Health Inequality and Health Types

Borella Bullano De Nardi Krueger Manresa

Health affects many key economic outcomes

- Labor supply, earnings, and retirement
- Medical expenses
- Life expectancy
- Savings

Health affects many key economic outcomes

- Labor supply, earnings, and retirement
- Medical expenses
- Life expectancy
- Savings

$\Rightarrow$ **Crucial to understand health and its dynamics**

Health affects many key economic outcomes: some references

- Labor supply, earnings, and retirement (French (2005); French and Jones (2011); Capatina and Keane (2023); Hosseini, Kopecky, and Zhao (2021); Blundell, Britton, Dias, and French (2023))
- Medical expenses (Jones, De Nardi, French, McGee, and Kirschner (2018))
- Life expectancy (Kopecky and Koreshkova (2014); De Nardi, French, and Jones (2010))
- Savings (De Nardi, French, and Jones (2010); De Nardi, Porapakkarm, and Paschenko (2017))

Our goals

Better understand, during middle and old age

- **How health and mortality evolve**

Our goals

Better understand, during middle and old age

- **How health and mortality evolve**
- **How unequal is their evolution**

Our goals

Better understand, during middle and old age

- How health and mortality evolve
- How unequal is their evolution
- How to better model the dynamics of health and mortality

To show that

- There are health types
- It is important to model health types and to understand them

Specific questions

- Q1. **Are there “health types” in adulthood? That is, do people have heterogeneous health trajectories?**

Specific questions

- Q1. **Are there “health types” in adulthood? That is, do people have heterogeneous health trajectories?**
- Q2. **What are those health types?**

Specific questions

- Q1. **Are there "health types" in adulthood? That is, do people have heterogeneous health trajectories?**
- Q2. **What are those health types?**
- Q3. **Can health types be captured by observables? Are we dealing with observed or unobserved heterogeneity?**

Specific questions

- Q1. **Are there "health types" in adulthood? That is, do people have heterogeneous health trajectories?**
- Q2. **What are those health types?**
- Q3. **Can health types be captured by observables? Are we dealing with observed or unobserved heterogeneity?**
- Q4. **How important are health types and what do we miss if we ignore them?**

Specific questions

- Q1. **Are there "health types" in adulthood? That is, do people have heterogeneous health trajectories?**
- Q2. **What are those health types?**
- Q3. **Can health types be captured by observables? Are we dealing with observed or unobserved heterogeneity?**
- Q4. **How important are health types and what do we miss if we ignore them?**
- Q5. **How can we parsimoniously model health and mortality dynamics?**

Q1. Are there “health types” in adulthood?

That is, do people have

heterogeneous health trajectories?

Measuring health

Health and Retirement Study (HRS) data, hence for the United States

- Individuals age 52 and older
- Biennial panel, use data from 1996 to 2018
- Rich and high-quality

Possible health deficits

ADLs
Difficulty bathing
Difficulty dressing
Difficulty eating
Difficulty getting in/out of bed
Difficulty using the toilet
Difficulty walking across a room
Difficulty walking one block
Difficulty walking several blocks

IADLs
Difficulty grocery shopping
Difficulty making phone calls
Difficulty managing money
Difficulty preparing a hot meal
Difficulty taking medication
Difficulty using a map

Other Functional Limitations
Difficulty climbing one flight of stairs
Difficulty climbing several flights of stairs
Difficulty getting up from a chair
Difficulty kneeling or crouching
Difficulty lifting a weight heavier than 10 lbs
Difficulty lifting arms over the shoulders
Difficulty picking up a dime
Difficulty pulling/pushing large objects
Difficulty sitting for two hours

Diagnoses
Diagnosed with high blood pressure
Diagnosed with diabetes
Diagnosed with cancer
Diagnosed with lung disease
Diagnosed with a heart condition
Diagnosed with a stroke
Diagnosed with psychological or psychiatric problems
Diagnosed with arthritis

Healthcare Utilization
Has stayed in the hospital in the previous two years
Has stayed in a nursing home in the previous two years

Addictive Diseases
Has BMI larger than 30
Has ever smoked cigarettes

Frailty, some references

- Health measure proposed in the **gerontology literature** (Mitnitski, Mogilner, and Rockwood (2001); Mitnitski, Mogilner, MacKnight, and Rockwood (2002); Mitnitski, Song, Skoog, Broe, Cox, Grunfeld, and Rockwood (2005); Goggins, Woo, Sham, and Ho (2005); Searle, Mitnitski, Gahbauer, Gill, and Rockwood (2008))
- **Advantages** over others health measure
  - Great predictor of economic and future outcomes (Hosseini, Kopecky, and Zhao (2022))
  - Including by race, ethnicity, and gender (Borella, De Nardi, McGee, Russo, and Abram (2023))

Frailty and our sample

- Health deficits are recorded as either present (=1) or not (=0)

Frailty and our sample

- Health deficits are recorded as either present (=1) or not (=0)
- **Frailty: the number of one's health deficits divided by the number of all possible health deficits (at each point in time)** Frailty distribution in our sample

Frailty and our sample

- Health deficits are recorded as either present (=1) or not (=0)
- **Frailty: the number of one's health deficits divided by the number of all possible health deficits (at each point in time)** Frailty distribution in our sample
- Include people from age 52-53 and until either death or 2018

Frailty and our sample

- Health deficits are recorded as either present (=1) or not (=0)
- **Frailty: the number of one's health deficits divided by the number of all possible health deficits (at each point in time)** Frailty distribution in our sample
- Include people from age 52-53 and until either death or 2018
- **Assign a frailty of 1 when people die (death is a manifestation of health)** Details

Extracting health types: K-means clustering

- Assign data to **clusters** (health types) so that
  - Observations in a cluster are as similar to each other as possible
  - Observations in different cluster are as different from each other as possible

Extracting health types: K-means clustering

- Assign data to **clusters** (health types) so that
  - Observations in a cluster are as similar to each other as possible
  - Observations in different cluster are as different from each other as possible
- **Method advantages**
  - Clustering provides a direct and intuitive mapping between types and people

Extracting health types: K-means clustering

- Assign data to **clusters** (health types) so that
  - Observations in a cluster are as similar to each other as possible
  - Observations in different cluster are as different from each other as possible
- **Method advantages**
  - Clustering provides a direct and intuitive mapping between types and people
  - Clustering is non-parametric

Extracting health types: K-means clustering

- Assign data to **clusters** (health types) so that
  - Observations in a cluster are as similar to each other as possible
  - Observations in different cluster are as different from each other as possible
- **Method advantages**
  - Clustering provides a direct and intuitive mapping between types and people
  - Clustering is non-parametric
  - K-means is only clustering method for which the statistical properties of identifying unobserved heterogeneity from discrete classification have been determined (Bonhomme, Lamadon, and Manresa (2022))

K-means clustering

- Cluster the data in a pre-specified number of groups (K)
- Associate each cluster (group) to a **centroid** (the cluster’s “representative agent")

K-means clustering

- Cluster the data in a pre-specified number of groups (K)
- Associate each cluster (group) to a **centroid** (the cluster's “representative agent")

- K-means output:
  - **Assignment**: cluster to which each data point is allocated
  - **Centroids for the K groups**: mean of observations belonging to each cluster

K-means definition

Our K-means algorithm implementation

- **Clustering period: from age 52 to 60, so the early part of our data**

Our K-means algorithm implementation

- **Clustering period: from age 52 to 60, so the early part of our data**
- Treat **health trajectory of each person over the clustering period as a vector**

$$h_i = [f_{i,52}, f_{i,54}, f_{i,56}, f_{i,58}, f_{i,60}]$$

where $f_{i,j}$ is frailty for person i at age j

Our K-means algorithm implementation

- **Clustering period: from age 52 to 60, so the early part of our data**
- Treat **health trajectory of each person over the clustering period as a vector**

$$h_i = [f_{i,52}, f_{i,54}, f_{i,56}, f_{i,58}, f_{i,60}]$$

  where $f_{i,j}$ is frailty for person i at age j

- **Cluster these health trajectories for each person**

Our K-means algorithm implementation

- **Clustering period: from age 52 to 60, so the early part of our data**
- Treat **health trajectory of each person over the clustering period as a vector**

$$h_i = [f_{i,52}, f_{i,54}, f_{i,56}, f_{i,58}, f_{i,60}]$$

  where $f_{i,j}$ is frailty for person i at age j

- **Cluster these health trajectories for each person**
- **As a result, people of each health type will have**
  - **Similar initial health**
  - **Similar health trajectories during this earlier period**

Choosing the number of clusters, or health types

Economic criteria

- Maximize predictive performance of health types for frailty and mortality during the clustering period
  - Choose K such that increasing K does not *improve* the predictive power of these regressions Predictive power
  - Estimate using cross-validation Details

Machine learning criteria

- Elbow Details and silhouette criteria Details

Choosing the number of clusters, or health types

Economic criteria

- Maximize predictive performance of health types for frailty and mortality during the clustering period
  - Choose K such that increasing K does not *improve* the predictive power of these regressions Predictive power
  - Estimate using cross-validation Details

Machine learning criteria

- Elbow Details and silhouette criteria Details

Obtain 5 health types Details

Choosing the number of clusters, or health types

Economic criteria

- Maximize predictive performance of health types for frailty and mortality during the clustering period
  - Choose K such that increasing K does not *improve* the predictive power of these regressions Predictive power
  - Estimate using cross-validation Details

Machine learning criteria

- Elbow Details and silhouette criteria Details

Obtain 5 health types Details

Clusters explain 84% of the variation in health trajectories

Are these really health types?

