Supplemental Material: Endogenous Health Groups and Heterogeneous Dynamics of the Elderly

Dante Amengual
CEMFI
<amengual@cemfi.es>

Jesús Bueren
European University Institute
<jesus.bueren@eui.eu>

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

Abstract

Section S.1 complements section 2 of the paper by providing a summary of all variables used, a visualization of the sample composition, and summary statistics of the health costs. Section S.2 provide similar details for the case of European data. Section S.3 discusses some technical details about the estimation. Section S.4 gathers the posterior of all the parameters of our model and provides evidence for convergence of the Metropolis-within-Gibbs in terms of effective sample sizes. Section S.5 complements Table 4 providing the whole distribution of medical expenses per health group. Section S.6 gathers all the parameters of the regressions in Table 5 and the results without covariates. Section S.7 complements Section 5.2 by gathering all the estimates for the logit estimation using self-reported case, showing the analogous figures to figure 4 using the remaining education-gender groups, and summarizing the transitions of all groups by education and gender. Finally, S.8 details the replication exercise.

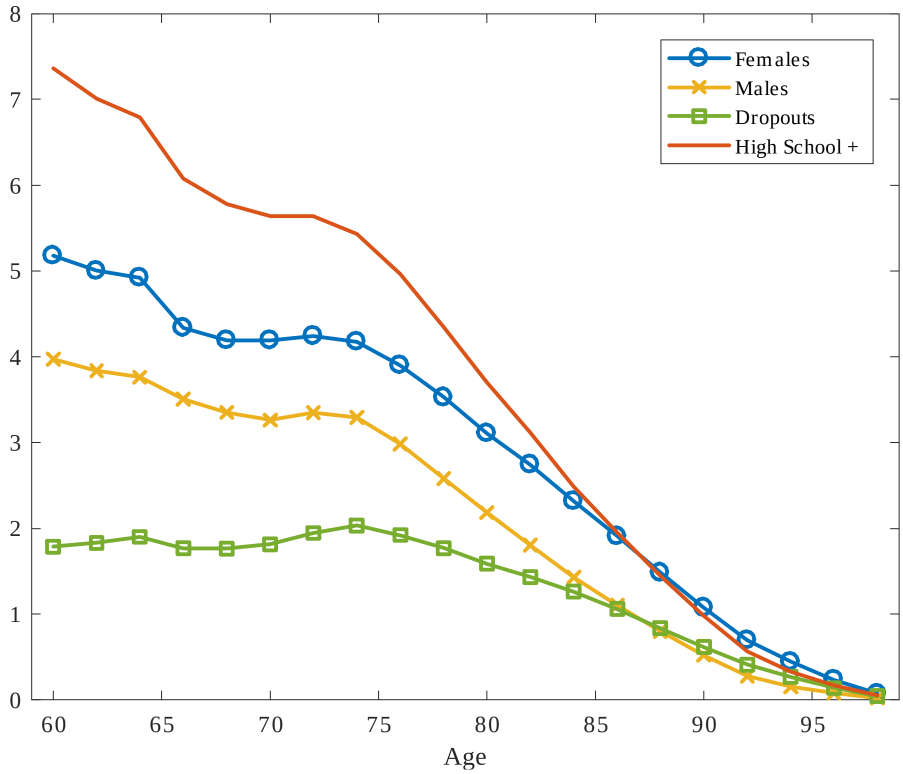
S.1 HRS Data

Table S.1: Variable definition

Variable	Description	Source
	ADLS	
	Respondent reports some difficulty with...	
DRESS	Dressing	DRESSA
TOILET	Using the toilet	TOILTA
BATH	Bathing	BATHA
BED	Getting in and out of bed	BEDA
WALK	Walking across the room	WALKA
EAT	Eating	EATA
	IADLS	
	Respondent reports some difficulty with...	
MEALS	Preparing meals	MEALSA
SHOP	Shopping for groceries	SHOPA
MONEY	Managing money	MONEYA
MEDS	Taking medications	MEDSA
PHONE	Using the phone	PHONEA
MAP	Using a map	MAPA
	Financial risk variables	
OOP†	Total out-of-pocket medical expenditures since the last interview, or the last 2 years for new interviewees. Measured in constant 2000 US dollars	OOPMD
Nurs-h resident	Respondent lives in a nursing home or other health care facility at the time of the interview.	NHMLIV
Received h-care	Respondent reports if any medically- trained person has come to respondent's home since the last interview, or the last 2 years for new interviewees.	HOMCAR
	Classification Method	
SRH	Respondent's self-reported general health status. Codes range from "1" for Excellent to "5" for Poor.	SHLT
Frailty index	Constructed using the variables based on Genworth and Mutual of Omaha LTCI underwriting guidelines	Braun et al. (2017) on-line appendix
4-I-ADL	Based on whether the respondent reports difficulty with any of the previous ADL and/or IADL.	

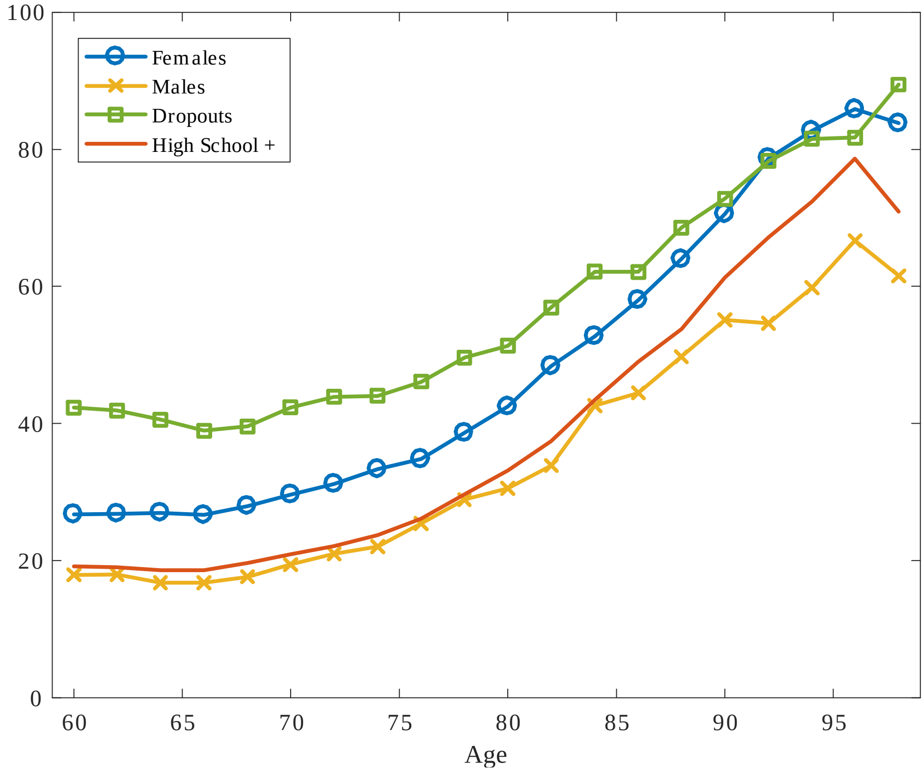
Notes: This table describes the main variables used in the analysis. The last column indicates the source of the data. Capital letters indicate the name of the variable in the HRS RAND v.P. † Includes hospital and nursing home stays, doctor visits, dental treatments, outpatient surgery, prescription drugs, home health care, and special facilities.

Figure S.1: Share of interviewees by age



Notes: This figure shows how the composition of individuals changes across age. 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 S.2: Share of interviewees reporting at least one difficulty with an I-ADL by age



Notes: This figure plots the incidence of problems with I-ADLs. 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.

Table S.2: Summary Statistics

	Wave	Mean	Std	Median	25%	75%
Estimation Sample						
OOP	Current	3,027	9,841	1,070	300	2,754
	Next	3,178	10,433	1,127	333	2,869
Nurs-h resident	Current	2.9	16.9	0.0	0.0	0.0
	Next	3.4	18.0	0.0	0.0	0.0
Received h-care	Current	9.7	29.6	0.0	0.0	0.0
	Next	10.3	30.4	0.0	0.0	0.0
Mortality	Next	8.5	27.9	0.0	0.0	0.0
Comparison Sample						
OOP	Current	2,897	9,349	1,100	334	2,761
	Next	3,043	10,037	1,145	355	2,852
Nurs-h resident	Current	2.1	14.2	0.0	0.0	0.0
	Next	2.7	16.1	0.0	0.0	0.0
Received h-care	Current	9.3	29.0	0.0	0.0	0.0
	Next	10.0	30.1	0.0	0.0	0.0
Mortality	Next	7.6	26.5	0.0	0.0	0.0

Notes: This table presents the summary statistics of the two samples used in the paper. The estimation sample corresponds to the observations that we incorporate into the estimation procedure. Due to missing data, we might not be able to classify individuals in this sample using the alternative classifications; hence we use a restricted sample in order to compare across classification methods. The summary statistics of this latter sample are included in the second panel.

S.2 The European Case

There are two main surveys that cover European retirees and their health: ELSA and SHARE. The former covers the data from people aged over 50 living in the UK. Meanwhile, SHARE covers similar data of people from 28 European countries and Israel. Each European country has a specific team running the survey which leads to small differences, but more importantly, different sample sizes. To avoid noise coming from undersampled countries, we focus on 17 countries which are significantly covered and we aggregate them in groups according to their similarity: Continental Europe (Austria, Germany, Netherlands, Switzerland, Belgium, Luxembourg); Nordic countries (Sweden, Denmark); Mediterranean (Spain, Italy, France, Israel, Portugal); Eastern Europe (Czech Republic, Poland, Hungary, Slovenia, Estonia). The European surveys cover very similar data as the HRS; hence, we mimic the definitions in the main paper (Table S.3 and S.4 detail the precise data). The only exception are out-of-pocket medical expenses, for which the ELSA does not provide data.

Table S.5 and S.6 provide the summary statistics corresponding to ELSA and SHARE respectively. An striking feature is the low proportion of nursing home residency compared to the US. This fact arises because of the attrition in these surveys resulting from the complications of reinterviewing people living in nursing homes (Banks et al., 2011). Table S.7 and S.8 summarize dynamics by the expected duration in each health group by education, gender, and region. We can draw the same conclusions as for the US case.

Table S.3: Variable definition in ELSA

Variable	Description	Source
	ADLS	
	Respondent reports some difficulty with...	
DRESS	Dressing	headldr
TOILET	Using the toilet	headlwc
BATH	Bathing	headlba
BED	Getting in and out of bed	headlbe
WALK	Walking across the room	headlwa
EAT	Eating	headlea
	IADLS	
	Respondent reports some difficulty with...	
MEALS	Preparing meals	headlpr
SHOP	Shopping for groceries	headlsh
MONEY	Managing money	headlmo
MEDS	Taking medications	headlme
PHONE	Using the phone	headlph
MAP	Using a map	headlma
	Financial risk variables	
OOP	Total out-of-pocket medical expenditures: NOT AVAILABLE	-
Nurs-h resident	Respondent lives in a nursing home or other health care facility at the time of the interview.	inst
Received h-care	Respondent reports if any medically- trained person has come to respondent's home since the last interview, or the last 2 years for new interviewees.	hehpb
	Classification Method	
SRH	Respondent's self-reported general health status. Codes range from "1" for Excellent to "5" for Poor.	shlt
Frailty index	Constructed using the variables based on Genworth and Mutual of Omaha LTCI underwriting guidelines	Braun et al. (2017) on-line appendix
4-I-ADL	Based on whether the respondent reports difficulty with any of the previous ADL and/or IADL.	

Notes: This table describes the main variables used in the analysis. The last column indicates the source of the data. Capital letters indicate the name of the variable in the Institute of Fiscal Studies contributed files.

Table S.4: Variable definition in Share

Variable	Description	Source
	ADLS	
	Respondent reports some difficulty with...	
DRESS	Dressing	ph049d1
TOILET	Using the toilet	ph049d6
BATH	Bathing	ph049d3
BED	Getting in and out of bed	ph049d5
WALK	Walking across the room	ph049d2
EAT	Eating	ph049d4
	IADLS	
	Respondent reports some difficulty with...	
MEALS	Preparing meals	ph049d8
SHOP	Shopping for groceries	ph049d9
MONEY	Managing money	ph049d13
MEDS	Taking medications	ph049d11
PHONE	Using the phone	ph049d10
MAP	Using a map	ph049d7
	Financial risk variables	
OOP†	Total out-of-pocket medical expenditures since the last interview, or the last 2 years for new interviewees. Measured in constant 2018 Euros dollars	hc045e-hc051e
Nurs-h resident	Respondent lives in a nursing home or other health care facility at the time of the interview.	hc029
Received h-care	Respondent reports if any medically- trained person has come to respondent's home since the last interview, or the last 2 years for new interviewees.	hc032d1-hc032d3
	Classification Method	
SRH	Respondent's self-reported general health status. Codes range from "1" for Excellent to "5" for Poor.	healthstat
Frailty index	Constructed using the variables based on Genworth and Mutual of Omaha LTCI underwriting guidelines	Braun et al. (2017) on-line appendix
4-I-ADL	Based on whether the respondent reports difficulty with any of the previous ADL and/or IADL.	

Notes: This table describes the main variables used in the analysis. The last column indicates the source of the data. Capital letters indicate the name of the variable in the SHARE release 7.1.0 † Includes inpatient care, outpatient care, prescribed drugs and nursing home/daycare/homecare during last 12 months.

Table S.5: Summary Statistics: ELSA

	Mean	Std
Estimation Sample		
Nurs-h resident	0.5	7.0
Received h-care	4.5	20.8
Mortality	5.1	22.1
Comparison Sample		
Nurs-h resident	0.1	2.8
Received h-care	4.5	20.7
Mortality	4.8	21.4

Notes: This table presents the summary statistics of the two samples used in the paper. The estimation sample corresponds to the observations that we incorporate into the estimation procedure. Due to missing data, we might not be able to classify individuals in this sample using the alternative classifications; hence we use a restricted sample in order to compare across classification methods. The summary statistics of this latter sample are included in the second panel.

Table S.6: Summary Statistics: SHARE

	Mean	Std	Median	25%	75%
Estimation Sample					
OOP	544.9	1,864.8	174.7	40.1	482.2
Nurs-h resident	0.5	7.2	0.0	0.0	0.0
Received h-care	10.7	30.9	0.0	0.0	0.0
Mortality	5.7	23.2	0.0	0.0	0.0
Comparison Sample					
OOP	555.6	1,882.0	182.2	44.1	500.8
Nurs-h resident	0.5	7.1	0.0	0.0	0.0
Received h-care	10.6	30.8	0.0	0.0	0.0
Mortality	5.7	23.2	0.0	0.0	0.0

Notes: This table presents the summary statistics of the two samples used in the paper. The estimation sample corresponds to the observations that we incorporate into the estimation procedure. Due to missing data, we might not be able to classify individuals in this sample using the alternative classifications; hence we use a restricted sample in order to compare across classification methods. The summary statistics of this latter sample are included in the second panel.

