Endogenous Health Groups and Heterogeneous Dynamics of the Elderly*

Dante Amengual
CEMFI
<amengual@cemfi.es>

Jesús Bueren
CEMFI
<jesus.bueren@cemfi.edu.es>

Julio A. Crego
Tilburg University
<jacrego@uvt.nl>

June 2017

Abstract

Health dynamics and its associated medical and care costs have been identified by the macro literature as a major concern of the elderly. Due to its multidimensionality, however, a difficult task faced by researchers is to summarize health parsimoniously into a single state variable. We propose a panel Markov switching model to identify patterns of health heterogeneity where individuals can move across health groups as they age. To estimate the model, we use Markov chain Monte Carlo techniques to exploit information from both the cross-sectional and time series dimensions. We identify health groups for individuals in the Health and Retirement Survey for the US. Results show that there exists four clearly differentiated groups depending on individual's physical and mental disabilities. Furthermore, we show that health groups outperform other measures of health commonly used in the literature at explaining the variance in the use of nursing homes, home health care, out of pocket medical expenses and predicted mortality.

*We would like to thank Manuel Arellano, Nezih Guner, Pedro Mira, Josep Pijoan and Enrique Sentana for their comments and suggestions. We have also benefitted from the feedback provided by seminar participants at CEMFI, as well as audiences at the VII Workshop in Time Series Econometrics (Zaragoza) 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.

1 Introduction

Longevity increase and health decline have recently drawn substantial attention as baby boomers claim pension benefits and put pressure over the health care system. As a result, panel data sets based on detailed surveys, which contain a wide array of variables about different aspects of elderly's health, are nowadays available.[1] Despite the richness of the data, researchers often need to rely on a discrete measure that summarizes most of the information about health in the data. For instance, structural economic models aimed to analyze the effects of survival risk, medical and long-term care (LTC) costs, and more generally welfare, involve health as a state variable. In those situations, researchers typically undertake an ad-hoc decision over which health variable to choose to keep the state space feasible. In that regard, there does not exist a consensus as many alternative classifications are adopted in the literature. Examples of the latter include Ameriks et al. (2015) who combine two groups of self-reported health with a variable which reflects if the individual reports a difficulty with the so-called activities of daily living; De Nardi et al. (2016) who focus on two different groups of self-assessed health, together with a nursing home residency status; and Barczyk and Kredler (2017) who classify individuals as disabled when they receive more than 90 monthly hours of care.

In this paper we propose a parsimonious health classification which exploits both the cross-sectional and the time-series dimension of the Health and Retirement Study (HRS). Specifically, for each individual and point in time, we consider twelve dummy variables that characterize her difficulty with Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) in order to extract her latent health status. Therefore, we reduce the dimensionality of the data from 4,096 potential groups (2^{12}) into a small but nonetheless representative number of latent health categories. To classify individuals, we do not impose any restriction between parameters across I-ADLs, thus activities can differ in their importance.[2] Additionally, we explicitly model health and survival dynamics as a hidden Markov chain, allowing for heterogeneity across age, gender and education. For this purpose, we assume that transitions across health groups are logistic functions of the aforementioned attributes whose parameters change depending on the current health status.

[1] Among the most widely used are the Health and Retirement Study in the U.S, the English Longitudinal Study of Aging, and the Survey of Health, Aging and Retirement in Europe.

[2] 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.

This modelling approach presents three desirable features. First, it considers the classification of individuals, and groups dynamics jointly. In this way, the health groups classification is not based solely on the information of the current period but on all the observations including potential 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 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 to classify an individual's health even in the case of missing information as long as we have past or future information.

Our empirical strategy requires the estimation of thousands of hidden Markov chains, one per individual, together with hundreds of parameters. In addition, since respondents eventually die, we deal with an unbalanced panel. As it is common with such models, in which the state variables are latent and non-Gaussian, the likelihood function may not be available or its computation could be burdensome. 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 Bernoulli processes through a Metropolis step; and likewise, the parameters ruling the dynamics. Then, conditional on these parameters we obtain, for each individual, a realization of the latent health group using Kim (1994)'s smoother algorithm.

Four groups, in which individuals are classified as physically frail, mentally frail, impaired or healthy, represent individuals' health status suitably. The impaired have both types of limitations, physical and cognitive, while the healthy have no or light difficulties with I-ADLs. 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, 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 and less than 5% to the mentally frail. 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. Additionally, and in line with Brown (2002) and Meara et al. (2008), we find a large educational gradient in life expectancy. Specifically, high-school graduates live three years more than dropouts. Likewise, education has a health protective effect, the most educated individuals show the lowest probabilities of moving to worse health groups, as in Pijoan-Mas and Ríos-Rull (2014). Moreover, despite living longer, more educated individuals spend less time impaired. Precisely, our results reveal the existence of a morbidity gradient as high-school graduates live on average around 30% less time physically frail, mentally frail or impaired. Aside from education, current health status is an important source of heterogeneity because of the groups' persistence. For instance, a respondent who is impaired at age 75 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 LTC needs.

We then compare access to medical and care services across health groups based on the estimated probabilities. On average, impaired (healthy) individuals spend around \$11,189 (\$2,322) per year in out-of-pocket medical spending. The use of LTC services also presents large differences across the physically and mentally frail. While 10% of the individuals mentally frail live in a nursing home at the time of the interview, only 1% of the physically frail do so. Regarding impaired individuals, 38% of them live in a nursing home, which according to Kopecky and Koreshkova (2014) is an important driver of savings.

Finally, we contrast our estimated health groups with other commonly used health classifications, namely, five different levels of self-reported health (Pijoan-Mas and Ríos-Rull (2014)) and whether the individual reports difficulty with any ADL (Bohacek et al. (2015)).[3] To do so, we consider three main variables associated to health-related spending, particularly, out-of-pocket medical expenditures, and indicators of residing in a nursing home and receiving care. Our four groups classification generates more differentiated groups without a substantial increase in the within group heterogeneity. Furthermore, our grouping method explains about three times more variance than self-reported health and twice as much as the use of an ADL indicator.

Our 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

[3] De Nardi et al. (2010), Kopecky and Koreshkova (2014), Dobrescu (2015), and De Nardi et al. (2016) use a subset of the self-reported health measure.

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 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.[4]

Lastly, our paper adds to the literature that compares self-reported versus objective measures of health. Our results suggest that self-reported health measures do not provide a good representation of individuals' health. In particular, respondents tend to answer good health too often, therefore health groups such as excellent or very good may be related to other considerations rather than health. Actually, and consistent with this interpretation, self-reported status is much less persistent than our measure. Additionally, we find suggestive evidence of benchmarking as older individuals that report excellent health are very similar to those reporting good health in terms of life expectancy. Nevertheless, self-reported health is better at predicting most of chronic conditions like cancer or diabetes which indicates individuals may overweight on past pathologies. These results are in line with several previous papers focused on the labor market (see Currie and Madrian (1999) for a survey).[5]

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

[4] A clear example is Braun et al. (2017) who needs to split a frailty index into five quintiles to introduce health in their insurance demand model.

[5] 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.

2 HRS and I-ADLs

Our data comes from the RAND HRS dataset which comprises a cleaned version of the Health and Retirement Survey.[6] This survey is a US representative biennial panel of households of age 50 and above conducted by the University of Michigan.[7] This dataset is widely used in the literature and is the basis of most of the papers mentioned in the introduction (see De Nardi et al. (2016) and references therein). It covers the periods from 1992 to 2014 and contains subjective and objective indicators of health, as well as demographic characteristics. Besides, the HRS exit interview records the death of the individual and includes the answers from a proxy informant.

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. To focus on health needs we select individuals over 60 years old. Moreover, 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 six waves (12 years).

Figure 1 shows that, 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.

Regarding health status, we focus on the ability of individuals to perform Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs).[8] Katz et al. (1963) proposed ADLs as a measure of how independent a patient is, and consequently they included very basic activities such as if they can walk or dress. IADLs, in contrast, consist of activities more closely related with cognition. Examples of the latter include the possibility of using a phone or controlling her medication. In the HRS, people are surveyed about whether they have any difficulty to perform these types of basic tasks.

[6] 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).

[7] The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

[8] 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).