- **Do health types predict future frailty and mortality dynamics**?

Are these really health types?

- **Do health types predict future frailty and mortality dynamics**?
- Forecast frailty and mortality **after age 60** (after our clustering period ends)
- Only include people still alive at 60

Are these really health types?

- **Do health types predict future frailty and mortality dynamics**?
- Forecast frailty and mortality **after age 60** (after our clustering period ends)
- Only include people still alive at 60
- Controls: education, race, gender, HRS cohort, marital status, 3rd-order polynomial in age

Are these really health types?

- **Do health types predict future frailty and mortality dynamics**?
- Forecast frailty and mortality **after age 60** (after our clustering period ends)
- Only include people still alive at 60
- Controls: education, race, gender, HRS cohort, marital status, 3rd-order polynomial in age
- Initial Health: Frailty and SRHS at 52
- Health types

Are these really health types?

- **Do health types predict future health and mortality dynamics**?

Are these really health types?

- **Do health types predict future health and mortality dynamics**?

	Frailty				Death			
Controls	x	x	x	x	x	x	x	x
Initial health			x	x			x	x
Health types		x		x		x		x
R^2	0.120	0.566	0.503	0.586				
Pseudo-R^2					0.138	0.199	0.177	0.202

Yes! Large increase in out-of-sample predictive power

Initial health important to explain future health outcomes and mortality, but outperformed by health types

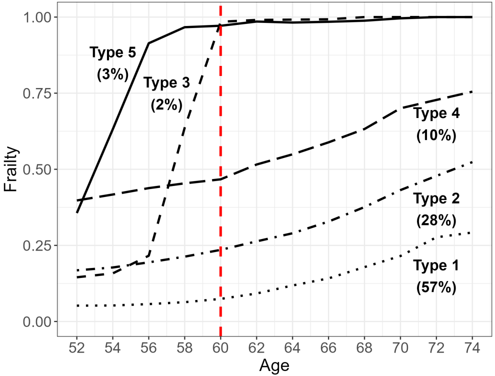
Frailty Mortality K robustness - frailty K robustness - mortality

Answers to Q1. Are there "health types" in adulthood? That is, do people have heterogeneous health trajectories?

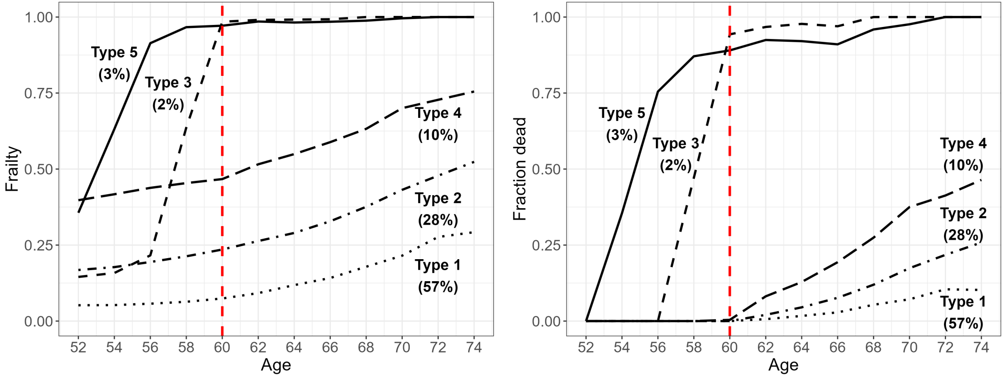
- Yes, we uncover 5 health types
- These health types
  - Help capture health and mortality dynamics during clustering period (age 52-60): Clusters explain 84% of the variation in health trajectories
  - Are key predictors of health and mortality after age 60

Q2. What are those health types?

Average frailty and fraction dying by health type and age

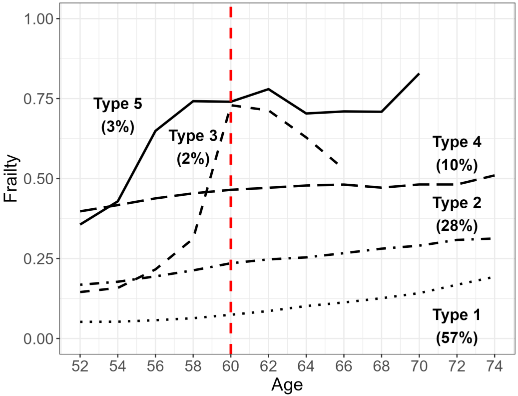


Average frailty and fraction dying by health type and age



- Different health dynamics, both during and after the clustering period
- Types 2 and 3, and types 4 and 5 start out similarly but evolve very differently
- Table Cause of death

Average frailty of survivors by health type and age



Even conditional on survival

- Different health dynamics by health types
- Types 2 and 3, and types 4 and 5 start out similarly but evolve very differently
- Frailty distribution

Answers to Q2. What are those health types?

- **At age 52 health is very unequally distributed.** On average,
  - Type 1: 2 health deficits
  - Types 2 and 3: 6 health deficits
  - Types 4 and 5: 14 health deficits

Answers to Q2. What are those health types?

- **At age 52 health is very unequally distributed.** On average,
  - Type 1: 2 health deficits
  - Types 2 and 3: 6 health deficits
  - Types 4 and 5: 14 health deficits
- **After age 52 heterogeneous trajectories**
  - Most people's frailty increases slowly
  - A small fraction of people (overall 5%) experiences fast health deterioration

Answers to Q2. What are those health types?

- **At age 52 health is very unequally distributed.** On average,
  - Type 1: 2 health deficits
  - Types 2 and 3: 6 health deficits
  - Types 4 and 5: 14 health deficits
- **After age 52 heterogeneous trajectories**
  - Most people's frailty increases slowly
  - A small fraction of people (overall 5%) experiences fast health deterioration
- **Our 5 health types**
  - Type 1: The vigorous resilient
  - Type 2: The fair-health resilient
  - Type 3: The fair-health vulnerable
  - Type 4: The frail resilient
  - Type 5: The frail vulnerable

Heatmaps

Q3. Can health types be captured by observables?

Are we dealing with observed or unobserved heterogeneity?

Health types and demographics

	All sample	Type 1	Type 2	Type 3	Type 4	Type 5
Fraction of people	1	0.57	0.28	0.02	0.10	0.03
Fraction women	0.63	0.59	0.69	0.57	0.73	0.55
Fraction black people	0.17	0.13	0.20	0.28	0.28	0.28
Mean years of education	13.01	13.60	12.46	12.72	11.52	12.27
Fraction partnered at 52	0.78	0.82	0.77	0.66	0.64	0.63
Mean individual income at 52	30,828	39,303	24,239	18,177	10,818	9,941
Mean household income at 52	56,322	70,156	45,660	34,925	22,211	26,710

- Women less likely to be healthy but do not tend to deteriorate quickly
- Black people less likely to be healthy but do not deteriorate faster
- More educated more likely to be of Type 1
- People in couples more likely to be of Type 1
- Clear gradient for individual income but not for household income

Health behaviors and health insurance status by health type

	All sample	Type 1	Type 2	Type 3	Type 4	Type 5
Fraction of people	1	0.57	0.28	0.02	0.10	0.03
Health behaviours						
Fraction ever smoked	0.56	0.49	0.64	0.72	0.67	0.76
Fraction vigorous activity at 52	0.50	0.61	0.44	0.46	0.21	0.22
Health insurance status						
Private health insurance at 52	0.76	0.85	0.74	0.61	0.42	0.41
Public health insurance at 52	0.13	0.04	0.13	0.19	0.45	0.49
Medicaid	0.06	0.01	0.06	0.07	0.24	0.29
Medicare	0.06	0.01	0.06	0.12	0.25	0.26
Uninsured at 52	0.14	0.12	0.16	0.22	0.20	0.17

- Smoking increasing in frailty type and more prevalent for fast deterioration types
- Exercise highest for type one and decreasing in frailty type, but similar for slow and fast deterioration types
- Private insurance decreasing in frailty type. Public insurance increasing

Can observables explain health types?