S.3 Estimation details

Letting $p(\mu, \beta)$ denote the priors on the parameter vectors, and $p(h_{it}|h_{it-1}, \mathbf{w}_{it}, \beta)$ the transition probabilities that characterize the health of individual i at time t conditional on her health on the previous period and $\mathbf{w}_{it}$, which we collect in $\mathbf{W}$ (i.e. age, gender and education level), we can write the priors for both the parameter and latent variables as

$$p(\mathbf{H}|\mathbf{W}, \mathbf{h}_0, \beta) \times p(\mu, \beta) = \prod_{i=1}^{N} \prod_{t=1}^{T} p(h_{it}|h_{it-1}, \mathbf{w}_{it}, \beta) \times p(\mathbf{h}_0) \times p(\mu, \beta),$$

where $\mathbf{h}_0 = (h_{10}, ..., h_{N0})'$ denotes the health of individual i in the period when she enters into the sample. The joint likelihood of the observables we explicitly model is given by

$$\mathcal{L}(\tilde{\mathbf{X}}\,|\mathbf{W}, \mathbf{H}, \mu, \mathbf{h}_0\,) = \prod_{i=1}^{N} \prod_{t=1}^{T} p(\mathbf{x}_{it}|h_{it}, \mu),$$

where $\tilde{\mathbf{X}} = \mathbf{X} \setminus \mathbf{W}$ and the expression for $p(\mathbf{x}_{it}|h_{it}, \mu)$ is given in (1). Therefore, we obtain,

$$p(\mu, \beta, \mathbf{H}) \propto \mathcal{L}(\mathbf{X}\,|\mathbf{W}, \mathbf{H}, \mu, \mathbf{h}_0\,) \times p(\mathbf{H}|\mathbf{W}, \mathbf{h}_0, \beta) \times p(\mathbf{h}_0) \times p(\mu, \beta).$$

We consider the initial health state of each individual as a new set of parameters.

The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed on average 6 waves. Regarding the number of parameters, consider for instance the case of four health status: since the probability of reporting difficulties varies with health status we have that μ is a vector of dimension $12 \times 4 = 48$, and similarly, the transitions are also varying across health types so that β is a vector of dimension $16 \times 6 = 96$. For each wave in which we observe an individual we have 12 observations (one for each I-ADL) which, conditional on parameter values, provide information about her current health status. Therefore, we have 144 parameters (or, if we counted latent variables, 159,168) and more than 1.9 millions of observations.

As for the consistency of parameter estimates, what we need is, at least, the size of panel individuals to increase in order to obtain consistency of the parameters. Finally, the model is a standard regime-switching Markov chain with the exception of a large number of idiosyncratic chains, but this fact does not imply violations on the regularity conditions for the consistency and asymptotic normality of the parameters μ and β. Moreover, the priors we use satisfy the requirements in Barron et al. (1999).

Table S.7: Expected forthcoming time in each health group by education and gender at age 60.

Education	Healthy +	Physically frail +	Mentally frail +	Impaired =	Life Expectancy
Continental Europe					
Females					
Dropouts	18.9	2.8	1.3	0.8	23.8
	(0.2)	(0.1)	(0.1)	(0.0)	(0.1)
Highschool	22.1	2.2	0.8	0.5	25.6
	(0.1)	(0.1)	(0.0)	(0.0)	(0.1)
Males					
Dropouts	18.0	1.7	0.7	0.5	20.9
	(0.2)	(0.1)	(0.0)	(0.0)	(0.2)
Highschool	20.9	1.4	0.5	0.4	23.2
	(0.1)	(0.1)	(0.0)	(0.0)	(0.1)
Nordic					
Females					
Dropouts	19.5	1.7	0.9	0.5	22.6
	(0.2)	(0.1)	(0.1)	(0.0)	(0.2)
Highschool	22.3	1.4	0.5	0.3	24.6
	(0.2)	(0.1)	(0.0)	(0.0)	(0.2)
Males					
Dropouts	17.6	1.0	0.5	0.2	19.3
	(0.2)	(0.1)	(0.0)	(0.0)	(0.2)
Highschool	20.4	0.8	0.3	0.2	21.8
	(0.2)	(0.1)	(0.0)	(0.0)	(0.2)
Mediterranean					
Females					
Dropouts	18.3	2.2	1.6	1.2	23.3
	(0.1)	(0.1)	(0.1)	(0.0)	(0.1)
Highschool	21.3	1.8	1.1	0.9	25.1
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Males					
Dropouts	17.2	1.3	0.9	0.7	20.2
	(0.1)	(0.0)	(0.0)	(0.0)	(0.2)
Highschool	20.1	1.1	0.6	0.6	22.5
	(0.2)	(0.1)	(0.0)	(0.0)	(0.2)

Notes: SHARE and ELSA Data. We select individuals over 60 years old and we drop individuals whose education, gender or age are missing. The final sample consists of 71,416 individuals. Results reported in years. In parentheses we report the standard deviation of the posterior distribution.

Table S.8: Expected forthcoming time in each health group by education and gender at age 60.

Education	Healthy +	Physically frail +	Mentally frail +	Impaired =	Life Expectancy
			Eastern Europe		
			Females		
Dropouts	16.2	2.8	1.3	0.8	21.1
	(0.2)	(0.1)	(0.1)	(0.0)	(0.2)
Highschool	19.1	2.3	0.9	0.7	23.1
	(0.1)	(0.1)	(0.1)	(0.0)	(0.1)
			Males		
Dropouts	18.0	1.7	0.7	0.5	20.9
	(0.2)	(0.1)	(0.0)	(0.0)	(0.2)
Highschool	20.9	1.4	0.5	0.4	23.2
	(0.1)	(0.1)	(0.0)	(0.0)	(0.1)
			U.K.		
			Females		
Dropouts	15.5	5.3	0.8	0.7	22.3
	(0.2)	(0.1)	(0.1)	(0.0)	(0.2)
Highschool	19.3	4.0	0.5	0.6	24.3
	(0.2)	(0.1)	(0.0)	(0.0)	(0.1)
			Males		
Dropouts	14.4	3.4	0.4	0.5	18.7
	(0.2)	(0.1)	(0.0)	(0.0)	(0.2)
Highschool	18.4	2.4	0.3	0.3	21.5
	(0.2)	(0.1)	(0.0)	(0.0)	(0.2)

Notes: SHARE and ELSA Data. We select individuals over 60 years old and we drop individuals whose education, gender or age are missing. The final sample consists of 71,416 individuals. Results reported in years. In parentheses we report the standard deviation of the posterior distribution.

S.4 Parameters posterior and effective sample sizes

Table S.9: Parameter estimates: 2 Groups

Variable	Healthy		Impaired	
	To Healthy			
Constant	-8.624	(0.278)	-6.920	(0.492)
Age	0.085	(0.004)	0.104	(0.007)
HighSchool	-0.975	(0.298)	-0.020	(0.490)
Female	-1.067	(0.279)	-0.019	(0.516)
Age × Female	0.007	(0.004)	-0.003	(0.007)
Age × HighSchool	0.008	(0.004)	0.004	(0.007)
	To Impaired			
Constant	-0.272	(0.369)	-6.334	(0.342)
Age	0.005	(0.005)	0.073	(0.004)
HighSchool	1.891	(0.386)	0.727	(0.332)
Female	-1.129	(0.374)	0.044	(0.352)
Age × Female	0.004	(0.005)	-0.007	(0.004)
Age × HighSchool	-0.020	(0.005)	-0.008	(0.004)
	Pr. of each I-ADL by health group			
WALK	0.018	(0.000)	0.528	(0.004)
DRESS	0.043	(0.001)	0.596	(0.004)
BATH	0.015	(0.000)	0.589	(0.004)
EAT	0.005	(0.000)	0.303	(0.003)
BED	0.018	(0.000)	0.431	(0.004)
TOILET	0.020	(0.000)	0.403	(0.004)
MAP	0.101	(0.001)	0.568	(0.004)
PHONE	0.013	(0.000)	0.388	(0.004)
MONEY	0.018	(0.000)	0.532	(0.004)
MED	0.008	(0.000)	0.310	(0.003)
SHOP	0.023	(0.001)	0.767	(0.004)
MEAL	0.008	(0.000)	0.621	(0.004)

Notes: Median and standard deviation (in parenthesis) of the posterior of each parameter in the estimation with two groups. estimates and standard errors. Each column refers to the current health group of the individual while each of the first two panels presents the parameters of the transition to a different health group. The last panel gathers the estimation results of the Bernouilli process that drives I-ADLs.

Table S.10: Parameter estimates: 3 Groups

Variable	Healthy		Frail		Impaired	
			To Healthy			
Constant	-8.693	(0.350)	-6.658	(0.627)	1.522	(5.529)
Age	0.085	(0.005)	0.099	(0.009)	0.107	(0.153)
HighSchool	-0.811	(0.370)	-1.072	(0.624)	-2.622	(6.042)
Female	-1.446	(0.342)	0.743	(0.620)	0.403	(5.980)
Age × Female	0.011	(0.005)	-0.015	(0.009)	0.021	(0.079)
Age × HighSchool	0.006	(0.005)	0.018	(0.009)	-0.050	(0.151)
			To Frail			
Constant	-0.452	(0.450)	-6.991	(0.410)	-7.978	(1.034)
Age	0.005	(0.006)	0.077	(0.005)	0.123	(0.015)
HighSchool	2.008	(0.478)	0.165	(0.402)	-0.262	(0.855)
Female	-1.591	(0.456)	-0.079	(0.423)	1.172	(1.027)
Age × Female	0.011	(0.006)	-0.006	(0.005)	-0.023	(0.015)
Age × HighSchool	-0.023	(0.006)	-0.003	(0.005)	0.012	(0.012)
			To Impaired			
Constant	4.019	(0.896)	-0.014	(0.576)	-5.888	(0.525)
Age	-0.028	(0.012)	0.008	(0.007)	0.074	(0.007)
HighSchool	2.168	(0.966)	0.538	(0.576)	1.555	(0.514)
Female	0.451	(0.915)	0.367	(0.596)	0.877	(0.551)
Age × Female	-0.015	(0.012)	-0.012	(0.007)	-0.018	(0.007)
Age × HighSchool	-0.025	(0.012)	-0.005	(0.007)	-0.017	(0.006)
			Pr. of each I-ADL by health group			
WALK	0.005	(0.000)	0.262	(0.003)	0.725	(0.006)
DRESS	0.021	(0.001)	0.337	(0.004)	0.807	(0.005)
BATH	0.004	(0.000)	0.249	(0.004)	0.872	(0.005)
EAT	0.002	(0.000)	0.072	(0.002)	0.573	(0.006)
BED	0.007	(0.000)	0.195	(0.003)	0.647	(0.006)
TOILET	0.009	(0.000)	0.186	(0.003)	0.617	(0.006)
MAP	0.084	(0.001)	0.343	(0.004)	0.808	(0.006)
PHONE	0.007	(0.000)	0.117	(0.002)	0.690	(0.006)
MONEY	0.008	(0.000)	0.206	(0.003)	0.837	(0.005)
MED	0.004	(0.000)	0.080	(0.002)	0.572	(0.006)
SHOP	0.005	(0.000)	0.397	(0.005)	0.971	(0.002)
MEAL	0.003	(0.000)	0.213	(0.004)	0.949	(0.003)

Notes: Median and standard deviation (in parenthesis) of the posterior of each parameter in the estimation with three groups. estimates and standard errors. Each column refers to the current health group of the individual while each of the first three panels presents the parameters of the transition to a different health group. The last panel gathers the estimation results of the Bernouilli process that drives I-ADLs.

Table S.11: Parameter estimates: 4 Groups

Variable	Healthy		Physically frail		Mentally frail		Impaired	
				To Healthy				
Constant	-8.802	(0.331)	-6.251	(0.734)	-7.972	(1.086)	1.415	(3.528)
Age	0.086	(0.004)	0.094	(0.010)	0.114	(0.014)	0.023	(0.042)
HighSchool	-0.797	(0.362)	-1.069	(0.689)	-0.648	(1.392)	0.464	(4.472)
Female	-1.454	(0.330)	0.816	(0.733)	2.007	(1.413)	1.857	(5.361)
Age × Female	0.011	(0.004)	-0.019	(0.010)	-0.021	(0.018)	-0.000	(0.066)
Age × HighSchool	0.006	(0.005)	0.017	(0.010)	0.016	(0.019)	-0.004	(0.054)
				To Physically frail				
Constant	-1.506	(0.508)	-6.646	(0.493)	-7.158	(7.180)	-10.056	(1.549)
Age	0.026	(0.007)	0.072	(0.006)	0.678	(0.946)	0.161	(0.023)
HighSchool	2.320	(0.523)	-0.007	(0.485)	-11.409	(6.099)	-1.023	(1.107)
Female	-2.346	(0.512)	0.120	(0.479)	-2.371	(7.272)	2.903	(1.566)
Age × Female	0.020	(0.007)	-0.009	(0.006)	-0.523	(0.944)	-0.051	(0.023)
Age × HighSchool	-0.029	(0.007)	-0.001	(0.006)	0.193	(0.098)	0.022	(0.015)
				To Mentally frail				
Constant	2.592	(0.636)	1.485	(1.484)	-7.756	(0.742)	-2.690	(1.819)
Age	-0.022	(0.008)	0.003	(0.019)	0.089	(0.009)	0.064	(0.023)
HighSchool	2.452	(0.682)	-0.426	(1.335)	1.448	(0.812)	1.326	(2.064)
Female	1.142	(0.663)	1.798	(1.421)	0.415	(0.822)	-0.305	(1.954)
Age × Female	-0.023	(0.008)	-0.031	(0.018)	-0.010	(0.010)	-0.004	(0.025)
Age × HighSchool	-0.025	(0.009)	0.009	(0.017)	-0.017	(0.010)	-0.002	(0.026)
				To Impaired				
Constant	3.570	(0.943)	0.839	(0.832)	-2.339	(0.927)	-5.845	(0.597)
Age	-0.019	(0.012)	0.002	(0.011)	0.037	(0.011)	0.076	(0.008)
HighSchool	3.228	(1.049)	0.007	(0.783)	0.581	(0.995)	1.338	(0.568)
Female	0.149	(1.026)	0.023	(0.833)	0.268	(1.059)	0.724	(0.619)
Age × Female	-0.010	(0.013)	-0.007	(0.011)	-0.011	(0.013)	-0.016	(0.008)
Age × HighSchool	-0.037	(0.013)	0.003	(0.010)	-0.008	(0.012)	-0.015	(0.007)
				Pr. of each I-ADL by health group				
WALK	0.005	(0.000)	0.320	(0.005)	0.170	(0.006)	0.848	(0.005)
DRESS	0.020	(0.001)	0.424	(0.005)	0.193	(0.007)	0.934	(0.004)
BATH	0.004	(0.000)	0.293	(0.004)	0.272	(0.008)	0.961	(0.003)
EAT	0.002	(0.000)	0.074	(0.002)	0.141	(0.005)	0.661	(0.007)
BED	0.007	(0.000)	0.248	(0.004)	0.079	(0.004)	0.795	(0.006)
TOILET	0.008	(0.000)	0.243	(0.004)	0.062	(0.004)	0.764	(0.007)
MAP	0.089	(0.001)	0.236	(0.005)	0.770	(0.009)	0.804	(0.007)
PHONE	0.009	(0.000)	0.049	(0.002)	0.463	(0.009)	0.703	(0.007)
MONEY	0.011	(0.000)	0.084	(0.003)	0.743	(0.009)	0.836	(0.006)
MED	0.005	(0.000)	0.037	(0.002)	0.337	(0.008)	0.584	(0.007)
SHOP	0.006	(0.000)	0.363	(0.005)	0.728	(0.009)	0.973	(0.002)
MEAL	0.004	(0.000)	0.176	(0.004)	0.580	(0.010)	0.955	(0.003)

Notes: Median and standard deviation (in parenthesis) of the posterior of each parameter in the estimation with four groups. estimates and standard errors. Each column refers to the current health group of the individual while each of the first four panels presents the parameters of the transition to a different health group. The last panel gathers the estimation results of the Bernouilli process that drives I-ADLs.