Example 1: Actual question and possible answers

```
G016
Q2725
E72.  Because of a health or memory problem do you have any difficulty with
      walking across a room?

      1. YES   5. NO   6. CAN'T DO   7. DON'T DO   8. DK   9. RF
```

Example 1 provides an illustration of one of the questions and the 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. In total, we observe 12 binary variables, denoted as I-ADLs, which includes six ADLs and six IADLs.

Table 1 defines the activities included in the HRS and provides the proportion of observations in which an individual declares to have difficulties realizing them. The most common ADL is not being capable of dressing whereas eating is the ADL that present less difficulties. The heterogeneity in IADLs is more extreme as more than 15% of respondents struggle reading a map but only 5% claim to face problems when taking medications. 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. Furthermore, 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. Meanwhile, 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 among these groups 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.

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 first panel of Table 2 we also compare this measure with the answers related to ADLs and IADLs. 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.

The study contains as well questions about chronic conditions such as if someone has ever been diagnosed with cancer or a chronic lung disease. The remaining panels of Table 2 describe the sample average of each of the I-ADL variables conditional on whether the individual has one these conditions or not.[9] It indicates that while all these diseases predict difficulties on the daily living, having had a stroke or psychiatric problems are the best predictors as the proportion of patients reporting difficulties is two and six times the ones who do not, respectively. On the other hand, a diagnosis of high blood pressure (HBP) almost does not change the probability of struggling with I-ADLs. Likewise, the relationship between HBP, diabetes, cancer and arthritis and IADLs is weak since this conditions are more related to physical problems than mental problems.

To sum up, I-ADLs are coherent with other health measures such as self-reported health or chronic conditions. Nevertheless, they aim to measure the degree of dependence for a given individual which relates to his need of LTC, and other mostly uninsured medical expenses. Furthermore, these expenses are quantitatively important as formal LTC related expenses accounted for \$310 billion in 2013 which add to the informal ones such as the necessity of family members to quit their jobs to take care of their elderly relatives. For those reasons, and since our aim is to characterize the financial health risk of individuals, we use difficulties with I-ADLs as our main input to classify respondents into groups. Although our model can incorporate in addition to I-ADLs, SRH, chronic diseases as any other variable in the HRS, we restrict to I-ADLs as reducing the set of variables eases the interpretation of the groups.[10] Additionally, by excluding these other variables, we can use them to compare the performance of our model against other alternatives.

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, K dummy variables corresponding to each I-ADL $(x_{1,i,t}, x_{2,i,t}, \ldots, x_{K,i,t})$ are observed across time provided

[9] Chronic conditions classification as follows. HBP: high blood pressure or hypertension; Diab.: diabetes or high blood sugar; Cancer: cancer or a malignant tumor of any kind except skin cancer; Lung: chronic lung disease except asthma such as chronic bronchitis or emphysema; Heart: heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems; Stroke: stroke or transient ischemic attack; Psyche: emotional, nervous, or psychiatric problems; Arthr.: arthritis or rheumatism.

[10] Due to the nature of I-ADLs, we identify health groups in terms of dependence.

the individual is alive. 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, current health cluster is independent of previous health clusters except the most recent one (Markov first-order property). In addition to transit 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

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

where $\boldsymbol{\mu}_g = (\mu_{1,g}, \mu_{2,g}, \ldots, \mu_{K,g})'$. 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 $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ 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. $p(\boldsymbol{\beta}^{(m)}|\boldsymbol{\mu}^{(m-1)}, \mathbf{H}^{(m-1)}, \mathbf{X})$: sampling the transition parameters (Metropolis).
2. $p(\boldsymbol{\mu}^{(m)}|\boldsymbol{\beta}^{(m)}, \mathbf{H}^{(m-1)}, \mathbf{X})$: sampling the Bernoulli mixture parameters (Metropolis).
3. $p(\mathbf{h}_i^{(m)}|\boldsymbol{\beta}^{(m)}, \boldsymbol{\mu}^{(m)}, \mathbf{X})$: sampling the latent health indicator for each $i = 1, ..., N$ using the Kim (1994)'s smoother.

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. Appendix A provides an analysis of the model convergence.

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 $p(h_{i,T} = g|\boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{X})$ for all $g \in \mathcal{H}$.[11]

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

3. Similarly, we sample $h_{i,t}$ conditional on $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, $\mathbf{X}$ and $h_{i,t+1}$, using the following result:

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

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.

3.1.2 Sampling the transition probabilities and the Bernoulli parameters

In this step we sample from the posterior of $(\boldsymbol{\mu}, \boldsymbol{\beta})$ conditional on the health groups, $\mathbf{H}$, and the data, $\mathbf{X}$.

Regrading priors, we consider a uniform on $[0, 1]$ for the elements of $\boldsymbol{\mu}$ and a diffuse Gaussian prior centered at $\mathbf{0}$ and covariance matrix $100 \cdot \mathbf{I}$ for $\boldsymbol{\beta}$. Hence, the posterior of the parameters governing the health dynamics and the one driving the Bernoulli distributions are conditionally independent. Precisely, their posterior distributions are given by

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

and

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

To form a complete likelihood, we need to know the unconditional distribution of $h_{i,1}$ for each i, $p(h_{i,1}|\boldsymbol{\beta})$. 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).

3.1.3 Starting the algorithm

To obtain the starting set of parameters $\boldsymbol{\mu}^0$ and $\boldsymbol{\beta}^0$ for the algorithm, we sample from an approximate model in two steps. First we obtain $\boldsymbol{\mu}^0$ as the mode of the posterior described in

[11] To start Hamilton's algorithm we need to provide the state at time 0. To do that we assume that individuals start from the stationary distribution given by the parameters of the first transition (from 60 to 62 years old).

equation (1) under the assumption that $h_{i,t}$ are independent across both dimensions.[12] Second, we use the same model to simulate $h_{i,t}$ from the posterior probability $p(h_{i,t}|\boldsymbol{\mu}, \mathbf{x}_{i,t})$. Given a sample of health groups, we get the mode of the posterior of $\boldsymbol{\beta}$, $\boldsymbol{\beta}^0$, under the assumption that groups follow the same multinomial logit specification as in the baseline model.

3.2 Obtaining moments

In most applications, and 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 enhance 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 individuals are classified into groups and then, how these evolve across age. Given a number of groups, their characteristics are such that individuals in a given cluster present a similar join distribution, and, at the same time, capturing the heterogeneity across different groups. It should be noticed that the join distribution not only refers to difficulties with I-ADLs at a point in time, but also to the entire history, hence capturing 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.

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

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.

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). Besides the variability across groups, we find large differences in the probabilities within groups which suggest that I-ADLs differ in their importance. For example, as regard 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.

Figure 3 reports the difficulties of individuals in a given cluster but it is silent about which I-ADLs are more relevant to classify 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.

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 3 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 6 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 with around 4% probability of reporting troubles with ADLs. 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%).

The second column indicates that most of individuals are classified as *healthy* whereas around 11% belong to the *physically frail* group and around 3% present *mental frailty*, and the same fraction are *impaired.*

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 6 shows they are indeed on average ten 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.

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 85, *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. This patterns are very similar across education and gender, although the initial composition of individuals varies with demographic characteristics.

Since mortality and health deterioration vary by education group, dropouts and high-school graduates encounter a very distinct health risk. Table 4 shows the expected time an individual at age 60 lives in each health group. Even if the more educated elderly live longer, they spend less years as *impaired* and *frail*, which suggests that richer individuals face lower health risk. For instance, in the case of males, dropouts stay 0.3 years more in the *impaired* state that translates in \$3,000 more spent in OOP medical expenses.[13] 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 consumption smoothing purposes.

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

[13] See Table 5.

health status. For example, a *healthy* elderly owns a very low probability to become impaired, thus a low health risk. 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, with the exception of *physically frail* at early ages, health is more likely to worsen than to improve.

In terms of mortality, results are in line with our interpretation of the groups. *Mentally frail* and *impaired* face the highest mortality risk. Moreover, individuals in these groups almost never transit to *physically frail* or *healthy*. Meanwhile, *healthy* elderly die very rarely, although it becomes more likely with age.

