Online Appendix to 'The Evolution of Health over the Life Cycle'

Roozbeh Hosseini
University of Georgia
Federal Reserve Bank of Atlanta
roozbeh@uga.edu

Karen A. Kopecky
Federal Reserve Bank of Atlanta
Emory University
karen.kopecky@atl.frb.org

Kai Zhao
University of Connecticut
kai.zhao@uconn.edu

February 22, 2021

Contents

A PSID 4

A.1 PSID sample and list of deficit variables . . . 4

A.1.1 Regression tables for construction of SIO in PSID . . . 6

A.2 Additional tables and figures For PSID sample . . . 9

A.2.1 Dispersion in frailty by gender . . . 9

A.2.2 Persistence of frailty by gender . . . 9

A.3 Additional figures and tables on comparisons between frailty and SRHS in PSID 12

A.3.1 Health dynamics by gender . . . 12

A.3.2 Dispersion by gender . . . 13

A.3.3 Health dynamics: robustness . . . 13

A.3.4 Persistences . . . 15

A.3.5 Predicting becoming SSDI beneficiary . . . 20

A.4 Probability Transition Matrices for Frailty Variants in PSID . . . 32

B MEPS 35

B.1 MEPS sample and list of deficit variables . . . 35

B.1.1 Regression tables for construction of SIO index in MEPS . . . 35

B.2 Facts on frailty dynamics over the life cycle in MEPS . . . 39

B.3 Comparison of frailty index to SRHS in MEPS . . . 42

B.3.1 Health dynamics . . . 42

B.3.2 Dispersion . . . 42

B.3.3 Persistence . . . 42

B.3.4 Predicting medical expenditures . . . 42

B.4 Comparisons to other variants of frailty in MEPS . . . 49

C HRS 54

C.1 HRS sample and list of deficit variables . . . 54

C.1.1 Regression tables for construction of SIO in HRS . . . 54

C.2 Facts on frailty dynamics over the life cycle in HRS . . . 59

C.3 Comparison of the frailty index to SRHS in HRS . . . 62

C.3.1 Health dynamics . . . 62

C.3.2 Dispersion . . . 62

C.3.3 Persistence . . . 62

C.3.4 Predicting mortality . . . 62

C.3.5 Predicting nursing home entry . . . 65

C.3.6 Predicting becoming SSDI beneficiary . . . 74

C.3.7 Predicting retirement . . . 78

C.4 Comparison to other variants of frailty in HRS . . . 82

C.5 Comparing frailty index across HRS, PSID, and MEPS . . . 88

C.5.1 Comparing wave-to-wave changes . . . 95

D Supplemental Material For Estimation of Frailty Process 98

D.1 Computation of Cohort-adjusted Empirical Moments . . . 98

D.2 Mortality Probit 102
D.3 Additional Estimation Results 103

E Frailty process estimation using variants of the frailty index 106
E.1 Frailty with SRHS included as a deficit variable 106
E.2 Frailty using first principal component weighting of deficits 110
E.3 Frailty using subjective measures to inform the weights 114

A PSID

In this section we briefly describe the PSID sample we use for our analysis and report the list of deficit variables we use for construction of our health measures. We also include additional details regarding baseline frailty and other measures.

A.1 PSID sample and list of deficit variables

The PSID is a longitudinal panel survey of U.S. families that was started in 1968. Its disability and health-related questions were expanded in 2003 to include questions on specific medical conditions, ADL's and IADL's. We rely on these questions to construct our health measures. For this reason we restrict our sample to the 2003 to 2017 period. We also restrict the sample to household heads and their spouses who are at least 25 years of age. Our sample consists of 96,759 observations of 19996 individuals (9198 men and 10,798 women). The PSID is biennial over this period.

We report of all deficit variables we use in PSID in Table 1. These variables can be divided into fours categories. There are 13 variables that are related to difficulty in activities of daily living (ADL) and instrumental ADL (IADL). These are the variables that can detect deterioration as well as improvement in health condition. There are 12 variables that are related to past and present diagnostics and serious medical conditions. These are variables that can only detect deterioration in health due to the framing of the question ("Have you ever had..."). We include BMI of higher than 30, whether the person ever smoked and whether the person is currently smoking as they are important indicators of health conditions in past and present.

Finally, as described in the paper, we also construct variants of frailty that uses self-reported health stats. As we describe in the text, we aggregate this variable into binary response. We denote 'fair' and 'poor' as indicator of *bad health* (and hence assign value of 1).

Overall, we use 28 variables to construct baseline frailty index. Also, frailty with SRHS includes all those 28 variables together with self-reported health status. FPC uses all 29 variables in Frailty with SRHS. The difference is that instead of equal weighting it uses weights associated to the first principal component of these 29 variables. These (normalized) weights are reported in Table 2.

To highlight the importance that first principal component optimal weighting scheme assigns to each variable, the deficit variables in this table are ordered according to their weight in the first principal component. The first column in the table reports FPC weights (corresponding to the first principal component). To ease comparison, we report the equal weights corresponding to frailty with SRHS and baseline frailty in the second and third column, respectively.

The FPC index gives the highest weights to ADL and IADL deficits. Ten highest ranked deficits (according to their weights) all belong to this group. The next highest ranked variable is the self-reported health status (whether the report is 'fair' or 'poor'). On the other hand, the weights on sever medical conditions and diagnostics are much lower.

Table 2 also includes weights associated with SIO. The idea behind the construction of this index is to let subjective health measures inform us about weights that is put on

Variable	Value
Some difficulty with ADL/IADLs:	
Eating	Yes=1, No=0
Dressing	Yes=1, No=0
Getting in/out of bed or chair	Yes=1, No=0
Using the toilet	Yes=1, No=0
Bathing/showering	Yes=1, No=0
Walking	Yes=1, No=0
Using the telephone	Yes=1, No=0
Managing money	Yes=1, No=0
Shopping for personal items	Yes=1, No=0
Preparing meals	Yes=1, No=0
Heavy housework	Yes=1, No=0
Light housework	Yes=1, No=0
Getting outside	Yes=1, No=0
Ever had one of following conditions:	
High Blood Pressure	Yes=1, No=0
Diabetes	Yes=1, No=0
Cancer	Yes=1, No=0
Lung disease	Yes=1, No=0
Heart disease	Yes=1, No=0
Heart attack	Yes=1, No=0
Stroke	Yes=1, No=0
Arthritis	Yes=1, No=0
Asthma	Yes=1, No=0
Loss of memory or mental ability	Yes=1, No=0
Psychological problems	Yes=1, No=0
Other serious, chronic condition	Yes=1, No=0
BMI ≥ 30	Yes=1, No=0
Has ever smoked	Yes=1, No=0
Smokes now	Yes=1, No=0
The following variable is only used in construction of Frailty-with-SRHS and the FPC index	
Self-reported health of 'fair' or 'poor'	Yes=1, No=0

Table 1: Health variables used to construct frailty in PSID. For "ever had one of following conditions", we make the following adjustment to the raw data: if in any wave an individual has a positive answer to any the conditions, we assign the value of 1 to that conditions when calculating frailty in all future waves.

each deficit. This is inspired by Blundell et al. (2020) who construct an index, similar to our SIO, to use as a measure of health for estimating the causal impact of health on labor supply. In construction of their index, they only use deficits related to severe health and past diagnostics. In construing SIO, we depart from them and use all available deficit variables in constructing the index.

Variable	FPC (1)	SIO (2)	Frailty with SRHS (3)	Baseline Frailty (4)
Difficulty with bathing/showering	0.056	0.038	0.034	0.036
Difficulty with some difficulty Walking	0.056	0.206	0.034	0.036
Difficulty with dressing	0.055	0.007	0.034	0.036
Difficulty getting outside	0.054	0.013	0.034	0.036
Difficulty getting in/out of bed or chair	0.054	0.066	0.034	0.036
Difficulty shopping for personal items	0.053	-0.010	0.034	0.036
Difficulty preparing meals	0.052	-0.018	0.034	0.036
Difficulty with light housework	0.050	-0.004	0.034	0.036
Difficulty using the toilet	0.048	-0.040	0.034	0.036
Difficulty with heavy housework	0.046	0.232	0.034	0.036
Self-reported health of 'fair' or 'poor'	0.042	n.a.	0.034	n.a.
Difficulty managing money	0.037	0.090	0.034	0.036
Ever had loss of memory or mental ability	0.036	0.055	0.034	0.036
Difficulty using the telephone	0.035	0.029	0.034	0.036
Ever had arthritis	0.034	0.035	0.034	0.036
Difficulty eating	0.033	0.048	0.034	0.036
Ever had heart disease	0.029	0.023	0.034	0.036
Ever had stroke	0.029	0.025	0.034	0.036
Ever had high blood pressure	0.027	0.020	0.034	0.036
Ever had other serious/chronic condition	0.026	0.032	0.034	0.036
Ever had psychological problems	0.025	0.026	0.034	0.036
Ever had heart attack	0.024	0.031	0.034	0.036
Ever had diabetes	0.023	0.028	0.034	0.036
Ever had lung disease	0.022	0.020	0.034	0.036
Ever had cancer	0.016	0.013	0.034	0.036
Ever had asthma	0.014	0.007	0.034	0.036
BMI >= 30	0.010	0.008	0.034	0.036
Ever smoked	0.010	0.004	0.034	0.036
Smokes now	0.004	0.019	0.034	0.036

Table 2: Weights assigned to each deficit in each variations of frailty. Column (1) show first principal component (FPC) weights. Column (2) shows SIO weights. Column (3) is the equal weight frailty index that includes SRHS. Column (4) is equal weight frailty index that excludes SRHS (baseline frailty index). Deficit variables are sorted according to FPC weights. All weights are based on PSID data on household heads and their spouses ages 25 and older.

A.1.1 Regression tables for construction of SIO in PSID

The construction closely follows Blundell et al. (2020) and it is done in two steps. In the first step we use principal component analysis to combine three subjective health measures: self-reported health status, whether health limits daily activities, and whether health limits ability to work. The principal component weights are reported in Table 3. These are the weights associated with the first principal component of these three variables.

In the second step we run the resulting first principal component on a list of objective

Health limits activities	0.349
Health limits work	0.345
Self-reported health of 'fair' or 'poor'	0.306

Table 3: Principal component weights for first stage calculation of SIO index (PSID).

deficit variables (equation (1) in the paper). The results of this regressions are reported in Table 4. As we describe in the paper, the normalized value of these coefficients are used as weights for construction of the SIO index. These weights are reported in Table 2.

Variable	Value	Variable	Value
Ever had cancer	0.021*** (0.002)	Difficulty with walking	0.328*** (0.003)
Ever had diabetes	0.044*** (0.002)	Difficulty with getting outside	0.021*** (0.005)
Ever had high blood pressure	0.031*** (0.001)	Difficulty with using the toilet	-0.064*** (0.006)
Ever had arthritis	0.056*** (0.001)	Smokes now	0.030*** (0.002)
Ever had psychological problems	0.050*** (0.002)	Ever smoked	0.006*** (0.001)
Ever had lung disease	0.032*** (0.002)	Difficulty with preparing meals	-0.029*** (0.005)
Ever had stroke	0.040*** (0.003)	Difficulty with shopping for personal items	-0.016*** (0.005)
Ever had heart disease	0.036*** (0.002)	Difficulty with managing money	0.144*** (0.004)
Ever had heart attack	0.050*** (0.003)	Difficulty with using the telephone	0.047*** (0.006)
Ever had asthma	0.011*** (0.002)	Difficulty with heavy housework	0.369*** (0.002)
Ever had loss of memory or mental ability	0.087*** (0.003)	Difficulty with light housework	-0.007 (0.004)
Difficulty with bathing, showering	0.060*** (0.005)	Other serious, chronic condition	0.041*** (0.001)
Difficulty with dressing	0.012** (0.005)	BMI ≥ 30	0.012*** (0.001)
Difficulty with eating	0.076*** (0.006)	cons	0.007*** (0.001)
Difficulty with getting in, out of bed or chair	0.105*** (0.004)		
Observations	92,693		
Adjusted R^2	0.682		

Table 4: SIO index second step regression (PSID). Note: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

A.2 Additional tables and figures For PSID sample

A.2.1 Dispersion in frailty by gender

Figure 1 shows box plots of frailty by 10 year age group, histogram of frailty and fraction of zeros by 10 year age group for men and women. Overall, and consistent with summary stats in Table 1 of main text, there are very little differences across gender. However, women have slightly lower fraction at zeros (specially at younger ages) and higher non-zero values. This results in slightly higher mean as well as higher standard deviation for women (as reported in Table 1 of the paper).

A.2.2 Persistence of frailty by gender

Tables 5 and 6 report transition probabilities across frailty quintiles by 25 year age groups for men and women, respectively. These transition probabilities are very similar. Both sets of transition probabilities exhibit the same three patterns as those of the overall sample. Frailty is highly persistent for both men and women. However, its slightly more persistent for women relative to men. Persistence decline with age. However, this decline is more pronounced for women relative men. Finally, within each age group and for both men and women persistence is highest at the bottom and top quintiles.

			Ages 25 to 49		
	Bottom	2nd	3rd	4th	Top
Bottom	79.7	13.5	4.3	2.1	0.4
2nd	4.1	72.2	15.8	6.5	1.3
3rd	1.4	8.6	66.4	18.2	5.4
4th	0.2	2.9	9.2	67.3	20.5
Top	0.0	0.1	1.0	9.3	89.6
			Ages 50 to 74		
	Bottom	2nd	3rd	4th	Top
Bottom	79.7	15.3	3.6	1.1	0.4
2nd	4.7	66.8	21.8	5.2	1.5
3rd	0.2	7.2	65.9	23.8	3.0
4th	0.0	0.9	7.0	69.1	23.0
Top	0.1	0.1	0.4	9.0	90.4
			Ages 75 and older		
	Bottom	2nd	3rd	4th	Top
Bottom	73.4	18.0	5.6	1.2	1.8
2nd	3.9	57.3	23.6	10.3	4.8
3rd	1.2	6.9	50.9	27.5	13.4
4th	0.4	0.7	12.0	55.4	31.5
Top	0.0	1.0	2.4	14.4	82.2

Table 5: Two-year transition probabilities (%) across frailty quintiles for men: 25-49 year-olds, 50-74 year-olds and ages 75 and older constructed using the PSID sample. Rows sum to one.

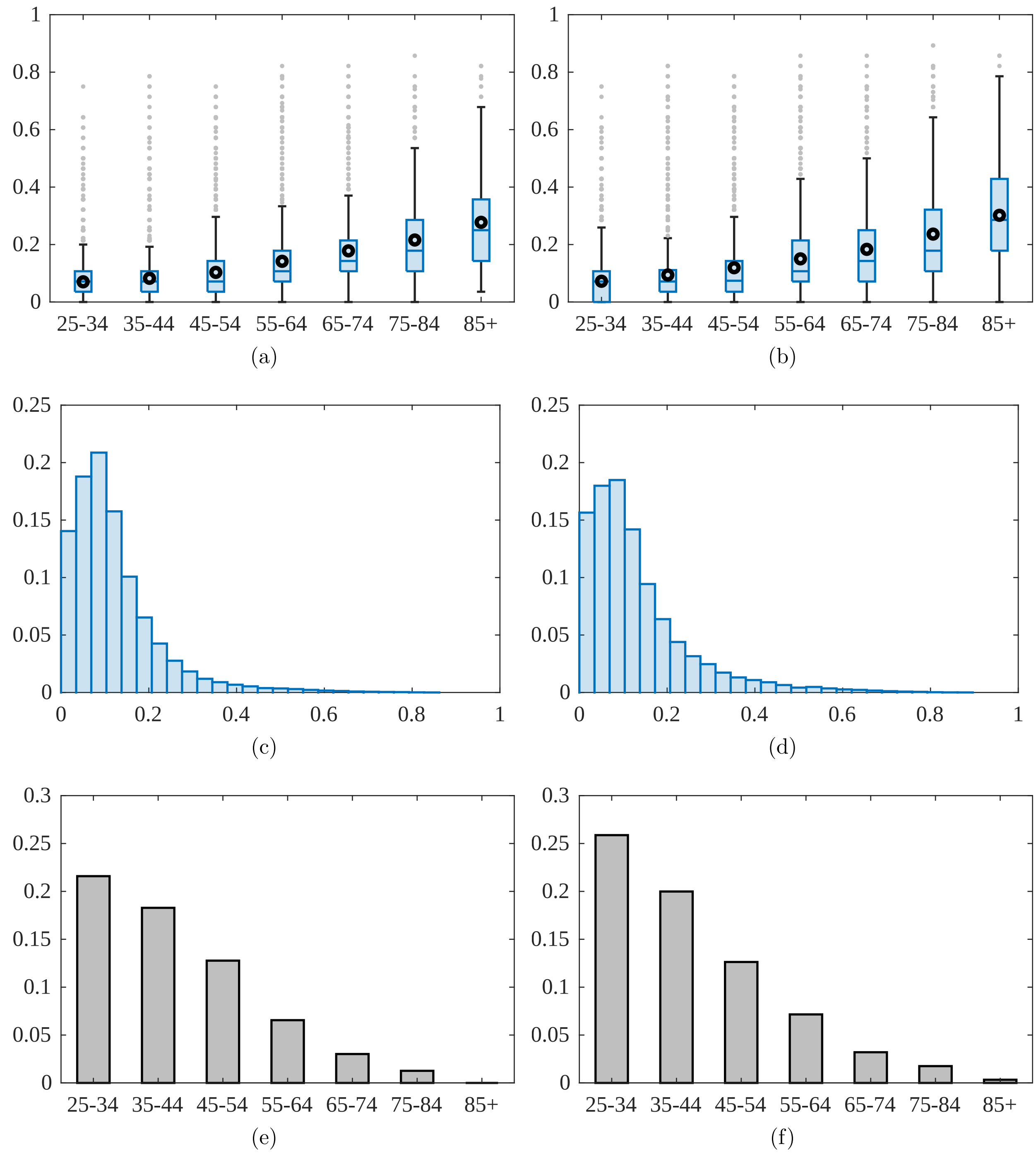


Figure 1: Panel (a) and (b) are box plots of frailty by 10-year age groups for men and women, respectively, in the PSID sample. The bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The middle lines are medians and the open circles are means. The upper whiskers extend to 1.5 times the inner quartile range. The lower whiskers extend to the minimum value of frailty in the age group. The dots are data points that lie beyond the whiskers. Panel (c) and (d) are histograms showing the cross-sectional distribution of frailty for men and women, respectively, in the sample. Panel (e) and (f) are fraction of zeros by by 10-year age groups for men ad women, respectively.

	Bottom	2nd	3rd	4th	Top
Ages 25 to 49					
Bottom	81.2	13.6	3.5	1.3	0.4
2nd	4.0	70.9	19.1	4.9	1.1
3rd	0.4	9.5	66.3	19.8	4.1
4th	0.0	1.0	9.0	69.9	20.1
Top	0.1	0.2	0.6	7.6	91.5
Ages 50 to 74					
	Bottom	2nd	3rd	4th	Top
Bottom	79.1	16.7	3.1	0.8	0.3
2nd	5.9	68.2	20.1	4.8	0.9
3rd	0.2	7.6	68.2	20.9	3.1
4th	0.1	0.8	8.5	68.2	22.4
Top	0.0	0.2	1.0	10.0	88.7
Ages 75 and older					
	Bottom	2nd	3rd	4th	Top
Bottom	69.2	19.0	6.7	3.6	1.5
2nd	6.1	56.5	24.9	7.8	4.7
3rd	1.2	10.1	45.8	28.8	14.1
4th	0.0	2.1	16.2	54.0	27.6
Top	0.0	0.3	2.3	15.2	82.2

Table 6: Two-year transition probabilities (%) across frailty quintiles for women: 25-49 year-olds, 50-74 year-olds and ages 75 and older constructed using the PSID sample. Rows sum to one.

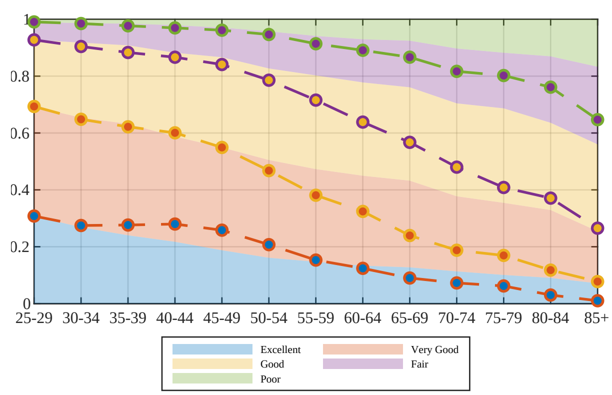
A.3 Additional figures and tables on comparisons between frailty and SRHS in PSID

A.3.1 Health dynamics by gender

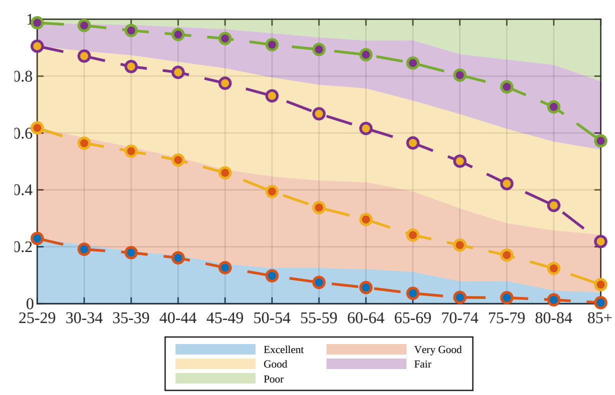
In the paper we compare the evolution of frailty (a continuous variable) with categorical SRHS by partitioning individuals into categories using frailty. Here, we construct these partitions separately for men and women and report our findings. The construction is very similar to what is outline in the main text. The only difference is that determination of frailty cut offs (and hence defining frailty health categories) is done with gender.

The results are reported in Figure 2. Panel (a) and (b) show how the distribution of each health measure evolves with age, for men and women respectively, when individuals are assigned to health categories using the method described above. The share of individuals in each self-reported health category at age 25 to 29 and corresponding frailty cutoffs are reported in the table below the figures.

The pictures are very similar. We first note that the self-reported health category evolves very similarly over the age groups for men and women. Although, there are fewer reports of 'excellent' and 'very good' health among women at age 25 to 29. However, as individuals get older, the decline in self-reported health categories and frailty categories are very similar across men and women. For both men and women, the decline in 'excellent'/'very good' shares and rise in 'fair'/'poor' shares happens more rapidly with age when health is measured by frailty instead of SRHS, both for men and women. This is especially true after age 49.



(a)



(b)

	excellent	very good	good	fair	poor
% of SRHS in age 25 - 29, men	30.8	38.5	23.4	6.3	1.0
Frailty range, men	[0, 0.036)	[0.036, 0.071)	[0.071, 0.143)	[0.143, 0.286)	[0.286, 1]
% of SRHS in age 25 - 29, women	22.9	38.9	28.7	8.2	1.2
Frailty range, women	[0, 0.000)	[0.000, 0.071)	[0.071, 0.143)	[0.143, 0.286)	[0.286, 1]

Figure 2: Panel (a) and (b) are the distributions of health status by age for men and women, respectively. The colored areas show the fraction of individuals by SRHS at each age. The dashed lines show the fraction of individuals by the corresponding health measure category at each age. Source: authors' calculation using PSID.

A.3.2 Dispersion by gender

Figure 3 shows the box plots for the distribution of frailty by self-reported health categories for each 10 year age group, for men (panel (a)) and women (panel (b). Bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points that are not outlines. The middle line indicates median. There is not noticeable difference between men and women here. The big picture here is that for, both men and women, there is considerable variations in frailty at each group and within each self-reported health category. Also, the overall patterns of distribution of frailty by age and self-reported health category seem to be very similar across men and women.

A.3.3 Health dynamics: robustness

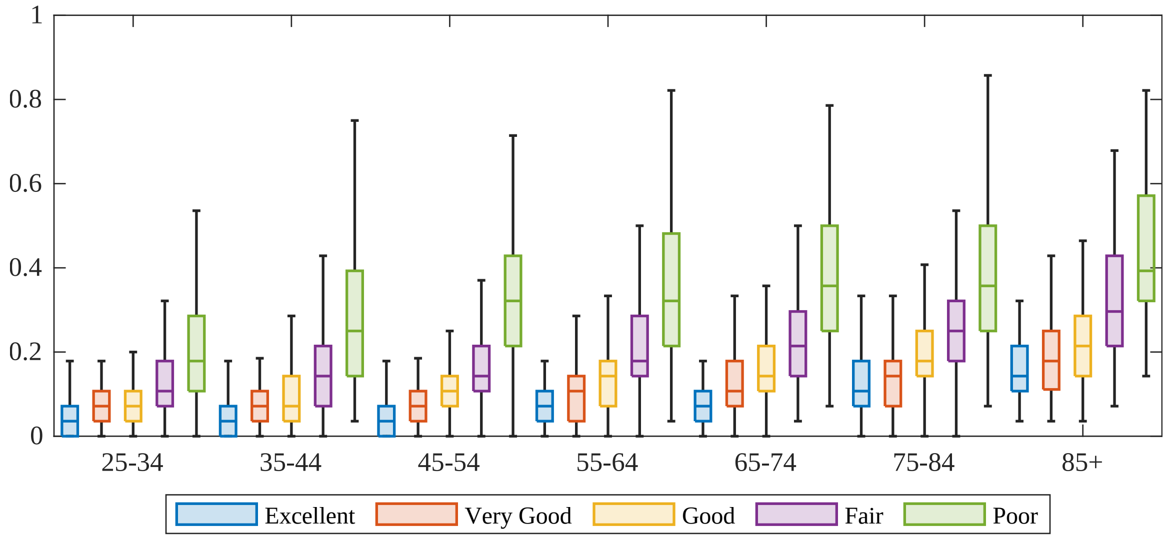
In the paper we draw a comparison between the frailty index and self-reported health status by keeping track of the distribution of health status over the life cycle. We do this by partitioning the frailty distribution into 5 health status categories. To do this, we find frailty cut offs that partition the cross-sectional distribution of frailty at age 25 to 29 into 5 groups. We choose these cut-offs such that the share of these frailty categories match exactly with the self-reported health status categories. We then define frailty status of 'excellent' at any age to be a frailty level below the lowest cut-off. Frailty status of 'very good' is any frailty level between the lowest and second lowest cut off, and so on. We report these cut-offs and resulting area plot in Figure 2 of the main text.

We observe that the fraction of 'excellent' and 'very good' frailty categories decline faster than that of self-reported health status. Also, the fraction of 'fair' and 'poor' frailty categories grow faster than than that of self-reported health status.

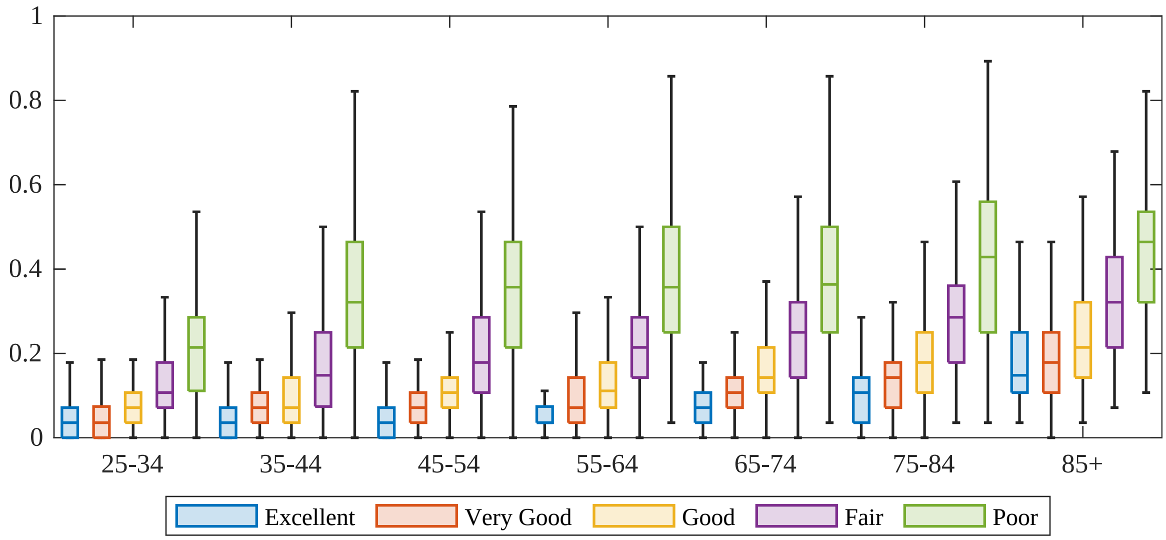
There are two concerns here. One is the random nature of the samples. For example, is the fraction of 'excellent' frailty category different from fraction 'excellent' self-reported health category in a statistical meaningful and significant way? The other concern is whether the findings are robust to the choice of age group for constructing the cut offs. For example, what would be the outcome if we use age group 50 to 55 as reference for choosing the frailty cut offs.

To address the first concern, we construct confidence intervals. To do this we draw 1000 independent joint samples of age, frailty, and self-reported health status (similar to one would do for a bootstrap). For each sample we do the following. First, we calculate the fraction of each self-reported health categories for each of the draws. Next, for each of the draws, we find frailty cut offs such that they partition frailty distribution into 5 health categories. The cut offs are such that the share of each frailty health category at age 25 to 29 matches exactly with the share of corresponding self-reported health category (within the same draw). Finally, we use the calculate cut offs to find share of each frailty health category at all age groups. We plot the results in Figure 4. The blue line shows the average fraction each self-reported health category age (across all the draws). The red line shows the corresponding average for frailty health categories. The shaded areas are 95 percent confidence intervals.

Notice that the share of frailty health categories 'excellent' and 'very good' are noisier and have larger confidence intervals. Nevertheless, fraction of 'excellent' and 'very good' frailty



(a)



(b)

Figure 3: Box plots of frailty by age and SRHS catgories for man (panel (a)) and women (panel (b)).Bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points that are not outliers. The middle line indicates median. Data source: PSID.

categories are clearly below that of self-reported health status. The pattern for fraction of 'fair' and 'poor' frailty categories are more precisely calculated and show similar pattern. The share of 'fair' and 'poor' health categories are clearly above that of self-reported health categories. Health declines faster when measured by frailty and this is not driven by noise.

To check sensitivity with respect to references age, we repeat the procedure above, but choose frailty cut offs to match the distribution of self-reported health status at age 50 to 55. The results are reported in Figure 5. There are three noticeable differences. One, the averages of fraction frailty health category and self-reported health categories cross at age 50 to 55. This is, of course, by construction. Two, The fractions for frailty health categories are more precisely calculated (confidence bands are narrower related to Figure 4). This is due to fact that at older age there is more variations in frailty. This is leads to less noise in calculation of the cut offs. Three, share of 'fair' and 'poor' frailty categories track those of self-reported health categories closely up to age 50 (for 'fair') and 75 (for 'poor').

However, overall these Figures convey similar message. Health declines faster when measured by frailty. This is not driven by the choice of reference age for calculating the cut offs. However, the rate at which the share of 'fair' and 'poor' frailty categories rise appear to be sensitive to the cut offs.

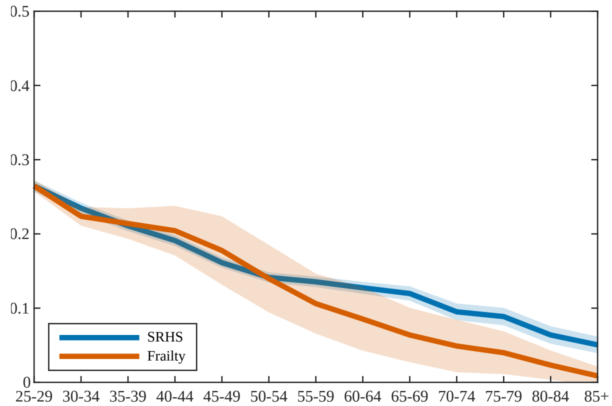
A.3.4 Persistences

As part of comparison between frailty and self-reported health status, we look at the transition probabilities across different health categories. This allows us to also compare persistence of of self-reported health categories with that of frailty health categories. To this end we report transition probabilities across 5 health categories of self-reported health status (Table 7) and frailty (Table 8) by 25 year age group. The frailty health categories are defined by cut offs that equate the share of each health categories across frailty and self-reported health status at age 25 to 29 . These cut offs are reported in Figure 2 of the paper.