Table S.12: Parameter estimates: 5 Groups

Variable	Healthy		Map		Physically frail		Mentally frail		Impaired	
					To Healthy					
Constant	-8.502	(0.362)	-12.973	(1.500)	-5.347	(0.701)	-19.282	(3.662)	-6.126	(4.478)
Age	0.082	(0.005)	0.194	(0.023)	0.086	(0.010)	3.565	(0.356)	1.815	(0.543)
HighSchool	-0.988	(0.392)	-0.855	(0.963)	-0.954	(0.690)	6.729	(2.325)	-15.305	(1.543)
Female	-1.688	(0.315)	2.511	(1.254)	0.111	(0.539)	-7.691	(2.239)	-2.373	(1.877)
Age $\times$ F	0.014	(0.004)	-0.035	(0.019)	-0.004	(0.008)	-2.507	(0.306)	1.252	(0.209)
Age $\times$ HS	0.009	(0.005)	0.004	(0.014)	0.010	(0.010)	1.186	(0.595)	-0.055	(0.388)
					To Map					
Constant	-2.566	(1.080)	-9.263	(0.636)	-4.353	(1.345)	-6.960	(0.710)	-2.142	(2.509)
Age	0.049	(0.014)	0.096	(0.008)	0.087	(0.020)	0.100	(0.009)	0.063	(0.032)
HighSchool	3.055	(0.971)	-0.434	(0.648)	-3.446	(1.019)	0.207	(0.915)	14.270	(2.579)
Female	-0.689	(0.782)	-1.089	(0.676)	-0.013	(1.229)	-0.708	(0.807)	8.107	(1.709)
Age $\times$ F	-0.011	(0.011)	0.006	(0.009)	-0.021	(0.018)	0.014	(0.011)	3.348	(0.901)
Age $\times$ HS	-0.031	(0.013)	0.002	(0.008)	0.070	(0.016)	-0.000	(0.012)	0.382	(0.382)
					To Physically frail					
Constant	0.219	(0.606)	-3.842	(1.023)	-6.554	(0.445)	1.249	(2.357)	-9.854	(1.113)
Age	0.005	(0.008)	0.063	(0.014)	0.072	(0.006)	0.928	(0.348)	0.160	(0.017)
HighSchool	0.899	(0.640)	2.259	(0.944)	0.030	(0.471)	-9.488	(2.319)	-2.716	(0.894)
Female	-2.043	(0.500)	-2.746	(1.069)	0.212	(0.367)	-8.291	(1.090)	4.153	(0.965)
Age $\times$ F	0.018	(0.007)	0.022	(0.014)	-0.010	(0.005)	-0.798	(0.345)	-0.070	(0.015)
Age $\times$ HS	-0.014	(0.008)	-0.022	(0.013)	-0.001	(0.006)	3.051	(0.550)	0.043	(0.013)
					To Mentally frail					
Constant	4.789	(1.041)	-0.899	(0.814)	0.397	(1.217)	-5.991	(0.595)	-0.978	(1.363)
Age	-0.039	(0.014)	0.012	(0.010)	0.018	(0.016)	0.070	(0.007)	0.041	(0.017)
HighSchool	1.701	(1.018)	0.048	(0.875)	0.640	(1.063)	0.760	(0.647)	2.717	(1.225)
Female	0.388	(0.993)	3.708	(0.920)	2.310	(0.889)	-0.616	(0.657)	-3.270	(1.224)
Age $\times$ F	-0.011	(0.013)	-0.048	(0.012)	-0.037	(0.011)	0.001	(0.008)	0.034	(0.015)
Age $\times$ HS	-0.020	(0.013)	-0.003	(0.011)	-0.005	(0.013)	-0.009	(0.008)	-0.019	(0.015)
					To Impaired					
Constant	3.121	(1.015)	3.580	(1.723)	1.186	(0.727)	-1.742	(0.761)	-5.805	(0.555)
Age	-0.011	(0.013)	-0.021	(0.022)	-0.001	(0.009)	0.031	(0.009)	0.076	(0.007)
HighSchool	3.637	(0.903)	2.067	(1.930)	0.468	(0.764)	-0.722	(0.785)	1.123	(0.536)
Female	1.744	(1.033)	-0.769	(1.727)	-0.594	(0.657)	-0.005	(0.846)	0.974	(0.534)
Age $\times$ F	-0.030	(0.013)	0.004	(0.022)	0.000	(0.008)	-0.008	(0.010)	-0.019	(0.007)
Age $\times$ HS	-0.043	(0.012)	-0.022	(0.024)	-0.003	(0.010)	0.008	(0.009)	-0.013	(0.007)
					Pr. of each I-ADL by health group					
WALK	0.005	(0.000)	0.016	(0.001)	0.360	(0.005)	0.221	(0.007)	0.878	(0.005)
DRESS	0.020	(0.001)	0.041	(0.002)	0.463	(0.005)	0.262	(0.008)	0.952	(0.003)
BATH	0.004	(0.000)	0.014	(0.001)	0.334	(0.005)	0.359	(0.008)	0.973	(0.003)
EAT	0.002	(0.000)	0.007	(0.001)	0.086	(0.003)	0.182	(0.006)	0.693	(0.007)
BED	0.006	(0.000)	0.026	(0.001)	0.273	(0.004)	0.104	(0.005)	0.843	(0.006)
TOILET	0.008	(0.000)	0.024	(0.001)	0.265	(0.004)	0.089	(0.005)	0.811	(0.006)
MAP	0.012	(0.001)	0.588	(0.008)	0.178	(0.004)	0.802	(0.008)	0.822	(0.007)
PHONE	0.006	(0.000)	0.041	(0.002)	0.052	(0.002)	0.517	(0.008)	0.728	(0.007)
MONEY	0.006	(0.000)	0.063	(0.002)	0.091	(0.003)	0.792	(0.008)	0.851	(0.006)
MED	0.004	(0.000)	0.021	(0.001)	0.039	(0.002)	0.390	(0.008)	0.604	(0.007)
SHOP	0.005	(0.000)	0.041	(0.002)	0.401	(0.005)	0.830	(0.007)	0.977	(0.002)
MEAL	0.004	(0.000)	0.011	(0.001)	0.209	(0.004)	0.690	(0.009)	0.962	(0.003)

Notes: Median and standard deviation (in parenthesis) of the posterior of each parameter in the estimation with five groups. estimates and standard errors. Each column refers to the current health group of the individual while each of the first five panels presents the parameters of the transition to a different health group. The last panel gathers the estimation results of the Bernouilli process that drives I-ADLs.

Table S.13: Effective sample sizes: 4 Groups, $k = 20$

Variable	Healthy		Physically frail		Mentally frail		Impaired	
			To Healthy					
Constant	125	[361]	208	[431]	604	[581]	168	[255]
Age	126	[348]	201	[377]	673	[631]	117	[109]
HighSchool	913	[655]	313	[768]	611	[769]	430	[297]
Female	264	[520]	193	[317]	739	[787]	544	[337]
Age × Female	306	[572]	173	[269]	686	[797]	684	[389]
Age × HighSchool	975	[591]	316	[761]	737	[704]	460	[332]
			To Physically frail					
Constant	94	[545]	144	[918]	956	[580]	61	[467]
Age	97	[545]	127	[727]	942	[685]	57	[500]
HighSchool	692	[760]	572	[798]	632	[859]	124	[410]
Female	105	[483]	325	[616]	437	[564]	157	[462]
Age × Female	115	[569]	256	[580]	567	[600]	160	[488]
Age × HighSchool	746	[707]	531	[748]	682	[931]	98	[394]
			To Mentally frail					
Constant	260	[572]	724	[523]	672	[418]	921	[313]
Age	248	[573]	756	[549]	730	[426]	890	[305]
HighSchool	147	[653]	287	[447]	287	[577]	337	[892]
Female	489	[564]	454	[464]	258	[494]	742	[326]
Age × Female	480	[549]	450	[504]	332	[459]	711	[310]
Age × HighSchool	130	[606]	288	[462]	309	[565]	324	[950]
			To Impaired					
Constant	834	[622]	1961	[1137]	1140	[423]	786	[781]
Age	705	[603]	1986	[979]	1148	[423]	817	[863]
HighSchool	377	[518]	593	[1172]	283	[1052]	967	[645]
Female	194	[698]	296	[904]	326	[464]	436	[437]
Age × Female	185	[689]	336	[953]	350	[420]	389	[471]
Age × HighSchool	303	[586]	603	[1173]	289	[928]	1013	[682]
			Pr. of each I-ADL by health group					
WALK	4272	[1981]	2636	[1444]	2654	[1802]	2911	[1374]
DRESS	2818	[1289]	4161	[1640]	2711	[2042]	2800	[1888]
BATH	4837	[5029]	5344	[2215]	2809	[1942]	3297	[2110]
EAT	5725	[2009]	2634	[1503]	2151	[2860]	3555	[1153]
BED	3506	[1709]	3235	[1390]	4048	[1495]	1768	[3609]
TOILET	4114	[2093]	5863	[2254]	3371	[1686]	2774	[1467]
MAP	4794	[2848]	4461	[2077]	1946	[1059]	1891	[1227]
PHONE	3025	[932]	4088	[1819]	4046	[1331]	4154	[2274]
MONEY	3490	[2007]	3605	[820]	1703	[777]	3139	[1191]
MED	3729	[1564]	4780	[2683]	2325	[1336]	3127	[1765]
SHOP	4575	[2182]	3231	[1433]	1875	[1933]	3301	[2710]
MEAL	5789	[2100]	3204	[2088]	1267	[1710]	4351	[1773]

Notes: We compute the effective sample size (ESS) by dividing the chain in k different subchains and computing the average parameter in each subsample i: $\bar{g}_i$. Then, the effective sample size equals $\frac{ks_N}{s_k}$ where s_k is the standard deviation of $\bar{g}_i$ and s_N is the standard deviation of the parameters over the whole chain. If the model has converged the ESS should increase linearly with the length of the chain. We include the ESS of half of the sample between square brackets.

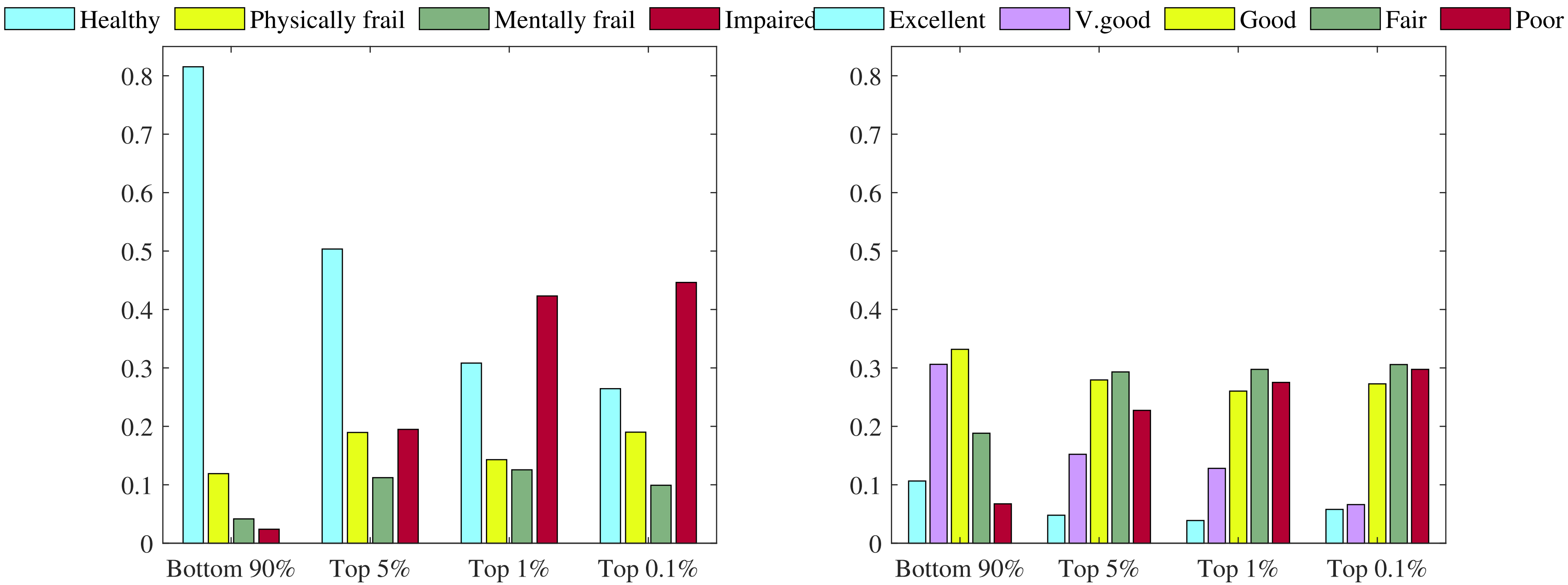
Table S.14: Effective sample sizes: 4 Groups, $k = 40$

Variable	Healthy		Physically frail		Mentally frail		Impaired	
			To Healthy					
Constant	1491	[924]	1663	[767]	937	[568]	670	[236]
Age	1480	[937]	1787	[789]	896	[585]	730	[253]
HighSchool	1881	[782]	1363	[679]	767	[553]	481	[244]
Female	1103	[554]	1320	[793]	896	[537]	447	[287]
Age × Female	1045	[525]	1307	[837]	888	[532]	389	[270]
Age × HighSchool	1892	[809]	1388	[685]	779	[549]	487	[241]
			To Physically frail					
Constant	1040	[967]	1112	[1055]	410	[283]	1005	[544]
Age	997	[934]	1181	[1189]	66	[347]	989	[528]
HighSchool	1562	[791]	1250	[1081]	381	[265]	996	[704]
Female	1398	[556]	1633	[698]	358	[245]	851	[480]
Age × Female	1382	[522]	1679	[714]	66	[281]	844	[470]
Age × HighSchool	1502	[778]	1329	[1133]	397	[269]	968	[660]
			To Mentally frail					
Constant	939	[469]	1070	[713]	954	[574]	682	[327]
Age	961	[470]	1119	[732]	943	[590]	655	[324]
HighSchool	750	[427]	984	[733]	1038	[591]	742	[640]
Female	1222	[738]	1094	[694]	1441	[600]	756	[443]
Age × Female	1245	[738]	1123	[685]	1403	[596]	695	[421]
Age × HighSchool	744	[421]	1006	[753]	1030	[606]	774	[667]
			To Impaired					
Constant	1576	[798]	1434	[934]	949	[799]	1472	[888]
Age	1605	[843]	1414	[979]	937	[801]	1449	[923]
HighSchool	1466	[764]	1097	[742]	1027	[945]	1487	[771]
Female	1586	[592]	1173	[876]	1326	[769]	1405	[783]
Age × Female	1602	[610]	1149	[859]	1253	[762]	1429	[771]
Age × HighSchool	1450	[796]	1081	[727]	1029	[977]	1520	[789]
			Pr. of each I-ADL by health group					
WALK	4586	[3212]	4708	[1969]	5828	[2961]	3906	[2235]
DRESS	5414	[2721]	6543	[2272]	5300	[3051]	4816	[2281]
BATH	8546	[3201]	4542	[2768]	3567	[2175]	5337	[3998]
EAT	5694	[2788]	7084	[3687]	3955	[3117]	8876	[3955]
BED	7843	[2566]	4817	[2110]	7490	[4200]	3719	[2443]
TOILET	4783	[4090]	4107	[2877]	5249	[2016]	3765	[2594]
MAP	5051	[2471]	3913	[2385]	4253	[2836]	7083	[2911]
PHONE	5381	[2414]	4370	[2357]	3082	[2037]	5716	[3867]
MONEY	8617	[4346]	5531	[3403]	2953	[1629]	7841	[3224]
MED	5726	[3077]	9829	[4552]	3761	[2051]	6567	[3550]
SHOP	4404	[1975]	4384	[2333]	2766	[2072]	6115	[4007]
MEAL	5455	[3671]	5234	[2740]	3760	[1677]	5081	[2920]

Notes: We compute the effective sample size (ESS) by dividing the chain in k different subchains and computing the average parameter in each subsample i: $\bar{g}_i$. Then, the effective sample size equals $\frac{ks_N}{s_k}$ where s_k is the standard deviation of $\bar{g}_i$ and s_N is the standard deviation of the parameters over the whole chain. If the model has converged the ESS should increase linearly with the length of the chain. We include the ESS of half of the sample between square brackets.