- Health types are often ignored. Exceptions in structural models: De Nardi, Pashchenko, and Porapakkarm (2017); Bolt (2021); Bairoliya, Miller, and Nygaard (2024); Capatina and Keane (2023)
- Observables that are correlated with health are typically considered

Can observables explain health types?

- Health types are often ignored. Exceptions in structural models: De Nardi, Pashchenko, and Porapakkarm (2017); Bolt (2021); Bairoliya, Miller, and Nygaard (2024); Capatina and Keane (2023)
- Observables that are correlated with health are typically considered
- This partial analysis shows some interesting correlations

Can observables explain health types?

- Health types are often ignored. Exceptions in structural models: De Nardi, Pashchenko, and Porapakkarm (2017); Bolt (2021); Bairoliya, Miller, and Nygaard (2024); Capatina and Keane (2023)
- Observables that are correlated with health are typically considered
- This partial analysis shows some interesting correlations
- But to what extent do observables capture health type heterogeneity? To what extent is this unobserved heterogeneity?

Can observables explain health types?

- Health types are often ignored. Exceptions in structural models: De Nardi, Pashchenko, and Porapakkarm (2017); Bolt (2021); Bairoliya, Miller, and Nygaard (2024); Capatina and Keane (2023)
- Observables that are correlated with health are typically considered
- This partial analysis shows some interesting correlations
- But to what extent do observables capture health type heterogeneity? To what extent is this unobserved heterogeneity?
- Move to a more systematic exercise to understand the relationship between health types and observables

Run multinomial logistic regression of health types on

- Initial health
- Many other observables

Can observables explain health types?

	Health Types		
	(1)	(2)	(3)
Initial Frailty		x	x
Demographics	x		x
Health behaviours	x		x
Health insurance	x		x
Pseudo R2	0.133	0.434	0.451

Demographics: Education, race, gender, HRS cohort, marital status, and household total income. Health behaviors: Ever Smoked and vigorous activity dummies. Health insurance: Private and public health insurance dummies.

- Model with rich set of observables has poor performance
- Initial frailty alone substantially increases predictive power
- Adding observables to initial frailty has a small effect

Can observables explain health types?

	Health Types		
	(1)	(2)	(3)
Initial Frailty		x	x
Demographics	x		x
Health behaviours	x		x
Health insurance	x		x
Pseudo R2	0.133	0.434	0.451

Demographics: Education, race, gender, HRS cohort, marital status, and household total income. Health behaviors: Ever Smoked and vigorous activity dummies. Health insurance: Private and public health insurance dummies.

- Model with rich set of observables has poor performance
- Initial frailty alone substantially increases predictive power
- Adding observables to initial frailty has a small effect

⇒ **Health types parsimonious way to capture health heterogeneity**

Other determinants Frailty composition

Answers to Q3. Can health types be captured by observables? Are we dealing with observed or unobserved heterogeneity?

Health types

- Are not captured by observables
- Reflect unobserved heterogeneity
- Are a very parsimononious way of capturing health heterogeneity

Q4. How important are health types and what do we miss if we ignore them?

How important are health types?

Model self-reported health status, from age 52 to death, as

- Excellent, Very good, Good, Fair, Poor, Dead

State-of-the-art Markov 1 model for health dynamics

- **Rich set of observables**
  - Age and age squared
  - Current health
  - Couple status
  - Education
  - ... all interacted with gender
- **Health types**

Model details

Do health types help explain SRHS from age 52 and until death?

	Future SRHS	
	(1)	(2)
Observables	x	x
Health types		x
Pseudo R^2	0.257	0.292

Observables: Current SRHS, education, couple status and 2^{nd} order polynomial in age, interacted with gender

- Yes! Even when controlling for health and a rich set of observables, reject the hypothesis that health types do not affect health
- **Health types are important drivers of health dynamics, even when we include a rich set of observables**

Health types and their implications for health dynamics

- Use state-of-the-art multinomial logit models for SRHS and mortality
- Simulate health and mortality paths
- Conditional on one's initial health type and other characteristics

Health types and their implications for health dynamics

- Use state-of-the-art multinomial logit models for SRHS and mortality
- Simulate health and mortality paths
- Conditional on one's initial health type and other characteristics

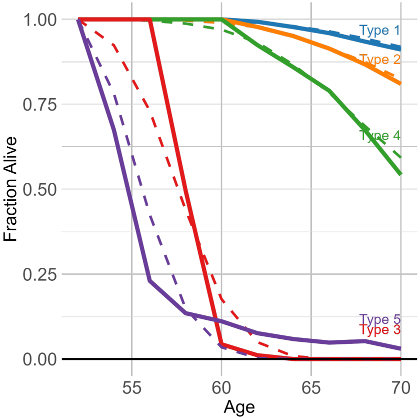
- **Display paths by one's health type and**
  - State-of-the-art model **without health types**
  - State-of-the-art model **with health types**

Health types and their implications for health dynamics

- Use state-of-the-art multinomial logit models for SRHS and mortality
- Simulate health and mortality paths
- Conditional on one's initial health type and other characteristics
- **Display paths by one's health type and**
  - State-of-the-art model **without health types**
  - State-of-the-art model **with health types**
- Comparing data and model for
  - Fraction of people alive by age
  - Fraction of people in *Good health* (good, very good or excellent), conditional on being alive

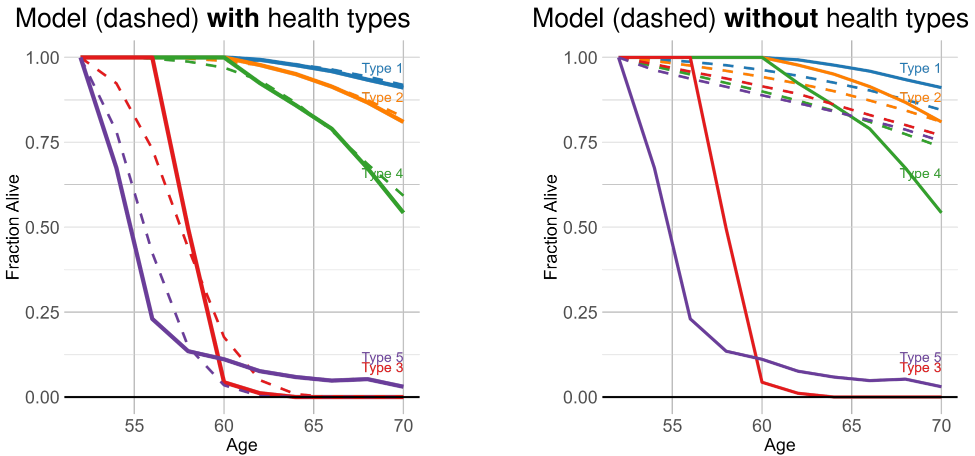
Fraction of people alive by health type

Model (dashed) **with** health types



Model (dashed) **without** health types

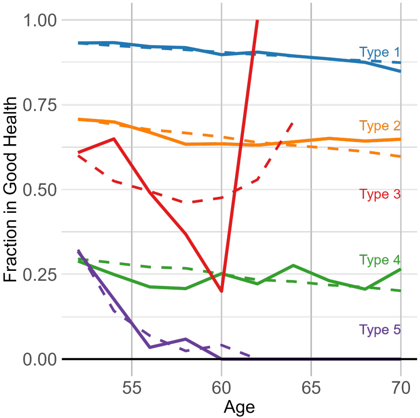
Fraction of people alive by health type



- State-of-the-art model without health types misses timing and heterogeneity in mortality

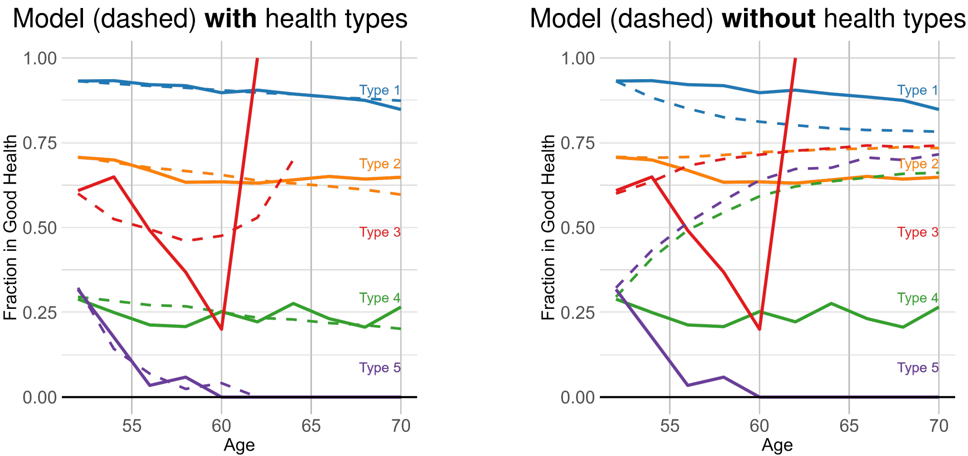
Fraction of people in good health by health Type

Model (dashed) **with** health types



Model (dashed) **without** health types

Fraction of people in good health by health type



- State-of-the-art model no health types misses fraction in Good Health by health type

Answers to Q4. What do we miss if we ignore health types?

