Health Inequality and Health Types

Borella  Bullano  De Nardi  Krueger  Manresa

The views expressed are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, the CEPR, the NBER, any agency of the federal government, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.
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

⇒ 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

### 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 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)

- 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)

- 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)**
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
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
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
  - Estimate using cross-validation

Machine learning criteria

- Elbow and silhouette criteria
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
  - Estimate using cross-validation

**Machine learning criteria**

- Elbow and silhouette criteria

**Obtain 5 health types**
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
  - Estimate using cross-validation

**Machine learning criteria**

- Elbow and silhouette criteria

**Obtain 5 health types**

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?

<table>
<thead>
<tr>
<th>Controls</th>
<th>Frailty Next Wave</th>
<th>Death Next Wave</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Initial health</th>
<th>Frailty Next Wave</th>
<th>Death Next Wave</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health types</th>
<th>Frailty Next Wave</th>
<th>Death Next Wave</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| R²             | 0.120            | 0.566           |
| Pseudo-R²      | 0.138            | 0.199           |

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

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

Types 2 and 3, and types 4 and 5 start out similarly. Evolve very differently. Different health dynamics, both during and after the clustering period.
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.
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
Q3. Can health types be captured by observables?

Are we dealing with observed or unobserved heterogeneity?
Health types and demographics

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

- 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

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

- 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?

<table>
<thead>
<tr>
<th>Health Types</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Frailty</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Health behaviours</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Health insurance</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.133</td>
<td>0.434</td>
<td>0.451</td>
</tr>
</tbody>
</table>


- 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?

<table>
<thead>
<tr>
<th>Health Types</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Frailty</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Demographics</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Health behaviours</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Health insurance</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.133</td>
<td>0.434</td>
<td>0.451</td>
</tr>
</tbody>
</table>


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

<table>
<thead>
<tr>
<th></th>
<th>Future SRHS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Observables</strong></td>
<td>x</td>
</tr>
<tr>
<td><strong>Health types</strong></td>
<td>x</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.257</td>
</tr>
</tbody>
</table>

Observables: Current SRHS, education, couple status and 2\textsuperscript{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

Model (dashed) **with** health types

Model (dashed) **without** health types

▶ 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 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?

<table>
<thead>
<tr>
<th>Observed Variables</th>
<th>Future SRHS and mortality</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observables</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Health</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>2\textsuperscript{nd} order polynomial in age</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Health types</td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

Pseudo $R^2$  

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo $R^2$</td>
<td></td>
<td>0.257</td>
</tr>
</tbody>
</table>

- First column: observables include education, couple, and 2\textsuperscript{nd} order polynomial in age. All regressors are interacted with gender.
What if we only include health types and initial health?

<table>
<thead>
<tr>
<th></th>
<th>Future SRHS and mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Observables</strong></td>
<td></td>
</tr>
<tr>
<td>Current Health</td>
<td>x</td>
</tr>
<tr>
<td>2(^{nd}) order polynomial in age</td>
<td>x</td>
</tr>
<tr>
<td>Health types</td>
<td></td>
</tr>
<tr>
<td><strong>Pseudo (R^2)</strong></td>
<td>0.257</td>
</tr>
</tbody>
</table>

- **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 counterfactuals
- 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).


References III

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


Additional Material
Frailty distribution in our sample

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

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>P1</th>
<th>P50</th>
<th>P99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alive</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.09</td>
<td>0</td>
<td>0.17</td>
</tr>
<tr>
<td>Dead this period</td>
<td>0.66</td>
<td>0.21</td>
<td>0.17</td>
<td>0.69</td>
<td>0.97</td>
</tr>
</tbody>
</table>
## Cause of death

<table>
<thead>
<tr>
<th>Cancer</th>
<th>Heart</th>
<th>Other Health-related</th>
<th>Non-health related</th>
<th>Death expected?</th>
<th>Death during clustering period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Expected</td>
<td>Unexpected</td>
</tr>
<tr>
<td>Type 1</td>
<td>0.49</td>
<td>0.25</td>
<td>0.23</td>
<td>0.60</td>
<td>0.40</td>
</tr>
<tr>
<td>Type 2</td>
<td>0.35</td>
<td>0.31</td>
<td>0.32</td>
<td>0.49</td>
<td>0.51</td>
</tr>
<tr>
<td>Type 3</td>
<td>0.41</td>
<td>0.21</td>
<td>0.30</td>
<td>0.47</td>
<td>0.53</td>
</tr>
<tr>
<td>Type 4</td>
<td>0.18</td>
<td>0.26</td>
<td>0.55</td>
<td>0.38</td>
<td>0.63</td>
</tr>
<tr>
<td>Type 5</td>
<td>0.28</td>
<td>0.29</td>
<td>0.37</td>
<td>0.44</td>
<td>0.56</td>
</tr>
<tr>
<td>Overall</td>
<td>0.35</td>
<td>0.27</td>
<td>0.34</td>
<td>0.48</td>
<td>0.52</td>
</tr>
</tbody>
</table>

### 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
K-means algorithm

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

\[
\left( \hat{h}(1), \ldots, \hat{h}(K), \{\hat{k}_i\}_{i=1}^N \right) = \text{argmin} \left( \tilde{h}(1), \ldots, \tilde{h}(K), \{k_i\}_{i=1}^N \right) \sum_{i=1}^N \left\| h_i - \tilde{h}(k_i) \right\|^2
\]