5 Comparison with alternative indices

Although the best health classification depends on the researcher's objective, comparing different methods sheds some light on when to use them. In particular, we compare our health classification against three of the most commonly classifications used in the literature: self-reported health, whether the individuals reports difficulty with an ADL or not, and if the respondent needs more than three hours of medical care per day.[14] In addition, we also consider the cartesian product of whether the individuals reports difficulty with i) at least one ADL and ii) IADLs (excluding MAP) as an unsophisticated proxy of our endogenous classification.

Self-reported health relies on individuals subjectively assessing their health status which creates several patterns that might not be related with health spending. Table 5 shows that the most common answer is *good* which corresponds to the middle group. Since our measures of health costs are very similar across the three healthier groups, we conceive reporting *excellent* or *very good* as mainly driven by other characteristics such as optimism and benchmarking. Nonetheless, we observe that the worse they feel, the more they spend in medical services, the more likely they receive home care or they live in a nursing home, especially the *fair* and the *poor*. Additionally, the probability of facing difficulties with an ADL or IADL is also higher in worse health groups.

Grouping individuals according if they have an ADL or not is similar to our approach, specifically to identify the *healthy* respondents, thus the proportions of *healthy* and No-ADL almost

[14] For instance, as for self-reported health, Pijoan-Mas and Ríos-Rull (2014) use the five different levels, while De Nardi et al. (2010), Kopecky and Koreshkova (2014), Dobrescu (2015), and De Nardi et al. (2016) use a subset of that measure; on the other hand, Bohacek et al. (2015) consider whether the individual reports difficulty with any ADL;.finally, Barczyk and Kredler (2017) look at hours of medical care.

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 eating are usually more dependent than those who walk uneasily.

Using hours of care to split individuals into two groups, in contrast, leads to an even more unbalanced division, with only 2.5% of elderly in the worst health group. This classification, nevertheless, presents an endogeneity issue since hours of care are indeed a decision taken by the respondent.

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 that regard, the group of respondents reporting no difficulty (difficulties with both types of I-ADLs), a natural substitute for the *healthy* (*impaired*), is underrepresented (overrepresented).

5.1 A horse race

Table 5 presents the distribution of individuals across groups, indicators of their difficulties with I-ADLs, and three variables related to the financial risk due to health: out-of-pocket (OOP) medical expenditures, if they receive home-care and if they reside 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. Likewise, received home care equals 1 if a medically-trained person has come to the respondent home to help her, and nursing home resident takes value 1 for those individuals who live in a nursing home.

Even if the five aforementioned classifications present an increasing pattern in these three variables as health deteriorates, their increments are very different across methods. For instance, using our methodology, the average difference in terms of OOP between *healthy* and *impaired* elderly is $8,867. According to self-reported health, however, an individual belonging to the worst group only expends $3,609 more than one in the best group. Similarly, the fact that you report an ADL implies that your average OOP medical spending is $2,806 higher. We find a similar figure when we classify individuals according to their care needs. Not surprisingly, 4-I-ADL is the closest to our classification. As for the intermediate groups, they are again more similar in the case of either self-reported health or 4-I-ADL as their increment in spending is below \$1,000 except from *frail* to *poor*, compared to \$1,296 which is the minimum difference between our groups.

Regarding the probability of residing in a nursing home, the same pattern also arises as the difference between the best and worst of our health groups at least duplicates the same difference using the alternative methods. The same is true for home care when we look at self-reported health or struggling with at least one ADL; our four groups outranks 4-I-ADL, although mildly. As expected, by construction the grouping based on hours of care is the most accurate at explaining required care.

It is important to consider the incidence of chronic conditions as they are suffered by 88% of individuals in our sample. Table 6 details the probability of having been diagnosed with one chronic condition by classification method and pathology. In general self-reported health generates more differentiated groups which suggests that respondents might overweight their previous conditions when assessing their health status. STROKE is the only exception: it is more important to categorize individuals for our classification than for self-reported health. It should be noticed that this disease is the second cause of death worldwide, and only 50% of patients diagnosed with a stroke survive more than 5 years (Ingall, 2004). Additionally, and in line with our interpretation of the health groups, the *mentally frail* individuals are more likely to have been diagnosed with a stroke or psychiatric problems whereas those *psychically frail* tend to have arthritis in their clinical history.

Overall, our groups are more distant than the alternative ones regarding medical spending and I-ADLs, but not concerning chronic diseases. Distance in terms of the average is a desirable feature of a grouping method, however, it is just one side of the coin. The other side is the within group similarity as we would like individuals inside a group to be alike. To formalize this comparison, we compute the heterogeneity distance proposed by Kurczynski (1970):

$$H^2 = \frac{2}{Q-1} \sum_{g \in \mathcal{Q}} (\boldsymbol{\pi}_g - \overline{\boldsymbol{\pi}})' S^{-1} (\boldsymbol{\pi}_g - \overline{\boldsymbol{\pi}})$$

where $\mathcal{Q}$ is the set that includes all possible Q groups, π_g is a $D \times 1$ vector which contains the average each of the D variables considered in the comparison, S is the within-group variability and $\overline{\boldsymbol{\pi}}$ is the average of $\boldsymbol{\pi}_g$. The upper panel of Table 7 presents the distance measure using the previously considered variables, as well as ADLs and IADLs, as references. As expected, our classification improves the division of individuals according to every set of variables except chronic conditions. Moreover, self-reported health performs poorly across all measures we consider, and it is outranked by dividing the sample in those who have difficulties with an ADL and those who do not.

Sometimes, though, the researcher's concern might not be to classify individuals in homogeneous and 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, the second panel of Table 7 displays the R^2 of the following regression:

$$y_{i,t} = \mathbf{d}'_{i,t}\boldsymbol{\beta} + \varepsilon_{i,t}$$

where $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs, and $y_{i,t}$ is the variable used as reference. 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). With the exception of OOP medical expenses, and to ease the exposition, we average the resulting R^2 corresponding to different categories. Our classification method explains more variance than the alternative methods, except once again for the case of chronically conditions. In particular, our four groups explain three times more variance than self-reported health and twice as much as just considering if you have an ADL or not. As in the case of Kurczynski's distance, this last classification outperforms self-reported health in most categories.

The lower panel repeats the same exercise but using as dependent variable the residual of the following regression:

$$y_{i,t} = \mathbf{z}'_{i,t}\gamma + v_{i,t}$$

where $\mathbf{z}_{i,t}$ includes gender, age and education. The resulting R^2 has therefore the interpretation of the percentage of variance explained by the index of health on top of that explained by covariates. Results are very similar to the previous ones. If we do not consider chronic conditions, self-reported health is the worst classification whereas our method explains the highest percentage of variance. This ranking reverses in the case of chronic conditions though.

5.2 Dynamics: self-reported health versus endogenous classification

Besides comparing the classification of individuals, 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 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*.

There are two main risks associated with health transitions which increase the incentives to save. The first one is survival risk, if individuals were certain about their death day, the will consume everything but the amount they desire to leave as bequest. In reality, however, this day is not know, hence they have to save in case they live more than expected. The second one is health risk, individuals save more if they are very likely to enter a health status with high medical costs.

Figure 7 reports the median probability of dying and its 95% highest density interval for female dropouts. The left panel corresponds to the healthiest groups, whereas the right panel presents the results for the unhealthiest 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*. One possibility is that individuals compare themselves with relatives and friends of the same age to assess their health status. Hence respondents of age 65 and 90 have a different benchmark. On the other hand, age is not as important for the *healthy* group as mortality less than doubles between age 80 and 98. 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 suggest that this method is not good to predict mortality, specially at older ages. In addition, *impaired* individuals feature a higher death probability than those who assess themselves as in *poor* health.

Regarding health risk, the first important characteristic is the time individuals expect to be in the most unhealthy status. Figure 8 describes the distribution of this variable for an individual in the best health group who is 60 years old (left) or 80 years old (right). In both cases we observe that self-reported health imply a much higher mean duration, and a fatter right tail. Therefore, being *impaired* is more costly but it is short-lived; meanwhile reporting a *poor* health implies a lower cost but individuals spend more time in this state. If we compare the right and left panel, the differences in uncertainty increase as individuals age. In particular, the probability of living 10 years or more in the unhealthy group decreases for the *impaired* while it increases for poor health. The distribution changes as individual age mainly because mortality. As we show before self-reported health is not a good predictor of mortality, hence individuals with *poor* health might not die for several waves. Using our method, however, at old ages these respondents present a high likelihood of dying, thus they do not expect to be a long time in the worse state.