Even a state-of-the-art model model of health and mortality without health types misses

- **Most heterogeneity in the timing of death by health type**
- **The evolution of health by health type, even conditional on survival**

Q5. How can we parsimoniously model health and mortality

What if we only include health types and initial health?

	Future SRHS and mortality	
	(1)	(2)
Observables	x	
Current Health	x	x
2^{nd} *order polynomial in age*	x	x
Health types		x
Pseudo R^2	0.257	0.285

- First column: observables include education, couple, and 2^{nd} order polynomial in age. All regressors are interacted with gender

What if we only include health types and initial health?

	Future SRHS and mortality	
	(1)	(2)
Observables	x	
Current Health	x	x
2^{nd} *order polynomial in age*	x	x
Health types		x
Pseudo R^2	0.257	0.285

- First column: observables include education, couple, and 2^{nd} order polynomial in age. All regressors are interacted with gender
- **Simple model with health types, previous health, and age outperforms model with rich observables and no health types**

Answers to Q5. How can we parsimoniously model health and mortality?

- **Identify health types**
- **Use simple model including age, current health, and health types. No need for other observables**

Conclusions

- Propose a new method to evaluate health outcomes, based on *health trajectories*
- Find health types that have heterogeneous health deterioration and mortality

Conclusions

- Propose a new method to evaluate health outcomes, based on *health trajectories*
- Find health types that have heterogeneous health deterioration and mortality
- Health types are unobservable but easily attributed to people using K-means clustering

Conclusions

- Propose a new method to evaluate health outcomes, based on *health trajectories*
- Find health types that have heterogeneous health deterioration and mortality
- Health types are unobservable but easily attributed to people using K-means clustering
- Ignoring health types misses the dynamics of both health and mortality

Directions for future research

- Modelling health types important to better
  - Understand health inequality
  - Evaluate to what extent health inequality drives inequality in economic outcomes
  - Study the effects of policy countefactuals

Directions for future research

- Modelling health types important to better
  - Understand health inequality
  - Evaluate to what extent health inequality drives inequality in economic outcomes
  - Study the effects of policy countefactuals
- Quantify how long of a history we need to identify health types

Directions for future research

- Modelling health types important to better
  - Understand health inequality
  - Evaluate to what extent health inequality drives inequality in economic outcomes
  - Study the effects of policy countefactuals
- Quantify how long of a history we need to identify health types
- Assess to what extent people know their health type and when

Directions for future research

- Modelling health types important to better
  - Understand health inequality
  - Evaluate to what extent health inequality drives inequality in economic outcomes
  - Study the effects of policy countefactuals
- Quantify how long of a history we need to identify health types
- Assess to what extent people know their health type and when
- Evaluate health types earlier in life

Directions for future research

- Modelling health types important to better
  - Understand health inequality
  - Evaluate to what extent health inequality drives inequality in economic outcomes
  - Study the effects of policy countefactuals
- Quantify how long of a history we need to identify health types
- Assess to what extent people know their health type and when
- Evaluate health types earlier in life
- Study to what extent health types relate to key economic outcomes
  - Education, marriage, and fertility decisions
  - Disability, length of working life, and retirement
  - Medical expenses
- What contributes to types formation and when? Bolt (2021)

References I

Bairoliya, N., R. Miller, and V. Nygaard (2024). Exercise or extra fries? how behavior impacts health over the life cycle. How Behavior Impacts Health Over the Life Cycle (February 18, 2024).

Blundell, R., J. Britton, M. C. Dias, and E. French (2023). The impact of health on labor supply near retirement. Journal of Human Resources 58(1), 282–334.

Bolt, U. (2021). What is the source of the health gradient? the case of obesity. Technical report, Mimeo, UCL.

Bonhomme, S., T. Lamadon, and E. Manresa (2022). Discretizing unobserved heterogeneity. Econometrica 90(2), 625–643.

Borella, M., M. De Nardi, R. McGee, N. Russo, and R. Abram (2023). Measuring and understanding health inequality by race and ethnicity.

Capatina, E. and M. Keane (2023). Health shocks, health insurance, human capital, and the dynamics of earnings and health. Working Paper 080, Federal Reserve Bank of Minneapolis.

References II

De Nardi, M., E. French, and J. Jones (2010). Why do the elderly save? Journal of Political Economy 118, 39–75.

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2017). The lifetime costs of bad health. Technical report, National Bureau of Economic Research.

De Nardi, M., P. Porapakkarm, and S. Paschenko (2017). The lifetime costs of bad health. NBER. Revised March 2022.

French, E. (2005, April). The effects of health, wealth, and wages on labour supply and retirement behaviour. The Review of Economic Studies 2, 395–427.

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

Goggins, W. B., J. Woo, A. Sham, and S. C. Ho (2005). Frailty index as a measure of biological age in a chinese population. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 60, 1046–1051.

References III

Hosseini, R., K. Kopecky, and K. Zhao (2021). How important is health inequality for lifetime earnings inequality?

Hosseini, R., K. Kopecky, and K. Zhao (2022). The evolution of health over the life cycle. Review of Economic Dynamics 45, 237–263.

Jones, J., M. De Nardi, E. French, R. McGee, and J. Kirschner (2018). The lifetime medical spending of retirees. Economic Quarterly 104.

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

Mitnitski, A., A. Mogilner, C. MacKnight, and K. Rockwood (2002). The accumulation of deficits with age and the possible invariants of aging. The Scientific World 2, 1816–1822.

Mitnitski, A., A. Mogilner, and K. Rockwood (2001). Accumulation of deficits as a proxy measure of aging. The Scientific World 1, 323–336.

References IV

Mitnitski, A., X. Song, I. Skoog, G. Broe, J. Cox, E. Grunfeld, and K. Rockwood (2005). Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality. Journal of American Geriatrics Society 53, 2184–2189.

Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65.

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

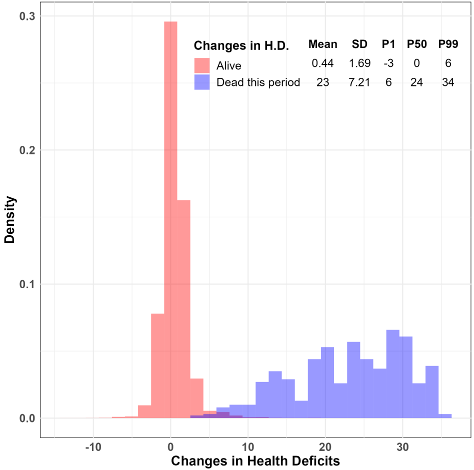
Thorndike, R. (1953, December). Who belongs in the family? Psychometrika.

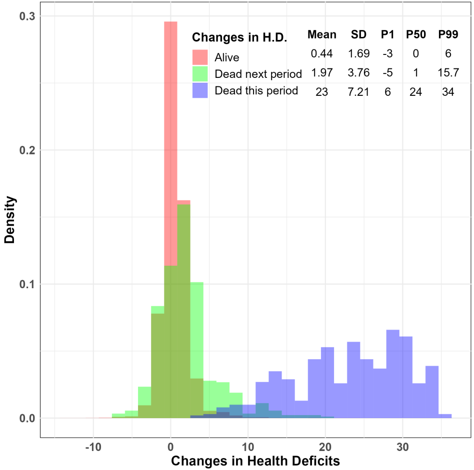
Additional Material

Frailty distribution in our sample

Number of Deficits	Average Frailty	Freq.	Percent.	Cumul Percent.
0	0.00	2141	5.78	5.78
1	0.03	5042	13.62	19.40
2	0.06	5257	14.20	33.60
3	0.09	4340	11.72	45.33
4	0.11	3660	9.89	55.21
5	0.14	2998	8.10	63.31
6	0.17	2249	6.07	69.38
7	0.20	1830	4.94	74.33
8	0.23	1414	3.82	78.15
9	0.26	1367	3.69	81.84
10	0.29	1077	2.91	84.75
11	0.31	899	2.43	87.18
12	0.34	687	1.86	89.03
13	0.37	700	1.89	90.92
14	0.40	596	1.61	92.53
15	0.43	531	1.43	93.97
16	0.46	445	1.20	95.17
17	0.49	352	0.95	96.12
18	0.51	269	0.73	96.85
19	0.54	214	0.58	97.43
20	0.57	194	0.52	97.95
21	0.60	188	0.51	98.46
22	0.63	156	0.42	98.88
23	0.66	126	0.34	99.22
24	0.69	73	0.20	99.42
25	0.71	65	0.18	99.59
26	0.74	39	0.11	99.70
27	0.77	40	0.11	99.81
28	0.80	33	0.09	99.89
29	0.83	17	0.05	99.94
30	0.86	15	0.04	99.98
31	0.89	6	0.02	100.00
32	0.91	1	0.00	100.00