S.5 Distribution of medical expenses by health group

Figure S.3: Group distribution by medical expenditure quantile



Notes: This graph depicts the proportion of individuals in each health group given a particular medical expenses quantile (Bottom 90%: $< \$8,045$; Top 5%: $> \$13,054$; Top 1%: $> \$46,718$; Top 0.1%: $> \$154,260$). Medical expenditures correspond to out-of-pocket medical expenditures over a 2 year period in dollars of 2018. The graph on the left corresponds to our endogenous classification while the graph on the rights corresponds to self-reported health. RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old out of Medicaid 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. .

S.6 Parameters Table 5

Table S.15: Parameters of the regression in Table 5: No health

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	0.030***	0.030**	0.003***	0.004***	0.006***	0.007***	0.010***
	(0.007)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-2.413***	-2.919***	0.068***	0.082***	0.087***	0.084***	0.130***
	(0.527)	(0.661)	(0.008)	(0.010)	(0.016)	(0.020)	(0.016)
HS×Age	0.042***	0.051***	-0.001***	-0.001***	-0.002***	-0.002***	-0.002***
	(0.007)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-2.063***	-2.781***	-0.123***	-0.146***	-0.090***	-0.082***	0.077***
	(0.476)	(0.599)	(0.007)	(0.009)	(0.015)	(0.018)	(0.014)
Female×Age	0.035***	0.046***	0.002***	0.002***	0.002***	0.001***	-0.001***
	(0.007)	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.021	0.091	-0.181***	-0.237***	-0.336***	-0.367***	-0.580***
	(0.547)	(0.691)	(0.008)	(0.011)	(0.017)	(0.021)	(0.016)
R2	0.007	0.008	0.043	0.051	0.036	0.035	0.059

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) of the following regression:

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

where $y_{i,t}$ is the variable used as a reference, $\mathbf{z}$ includes gender and education, and $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs. ***,**,* indicate significance at the 99%, 95% and 90% confidence level.

Table S.16: Parameters of the regression in Table 5: Self-reported health (2 groups)

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	0.006	0.016	0.001***	0.002***	0.005***	0.006***	0.007***
	(0.008)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-2.448***	-2.857***	0.027***	0.050***	0.102***	0.115***	0.105***
	(0.542)	(0.681)	(0.008)	(0.011)	(0.017)	(0.020)	(0.016)
HS×Age	0.041***	0.049***	-0.000***	-0.001***	-0.002***	-0.002***	-0.002***
	(0.007)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-1.909***	-2.641***	-0.116***	-0.139***	-0.083***	-0.078***	0.081***
	(0.475)	(0.598)	(0.007)	(0.009)	(0.014)	(0.018)	(0.014)
Female×Age	0.032***	0.043***	0.002***	0.002***	0.001***	0.001***	-0.001***
	(0.007)	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Bad	-3.009***	-1.902**	-0.229***	-0.197***	-0.175***	-0.083***	-0.261***
	(0.523)	(0.667)	(0.008)	(0.010)	(0.016)	(0.020)	(0.015)
Bad×HS	1.202***	1.157***	0.009***	0.010***	0.022***	0.015**	0.023***
	(0.131)	(0.160)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)
Bad×Age	0.052***	0.036***	0.003***	0.003***	0.004***	0.002***	0.005***
	(0.007)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Bad×Female	0.252*	0.125	0.005**	0.006**	0.028***	0.031***	-0.028***
	(0.121)	(0.150)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)
Constant	1.378*	0.780	-0.071***	-0.153***	-0.265***	-0.353***	-0.438***
	(0.602)	(0.755)	(0.009)	(0.012)	(0.018)	(0.022)	(0.018)
R2	0.015	0.013	0.060	0.062	0.073	0.062	0.093

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) 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, $\mathbf{z}$ includes gender and education. $\mathbf{d}_{i,t}$ is a dummy variable that takes value one if the individual reports poor or very poor health (Bad). Individuals reporting excellent, very good, good compose the excluded category. ***,**,* indicate significance at the 99%, 95% and 90% confidence level.

Table S.17: Parameters of the regression in Table 5: Self-reported health (5 groups)

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	-0.003	0.014	0.001**	0.002***	0.003***	0.004***	0.006***
	(0.013)	(0.016)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-2.436***	-2.786***	0.013	0.034**	0.091***	0.104***	0.091***
	(0.592)	(0.737)	(0.009)	(0.012)	(0.018)	(0.022)	(0.017)
HS×Age	0.042***	0.049***	-0.000**	-0.001***	-0.002***	-0.002***	-0.002***
	(0.007)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-1.927***	-2.704***	-0.113***	-0.136***	-0.085***	-0.072***	0.088***
	(0.496)	(0.620)	(0.007)	(0.010)	(0.015)	(0.018)	(0.015)
Female×Age	0.030***	0.041***	0.002***	0.002***	0.001***	0.001***	-0.001***
	(0.007)	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Very good	0.489	0.826	-0.015	0.012	-0.066*	-0.059	-0.034
	(0.987)	(1.181)	(0.015)	(0.018)	(0.030)	(0.035)	(0.029)
Good	-0.600	-0.098	-0.056***	-0.043*	-0.142***	-0.155***	-0.111***
	(0.960)	(1.153)	(0.014)	(0.018)	(0.029)	(0.034)	(0.028)
Fair	-1.857	-1.108	-0.176***	-0.163***	-0.278***	-0.214***	-0.258***
	(0.999)	(1.214)	(0.015)	(0.019)	(0.030)	(0.036)	(0.029)
Poor	-5.767***	-3.562*	-0.436***	-0.347***	-0.246***	-0.151***	-0.441***
	(1.136)	(1.450)	(0.017)	(0.023)	(0.035)	(0.043)	(0.033)
Very good×HS	0.022	-0.235	0.007	0.006	0.016	0.009	0.004
	(0.300)	(0.352)	(0.004)	(0.006)	(0.009)	(0.010)	(0.009)
Good×HS	0.100	0.148	0.004	0.007	0.016	0.020*	0.006
	(0.287)	(0.337)	(0.004)	(0.005)	(0.009)	(0.010)	(0.008)
Fair×HS	1.023***	0.953**	0.011*	0.017**	0.035***	0.029**	0.020*
	(0.290)	(0.342)	(0.004)	(0.005)	(0.009)	(0.010)	(0.008)
Poor×HS	1.934***	1.801***	0.029***	0.025***	0.061***	0.051***	0.071***
	(0.317)	(0.386)	(0.005)	(0.006)	(0.010)	(0.011)	(0.009)
Very good×Age	-0.004	-0.007	0.000	-0.000	0.001*	0.001*	0.001
	(0.013)	(0.016)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Good×Age	0.017	0.008	0.001***	0.001*	0.002***	0.002***	0.002***
	(0.013)	(0.015)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fair×Age	0.039**	0.028	0.003***	0.002***	0.005***	0.004***	0.005***
	(0.013)	(0.016)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Poor×Age	0.103***	0.068***	0.007***	0.005***	0.006***	0.004***	0.009***
	(0.015)	(0.019)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Very good×Female	0.101	0.174	0.000	0.000	0.003	-0.007	-0.003
	(0.207)	(0.241)	(0.003)	(0.004)	(0.006)	(0.007)	(0.006)
Good×Female	0.280	0.319	-0.001	0.002	0.011	0.011	-0.011
	(0.204)	(0.238)	(0.003)	(0.004)	(0.006)	(0.007)	(0.006)
Fair×Female	0.241	0.154	0.003	0.005	0.025***	0.024**	-0.023***
	(0.216)	(0.256)	(0.003)	(0.004)	(0.006)	(0.007)	(0.006)
Poor×Female	0.821**	0.763*	0.010**	0.008	0.051***	0.050***	-0.063***
	(0.256)	(0.321)	(0.004)	(0.005)	(0.008)	(0.009)	(0.007)
Constant	1.574	0.587	-0.031*	-0.130***	-0.165***	-0.250***	-0.359***
	(1.003)	(1.220)	(0.015)	(0.019)	(0.030)	(0.036)	(0.029)
R2	0.019	0.015	0.071	0.068	0.090	0.074	0.112

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) 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, $\mathbf{z}$ includes gender and education. $\mathbf{d}_{i,t}$ includes four dummy variables that takes value one if the individual report very good, good, poor or very poor health. Individuals reporting excellent health compose the excluded category. $^{***},^{**},^{*}$ indicate significance at the 99%, 95% and 90% confidence level.

Table S.18: Parameters of the regression in Table 5: ADL: Yes/No

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	-0.025**	-0.005	0.000	0.002***	0.003***	0.005***	0.006***
	(0.008)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-3.096***	-3.400***	0.002	0.029**	0.062***	0.083***	0.074***
	(0.533)	(0.667)	(0.008)	(0.010)	(0.016)	(0.020)	(0.016)
HS×Age	0.050***	0.055***	-0.000	-0.000**	-0.001***	-0.001***	-0.001***
	(0.007)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.495	-1.257*	-0.068***	-0.087***	-0.038**	-0.036*	0.116***
	(0.479)	(0.602)	(0.007)	(0.009)	(0.014)	(0.018)	(0.014)
Female×Age	0.010	0.021*	0.001***	0.001***	0.001**	0.001*	-0.002***
	(0.007)	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ADL>0	-9.861***	-9.286***	-0.446***	-0.437***	-0.159***	-0.070**	-0.368***
	(0.568)	(0.737)	(0.008)	(0.011)	(0.017)	(0.022)	(0.017)
ADL>0×HS	1.469***	1.686***	0.002	0.004	-0.006	-0.004	0.009*
	(0.145)	(0.182)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)
ADL>0×Age	0.148***	0.133***	0.007***	0.007***	0.005***	0.003***	0.007***
	(0.007)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ADL>0×Female	0.653***	0.805***	0.008***	0.012***	0.014***	0.021***	-0.040***
	(0.143)	(0.183)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)
Constant	3.644***	2.474***	-0.000	-0.100***	-0.192***	-0.306***	-0.380***
	(0.574)	(0.716)	(0.008)	(0.011)	(0.017)	(0.021)	(0.017)
R2	0.022	0.017	0.114	0.098	0.099	0.075	0.098

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) 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, $\mathbf{z}$ includes gender and education. $\mathbf{d}_{i,t}$ is a dummy variable that takes value one if the individual presents difficulties with an ADL. Individuals without difficulties compose the excluded category. ***,**,* indicate significance at the 99%, 95% and 90% confidence level.

Table S.19: Parameters of the regression in Table 5: Frailty index quintiles

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	-0.049***	-0.057***	-0.001***	-0.000	0.001***	0.003***	0.003***
	(0.011)	(0.014)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-3.436***	-4.161***	-0.028***	-0.018	0.046**	0.063**	0.031
	(0.555)	(0.696)	(0.008)	(0.011)	(0.017)	(0.020)	(0.016)
HS×Age	0.058***	0.069***	0.000***	0.000	-0.001**	-0.001**	-0.000*
	(0.007)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.662	-1.257*	-0.062***	-0.075***	-0.038**	-0.032	0.133***
	(0.481)	(0.604)	(0.007)	(0.009)	(0.014)	(0.018)	(0.014)
Female×Age	0.009	0.019*	0.001***	0.001***	0.001**	0.001*	-0.002***
	(0.007)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Second	-0.229	-1.491	-0.008	-0.047**	-0.033	-0.083**	-0.055*
	(0.864)	(1.031)	(0.012)	(0.016)	(0.026)	(0.030)	(0.025)
Third	-0.093	-2.730**	-0.019	-0.100***	-0.065**	-0.107***	-0.125***
	(0.842)	(1.002)	(0.012)	(0.015)	(0.025)	(0.029)	(0.024)
Fourth	-0.402	-3.767***	-0.060***	-0.175***	-0.154***	-0.203***	-0.230***
	(0.836)	(1.002)	(0.012)	(0.015)	(0.025)	(0.029)	(0.024)
Top	-10.973***	-12.068***	-0.510***	-0.596***	-0.349***	-0.284***	-0.542***
	(0.798)	(1.003)	(0.011)	(0.015)	(0.024)	(0.029)	(0.024)
Second×HS	-0.129	-0.148	0.001	-0.002	0.007	0.003	0.004
	(0.244)	(0.286)	(0.004)	(0.004)	(0.007)	(0.008)	(0.007)
Third×HS	0.002	-0.028	-0.001	-0.002	-0.002	0.004	-0.009
	(0.235)	(0.275)	(0.003)	(0.004)	(0.007)	(0.008)	(0.007)
Fourth×HS	0.006	0.111	0.002	-0.001	0.011	0.015*	-0.000
	(0.232)	(0.273)	(0.003)	(0.004)	(0.007)	(0.008)	(0.007)
Top×HS	1.726***	2.247***	0.013***	0.022***	0.008	0.012	0.017**
	(0.224)	(0.272)	(0.003)	(0.004)	(0.007)	(0.008)	(0.007)
Second×Age	0.008	0.027	0.000	0.001***	0.001	0.001***	0.001**
	(0.012)	(0.014)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Third×Age	0.011	0.050***	0.000	0.002***	0.001***	0.002***	0.002***
	(0.011)	(0.014)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fourth×Age	0.022	0.072***	0.001***	0.003***	0.003***	0.004***	0.004***
	(0.011)	(0.013)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Top×Age	0.175***	0.186***	0.008***	0.009***	0.008***	0.007***	0.010***
	(0.011)	(0.013)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Second×Female	0.209	0.033	-0.001	-0.002	-0.003	-0.009	-0.007
	(0.173)	(0.202)	(0.002)	(0.003)	(0.005)	(0.006)	(0.005)
Third×Female	0.199	0.130	-0.002	-0.000	0.001	-0.006	-0.013**
	(0.171)	(0.201)	(0.002)	(0.003)	(0.005)	(0.006)	(0.005)
Fourth×Female	0.285	-0.018	-0.001	-0.001	-0.006	0.001	-0.038***
	(0.179)	(0.211)	(0.003)	(0.003)	(0.005)	(0.006)	(0.005)
Top×Female	0.706***	0.813***	0.007**	0.008*	0.016**	0.011	-0.059***
	(0.180)	(0.226)	(0.003)	(0.003)	(0.005)	(0.007)	(0.005)
Constant	4.683***	5.407***	0.050***	0.022	-0.078**	-0.169***	-0.206***
	(0.822)	(1.001)	(0.012)	(0.015)	(0.024)	(0.029)	(0.024)
R2	0.027	0.023	0.126	0.115	0.116	0.096	0.121

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) 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, $\mathbf{z}$ includes gender and education. $\mathbf{d}_{i,t}$ is a vector that includes 4 dummy variables that take value one if the individual belongs to the second, third, fourth of fifth quantile of the frailty index proposed by Braun et al. (2017). Individuals in the quintile with the lowest frailness compose the excluded category. ***,**,* indicate significance at the 99%, 95% and 90% confidence level.