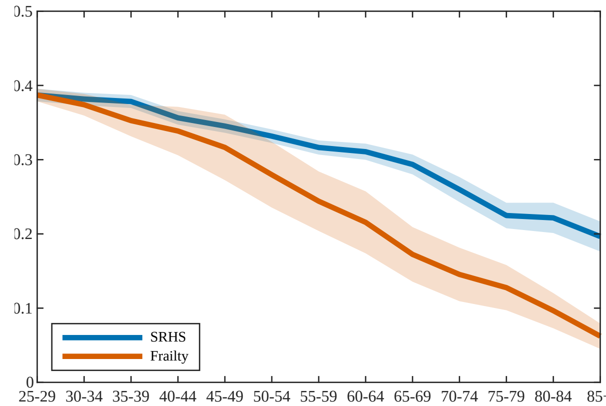
The table shows that the frailty index is more persistent than SRHS. Notice that, for all age groups, the diagonal values are all higher for the frailty index relative to SRHS. For example, individuals between age of 50 and 75, with frailty category 'excellent' have a 78.6% chance of maintaining this status while individuals in the same age group, and with 'excellent' SRHS have only a 57.1% chance. The difference in persistence is largest at the poor health end of the spectrum. Once an individual's frailty index is high enough that s/he is assigned to the 'poor' frailty category the probability s/he is there two years later is 73.2%, 83.4%, and 86.86% respectively for age groups 25–49, 50–74, and 75+. In contrast, individuals who report a SRHS status of 'poor' have only 41%, 59.3%, and 59.7% chance, respectively, of reporting poor health two years later.

The take away from these tables is that that frailty health categories are significantly more persistent than self-reported health status categories. In other words, at any age group, it is much more likely for individuals to change their assessment about their own health than it is for their frailty index to cross fixed thresholds that define frailty health categories.

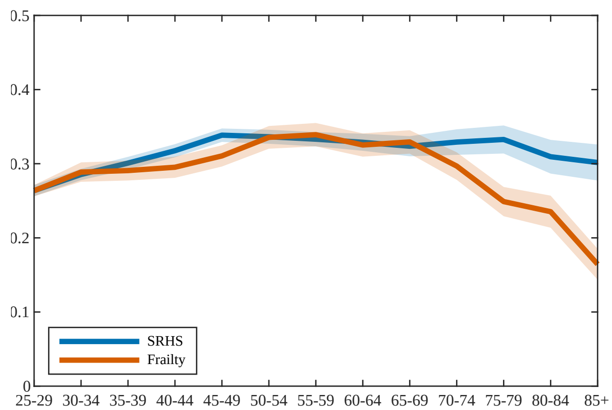
[1]We repeat the same exercise in HRS. See Section C.3.6 of this appendix.



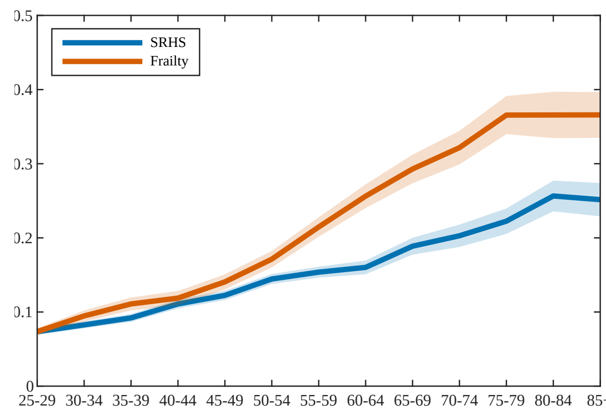
(a) Excellent



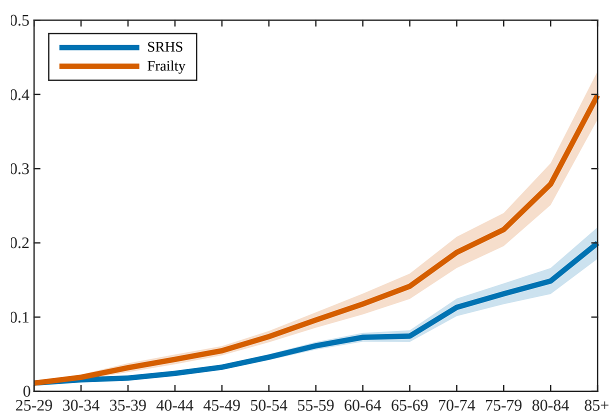
(b) Very Good



(c) Good



(d) Fair



(e) Poor

Figure 4: Distribution of health status by age. Blue like shows the share of each SRHS at each 5 year age group. Red line shows the share of each frailty category at each 5 year age group. Cut offs for frailty categories are chosen so that the distributions are aligned at age 25 to 29. Shaded areas are 95 percent confidence intervals based on 1000 independent draws of the joint distribution of age, frailty, and SRHS. Source: authors' calculations using PSID.

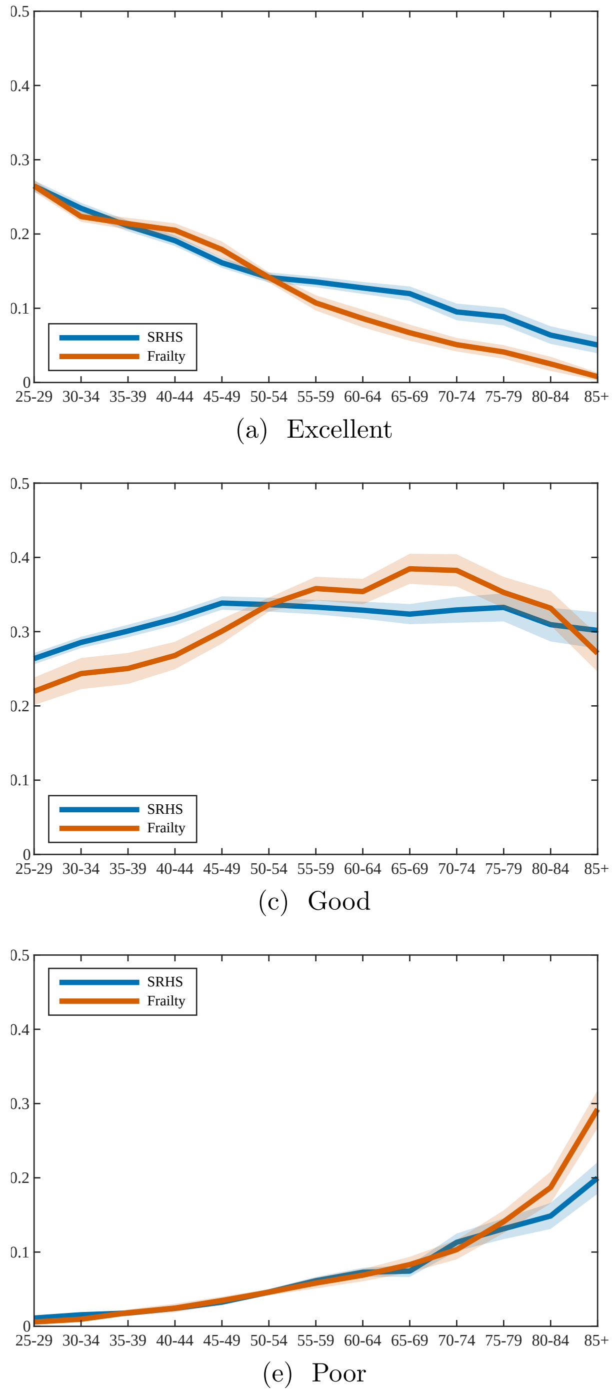


Figure 5: Distribution of health status by age. Blue line shows the share of each SRHS at each 5 year age group. Red line shows the share of each frailty category at each 5 year age group. Cut offs frailty categories are chosen so that the distributions are aligned at age 50 to 55. Shaded areas are 95 percent confidence intervals based on 1000 independent draws of the joint distribution of age, frailty, and SRHS. Source: authors' calculations using PSID.

	Excellent	Very Good	Good	Fair	Poor
Ages 25 to 49					
Excellent	53.4	33.1	11.1	2.0	0.3
Very Good	15.7	55.2	25.3	3.3	0.4
Good	6.2	27.7	52.0	12.6	1.4
Fair	3.0	10.9	34.7	43.5	7.9
Poor	2.4	3.8	14.6	38.3	41.0
Ages 50 to 74					
	Excellent	Very Good	Good	Fair	Poor
Excellent	57.1	32.2	7.9	2.2	0.6
Very Good	11.4	58.7	25.9	3.4	0.7
Good	2.6	21.5	57.1	16.1	2.7
Fair	1.4	5.3	28.6	51.1	13.6
Poor	0.9	2.8	8.2	28.9	59.3
Ages 75 and older					
	Excellent	Very Good	Good	Fair	Poor
Excellent	42.2	31.1	18.0	5.2	3.5
Very Good	8.4	48.3	28.7	9.8	4.9
Good	2.7	17.1	53.2	20.2	6.9
Fair	1.0	5.9	20.3	50.2	22.6
Poor	0.9	4.7	10.1	24.6	59.7

Table 7: Two-year transition probabilities (%) across self-reported health status categories within three age groups: 25–49 year-olds, 50–74 year-olds, and ages 75 and older constructed using the PSID sample. Rows sum to one.

	Excellent	Very Good	Good	Fair	Poor
Ages 25 to 49					
Excellent	80.5	17.0	2.3	0.2	0.1
Very Good	3.1	80.2	15.4	1.1	0.2
Good	0.1	8.0	79.0	11.7	1.3
Fair	0.1	0.6	11.7	75.5	12.1
Poor	0.1	0.8	2.6	23.3	73.2
Ages 50 to 74					
	Excellent	Very Good	Good	Fair	Poor
Excellent	78.6	17.8	3.0	0.5	0.1
Very Good	1.8	75.6	20.2	2.0	0.4
Good	0.0	4.3	78.6	15.9	1.1
Fair'	0.0	0.2	7.7	76.4	15.7
Poor	0.0	0.1	0.6	16.0	83.4
Ages 75 and older					
	Excellent	Very Good	Good	Fair	Poor
Excellent	64.5	25.5	5.5	2.7	1.8
Very Good	0.7	60.9	28.1	7.3	3.0
Good	0.0	3.4	61.0	30.6	5.1
Fair	0.0	0.2	6.2	67.2	26.4
Poor	0.0	0.0	0.5	12.8	86.8

Table 8: Two-year transition probabilities (%) across frailty health categories within three age groups: 25–49 year-olds, 50–74 year-olds, and ages 75 and older constructed using the PSID sample. Rows sum to one.

A.3.5 Predicting becoming SSDI beneficiary

In this section we report the relationship between health status and becoming a Social Security Disability Insurance (SSDI) beneficiary using PSID data.[1]

We restrict our PSID sample to those younger than 66 years old and not receiving SSDI. The dependent variable is one if they become SSDI beneficiary between wave t-1 and wave t and zero otherwise. Tables 9 and 10 show the regression results for those younger than 66, and those younger than 45, respectively. These tables show that a frailty polynomial has more power in predicting transition to disability, one wave (two years) ahead, relative to self-reported health status. Moreover, the frailty maintain a significant predicting power even with in each self-reporting health category. This is not driven by older individuals in the sample. Table 10 shows that even among the younger individuals the same results hold. Tables 11 and 12 show that adding health-age interactions does not change the findings.

Tables 13,14, 15 , 16, 17, 18, 19, and 20 repeat the exercises above with two and three lags of health indicator as explanatory variables. They show that frailty maintains its predictive edge over self-reported health status 4 and 6 years ahead.

	Panel A: younger than 66			**Panel B:** by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.080		-0.054					
	(0.073)		(0.077)					
good$_{t-1}$	0.487***		0.208***					
	(0.067)		(0.072)					
fair$_{t-1}$	1.013***		0.484***					
	(0.069)		(0.076)					
poor$_{t-1}$	1.622***		0.745***					
	(0.078)		(0.089)					
frailty$_{t-1}$		7.380***	5.992***	4.663**	5.223***	4.868***	6.578***	3.993***
		(0.385)	(0.408)	(2.064)	(1.208)	(0.841)	(0.805)	(1.086)
frailty$^2_{t-1}$		-5.558***	-4.879***	2.330	-0.958	-1.317	-5.681***	-2.478*
		(0.654)	(0.676)	(4.584)	(2.933)	(1.777)	(1.304)	(1.473)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	45,906	45,906	45,906	8,347	16,942	14,415	4,952	1,250
Pseudo R^2	0.187	0.232	0.251	0.220	0.156	0.156	0.121	0.101

Table 9: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on frailty and SRHS using data from a sub-sample of individuals younger than 66 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: younger than 45			**Panel B:** by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.113		-0.024					
	(0.100)		(0.105)					
good$_{t-1}$	0.330***		0.055					
	(0.097)		(0.104)					
fair$_{t-1}$	0.999***		0.479***					
	(0.099)		(0.110)					
poor$_{t-1}$	1.550***		0.627***					
	(0.125)		(0.146)					
frailty$_{t-1}$		6.964***	5.838***	8.327***	3.910**	5.121***	8.175***	4.460*
		(0.651)	(0.687)	(3.184)	(1.715)	(1.325)	(1.474)	(2.393)
frailty$^2_{t-1}$		-4.370***	-3.910***	-9.776	2.549	-1.878	-6.557***	-2.465
		(1.175)	(1.209)	(8.619)	(4.386)	(2.755)	(2.305)	(3.345)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	23,475	23,475	23,475	5,084	9,178	6,847	2,026	340
Pseudo R^2	0.153	0.218	0.237	0.167	0.144	0.157	0.193	0.168

Table 10: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on frailty and SRHS using data from a sub-sample of individuals younger than 45 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

	Panel A: younger than 66			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	-0.007		-0.141					
	(0.280)		(0.294)					
good$_{t-1}$	-0.172		-0.407					
	(0.266)		(0.285)					
fair$_{t-1}$	0.777***		0.288					
	(0.268)		(0.300)					
poor$_{t-1}$	1.467***		0.509					
	(0.322)		(0.374)					
excellent$_{t-1}\times$ age	-0.005		-0.007					
	(0.024)		(0.024)					
very good$_{t-1}\times$ age	-0.002		-0.005					
	(0.024)		(0.023)					
good$_{t-1}\times$ age	0.010		0.006					
	(0.023)		(0.023)					
fair$_{t-1}\times$ age	0.001		-0.002					
	(0.024)		(0.024)					
poor$_{t-1}\times$ age	-0.000		-0.001					
	(0.024)		(0.024)					
frailty$_{t-1}$		6.426***	5.946***	12.823	3.459	6.770**	10.379***	0.839
		(1.685)	(1.835)	(8.511)	(4.747)	(3.422)	(3.885)	(5.571)
frailty$^2_{t-1}$		-2.366	-2.754	-40.671	8.237	-3.401	-7.450	3.484
		(3.087)	(3.243)	(25.295)	(12.303)	(7.153)	(6.427)	(8.063)
frailty$_{t-1}\times$ age		0.019	-0.002	-0.178	0.038	-0.042	-0.081	0.062
		(0.035)	(0.038)	(0.202)	(0.106)	(0.073)	(0.080)	(0.112)
frailty$^2_{t-1}\times$ age		-0.064	-0.040	0.942*	-0.202	0.047	0.038	-0.117
		(0.063)	(0.066)	(0.563)	(0.272)	(0.152)	(0.133)	(0.160)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	45,906	45,906	45,906	8,347	16,942	14,415	4,952	1,250
Pseudo R^2	0.188	0.232	0.253	0.243	0.157	0.157	0.123	0.102

Table 11: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on frailty, SRHS and age interactions using data from a sub-sample of individuals younger than 66 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: younger than 45			**Panel B:** by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.031		0.070					
	(0.656)		(0.693)					
good$_{t-1}$	-0.457		-0.220					
	(0.641)		(0.692)					
fair$_{t-1}$	0.681		1.139					
	(0.640)		(0.725)					
poor$_{t-1}$	0.977		1.901*					
	(0.839)		(0.973)					
excellent$_{t-1}\times$ age	-0.028		-0.052					
	(0.049)		(0.043)					
very good$_{t-1}\times$ age	-0.026		-0.055					
	(0.047)		(0.042)					
good$_{t-1}\times$ age	-0.005		-0.044					
	(0.046)		(0.041)					
fair$_{t-1}\times$ age	-0.019		-0.072					
	(0.047)		(0.044)					
poor$_{t-1}\times$ age	-0.011		-0.090*					
	(0.050)		(0.048)					
frailty$_{t-1}$		5.794*	-0.713	14.058	4.242	6.404	2.668	-27.267*
		(3.149)	(4.644)	(16.780)	(12.125)	(8.350)	(9.994)	(14.679)
frailty$^2_{t-1}$		-5.415	0.910	-84.103	23.131	1.490	-8.158	43.842*
		(6.963)	(8.601)	(75.094)	(34.773)	(17.711)	(17.389)	(23.396)
frailty$_{t-1}\times$ age		0.038	0.199	-0.126	-0.017	-0.040	0.175	0.909**
		(0.088)	(0.134)	(0.496)	(0.366)	(0.245)	(0.284)	(0.423)
frailty$^2_{t-1}\times$ age		0.021	-0.152	2.050	-0.608	-0.093	0.015	-1.303**
		(0.191)	(0.241)	(2.073)	(1.047)	(0.518)	(0.483)	(0.645)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	23,475	23,475	23,475	5,084	9,178	6,847	2,026	340
Pseudo R^2	0.154	0.218	0.240	0.183	0.149	0.158	0.199	0.190

Table 12: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on frailty, SRHS and age interactions using data from a sub-sample of individuals younger than 45 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: younger than 66			**Panel B:** by SRHS health at $t-2$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-2}$	0.173**		0.041					
	(0.069)		(0.073)					
good$_{t-2}$	0.545***		0.281***					
	(0.065)		(0.069)					
fair$_{t-2}$	1.111***		0.597***					
	(0.068)		(0.074)					
poor$_{t-2}$	1.680***		0.803***					
	(0.080)		(0.092)					
frailty$_{t-2}$		7.478***	6.035***	8.611***	6.644***	4.865***	3.708***	4.766***
		(0.383)	(0.405)	(2.374)	(1.253)	(0.791)	(0.737)	(1.067)
frailty$^2_{t-2}$		-5.741***	-4.989***	-7.259	-4.013	0.373	-1.442	-4.219***
		(0.688)	(0.712)	(4.883)	(3.006)	(1.691)	(1.272)	(1.555)
Observations	33,763	33,763	33,763	5,669	12,355	10,866	3,845	1,028
Pseudo R^2	0.183	0.225	0.244	0.229	0.173	0.191	0.091	0.077

Table 13: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on two periods lagged values of frailty and SRHS using data from a sub-sample of individuals younger than 66 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-2$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: younger than 45			**Panel B:** by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-2}$	0.182		0.078					
	(0.111)		(0.116)					
good$_{t-2}$	0.483***		0.266**					
	(0.106)		(0.113)					
fair$_{t-2}$	1.066***		0.602***					
	(0.112)		(0.123)					
poor$_{t-2}$	1.671***		0.801***					
	(0.145)		(0.167)					
frailty$_{t-2}$		6.684***	5.372***	6.631**	7.818***	4.503***	4.786***	5.053*
		(0.710)	(0.744)	(3.095)	(2.179)	(1.411)	(1.377)	(2.739)
frailty$^2_{t-2}$		-3.847***	-3.168**	-4.376	-9.287	0.505	-1.329	-3.830
		(1.365)	(1.407)	(6.032)	(5.875)	(3.016)	(2.414)	(4.365)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	15,686	15,686	15,686	3,162	6,128	4,730	1,420	246
Pseudo R^2	0.157	0.210	0.230	0.207	0.134	0.191	0.143	0.119

Table 14: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on two period lagged values of frailty and SRHS using data from a sub-sample of individuals younger than 45 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-2$. Controls are gender, education, marital status and a quadratic in age. Note: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

	Panel A: younger than 66			Panel B: by SRHS health at $t-2$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-2}$	0.201		0.148					
	(0.291)		(0.305)					
good$_{t-2}$	0.373		0.188					
	(0.276)		(0.294)					
fair$_{t-2}$	0.948***		0.485					
	(0.287)		(0.318)					
poor$_{t-2}$	2.019***		1.018**					
	(0.355)		(0.409)					
excellent$_{t-2}\times$ age	-0.001		-0.002					
	(0.028)		(0.027)					
very good$_{t-2}\times$ age	-0.002		-0.005					
	(0.028)		(0.027)					
good$_{t-2}\times$ age	0.002		-0.000					
	(0.027)		(0.027)					
fair$_{t-2}\times$ age	0.002		-0.000					
	(0.028)		(0.027)					
poor$_{t-2}\times$ age	-0.008		-0.007					
	(0.028)		(0.027)					
frailty$_{t-2}$		6.474***	5.714***	-4.290	12.856**	5.667	7.247**	3.507
		(1.787)	(1.925)	(8.198)	(5.694)	(3.640)	(3.558)	(5.863)
frailty$^2_{t-2}$		-1.331	-1.442	14.049	-23.217	1.455	-2.464	0.617
		(3.439)	(3.596)	(18.466)	(14.932)	(7.736)	(6.490)	(9.057)
frailty$_{t-2}\times$ age		0.019	0.004	0.340	-0.136	-0.018	-0.079	0.021
		(0.037)	(0.041)	(0.219)	(0.123)	(0.076)	(0.076)	(0.120)
frailty$^2_{t-2}\times$ age		-0.088	-0.069	-0.570	0.416	-0.021	0.027	-0.091
		(0.071)	(0.074)	(0.508)	(0.317)	(0.160)	(0.137)	(0.182)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	33,763	33,763	33,763	5,669	12,355	10,866	3,845	1,028
Pseudo R^2	0.184	0.225	0.245	0.241	0.175	0.191	0.094	0.078

Table 15: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on two periods lagged values of frailty, SRHS and age interactions using data from a sub-sample of individuals younger than 66 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-2$. Controls are gender, education, marital status and a quadratic in age. Note: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

	Panel A: younger than 45			Panel B: by SRHS health at $t-2$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-2}$	-0.155		-0.165					
	(0.786)		(0.813)					
good$_{t-2}$	-0.183		-0.137					
	(0.750)		(0.787)					
fair$_{t-2}$	0.215		0.442					
	(0.790)		(0.865)					
poor$_{t-2}$	0.886		1.540					
	(1.066)		(1.223)					
excellent$_{t-2}\times$ age	-0.094		-0.092					
	(0.072)		(0.076)					
very good$_{t-2}\times$ age	-0.083		-0.084					
	(0.069)		(0.075)					
good$_{t-2}\times$ age	-0.073		-0.079					
	(0.069)		(0.074)					
fair$_{t-2}\times$ age	-0.067		-0.087					
	(0.069)		(0.076)					
poor$_{t-2}\times$ age	-0.069		-0.114					
	(0.073)		(0.081)					
frailty$_{t-2}$		5.847	-0.011	12.651	7.461	6.051	-1.303	-28.404
		(3.679)	(5.444)	(27.089)	(9.451)	(11.034)	(9.502)	(20.174)
frailty$^2_{t-2}$		-3.672	2.850	-45.300	-5.794	10.497	9.056	39.565
		(8.980)	(11.151)	(66.703)	(31.319)	(25.407)	(19.356)	(34.854)
frailty$_{t-2}\times$ age		0.026	0.165	-0.196	0.011	-0.062	0.183	1.034*
		(0.107)	(0.163)	(0.909)	(0.281)	(0.320)	(0.281)	(0.615)
frailty$^2_{t-2}\times$ age		-0.008	-0.185	1.326	-0.107	-0.255	-0.308	-1.333
		(0.256)	(0.325)	(2.197)	(0.943)	(0.719)	(0.555)	(1.043)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	15,686	15,686	15,686	3,162	6,128	4,730	1,420	246
Pseudo R^2	0.158	0.210	0.232	0.213	0.134	0.194	0.143	0.139

Table 16: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on two periods lagged values of frailty, SRHS and age interactions using data from a sub-sample of individuals younger than 45 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-2$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: younger than 66			**Panel B:** by SRHS health at $t-3$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-3}$	0.157**		0.035					
	(0.069)		(0.073)					
good$_{t-3}$	0.554***		0.303***					
	(0.065)		(0.070)					
fair$_{t-3}$	1.107***		0.604***					
	(0.069)		(0.076)					
poor$_{t-3}$	1.548***		0.683***					
	(0.088)		(0.100)					
frailty$_{t-3}$		7.792***	6.372***	3.770	7.372***	4.831***	4.898***	3.802***
		(0.426)	(0.447)	(2.458)	(1.449)	(0.837)	(0.858)	(1.154)
frailty$^2_{t-3}$		-6.730***	-5.769***	5.316	-4.776	-0.213	-3.915**	-3.272*
		(0.822)	(0.849)	(6.074)	(3.579)	(1.880)	(1.625)	(1.772)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	23,103	23,103	23,103	3,617	8,434	7,469	2,793	790
Pseudo R^2	0.172	0.214	0.231	0.212	0.201	0.171	0.096	0.061

Table 17: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on three periods lagged values of frailty and SRHS using data from a sub-sample of individuals younger than 66 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-3$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: younger than 45			Panel B: by SRHS health at $t-3$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-3}$	0.192		0.078					
	(0.124)		(0.130)					
good$_{t-3}$	0.490***		0.246*					
	(0.120)		(0.127)					
fair$_{t-3}$	1.113***		0.652***					
	(0.127)		(0.139)					
poor$_{t-3}$	1.658***		0.857***					
	(0.174)		(0.195)					
frailty$_{t-3}$		7.629***	6.150***	1.661	5.547**	4.900***	6.741***	8.019**
		(0.870)	(0.910)	(4.101)	(2.829)	(1.594)	(1.741)	(3.234)
frailty$^2_{t-3}$		-6.294***	-5.373***	8.634	-1.619	-0.128	-6.165*	-11.082**
		(1.813)	(1.857)	(11.957)	(7.978)	(3.573)	(3.320)	(5.406)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	9,656	9,656	9,656	1,835	3,760	2,963	915	183
Pseudo R^2	0.167	0.208	0.232	0.116	0.177	0.205	0.148	0.103

Table 18: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on three periods lagged values of frailty and SRHS using data from a sub-sample of individuals younger than 45 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-3$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: younger than 66			Panel B: by SRHS health at $t-3$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-3}$	0.180		0.115					
	(0.312)		(0.328)					
good$_{t-3}$	0.156		-0.079					
	(0.297)		(0.319)					
fair$_{t-3}$	1.072***		0.580*					
	(0.310)		(0.346)					
poor$_{t-3}$	1.788***		0.880*					
	(0.415)		(0.474)					
excellent$_{t-3}\times$ age	-0.003		-0.004					
	(0.019)		(0.019)					
very good$_{t-3}\times$ age	-0.003		-0.006					
	(0.018)		(0.019)					
good$_{t-3}\times$ age	0.006		0.004					
	(0.018)		(0.019)					
fair$_{t-3}\times$ age	-0.002		-0.004					
	(0.018)		(0.019)					
poor$_{t-3}\times$ age	-0.008		-0.008					
	(0.019)		(0.020)					
frailty$_{t-3}$		8.069***	6.937***	-7.971	9.166	5.621	7.970*	10.099
		(2.028)	(2.204)	(9.339)	(6.573)	(3.888)	(4.442)	(6.326)
frailty$^2_{t-3}$		-4.990	-4.658	26.368	-9.326	3.702	-6.838	-13.763
		(4.141)	(4.344)	(24.927)	(18.201)	(8.907)	(8.632)	(10.043)
frailty$_{t-3}\times$ age		-0.008	-0.014	0.287	-0.041	-0.019	-0.068	-0.132
		(0.043)	(0.047)	(0.214)	(0.147)	(0.084)	(0.096)	(0.132)
frailty$^2_{t-3}\times$ age		-0.034	-0.019	-0.507	0.104	-0.083	0.066	0.217
		(0.087)	(0.091)	(0.545)	(0.408)	(0.192)	(0.187)	(0.204)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	23,103	23,103	23,103	3,617	8,434	7,469	2,793	790
Pseudo R^2	0.173	0.214	0.233	0.221	0.201	0.172	0.096	0.062

Table 19: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on three periods lagged values of frailty, SRHS and age interactions using data from a sub-sample of individuals younger than 66 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-3$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: younger than 45			Panel B: by SRHS health at $t-3$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-3}$	-0.272		-0.321					
	(0.980)		(1.006)					
good$_{t-3}$	-0.186		-0.069					
	(0.940)		(0.989)					
fair$_{t-3}$	0.220		0.179					
	(0.980)		(1.077)					
poor$_{t-3}$	0.391		0.520					
	(1.450)		(1.618)					
excellent$_{t-3}\times$ age	-0.041		-0.059					
	(0.045)		(0.047)					
very good$_{t-3}\times$ age	-0.026		-0.046					
	(0.040)		(0.044)					
good$_{t-3}\times$ age	-0.019		-0.049					
	(0.039)		(0.045)					
fair$_{t-3}\times$ age	-0.012		-0.044					
	(0.037)		(0.046)					
poor$_{t-3}\times$ age	-0.001		-0.048					
	(0.053)		(0.063)					
frailty$_{t-3}$		4.312	-3.640	-36.342	29.355	-3.457	-16.180	8.523
		(4.288)	(7.249)	(27.374)	(22.687)	(11.461)	(13.698)	(28.852)
frailty$^2_{t-3}$		5.376	16.683	286.455*	-57.404	24.101	49.123*	4.024
		(11.354)	(15.165)	(172.511)	(68.664)	(28.659)	(28.997)	(51.338)
frailty$_{t-3}\times$ age		0.102	0.306	1.416	-0.774	0.248	0.693*	0.013
		(0.130)	(0.226)	(0.957)	(0.732)	(0.346)	(0.415)	(0.926)
frailty$^2_{t-3}\times$ age		-0.359	-0.685	-10.263	1.836	-0.712	-1.672*	-0.542
		(0.346)	(0.469)	(6.408)	(2.254)	(0.848)	(0.880)	(1.680)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	9,656	9,656	9,656	1,835	3,760	2,963	915	183
Pseudo R^2	0.168	0.209	0.234	0.161	0.184	0.206	0.157	0.111

Table 20: Results of probit regressions of transitioning to SSDI beneficiary status between wave t-1 and wave t, on three periods lagged values of frailty, SRHS and age interactions using data from a sub-sample of individuals younger than 66 in PSID. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-3$. Controls are gender, education, marital status and a quadratic in age. Note: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

A.4 Probability Transition Matrices for Frailty Variants in PSID

		Ages 25 to 49			
	Bottom	2nd	3rd	4th	Top
Bottom	79.3	14.6	4.0	1.7	0.4
2nd	5.0	69.4	18.5	5.6	1.4
3rd	0.7	10.6	63.2	20.6	5.0
4th	0.1	1.7	11.3	65.6	21.3
Top	0.1	0.2	0.8	9.5	89.4
		Ages 50 to 74			
	Bottom	2nd	3rd	4th	Top
Bottom	79.3	15.7	3.8	0.9	0.3
2nd	5.5	67.4	20.7	5.3	1.2
3rd	0.3	8.6	64.5	23.2	3.3
4th	0.0	1.0	9.8	66.4	22.8
Top	0.0	0.1	0.7	10.5	88.6
		Ages 75 and older			
	Bottom	2nd	3rd	4th	Top
Bottom	70.0	20.2	5.9	2.5	1.4
2nd	5.7	55.5	25.2	8.9	4.6
3rd	1.0	10.3	47.1	28.7	13.0
4th	0.1	1.8	15.1	51.2	31.8
Top	0.0	0.2	1.6	16.4	81.9

Table 21: Two-year transition probabilities (%) across quintiles of the frailty with SRHS distribution for three age groups: 25-49 year-olds, 50-74 year-olds, and 75 and older constructed using the PSID sample. Rows sum to one.

			Ages 25 to 49		
	Bottom	2nd	3rd	4th	Top
Bottom	80.1	8.2	7.3	3.3	1.1
2nd	3.7	75.0	10.6	8.4	2.3
3rd	0.1	2.2	74.0	17.5	6.2
4th	1.8	3.2	5.9	68.5	20.7
Top	0.2	0.3	2.2	10.5	86.9
			Ages 50 to 74		
	Bottom	2nd	3rd	4th	Top
Bottom	80.2	13.4	3.8	2.0	0.7
2nd	2.5	71.3	18.5	5.8	1.9
3rd	1.2	5.4	67.1	22.1	4.2
4th	0.3	1.8	11.9	62.9	23.1
Top	0.1	0.2	1.8	14.0	83.9
			Ages 75 and older		
	Bottom	2nd	3rd	4th	Top
Bottom	66.5	20.4	7.2	3.5	2.3
2nd	6.3	53.1	24.7	10.6	5.3
3rd	1.0	15.8	42.0	27.7	13.5
4th	1.1	2.8	18.6	47.9	29.6
Top	0.0	1.2	3.2	16.5	79.1

Table 22: Two-year transition probabilities (%) across quintiles of the FPC index distribution for three age groups: 25-49 year-olds, 50-74 year-olds, and 75 and older constructed using the PSID sample. Rows sum to one.