The second relevant element of health risk is persistence. If the process is not persistent, health today does not have an impact on individuals' behavior; at the same time, lower savings are needed to smooth the consumption that small period. Furthermore, persistence sheds some light on the type of health process. For instance we aim to create an indicator of dependence which is by definition persistent in contrast to others such as a flu or a sprained ankle. Figure 9 depicts the probability of remaining in the same cluster conditional on the cluster 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, the same pattern emerges; our classification displays more persistence. Moreover, as people age they tend to report an improve in their health status, in contrast according to our grouping method they are more likely to remain *impaired* which is closer to reality, and probably points to changes in their health benchmark.

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.

Barczyk, D. and Kredler, M. (2017). Evaluating long-term care policy options, taking the family seriously. *The Review of Economic Studies, Forthcoming.*

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.

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.

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.

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., 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.

Ingall, T. (2004). Stroke-incidence, mortality, morbidity and risk. *Journal of Insurance Medicine*, 36:143–152.

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. *JAMA*, 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.

Kurczynski, T. W. (1970). Generalized distance and discrete variables. *Biometrics*, 26(3):525–534.

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.

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.

Table 1: Sample summary statistics

Variable	Definition	# Obs	Yes (percentage)
	Activities of daily living (ADLs): Some difficulty...		
DRESS	Dressing	134,980	12.4
TOILET	Using the toilet	134,785	7.6
BATH	Bathing (shower)	134,949	10.0
BED	Getting in or out of bed	134,900	7.9
WALK	To walk across a room	134,913	9.4
EAT	Eating	134,908	4.9
	Instrumental activities of daily living (IADLs): Some difficulty...		
MEALS	Preparing hot meal	127,840	9.6
SHOP	Shopping for groceries	130,313	12.8
MONEY	Managing money	130,013	9.2
MEDS	Taking medications	131,264	5.3
PHONE	Using a phone	134,259	6.8
MAP	Using a map	117,200	15.7
	Some difficulties with...		
ADL	At least one ADL	134,366	21.1
IADL	At least one IADL	103,910	23.2
I-ADL	At least one ADL or IADL	103,663	29.6

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).

Table 2: Share of individuals reporting difficulties with I-ADLs by self-reported health and whether individual displays chronic conditions

Classification	DRESS	TOILET	BATH	BED	WALK	EAT	MEALS	SHOP	MONEY	MEDS	PHONE	MAP
Self-reported health (SRH)												
Excellent	2.2	1.0	1.6	1.0	1.1	0.8	1.8	2.2	2.5	1.2	1.6	6.5
Very good	3.5	2.1	2.3	1.4	1.9	1.0	2.4	3.1	3.1	1.5	2.2	8.7
Good	8.1	4.8	5.7	4.3	5.2	2.5	5.6	7.7	6.2	3.1	4.4	13.6
Fair	20.2	12.1	16.0	13.0	14.8	7.4	14.7	21.0	14.1	7.9	10.2	23.8
Poor	44.1	29.2	40.3	33.2	39.4	21.5	39.3	50.2	32.2	20.4	24.7	39.3
High blood pressure or hypertension (HBP)												
No	9.1	5.4	6.9	5.7	6.4	3.4	6.6	8.8	6.6	3.7	5.1	12.3
Yes	14.9	9.3	12.2	9.6	11.5	5.9	11.7	15.7	11.0	6.3	8.0	18.2
Diabetes or high blood sugar (Diab.)												
No	10.6	6.7	8.6	6.8	8.0	4.3	8.5	11.2	8.2	4.7	6.1	14.5
Yes	19.4	11.3	15.2	12.3	14.7	7.0	13.9	18.9	12.9	7.6	9.2	20.4
Cancer or a malignant tumor of any kind except skin cancer (Cancer)												
No	12.0	7.3	9.5	7.7	8.9	4.6	9.1	12.2	8.8	5.0	6.5	15.6
Yes	14.8	9.3	12.5	9.3	11.9	6.3	11.9	15.7	11.0	6.6	8.3	15.8
Chronic lung disease except asthma such as chronic bronchitis or emphysema (Lung)												
No	11.4	7.0	8.9	7.2	8.4	4.5	8.8	11.5	8.7	5.0	6.5	15.0
Yes	20.2	12.1	17.8	12.9	16.4	7.4	15.6	22.4	12.4	7.5	9.0	20.7
Heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems (Heart)												
No	9.8	5.9	7.4	6.1	6.9	3.6	7.1	9.6	7.0	3.9	5.1	14.4
Yes	19.4	12.2	16.6	12.6	15.8	8.0	16.3	21.4	14.7	8.9	11.0	19.0
Stroke or transient ischemic attack (Stroke)												
No	10.2	6.1	7.7	6.3	7.2	3.2	7.1	10.1	6.8	3.8	5.0	14.1
Yes	31.4	21.3	30.1	22.2	27.8	18.9	31.6	37.2	30.1	17.8	21.8	30.8
Emotional, nervous, or psychiatric problems (Psyche)												
No	10.2	6.0	7.7	6.0	7.5	3.6	7.3	10.0	6.8	3.7	5.2	13.2
Yes	25.7	17.4	23.2	19.1	20.6	12.4	23.3	29.6	23.3	14.3	16.0	31.3
Arthritis or rheumatism (Arthr.)												
No	6.1	3.2	5.2	3.7	4.4	2.8	5.8	7.2	6.2	3.5	4.8	11.5
Yes	16.4	10.4	13.0	10.6	12.5	6.2	11.9	16.3	11.0	6.4	8.0	18.4

Notes: RAND HRS Data; Sample from 1996 to 2014 (10 waves). Results reported in percentage points.

Table 3: Summary statistics for estimated health clusters

Group	Share	ADL	IADL	Age	Female	Dropout
			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). Results reported in percentage points. See Section 3 for details about the econometric model and the estimation procedure.

Table 4: Average duration by health group, education, and gender conditional on reaching 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). Results reported in years. See Section 3 for details about the econometric model and the estimation procedure.

Table 5: Long-term care needs

	Share	OOP med spending	Recieved h-care	Nurs-h resident	ADL > 0	IADL > 0	IADL > 0 w/o map
				Self-reported health			
Excellent	10.0	1,818	2.3	0.5	3.0	8.1	3.3
Very good	29.8	2,100	3.6	0.6	5.7	11.3	4.7
Good	32.6	2,738	6.9	1.1	13.3	19.1	10.2
Fair	19.4	3,639	13.2	2.7	31.7	37.4	25.6
Poor	8.1	5,427	27.3	8.8	62.6	64.4	55.9
				ADL: Yes/No			
No	82.4	2,355	5.0	0.3	0.0	14.5	6.1
Yes	17.6	5,161	24.4	9.0	100.0	62.1	54.3
				Care			
$<$ 3 hours	97.5	2,782	7.2	1.7	15.8	21.0	12.5
$>$ 3 hours	2.5	5,435	48.1	4.6	85.7	96.5	95.1
				4-I-ADL (i, j): ADL=i & IADL=j, IADL without MAP			
(0,0)	77.4	2,291	4.4	0.1	0.0	9.0	0.0
(1,0)	8.0	3,092	13.5	0.6	100.0	17.1	0.0
(0,1)	5.0	3,350	13.3	2.5	0.0	100.0	100.0
(1,1)	9.5	6,902	34.3	16.2	100.0	100.0	100.0
				4 groups			
Healthy	82.0	2,322	4.7	0.1	4.1	11.4	3.0
Physically frail	11.5	3,649	19.4	1.3	79.7	61.5	49.7
Mentally frail	3.2	4,974	23.9	9.7	54.9	99.7	98.0
Impaired	3.3	11,189	40.8	38.2	100.0	99.9	99.9

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). Results reported in percentage points, except for OOP med spending which is reported in US dollars. See Section 3 for details about the econometric model and the estimation procedure.