Table S.20: Parameters of the regression in Table 5: I-ADLs dummies

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	-0.031**	0.002	0.001***	0.002***	0.003***	0.005***	0.007***
	(0.011)	(0.014)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-3.544***	-2.845**	0.046***	0.072***	0.039*	0.084***	0.112***
	(0.784)	(0.963)	(0.007)	(0.010)	(0.017)	(0.022)	(0.016)
HS×Age	0.066***	0.057***	-0.001***	-0.001***	-0.001*	-0.001***	-0.002***
	(0.011)	(0.013)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.750	-1.581	-0.083***	-0.073***	-0.050***	-0.050**	0.109***
	(0.666)	(0.813)	(0.006)	(0.008)	(0.014)	(0.018)	(0.014)
Female×Age	0.016	0.026*	0.001***	0.001***	0.001***	0.001**	-0.002***
	(0.009)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Walk	-0.331	1.637	-0.062***	-0.110***	0.086**	0.147**	0.030
	(1.537)	(1.985)	(0.014)	(0.020)	(0.034)	(0.045)	(0.033)
Dress	0.437	0.011	-0.008	-0.004	0.034	-0.019	-0.055
	(1.357)	(1.714)	(0.012)	(0.017)	(0.029)	(0.038)	(0.029)
Bath	-4.727**	-8.337***	-0.141***	-0.147***	-0.035	0.009	-0.058
	(1.653)	(2.120)	(0.015)	(0.021)	(0.036)	(0.048)	(0.035)
Eat	1.944	-6.313*	0.040*	-0.010	0.040	0.099	-0.083
	(2.063)	(2.785)	(0.018)	(0.028)	(0.046)	(0.064)	(0.044)
Bed	-6.543***	3.756	-0.167***	-0.099***	-0.068*	-0.041	-0.070*
	(1.579)	(2.024)	(0.014)	(0.020)	(0.035)	(0.046)	(0.034)
Toilet	-0.366	-0.575	-0.117***	-0.012	0.013	0.008	-0.065
	(1.610)	(2.061)	(0.014)	(0.021)	(0.035)	(0.047)	(0.034)
Map	0.376	-2.127	-0.020*	-0.035**	-0.011	-0.041	0.051*
	(1.031)	(1.274)	(0.009)	(0.013)	(0.022)	(0.028)	(0.022)
Phone	-0.343	-2.188	0.087***	0.069**	0.002	-0.041	-0.170***
	(1.945)	(2.590)	(0.017)	(0.026)	(0.043)	(0.059)	(0.042)
Money	-4.167*	-4.429*	-0.150***	-0.270***	-0.010	-0.013	-0.135***
	(1.680)	(2.197)	(0.015)	(0.022)	(0.037)	(0.049)	(0.037)
Med	-9.306***	-8.128**	-0.188***	-0.102***	0.011	0.046	-0.034
	(1.965)	(2.665)	(0.017)	(0.027)	(0.044)	(0.061)	(0.043)
Shop	5.321***	2.494	0.050***	-0.081***	-0.035	0.017	-0.028
	(1.500)	(1.902)	(0.013)	(0.019)	(0.032)	(0.042)	(0.032)
Meals	-2.612	0.921	-0.169***	-0.111***	0.181***	0.151**	0.000
	(1.825)	(2.341)	(0.016)	(0.024)	(0.040)	(0.052)	(0.039)
Constant	4.956***	3.593***	0.061***	-0.044***	-0.130***	-0.275***	-0.303***
	(0.858)	(1.037)	(0.008)	(0.010)	(0.019)	(0.023)	(0.018)
R2	0.041	0.027	0.315	0.202	0.131	0.088	0.132

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) 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, $\mathbf{z}$ includes gender and education. $\mathbf{d}_{i,t}$ is a dummy variable that takes value one if the individual presents difficulties with an ADL. Individuals without difficulties compose the excluded category. ***,**,* indicate significance at the 99%, 95% and 90% confidence level.

Table S.21: Parameters of the regression in Table 5: 4-I-ADL method

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	-0.027**	-0.019	-0.001***	0.001***	0.003***	0.005***	0.005***
	(0.008)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-2.868***	-3.515***	-0.029***	-0.012	0.031	0.058**	0.037*
	(0.541)	(0.674)	(0.008)	(0.010)	(0.016)	(0.020)	(0.016)
HS×Age	0.047***	0.056***	0.000***	0.000	-0.000*	-0.001***	-0.001***
	(0.008)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.077	-0.784	-0.052***	-0.068***	-0.021	-0.022	0.120***
	(0.483)	(0.604)	(0.007)	(0.009)	(0.014)	(0.018)	(0.014)
Female×Age	0.001	0.014	0.001***	0.001***	0.000	0.000	-0.002***
	(0.007)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(1,0)	0.923	-1.840	-0.020	-0.087***	-0.016	-0.053	-0.107***
	(0.871)	(1.073)	(0.012)	(0.016)	(0.026)	(0.031)	(0.026)
(0,1)	-1.468	-6.627***	-0.154***	-0.290***	-0.032	-0.095**	-0.283***
	(0.952)	(1.215)	(0.013)	(0.018)	(0.028)	(0.036)	(0.028)
(1,1)	-13.015***	-13.261***	-0.584***	-0.628***	-0.088***	-0.012	-0.433***
	(0.697)	(0.932)	(0.010)	(0.014)	(0.021)	(0.028)	(0.021)
(1,0)×HS	-0.273	0.234	0.002	0.005	0.022***	0.008	0.008
	(0.216)	(0.260)	(0.003)	(0.004)	(0.006)	(0.008)	(0.006)
(0,1)×HS	0.542*	1.424***	0.003	0.011*	-0.003	0.002	0.008
	(0.237)	(0.294)	(0.003)	(0.004)	(0.007)	(0.009)	(0.007)
(1,1)×HS	3.177***	3.620***	0.023***	0.030***	0.005	0.014*	0.032***
	(0.179)	(0.233)	(0.003)	(0.004)	(0.005)	(0.007)	(0.005)
(1,0)×Age	-0.001	0.031*	0.000*	0.001***	0.001**	0.002***	0.002***
	(0.012)	(0.015)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(0,1)×Age	0.026*	0.096***	0.002***	0.004***	0.001**	0.002***	0.005***
	(0.012)	(0.016)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(1,1)×Age	0.192***	0.189***	0.009***	0.010***	0.005***	0.003***	0.009***
	(0.009)	(0.012)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(1,0)×Female	0.088	0.216	-0.002	0.001	0.009	0.014*	-0.027***
	(0.204)	(0.249)	(0.003)	(0.004)	(0.006)	(0.007)	(0.006)
(0,1)×Female	0.513*	0.276	0.006	0.009*	0.047***	0.043***	0.003
	(0.228)	(0.284)	(0.003)	(0.004)	(0.007)	(0.008)	(0.007)
(1,1)×Female	0.939***	1.079***	0.010***	0.012***	0.014*	0.019*	-0.059***
	(0.183)	(0.246)	(0.003)	(0.004)	(0.006)	(0.007)	(0.005)
Constant	3.705***	3.384***	0.039***	-0.036**	-0.144***	-0.257***	-0.294***
	(0.593)	(0.736)	(0.008)	(0.011)	(0.018)	(0.022)	(0.017)
R2	0.032	0.025	0.162	0.138	0.123	0.091	0.119

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) 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, $\mathbf{z}$ includes gender and education. $\mathbf{d}_{i,t}$ is a vector that includes 3 dummy variables that take value one if the individual presents difficulties with an ADL but no IADL (1,0), if she struggles with an IADL but no ADL (0,1) and if she has difficulties with at least one of each (1,1). Individuals without difficulties compose the excluded category. ***,**,* indicate significance at the 99%, 95% and 90% confidence level.

Table S.22: Parameters of the regression in Table 5: 2 Health clusters

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	-0.036***	-0.018	-0.000***	0.001***	0.003***	0.005***	0.006***
	(0.011)	(0.013)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-3.980***	-4.645***	-0.031***	-0.011	0.027	0.057**	0.036*
	(0.721)	(0.885)	(0.008)	(0.010)	(0.016)	(0.020)	(0.016)
HS×Age	0.062***	0.073***	0.000***	0.000	-0.000	-0.001***	-0.001***
	(0.010)	(0.012)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.021	-0.939	-0.048***	-0.061***	-0.027	-0.027	0.127***
	(0.643)	(0.793)	(0.007)	(0.009)	(0.014)	(0.018)	(0.014)
Female×Age	0.004	0.017	0.001***	0.001***	0.000*	0.000	-0.002***
	(0.009)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Impaired	-14.303***	-16.501***	-0.528***	-0.576***	-0.041*	0.019	-0.403***
	(0.847)	(1.107)	(0.009)	(0.013)	(0.019)	(0.025)	(0.019)
Impaired×HS	3.449***	4.572***	0.024***	0.037***	0.012*	0.025***	0.037***
	(0.217)	(0.277)	(0.002)	(0.003)	(0.005)	(0.006)	(0.005)
Impaired×Age	0.214***	0.237***	0.008***	0.009***	0.004***	0.002***	0.008***
	(0.011)	(0.015)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Impaired×Female	0.879***	0.874**	0.010***	0.010**	0.013** 0.016*	-0.056***	
	(0.223)	(0.294)	(0.002)	(0.003)	(0.005)	(0.007)	(0.005)
Constant	5.156***	4.090***	0.032***	-0.064***	-0.167***	-0.290***	-0.329***
	(0.767)	(0.940)	(0.008)	(0.011)	(0.017)	(0.021)	(0.017)
R2	0.026	0.023	0.155	0.144	0.114	0.09	0.12

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) 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, $\mathbf{z}$ includes gender and education. $\mathbf{d}_{i,t}$ is a vector that includes the probabilities of being physically frail, mentally frail, or impaired. Healthy is the excluded category. ***,**,* indicate significance at the 99%, 95% and 90% confidence level.

Table S.23: Parameters of the regression in Table 5: 4 Health clusters

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
Age	-0.024*	-0.018	-0.000***	0.001***	0.003***	0.005***	0.005***
	(0.011)	(0.013)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HS	-3.052***	-4.142***	-0.036***	-0.020*	0.030	0.050*	0.023
	(0.726)	(0.892)	(0.007)	(0.010)	(0.016)	(0.020)	(0.016)
HS×Age	0.050***	0.066***	0.001***	0.000	-0.000	-0.001**	-0.001*
	(0.010)	(0.012)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.532	-0.656	-0.037***	-0.050***	-0.019	-0.011	0.132***
	(0.644)	(0.795)	(0.006)	(0.009)	(0.014)	(0.018)	(0.014)
Female×Age	-0.005	0.012	0.001***	0.001***	0.000	0.000	-0.002***
	(0.009)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Physically	1.056	-2.229*	-0.063***	-0.161***	-0.051**	-0.027	-0.164***
	(0.892)	(1.101)	(0.009)	(0.012)	(0.020)	(0.024)	(0.019)
Mentally	-5.734***	-15.558***	-0.296***	-0.509***	-0.070*	-0.125**	-0.352***
	(1.320)	(1.691)	(0.013)	(0.019)	(0.030)	(0.038)	(0.029)
Impaired	-28.442***	-24.854***	-0.777***	-0.833***	0.122***	0.130*	-0.480***
	(1.472)	(2.147)	(0.015)	(0.024)	(0.035)	(0.052)	(0.033)
Physically×HS	0.653**	1.111***	0.005*	0.009**	0.013**	0.006	0.020***
	(0.226)	(0.272)	(0.002)	(0.003)	(0.005)	(0.006)	(0.005)
Mentally×HS	2.121***	4.519***	0.019***	0.041***	-0.002	0.008	0.017*
	(0.332)	(0.421)	(0.003)	(0.005)	(0.007)	(0.009)	(0.007)
Impaired×HS	8.676***	9.251***	0.087***	0.122***	0.040***	0.026*	0.065***
	(0.362)	(0.527)	(0.004)	(0.006)	(0.009)	(0.013)	(0.008)
Physically×Age	-0.001	0.042**	0.001***	0.003***	0.002***	0.002***	0.004***
	(0.012)	(0.015)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mentally×Age	0.086***	0.219***	0.004***	0.008***	0.003***	0.003***	0.006***
	(0.017)	(0.022)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Impaired×Age	0.412***	0.367***	0.013***	0.014***	0.004***	0.003***	0.011***
	(0.018)	(0.028)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Physically×Female	0.349	0.459	-0.001	0.001	0.019** *	0.011	-0.035***
	(0.222)	(0.273)	(0.002)	(0.003)	(0.005)	(0.006)	(0.005)
Mentally×Female	1.217***	1.399**	0.016***	0.018***	0.041***	0.046***	-0.026***
	(0.339)	(0.435)	(0.003)	(0.005)	(0.008)	(0.010)	(0.007)
Impaired×Female	1.230**	0.248	0.006	-0.010	-0.004	-0.003	-0.088***
	(0.386)	(0.591)	(0.004)	(0.007)	(0.009)	(0.014)	(0.008)
Constant	4.248***	3.929***	0.036***	-0.042***	-0.145***	-0.261***	-0.287***
	(0.786)	(0.963)	(0.008)	(0.011)	(0.018)	(0.021)	(0.017)
R2	0.039	0.028	0.262	0.191	0.131	0.099	0.138

Notes: Numbers correspond to the estimates and standard errors (in parenthesis) 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, $\mathbf{z}$ includes gender and education. $\mathbf{d}_{i,t}$ is a vector that includes 3 dummy variables that take value one if the most likely health group is physically frail, mentally frail, or impaired. Healthy is the excluded category. ***,**,* indicate significance at the 99%, 95% and 90% confidence level.

Table S.24: Fraction of explained variance by health classification without covariates

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
SRH (2 groups)	0.7	0.4	1.3	1.1	4.3	3.1	3.8
SRH (5 groups)	1.0	0.6	2.1	1.6	6.1	4.3	5.7
ADL: Yes/No	1.3	0.7	5.8	4.2	7.9	5.2	4.8
Frailty index	1.6	1.1	6.0	4.9	9.3	7.0	6.6
4-I-ADL	2.0	1.3	10.7	8.2	11.0	7.2	7.7
2 groups (mode)	1.7	1.2	10.1	8.6	10.0	6.7	7.8
4 groups (mode)	2.6	1.6	21.4	13.8	12.0	8.1	10.0
Observations	118,706	94,544	118,706	94,544	117.408	93.268	102.292

Notes: Numbers correspond to the R^2 of the following regression:

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

where $y_{i,t}$ is the variable used as a reference and $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs. 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 health group (which we label as Mode).