	Bottom	2nd	3rd	4th	Top
Ages 25 to 49					
Bottom	81.4	9.0	6.1	2.0	1.5
2nd	3.9	76.3	11.1	6.5	2.1
3rd	0.0	7.1	73.3	14.6	5.0
4th	0.1	0.4	3.4	77.3	18.8
Top	0.6	0.9	2.1	7.3	89.0
Ages 50 to 74					
	Bottom	2nd	3rd	4th	Top
Bottom	80.8	13.2	2.6	2.4	0.9
2nd	2.0	73.2	16.7	5.4	2.6
3rd	0.0	2.4	73.0	16.8	7.8
4th	1.1	3.3	5.5	65.5	24.6
Top	0.1	0.6	4.2	19.9	75.1
Ages 75 and older					
	Bottom	2nd	3rd	4th	Top
Bottom	67.3	10.9	15.4	4.4	2.0
2nd	1.1	56.6	22.6	12.8	7.0
3rd	7.6	16.1	35.4	21.0	20.0
4th	0.0	6.5	16.0	42.4	35.1
Top	0.3	4.9	10.3	28.9	55.7

Table 23: Two-year transition probabilities (%) across quintiles of the SIO index distribution for three age groups: 25-49 year-olds, 50-74 year-olds, and 75 and older constructed using the PSID sample. Rows sum to one.

B MEPS

In this section we briefly describe the MEPS sample we use for our analysis and report the list of deficit variables we use for construction of our health measures. We also include additional details regarding baseline Frailty and other measures in MEPS.

B.1 MEPS sample and list of deficit variables

The MEPS consists of a collection of rotating two-year panels. We use MEPS data from the 2000 to 2016 period. Our sample consists of respondents aged 25 years and above. The sample contains 353,487 observations on 195,887 individuals (90,230 men and 105,657 women). Table 24 lists the 27 variables we use to construct respondents' frailty index values. This list includes 12 ADL/IADL variables, 13 diagnosis, chronic conditions, and measurement variables, and two mental health and cognition variables.

The index is constructed in the same way as for PSID. One advantage of the MEPS sample is its large number of observations. However, because MEPS is a two-year rotating panel, it only has two frailty observations, at most, per individual.

We also construct all other variants of frailty in MEPS (as we in PSID). Table 25 reports the weights associated with each deficit for each of the alternative health indices. The deficits are order according to their weights in the first principal component (FPC) index. We observe that in MEPS, similar to PSID, first principal component puts highest weight on ADLs and IADLs variables. In particular, aggregating the weights across deficits by category we find that it puts 64 percent of the total weight on the ADL/IADL variables, 26.5 percent on the medical diagnosis/measurement variables, and 6 percent on the mental/cognitive health variables with the remaining 3.5 percent on SRHS. In contrast, frailty (and frailty with SRHS) equally weight each deficit. Consequently, frailty puts 44.5 percent of the weight on the ADL/IADL variables, 48 percent on the medical variables, 7 percent on mental health, and (since it is omitted) zero weight on SRHS. Similarly, the distribution of weights for frailty-with-SRHS is 43, 46, 7, and 4 percent, respectively. This break downs are very similar to those reported for PSID in the paper.

B.1.1 Regression tables for construction of SIO index in MEPS

We follow the same two step procedure to construct SIO in MEPS, as we do in PSID. The principal component weights for the first step are reported in Table 26. The second step regression results are reported in Table 27.

Variable	Value
Need help with ADLs	Yes=1, No=0
Need help with IADLs	Yes=1, No=0
Use assistive technology	Yes=1, No=0
Limitation impacts work/housework/school	Yes=1, No=0
Any difficultly with the following:	
Walking three blocks	Yes=1, No=0
Walking a mile	Yes=1, No=0
Standing for 20 minutes	Yes=1, No=0
Bending/Stooping	Yes=1, No=0
Lifting 10 lbs	Yes=1, No=0
Walking up 10 steps	Yes=1, No=0
Using fingers to grasp	Yes=1, No=0
Reaching over head	Yes=1, No=0
Ever been diagnosed with:	
High Blood Pressure	Yes=1, No=0
Diabetes	Yes=1, No=0
Cancer	Yes=1, No=0
Emphysema	Yes=1, No=0
Angina	Yes=1, No=0
Coronary Heart Disease	Yes=1, No=0
Other Heart Disease	Yes=1, No=0
Heart Attack	Yes=1, No=0
Stroke	Yes=1, No=0
Asthma	Yes=1, No=0
Arthritis	Yes=1, No=0
High Cholesterol	Yes=1, No=0
BMI ≥ 30	Yes=1, No=0
Cognitive Limitations	Yes=1, No=0
K6 Depression Score	0–24, rescaled to 0–1
The following variable is only used in construction of Frailty-with-SRHD and FPC	
Self-reported health of 'fair' or 'poor'	Yes=1, No=0

Table 24: Health variables used to construct frailty in MEPS. For "ever had one of following conditions", we make the following adjustment to the raw data: if in any wave an individual has a positive answer to any the conditions, we assign the value of 1 to that conditions when calculating frailty in all future waves.

Variable	FPC (1)	SIO (2)	Frailty with SRHS (3)	Baseline Frailty (4)
Difficulty walking three blocks	0.062	0.190	0.036	0.037
Difficulty walking a mile	0.062	0.021	0.036	0.037
Difficulty standing for 20 minutes	0.060	0.080	0.036	0.037
Difficulty bending/Stooping	0.060	-0.002	0.036	0.037
Difficulty walking up 10 steps	0.059	0.016	0.036	0.037
Difficulty lifting 10 lbs	0.058	0.006	0.036	0.037
Difficulty reaching over head	0.056	0.270	0.036	0.037
Limitation impacts work/housework/school	0.051	0.036	0.036	0.037
Use assistive technology	0.048	0.008	0.036	0.037
Difficulty using fingers to grasp	0.046	0.065	0.036	0.037
Need help with IADLs	0.040	0.039	0.036	0.037
Cognitive limitations	0.037	-0.050	0.036	0.037
Need help with ADLs	0.035	0.000	0.036	0.037
Self-reported health of 'fair' or 'poor'	0.035	n.a.	0.036	n.a.
Ever had arthritis	0.033	0.025	0.036	0.037
Ever had high blood pressure	0.027	0.126	0.036	0.037
K6 depression score	0.026	0.021	0.036	0.037
Ever had stroke	0.024	0.013	0.036	0.037
Ever had coronary heart disease	0.023	0.006	0.036	0.037
Ever had high cholesterol	0.023	0.006	0.036	0.037
Ever had other heart disease	0.022	0.034	0.036	0.037
Ever had diabetes	0.021	0.011	0.036	0.037
Ever had heart attack	0.021	0.006	0.036	0.037
Ever had angina	0.018	0.006	0.036	0.037
Ever had emphysema	0.017	0.005	0.036	0.037
Ever had cancer	0.015	0.012	0.036	0.037
Ever had asthma	0.011	0.012	0.036	0.037
BMI >= 30	0.010	0.000	0.036	0.037

Table 25: Weights assigned to each deficit in each variations of frailty. Column (1) show first principal component (FPC) weights. Column (2) shows SIO wieghts. Column (3) is the equal weight frailty index that includes SRHS. Column (4) is equal weight frailty index that excludes SRHS (baseline frailty index). Deficit variables are sorted according to FPC weights. All calculations are based on MEPS data on household heads and their spouses ages 25 and older.

Health limits activities	0.347
Health limits work	0.352
Self-reported health of 'fair' or 'poor'	0.301

Table 26: Principal component weights for first stage calculation of SIO.

Variable	Value	Variable	Value
Ever had diabetes	0.043***	Difficulty standing for 20 minutes	0.026***
	(0.001)	(0.002)	
Ever had high blood pressure	0.020***	Difficulty bending, Stooping	0.100***
	(0.001)	(0.002)	
Ever had arthritis	0.031***	Difficulty lifting 10 lb	0.020***
	(0.001)		(0.002)
Ever had stroke	0.026***	Difficulty walking up 10 steps	-0.002
	(0.002)		(0.002)
Ever had emphysema	0.048***	Difficulty using fingers to grasp	0.010***
	(0.002)		(0.002)
Ever had angina	0.008***	Difficulty reaching over head	0.007***
	(0.002)		(0.002)
Ever had coronary heart disease	0.017***	Ever had cancer	0.006***
	(0.002)		(0.001)
Ever had asthma	0.016***	Ever had heart attack	0.014***
	(0.001)		(0.002)
Ever had other heart disease	0.008***	Ever had high cholesterol	0.008***
	(0.001)		(0.001)
Need help with ADLs	-0.063***	Limitation impacts work, housework, school	0.338***
	(0.003)		(0.002)
Need help with IADLs	0.082***	Cognitive limitations	0.049***
	(0.002)	(0.002)	
Use assistive technology	0.045***	BMI $\geq$ 30	0.011***
	(0.002)	(0.001)	
Difficulty walking three blocks	-0.011***	K6 depression score	0.158***
	(0.004)	(0.002)	
Difficulty walking a mile	0.238***	cons	0.006***
	(0.004)		(0.001)
Observations	166,045		
Adjusted R^2	0.781		

Table 27: SIO index second step regression (MEPS). Note: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

B.2 Facts on frailty dynamics over the life cycle in MEPS

We report summary statistics on the distribution of frailty in our MEPS sample in Table 28. The numbers reported in this table are very similar to the ones reported in the paper for the PSID sample (Table 1 in the paper). [2] However, the frailty in MEPS is slightly lower, on average, at young ages and slight higher, on average, for those older than 75. It also has higher standard deviation and larger fraction at zero, relative to the frailty index in PSID. However, other patters are remarkably similar to PSID. In all three age groups presented in the table, women have higher frailty on average than men but only slightly. Also, differences in frailty are larger by education than by gender. In summary, in our MEPS sample, mean frailty increases with age, decreases with education, and is slightly higher for women than for men.

Figure 6a shows box and whisker plots of frailty by 10-year age groups constructed using the MEPS sample. This figure shows that in MEPS, similar to PSID, the gap between the top and bottom quantiles of the distributions widens with age. Also, Even though dispersion in frailty is increasing with age, the figure also shows significant variation in frailty in all of the age groups. Therefore, frailty index is capturing variation in health even among the younger individuals in the sample.

Figure 6b shows the histogram of frailty for the entire MEPS sample. This figures demonstrates that, similar to PSID, the cross-sectional distribution of frailty in MEPS is right-skewed. This is for two reasons. One, there is a large mass at or near zero. As it is reported in Table 28 19 percent of MEPS sample have frailty of precisely zero. About another 10 percent have a non-zero K6 Depression Score. This score measures nonspecific psychological distress by asking respondents how often they felt symptoms of mental illnesses (eg, hopelessness and depression) within the past 30 days. Responses were categorized on a 5-point scale (none of the time, a little of the time, some of the time, most of the time, or all of the time. The scale ranges from 0 to 24, with higher K6 scores indicating the greater likelihood of psychological distress. We incorporate this score by normalizing it to a number between zero and one and treating it like other deficits. As a result, a low, but non-zero K6 score contributes a small amount to frailty index. This results to a mass near zero.

The other reason for skewness in the frailty index in MESP, similar to PSID, is the substantial variation in frailty among highly frail individuals. As the box plots in Figure 6a show, the distribution of frailty is right-skewed within each of the 10-year age groups.

Finally Table 29 shows two-year transition probabilities across frailty quintiles for three age groups: 25 to 49 year-olds, 50 to 74 year-olds, and individuals 75 and older. The tables shows that, similar to PSID, frailty in MEPS is highly persistent. Within all three age groups, individuals are most likely to remain in the same frailty quintile and the probability of moving to another quintile is rapidly declining in its distance from the previous one. Alos, frailty is more persistent at the top and bottom quintiles. However, unlike PSID, we don't observer a clear pattern of decline (or increase) in persistence with age. Also, the persistence is in general lower at each age group (judged by the values on the diagonal) relative to PSID.

[2] This is consistent with findings of Kulminski et al. (2007), among others, who report consistency in properties of frailty index across various datasets.

	All	Men	Women	Less Than HS	HS Grad	College Grad
Mean	0.11	0.10	0.12	0.15	0.11	0.07
ages 25-49	0.06	0.05	0.06	0.06	0.06	0.04
ages 50-74	0.15	0.14	0.16	0.20	0.15	0.10
ages 75+	0.28	0.25	0.30	0.31	0.26	0.22
Standard deviation	0.15	0.14	0.16	0.19	0.15	0.11
ages 25-49	0.09	0.08	0.10	0.11	0.10	0.06
ages 50-74	0.17	0.15	0.18	0.20	0.16	0.12
ages 75+	0.21	0.20	0.22	0.22	0.20	0.19
Min	0.00	0.00	0.00	0.00	0.00	0.00
5th percentile	0.00	0.00	0.00	0.00	0.00	0.00
50th percentile	0.04	0.04	0.05	0.07	0.04	0.04
95th percentile	0.49	0.43	0.52	0.58	0.47	0.30
Max	1.00	1.00	0.98	1.00	0.95	0.92
Fraction at zero	0.19	0.21	0.17	0.18	0.19	0.22
Observations	353509	162339	191170	65562	149280	69055
Individuals	195888	90230	105658	36574	83656	38531
% of total	100	46	54	19	43	20

Table 28: Summary statistics for frailty constructed using 2000–2016 MEPS respondents over the age 25.

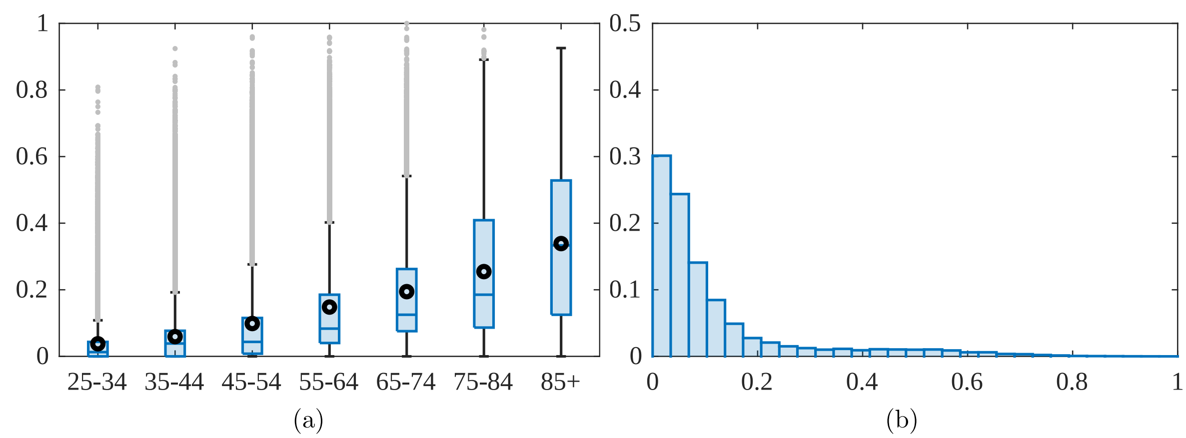


Figure 6: Panel (a) is box plots of frailty by 10-year age groups in the MEPS sample. The bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The middle lines are medians and the open circles are means. The upper whiskers extend to 1.5 times the inner quartile range. The lower whiskers extend to the minimum value of frailty in the age group. The dots are data points that lie beyond the whiskers. Panel (b) is a histogram showing the cross-sectional distribution of frailty in the sample.

	Bottom	2nd	3rd	4th	Top
Ages 25 to 49					
Bottom	68.6	12.6	14.0	2.9	1.9
2nd	9.2	67.8	12.5	8.8	1.7
3rd	11.1	14.8	49.8	17.4	6.9
4th	11.0	6.1	17.0	50.6	15.3
Top	2.5	1.3	7.1	15.3	73.8
Ages 50 to 74					
	Bottom	2nd	3rd	4th	Top
Bottom	75.1	17.6	4.5	1.5	1.3
2nd	17.5	55.2	19.5	4.9	2.9
3rd	2.7	13.3	57.5	20.1	6.3
4th	1.1	3.5	15.5	59.1	20.8
Top	0.8	1.9	5.0	17.5	74.9
Ages 75 and older					
	Bottom	2nd	3rd	4th	Top
Bottom	67.9	16.6	7.4	6.3	1.8
2nd	13.8	45.8	21.8	13.5	5.0
3rd	5.0	13.3	45.2	24.1	12.4
4th	3.4	7.1	16.3	44.1	29.1
Top	0.3	1.4	5.4	15.9	76.9

Table 29: Two-year transition probabilities (%) across frailty quintiles for three age groups: 25–49 year-olds, 50–74 year-olds, and ages 75 and older constructed using the MEPS sample. Rows sum to one.

B.3 Comparison of frailty index to SRHS in MEPS

B.3.1 Health dynamics

We compare dynamics of the distribution of frailty over the life cycle to that of SRHS in MEPS. To do this, we follow the same procedures outlines in the paper and reported for the PSID sample. We find cut offs of frailty so that the partitions in the distribution of frailty have the same share as SRHS for age group 25 to 29. We refer to these partitions as frailty health category we label them 'excellent', 'very good', 'good', 'fair', and 'poor'. The we label an individual in 'excellent' frailty category if their frailty falls below the lower cut off. Other categories are define accordingly as laid out in details in the paper.

The results are reported in Figure 7. For each age group, the height of each shaded area is the fraction of individuals in the corresponding SRHS category. The dashed lines demonstrate the boundaries of each frailty category at each age group. The cut offs are reported below the figure.

The patter here is remarkably similar to the one in PSID (reported in the paper). Share of SRHS 'excellent' and 'very good' goes from 67 percent in age group 25-29 to close to 40 percent in in age group 85+. At the same time, the fraction of 'excellent' and 'very good' frailty category declines sharply after age 40-44 and falls to less than 10 percent in age group 85+.

Similarly the fraction of 'fair' or 'poor' frailty category expands at more rapid pace than that of SRHS. While there are only 30 percent of 85+ age group have SRHS of 'fair' or 'poor', more than 80 percent have frailty category of 'fair' or 'poor'. Health, measured by frailty, deteriorates faster than health measured by SRHS.

B.3.2 Dispersion

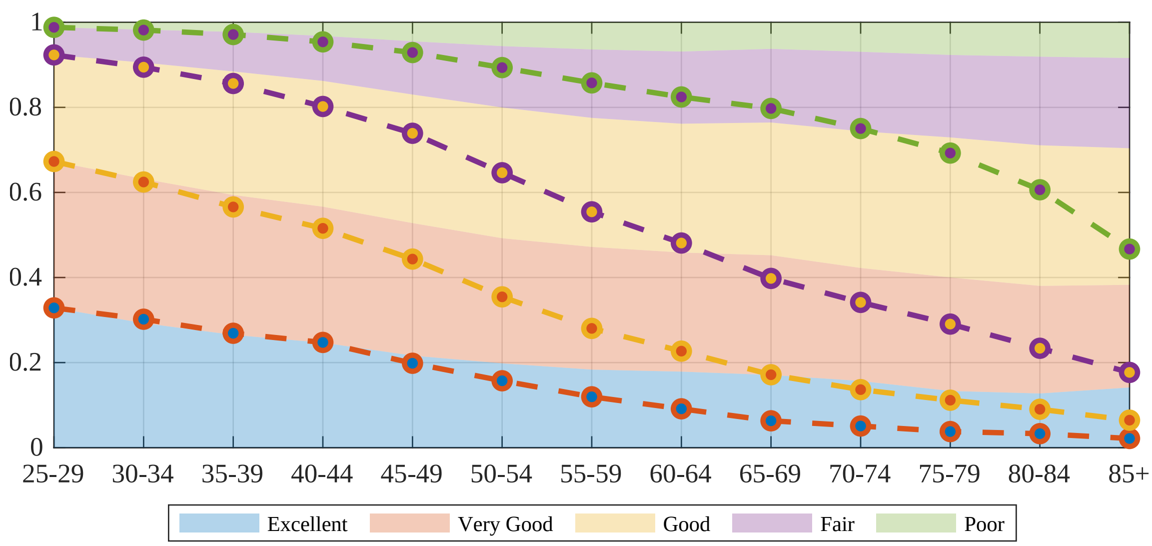
Figure 8 shows box plots for the distribution of frailty by SRHS categories and 10-year age groups. The bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The middle line indicates the median. Within each age group, median frailty increases across SRHS categories indicating that the two health measures are positively correlated. Notice that in all the age groups the inter-quartile range also increases as SRHS declines indicating that there is more variation in frailty among individuals who self-report worse health.

B.3.3 Persistence

Table 30 shows two-year transition probabilities across SRHS categories by 25 year age groups in MEPS. Table 31 shows the same, but for transition across frailty health categories, as defined above. The patterns here are very similar to PSID. Frailty health categories exhibit more more persistence than SRHS categories.

B.3.4 Predicting medical expenditures

MEPS is the only dataset we use that contains rich and comprehensive data on total medical expenditures. We take advantage of this feature to examine the extent to which frailty can predict total medical expenditures two years ahead.



	excellent	very good	good	fair	poor
% of SRHS in age 25 - 29	32.9	34.4	25.0	6.5	1.2
Frailty range	[0, 0.007)	[0.007, 0.043)	[0.043, 0.090)	[0.090, 0.293)	[0.293, 1]

Figure 7: Distribution of health status by age. The colored areas show the fraction of individuals by SRHS at each age. The dashed lines show the fraction of individuals by the corresponding frailty category at each age. Source: authors' calculations using MEPS.

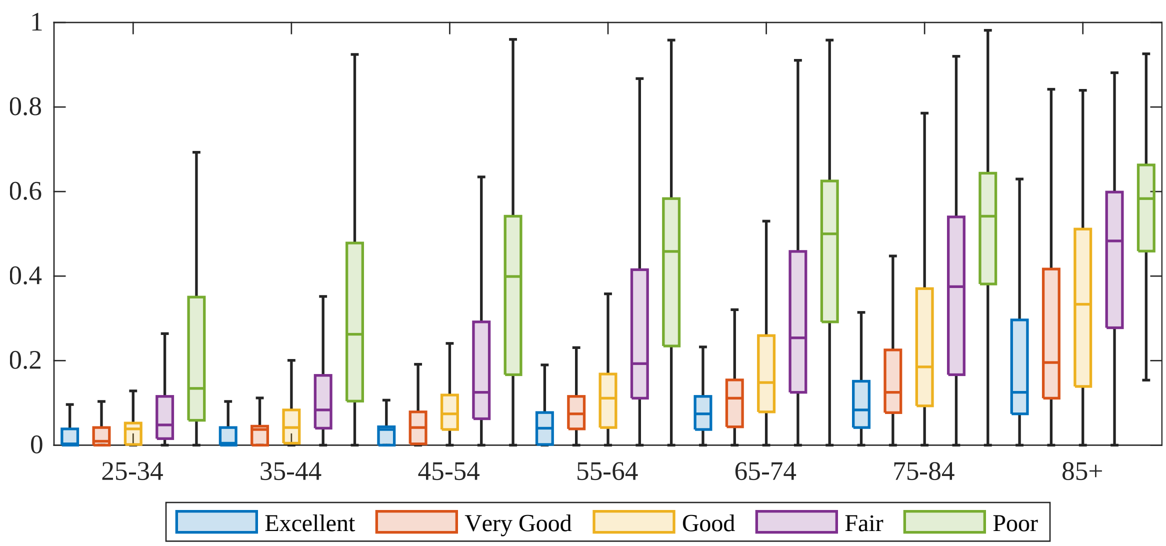


Figure 8: Box plots of frailty by SRHS and 10-year age groups in the MEPS sample. The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The middle line are medians. The upper whiskers extend to 1.5 times the inner quartile range. The lower whiskers extend to the minimum value of frailty in the subgroup.

	Excellent	Very Good	Good	Fair	Poor
Ages 25 to 49					
Excellent	52.0	31.1	14.5	2.1	0.3
Very Good	22.1	48.4	25.2	3.8	0.5
Good	12.0	30.1	46.6	9.9	1.5
Fair	6.7	16.0	40.4	30.6	6.2
Poor	2.8	6.4	22.6	34.2	34.0
Ages 50 to 74					
	Excellent	Very Good	Good	Fair	Poor
Excellent	48.1	33.1	15.5	2.6	0.7
Very Good	15.3	48.9	29.3	5.3	1.1
Good	7.1	25.7	50.0	14.7	2.5
Fair	3.0	10.5	34.7	41.6	10.2
Poor	1.0	4.1	17.0	35.8	42.1
Ages 75 and older					
	Excellent	Very Good	Good	Fair	Poor
Excellent	37.2	33.9	21.8	5.2	1.9
Very Good	12.2	41.6	32.9	10.6	2.8
Good	5.4	22.5	47.6	19.1	5.4
Fair	2.6	11.0	30.4	41.0	14.9
Poor	0.7	6.3	17.7	36.4	38.9

Table 30: Two-year transition probabilities (%) across self-reported health status categories within three age groups: 25–49, 50–74 year-olds, and ages 75 and older constructed using the MEPS sample. Rows sum to one.

	Excellent	Very Good	Good	Fair	Poor
Ages 25 to 49					
Excellent	72.6	20.9	5.2	1.0	0.3
Very Good	16.9	63.9	16.2	2.5	0.6
Good	11.2	18.3	58.5	10.7	1.3
Fair	2.0	5.0	20.7	63.9	8.4
Poor	1.8	4.2	7.2	29.9	56.9
Ages 50 to 74					
	Excellent	Very Good	Good	Fair	Poor
Excellent	66.9	21.3	8.7	2.4	0.6
Very Good	11.9	55.6	24.7	6.2	1.5
Good	5.2	11.3	60.3	20.5	2.6
Fair	0.5	1.8	10.4	73.3	14.1
Poor	0.3	0.8	3.0	24.5	71.5
Ages 75 and older					
	Excellent	Very Good	Good	Fair	Poor
Excellent	52.1	18.7	14.1	10.2	4.9
Very Good	8.9	43.3	28.0	13.5	6.3
Good	3.5	9.5	46.3	29.4	11.3
Fair	0.5	1.7	7.2	62.4	28.2
Poor	0.1	0.4	1.7	15.5	82.3

Table 31: Two-year transition probabilities (%) across frailty health categories within three age groups: 25–49, 50–74 year-olds, and ages 75 and older constructed using the MEPS sample. Rows sum to one.

In this section we report the results of a series of OLS regressions with log total medial expenditures as the dependent variable. All the regressions include gender, education, marital status and age polynomials as controls. Panel A in each table shows regression results for the entire sample. In column (1) the regressors include dummies for lagged self-reported health categories. In column (2), the SRHS dummies are replaced with a quadratic in lagged frailty. Column (3) includes both sets of health indicators.

In panel B in each table repeats the same regression as column (2) of panel A. However, instead of running the regression on the entire sample, it reports results of five different regressions, one for each SRHS category. In each these regressions, SRHS has no explanatory/predictive power.

We report the main regression which is run on entire MEPS sample in Table 3 of the paper. In Table 32 we report the same regressions, but on a restricted sample who is younger than 45. The results show that restricting to a younger subs-ample does not alter the results in any way. Frailty remains more powerful than SRHS in predicting medical expenditure, even within each SRHS categories.

Table 33 repeats the same regression as the one reported in Table 3 of the paper, but it includes age interactions. Table 34 repeats the same regression as the one reported in Table 32 with age interactions. Including age interaction does not alter any of the main takeaways.

	Panel A: younger than 45			**Panel B:** by SRHS health at $t-1$				
				'Excellent'	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.344***		0.210***					
	(0.027)		(0.026)					
good$_{t-1}$	0.550***		0.227***					
	(0.028)		(0.028)					
fair$_{t-1}$	1.189***		0.422***					
	(0.040)		(0.042)					
poor$_{t-1}$	2.623***		1.048***					
	(0.072)		(0.078)					
frailty$_{t-1}$		15.355***	14.188***	13.067***	15.710***	14.692***	10.239***	5.848***
		(0.298)	(0.312)	(0.832)	(0.626)	(0.573)	(0.817)	(1.247)
frailty$^2_{t-1}$		-16.608***	-16.256***	-19.398***	-23.800***	-18.660***	-9.755***	-5.080**
		(0.660)	(0.675)	(2.908)	(1.918)	(1.407)	(1.531)	(2.001)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	52,571	52,571	52,571	14,228	18,378	15,021	3,971	973
Adjusted R^2	0.153	0.197	0.201	0.183	0.186	0.185	0.164	0.109

Table 32: Results of OLS regressions of log total wave t medical expenditures on frailty and SRHS using data from a younger (below 45 years old) sub-sample in MEPS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.249***		0.237***					
	(0.057)		(0.055)					
good$_{t-1}$	0.251***		0.111*					
	(0.058)		(0.057)					
fair$_{t-1}$	1.018***		0.250***					
	(0.077)		(0.078)					
poor$_{t-1}$	3.312***		0.992***					
	(0.128)		(0.134)					
excellent$_{t-1}\times$ age	-0.214***		-0.186***					
	(0.039)		(0.038)					
very good$_{t-1}\times$ age	-0.211***		-0.187***					
	(0.039)		(0.038)					
good$_{t-1}\times$ age	-0.205***		-0.184***					
	(0.039)		(0.038)					
fair$_{t-1}\times$ age	-0.207***		-0.184***					
	(0.039)		(0.038)					
poor$_{t-1}\times$ age	-0.231***		-0.191***					
	(0.039)		(0.038)					
frailty$_{t-1}$		21.412***	20.302***	12.920***	21.303***	20.983***	15.381***	8.625***
		(0.488)	(0.509)	(1.522)	(1.023)	(0.910)	(1.188)	(1.681)
frailty$^2_{t-1}$		-22.539***	-22.573***	-17.120***	-32.593***	-24.966***	-14.756***	-6.756***
		(0.901)	(0.919)	(4.310)	(2.467)	(1.821)	(1.958)	(2.382)
frailty$_{t-1}\times$ age		-0.156***	-0.149***	0.030	-0.128***	-0.150***	-0.134***	-0.107***
		(0.009)	(0.009)	(0.028)	(0.018)	(0.016)	(0.020)	(0.029)
frailty$^2_{t-1}\times$ age		0.182***	0.185***	-0.047	0.246***	0.195***	0.142***	0.106***
		(0.015)	(0.015)	(0.069)	(0.038)	(0.029)	(0.032)	(0.039)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	124,706	124,706	124,706	26,224	39,865	38,543	14,915	5,159
Adjusted R^2	0.247	0.310	0.312	0.253	0.278	0.291	0.253	0.119

Table 33: Results of OLS regressions of log total wave t medical expenditures on frailty and SRHS using data from MEPS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

	Panel A: younger than 45			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.250		0.307*					
	(0.167)		(0.163)					
good$_{t-1}$	0.220		0.405**					
	(0.176)		(0.175)					
fair$_{t-1}$	0.589**		0.862***					
	(0.258)		(0.264)					
poor$_{t-1}$	1.855***		2.267***					
	(0.497)		(0.524)					
excellent$_{t-1}\times$ age	-0.035		-0.038					
	(0.199)		(0.193)					
very good$_{t-1}\times$ age	-0.032		-0.041					
	(0.199)		(0.193)					
good$_{t-1}\times$ age	-0.025		-0.043					
	(0.199)		(0.193)					
fair$_{t-1}\times$ age	-0.017		-0.051					
	(0.199)		(0.193)					
poor$_{t-1}\times$ age	-0.013		-0.072					
	(0.199)		(0.194)					
frailty$_{t-1}$		11.964***	9.307***	1.629	2.830	10.537***	10.071*	13.992
		(1.966)	(2.047)	(5.093)	(4.036)	(3.783)	(5.675)	(9.036)
frailty$^2_{t-1}$		-16.390***	-14.370***	-3.408	3.343	-16.825*	-14.812	-18.016
		(4.829)	(4.904)	(18.431)	(13.512)	(10.195)	(11.399)	(15.784)
frailty$_{t-1}\times$ age		0.100*	0.142**	0.336**	0.371***	0.122	0.009	-0.222
		(0.055)	(0.057)	(0.147)	(0.114)	(0.105)	(0.155)	(0.243)
frailty$^2_{t-1}\times$ age		-0.022	-0.066	-0.476	-0.782**	-0.068	0.128	0.350
		(0.130)	(0.132)	(0.522)	(0.374)	(0.275)	(0.306)	(0.417)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	52,555	52,555	52,555	14,221	18,374	15,017	3,970	973
Adjusted R^2	0.153	0.197	0.201	0.183	0.187	0.185	0.164	0.107

Table 34: Results of OLS regressions of log total wave t medical expenditures on frailty and SRHS and age interactions using data from a younger (below 45 years old) sub-sample in MEPS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

B.4 Comparisons to other variants of frailty in MEPS

Table 35 shows summary statistics for the frailty along with its three variants in MEPS sample. In column (1) we repeat the summary stats for frailty. Columns (2) through (4) show stats for frailty with SRHS, the FPC index, and the SIO index, respectively. Figure 9 shows boxplots of these indices by age and histograms showing their cross-sectional distributions.