Table 6: Summary statistics for estimated health clusters

	Share	HBP	Diab.	Cancer	Lung	Heart	Stroke	Psyche	Arthr.
				Self-reported health					
Excellent	10.0	33.2	5.5	9.9	3.3	8.8	2.9	5.1	37.1
Very good	29.8	51.0	11.3	13.0	5.7	16.4	4.7	7.8	53.3
Good	32.6	61.5	21.6	15.7	10.7	27.6	7.9	12.2	64.3
Fair	19.4	67.8	30.2	19.1	18.3	39.3	13.7	20.6	72.6
Poor	8.1	72.0	35.9	24.0	29.8	51.4	24.3	33.5	78.4
				ADL: Yes/No					
No	82.4	55.2	17.5	14.8	9.6	23.5	6.3	10.5	56.5
Yes	17.6	68.6	30.0	19.5	20.3	41.0	21.3	28.0	82.3
				Care					
$<$ 3 hours	97.5	57.1	19.4	15.5	11.3	26.0	8.3	13.0	60.5
$>$ 3 hours	2.5	76.2	34.2	20.4	20.7	48.4	32.4	36.0	80.5
				4-I-ADL (i, j): ADL=i & IADL=j, IADL without MAP					
(0,0)	77.4	54.6	17.0	14.5	9.2	22.5	5.7	9.6	55.9
(1,0)	8.0	64.7	27.2	18.9	17.9	34.8	12.1	18.5	82.7
(0,1)	5.0	64.8	25.9	19.1	16.8	38.5	15.8	24.2	66.3
(1,1)	9.5	71.9	32.3	20.0	22.2	46.3	29.0	36.0	81.9
				4 groups					
Healthy	82.0	54.9	17.3	14.7	9.4	23.0	5.8	9.9	56.7
Physically frail	11.5	69.6	30.7	19.9	22.6	41.3	16.4	25.7	83.8
Mentally frail	3.2	69.6	28.6	18.7	17.3	42.3	26.8	35.7	72.1
Impaired	3.3	71.5	32.2	20.2	19.3	49.3	41.9	41.9	78.5

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). HBP: high blood pressure or hypertension; Diab.: diabetes or high blood sugar; Cancer: cancer or a malignant tumor of any kind except skin cancer; Lung: chronic lung disease except asthma such as chronic bronchitis or emphysema; Heart: heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems; Stroke: stroke or transient ischemic attack; Psyche: emotional, nervous, or psychiatric problems; Arthr.: arthritis or rheumatism. Results reported in percentage points. See Section 3 for details about the econometric model and the estimation procedure.

Table 7: Statistical comparison of alternative health measures

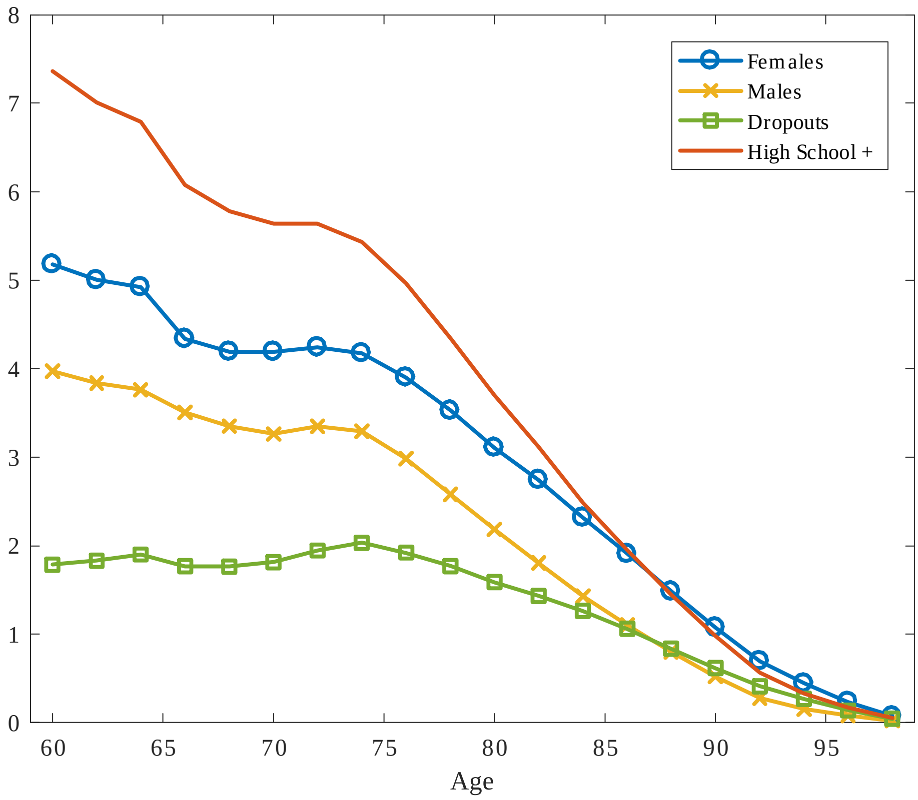
	SRH	ADL > 0	Care	4 I-ADL	4 groups	
					Probs	Mode
		Kurczynski distance				
OOP	0.0	0.1	0.0	0.1	0.3	0.3
LTC needs	0.4	0.9	0.7	1.0	4.0	3.0
Chronical conditions	1.2	0.9	0.5	0.6	1.1	0.9
ADL	3.0	13.2	6.0	10.8	24.7	19.2
IADL	2.4	5.9	8.5	10.7	25.2	18.7
		Percentage of explained variance				
OOP	1.2	1.6	1.9	2.2	3.9	3.7
LTC needs	4.5	7.6	8.5	10.9	20.0	18.9
Chronical conditions	4.9	2.6	1.1	3.0	3.9	3.6
ADL	11.9	35.2	16.3	38.9	49.7	46.7
IADL	9.9	16.7	19.2	30.0	51.8	49.0
		Percentage of explained (conditional) variance				
OOP	1.1	1.2	1.4	1.6	3.0	2.9
LTC needs	3.4	5.1	6.1	7.5	15.4	14.5
Chronical conditions	3.9	1.8	0.7	2.1	2.7	2.5
ADL	9.5	30.1	13.2	33.1	42.9	40.2
IADL	7.3	12.4	15.3	23.1	42.5	40.3

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). The first panel (Kurczynski distance) reports the heterogeneity distance proposed by Kurczynski (1970):

$$H^2 = \frac{2}{Q-1} \sum_{g \in \mathcal{Q}} (\boldsymbol{\pi}_g - \overline{\boldsymbol{\pi}})' S^{-1} (\boldsymbol{\pi}_g - \overline{\boldsymbol{\pi}}),$$

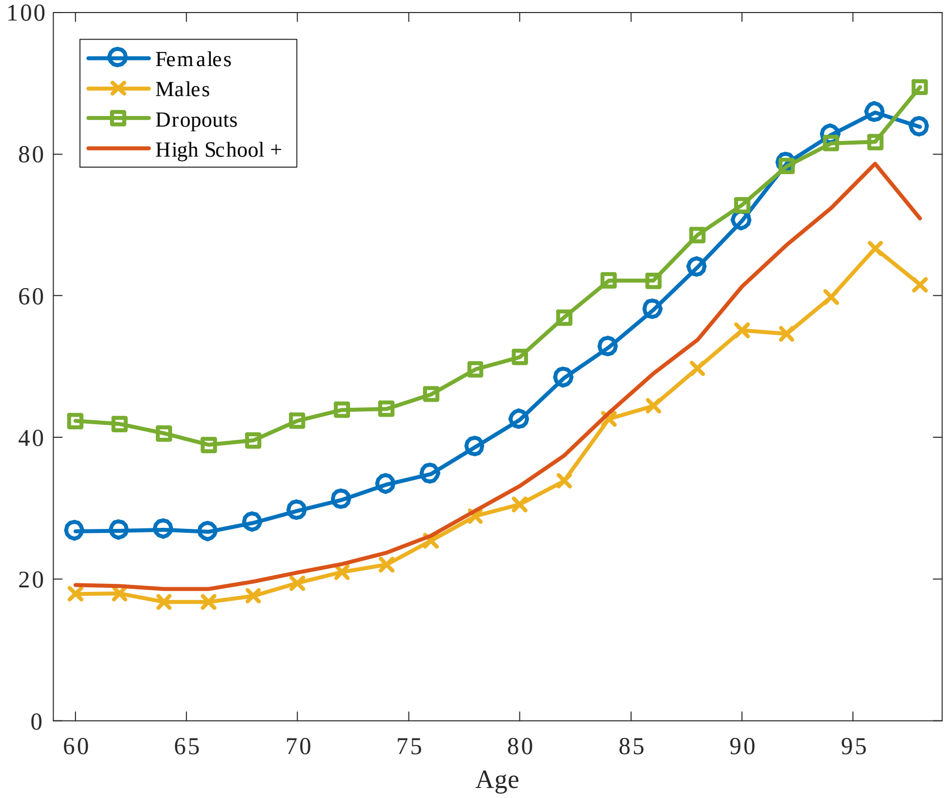
where $\mathcal{Q}$ is the set that include all possible Q groups, π_g is a $D \times 1$ vector which contains the average each of the D variables considered in the comparison, S is the within-group variability and $\overline{\boldsymbol{\pi}}$ is the average of $\boldsymbol{\pi}_g$. The second panel (R^2) displays the proportion of variance explained by each grouping method measured by the R^2 of the following regression $y_{i,t} = \mathbf{d}'_{i,t}\boldsymbol{\beta} + \varepsilon_{i,t}$, where $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs, and $y_{i,t}$ is the variable used as reference. The lower panel (Conditional R^2) repeats the same exercise but using as dependent variable the residual of $y_{i,t} = \mathbf{z}'_{i,t}\gamma + v_{i,t}$, where $\mathbf{z}_{i,t}$ includes gender, age and education.