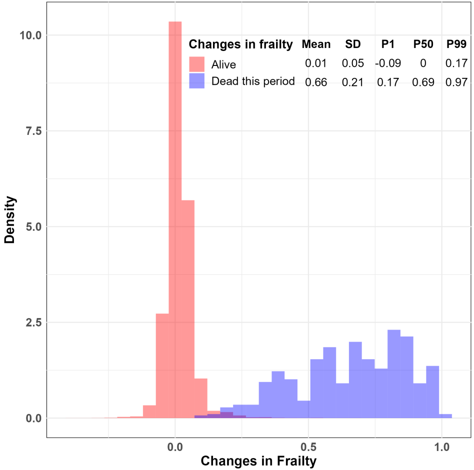
Change in health deficits between periods

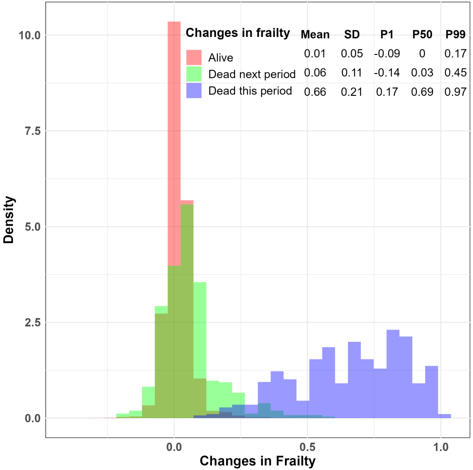




Back

Changes in frailty between periods





Back

Cause of death

	Death Cause				Death expected?		Death
	Cancer	Heart	Other Health-related	Non-health related	Expected	Unexpected	during clustering period
Type 1	0.49	0.25	0.23	0.03	0.60	0.40	0.00
Type 2	0.35	0.31	0.32	0.02	0.49	0.51	0.00
Type 3	0.41	0.21	0.30	0.09	0.47	0.53	0.94
Type 4	0.18	0.26	0.55	0.01	0.38	0.63	0.00
Type 5	0.28	0.29	0.37	0.05	0.44	0.56	0.89
Overall	0.35	0.27	0.34	0.04	0.48	0.52	0.051

Overall:

- Two major causes of death
  Cancer/Tumors and *Heart conditions* represent 62% of total deaths
- *Other health conditions* and *Non-health related* accounts for 34% and 4%
- 48% of death were *expected*

By health types:

- Low heterogeneity across types
- Types 3 and 5 depict patterns similar to the overall sample

Back

K-means algorithm

Unsupervised clustering algorithm designed to partition data into “K” groups

$$\left(\hat{h}(1), ..., \hat{h}(K), \{\hat{k}_i\}_{i=1}^{N}\right) = argmin_{\left(\tilde{h}(1),...,\tilde{h}(K),\{k_i\}_{i=1}^{N}\right)} \sum_{i=1}^{N} \left\| h_i - \tilde{h}(k_i) \right\|^2$$

- $\hat{h}(j)$ is the cluster j *centroid* (mean of data point belonging to j)
- $\{\hat{k}_i\}_{i=1}^{N}$ is a partition of the N data points, h_i, into K groups
- h_i is a data point and $\tilde{h}(k_i)$ is a possible centroid for cluster k_i

Back

Traditional machine learning methods - Elbow method

Elbow method Thorndike (1953):

- Calculate the proportion of the total variance explained by the clusters

$$\omega(k) = 1 - \frac{\sum_{i=1}^{N} \left\| h_i - \tilde{h}(k_i) \right\|^2}{\sum_{i=1}^{N} \left\| h_i - \bar{h} \right\|^2}$$

- Plot $\omega(k)$
- Choose k when the increase in this ratio using $k+1$ cluster is *small*
- Plot depicts an *elbow* at k

Back

Traditional machine learning methods - Silhouette method

- Silhouette measure (Rousseeuw (1987)) increases with average distance between clusters and decreases with variance within clusters

$$s(i) = \begin{cases} 0 & |C_I| = 1 \\ \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} & \text{otherwise} \end{cases}$$

  $a(i)$: mean distance between i and other points within the same cluster, $b(i)$: mean distance between i and the points in the nearest cluster, $|C_I|$ is cluster size

  Details

- Criterion: select the number of clusters that maximizes the average silhouette of the clustering

Back

Traditional machine learning methods - Silhouette method

Given some point i, letting $i \in C_I$ for some cluster C_I, define:

$$a(i) = \frac{1}{|C_I| - 1} \sum_{j \in C_I, j \neq i} d(i,j)$$

$$b(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{j \in C_J} d(i,j)$$

Where $|\cdot|$ gives set size and d is the euclidean distance, so that $a(i)$ is the mean distance between i and other points within the same cluster and $b(i)$ is the mean distance between i and the points in the nearest cluster. Then the silhouette at point i is given by:

$$s(i) = \begin{cases} 0 & |C_I| = 1 \\ \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} & \text{otherwise} \end{cases}$$

Back Silhouette

Back

Regressions for frailty and mortality between age 52 and 60

$$f_{it} = \mathbf{a}\mathbf{X}_{it} + f_{age}(t) + \sum_{\eta=1}^{k} a_\eta D_{i\eta} + \epsilon_{it}^{u} \tag{1a}$$

$$f_{it} = \mathbf{a}\mathbf{X}_{it} + f_{age}(t) + \epsilon_{it}^{u} \tag{1b}$$

$$P(D_{it}|\mathbf{X}_{it}, \eta) = \Lambda(\mathbf{b}\mathbf{X}_{it} + g_{age}(t) + \sum_{\eta=1}^{k} b_\eta D_{i\eta}) \tag{2a}$$

$$P(D_{it}|\mathbf{X}_{it}) = \Lambda(\mathbf{b}\mathbf{X}_{it} + g_{age}(t)) \tag{2b}$$

$\mathbf{X}_{it}$: education, race, gender, HRS cohort, marital status, age
$D_{i\eta}$ health types dummies

AME Details

Back

Absolute Mean Error

For a given number of cluster k

- Estimate the *absolute mean error* (AME)

$$\underbrace{AME(k) = \frac{1}{N}\sum_{i}^{N} |y_{it} - f(x_{it}, \eta_k; \theta)|}_{\text{with cluster information}} \qquad \underbrace{AME = \frac{1}{N}\sum_{i}^{N} |y_{it} - f(x_{it}; \theta)|}_{\text{without cluster information}}$$

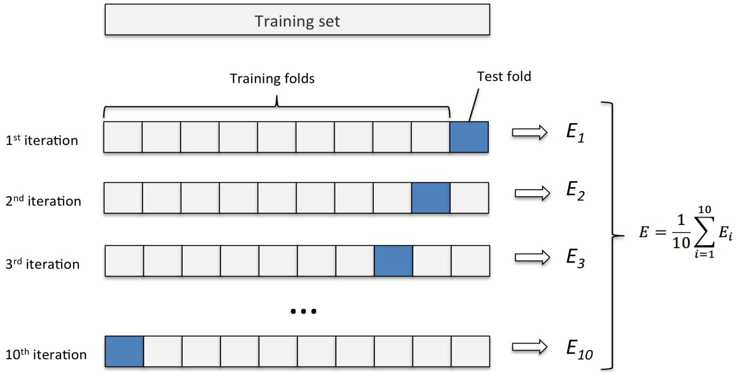
- Calculate $r(k)$

$$r(k) = \frac{\sum_i^N |y_{it} - f(x_{it}, \eta_k; \theta)|}{\sum_i^N |y_{it} - f(x_{it}; \theta)|}$$