Finally, Tables 36 , 37 , and 38 show two-year transition probabilities across quintiles of frailty-with-SRHS, FPC, and SIO, respectively for three age groups of 25–49, 50–74 year-olds, and ages 75 in the MEPS sample.

	Frailty (1)	Frailty w/SRHS (2)	FPC (3)	SIO (4)
Mean	0.11	0.11	0.10	0.11
men	0.10	0.10	0.09	0.10
women	0.12	0.12	0.12	0.13
ages 25-49	0.06	0.06	0.05	0.06
ages 50-74	0.15	0.15	0.15	0.15
ages 75+	0.28	0.28	0.30	0.30
less than HS	0.15	0.15	0.15	0.17
HS grad	0.11	0.11	0.10	0.11
college grad	0.07	0.07	0.06	0.06
Standard Dev	0.15	0.15	0.19	0.21
men	0.14	0.14	0.17	0.19
women	0.16	0.17	0.20	0.22
ages 25-49	0.09	0.09	0.11	0.14
ages 50-74	0.17	0.17	0.21	0.23
ages 75+	0.21	0.21	0.28	0.30
less than HS	0.19	0.19	0.23	0.26
HS grad	0.15	0.15	0.18	0.20
college grad	0.11	0.11	0.13	0.14
Min	0.00	0.00	0.00	0.00
5th percentile	0.00	0.00	0.00	0.00
50th percentile	0.04	0.04	0.03	0.03
95th percentile	0.49	0.50	0.62	0.71
Max	1.00	1.00	1.00	1.00
Fraction at zero	0.19	0.18	0.18	0.19
Observations	353509	353509	353509	335275
Individuals	195888	195888	195888	185990

Table 35: Summary statistics for four versions of the frailty index: frailty, frailty with SRHS, FPC index, and SIO index using MEPS data on respondents over the age 25.

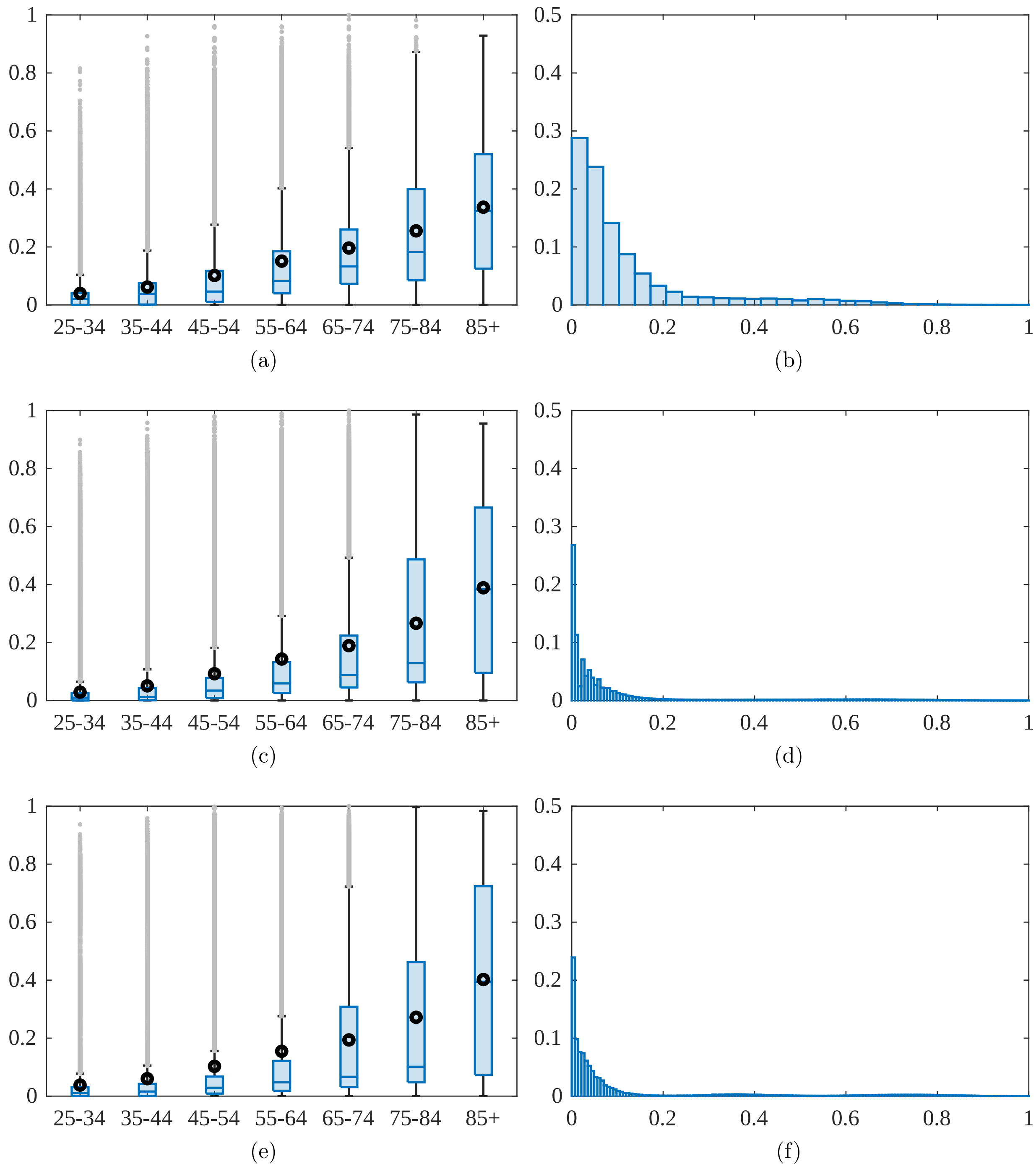


Figure 9: Box plots of frailty with SRHS (a), FPC index (c), and SIO index (e) by 10-year age groups in the MEPS sample. The bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The middle lines are medians and the open circles are means. The upper whiskers extend to 1.5 times the inner quartile range. The lower whiskers extend to the minimum value of frailty in the age group. The dots are data points that lie beyond the whiskers. Histograms showing the cross-sectional distribution of frailty with SRHS (b), FPC index (d), and SIO index (f) in the sample.

	Bottom	2nd	3rd	4th	Top
Ages 25 to 49					
Bottom	69.6	11.2	14.1	3.5	1.6
2nd	9.6	68.9	9.3	10.4	1.9
3rd	21.8	10.3	45.1	17.5	5.4
4th	2.6	8.9	15.9	55.4	17.2
Top	2.4	1.5	5.9	18.8	71.4
Ages 50 to 74					
	Bottom	2nd	3rd	4th	Top
Bottom	75.1	16.8	5.5	1.5	1.2
2nd	17.7	55.6	18.7	5.5	2.5
3rd	4.0	15.6	54.6	19.6	6.1
4th	1.1	3.9	17.0	57.4	20.7
Top	0.6	1.6	5.0	18.1	74.7
Ages 75 and older					
	Bottom	2nd	3rd	4th	Top
Bottom	67.6	17.0	7.5	5.9	1.9
2nd	14.8	45.3	22.2	13.4	4.4
3rd	6.2	14.6	42.9	24.6	11.7
4th	2.6	6.6	15.8	44.8	30.2
Top	0.3	0.8	5.8	16.9	76.3

Table 36: Two-year transition probabilities (%) across quintiles of frailty-with-SRHS for three age groups: 25–49, 50–74 year-olds, and ages 75 and older constructed using the MEPS sample. Rows sum to one.

Ages 25 to 49					
	Bottom	2nd	3rd	4th	Top
Bottom	70.3	13.5	7.2	6.8	2.3
2nd	10.8	65.0	14.2	7.4	2.6
3rd	13.4	15.3	52.0	13.4	5.8
4th	4.1	4.5	8.5	64.3	18.5
Top	1.7	1.9	4.4	15.4	76.6
Ages 50 to 74					
	Bottom	2nd	3rd	4th	Top
Bottom	74.9	16.8	4.8	1.7	1.7
2nd	4.3	66.3	19.6	5.7	4.1
3rd	1.8	7.5	62.2	20.8	7.6
4th	0.8	2.6	13.0	62.0	21.6
Top	0.8	2.4	6.1	21.1	69.5
Ages 75 and older					
	Bottom	2nd	3rd	4th	Top
Bottom	63.2	18.7	8.1	7.4	2.6
2nd	6.9	49.7	23.5	14.9	5.0
3rd	4.3	13.1	45.4	22.2	15.0
4th	3.3	7.9	18.8	41.6	28.4
Top	0.2	1.6	7.3	14.1	76.9

Table 37: Two-year transition probabilities (%) across quintiles of FPC for three age groups: 25–49, 50–74 year-olds, and ages 75 and older constructed using the MEPS sample. Rows sum to one.

	Ages 25 to 49				
	Bottom	2nd	3rd	4th	Top
Bottom	65.5	14.8	10.1	5.9	3.7
2nd	10.2	56.9	18.9	9.7	4.4
3rd	5.8	18.7	48.4	20.4	6.8
4th	2.9	7.5	18.9	51.4	19.3
Top	2.3	3.2	6.1	18.5	69.9
	Ages 50 to 74				
	Bottom	2nd	3rd	4th	Top
Bottom	70.9	18.8	6.2	2.5	1.7
2nd	10.9	59.0	20.4	6.3	3.5
3rd	2.5	12.4	57.0	20.4	7.8
4th	1.5	3.7	15.0	58.8	21.0
Top	1.0	2.4	6.5	19.7	70.4
	Ages 75 and older				
	Bottom	2nd	3rd	4th	Top
Bottom	64.4	18.0	8.3	6.8	2.5
2nd	7.1	49.5	23.8	14.1	5.4
3rd	5.7	15.3	42.8	23.7	12.4
4th	2.6	7.9	17.9	38.1	33.5
Top	0.5	1.9	6.1	22.0	69.5

Table 38: Two-year transition probabilities (%) across quintiles of SIO for three age groups: 25–49, 50–74 year-olds, and ages 75 and older constructed using the MEPS sample. Rows sum to one.

C HRS

In this section we briefly describe the HRS sample we use for our analysis and report the list of deficit variables we use for construction of our health measures. We also include additional details regarding baseline Frailty and other measures in HRS.

C.1 HRS sample and list of deficit variables

The HRS is a biennial longitudinal survey of Americans over age 50. We use the HRS waves spanning the period 1996 to 2016. Our sample consists of respondents aged 25 years and above. Our sample consists of 226,402 observations of 40,288 individuals (17,264 men and 23,024 women). Table 43 lists the 36 variables we use to construct the respondents' frailty index and it variants. This list consists of 20 ADL/IADL variables, 12 severe health diagnosis, body measurement and smoking variables, 1 cognitive test variable, and 4 health care utilization (doctor visits, etc) variables. The indices are constructed in the same way as for PSID respondents.

The advantage of HRS over PSID is that it contains a larger number of deficit variables, and includes information on cognitive limitations. Specifically, the HRS includes a variable called Cognitive Functioning. This is a RAND-HRS variable that aggregates scores on the following cognitive tests: Immediate Word Recall, Delayed Word Recall, Serial 7 Test, Backwards Count, Identifying Objects, Identifying President/Vice President, and Identifying Date. The result is a score between 0 and 35, with 35 meaning no cognitive impairment. We first inverse the score, so higher score imply more severe impairment. We then normalize it to a number between 0 and 1 and use the resulting variable as a deficit variable.

We also construct all other variants of frailty in HRS (as we do in PSID). Table 40 reports the weights associated with each deficit for each of the alternative health indices. The deficits are order according to their weights in the first principal component (FPC) index. We observe that in HRS, similar to PSID, first principal component puts highest weight on ADLs and IADLs variables. In particular, aggregating the weights across deficits by category we find that it puts 70 percent of the total weight on the ADL/IADL variables, 17 percent on the medical diagnosis/measurement variables, 2 percent on the mental/cognitive health variables, and 4 percent on SRHS status. The remaining 7 percent is the weight in health care utilization (such as hospital and doctor visits).

In contrast, frailty (and frailty with SRHS) equally weight each deficit. Consequently, frailty puts 56 percent of the weight on the ADL/IADL variables, 31 percent on the medical variables, 3 percent on mental health, and (since it is omitted) zero weight on SRHS. Frailty also puts 11 percent on health care utilization. These breakdowns, are a little different than those in PSID. This is due to the fact that HRS includes larger number ADL/IADL deficits, relative to PSID (and MEPS).

C.1.1 Regression tables for construction of SIO in HRS

We follow the same two step procedure to construct SIO in HRS, as we do in PSID. The principal component weights for the first step are reported in Table 41. The second step regression results are reported in 42

Variable	Value
Some difficulty with ADL/IADLs:	
Eating	Yes=1, No=0
Dressing	Yes=1, No=0
Getting in/out of bed	Yes=1, No=0
Using the toilet	Yes=1, No=0
Bathing/shower	Yes=1, No=0
Walking across room	Yes=1, No=0
Walking several blocks	Yes=1, No=0
Using the telephone	Yes=1, No=0
Managing money	Yes=1, No=0
Shopping for groceries	Yes=1, No=0
Preparing meals	Yes=1, No=0
Getting up from chair	Yes=1, No=0
Stooping/kneeling/crouching	Yes=1, No=0
Lift/carry 10 lbs	Yes=1, No=0
Using a map	Yes=1, No=0
Taking medications	Yes=1, No=0
Climbing 1 flight of stairs	Yes=1, No=0
Picking up a dime	Yes=1, No=0
Reaching/ extending arms up	Yes=1, No=0
Pushing/pulling large objects	Yes=1, No=0
Ever had one of following conditions:	
High Blood Pressure	Yes=1, No=0
Diabetes	Yes=1, No=0
Cancer	Yes=1, No=0
Lung disease	Yes=1, No=0
Heart disease	Yes=1, No=0
Stroke	Yes=1, No=0
Psychological problems	Yes=1, No=0
Arthritis	Yes=1, No=0
BMI $\geq$ 30	Yes=1, No=0
Has ever smoked	Yes=1, No=0
Back pain	Yes=1, No=0
Doctor visit	Yes=1, No=0
Hospital visit	Yes=1, No=0
Home care visit	Yes=1, No=0
Nursing home stay	Yes=1, No=0
Cognitive impairment score	0-35, inverted and rescaled to 0–1
The following variable is only used in construction of Frailty-with-SRHD and FPC	
Self-reported health of 'fair' or 'poor'	Yes=1, No=0

Table 39: Health variables used to construct frailty in HRS.

Variable	FPC (1)	SIO (2)	Frailty with SRHS (3)	Baseline Frailty (4)	BBDF (5)
Difficulty walking several blocks	0.045	0.107	0.027	0.028	n.a.
Difficulty lift/carry 10 lbs	0.044	0.061	0.027	0.028	n.a.
Difficulty shopping for groceries	0.043	0.039	0.027	0.028	n.a.
Difficulty climbing 1 flight of stairs	0.043	0.037	0.027	0.028	n.a.
Difficulty pushing/pulling large objects	0.042	0.078	0.027	0.028	n.a.
Difficulty bathing/showering	0.040	-0.013	0.027	0.028	n.a.
Difficulty dressing	0.040	0.017	0.027	0.028	n.a.
Difficulty walk across room	0.039	-0.018	0.027	0.028	n.a.
Difficulty preparing meals	0.037	0.019	0.027	0.028	n.a.
Difficulty getting up from chair	0.036	0.110	0.027	0.028	n.a.
Difficulty stooping/kneeling/crouching	0.036	0.145	0.027	0.028	n.a.
Self-reported health of ''fair' or 'poor'	0.036	n.a.	0.027	n.a.	n.a.
Difficulty getting in/out of bed	0.036	0.000	0.027	0.028	n.a.
Reaching/extending arms up	0.034	0.063	0.027	0.028	n.a.
Difficulty using the toilet	0.033	-0.036	0.027	0.028	n.a.
Difficulty eating	0.028	-0.019	0.027	0.028	n.a.
Difficulty managing money	0.027	0.017	0.027	0.028	n.a.
Difficulty picking up a dime	0.025	0.030	0.027	0.028	n.a.
Ever had arthritis	0.025	0.014	0.027	0.028	0.187
Back pain	0.024	0.024	0.027	0.028	n.a.
Home care visit	0.024	0.010	0.027	0.028	n.a.
Difficulty using the telephone	0.024	0.012	0.027	0.028	n.a.
Difficulty taking medications	0.023	0.006	0.027	0.028	n.a.
Cognitive impairment score	0.023	0.097	0.027	0.028	n.a.
Hospital visit	0.023	0.024	0.027	0.028	n.a.
Difficulty using a map	0.022	0.016	0.027	0.028	n.a.
Ever had psychological problems	0.020	0.021	0.027	0.028	0.133
Ever had heart disease	0.019	0.026	0.027	0.028	0.119
Nursing home stay	0.018	-0.019	0.027	0.028	n.a.
Ever had lung disease	0.017	0.049	0.027	0.028	0.175
Ever had stroke	0.016	0.020	0.027	0.028	0.155
Ever had high blood pressure	0.015	0.008	0.027	0.028	0.056
Ever had diabetes	0.015	0.029	0.027	0.028	0.118
BMI >= 30	0.011	0.004	0.027	0.028	n.a.
Ever had cancer	0.007	0.017	0.027	0.028	0.056
Doctor visit	0.006	-0.003	0.027	0.028	n.a.
Ever smoked	0.004	0.007	0.027	0.028	n.a.

Table 40: Weights assigned to each deficit in each variations of frailty. Column (1) show first principal component (FPC) weights. Column (2) shows SIO wieghts. Column (3) is the equal weight frailty index that includes SRHS. Column (4) is equal weight frailty index that excludes SRHS (baseline frailty index). Column (5) is weights assigned to deficits in BBDF index. Deficist variables are sorted according to FPC weights. All weights are based on HRS data on household heads and their spouses ages 25 and older.

Health limits activities	0.333
Health limits work	0.343
Self-reported health of ‘fair’ or ‘poor’	0.325

Table 41: Principal component weights for first stage calculation of SIO index (HRS).

Variable	Value	Variable	Value
Ever had cancer	0.023*** (0.002)	Difficulty bathing, showering	-0.008 (0.007)
Ever had diabetes	0.053*** (0.002)	Difficulty eating	-0.017 (0.011)
Ever had high blood pressure	0.030*** (0.002)	Difficulty getting in, out of bed	0.012* (0.007)
Ever had arthritis	0.017*** (0.002)	Difficulty using the toilet	-0.033*** (0.007)
Ever had psychological problems	0.026*** (0.003)	Difficulty using a map	0.018*** (0.003)
Ever had lung disease	0.056*** (0.003)	Difficulty preparing meals	0.007 (0.008)
Ever had stroke	0.020*** (0.004)	Difficulty shopping for groceries	0.023*** (0.007)
Ever had heart disease	0.039*** (0.002)	Difficulty using the telephone	0.013* (0.008)
Difficulty walking several blocks	0.080*** (0.003)	Difficulty taking medications	0.001 (0.008)
Difficulty getting up from chair	0.068*** (0.002)	Difficulty managing money	0.016*** (0.006)
Difficulty climbing 1 flight of stairs	0.032*** (0.003)	BMI $\geq$ 30	0.018*** (0.002)
Difficulty stooping, kneeling, crouching	0.087*** (0.002)	Cognitive impairment score	0.155*** (0.006)
Difficulty reaching, extending arms up	0.046*** (0.003)	Ever smoked	0.008*** (0.001)
pushing, pulling large objects	0.053*** (0.003)	Back pain	0.032*** (0.002)
Difficulty lift, carry 10 lbs	0.042*** (0.003)	Hospital visit	0.034*** (0.002)
Difficulty picking up a dime	0.015*** (0.004)	Nursing home stay	-0.031*** (0.008)
Difficulty dressing	0.019*** (0.005)	Doctor visit	-0.008*** (0.003)
Difficulty walk across room	-0.016** (0.007)	Home care visit	0.003 (0.004)
Observations	34,461		
Adjusted R^2	0.525		

Table 42: SIO index second step regression (HRS). Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

C.2 Facts on frailty dynamics over the life cycle in HRS

We report summary statistics on distribution of frailty in our HRS sample in Table 43. Note that HRS is an older sample relative to PSID. It only includes respondent who are older than 50. For this reason frailty in HRS sample has higher mean and higher standard deviation relative to PSID. [3] Also, the fraction of observations with frailty of zero is less than 1 percent.

However, other patters are remarkably similar to PSID. In all three age groups presented in the table, women have higher frailty on average than men but only slightly. Also, differences in frailty are larger by education than by gender. In summary, in our HRS sample, mean frailty increases with age, decreases with education, and is slightly higher for women than for men.

Figure 10a shows box and whisker plots of frailty by 10-year age groups constructed using the HRS sample. This figure shows that in HRS, similar to PSID, the gap between the top and bottom quantiles of the distributions widens with age. Also, the figure shows significant variation in frailty in all of the age groups above 50 year of age.

Figure 10b shows the histogram of frailty for the entire HRS sample. This figures demonstrates that, similar to PSID, the cross-sectional distribution of frailty in HRS is right-skewed. However, due to lower fraction of zeros (and very small values) the skewness is less pronounced.

Finally Table 44 shows two-year transition probabilities across frailty quintiles for two age groups: 50 to 74 year-olds, and individuals 75 and older. The tables shows that, similar to PSID, frailty in HRS is highly persistent. Within both age groups, individuals are most likely to remain in the same frailty quintile and the probability of moving to another quintile is rapidly declining in its distance from the previous one. Also, frailty is more persistent at the top and bottom quintiles. However, unlike PSID, we don't observer a clear pattern of decline (or increase) in persistence with age. Also, the persistence is in general lower at each age group (judged by the values on the diagonal) relative to PSID.

[3] See section C.5 of this appendix for a comparison between frailty in HRS and an older than 50 sub sample of PSID.

	All	Men	Women	Less Than HS	HS Grad	College Grad
Mean	0.22	0.20	0.23	0.28	0.21	0.16
ages 50-74	0.19	0.18	0.20	0.25	0.19	0.14
ages 75+	0.29	0.26	0.30	0.33	0.27	0.24
Standard deviation	0.16	0.15	0.17	0.18	0.15	0.13
ages 50-74	0.14	0.13	0.15	0.17	0.14	0.11
ages 75+	0.18	0.17	0.19	0.20	0.17	0.16
Min	0.00	0.00	0.00	0.00	0.00	0.00
5th percentile	0.04	0.04	0.04	0.06	0.06	0.03
50th percentile	0.17	0.16	0.18	0.23	0.17	0.12
95th percentile	0.56	0.51	0.58	0.65	0.52	0.42
Max	0.97	0.97	0.97	0.97	0.97	0.91
Fraction at zero	0.00	0.01	0.00	0.00	0.00	0.01
Observations	226402	95598	130803	54864	127236	44255
Individuals	40288	17264	23024	9360	22525	8397
% of total	100	43	57	23	56	21

Table 43: Summary statistics for frailty constructed using 1996–2016 HRS respondents and spouses over the age 50.

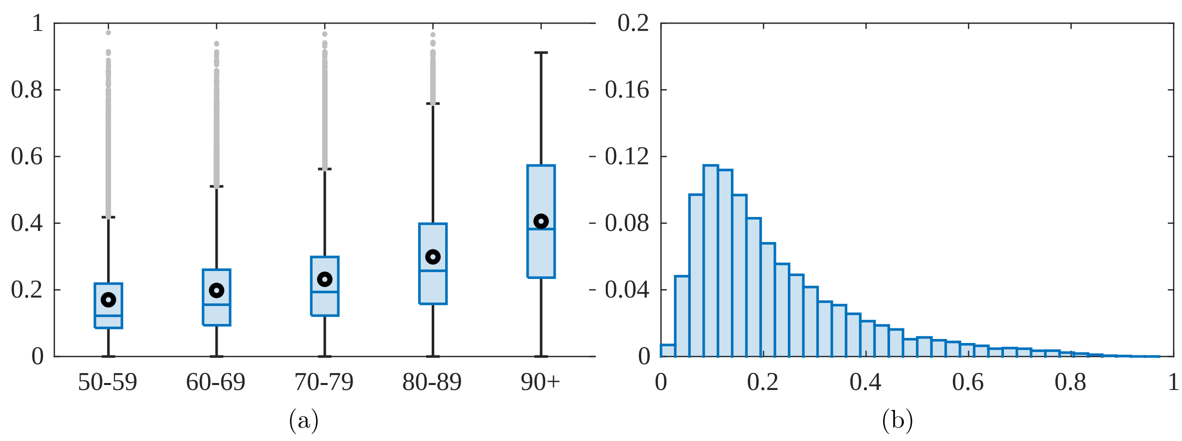


Figure 10: Panel (a) is box plots of frailty by 10-year age groups in the HRS sample. The bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The middle lines are medians and the open circles are means. The upper whiskers extend to 1.5 times the inner quartile range. The lower whiskers extend to the minimum value of frailty in the age group. The dots are data points that lie beyond the whiskers. Panel (b) is a histogram showing the cross-sectional distribution of frailty in the sample.

	Bottom	2nd	3rd	4th	Top
Ages 50 to 74					
Bottom	64.8	24.4	7.3	2.6	0.8
2nd	18.6	45.2	25.7	8.6	1.9
3rd	4.1	21.6	41.1	27.3	5.9
4th	0.8	5.1	20.8	49.2	24.0
Top	0.1	0.5	2.6	17.1	79.8
Ages 75 and older					
	Bottom	2nd	3rd	4th	Top
Bottom	58.5	25.1	9.9	3.9	2.6
2nd	13.3	38.3	29.8	12.8	5.8
3rd	2.6	17.3	37.9	30.7	11.5
4th	0.4	3.6	16.3	45.7	34.0
Top	0.0	0.5	2.2	13.3	84.0

Table 44: Two-year transition probabilities (%) across frailty quintiles for two age groups: 50–74 year-olds, and ages 75 and older constructed using the HRS sample. Rows sum to one.

C.3 Comparison of the frailty index to SRHS in HRS

C.3.1 Health dynamics

We compare dynamics of the distribution of frailty over the life cycle to that of SRHS in HRS. To do this, we follow the same procedures outlines in the paper and reported for PSID sample. The difference is that we choose cut offs of frailty so that the partitions in the distribution of frailty have the same share as SRHS for age group 50 to 55 (instead of 25 to 29). We refer to these partitions as frailty health category we label them 'excellent', 'very good', 'good', 'fair', and 'poor'. The we label an individual in 'excellent' frailty category if their frailty falls below the lower cut off. Other categories are define accordingly as laid out in details in the paper.

The results are reported in Figure 11. For each age group, the height of each shaded area is the fraction of individuals in the corresponding SRHS category. The dashed lines demonstrate the boundaries of each frailty category at each age group. The cut offs are reported below the figure.

The patter here is very similar to the one in PSID (reported in the paper). Share of SRHS 'excellent' and 'very good' goes from 57 percent in age group 50-55 to close to 27 percent in in age group 85+. At the same time, the fraction of 'excellent' and 'very good' frailty category falls to less than 10 percent in age group 85+.

Similarly the fraction of 'fair' or 'poor' frailty category expands at more rapid pace than that of SRHS. While there are only 42 percent of 85+ age group have SRHS of 'fair' or 'poor', more than 70 percent have frailty category of 'fair' or 'poor'. Health, measured by frailty, deteriorates faster than health measured by SRHS.

C.3.2 Dispersion

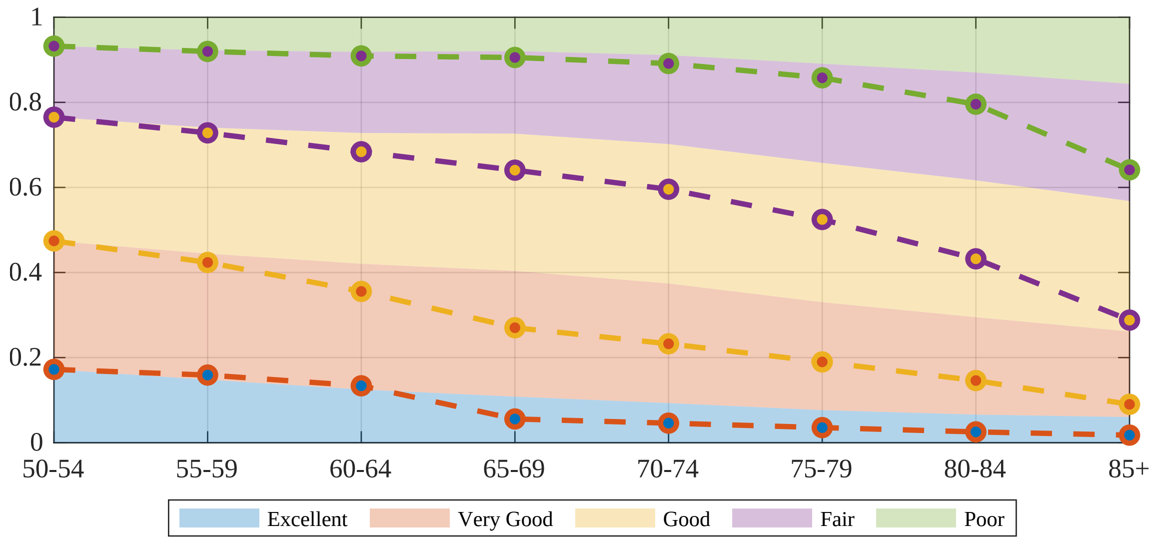
Figure 12 shows box plots for the distribution of frailty by SRHS categories and 10-year age groups. The bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The middle line indicates the median. Within each age group, median frailty increases across SRHS categories indicating that the two health measures are positively correlated. Notice that in all the age groups the inter-quartile range also increases as SRHS declines indicating that there is more variation in frailty among individuals who self-report worse health.

C.3.3 Persistence

Table 45 shows two-year transition probabilities across SRHS categories by 25 year age groups in MEPS. Table 46 shows the same, but for transition across frailty health categories, as defined above. The patterns here are similar to PSID. Frailty health categories exhibit more more persistence than SRHS categories, specially for 'poor' health category. However, the difference in persistence (judged by values on the diagonal) is small.

C.3.4 Predicting mortality

Tables 47, 48, 49, and 50 report additional results for probit regressions for mortality. In each probit estimation we regress a variable that is equal to one if an individual has died



	excellent	very good	good	fair	poor
% of SRHS in age 50 - 55	17.22	30.21	29.1	16.7	6.8
Frailty range	[0, 0.059)	[0.059, 0.117)	[0.117, 0.212)	[0.212, 0.406)	[0.406, 1]

Figure 11: Distribution of health status by age. The colored areas show the fraction of individuals by SRHS at each age. The dashed lines show the fraction of individuals by the corresponding frailty category at each age. Source: authors' calculations using HRS.