Figure 1: Share of interviewed indiviuals by age



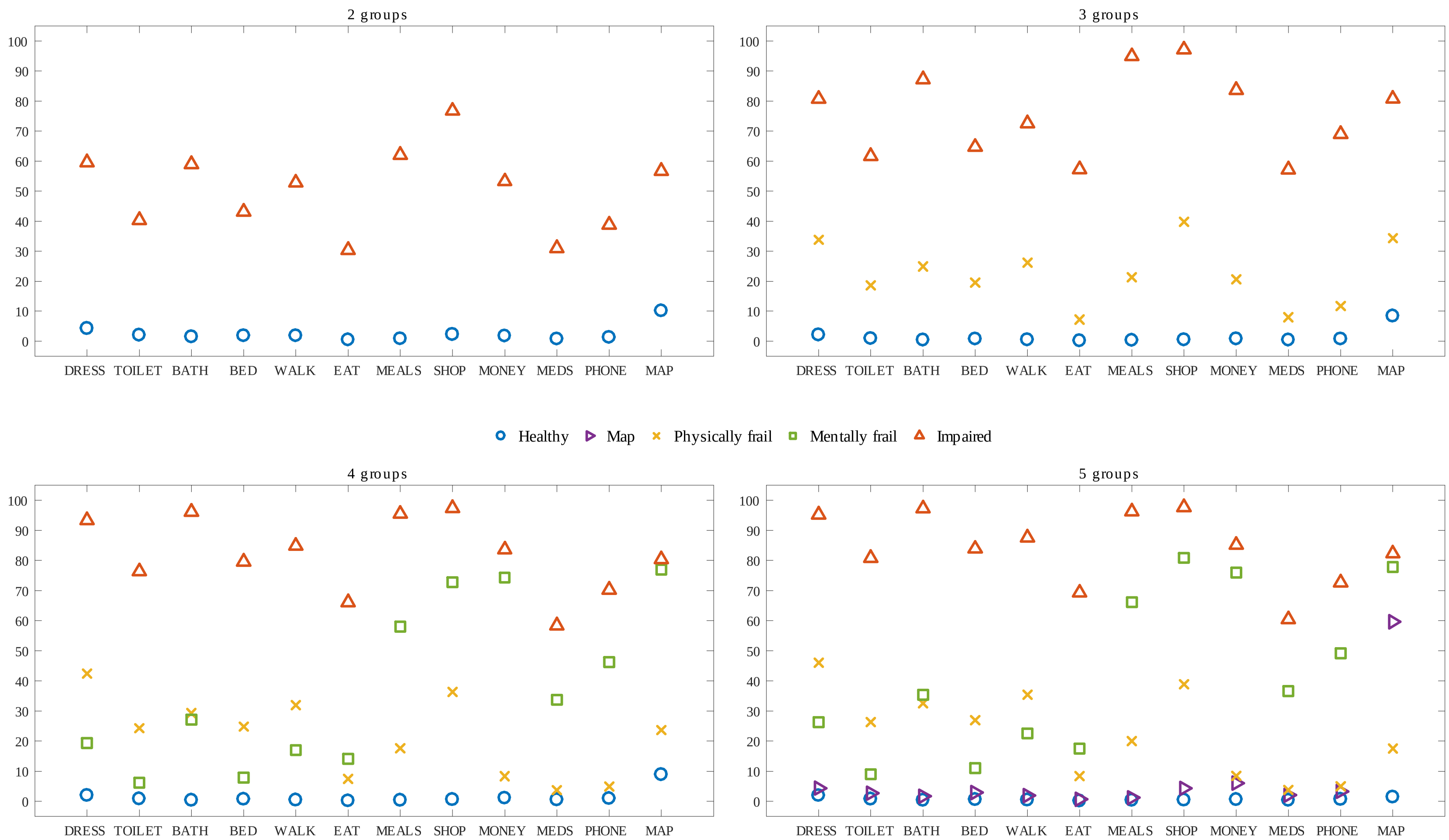
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals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

Figure 2: Share of interviewed indiviuals reporting at least one difficulty with any I-ADL by age



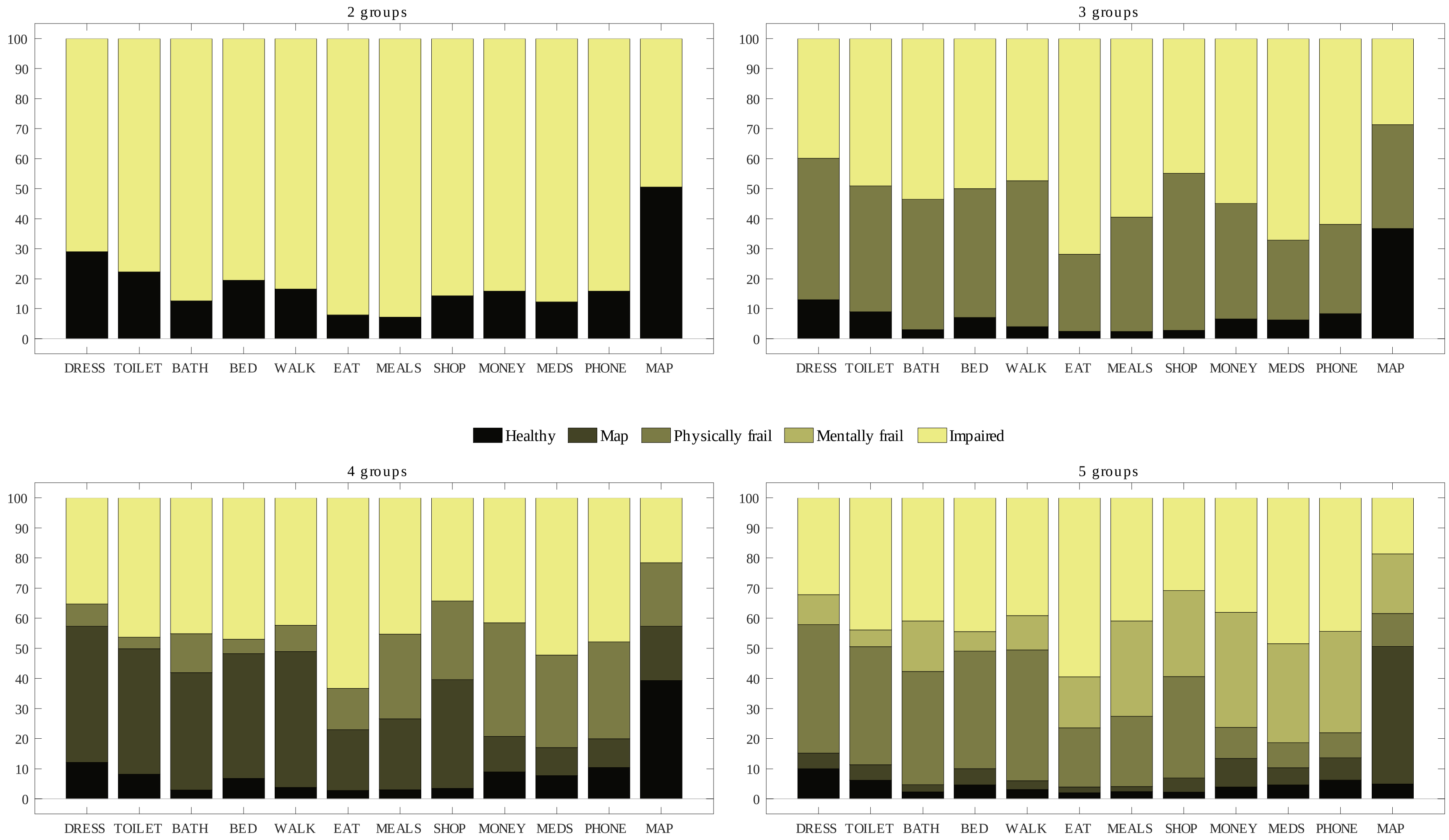
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals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