S.7 Health dynamics

Table S.25: Parameter estimates: self-reported health (2 groups)

Variable	Exc-V.good-Good		Fair-Poor	
	To Exc-V.good-Good			
Constant	-10.310	(0.352)	-6.003	(0.297)
Age	0.109	(0.004)	0.079	(0.004)
High School	-1.435	(0.365)	-0.412	(0.306)
Female	-1.682	(0.336)	-0.353	(0.304)
Age × Female	0.016	(0.004)	0.001	(0.004)
Age × High School	0.012	(0.005)	0.003	(0.004)
	To Fair-Poor			
Constant	-7.675	(0.370)	-7.709	(0.260)
Age	0.086	(0.005)	0.088	(0.003)
High School	0.815	(0.383)	1.213	(0.271)
Female	-1.324	(0.355)	-0.959	(0.272)
Age × Female	0.012	(0.005)	0.007	(0.004)
Age × High School	-0.008	(0.005)	-0.014	(0.003)

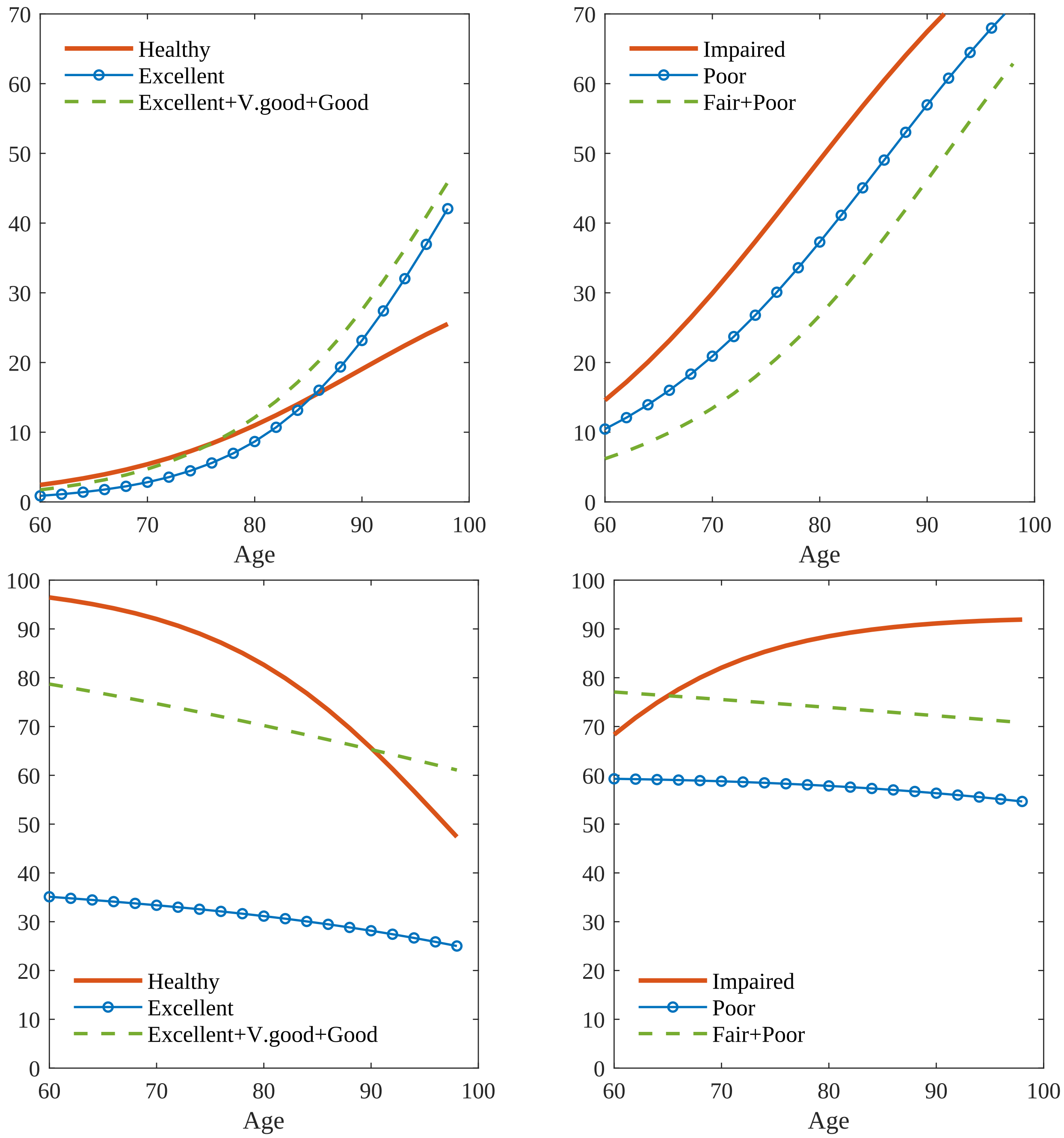
Notes: Parameter estimates and standard errors (in parenthesis) of the logit estimation using two groups of self-reported health. The most healthy group is composed for those individuals who report excellent, very good, or good health; meanwhile, the least healthy one includes those respondents who report poor or very poor health. The second column corresponds to the parameter of an individual who is currently in the healthiest group while the fourth column refers to unhealthy individuals. The first panel shows the estimation results for the transitions to the healthy group whereas the second includes those of the unhealthiest group.

Table S.26: Parameter estimates: self-reported health (5 groups)

Variable	Excellent		V.Good		Good		Fair		Poor	
					To Excellent					
Constant	-11.029	(1.222)	-9.068	(0.773)	-7.446	(0.725)	-6.546	(1.007)	2.149	(2.534)
Age	0.123	(0.015)	0.117	(0.010)	0.111	(0.010)	0.116	(0.014)	0.027	(0.033)
HighSchool	-1.704	(1.240)	-2.279	(0.793)	-0.280	(0.760)	3.014	(1.094)	-1.437	(2.710)
Female	-2.587	(1.089)	-2.374	(0.670)	-0.412	(0.722)	1.011	(1.105)	1.732	(2.679)
Age × F	0.028	(0.014)	0.026	(0.009)	0.002	(0.010)	-0.015	(0.015)	-0.020	(0.035)
Age × HS	0.011	(0.016)	0.022	(0.010)	0.004	(0.010)	-0.036	(0.015)	0.018	(0.035)
					To V.Good					
Constant	-10.463	(1.221)	-10.201	(0.668)	-8.018	(0.488)	-5.456	(0.567)	-3.770	(1.095)
Age	0.117	(0.015)	0.112	(0.008)	0.100	(0.006)	0.084	(0.008)	0.083	(0.015)
HighSchool	-0.442	(1.239)	-2.430	(0.691)	-1.666	(0.506)	-0.718	(0.574)	1.340	(1.150)
Female	-2.556	(1.087)	-1.685	(0.612)	-1.590	(0.466)	0.011	(0.573)	2.082	(1.183)
Age × F	0.028	(0.014)	0.015	(0.008)	0.014	(0.006)	-0.004	(0.008)	-0.033	(0.016)
Age × HS	-0.003	(0.016)	0.023	(0.009)	0.016	(0.006)	0.010	(0.008)	-0.009	(0.015)
					To Good					
Constant	-10.509	(1.252)	-9.849	(0.676)	-9.537	(0.449)	-6.937	(0.416)	-2.591	(0.693)
Age	0.121	(0.016)	0.110	(0.009)	0.107	(0.006)	0.089	(0.005)	0.055	(0.009)
HighSchool	2.309	(1.271)	-0.290	(0.698)	-0.699	(0.471)	-0.596	(0.427)	0.456	(0.670)
Female	-1.591	(1.137)	-1.565	(0.621)	-1.578	(0.439)	-0.785	(0.422)	-0.430	(0.698)
Age × F	0.016	(0.015)	0.015	(0.008)	0.014	(0.006)	0.006	(0.005)	0.000	(0.009)
Age × HS	-0.029	(0.016)	-0.001	(0.009)	0.004	(0.006)	0.005	(0.006)	-0.006	(0.009)
					To Fair					
Constant	-8.684	(1.369)	-7.133	(0.734)	-7.733	(0.470)	-8.588	(0.376)	-5.989	(0.456)
Age	0.106	(0.018)	0.084	(0.009)	0.090	(0.006)	0.100	(0.005)	0.084	(0.006)
HighSchool	4.448	(1.430)	0.885	(0.762)	0.355	(0.490)	1.274	(0.390)	1.085	(0.464)
Female	-1.801	(1.316)	-1.278	(0.694)	-1.442	(0.460)	-1.053	(0.390)	-0.645	(0.475)
Age × F	0.021	(0.017)	0.013	(0.009)	0.013	(0.006)	0.008	(0.005)	0.002	(0.006)
Age × HS	-0.049	(0.018)	-0.007	(0.010)	-0.005	(0.006)	-0.016	(0.005)	-0.012	(0.006)
					To Poor					
Constant	-3.444	(1.717)	-4.807	(0.938)	-4.710	(0.597)	-5.924	(0.425)	-6.479	(0.394)
Age	0.050	(0.022)	0.068	(0.012)	0.067	(0.008)	0.077	(0.006)	0.081	(0.005)
HighSchool	3.334	(1.803)	2.731	(0.996)	0.464	(0.622)	1.565	(0.440)	1.253	(0.413)
Female	-2.130	(1.725)	-0.201	(0.947)	-1.116	(0.601)	-1.330	(0.446)	-0.758	(0.415)
Age × F	0.024	(0.022)	-0.002	(0.012)	0.009	(0.008)	0.012	(0.006)	0.004	(0.005)
Age × HS	-0.035	(0.023)	-0.028	(0.013)	-0.002	(0.008)	-0.018	(0.006)	-0.012	(0.005)

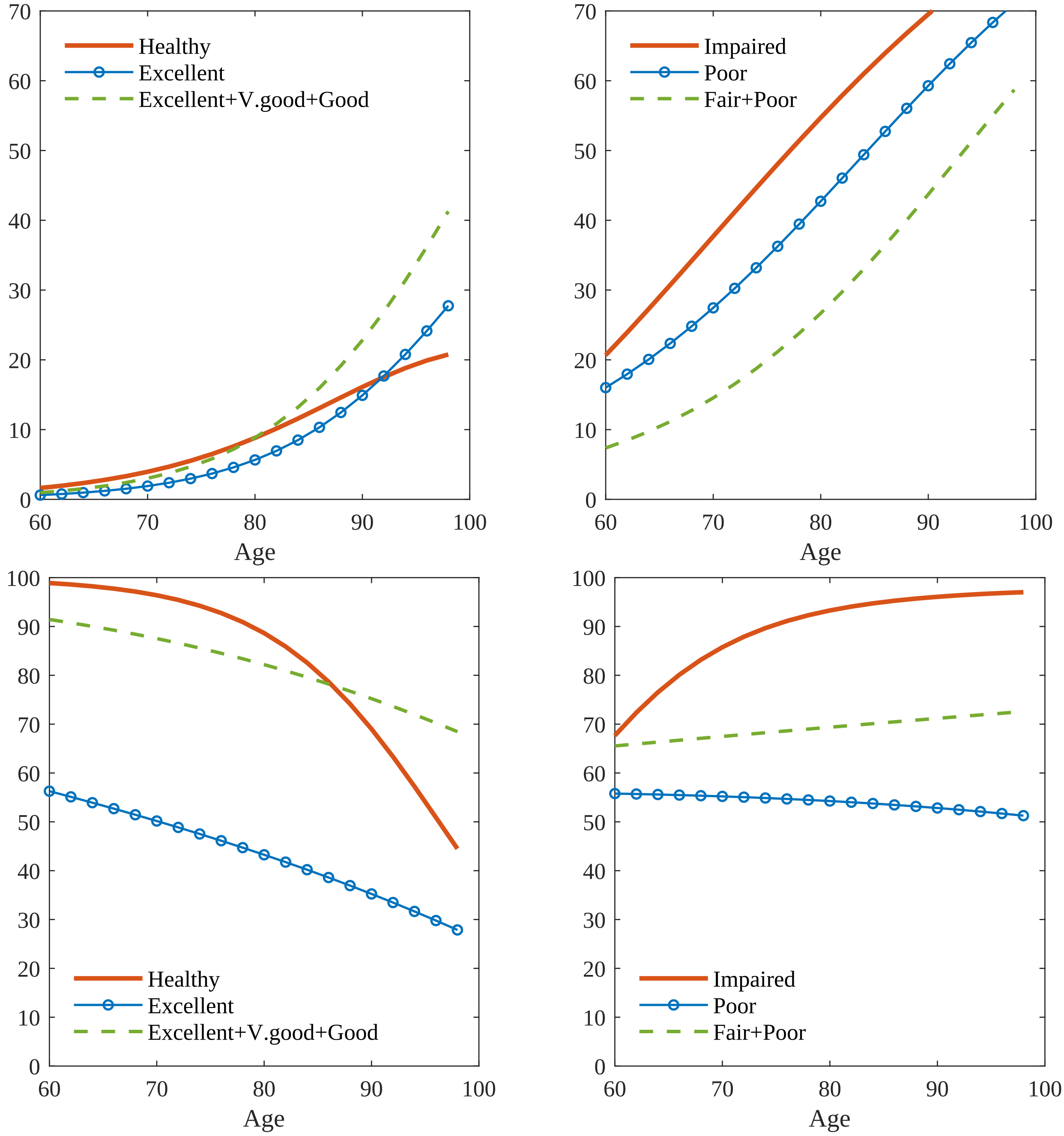
Notes: Parameter estimates and standard errors (in parenthesis) of the logit estimation using the five groups of self-reported health. Each column refers to the current health group of the individual while each panel presents the parameters of the transition to a different health group. For instance the fourth column of the third row of the first panel (-2.279) indicates that high school graduates who currently report very good health are less likely to report excellent in the next wave compared with dropouts.

Figure S.8: Transition to death and persistence: Dropout Males



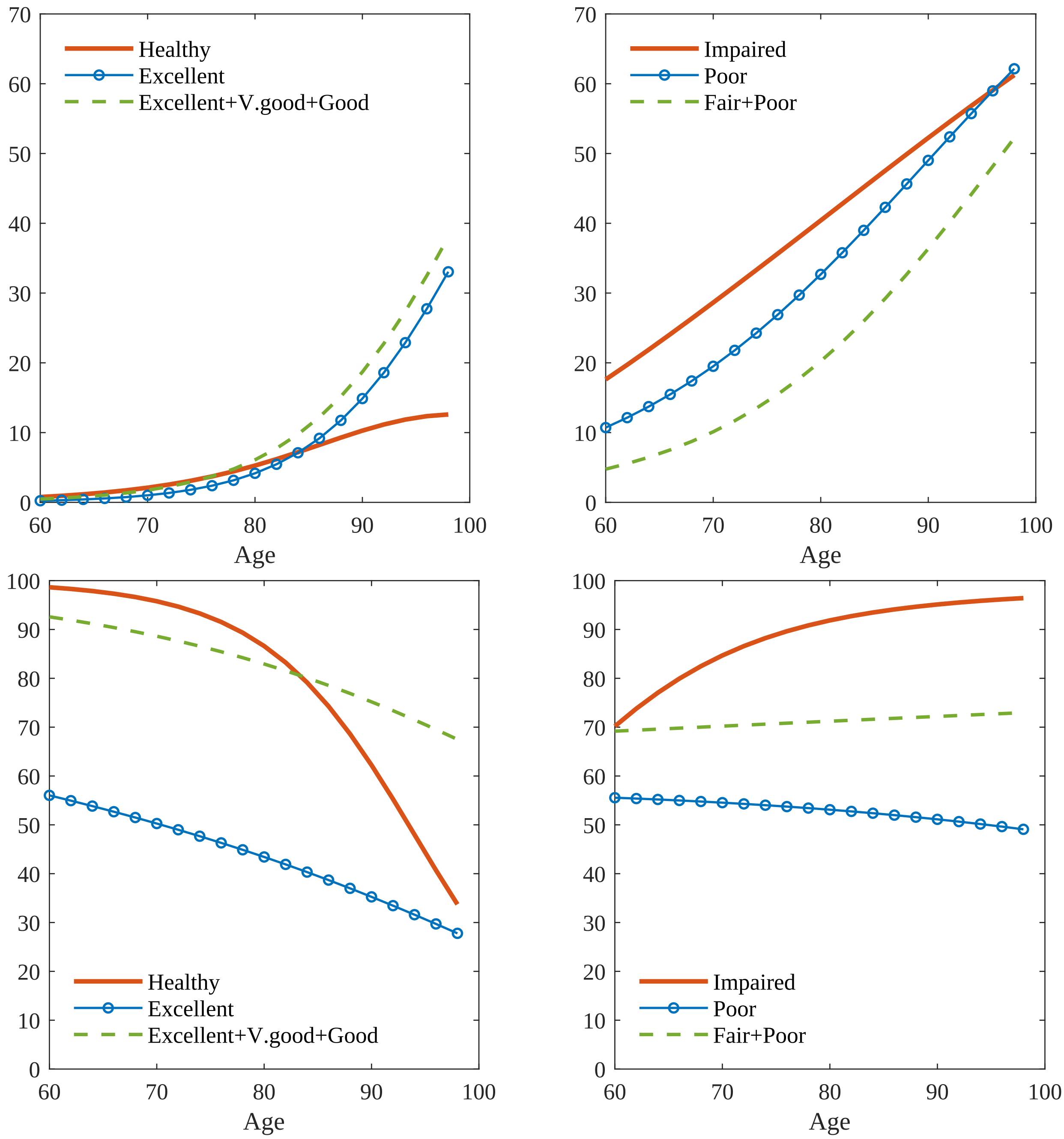
Notes: Upper plots: probability of dying per health group. Lower plots: Probability of maintaining the same health state. RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations).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