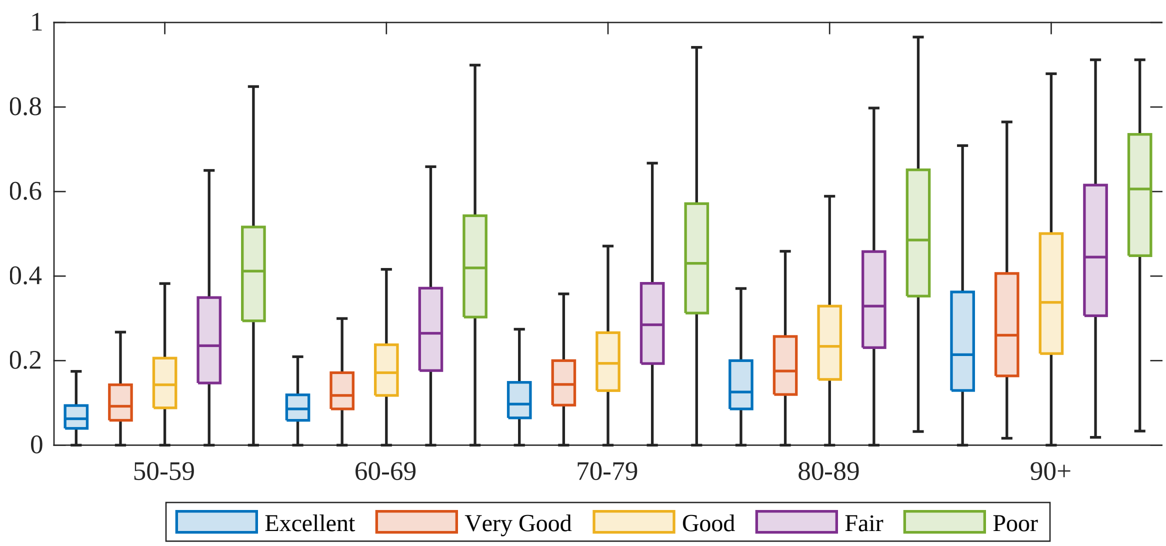


Figure 12: Box plots of frailty by SRHS and 10-year age groups in the HRS sample. The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The middle line are medians. The upper whiskers extend to 1.5 times the inner quartile range. The lower whiskers extend to the minimum value of frailty in the subgroup.

	Ages 50 to 74				
	Excellent	Very Good	Good	Fair	Poor
Excellent	50.8	35.1	10.8	2.7	0.6
Very Good	11.9	56.0	26.0	5.1	0.9
Good	3.3	22.4	53.3	17.9	3.1
Fair	1.5	6.4	27.1	51.4	13.7
Poor	0.6	2.1	8.2	33.0	56.0
	Ages 75 and older				
	Excellent	Very Good	Good	Fair	Poor
Excellent	38.7	35.5	16.3	6.1	3.4
Very Good	8.8	45.8	31.4	10.2	3.7
Good	2.6	19.0	48.1	23.8	6.6
Fair	1.3	6.8	24.0	48.7	19.3
Poor	0.9	3.3	11.2	30.5	54.1

Table 45: Two-year transition probabilities (%) across self-reported health status categories within two age groups: 50–74 year-olds, and ages 75 and older constructed using the HRS sample. Rows sum to one.

	Ages 50 to 74				
	Excellent	Very Good	Good	Fair	Poor
Excellent	57.6	32.5	8.5	1.1	0.2
Very Good	15.5	49.6	30.9	3.6	0.4
Good	1.5	16.5	60.0	20.5	1.5
Fair	0.1	1.4	21.3	63.1	14.1
Poor	0.0	0.2	1.6	24.0	74.1
	Ages 75 and older				
	Excellent	Very Good	Good	Fair	Poor
Excellent	39.7	36.1	14.9	6.5	2.7
Very Good	5.3	44.8	36.7	10.0	3.2
Good	0.3	8.5	51.3	34.1	5.8
Fair	0.0	0.6	13.2	60.3	25.8
Poor	0.0	0.1	0.7	15.0	84.3

Table 46: Two-year transition probabilities (%) across frailty health categories within two age groups: 50–74 year-olds, and ages 75 and older constructed using the HRS sample. Rows sum to one.

between wave t -1 and wave t on lagged SRHS categories and/or lagged frailty polynomial. All regression also include the following controls: gender, education, marital status and a quadratic in age.

In Panel A of Table 47 reports results of entire sample. Column (1) shows result of regression in lagged SRHS categories. Column (2) shows results of regression on lagged frailty polynomial. Column (3) report the results when both are included. To evaluate and compare the performance of each health measure we report the *pseudo R-squared* for each regression.

Columns (1) and (2) indicate that variations in both frailty and SRHS are informative about mortality. However, the pseudo R-squared in column (2) is higher than in column (1), indicating that frailty is more informative about mortality than SRHS. Column (3) shows that both health measures remain statistically significant when included together indicating that they both have independent predictive power. However, notice the large decline in the coefficients on the SRHS dummies moving from column (1) to (3), as well as, the magnitude of the changes in the pseudo R-squared's across columns. Both indicate that the SRHS's additional impact is relatively small.

In panel B of Table 47 we repeat the same regression as column (2) of panel A. However, instead of running the regression on the entire sample, we run five different regressions, one for each SRHS category. In each these regressions, SRHS has no explanatory/predictive power. However, in each category, the coefficients on the frailty index are highly significant. Notice that frailty is highly informative about mortality even among individuals who report 'excellent' health. This result is suggestive evidence that, even among healthier individuals, frailty is informative about true underlying health.

In Tables 48 and49 we repeat the same regressions, except we use 2 and 3 lagged values for SRHS and frailty, respectively. As expected, with 2 and 3 lagged values of SRHS and frailty, the predictive power of both health measures decline. However, our main finding stands.

Finally, in Table 50 we repeat the same exercise by including age interactions. Although these age interactions are significant, they including them does not alter the results.

C.3.5 Predicting nursing home entry

Tables 51, 52, 53, and 54 report additional results for probit regressions for transition to nursing home. In each probit estimation we regress a variable that is equal to one if an individual has moved to nursing home between wave t-1 and wave t on lagged SRHS categories and/or lagged frailty polynomial. All regression also include the following controls: gender, education, marital status and a quadratic in age.

In Panel A of Table 51 reports results of entire sample. Column (1) shows result of regression in lagged SRHS categories. Column (2) shows results of regression on lagged frailty polynomial. Column (3) report the results when both are included. To evaluate and compare the performance of each health measure we report the *pseudo R-squared* for each regression.

Columns (1) and (2) indicate that variations in both frailty and SRHS are informative about nursing home entry. However, the pseudo R-squared in column (2) is higher than in column (1), indicating that frailty is more informative about nursing home entry than SRHS.

	Panel A: everyone			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.053**		-0.007					
	(0.024)		(0.025)					
good$_{t-1}$	0.293***		0.120***					
	(0.023)		(0.024)					
fair$_{t-1}$	0.649***		0.300***					
	(0.023)		(0.025)					
poor$_{t-1}$	1.186***		0.570***					
	(0.024)		(0.027)					
frailty$_{t-1}$		2.970***	1.886***	2.871***	1.936***	1.603***	1.406***	-0.169
		(0.098)	(0.107)	(0.529)	(0.299)	(0.219)	(0.216)	(0.287)
frailty$^2_{t-1}$		-0.490***	0.105	-0.510	0.518	0.729**	0.570**	1.989***
		(0.120)	(0.126)	(0.861)	(0.453)	(0.300)	(0.254)	(0.287)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	212,978	212,978	212,978	24,273	59,674	66,083	43,342	19,606
Pseudo R^2	0.217	0.241	0.251	0.234	0.216	0.189	0.165	0.160

Table 47: Results of probit regressions of death between wave t-1 a t, on frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-2$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-2}$	0.114***		0.053**					
	(0.024)		(0.024)					
good$_{t-2}$	0.329***		0.173***					
	(0.023)		(0.024)					
fair$_{t-2}$	0.621***		0.314***					
	(0.023)		(0.025)					
poor$_{t-2}$	0.974***		0.450***					
	(0.025)		(0.028)					
frailty$_{t-2}$		2.831***	1.956***	2.810***	1.845***	1.346***	0.711***	-1.046***
		(0.110)	(0.119)	(0.611)	(0.325)	(0.241)	(0.223)	(0.267)
frailty$^2_{t-2}$		-0.887***	-0.316**	-0.873	0.141	0.733**	0.809***	2.217***
		(0.147)	(0.153)	(1.077)	(0.525)	(0.355)	(0.291)	(0.306)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	172,030	172,030	172,030	18,198	48,880	53,906	35,251	15,680
Pseudo R^2	0.186	0.203	0.208	0.224	0.202	0.176	0.140	0.125

Table 48: Results of probit regressions of death between wave t-1 a t, on two periods lagged values of frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-2$. Controls are gender, education, marital status and a quadratic in age. Note: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

	Panel A: everyone			**Panel B:** by SRHS health at $t-3$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-3}$	0.089***		0.030					
	(0.023)		(0.024)					
good$_{t-3}$	0.311***		0.167***					
	(0.023)		(0.024)					
fair$_{t-3}$	0.544***		0.269***					
	(0.023)		(0.026)					
poor$_{t-3}$	0.884***		0.423***					
	(0.026)		(0.030)					
frailty$_{t-3}$		2.743***	1.946***	2.743***	1.855***	1.303***	0.617**	-1.011***
		(0.124)	(0.134)	(0.690)	(0.373)	(0.269)	(0.250)	(0.282)
frailty$^2_{t-3}$		-1.080***	-0.587***	-1.268	-0.266	0.615	0.646*	2.050***
		(0.180)	(0.187)	(1.241)	(0.651)	(0.423)	(0.356)	(0.352)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	139,757	139,757	139,757	13,536	39,295	44,423	29,350	13,048
Pseudo R^2	0.174	0.185	0.189	0.220	0.196	0.168	0.129	0.111

Table 49: Results of Probit regressions of death between wave t-1 a t, on three periods lagged values of frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-3$. Controls are gender, education, marital status and a quadratic in age. Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.105		0.136					
	(0.162)		(0.166)					
good$_{t-1}$	0.598***		0.649***					
	(0.152)		(0.159)					
fair$_{t-1}$	1.219***		1.185***					
	(0.150)		(0.165)					
poor$_{t-1}$	1.780***		1.578***					
	(0.154)		(0.178)					
very good$_{t-1}\times$ age	-0.001		-0.002					
	(0.002)		(0.002)					
good$_{t-1}\times$ age	-0.004**		-0.007***					
	(0.002)		(0.002)					
fair$_{t-1}\times$ age	-0.008***		-0.012***					
	(0.002)		(0.002)					
poor$_{t-1}\times$ age	-0.008***		-0.013***					
	(0.002)		(0.002)					
frailty$_{t-1}$		5.131***	0.622	5.186	-1.794	-0.591	1.180	
		(0.653)	(0.716)	(3.415)	(2.037)	(1.455)	(1.422)	
frailty$^2_{t-1}$		-3.371***	-0.228	-3.864	6.109*	4.161**	-0.982	
		(0.826)	(0.866)	(5.898)	(3.299)	(2.100)	(1.753)	
frailty$_{t-1}\times$ age		-0.029***	0.017*	-0.031	0.049*	0.028	0.006	0.000
		(0.009)	(0.010)	(0.045)	(0.026)	(0.019)	(0.019)	(0.004)
frailty$^2_{t-1}\times$ age		0.038***	0.004	0.044	-0.072*	-0.044	0.016	0.024***
		(0.011)	(0.011)	(0.074)	(0.041)	(0.027)	(0.023)	(0.004)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	212,978	212,978	212,978	24,273	59,674	66,083	43,342	19,606
Pseudo R^2	0.217	0.241	0.252	0.235	0.216	0.190	0.165	0.160

Table 50: Results of Probit regressions of death between wave t-1 a t, on frailty, SRHS and age interactions using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Column (3) shows that SRHS lose significant when it is included together with frailty, except for the'poor' SHRS category. There is also large decline in the coefficients on the 'poor' SRHS dummy moving from column (1) to (3). Moreover, the magnitude of the changes in the pseudo R-squared's across columns is very small. Thesefacts indicate that the SRHS's additional impact is very small, when frailty is included.

In panel B of Table 51 we repeat the same regression as column (2) of panel A. However, instead of running the regression on the entire sample, we run five different regressions, one for each SRHS category. In each these regressions, SRHS has no explanatory/predictive power. However, in each category, the coefficients on the frailty index are highly significant. Notice that frailty is highly informative about nursing home entry even among individuals who report 'excellent' health. This result is suggestive evidence that, even among healthier individuals, frailty is informative about true underlying health.

In Tables 52 and53 we repeat the same regressions, except we use 2 and 3 lagged values for SRHS and frailty, respectively. As expected, with 2 and 3 lagged values of SRHS and frailty, the predictive power of both health measures decline. However, our main finding stands.

Finally, in Table 54 we repeat the same exercise by including age interactions. They deliver similar results as reported in Table 51.

	Panel A: everyone			**Panel B:** by SRHS health at $t-1$				
				'Excellent'	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.008		-0.064					
	(0.044)		(0.046)					
good$_{t-1}$	0.139***		-0.044					
	(0.042)		(0.045)					
fair$_{t-1}$	0.360***		0.012					
	(0.043)		(0.047)					
poor$_{t-1}$	0.700***		0.125**					
	(0.045)		(0.052)					
frailty$_{t-1}$		1.975***	1.798***	5.796***	2.682***	1.277**	-0.462	-1.431***
		(0.211)	(0.227)	(1.327)	(0.644)	(0.496)	(0.419)	(0.441)
frailty$^2_{t-1}$		0.160	0.160	-6.023***	-0.401	1.158*	2.619***	2.660***
		(0.269)	(0.279)	(2.235)	(0.937)	(0.685)	(0.517)	(0.499)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	168,412	168,412	168,412	16,778	48,162	53,726	35,028	14,599
Pseudo R^2	0.231	0.261	0.263	0.338	0.318	0.257	0.223	0.169

Table 51: Results of probit regressions of transition to nursing home between wave t-1 and t, on frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

	Panel A: everyone			**Panel B:** by SRHS health at $t-2$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-2}$	0.121***		0.066					
	(0.043)		(0.044)					
good$_{t-2}$	0.227***		0.091**					
	(0.041)		(0.043)					
fair$_{t-2}$	0.361***		0.102**					
	(0.043)		(0.046)					
poor$_{t-2}$	0.680***		0.246***					
	(0.046)		(0.053)					
frailty$_{t-2}$		1.860***	1.627***	3.805***	2.423***	1.371***	-0.947**	-1.102**
		(0.215)	(0.231)	(1.443)	(0.678)	(0.503)	(0.409)	(0.432)
frailty$^2_{t-2}$		-0.235	-0.188	-5.067*	-0.119	0.565	2.879***	1.701***
		(0.303)	(0.314)	(2.806)	(1.081)	(0.752)	(0.554)	(0.550)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	150,136	150,136	150,136	14,327	42,736	48,306	31,486	13,173
Pseudo R^2	0.216	0.230	0.231	0.298	0.305	0.235	0.199	0.148

Table 52: Results of probit regressions of transition to nursing home between wave t-1 and t, on two periods lagged values of frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-2$. Controls are gender, education, marital status and a quadratic in age. Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-3$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-3}$	0.073*		0.030					
	(0.043)		(0.044)					
good$_{t-3}$	0.203***		0.096**					
	(0.042)		(0.043)					
fair$_{t-3}$	0.352***		0.139***					
	(0.043)		(0.047)					
poor$_{t-3}$	0.586***		0.232***					
	(0.050)		(0.057)					
frailty$_{t-3}$		1.780***	1.414***	4.924***	3.404***	0.779	-0.300	-1.539***
		(0.239)	(0.256)	(1.785)	(0.860)	(0.555)	(0.450)	(0.469)
frailty$^2_{t-3}$		-0.419	-0.227	-8.601**	-2.610	1.054	1.855***	2.102***
		(0.361)	(0.374)	(4.088)	(1.598)	(0.887)	(0.650)	(0.636)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	121,334	121,334	121,334	11,191	34,826	39,261	25,571	10,388
Pseudo R^2	0.203	0.211	0.213	0.282	0.291	0.217	0.184	0.138

Table 53: Results of probit regressions of transition to nursing home between wave t-1 and t, on three periods lagged values of frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-3$. Controls are gender, education, marital status and a quadratic in age. Note: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-1$				
				‘Excellent	‘Very good’	‘Good’	‘Fair’	‘Poor’
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.758*		0.560					
	(0.401)		(0.412)					
good$_{t-1}$	1.149***		0.656					
	(0.384)		(0.403)					
fair$_{t-1}$	1.884***		0.771*					
	(0.381)		(0.416)					
poor$_{t-1}$	2.690***		0.839*					
	(0.391)		(0.444)					
excellent$_{t-1}\times$ age	-0.149***		-0.165***					
	(0.020)		(0.021)					
very good$_{t-1}\times$ age	-0.159***		-0.173***					
	(0.020)		(0.021)					
good$_{t-1}\times$ age	-0.162***		-0.173***					
	(0.020)		(0.021)					
fair$_{t-1}\times$ age	-0.169***		-0.174***					
	(0.020)		(0.021)					
poor$_{t-1}\times$ age	-0.175***		-0.174***					
	(0.020)		(0.021)					
frailty$_{t-1}$		7.511***	3.794**	16.226	5.493	6.158	7.729**	-2.976
		(1.630)	(1.759)	(12.963)	(5.215)	(3.845)	(3.269)	(3.180)
frailty$^2_{t-1}$		-2.558	0.955	-14.730	3.882	0.706	-3.141	6.821*
		(2.004)	(2.083)	(21.969)	(7.343)	(5.089)	(3.913)	(3.555)
frailty$_{t-1}\times$ age		-0.072***	-0.026	-0.132	-0.038	-0.064	-0.105**	0.021
		(0.021)	(0.022)	(0.156)	(0.065)	(0.048)	(0.042)	(0.042)
frailty$^2_{t-1}\times$ age		0.035	-0.010	0.114	-0.047	0.009	0.073	-0.056
		(0.026)	(0.027)	(0.261)	(0.091)	(0.064)	(0.050)	(0.047)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	168,412	168,412	168,412	16,778	48,162	53,726	35,028	14,599
Pseudo R^2	0.236	0.264	0.268	0.340	0.323	0.261	0.226	0.171

Table 54: Results of probit regressions of transition to nursing home between wave t-1 and t, on frailty, SRHS and age interactions using data from HRS. Panel A are results from the full sample while panel B are results obtained using subs-amples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

C.3.6 Predicting becoming SSDI beneficiary

Tables 55, 56, 57, and 58 report additional results for probit regressions for becoming SSDI beneficiary. In each probit estimation we regress a variable that is equal to one if an individual has become SSDI beneficiary between wave t-1 and wave t on lagged SRHS categories and/or lagged frailty polynomial. All regression also include the following controls: gender, education, marital status and a quadratic in age.

In Panel A of Table 55 reports results of entire sample. Column (1) shows result of regression in lagged SRHS categories. Column (2) shows results of regression on lagged frailty polynomial. Column (3) report the results when both are included. To evaluate and compare the performance of each health measure we report the *pseudo R-squared* for each regression.

Columns (1) and (2) indicate that variations in both frailty and SRHS are informative about disability status. However, the pseudo R-squared in column (2) is higher than in column (1), indicating that frailty is more informative in this regard than SRHS. Column (3) shows that frailty, as well as 'frair' and 'poor' SRHS dummies remain statistically significant when included together indicating that they have independent predictive power. However, notice the large decline in the coefficients on the SRHS dummies moving from column (1) to (3), as well as, the magnitude of the changes in the pseudo R-squared's across columns. Both indicate that the SRHS's additional impact is relatively small.

In panel B of Table 55 we repeat the same regression as column (2) of panel A. However, instead of running the regression on the entire sample, we run five different regressions, one for each SRHS category. In each these regressions, SRHS has no explanatory/predictive power. However, in each category, the coefficients on the frailty index are highly significant. Notice that frailty is highly informative about disability status even among individuals who report 'excellent' health. This result is suggestive evidence that, even among healthier individuals, frailty is informative about true underlying health.

In Tables 56 and57 we repeat the same regressions, except we use 2 and 3 lagged values for SRHS and frailty, respectively. As expected, with 2 and 3 lagged values of SRHS and frailty, the preditive power of both health measures decline. However, our main finding stands.

Finally, in Table 58 we repeat the same exercise by including age interactions. This does not alter the main finding reported above.

	Panel A: everyone			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.070		-0.085					
	(0.049)		(0.053)					
good$_{t-1}$	0.418***		0.015					
	(0.046)		(0.051)					
fair$_{t-1}$	0.984***		0.306***					
	(0.046)		(0.053)					
poor$_{t-1}$	1.597***		0.555***					
	(0.049)		(0.058)					
frailty$_{t-1}$		7.275***	6.098***	8.425***	5.810***	5.159***	5.134***	4.171***
		(0.253)	(0.273)	(1.765)	(0.794)	(0.556)	(0.465)	(0.640)
frailty$^2_{t-1}$		-4.929***	-4.387***	-5.888*	-2.030	-1.730*	-2.901***	-2.690***
		(0.368)	(0.384)	(3.535)	(1.407)	(0.905)	(0.662)	(0.780)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	76,513	76,513	76,513	8,975	25,334	23,925	12,378	3,846
Pseudo R^2	0.178	0.239	0.252	0.281	0.206	0.168	0.128	0.075

Table 55: Results of probit regressions of transition to SSDI beneficiary status between wave t-1 and t, on frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-2$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good_{t-2}	0.168***		0.049					
	(0.053)		(0.055)					
good_{t-2}	0.463***		0.187***					
	(0.050)		(0.053)					
fair_{t-2}	0.889***		0.405***					
	(0.051)		(0.057)					
poor_{t-2}	1.403***		0.631***					
	(0.057)		(0.066)					
frailty_{t-2}		5.591***	4.430***	5.940***	3.705***	3.710***	3.032***	1.682**
		(0.279)	(0.301)	(1.911)	(0.959)	(0.576)	(0.509)	(0.684)
frailty^2_{t-2}		-3.758***	-3.159***	-2.073	-0.206	-0.747	-1.252	-0.787
		(0.440)	(0.459)	(3.649)	(1.921)	(0.967)	(0.788)	(0.914)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	55,088	55,088	55,088	5,328	18,275	17,624	9,131	2,776
Pseudo R^2	0.136	0.166	0.177	0.220	0.133	0.116	0.075	0.040

Table 56: Results of probit regressions of transition to SSDI beneficiary status between wave t-1 and t, on two periods lagged values of frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-2$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-3$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-3}$	0.169***		0.055					
	(0.057)		(0.058)					
good$_{t-3}$	0.452***		0.189***					
	(0.054)		(0.058)					
fair$_{t-3}$	0.806***		0.354***					
	(0.056)		(0.063)					
poor$_{t-3}$	1.244***		0.519***					
	(0.065)		(0.076)					
frailty$_{t-3}$		5.134***	4.147***	1.407	5.277***	3.430***	2.568***	0.410
		(0.335)	(0.362)	(2.345)	(1.218)	(0.737)	(0.609)	(0.800)
frailty$^2_{t-3}$		-3.521***	-2.958***	7.314	-3.164	-1.713	-0.779	0.564
		(0.549)	(0.574)	(4.840)	(2.552)	(1.323)	(0.986)	(1.103)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	38,125	38,125	38,125	3,843	12,891	12,125	6,244	1,868
Pseudo R^2	0.115	0.144	0.151	0.217	0.130	0.084	0.059	0.037

Table 57: Results of probit regressions of transition to SSDI beneficiary status between wave t-1 and t, on three periods lagged values of frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-3$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	-0.213		-0.389					
	(0.524)		(0.550)					
good$_{t-1}$	-0.899*		-1.305**					
	(0.493)		(0.538)					
fair$_{t-1}$	0.204		-0.716					
	(0.476)		(0.547)					
poor$_{t-1}$	1.423***		-0.444					
	(0.523)		(0.625)					
excellent$_{t-1}\times$ age	-0.052**		-0.043*					
	(0.023)		(0.025)					
very good$_{t-1}\times$ age	-0.047**		-0.037					
	(0.022)		(0.024)					
good$_{t-1}\times$ age	-0.029		-0.020					
	(0.022)		(0.025)					
fair$_{t-1}\times$ age	-0.038*		-0.025					
	(0.022)		(0.025)					
poor$_{t-1}\times$ age	-0.049**		-0.025					
	(0.022)		(0.025)					
frailty$_{t-1}$		5.880**	6.651**	5.502	-3.081	-4.958	6.571	11.609
		(2.846)	(3.171)	(20.965)	(8.903)	(5.718)	(5.407)	(7.515)
frailty$^2_{t-1}$		1.404	-0.318	22.591	16.046	24.431**	1.095	-9.039
		(4.368)	(4.649)	(50.680)	(16.197)	(10.107)	(8.104)	(9.463)
frailty$_{t-1}\times$ age		0.023	-0.011	0.031	0.156	0.176*	-0.028	-0.132
		(0.050)	(0.056)	(0.356)	(0.156)	(0.100)	(0.094)	(0.131)
frailty$^2_{t-1}\times$ age		-0.109	-0.069	-0.449	-0.316	-0.455***	-0.066	0.113
		(0.076)	(0.081)	(0.846)	(0.283)	(0.177)	(0.141)	(0.165)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	76,513	76,513	76,513	8,975	25,334	23,925	12,378	3,846
Pseudo R^2	0.179	0.239	0.253	0.284	0.207	0.170	0.129	0.076

Table 58: Results of probit regressions of transition to SSDI beneficiary status between wave t-1 and t, on frailty, SRHS and age interactions using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

C.3.7 Predicting retirement

Tables 59 and 60 report additional results for probit regressions for retirement. In each probit estimation we regress a variable that is equal to one if an individual is retired between wave t-1 and wave t on lagged SRHS categories and/or lagged frailty polynomial. All regression also include the following controls: gender, education, marital status and a quadratic in age.

Although, retirement is not entirely a health related event, health status may be a factor in deciding when to retire. The results below show that both SRHS and frailty are both informative about retirement in near future.

In Panel A of Table 59 reports results of entire sample. Column (1) shows result of regression in lagged SRHS categories. Column (2) shows results of regression on lagged frailty polynomial. Column (3) report the results when both are included. To evaluate and compare the performance of each health measure we report the *pseudo R-squared* for each regression.

Columns (1) and (2) indicate that variations in both frailty and SRHS are informative about retirement. However, the pseudo R-squared in column (2) is slightly higher than in column (1), indicating that frailty is slightly more informative about mortality than SRHS. However, not in a meaningful way.

Column (3) shows that both health measures remain statistically significant when included together indicating that they both have independent predictive power. However, notice the large decline in the coefficients on the SRHS dummies moving from column (1) to (3), as well as, the magnitude of the changes in the pseudo R-squared's across columns. Both indicate that the SRHS's additional impact is relatively small.

In panel B of Table 59 we repeat the same regression as column (2) of panel A. However, instead of running the regression on the entire sample, we run five different regressions, one for each SRHS category. In each these regressions, SRHS has no explanatory/predictive power. However, in each category, the coefficients on the frailty index are highly significant. Notice that frailty is highly informative about mortality even among individuals who report 'excellent' health.

Finally, in Table 60 we repeat the same exercise by including age interactions. Although these age interactions are significant, they including them does not alter the results.

	Panel A: everyone			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.115***		0.073***					
	(0.020)		(0.021)					
good$_{t-1}$	0.172***		0.074***					
	(0.021)		(0.023)					
fair$_{t-1}$	0.354***		0.179***					
	(0.026)		(0.029)					
poor$_{t-1}$	0.597***		0.315***					
	(0.050)		(0.054)					
frailty$_{t-1}$		1.657***	1.383***	2.069**	0.608	1.335***	1.526***	0.858
		(0.220)	(0.232)	(0.945)	(0.517)	(0.461)	(0.567)	(1.033)
frailty$^2_{t-1}$		-0.119	-0.189	-1.520	1.418	-0.044	-1.253	-0.259
		(0.503)	(0.520)	(3.483)	(1.547)	(1.097)	(1.090)	(1.687)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	51,841	51,841	51,841	8,240	18,818	16,481	6,961	1,325
Pseudo R^2	0.120	0.123	0.124	0.113	0.117	0.137	0.120	0.080

Table 59: Results of probit regressions of transition to retirement between wave t-1 and t, on frailty and SRHS using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: everyone			Panel B: by SRHS health at $t-1$				
				'Excellent	'Very good'	'Good'	'Fair'	'Poor'
	(1)	(2)	(3)	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	-0.346*		-0.508***					
	(0.190)		(0.194)					
good$_{t-1}$	-0.551***		-0.966***					
	(0.196)		(0.209)					
fair$_{t-1}$	-0.156		-0.898***					
	(0.238)		(0.265)					
poor$_{t-1}$	1.214**		-0.027					
	(0.475)		(0.519)					
excellent$_{t-1}\times$ age	-0.005		0.003					
	(0.015)		(0.015)					
very good$_{t-1}\times$ age	0.003		0.013					
	(0.015)		(0.015)					
good$_{t-1}\times$ age	0.007		0.020					
	(0.015)		(0.015)					
fair$_{t-1}\times$ age	0.004		0.021					
	(0.015)		(0.016)					
poor$_{t-1}\times$ age	-0.015		0.008					
	(0.016)		(0.017)					
frailty$_{t-1}$		3.221	6.216***	18.811**	6.344	2.022	0.807	-1.253
		(2.097)	(2.241)	(9.322)	(4.787)	(4.524)	(5.296)	(9.099)
frailty$^2_{t-1}$		3.313	-1.495	-49.087	-5.354	1.132	9.548	3.496
		(4.797)	(5.038)	(33.818)	(13.962)	(10.846)	(10.327)	(15.377)
frailty$_{t-1}\times$ age		-0.026	-0.081**	-0.284*	-0.097	-0.012	0.010	0.036
		(0.035)	(0.038)	(0.157)	(0.080)	(0.076)	(0.089)	(0.156)
frailty$^2_{t-1}\times$ age		-0.058	0.021	0.811	0.118	-0.020	-0.178	-0.065
		(0.081)	(0.085)	(0.571)	(0.232)	(0.182)	(0.173)	(0.265)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	51,841	51,841	51,841	8,240	18,818	16,481	6,961	1,325
Pseudo R^2	0.121	0.124	0.125	0.114	0.117	0.137	0.121	0.080

Table 60: Results of probit regressions of transition to retirement between wave t-1 and t, on frailty, SRHS and age interactions using data from HRS. Panel A are results from the full sample while panel B are results obtained using sub-samples based on SRHS in wave $t-1$. Controls are gender, education, marital status and a quadratic in age. Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

C.4 Comparison to other variants of frailty in HRS

Table 61 shows summary statistics for the frailty along with its three variants in HRS sample. In column (1) we repeat the summary stats for frailty. Columns (2) through (4) show stats for frailty with SRHS, the FPC index, and the SIO index, respectively. Figure 13 shows boxplots of these indices by age and histograms showing their cross-sectional distributions.

Tables 62 , 63 , and 63 show two-year transition probabilities across quintiles of frailty-with-SRHS, FPC, and SIO, respectively for two age groups of 50–74 year-olds, and ages 75 and above in the HRS sample.

Finally, we report results of probit regressions of transition to retirement between wave t-1 and t on SRHS, frailty, frailty with SRHS, the FPC index, and the SIO index at $t-1$ using data from HRS in Table 65. Similar to results reported in Table 59 we see no significant different across indices in terms of predicting retirement.

	Frailty (1)	Frailty w/SRHS (2)	FPC (3)	SIO (4)
	frailty	Frailty w/SRHS	FPC	SIO
Mean	0.22	0.22	0.20	0.20
men	0.20	0.20	0.17	0.17
women	0.23	0.23	0.22	0.22
ages 50-74	0.19	0.19	0.16	0.17
ages 75+	0.29	0.29	0.28	0.30
less than HS	0.28	0.28	0.27	0.26
HS grad	0.21	0.21	0.19	0.20
college grad	0.16	0.16	0.13	0.15
Standard Dev	0.16	0.16	0.19	0.20
men	0.15	0.15	0.17	0.17
women	0.17	0.17	0.20	0.21
ages 50-74	0.14	0.14	0.17	0.18
ages 75+	0.18	0.18	0.22	0.22
less than HS	0.18	0.19	0.22	0.23
HS grad	0.15	0.15	0.18	0.19
college grad	0.13	0.13	0.15	0.16
Min	0.00	0.00	0.00	0.00
5th percentile	0.04	0.04	0.01	0.01
50th percentile	0.17	0.17	0.13	0.13
95th percentile	0.56	0.56	0.61	0.61
Max	0.97	0.97	0.99	1.00
Fraction at zero	0.00	0.00	0.00	0.00
Observations	226402	226402	226402	128574
Individuals	40288	40288	40288	29299

Table 61: Summary statistics for four versions of the frailty index: frailty, frailty with SRHS, FPC index, and SIO index using HRS data on respondents and spouses over the age 50.