Back Regressions

Back

Cross Validation: predicting over a sample not used for estimation



Back

Choosing the number of clusters/health types

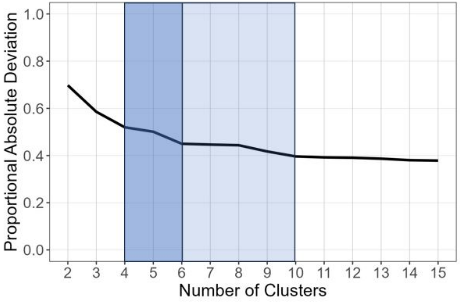


Figure: Frailty

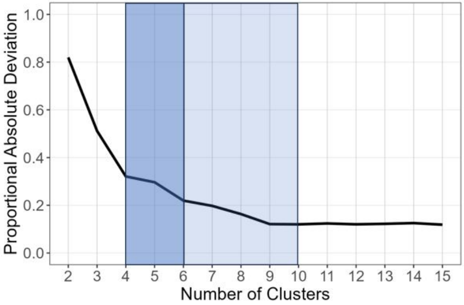
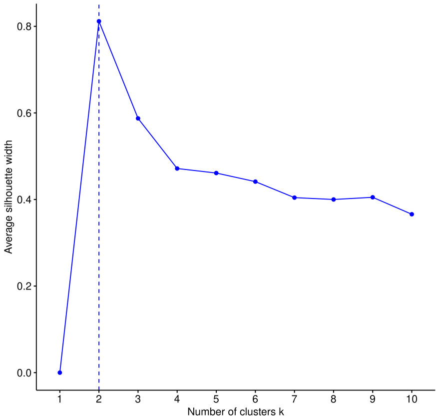


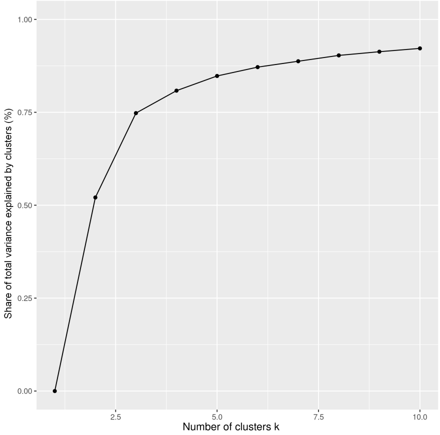
Figure: Mortality

- Elbow shows up between 4-6 cluster
- Traditional machine learning techniques indicate 2 to 5 clusters Traditional methods
- Choose 5 clusters

Back

Traditional Methods





The graph on the left shows the average silhouette of a clustering against the number of clusters. The graph on the right shows proportion of total variance explained by clusters against the number of clusters.

Back Number of Cluster

Out-of-sample frailty regressions

- We evaluate the out-of-sample predictive power by comparing (3) and (4)

$$f_{it} = X_{it}\beta + \epsilon_{it} \tag{3}$$

$$f_{it} = X_{it}\beta + \mathcal{D}_{i\eta}\beta^{\mathcal{D}} + \epsilon_{it} \tag{4}$$

- X_{it} is a rich set of controls, and $\mathcal{D}_{i\eta}$ are health types dummies
- X_{it}: age, (t_i), age squared (t_i^2), age cubed (t_i^3), Educational attainment (EA_i), race $(race_i)$, HRS cohort (HRS_i), women and marital status (c_{it}) dummies
- Alternative specification: X_{it} also include Initial frailty (f_{i52}) and initial SRHS (s_{i52}).

Back

Out-of-sample mortality regressions

- We evaluate the out-of-sample predictive power by comparing (5) and (6)

$$Pr(D_{i,t+2} = 1 | X_{it}) = \frac{e^{X_{it}\beta}}{1 + e^{X_{it}\beta}} \tag{5}$$

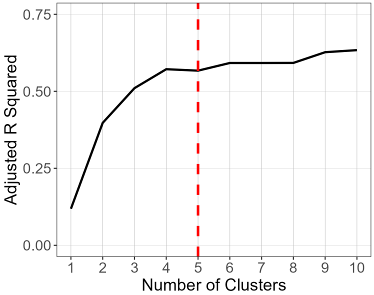
$$Pr(D_{i,t+2} = 1 | X_{it}, \mathcal{D}_{i\eta}) = \frac{e^{X_{it}\beta + \mathcal{D}_{i\eta}\beta^{\mathcal{D}}}}{1 + e^{X_{it}\beta + \mathcal{D}_{i\eta}\beta^{\mathcal{D}}}} \tag{6}$$

- X_{it} is a rich set of controls, and $\mathcal{D}_{i\eta}$ are health types dummies
- X_{it}: age, (t_i), age squared (t_i^2), age cubed (t_i^3), Educational attainment (EA_i), race $(race_i)$, HRS cohort (HRS_i), women and marital status (c_{it}) dummies
- Alternative specification: X_{it} also include Initial frailty (f_{i52}) and initial SRHS (s_{i52}).

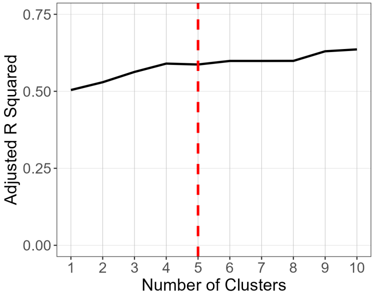
Back

Out-of-sample robustness to number of health types

Figure: Frailty next wave



(a) Demographics and Health types



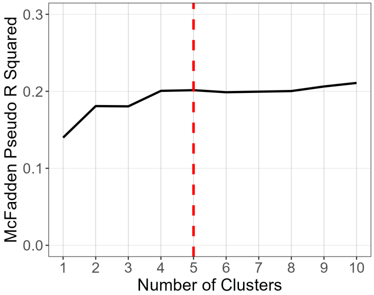
(b) Demographics, initial health and Health types

The red dotted line is our benchmark number of health types

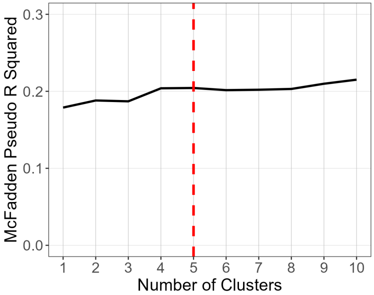
Back

Out-of-sample robustness to number of health types

Figure: Mortality next wave



(a) Demographics and Health types

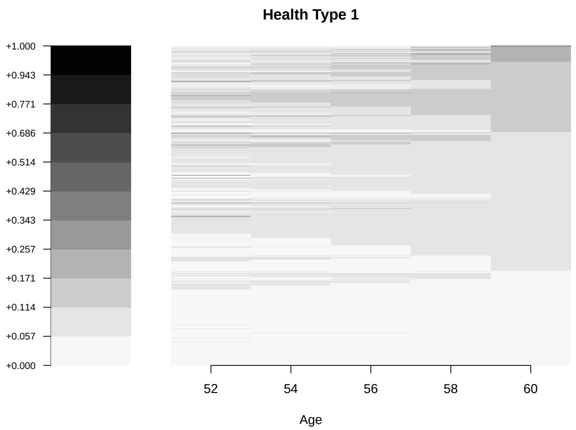


(b) Demographics, initial health and Health types

The red dotted line is our benchmark number of health types

Back

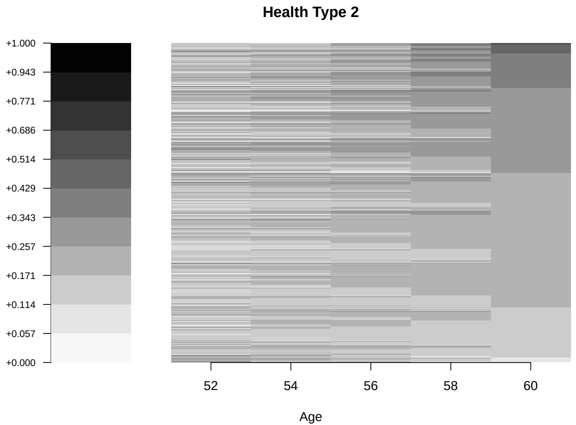
Evolution of frailty for each person



Type 1. The vigorous resilient: healthiest and unlikely to die (even after age 60)

Back

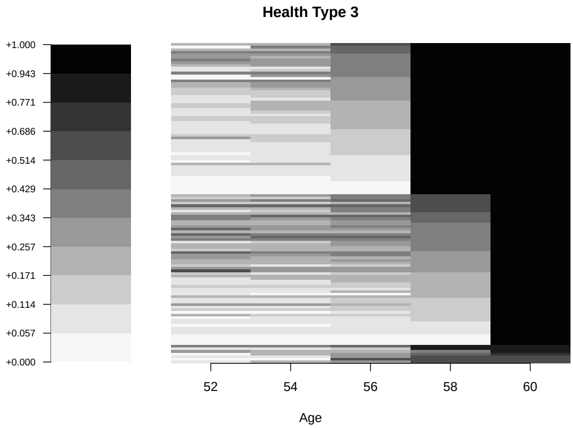
Evolution of frailty for each person



Type 2. The fair-health resilient: less healthy but still unlikely to die (even after age 60)