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



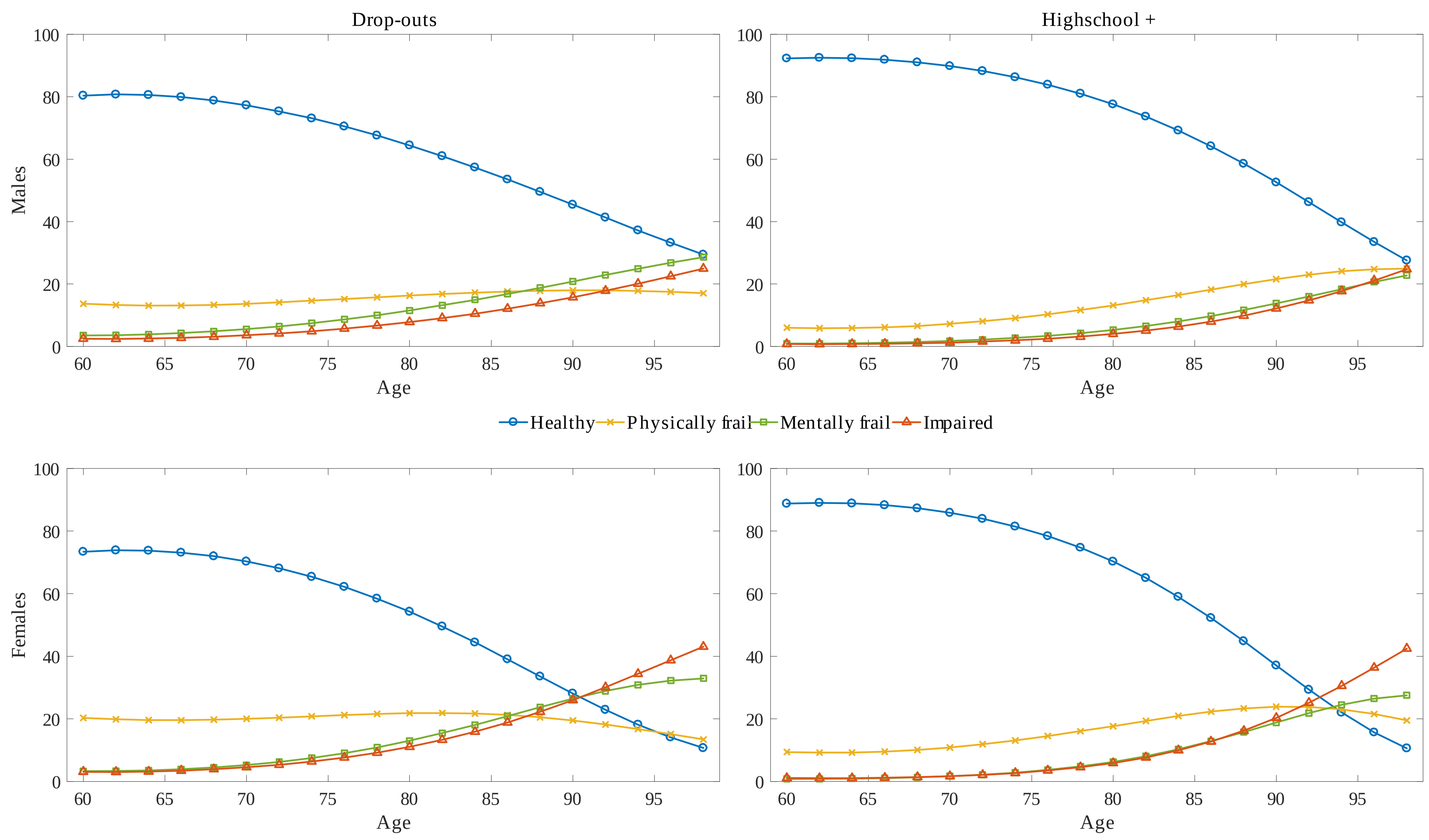
Notes: ADLs: Some difficulty with dressing (DRESS), using the toilet (TOILET), bathing (shower, BATH), getting in or out of bed (BED), to walk across a room (WALK) and eating (EAT). IADLs: Some difficulty with preparing hot meal (MEALS), shopping for groceries (SHOP), managing money (MONEY), taking medications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

Figure 4: Probability of reporting a difficulty with a given I-ADL by health group



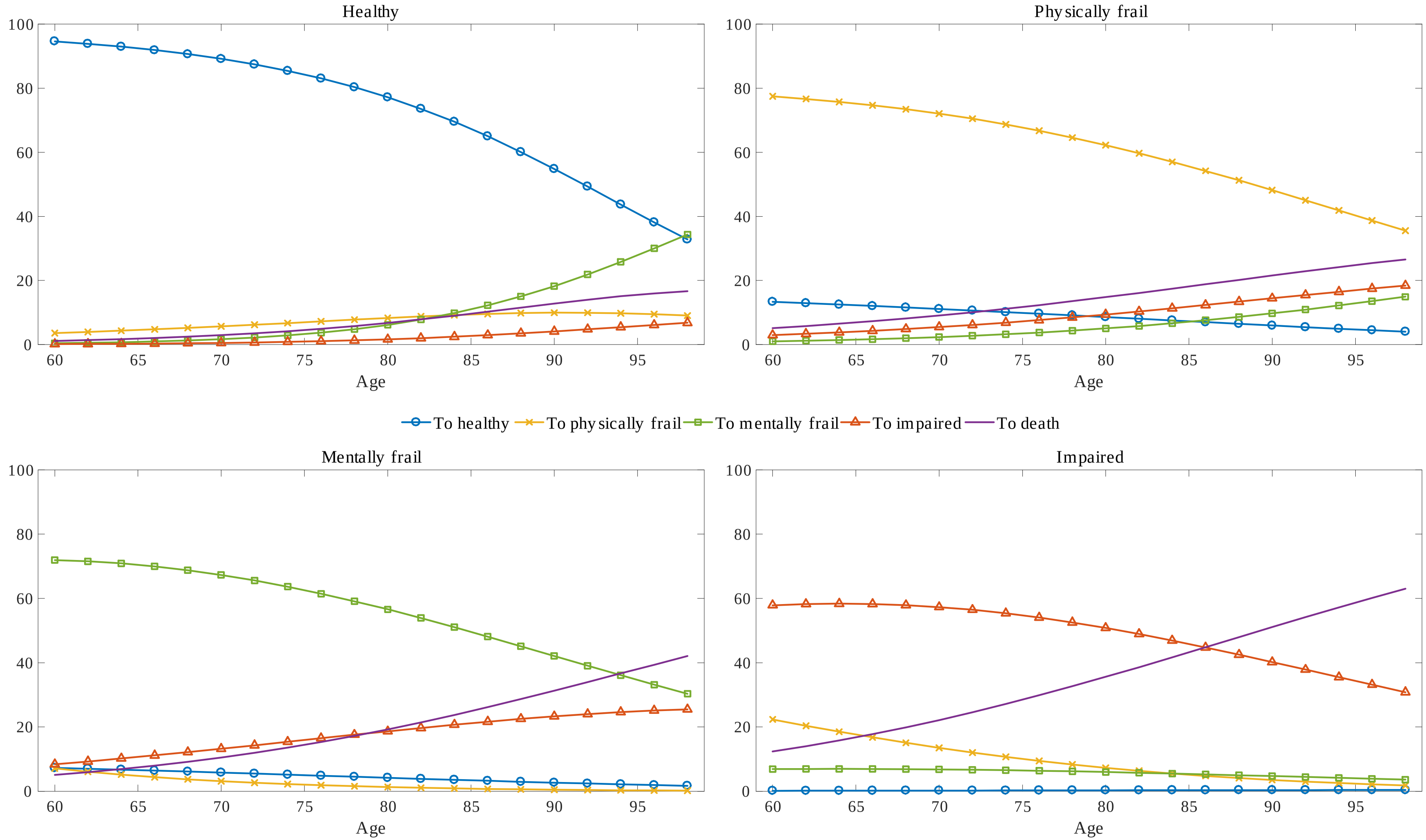
Notes: ADLs: Some difficulty with dressing (DRESS), using the toilet (TOILET), bathing (shower, BATH), getting in or out of bed (BED), to walk across a room (WALK) and eating (EAT). IADLs: Some difficulty with preparing hot meal (MEALS), shopping for groceries (SHOP), managing money (MONEY), taking medications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