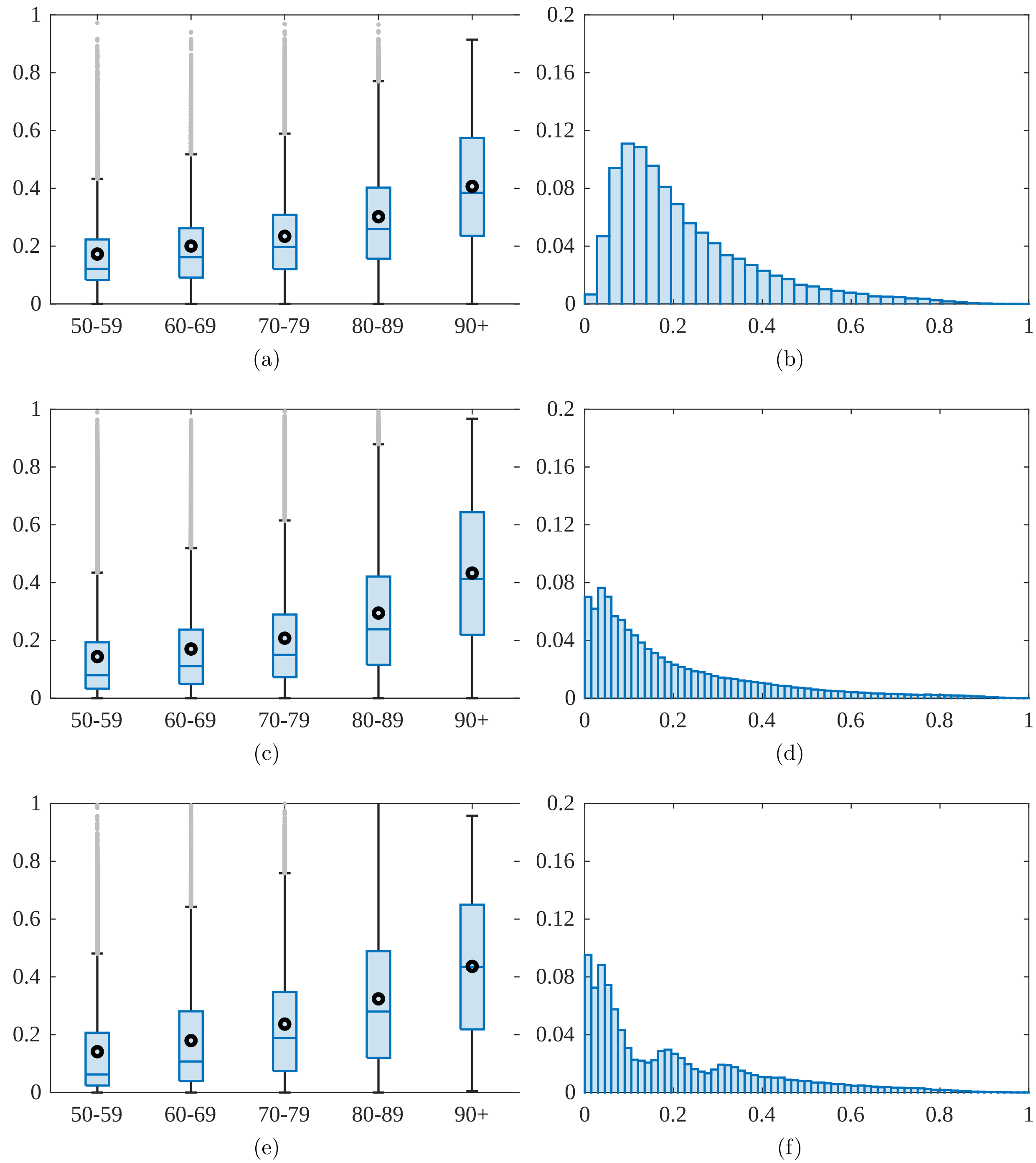


Figure 13: Box plots of frailty with SRHS (a), FPC index (c), and SIO index (e) by 10-year age groups in the HRS sample. The bottom and top edges of the boxes indicate the 25th and 75th percentiles, respectively. The middle lines are medians and the open circles are means. The upper whiskers extend to 1.5 times the inner quartile range. The lower whiskers extend to the minimum value of frailty in the age group. The dots are data points that lie beyond the whiskers. Histograms showing the cross-sectional distribution of frailty with SRHS (b), FPC index (d), and SIO index (f) in the sample.

	Ages 50 to 74				
	Bottom	2nd	3rd	4th	Top
Bottom	64.6	24.9	7.3	2.4	0.7
2nd	20.0	44.0	25.5	8.7	1.8
3rd	4.1	21.3	41.8	27.0	5.7
4th	0.7	4.8	20.5	50.3	23.7
Top	0.1	0.4	2.3	16.8	80.4
	Ages 75 and older				
	Bottom	2nd	3rd	4th	Top
Bottom	57.9	25.8	10.0	4.0	2.4
2nd	12.6	39.3	30.1	12.4	5.5
3rd	2.4	17.0	38.0	31.0	11.6
4th	0.5	3.6	15.7	46.4	33.8
Top	0.0	0.4	1.9	13.7	83.9

Table 62: Two-year transition probabilities (%) across quintiles of frailty-with-SRHS for two age groups: 50–74 year-olds, and ages 75 and older constructed using the HRS sample. Rows sum to one.

	Ages 50 to 74				
	Bottom	2nd	3rd	4th	Top
Bottom	62.5	23.6	9.7	3.3	0.9
2nd	18.6	42.9	26.1	10.0	2.3
3rd	6.0	22.3	39.2	26.7	5.8
4th	1.4	6.3	21.9	47.0	23.4
Top	0.2	0.8	3.0	18.1	77.9
	Ages 75 and older				
	Bottom	2nd	3rd	4th	Top
Bottom	54.8	27.3	10.9	4.5	2.5
2nd	15.6	37.8	29.1	12.0	5.4
3rd	3.4	17.2	36.5	31.3	11.6
4th	0.8	4.0	16.8	44.9	33.5
Top	0.1	0.6	2.5	14.1	82.7

Table 63: Two-year transition probabilities (%) across quintiles of FPC for two age groups: 50–74 year-olds, and ages 75 and older constructed using the HRS sample. Rows sum to one.

	Ages 50 to 74				
	Bottom	2nd	3rd	4th	Top
Bottom	51.2	24.0	12.2	9.7	3.0
2nd	22.9	37.2	21.1	13.6	5.3
3rd	10.9	18.5	38.3	21.3	11.0
4th	8.4	12.5	19.6	36.2	23.4
Top	2.4	4.9	10.1	25.7	57.0
	Ages 75 and older				
	Bottom	2nd	3rd	4th	Top
Bottom	55.0	23.8	13.8	5.8	1.5
2nd	16.3	39.6	24.4	14.6	5.0
3rd	10.2	20.6	32.8	26.5	9.9
4th	3.3	10.2	21.8	39.2	25.5
Top	0.8	3.7	7.6	21.5	66.4

Table 64: Two-year transition probabilities (%) across quintiles of SIO for two age groups: 50–74 year-olds, and ages 75 and older constructed using the HRS sample. Rows sum to one.

	(1)	(2)	(3)	(4)	(5)
very good$_{t-1}$	0.115***				
	(0.020)				
good$_{t-1}$	0.172***				
	(0.021)				
fair$_{t-1}$	0.354***				
	(0.026)				
poor$_{t-1}$	0.597***				
	(0.050)				
frailty$_{t-1}$		1.657***			
		(0.220)			
frailty$^2_{t-1}$		-0.119			
		(0.503)			
frailty w/ SRHS$_{t-1}$			1.662***		
			(0.215)		
frailty w/ SRHS$^2_{t-1}$			-0.122		
			(0.483)		
FPC$_{t-1}$				1.570***	
				(0.167)	
FPC$^2_{t-1}$				-0.567	
				(0.379)	
SIO$_{t-1}$					0.603***
					(0.130)
SIO$^2_{t-1}$					-0.207
					(0.229)
Controls	YES	YES	YES	YES	YES
Observations	51,841	51,841	51,841	51,841	45,057
Pseudo R^2	0.120	0.123	0.124	0.122	0.123

Table 65: Probit regressions of transition to retirement between wave t-1 and t on SRHS, frailty, frailty with SRHS, the FPC index, and the SIO index at $t-1$ using data from HRS. Controls are gender, education, marital status and a quadratic in age. Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

C.5 Comparing frailty index across HRS, PSID, and MEPS

In this section we report our stylized fact on frailty in HRS, and sub-samples of individuals older than 50 in PSID and MEPS. The purpose is to make the samples across datasets more comparable.

Table 66 shows the summary statistics for frailty across all datasets for individuals over the age of 50. In terms of averages, frailty in MEPS and PSID samples are very close. Both in overall samples, and with age group, gender, and education groups.However, MEPS sample has slightly higher dispersion. Standard deviations in MEPS are much higher than the ones in PSID (and HRS).

Average frailty in the HRS sample is higher than that of both MEPS and PSID. This is true overall, and within age, gender and education groups. Also, the fraction of zero frailty in HRS is very small (less than 1 percent). This is possibility due to larger number of deficit we use in HRS. Particularly, deficits variables on doctor visit, or hospital visits are set to one if the respondent had visited a doctor or hospital during the interview period. These variables (along with larger number if ADLs/IADLs variables) may be responsible for the higher averages in HRS. However, the larger averages may also be due to sample composition. HRS, unlike MEPS and PSID, includes individuals living in nursing homes (it also over samples individuals over the age of 85). However, within each sample, the age gradient across 25 year age groups are similar. Also, differences in frailty by education are larger than differences by gender. In all three samples, frailty is more dispersed for older age groups, for women, and for lower educated.

Figure 14 shows the area plots that are analogous to Figure 2 in the paper and Figures 7 and 11 in this appendix. All figures are constructed by choosing frailty cut offs to match the distribution of frailty categories to the distribution of SRHS at age 50-55 in each sample. These plots show that the rate of decline of frailty is very similar in the three datasets.

Figure 15 shows the same information as the area plots in Figure 14 for SRHS but reorganized so that the SRHS distributions in each dataset are all in the same graph. Figure 16 shows frailty CDF's for four 5-year age groups in each of the three datasets. These two figures are provided to allow for comparison of age-group specific SRHS distributions across the three datasets and frailty distributions across the three datasets. Notice that the distributions of both SRHS and frailty are very similar in HRS and PSID. Even though, mean frailty is slightly higher in HRS than in PSID the cumulative distributions of frailty have nearly identical shapes. Even though the differences are relatively small, the distributions of SRHS and frailty in MEPS however are less similar to those in HRS and PSID. First, the share of individuals with excellent health is higher and does not decline with age in MEPS as it does in the other samples. Second, in the MEPS sample, the distribution of frailty is more skewed at all ages. Notice that the CDF in MEPS is more concave as compared to PSID and HRS. The differences between both SRHS and frailty in MEPS versus HRS and PSID indicate that the health distribution in the MEPS sample is slightly different than in the PSID and HRS samples. In particular, the MEPS sample is more skewed. That is, it has a larger fraction of individuals who are either relatively healthier or unhealthier as compared to the PSID and HRS samples.

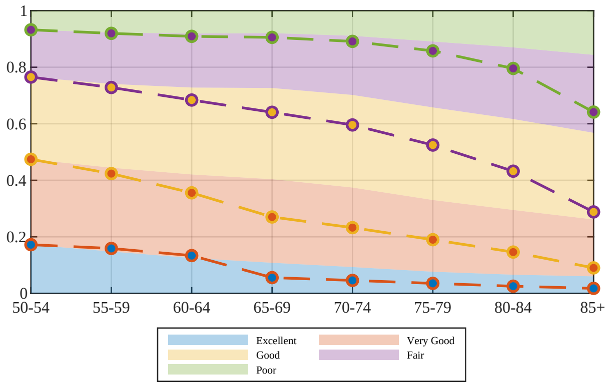
Figure 17 shows box plots for the distribution of frailty by SRHS categories and 10-year age groups. The box plots show very similar patterns for those under the age of 90.

However, not so much for older age groups. This is to be expected. MEPS and PSID sample of individuals older than 90 is very small and excludes those in nursing homes. For this reason, box plots for this samples are noisy and and look different than the ones in HRS.

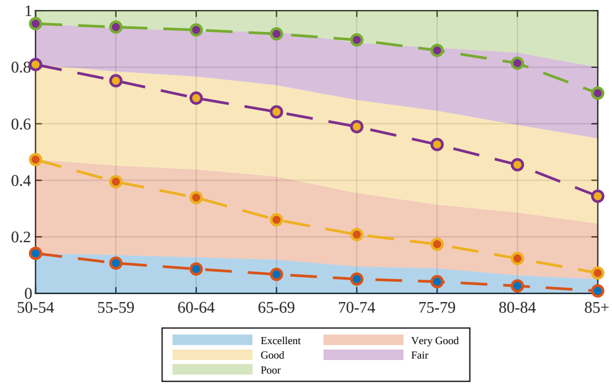
Finally, Tables 67, 68, 69, show two year transition probabilities across frailty quintiles for two age groups: 50–74 year-olds, and ages 75 and older. Judging by values on the diagonals, frailty in the PSID and MEPS sample is more persistent than in HRS. This is true in both age group and across all frailty quintiles. However, there is not clear pattern when comparing MEPS and HRS. For example, frailty in HRS seems to be more persistent than MEPS in the top quintile, but less persistent at the bottom.

	HRS (1)	PSID (50 and older) (2)	MEPS (50 and older) (3)
Mean	0.22	0.16	0.17
men	0.20	0.15	0.15
women	0.23	0.17	0.19
ages 50-74	0.19	0.15	0.15
ages 75+	0.29	0.25	0.28
less than HS	0.28	0.22	0.23
HS grad	0.21	0.16	0.17
college grad	0.16	0.12	0.12
Standard Dev	0.16	0.13	0.18
men	0.15	0.12	0.17
women	0.17	0.14	0.19
ages 50-74	0.14	0.12	0.17
ages 75+	0.18	0.16	0.21
less than HS	0.18	0.16	0.21
HS grad	0.15	0.13	0.18
college grad	0.13	0.11	0.14
Min	0.00	0.00	0.00
5th percentile	0.04	0.00	0.00
50th percentile	0.17	0.11	0.11
95th percentile	0.56	0.43	0.58
Max	0.97	0.89	1.00
Fraction at zero	0.00	0.06	0.09
Observations	226402	38355	160597
Individuals	40288	8218	89956

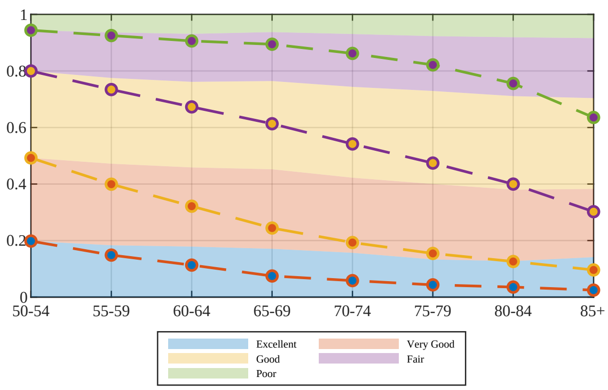
Table 66: Column (1): summary stats for frailty index in HRS sample. Column (2): summary stats for frailty index among those older than 50 in PSID. Column (3): summary stats for frailty index among those older than 50 in MEPS.



(a) HRS



(b) PSID over 50



(c) MEPS over 50

	excellent	very good	good	fair	poor
% of SRHS in age 50 - 55, HRS	17.22	30.21	29.1	16.7	6.8
Frailty range, HRS	[0, 0.059)	[0.059, 0.117)	[0.117, 0.212)	[0.212, 0.406)	[0.406, 1]
% of SRHS in age 50 - 55, PSID	14.1	33.2	33.6	14.5	4.6
Frailty range, PSID	[0, 0.036)	[0.036 0.071)	[0.071, 0.179)	[0.179, 0.357)	[0.357, 1]
% of SRHS in age 50 - 55, MEPS	19.8	29.4	30.7	14.4	5.6
Frailty range, MEPS	[0, 0.009)	[0.009, 0.076)	[0.076, 0.160)	[0.160, 0.459)	[0.459, 1]

Figure 14: Distribution of health status for respondents over 50 years old in HRS, PSID and MEPS. The colored areas show the fraction of individuals by SRHS at each age. The dashed lines show the fraction of individuals by the corresponding frailty category at each age.

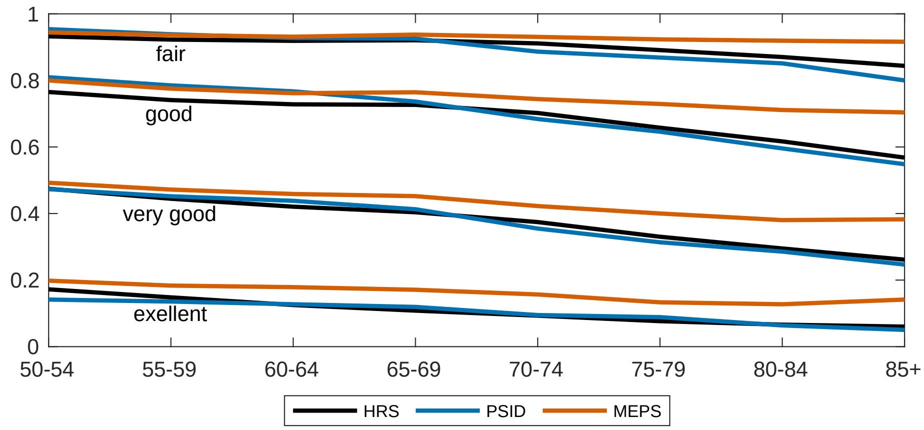


Figure 15: Distribution of SRHS by 5-year age groups for respondents over 50 years old in HRS, PSID and MEPS.

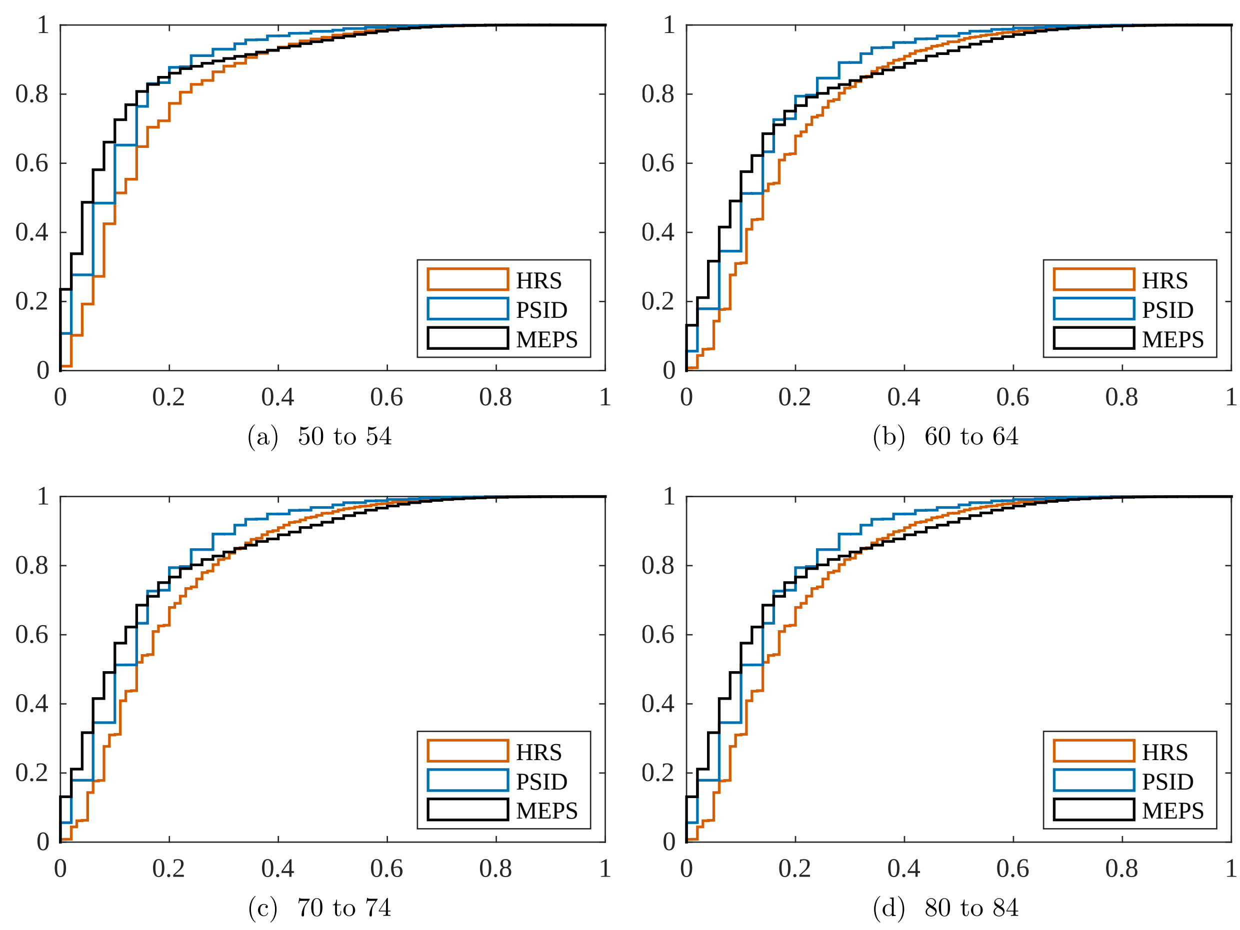
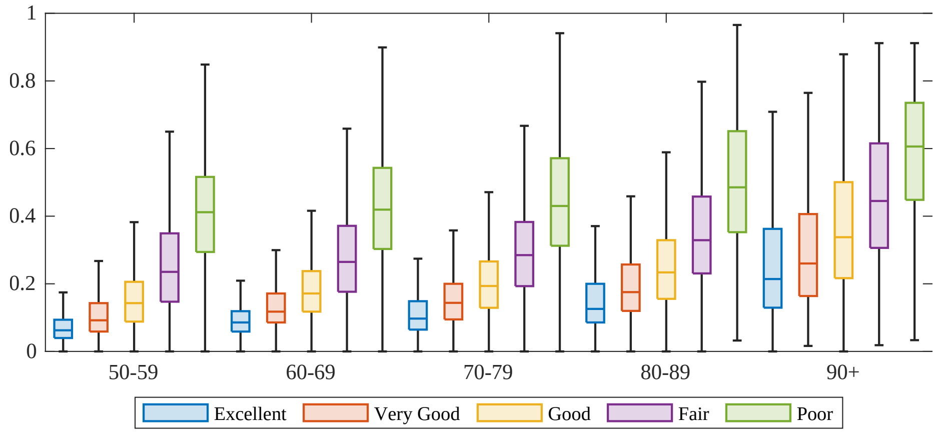
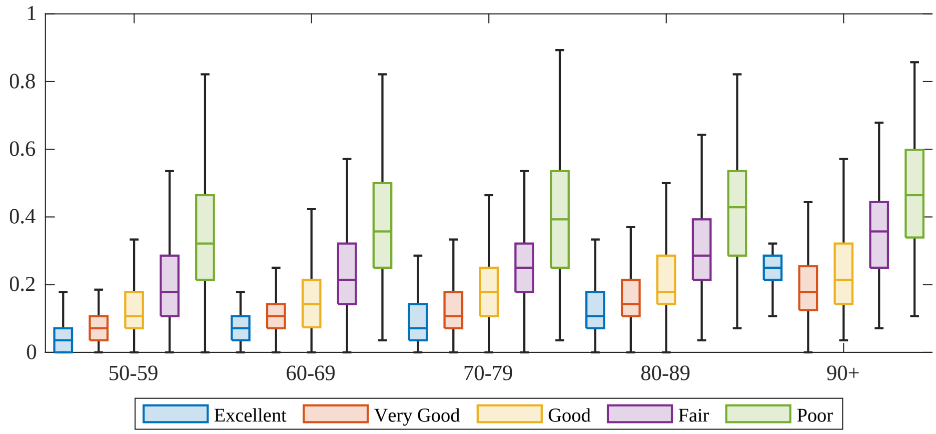


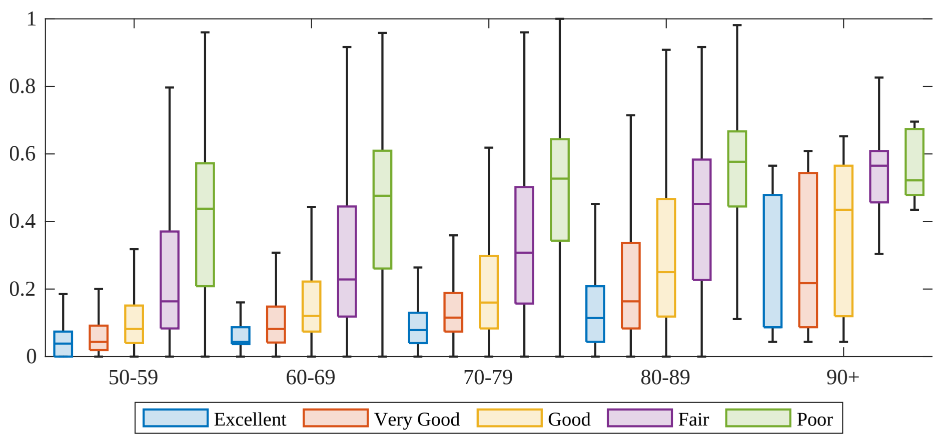
Figure 16: CDF of frailty at select 5 year age groups for HRS, PSID, and MEPS.



(a) HRS



(b) PSID over 50



(c) MEPS over 50

Figure 17: Box plots of frailty for respondents over 50 years of age in HRS, PSID, and MEPS. Bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points that are not outliers. The middle line indicates median.

	Ages 50 to 74				
	Bottom	2nd	3rd	4th	Top
Bottom	64.8	24.4	7.3	2.6	0.8
2nd	18.6	45.2	25.7	8.6	1.9
3rd	4.1	21.6	41.1	27.3	5.9
4th	0.8	5.1	20.8	49.2	24.0
Top	0.1	0.5	2.6	17.1	79.8
	Ages 75 and older				
	Bottom	2nd	3rd	4th	Top
Bottom	58.5	25.1	9.9	3.9	2.6
2nd	13.3	38.3	29.8	12.8	5.8
3rd	2.6	17.3	37.9	30.7	11.5
4th	0.4	3.6	16.3	45.7	34.0
Top	0.0	0.5	2.2	13.3	84.0

Table 67: Two-year transition probabilities (%) across frailty quintiles for two age groups: 50–74 year-olds, and ages 75 and older constructed using the HRS sample. Rows sum to one.

	Ages 50 to 74				
	Bottom	2nd	3rd	4th	Top
Bottom	79.2	16.1	3.4	0.8	0.4
2nd	5.4	67.8	20.7	5.0	1.1
3rd	0.2	7.4	67.1	22.1	3.2
4th	0.0	0.8	7.9	68.0	23.3
Top	0.1	0.2	0.8	10.0	89.0
	Ages 75 and older				
	Bottom	2nd	3rd	4th	Top
Bottom	70.8	18.9	5.8	3.0	1.5
2nd	5.4	57.1	24.3	8.9	4.3
3rd	1.1	8.7	46.5	30.1	13.5
4th	0.3	2.1	15.9	50.9	30.9
Top	0.0	0.2	1.9	15.0	82.9

Table 68: Two-year transition probabilities (%) across frailty quintiles for two age groups: 50–74 year-olds, and ages 75 and older constructed using the PSID sub sample of respondents over 50. Rows sum to one.

	Ages 50 to 74				
	Bottom	2nd	3rd	4th	Top
Bottom	74.9	17.8	4.6	1.5	1.3
2nd	13.6	57.5	20.8	5.2	3.0
3rd	1.0	10.2	60.5	21.9	6.4
4th	0.7	2.3	13.1	61.8	22.1
Top	0.6	1.5	4.6	17.1	76.2
	Ages 75 and older				
	Bottom	2nd	3rd	4th	Top
Bottom	66.2	17.2	7.8	6.8	2.0
2nd	8.6	49.3	23.4	13.9	4.8
3rd	4.1	11.1	47.2	24.2	13.4
4th	2.6	6.7	16.2	43.9	30.6
Top	0.2	0.8	5.5	14.1	79.4

Table 69: Two-year transition probabilities (%) across frailty quintiles for two age groups: 50–74 year-olds, and ages 75 and older constructed using the MEPS sub sample of respondents over 50. Rows sum to one.

C.5.1 Comparing wave-to-wave changes

HRS over samples population older than 85. It also includes nursing home population which are excluded in MEPS and PSID. In order to have a more comparable sub-sample, we further restrict our attention to individuals between 50 and 75. Table 70, reports summary stats for changes in frailty from one wave to another among this smaller sub-sample. Column (1) shows the numbers for the HRS sample. Column (2) and (3) show the numbers for PSID and MEPS.

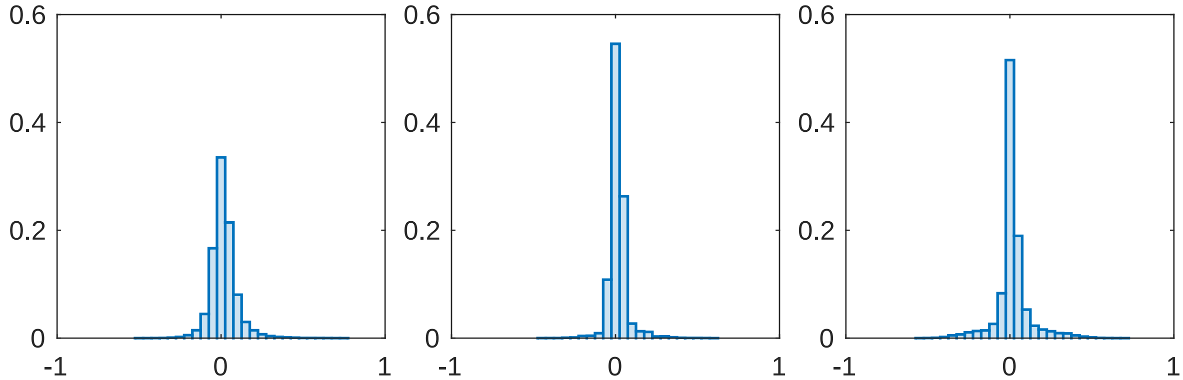
We first note that changes in HRS and MEPS are more frequent. 56 percent of the HRS sample in this age group show positive wave-to-wave changes and 43 percent show negative changes. In PSID 33 percent of the sample in this age group has positive wave to wave changes and only 14 percent have negative changes. For MEPS, the numbers are 45 percent and 34 percent, respectively for positive and negative changes.

We plot histograms of changes for this group in Figure 18. The difference is not as stark. In fact histograms for PSID and MEPS look very similar. For HRS the histograms demonstrate less concentration at zero, but a larger mass around zero. This is due to composition and number of deficit variables. For example "cognitive limitation score" is not a 0/1 variable. It can have many smaller values (same is true for "K6 depression score" in MEPS). This allows for possibility of small changes. While in PSID all wave-to-wave changes are due to adding or subtracting a 0/1 deficit variable. As a results HRS and MEPS have a higher fraction of positive/negative changes, relative to PSID. Notice in Table 70 that the mean positive changes in HRS in the age group 50 to 75 are smaller, relative to PSID and MEPS. Similarly, the mean negative changes are also smaller. Notice also that median negative changes are smallest in MEPS.

Overall, the distribution of changes in HRS, MEPS, and PSID differ for (at least) two reasons. One is the composition of the sample. We tried to control for that by looking at a smaller more comparable sub-sample. The other is the composition of the deficit variables. HRS (and also MEPS) have variables that move more frequently and have a finer grid. This allows for more frequent but smaller changes.

	HRS (1)	PSID (2)	MEPS (3)
Fraction with positive changes	0.56	0.33	0.45
mean positive change	0.06	0.07	0.07
median positive change	0.04	0.04	0.04
Fraction with negative changes	0.43	0.14	0.34
mean negative change	-0.04	-0.06	-0.06
median negative change	-0.03	-0.04	-0.02
Observations	226,402	38,355	160,597
Individuals	40,288	8,218	89,956

Table 70: Summary stats on wave-to-wave changes in frailty among 50 to 74 year olds. Column (1): wave-to-wave changes in frailty in the HRS sample. Column (2): wave-to-wave changes in frailty in the PSID sample. Column (3): wave-to-wave changes in frailty in the MEPS sample.



(a) HRS (50 to 74) (b) PSID (50 to 74) (c) MEPS (50 to 74)

Figure 18: Histogram of wave-to-wave changes in frailty for age group 50 to 75 in HRS, PSID, and MEPS.