Back

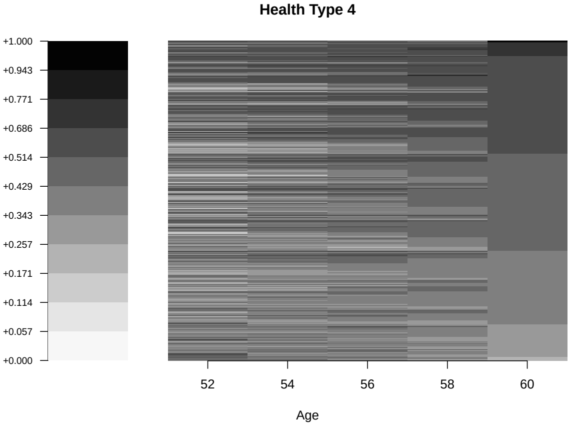
Evolution of frailty for each person



Type 3. The fair-health vulnerable: start in fair health but fast decline

Back

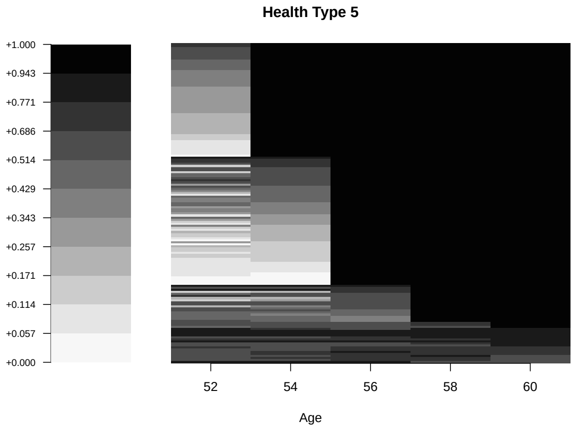
Evolution of frailty for each person



Type 4. The frale resilient: initially among the unhealthiest but resilient

Back

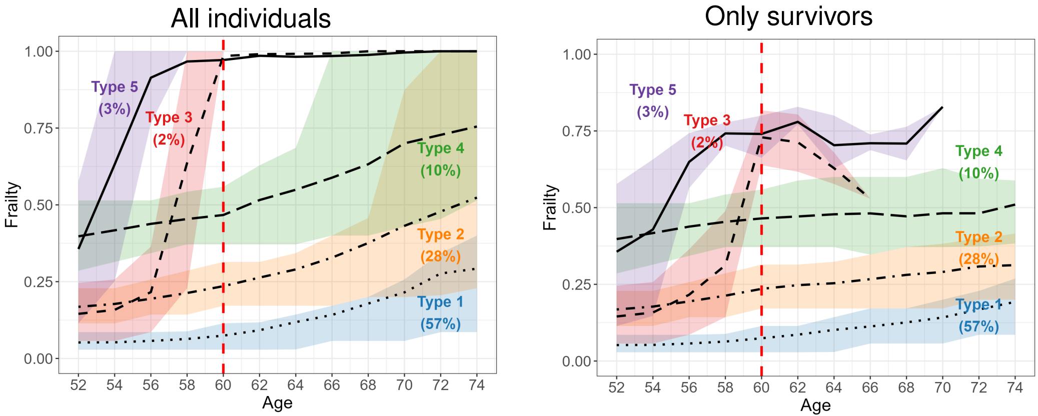
Evolution of frailty for each person



Type 5. The frail vulnerable: initially unhealthy and fast decline

Back

Frailty distribution by health types and age



Shaded area depicts the P80-P20 interval of frailty

Back

Main statistics by health type

	All sample	Type 1	Type 2	Type 3	Type 4	Type 5
Fraction of people	1	0.57	0.28	0.02	0.10	0.03
Health outcomes during clustering period						
Average frailty	0.17	0.06	0.20	0.43	0.44	0.77
Average health deficits	6.0	2.1	7.0	15.1	15.4	27.0
Fraction dead by 60	0.05	0	0	0.94	0	0.89
Health at 52						
Average frailty	0.13	0.05	0.17	0.15	0.40	0.36
Average health deficits	4.6	1.8	5.9	5.1	13.9	12.5
Average SRHS	2.64	2.12	3.01	3.15	4.03	3.95
Std. Dev. of frailty	0.14	0.04	0.08	0.12	0.13	0.23

Back

Health types and observable characteristics

	All sample	Type 1	Type 2	Type 3	Type 4	Type 5
Fraction of people	1	0.57	0.28	0.02	0.10	0.03
Health outcomes during clustering period						
Average frailty	0.17	0.06	0.20	0.43	0.44	0.77
Average health deficits	6.0	2.1	7.0	15.1	15.4	27.0
Fraction dead by 60	0.05	0	0	0.94	0	0.89
Health at 52						
Average frailty	0.13	0.05	0.17	0.15	0.40	0.36
Average health deficits	4.6	1.8	5.9	5.1	13.9	12.5
Average SRHS	2.64	2.12	3.01	3.15	4.03	3.95
Std. Dev. of frailty	0.14	0.04	0.08	0.12	0.13	0.23
Demographics						
Fraction women	0.63	0.59	0.69	0.57	0.73	0.55
Fraction black people	0.17	0.13	0.20	0.28	0.28	0.28
Mean years of education	13.01	13.60	12.46	12.72	11.52	12.27
Fraction partnered at 52	0.78	0.82	0.77	0.66	0.64	0.63
Mean individual income at 52	30, 828	39, 303	24, 239	18, 177	10, 818	9, 941
Mean household income at 52	56, 322	70, 156	45, 660	34, 925	22, 211	26, 710
Healthy behaviours						
Fraction ever smoked	0.56	0.49	0.64	0.72	0.67	0.76
Fraction vigorous activity at 52	0.50	0.61	0.44	0.46	0.21	0.22
Health insurance status						
Private health insurance at 52	0.76	0.85	0.74	0.61	0.42	0.41
Public health insurance at 52	0.13	0.04	0.13	0.19	0.45	0.49
Medicaid	0.06	0.01	0.06	0.07	0.24	0.29
Medicare	0.06	0.01	0.06	0.12	0.25	0.26
Uninsured at 52	0.14	0.12	0.16	0.22	0.20	0.17

Health type and observable characteristics: other determinants

	Health Types					
	(1)	(2)	(3)	(4)	(5)	(6)
Initial Frailty		x		x		x
Demographics	x			x	x	x
Healthy behaviours	x			x	x	x
Health insurance	x			x	x	x
Prob of living up to 75			x		x	x
Pseudo R2	0.133	0.434	0.032	0.451	0.147	0.456

Demographics: Education, race, gender, HRS cohort, marital status, and household total income. Health behaviors: Ever Smoked and vigorous activity dummies. Health insurance: Private and public health insurance dummies.

Back

What do we miss by using frailty instead of its underlying deficits?

Health deficits underlying frailty by type at age 52

- *ADLs*
- *IADLs*
- *Other functional limitations*
- *Health care utilization*
- *Diagnoses*
- *Addictive Diseases*

Back

What do we miss by using frailty instead of its underlying deficits?

	All Sample		Type 1		Type 2		Type 3		Type 4		Type 5	
Group of Deficits	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total
ADLs	10	0.4	1	0.0	6	0.4	7	0.4	18	2.5	20	2.5
IADLs	5	0.2	3	0.1	3	0.2	5	0.2	7	1.0	9	1.2
Other functional lim	37	1.7	23	0.4	41	2.4	36	1.8	43	6.0	36	4.5
Health care utilization	3	0.2	4	0.1	3	0.2	4	0.2	3	0.4	4	0.6
Diagnoses	25	1.1	30	0.5	27	1.6	28	1.4	19	2.6	21	2.7
Addictive	20	0.9	40	0.7	20	1.2	20	1.0	10	1.3	9	1.2
Deficits at 52	**100**	**4.6**	**100**	**1.8**	**100**	**5.9**	**100**	**5.1**	**100**	**13.9**	**100**	**12.5**

- Prevalence and number of deficit at 52 are heterogeneous between health types
- Types 2 and 3 and types 4 and 5 have **similar** frailty composition and levels
- "*Can observable explain health types?*" $\Rightarrow$ including frailty composition as observable characteristics does **not** help Details
- Frailty composition is **not key** in explaining health types

Back

Health type and observable characteristics: Frailty composition

	Health Types			
	(1)	(2)	(3)	(4)
Initial Frailty	x		x	
Initial Frailty composition		x		x
Demographics			x	x
Health behaviours			x	x
Health insurance			x	x
Pseudo R2	0.434	0.454	0.451	0.472

Demographics: Education, race, gender, HRS cohort, marital status, and household total income. Health behaviors: Ever Smoked and vigorous activity dummies. Health insurance: Private and public health insurance dummies. *Frailty composition*: ADls, IADLs, Other functional limitations, Health care utilization, diagnoses, and addictive diseases indexes.

