death



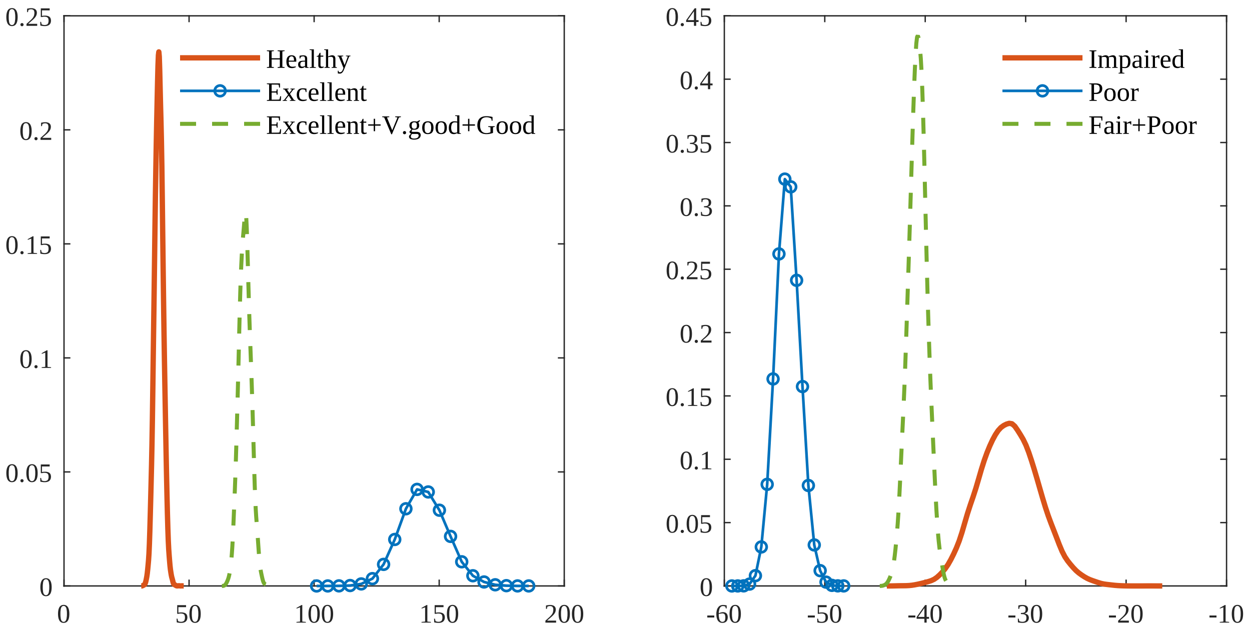
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. The 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 (see Supplemental Material)

Figure 8: Persistence of health status



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 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 (see Supplemental Material)

Figure 9: 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 units of the x-axis are percentage points.

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	17.6	22.9	71.1	57.7	21.4
		2 groups				
Healthy	89.1	8.9	14.3	70.4	56.5	18.9
Impaired	10.9	88.3	92.9	76.4	67.6	41.9
		3 groups				
Healthy	81.4	4.4	10.5	70.1	55.8	17.6
Physically frail	14.3	69.1	70.2	73.9	65.8	35.6
Impaired	4.3	96.5	99.9	79.6	67.7	45.6
		4 groups				
Healthy	82.0	4.1	11.4	70.1	55.8	17.9
Physically frail	11.5	79.7	61.5	73.0	66.4	32.5
Mentally frail	3.2	54.9	99.7	78.4	64.2	47.1
Impaired	3.3	100.0	99.9	79.5	67.8	45.0
		5 groups				
Healthy	71.3	3.9	3.6	70.0	52.1	14.7
Map	12.7	11.9	65.8	71.5	79.9	39.9
Physically frail	10.0	83.2	59.6	73.0	64.6	30.4
Mentally frail	3.1	65.5	99.8	78.6	64.4	47.4
Impaired	2.9	100.0	99.9	79.6	67.9	45.1

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 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	13.8		4.3		1.7		1.5		21.3
	(13.5, 14.1)		(4.1, 4.5)		(1.6, 1.8)		(1.4, 1.6)		(21.0, 21.7)
High school	19.0		3.3		1.1		1.1		24.5
	(18.7, 19.2)		(3.1, 3.4)		(1.0, 1.2)		(1.1, 1.2)		(24.2, 24.7)
				Males					
Dropouts	13.6		2.5		1.2		0.8		18.1
	(13.2, 13.9)		(2.4, 2.7)		(1.1, 1.3)		(0.7, 0.9)		(17.7, 18.5)
High school	18.1		1.9		0.7		0.5		21.2
	(17.8, 18.3)		(1.8, 2.0)		(0.6, 0.7)		(0.5, 0.6)		(20.9,21.4)

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 95% high-density intervals. 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
4 groups (mode)								
Healthy	78.8	2,311	0.2	4.8	4.0	4.2	12.5	3.7
Physically frail	13.0	3,592	1.7	20.8	14.0	84.2	69.7	55.9
Mentally frail	4.3	5,008	9.2	25.4	22.5	53.3	100.0	99.5
Impaired	3.9	10,062	33.8	51.5	41.7	100.0	100.0	99.9
4 groups (probabilities)								
Healthy	78.3	2,310	0.2	4.8	3.4	4.3	11.3	3.6
Physically frail	13.3	3,565	1.6	20.1	12.0	80.1	62.7	53.8
Mentally frail	4.4	4,908	9.0	24.4	19.4	54.2	99.5	98.5
Impaired	3.9	10,043	33.6	42.3	35.3	100.0	99.9	99.9

Notes: Results reported in percentage points, except for OOP med 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
4-I-ADL	3.2	2.5	16.2	13.8	12.3	9.1	11.9
4 groups (mode)	4.4	3.1	26.6	18.9	13.1	9.7	13.5
4 groups (pr.)	4.6	3.3	27.8	19.9	13.7	10.2	14.1
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 reported in percentage points. Numbers correspond to 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.