D Supplemental Material For Estimation of Frailty Process

D.1 Computation of Cohort-adjusted Empirical Moments

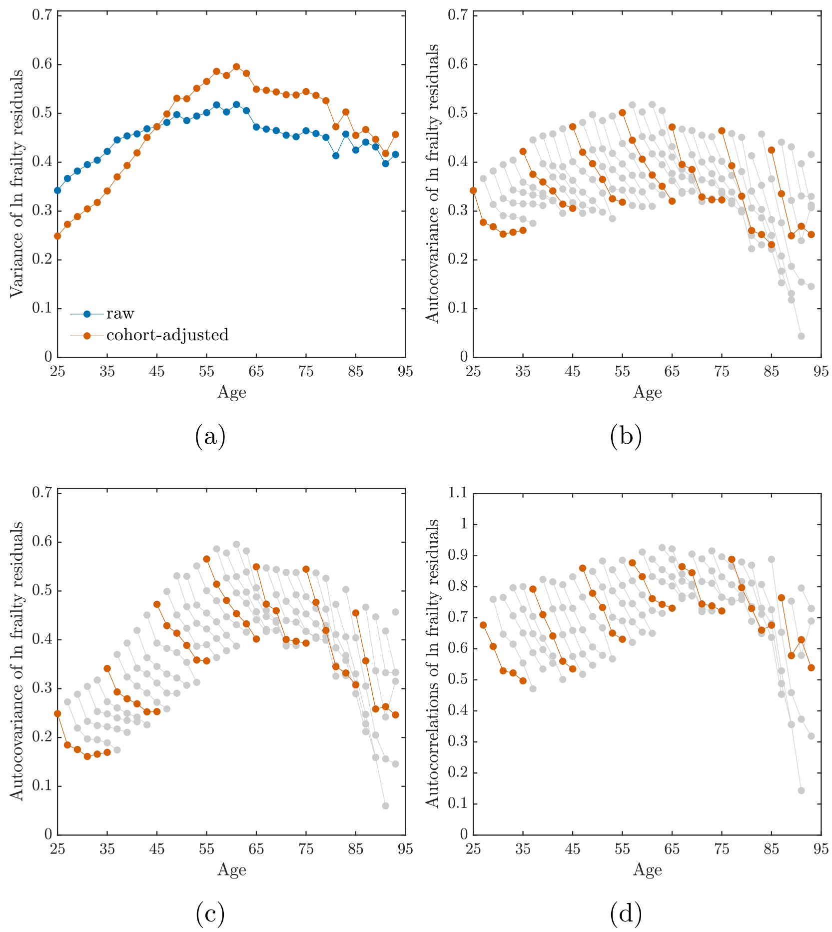


Figure 19: Raw and cohort-adjusted variances (a), raw variance-covariance matrix (b), cohort-adjusted variance-covariance matrix (c), and autocorrelations (d) of the log frailty residuals for the main PSID sample. Log frailty residuals are derived from an OLS regression of log frailty on a quartic in age, education dummies, and gender dummies.

This section describes how we obtain the cohort-adjusted variance and covariance moments that we target in the SMM estimations of the frailty process described in Section 4 of the paper. The cohort-adjusted moments for the main PSID sample are shown in panel (c) of Figure 19. The corresponding raw moments are in panel (b). Panel (a) shows the cohort-adjusted variances together with the raw variances for ease of comparison. Panel (d) shows the cohort-adjusted autocorrelations. Counterparts to Figure 19 panels (a)-(c) by gender and education are provided in Figures 20 and 21, respectively.

We construct the cohort-adjusted moments as follows. First, cohorts are defined. Cohort 1 consists of all individuals born before 1911 (this is 35 individuals with birth years from

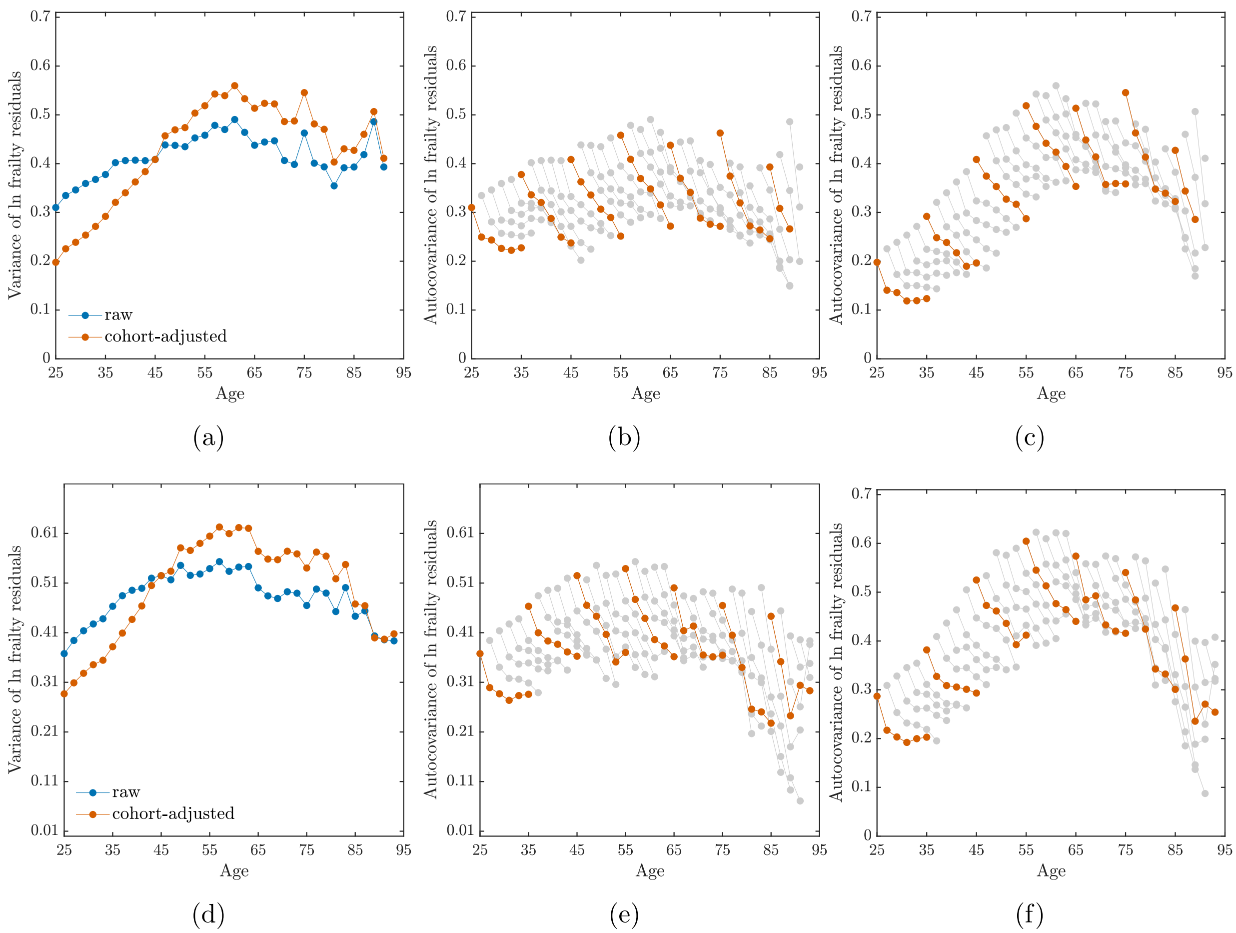


Figure 20: Raw and cohort-adjusted variances of log frailty residuals for men (a) and women (d), raw variance-covariance matrix for men (b) and women (e), and cohort-adjusted variance-covariance matrix for men (c) and women (f). Log frailty residuals are derived for each gender from an OLS regression of log frailty on a quartic in age and education dummies. Data is PSID.

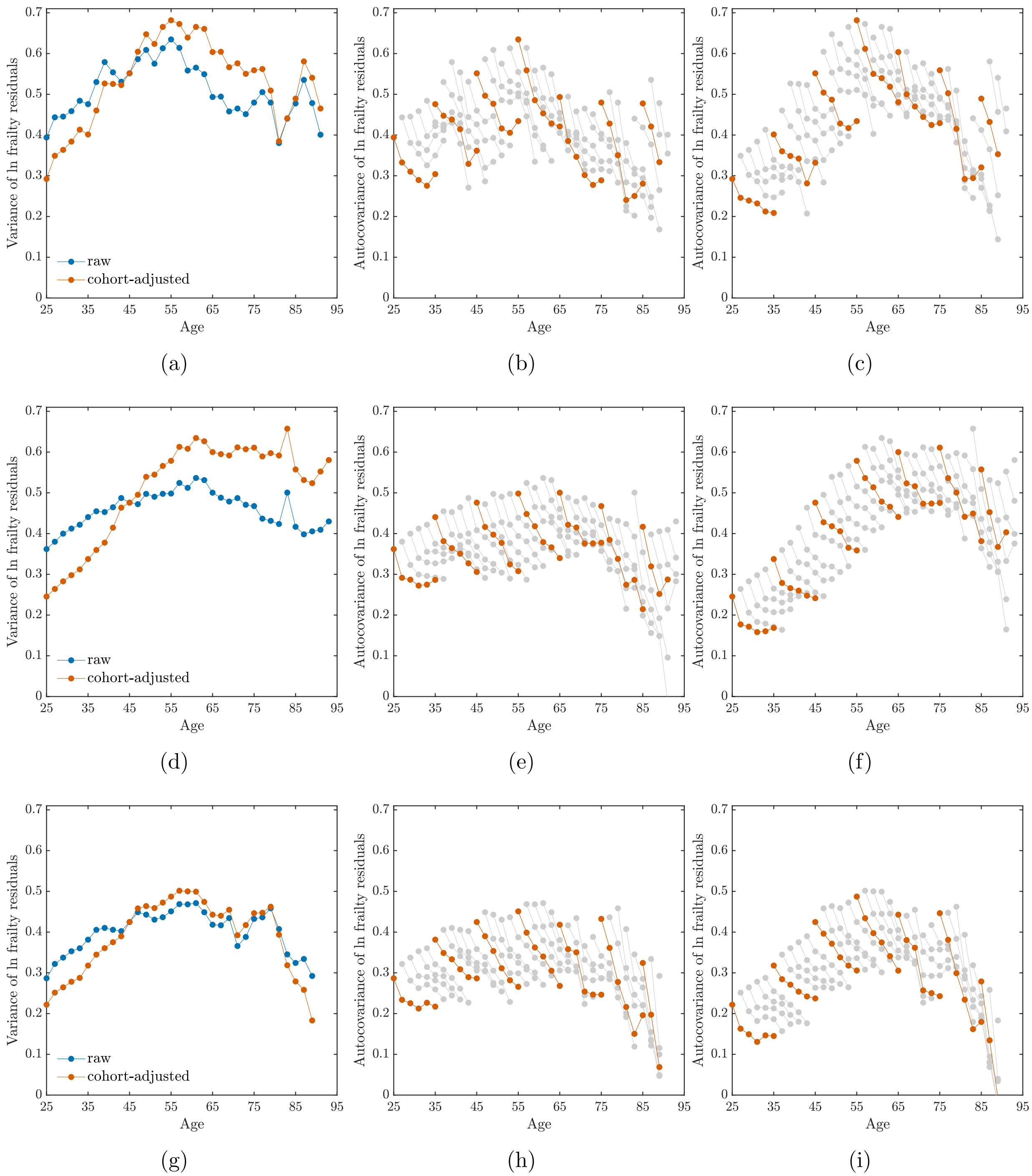


Figure 21: Raw and cohort-adjusted variances of log frailty residuals for less than HS (a) HS graduates (d) and college graduates (g), raw variance-covariance matrix for less than HS (b) HS graduates (e) and college graduates (h), and cohort-adjusted variance-covariance matrix for less than HS (c) HS graduates (f) and college graduates (i). Log frailty residuals are derived for each education group from an OLS regression of log frailty on a quartic in age and gender dummies. Data is PSID.

1905–1910). Cohorts then increment with every two birth years, so that the second cohort includes birth years 1911 and 1912, and the third 1913 and 1914. Denote the number of cohorts by n_c. Next, age groups are defined. Each age group spans 2 years and age groups are non-overlapping. Denote an individual's age group by age_f. The initial OLS regression is then run on the sample of individuals meeting the desired age restrictions for the particular covariance matrix being computed. The means of the squared residuals R_{it}^2 for each age group are the raw variance profile. Denote the raw variance of the 35–36 year old age group by R_{35}^2.

To obtain the cohort-adjusted variances and covariances, we first define the individual-specific moments

$$m_i^{t,t+k} = R_{it} R_{it+k},$$

for $k \in \{0, 1, 2, \ldots, 5\}$. We regress these moments on age and cohort dummies as follows:

$$m_i^{t,t+k} = \eta + \sum_{a=0}^{35} \sum_{j=0}^{5} \delta_j^t I[k = j] I[a = 25 + 2t] + \sum_{c=1}^{n_c} \theta_c I[C_i = c] + \varepsilon_{ik}^t.$$

Note that the omitted age effect dummy is δ_0^{35}. Individual-specific cohort-adjusted moments are then given by

$$\tilde{m}_i^{t,t+k} = \hat{\delta}_k^t + \varepsilon_{ik}^t + R_{35}^2 - \tilde{m}_0^{35},$$

where $\tilde{m}_{35}$ is the mean variance moment for individuals in the 35–36 year-old age group. The last two terms scale the moments so that the raw and cohort-adjusted variances are the same for this age group. The cohort-adjusted variance/covariance profiles are the means of these individual-specific cohort-adjusted moments for each age group:

$$\tilde{m}_k^t = \sum_i \tilde{m}_i^{t,t+k}.$$

D.2 Mortality Probit

Table 71 reports the results from the estimation of the mortality probit model, equation (7) in the paper. The model is estimated using 1998–2016 HRS data. The regression also includes year fixed effects. Frailty, age, and male gender all increase the probability of dying. We do not find a statistically significant effect of education on mortality.

	Coefficient estimates
frailty	3.001***
	(0.098)
frailty^2	-0.500***
	(0.120)
age	-0.021***
	(0.005)
age^2	3.91e-4***
	(3.45e-5)
education (years)	0.003
	(0.002)
male	0.308***
	(0.010)
constant	-2.919***
	(0.183)
year fixed effects	included
Observations	213,114
Pseudo R^2	0.239

Table 71: Results from mortality probit regression using the baseline frailty index and 1998–2016 HRS data. Standard errors are in parenthesis. Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

D.3 Additional Estimation Results

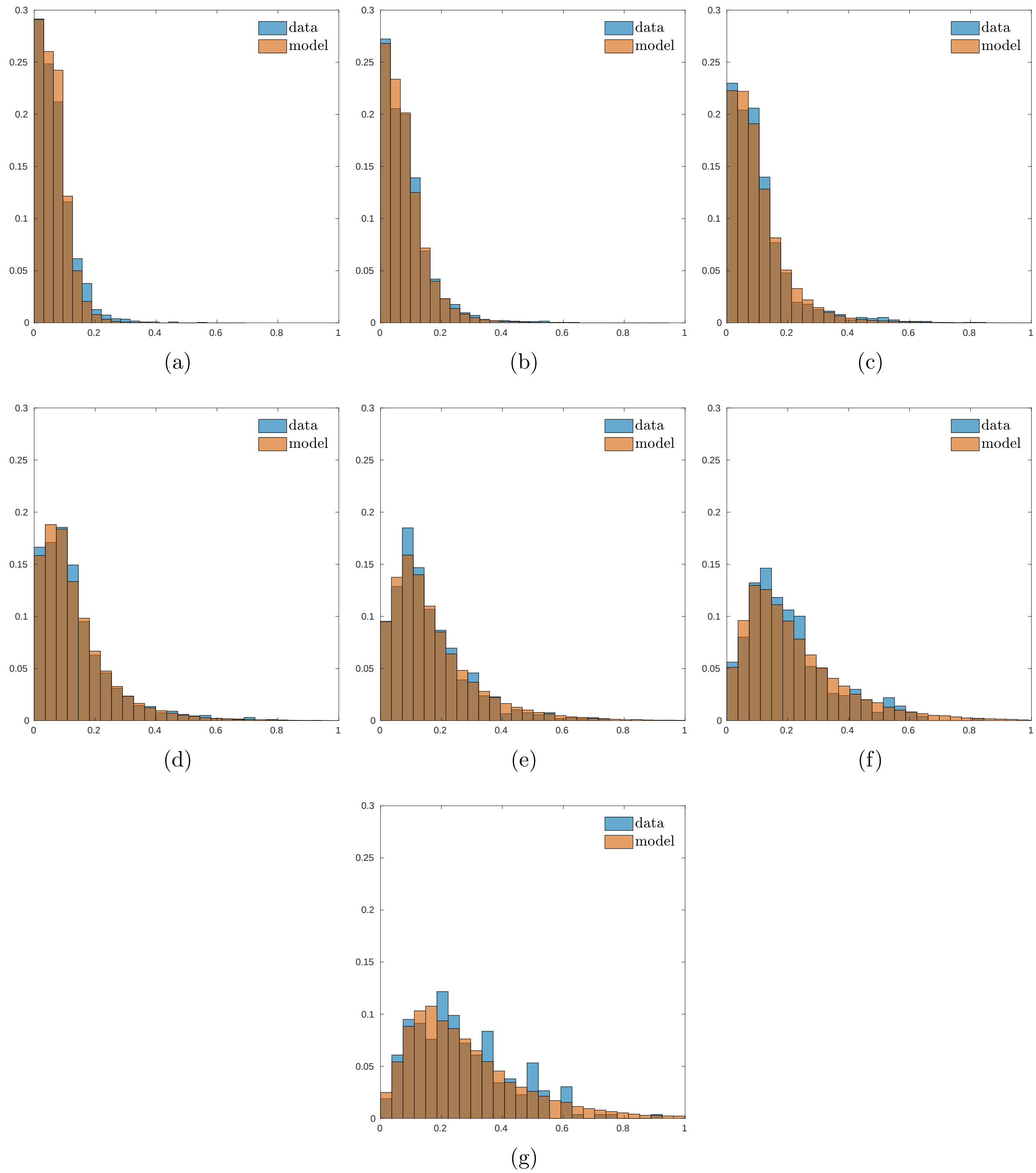


Figure 22: Histograms showing the cross-sectional distributions of frailty by age for 25–26 year-olds (a), 35–36 year-olds (b), 45–46 year-olds (c), 55–56 year-olds (d), 65–66 year-olds (e), 75–76 year-olds (f), and 85–86 year-olds (g) in the model using the baseline estimates and the data.

Figure 22 shows the histograms of frailty generated using the simulated data from the baseline estimation of the model and their data counterparts for 25 to 26 year-olds, 35 to 36 year-olds, 45 to 46 year-olds, 55 to 56 year-olds, 65 to 66 year-olds, 75 to 76 year-olds, and 85 to 86 year-olds. Notice that the distributions in the data match well those in the model in each 2-year age group.

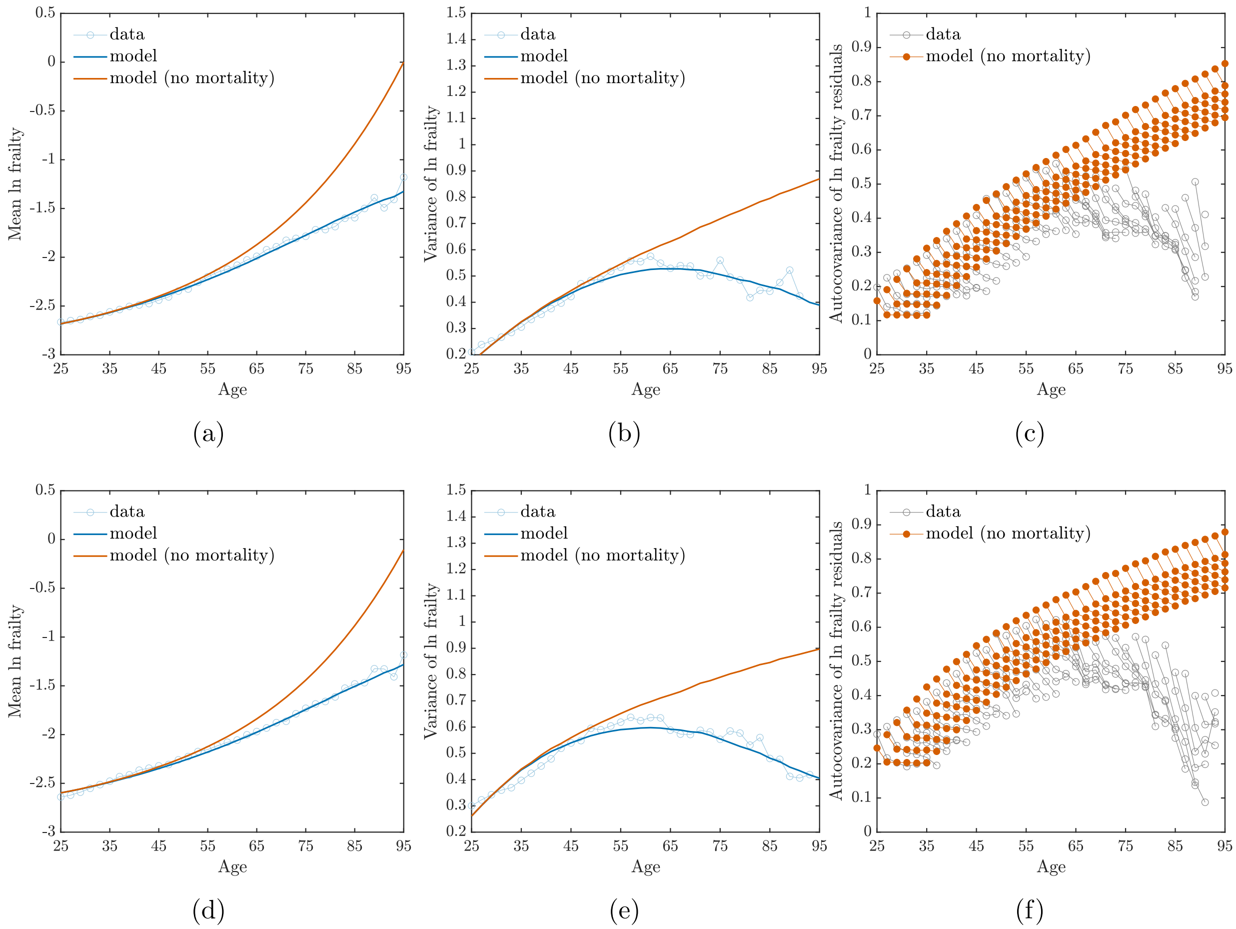


Figure 23: Moments illustrating the effect of mortality on estimations by gender. Mean log frailty by age in the data (open blue circles) and the model with mortality (solid blue line) and without mortality (solid orange line) for men (a) and women (d). Variance of log frailty by age in the data (open blue circles) and the model with mortality (solid blue line) and without mortality (solid orange line) for men (b) and women (e). Autocovariance profiles of log frailty residuals by age in the data (open gray circles) and the model without mortality (closed orange circles) for men (c) and women (f).

Tables 23 and 24 report the effects of shutting down mortality on the mean and variance of log frailty and the variance-covariance matrix of log frailty residuals for each gender separately and each education group separately. The figures are constructed using the results of estimating the frailty process separately for each gender and then for each education group presented in the paper.

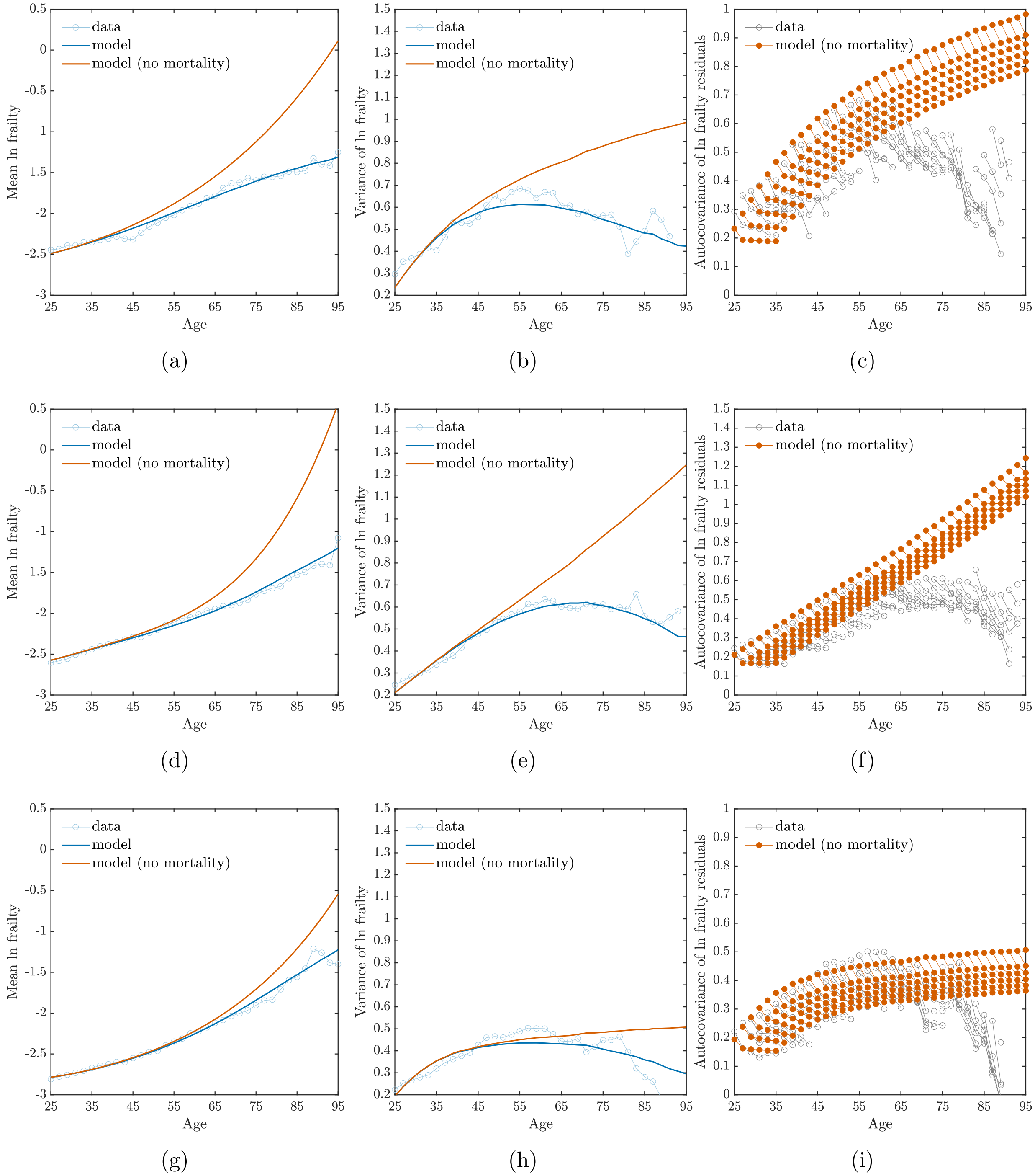


Figure 24: Moments illustrating the effect of mortality on estimations by education. Mean log frailty by age in the data (open blue circles) and the model with mortality (solid blue line) and without mortality (solid orange line) for less than HS (a), HS graduates (d), and college graduates (g). Variance of log frailty by age in the data (open blue circles) and the model with mortality (solid blue line) and without mortality (solid orange line) for less than HS (b), HS graduates (e), and college graduates (h). Autocovariance profiles of log frailty residuals by age in the data (open gray circles) and the model without mortality (closed orange circles) for less than HS (c), HS graduates (f), and college graduates (i).

E Frailty process estimation using variants of the frailty index

E.1 Frailty with SRHS included as a deficit variable

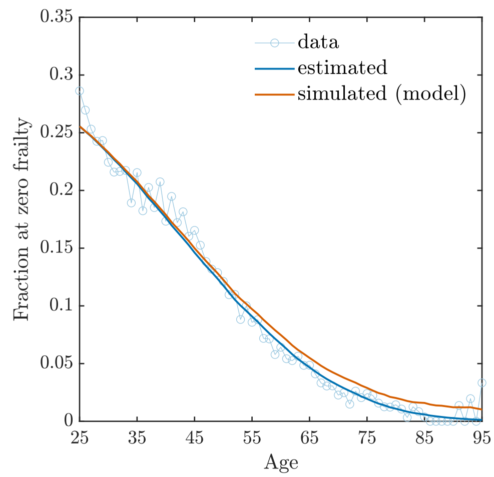


Figure 25: Fraction with zero frailty by age in the data (blue open circles), predicted from the probit regression (solid blue line), and predicted, after accounting for mortality, by the model (solid orange line) using the version of the frailty index that includes SRHS as a 0/1 deficit variable.

Variable	Age	Age^2	High school	College	Male	Constant
Value	6.37e-4	-2.97e-4***	0.069***	0.344***	-0.058***	-0.554***
Std. Error	(2.75e-3)	(2.97e-5)	(0.017)	(0.017)	(0.011)	(0.062)
Pseudo R^2 =0.0836, No. of observations=96,252						

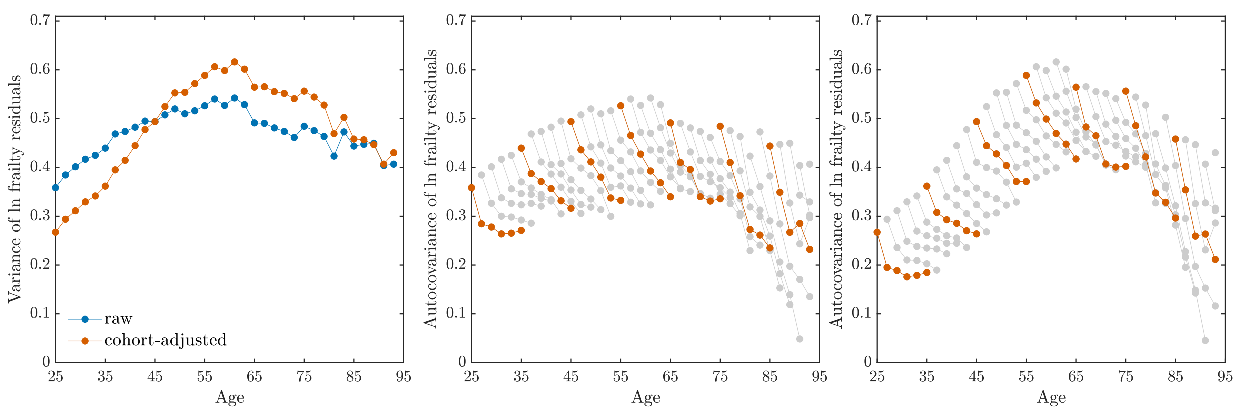
Table 72: Estimation results from the zero frailty probit regression using the version of the frailty index that includes SRHS as a 0/1 deficit variable. Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

In this section we repeat the estimation of the frailty process presented in Section 4.1 of the paper with the frailty index replaced by the frailty index that includes SRHS as a 0/1 deficit variable. As we did with frailty, we adjust the variance-covariance moments for cohort effects. The left panel of Figure 26 shows both the raw and cohort-adjusted variances. The right panel of the figure shows the entire variance-covariance matrix adjusted for cohort effects.

Table 72 shows the results of the zero frailty probit regression using frailty with SRHS as the health measure. Figure 25 shows the fraction at zero in the data, predicted by the frailty probit, and in the simulated model. Table 73 shows the results of the probit estimation of mortality in the HRS data using frailty with SRHS as the health measure. Figure 27 shows

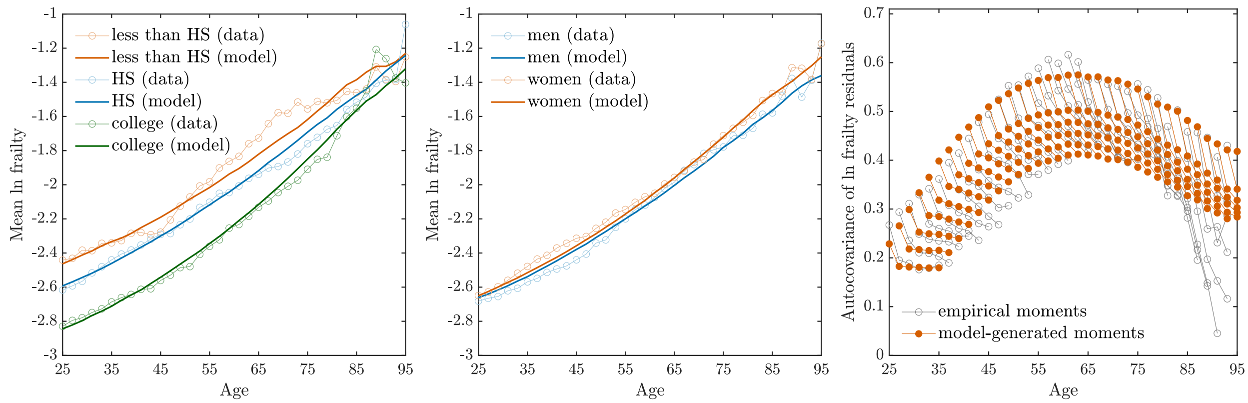
	Coefficient estimates
frailty	3.038***
	(0.098)
frailty^2	-0.530***
	(0.119)
age	-0.021***
	(0.005)
age^2	3.99e-4***
	(3.46e-5)
education (years)	0.005***
	(0.001)
male	0.308***
	(0.010)
constant	-2.940***
	(0.183)
year fixed effects	included
Observations	212,978
Pseudo R^2	0.243

Table 73: Results from mortality probit regression using the version of the frailty index that includes SRHS as a 0/1 deficit variable and 1998–2016 HRS data. Standard errors are in parenthesis. Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.