Back

Back Main

Multinomial Regression details

$$Pr(SRHS_{i,t+2} = k \mid X_{it}) = \frac{e^{X_{it}\beta_k}}{\sum_{n=0}^{5} e^{X_{it}\beta_n}} \tag{7}$$

Model **without** health types:

X_{it}: includes age (t_i) age squared (t_i^2), current SRHS dummies (DHS_{it}), couple dummy (c_{it}), educational attainment dummies (EA_i) interacted with a woman dummy (w_i)

$$\begin{aligned} X_{it} = &(1, t_i, t_i^2, DHS_{it}, EA_i, c_{it}, \\ &(w_i, w_i t_i, w_i t_i^2, w_i DHS_{it}, w_i EA_i, w_i c_{it}) \end{aligned}$$

Back

Additional Material - Not for presentation

Cluster Assignments: K=4 and K=5

		K = 4				
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Row total
	Type 1	2837	0	0	0	2837
	Type 2	64	1310	1	0	1375
K = 5	Type 3	0	20	42	61	123
	Type 4	0	12	494	0	506
	Type 5	0	0	3	152	155
	Column Total	2901	1342	540	213	

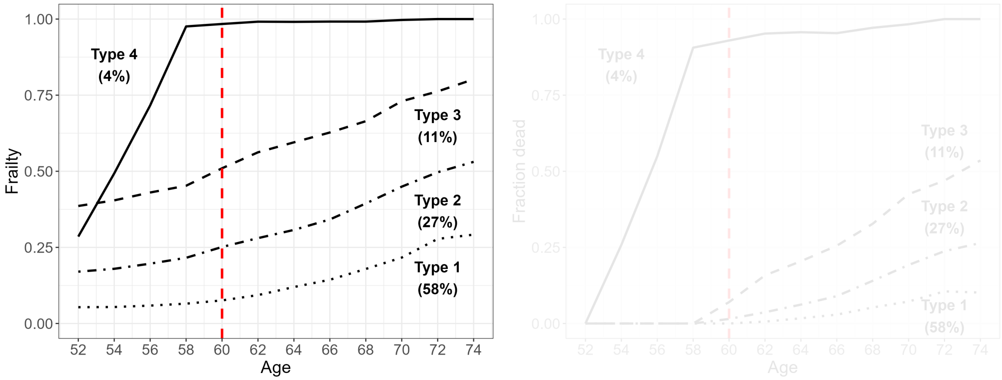
Back

Cluster Assignments: K=4 and K=5

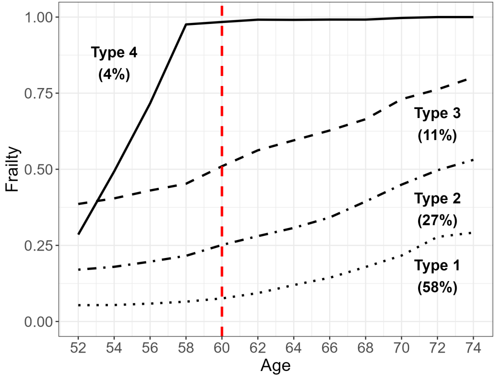
	All sample	Type 1	Type 2	Type 3	Type 4
Mean Frailty over clustering	0.17	0.06	0.20	0.44	0.69
Fraction dead by 60	0.05	0	0.01	0.07	0.93
Cluster size	1	0.58	0.27	0.11	0.04
Mean Frailty at 52	0.13	0.05	0.17	0.39	0.29
Mean SRHS at 52	2.64	2.13	3.03	4	3.68
Std. Dev. of Frailty at 52	0.14	0.04	0.08	0.14	0.23

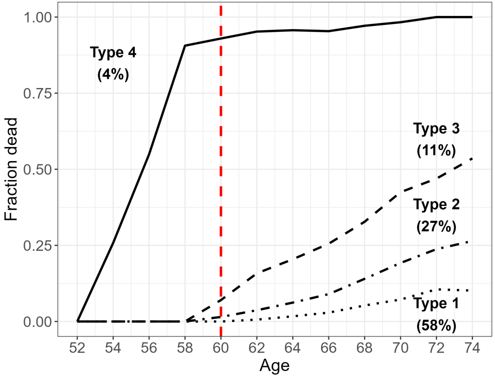
Back

Average frailty and fraction dying by health type and age

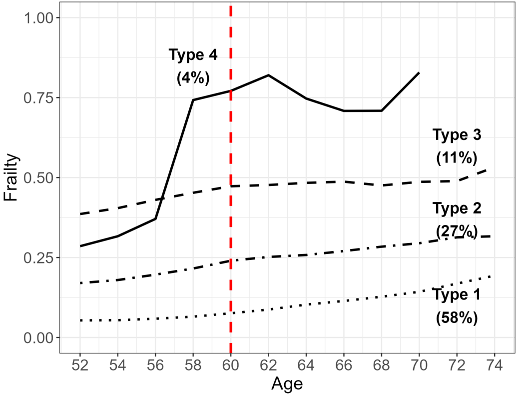


Average frailty and fraction dying by health type and age

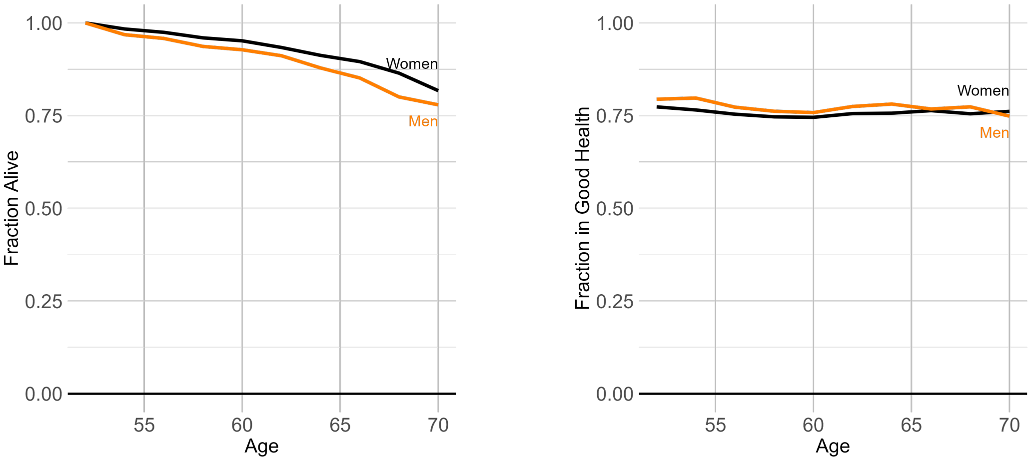




Average frailty of survivors by health type and age

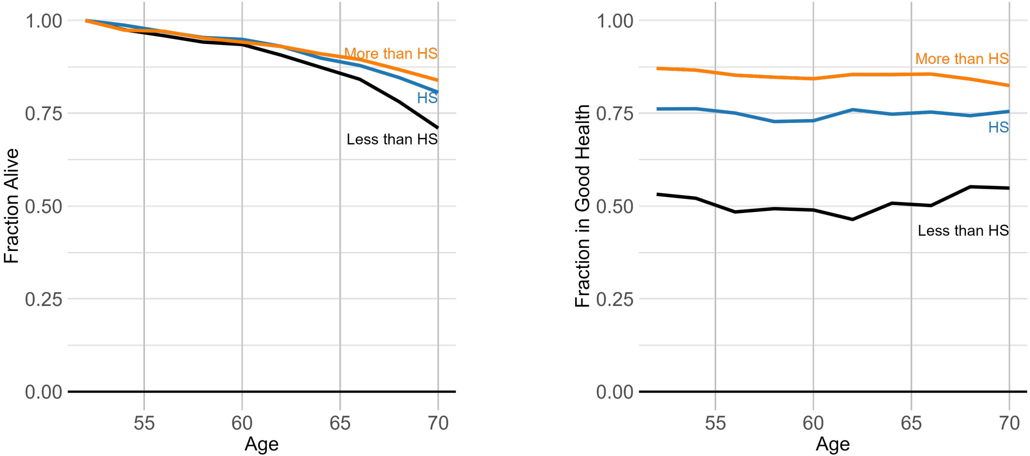


Difference in health outcomes by Sex



- **Fraction of people alive (left) and Fraction of people in good health (right)**
- **Much less variation by gender than by health type**

Difference in health outcomes by Education



- **Fraction of people alive (left) and Fraction of people in good health (right)**
- **Much less variation by education than by health type**