Figure S.9: Transition to death and persistence: High-school Males



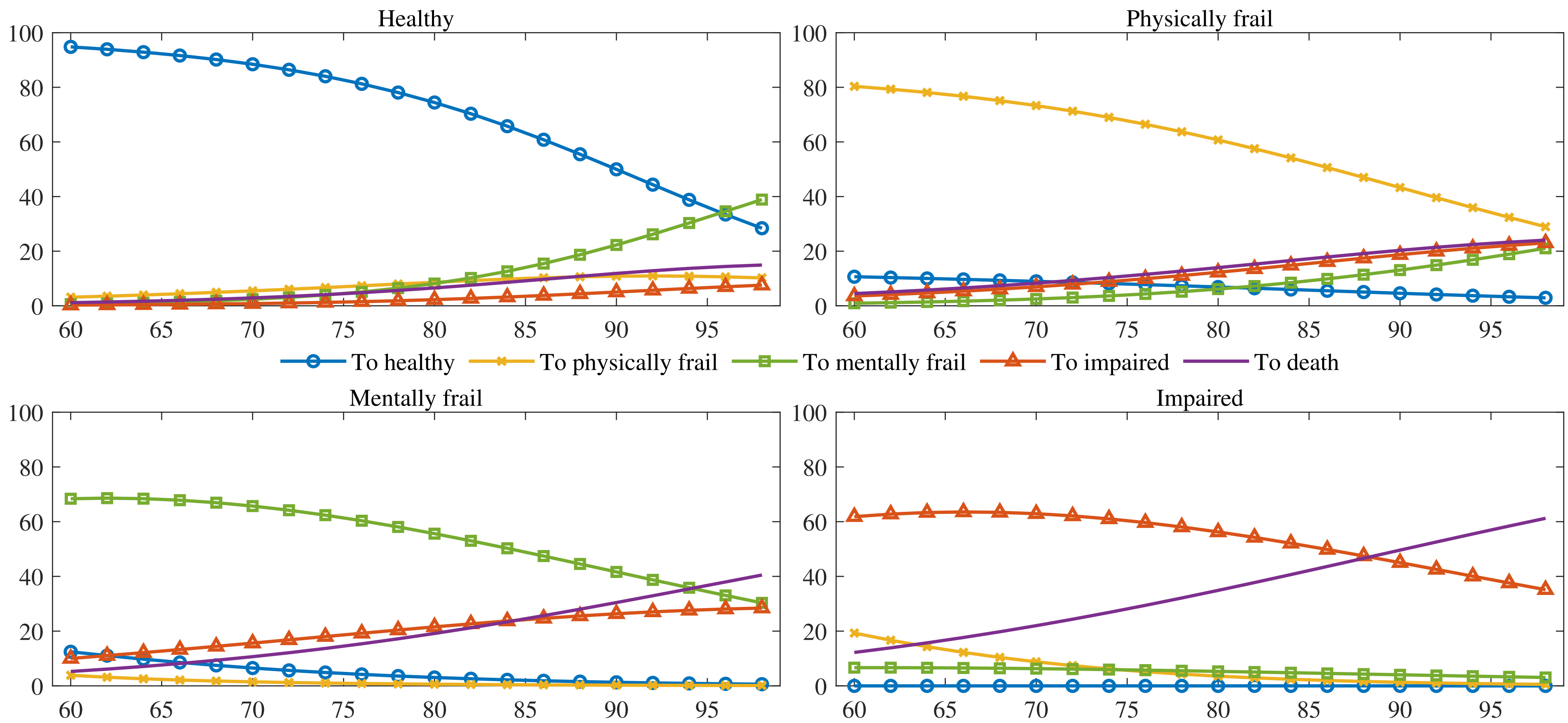
Notes: Upper plots: probability of dying per health group. Lower plots: Probability of maintaining the same health state. RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations).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

Figure S.10: Transition to death and persistence: High-school females



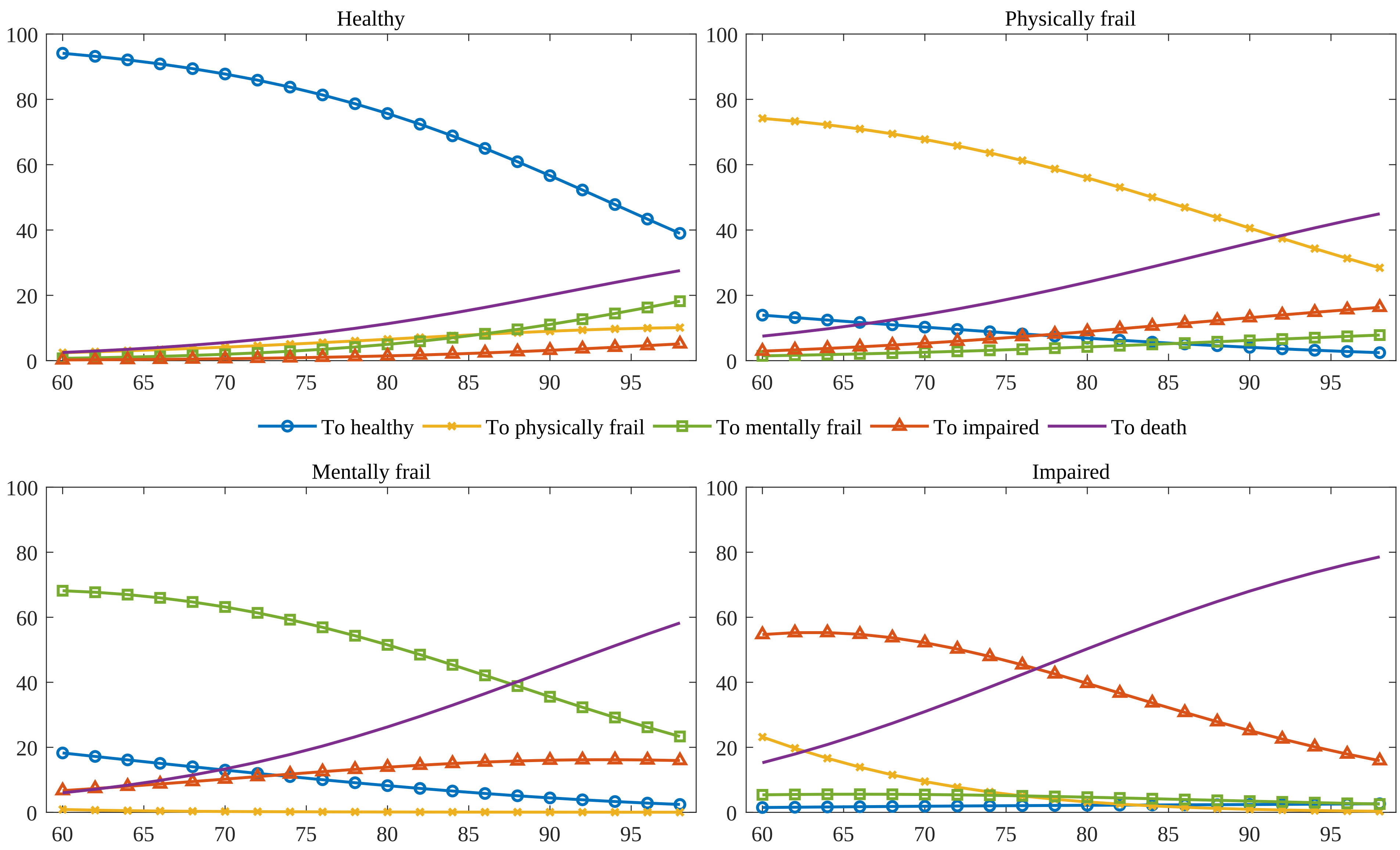
Notes: Upper plots: probability of dying per health group. Lower plots: Probability of maintaining the same health state. RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations).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

Figure S.4: Transitions by group as individuals age



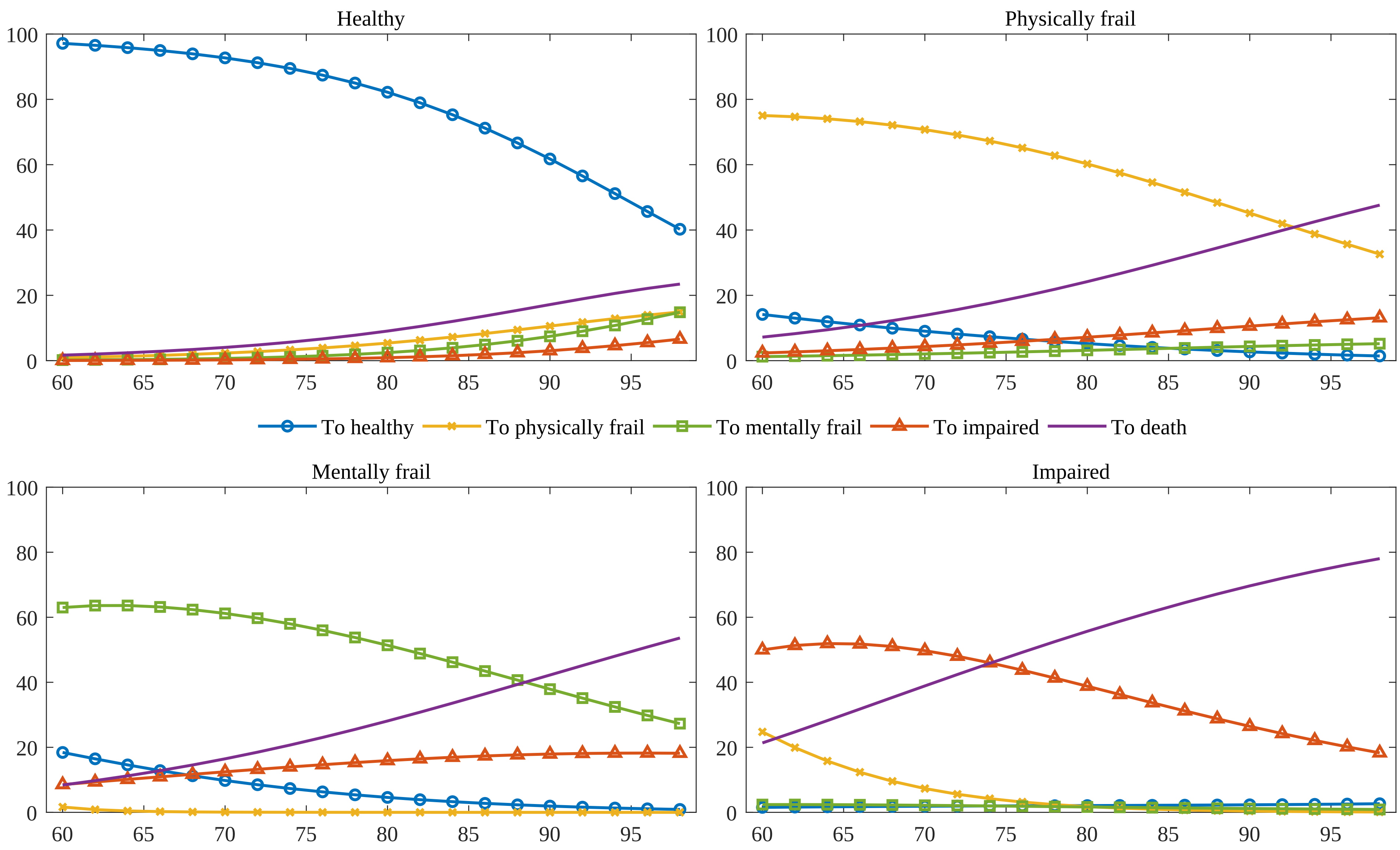
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals 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 S.5: Transitions by group as individuals age: Dropouts Males



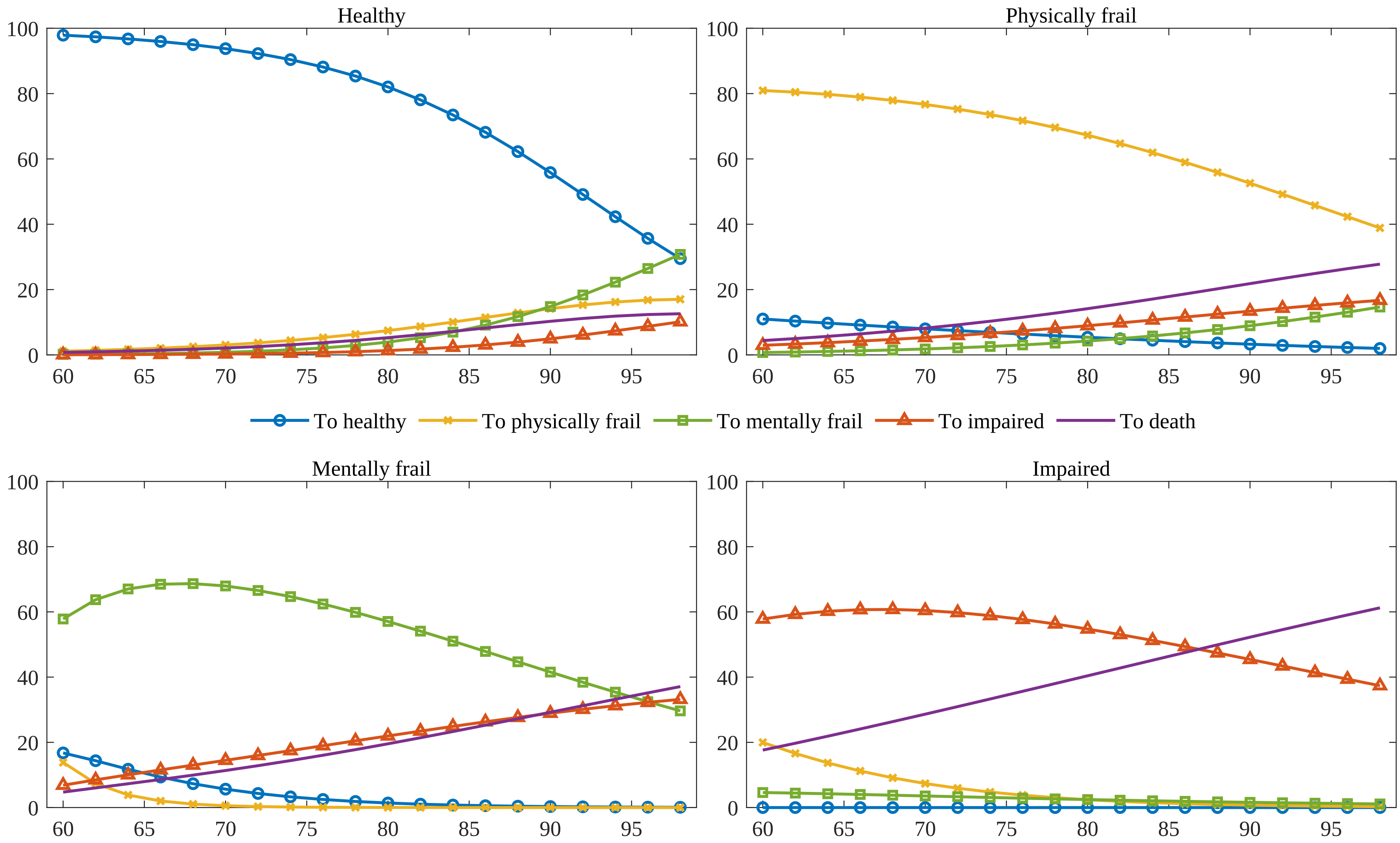
Notes: These figures depict the probability of moving to each health group according to the current health group. The units of the y-axis are percentage points and those of the x-axis are years. This graph corresponds to male dropouts.