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



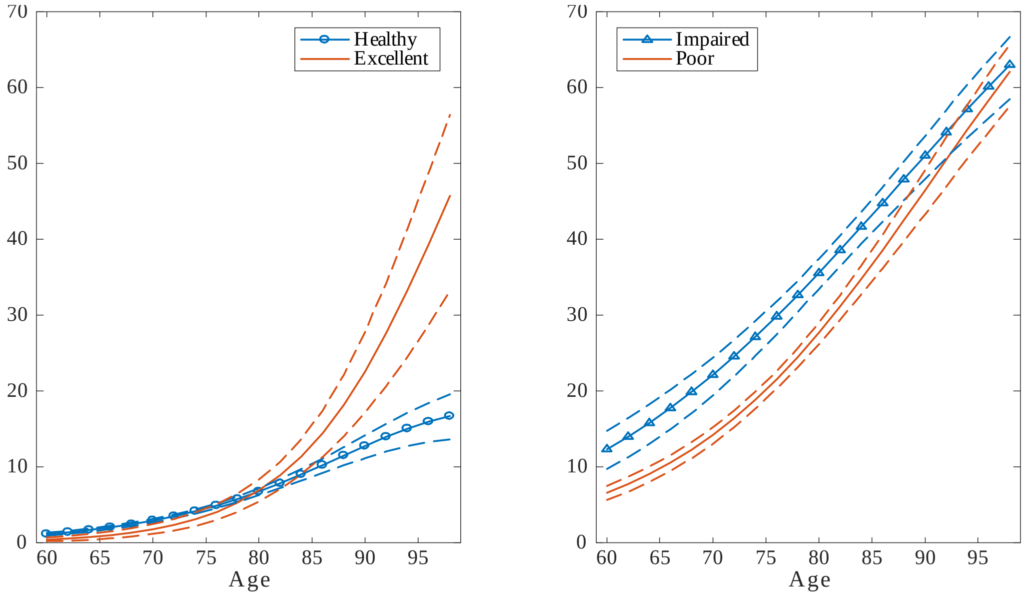
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). Results reported in years. See Section 3 for details about the econometric model and the estimation procedure.

Figure 6: Transitions by group as as individuals age: female drop-outs



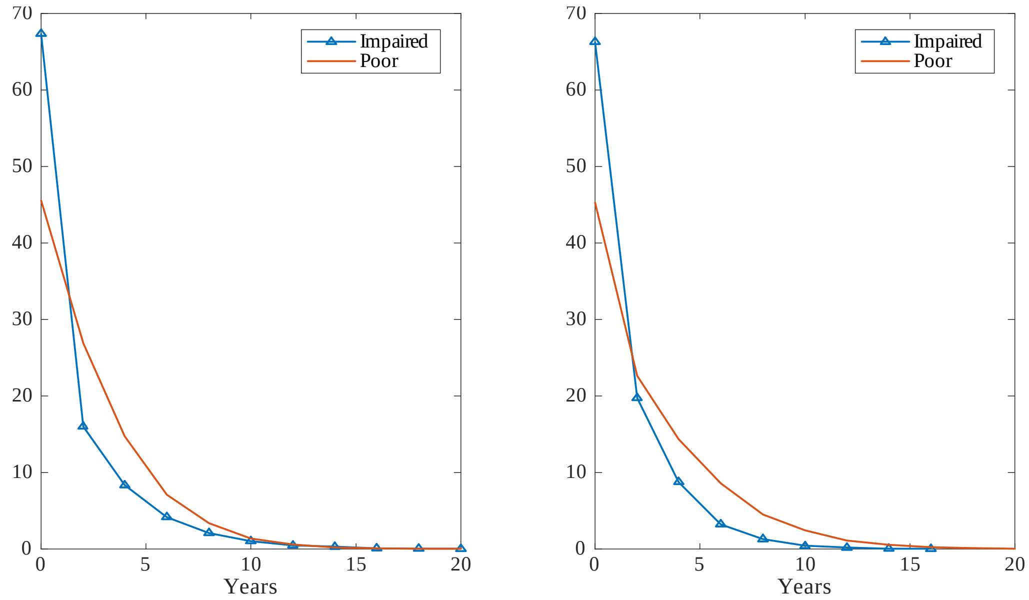
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). Results reported in years. See Section 3 for details about the econometric model and the estimation procedure.

Figure 7: Transition to death: female drop-outs



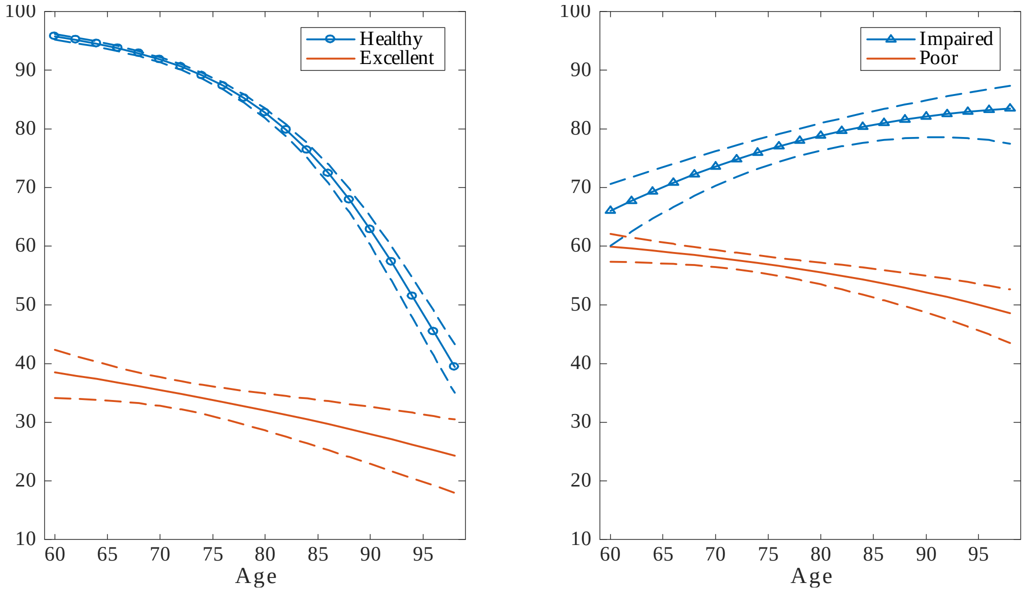
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). Results reported in years. See Section 3 for details about the econometric model and the estimation procedure.

Figure 8: Distribution of years in good and bad health conditional on being in good health at age 60 (left panel) and 80 (right panel): female drop-outs



Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). Results reported in years. See Section 3 for details about the econometric model and the estimation procedure.

Figure 9: Persistence of health status: female drop-outs



Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). Results reported in years. See Section 3 for details about the econometric model and the estimation procedure.