(a) (b) (c)

Figure 26: Raw and cohort-adjusted variances (a), raw variance-covariance matrix (b), and cohort-adjusted variance-covariance matrix (c) of the log frailty residuals for the main PSID sample using the version of the frailty index that includes SRHS as a 0/1 deficit variable. Log frailty residuals are derived from an OLS regression of log frailty on a quartic in age, education dummies, and gender dummies.



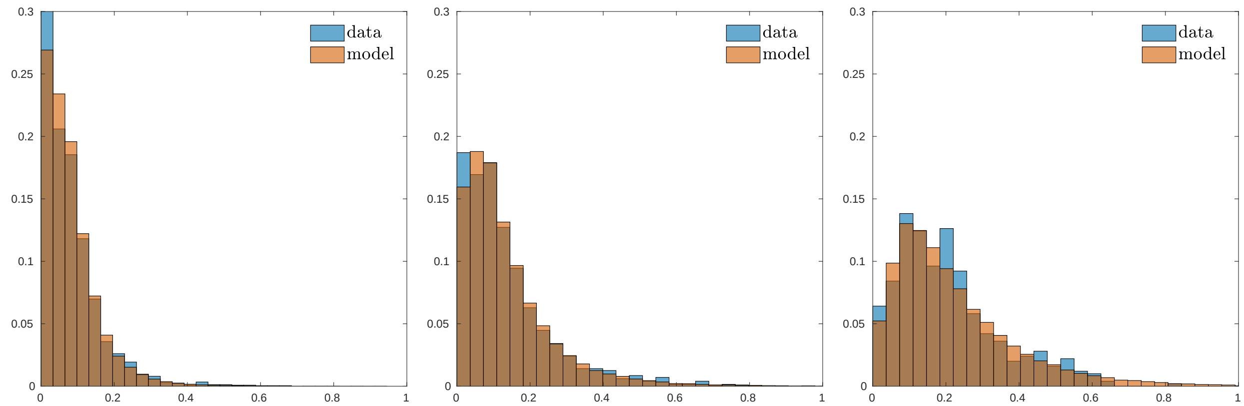
(a) (b) (c)

Figure 27: Fit of the estimated model using the version of the frailty index that includes SRHS as a 0/1 deficit variable: empirical moments targeted in the baseline SMM estimation and their model counterparts. Panel (a) is mean log frailty by age in the data (open circles) and the model (solid lines) for those without a HS degree (orange), HS graduates (blue), and college graduates (green). Panel (b) is mean log frailty by age in the data (open circles) and the model (solid lines) for women (orange) and men (blue). Panel (c) is autocovariance profiles of log frailty residuals by age in the data (open gray circles) and the model (closed orange circles).

Deterministic component			**Stochastic component**		
Variable	Value	Std. Error	Variable	Value	Std. Error
age*	1.183	(0.008)	σ_u^2	0.044	(0.002)
age*^{2}	1.866	(0.022)	σ_ε^2	0.020	(0.001)
age*^{3}	-1.370	(0.040)	σ_α^2	0.154	(0.004)
age*^{4}	5.694	(0.038)	ρ	0.989	(0.001)
HS degree	-0.133	(0.001)			
college degree	-0.374	(0.001)			
male	-0.006	(0.001)			
constant	-2.468	(0.001)			

Table 74: Results from estimating the log frailty process via SMM using the version of the frailty index that includes SRHS as a 0/1 deficit variable and a sample of individuals aged 25–95 from PSID. Note: age* $\equiv$ (age $-$ 25)/100.

the empirical moments targeted in the SMM estimation and the model counterparts. Table 74 presents the SMM estimation results. Comparing across tables and figures indicates that the target moments and estimation results are very similar to the baseline estimation results which are based on a frailty index that only includes objective health measures. One notable difference is that when SRHS is added as a deficit variable, the average effect on frailty of being male is even smaller than in the baseline. In addition, the effect of being more educated is slightly larger than in the baseline. Finally, Figure 28 shows histograms of frailty with



(a) (b) (c)

Figure 28: Histograms showing the cross-sectional distributions of frailty for 35–36 year-olds (a), 55–56 year-olds (b), and 75–76 year-olds (c) in the model and the data using the version of the frailty index that includes SRHS as a 0/1 deficit variable.

SRHS in the model and data for the 35-36 year-olds, 55-56 year-olds, and 75-76 year-olds.

E.2 Frailty using first principal component weighting of deficits

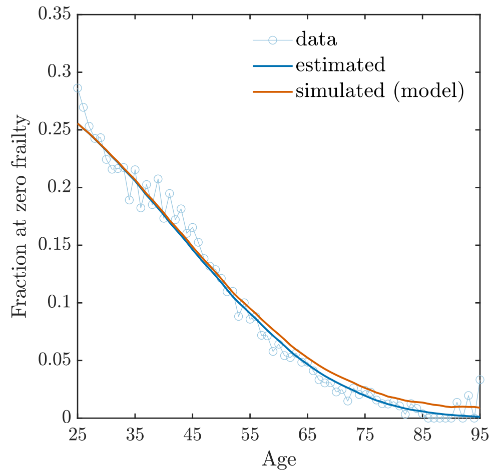


Figure 29: Fraction with zero frailty by age in the data (blue open circles), predicted from the probit regression (solid blue line), and predicted, after accounting for mortality, by the model (solid orange line) using the first principal component (FPC) frailty index.

	Coefficient estimates
frailty	2.380***
	(0.074)
frailty^2	-0.333***
	(0.088)
age	-0.012***
	(0.005)
age^2	3.34e-4***
	(3.46e-5)
education (years)	0.004***
	(0.001)
male	0.331***
	(0.010)
constant	-3.086***
	(0.183)
year fixed effects	included
Observations	213,114
Pseudo R^2	0.236

Table 75: Results from mortality probit regression using the first principal component (FPC) frailty index and 1998–2016 HRS data. Standard errors are in parenthesis. Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

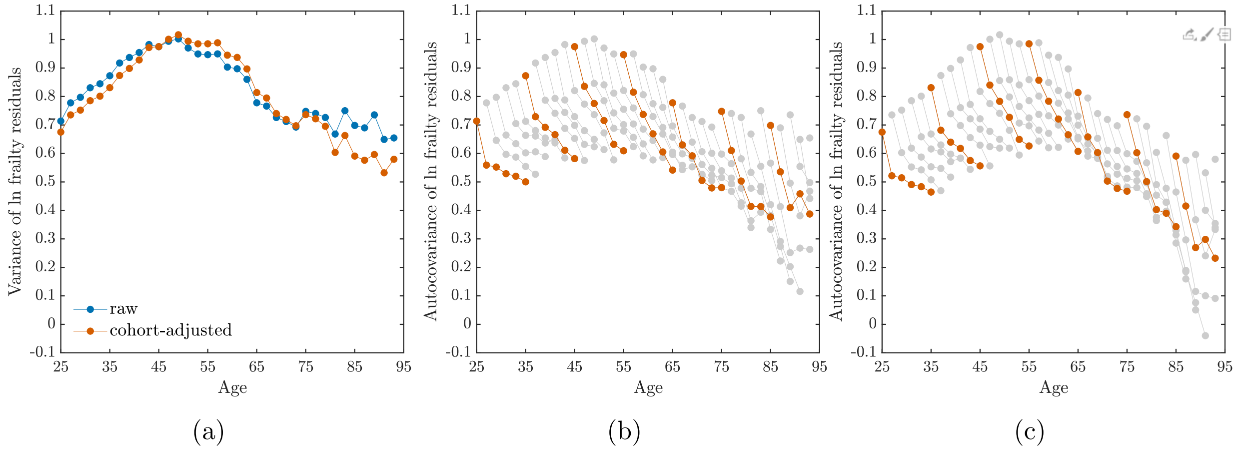


Figure 30: Raw and cohort-adjusted variances (a), raw variance-covariance matrix (b), and cohort-adjusted variance-covariance matrix (c) of the log frailty residuals for the main PSID sample using the first principal component (FPC) frailty index. Log frailty residuals are derived from an OLS regression of log frailty on a quartic in age, education dummies, and gender dummies.

Deterministic component			**Stochastic component**		
Variable	Value	Std. Error	Variable	Value	Std. Error
age*	2.196	(0.009)	σ_u^2	0.082	(0.004)
age*^{2}	2.234	(0.030)	σ_ε^2	0.053	(0.002)
age*^{3}	-1.487	(0.045)	σ_α^2	0.450	(0.007)
age*^{4}	5.407	(0.036)	ρ	0.950	(0.002)
HS degree	-0.217	(0.001)			
college degree	-0.441	(0.001)			
male	-0.137	(0.001)			
constant	-3.194	(0.001)			

Table 76: Results from estimating the log frailty process via SMM using the first principal component (FPC) frailty index and a sample of individuals aged 25–95 from PSID. Note: $\text{age}^* \equiv (\text{age} - 25)/100$.

In this section we show that the results of our main estimation are robust to using the FPC index as our health measure. Recall, the FPC index is the first principal component of the frailty index that includes SRHS as a 0/1 deficit variable. We do not repeat the results of the zero frailty probit regression because they are identical to those in Table 72. This is because zeros in the frailty with SRHS index coincide with zeros in the FPC index by construction. Figure 29 shows the fraction at zero in the data, predicted by the frailty probit, and in the simulated model. Table 75 shows the results of the probit estimation of mortality in the HRS data using the FPC index as the health measure.

As we did with frailty and frailty with SRHS, we adjust the variance-autocovariance moments for cohort effects. The left panel of Figure 30 shows both the raw and cohort-

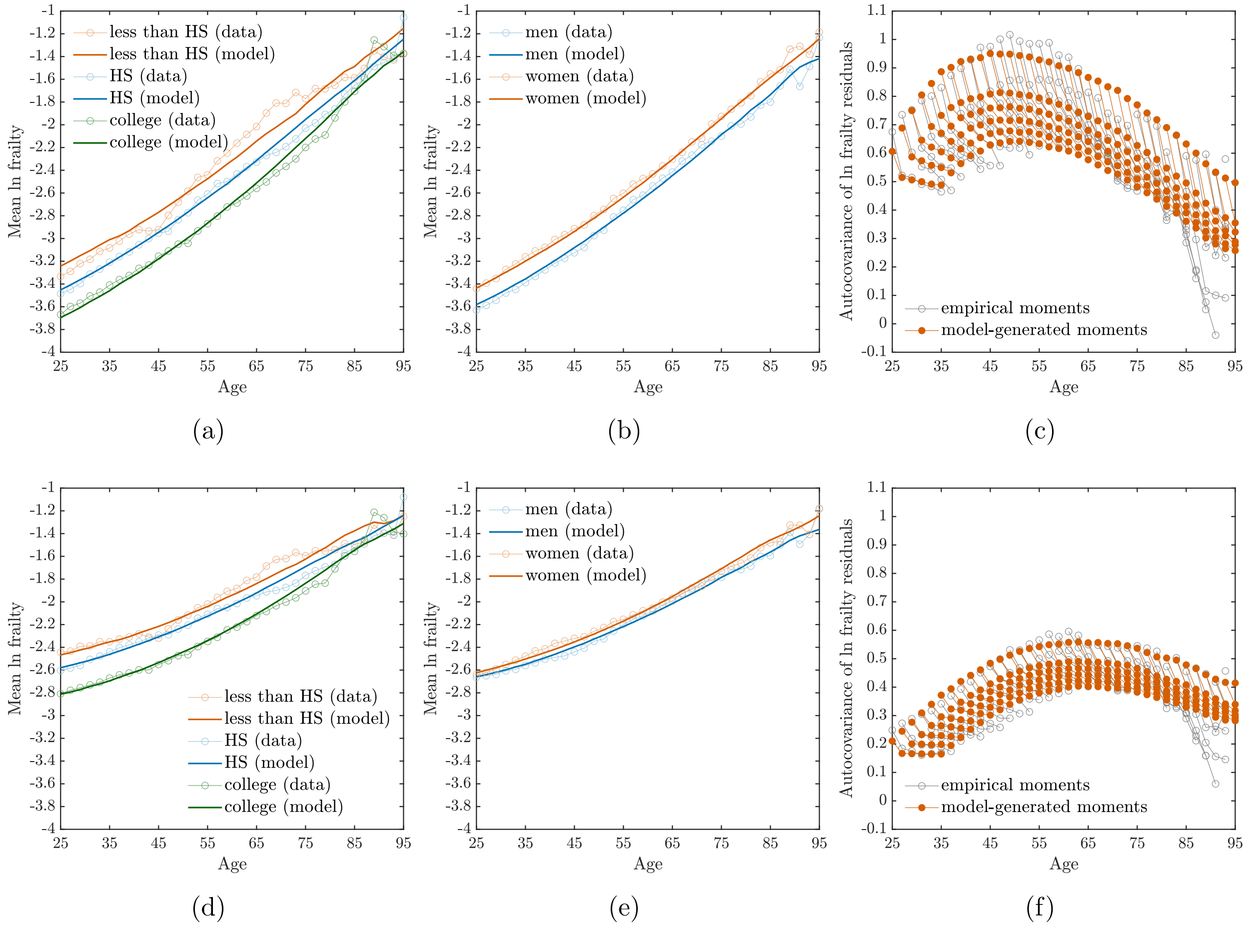


Figure 31: Fit of the estimated model using the first principal component (FPC) frailty index and baseline frailty (provided for comparison): empirical moments targeted in the baseline SMM estimation and their model counterparts. Panel (a) is mean log frailty by age in the data (open circles) and the model (solid lines) for those without a HS degree (orange), HS graduates (blue), and college graduates (green). Panel (b) is mean log frailty by age in the data (open circles) and the model (solid lines) for women (orange) and men (blue). Panel (c) is autocovariance profiles of log frailty residuals by age in the data (open gray circles) and the model (closed orange circles). Panels (d), (e), and (f) are the counterparts for the baseline frailty estimation results provided in the paper. Note that the y-axes have been changed here to ease comparability across the sets of moments for the two health measures.

adjusted variances. The right panel of the figure shows the entire variance-covariance matrix adjusted for cohort effects. Comparing Figure 30 to Figure 19 reveals that properties of the variance profile and autocovariance profiles for the principal component index are very similar to those for the frailty index. The variances of the log frailty residuals increase at earlier ages and then decline and the auto-covariances decay as the lag order increases. The biggest difference between the two set of moments is that the autocovariance profiles decline more rapidly with age for the FPC index as compared to frailty and the variance profile is more concave than the one for frailty. Notice also that the difference between the variance

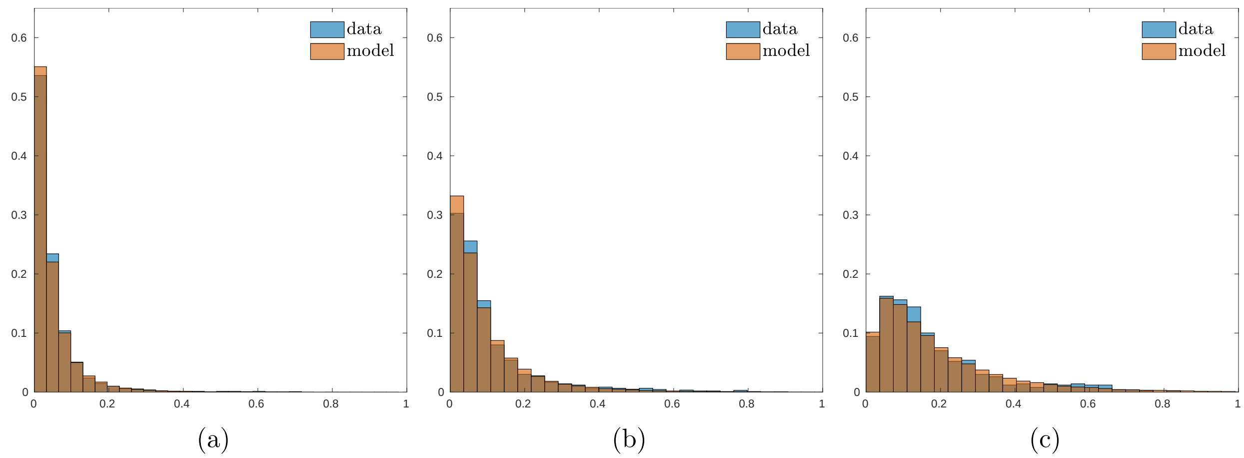


Figure 32: Histograms showing the cross-sectional distributions of frailty for 35–36 year-olds (a), 55–56 year-olds (b), and 75–76 year-olds (c) in the model and the data using the first principal component (FPC) frailty index.

and the first autocovariance at each age is larger for the principal component index than for the frailty index.

Figure 31 and Figure 32 show the empirical moments targeted in the SMM estimation and the model counterparts. Panels (a), (b), and (c) are the model and data moments using the FPC index. Panels (d), (e), and (f) are the baseline frailty moments (same as in the paper) but repeated here with the y-axes on the same scale as the FPC graphs to allow for easy comparison. Table 76 presents the SMM estimation results. Comparing across figures and tables indicates that the results are qualitatively similar to the baseline estimation results which are based on a frailty index that equally weights objective measures of health. Quantitatively, there are several differences. First, the FPC index is more variable and increases more rapidly with age than frailty. For this reason, the age profile of FPC is steeper and the overall variance of the stochastic component is larger. The estimation attributes a large share of the higher variance to ex ante heterogeneity. Since the FPC index has a more concave variance profile, the estimated value for ρ is lower (relative to the one estimated using the frailty index). Notice also that the education and gender effects are larger for the FPC index than for frailty. Finally, Figure 32 shows histograms of the FPC index in the model and data for 35-36 year-olds, 55-56 year-olds, and 75-76 year-olds.

E.3 Frailty using subjective measures to inform the weights

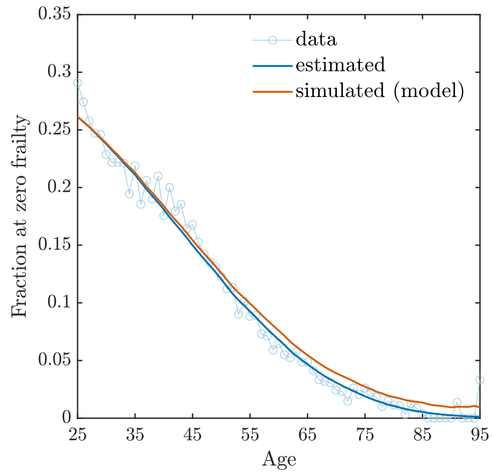


Figure 33: Fraction with zero frailty by age in the data (blue open circles), predicted from the probit regression (solid blue line), and predicted, after accounting for mortality, by the model (solid orange line) using the subjective instrumented by objective (SIO) frailty index.

Variable	Age	Age^2	High school	College	Male	Constant
Value	1.60e-3	-3.11e-4***	0.104***	0.299***	-0.063***	-0.516***
Std. Error	(2.75e-3)	(2.98e-5)	(0.017)	(0.017)	(0.010)	(0.062)
Pseudo R^2 =0.083, No. of observations=95,914						

Table 77: Estimation results from the zero frailty probit regression using the subjective instrumented by objective (SIO) frailty index. Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

In this section we show that the results of our main estimation are robust to using the SIO index as our health measure. Recall, the SIO index is a version of the frailty index where the weights are chosen to maximize the index's power in predicting subjective health measures. Table 77 shows the results of the zero frailty probit estimation. In theory, zeros in the SIO index coincide with zeros in the frailty with SRHS index and the FPC index. However, because the samples are not identical (due to missing observations on the subjective health measures) we repeated the estimation of the zero frailty probit using the SIO index sample. The results are very similar to those in Table 72. Figure 33 shows the fraction at zero in the data, predicted by the SIO frailty index probit, and in the simulated model. Table 78 shows the results of the probit estimation of mortality in the HRS data using the SIO index as the health measure.

As we did with the other variants of frailty, we adjust the variance-autocovariance moments for cohort effects. The left panel of Figure 34 shows both the raw and cohort-adjusted variances. The right panel of the figure shows the entire variance-covariance matrix adjusted for cohort effects. Comparing Figure 34 to Figure 19 and Figure 30 reveals that properties

	Coefficient estimates
frailty	0.356***
	(0.110)
frailty^2	1.189***
	(0.137)
age	-0.041***
	(0.008)
age^2	5.58e-4***
	(5.22e-5)
education (years)	-0.011***
	(0.002)
male	0.306***
	(0.015)
constant	-1.945***
	(0.268)
year fixed effects	included
Observations	133,707
Pseudo R^2	0.191

Table 78: Results from mortality probit regression using the subjective instrumented by objective (SIO) frailty index and 1998–2016 HRS data. Standard errors are in parenthesis. Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

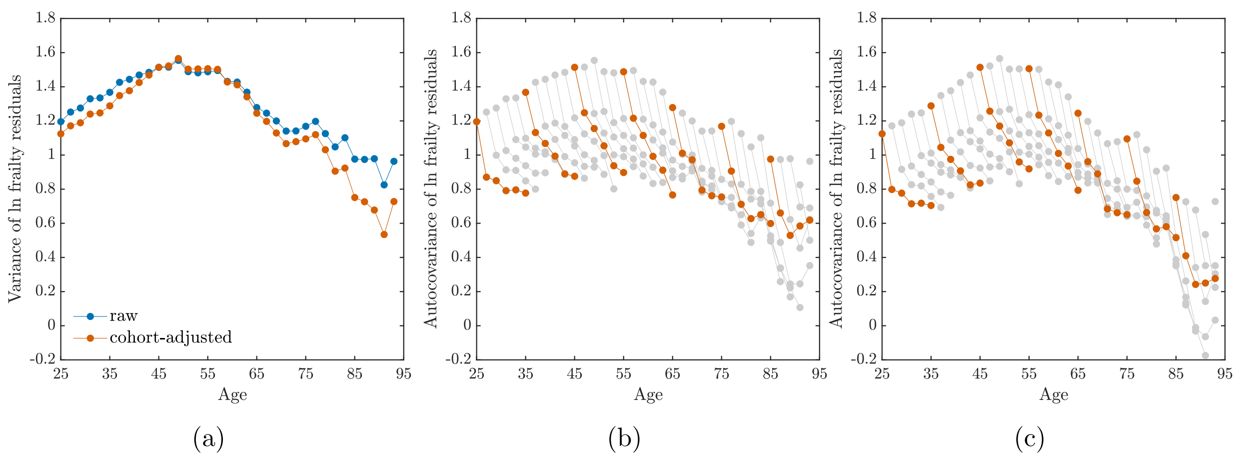


Figure 34: Raw and cohort-adjusted variances (a), raw variance-covariance matrix (b), and cohort-adjusted variance-covariance matrix (c) of the log frailty residuals for the main PSID sample using the subjective instrumented by objective (SIO) frailty index. Log frailty residuals are derived from an OLS regression of log frailty on a quartic in age, education dummies, and gender dummies.

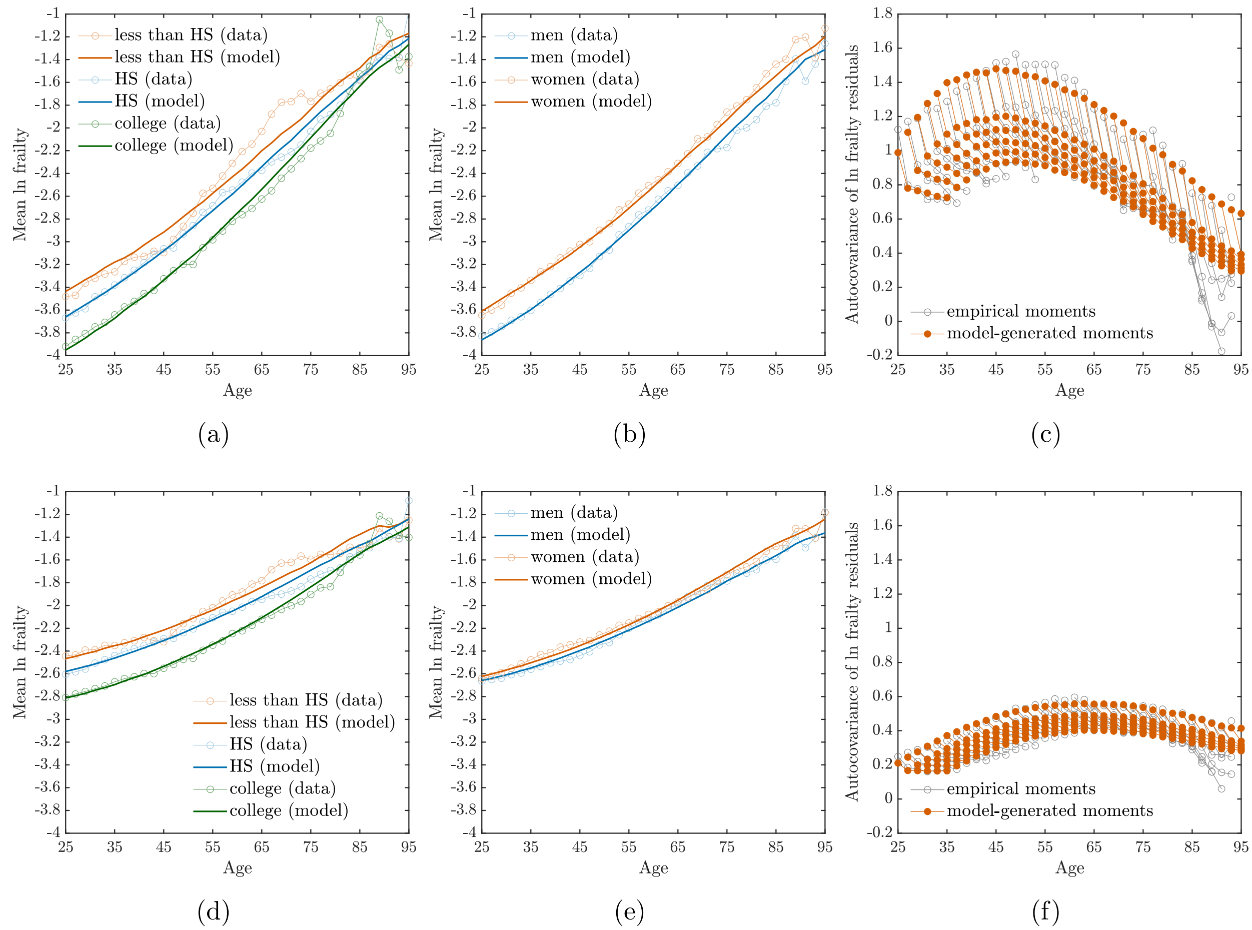


Figure 35: Fit of the estimated model using the subjective instrumented by objective (SIO) frailty index and baseline frailty (provided for comparison): empirical moments targeted in the baseline SMM estimation and their model counterparts. Panel (a) is mean log frailty by age in the data (open circles) and the model (solid lines) for those without a HS degree (orange), HS graduates (blue), and college graduates (green). Panel (b) is mean log frailty by age in the data (open circles) and the model (solid lines) for women (orange) and men (blue). Panel (c) is autocovariance profiles of log frailty residuals by age in the data (open gray circles) and the model (closed orange circles). Panels (d), (e), and (f) are the counterparts for the baseline frailty estimation results provided in the paper. Note that the y-axes have been changed here to ease comparability across the sets of moments for the two health measures.

of the variance profile and autocovariance profiles for the SIO index are very similar to those for the baseline frailty index but even more similar to those of the FPC index. The biggest difference between the moments of the SIO index and the FPC index is that the overall level of the variances and covariances is higher for the SIO index.

Figure 35 and Figure 36 show the empirical moments targeted in the SMM estimation and the model counterparts. Table 79 presents the SMM estimation results. Comparing across figures and tables indicates that the results are qualitatively similar to those of the

Deterministic component			Stochastic component		
Variable	Value	Std. Error	Variable	Value	Std. Error
age*	2.617	(0.009)	σ_u^2	0.192	(0.006)
age*2	3.048	(0.028)	σ_ε^2	0.075	(0.004)
age*3	-0.687	(0.043)	σ_α^2	0.687	(0.012)
age*4	5.594	(0.038)	ρ	0.961	(0.002)
HS degree	-0.230	(0.001)			
college degree	-0.499	(0.001)			
male	-0.244	(0.001)			
constant	-3.343	(0.001)			

Table 79: Results from estimating the log frailty process via SMM using the subjective instrumented by objective (SIO) frailty index and a sample of individuals aged 25–95 from PSID. Note: age* $\equiv$ (age $-$ 25)/100.

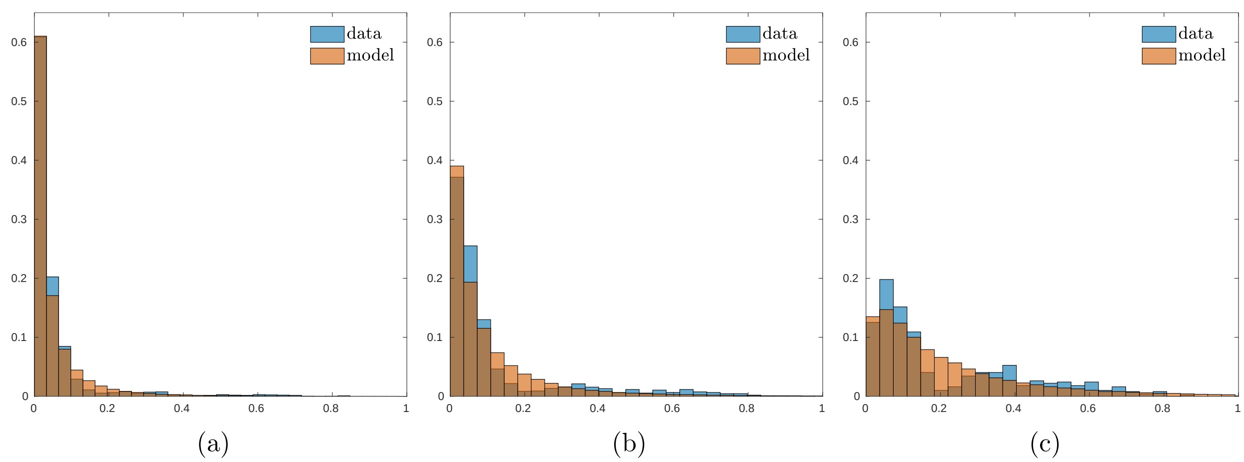


Figure 36: Histograms showing the cross-sectional distributions of frailty for 35–36 year-olds (a), 55–56 year-olds (b), and 75–76 year-olds (c) in the model and the data using the subjective instrumented by objective (SIO) frailty index.

other frailty variants and the most similar quantitatively to the FPC index estimates. Quantitatively, the two biggest differences between the estimation results using the FPC index and those using the SIO index are as follows. First, there is more variation in the SIO index and all three estimated variances are larger for this index than the FPC one. Second, the effect of being male on frailty is the largest (across all four frailty variants) in the SIO index. While being male reduces frailty on average by 14 percent according to the FPC index, it reduces the SIO index by 24 percent. The effects of age and education are very similar across the FPC and SIO indices as is the estimated persistence of the AR(1) shock. Finally, Figure 36 shows histograms of the FPC index in the model and data for the 35-36 year-olds, 55-56 year-olds, and 75-76 year-olds.

References

Blundell, R., M. C. Dias, J. Britton and E. French, "The impact of health on labor supply near retirement," *Journal of Human Resources* (2020), 1217–9240R4. A.1, A.1.1

Kulminski, A., A. Yashin, K. Arbeev, I. Akushevich, S. Ukraintseva, K. Land and K. Manton, "Cumulative index of health disorders as an indicator of aging-associated processes in the elderly: results from analyses of the National Long Term Care Survey," *Mechanisms of ageing and development* 128 (2007), 250–258. 2