Figure S.6: Transitions by group as individuals age: High-school males



Notes: These figures depict the probability of moving to each health group according to the current health group. The units of the y-axis are percentage points and those of the x-axis are years. This graph corresponds to male high-school graduates.

Figure S.7: Transitions by group as individuals age: High-school females



Notes: These figures depict the probability of moving to each health group according to the current health group. The units of the y-axis are percentage points and those of the x-axis are years. This graph corresponds to female high-school graduates.

S.8 Details of the replication exercise

In this appendix we describe the details of the replication exercise of De Nardi et al. (2010). The original model is estimated using a two-step strategy. In the first step the authors estimate the health transitions and medical expenses parameters without using the structural model and, in the second step, they estimate the vector of preference parameters and Medicaid generosity using the method of simulated moments. We reestimate the first step parameters and take their estimated preference parameters in their benchmark model. Following the original paper, we estimate the model using the AHEAD part of the Health and Retirement Study and we select only single retired individuals interviewed between 1993 to 2006.

S.8.1 Health transitions

We estimate health transition probabilities (including death) as a multinomial logistic regression. Future health is estimated as a multinomial logistic function of sex, a quadratic polynomial in age, sex interacted with age, a quadratic polynomial in permanent income rank, permanent income rank interacted with age, current health, current health interacted with age, and current health interacted with permanent income.

We estimate the model using two levels of self-reported health (excellent-very good-good and poor-very poor health) and the mode of our estimated classification when we set the number of groups to two (healthy and impaired). We face one complication during the estimation. While the HRS is a biannual survey, the model period is one year. For this purpose, we treat health between survey years as a latent variable and estimate the model using an integrated likelihood. For example if an individual is observed in 1996 and 1998 in good health and bad health, respectively, the likelihood will be given by the sum of the two possible transitions, properly weighted.

In comparison to the original paper, we obtain life expectancies across the permanent income distribution very close but not exactly equal. In our model, individuals in the bottom and top quintile of the permanent income distribution expect to live 11.3 and 13.7 years conditional on being alive at the age of 70 (versus 11.1 and 14.7 in the original paper). Transitions when using self-reported health are also very close to reported health transition in De Nardi et al. (2006).

S.8.2 Medical expenses

Following De Nardi et al. (2010), the mean of the logarithm of medical expenses is modeled as a function of a quadratic polynomial in age, sex, sex interacted with age, current health status, health status interacted with age, a quadratic in the individual's permanent income range, and permanent income range interacted with age. Following the code in Bailey Jone's website, we estimate these profiles using a fixed-effects estimator. The variance of the medical expense shock is modeled with the same variables and functional form as the mean and decomposed into a persistent and a transitory component. For the readers convenience, we rewrite the equation (6) of the original paper:

$$\ln m_t = m(g,h,I,t) + \sigma(g,h,I,t) \times (\xi_t + \zeta_t), \xi \sim N(0,\sigma_\xi^2)$$

$$\zeta_t = \rho\zeta_{t-1} + \epsilon_t, \epsilon \sim N(0,\sigma_\varepsilon^2)$$

Parameters in $m(g,h,I,t)$ are estimated by running an OLS regression using fixed effects. Then, we use the residuals of the regression first to estimate the parameters in $\sigma(g,h,I,t)$, and then we use a Kalman filter to estimate the persistent and transitory components. In contrast, De Nardi et al. (2010) use the persistence and transitory component found in French and Jones (2004), who follow a similar procedure as we do.

We introduce two differences in the medical expenditure process with respect to the original paper. First, (De Nardi et al., 2010) replace individuals having expenses of zero to \$250. In order not to overestimate the role of medical expenses, we allow agents in the model to receive zero expenses. We parametrize the probability of zero expenses by estimating a logit model of a dummy that takes the value of one in case there are zero expenses against the same set of covariates used to estimate the mean of medical expenses. Second, in order to capture the tail risk of medical expenses from nursing homes stays, we compute an order logit of nursing home residency against the same set of individuals. We introduce a nursing home shock in the model such that individuals are exposes to the average medical expenditure of individuals in nursing homes conditional on age, gender, PI and health status. None of the introduced differences affect the results but provide a better fit of the observed medical expenses.

Table S.27 presents the persistence and variance of innovations in the original paper, our estimation using self-reported health and using our two groups classification. Table S.27 shows that our estimation procedure delivers slightly more persistent medical expenditure shocks. Moreover, the parameters driving the residual medical expenses if we use self-reported health

Table S.27: Persistence and Variance of Innovations to Medical Expenses (Variances as Fractions of Total Cross-Sectional Variance)

Parameter	Variable	Estimate
	De Nardi et al. (2010)	
ρ	Autocorrelation, persistent part	0.922
σ_ϵ^2	Innovation variance, persistent part	0.050
σ_ξ^2	Innovation variance, transitory part	0.665
	Self-reported Health	
ρ	Autocorrelation, persistent part	0.935
σ_ϵ^2	Innovation variance, persistent part	0.064
σ_ξ^2	Innovation variance, transitory part	0.515
	Two Groups: mode	
ρ	Autocorrelation, persistent part	0.932
σ_ϵ^2	Innovation variance, persistent part	0.067
σ_ξ^2	Innovation variance, transitory part	0.512
	Two Groups: probabilities	
ρ	Autocorrelation, persistent part	0.932
σ_ϵ^2	Innovation variance, persistent part	0.067
σ_ξ^2	Innovation variance, transitory part	0.523
	Four Groups: mode	
ρ	Autocorrelation, persistent part	0.933
σ_ϵ^2	Innovation variance, persistent part	0.066
σ_ξ^2	Innovation variance, transitory part	0.511

or our classification, almost coincide. Not surprisingly, as health enters the conditional mean and variance.

Then, we replicate Figure 3 of the original paper by simulating medical expense histories for the AHEAD birth year cohort whose members were aged 72-76 (with an average age of 74) in 1996. We begin the simulations with draws from the joint distribution of age, health, permanent income, and sex observed in 1996. As figure S.12 shows, we find very similar mean expenditures for the simulated individuals. The similarity is due to the small proportion of variance of medical expenses explained by any measure of health.

S.9 The role of medical expenses

Finally, we replicate the main result of De Nardi et al. (2010) to assess the accuracy of our replication. For this purpose we select the AHEAD birth year cohort whose members were aged 72-76 (with an average age of 74) in 1996. Then, we compute the optimal saving decisions, simulate the model, and compare the resulting asset accumulation profiles in a world without medical expenses to the asset profiles generated by the baseline model. Figure S.13 shows that we are able to reproduce the main figure of the paper and that there are little differences across health classifications.

References

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

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

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.

De Nardi, M., French, E., and Jones, J. B. (2006). Differential mortality, uncertain medical expenses, and the saving of elderly singles. Technical report, National Bureau of Economic Research.

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.

French, E. and Jones, J. B. (2004). On the distribution and dynamics of health care costs. *Journal of Applied Econometrics*, 19(6):705–721.

Figure S.11: Estimated death probabilities across health classifications for females in the median of the permanent income distribution

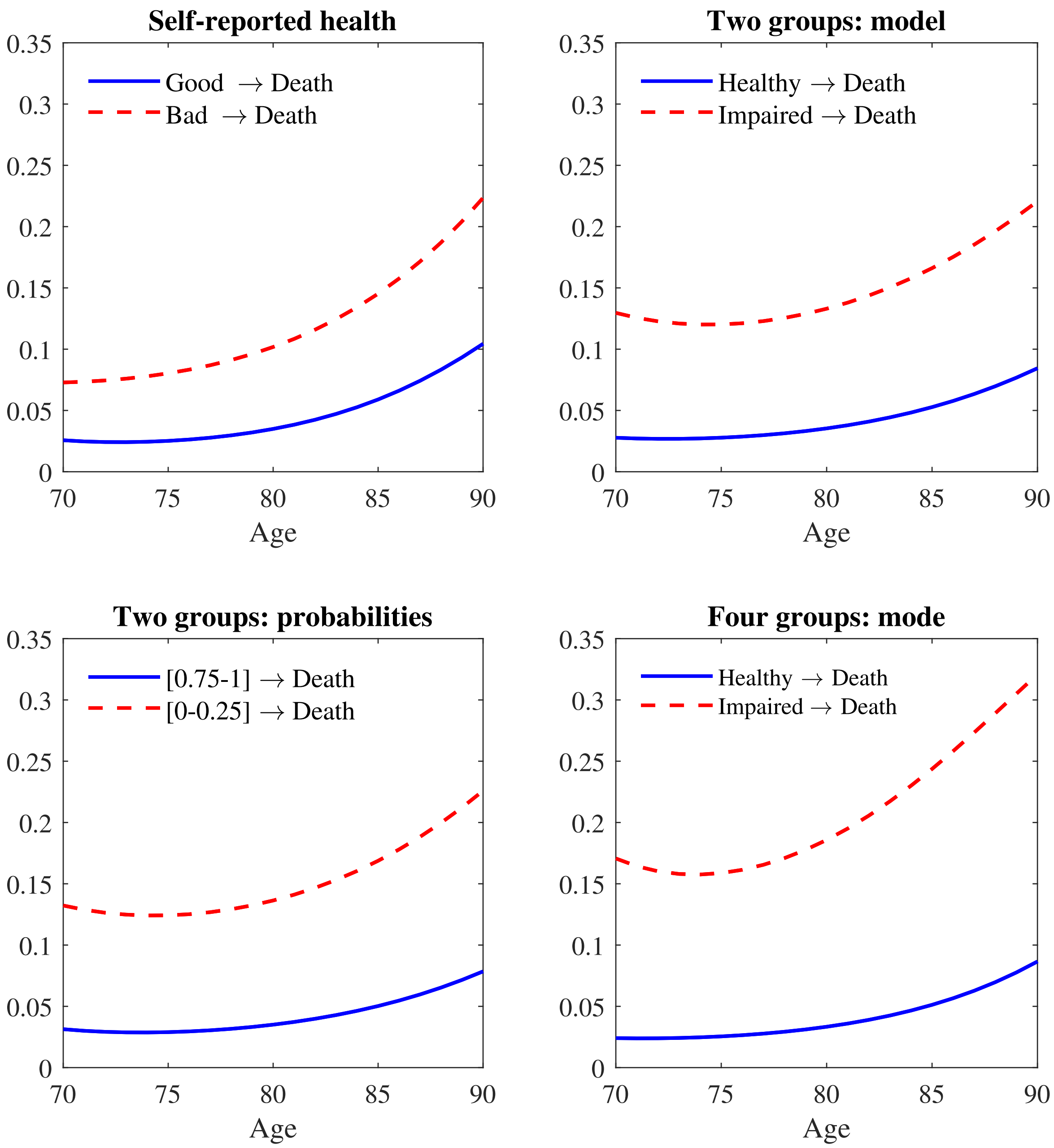


Figure S.12: Average medical expenses, by permanent income quintile

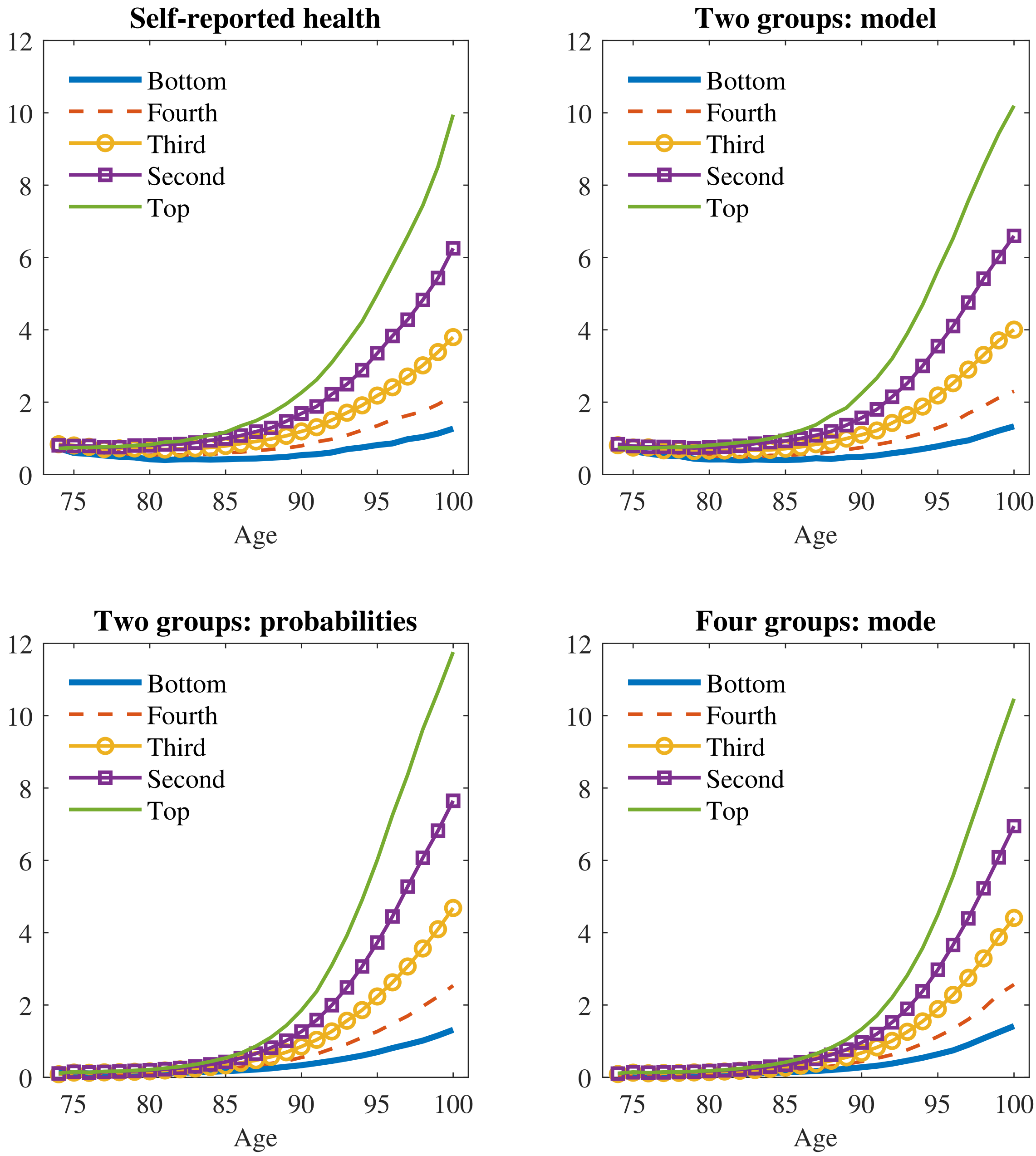


Figure S.13: Median assets by cohort and permanent income quintile: baseline model (dashed lines) and model with no medical expenses (solid lines).