- \( \hat{h}(j) \) 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 \( k_i \)
Traditional machine learning methods - Elbow method

Elbow method Thorndike (1953):

- Calculate the proportion of the total variance explained by the clusters
  \[
  \omega(k) = 1 - \frac{\sum_{i=1}^{N} \| h_i - \tilde{h}(k_i) \|^2}{\sum_{i=1}^{N} \| h_i - \bar{h} \|^2}
  \]

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

- Criterion: select the number of clusters that maximizes the average silhouette of the clustering
Traditional machine learning methods - Silhouette method

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

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

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

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:

$$s(i) = \begin{cases} 
0 & |C_l| = 1 \\
\frac{b(i) - a(i)}{\max\{a(i), b(i)\}} & \text{otherwise}
\end{cases}$$
Regressions for frailty and mortality between age 52 and 60

\[ f_{it} = aX_{it} + f_{age}(t) + \sum_{\eta=1}^{k} a_{\eta}D_{i\eta} + \epsilon_{it} \]  
(1a)

\[ f_{it} = aX_{it} + f_{age}(t) + \epsilon_{it} \]  
(1b)

\[ P(D_{it}|X_{it}, \eta) = \Lambda(bX_{it} + g_{age}(t) + \sum_{\eta=1}^{k} b_{\eta}D_{i\eta}) \]  
(2a)

\[ P(D_{it}|X_{it}) = \Lambda(bX_{it} + g_{age}(t)) \]  
(2b)

\( X_{it} \): education, race, gender, HRS cohort, marital status, age

\( D_{i\eta} \): health types dummies
Absolute Mean Error

For a given number of cluster $k$

- Estimate the absolute mean error (AME)

\[
AME(k) = \frac{1}{N} \sum_{i}^{N} |y_{it} - f(x_{it}, \eta_k; \theta)|
\]

with cluster information

\[
AME = \frac{1}{N} \sum_{i}^{N} |y_{it} - f(x_{it}; \theta)|
\]

without cluster information

- Calculate $r(k)$

\[
r(k) = \frac{\sum_{i}^{N} |y_{it} - f(x_{it}, \eta_k; \theta)|}{\sum_{i}^{N} |y_{it} - f(x_{it}; \theta)|}
\]
Cross Validation: predicting over a sample not used for estimation

\[ E = \frac{1}{10} \sum_{i=1}^{10} E_i \]
Choosing the number of clusters/health types

**Figure: Frailty**

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

**Figure: Mortality**
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.
Out-of-sample frailty regressions

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

\[ f_{it} = X_{it}\beta + \epsilon_{it} \]  

\[ f_{it} = X_{it}\beta + D_{i}\eta\beta^D + \epsilon_{it} \]  

- \( X_{it} \) is a rich set of controls, and \( D_{i}\eta \) are health types dummies

- \( X_{it} \): age, \((t_i)\), age squared \((t_i^2)\), age cubed \((t_i^3)\), Educational attainment \((EA_i)\), race \((\text{race}_i)\), HRS cohort \((\text{HRS}_i)\), women and marital status \((c_{it})\) dummies

- Alternative specification: \( X_{it} \) also include Initial frailty \((f_{i52})\) and initial SRHS \((s_{i52})\).
Out-of-sample mortality regressions

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

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

(5)

\[
Pr(D_{i,t+2} = 1|X_{it}, D_{i\eta}) = \frac{e^{X_{it}\beta + D_{i\eta}\beta_D}}{1 + e^{X_{it}\beta + D_{i\eta}\beta_D}}
\]  

(6)

- \(X_{it}\) is a rich set of controls, and \(D_{i\eta}\) 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 \((s_{i52})\).
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
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
Evolution of frailty for each person

<table>
<thead>
<tr>
<th>Age</th>
<th>Health Type 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>+0.000</td>
</tr>
<tr>
<td>54</td>
<td>+0.057</td>
</tr>
<tr>
<td>56</td>
<td>+0.114</td>
</tr>
<tr>
<td>58</td>
<td>+0.171</td>
</tr>
<tr>
<td>60</td>
<td>+0.257</td>
</tr>
</tbody>
</table>

**Type 1. The vigorous resilient:** healthiest and unlikely to die (even after age 60)
Evolution of frailty for each person

**Type 2. The fair-health resilient**: less healthy but still unlikely to die (even after age 60)
Evolution of frailty for each person

Type 3. The *fair-health vulnerable*: start in fair health but fast decline
Evolution of frailty for each person

Type 4. The frale resilient: initially among the unhealthiest but resilient
Evolution of frailty for each person

Type 5. The frail vulnerable: initially unhealthy and fast decline
Frailty distribution by health types and age

**All individuals**

- Type 1: (57%)
- Type 2: (28%)
- Type 3: (2%)
- Type 4: (10%)
- Type 5: (3%)

**Only survivors**

- Type 1: (57%)
- Type 2: (28%)
- Type 3: (2%)
- Type 4: (10%)
- Type 5: (3%)

Shaded area depicts the P80-P20 interval of frailty.
# Main statistics by health type

<table>
<thead>
<tr>
<th></th>
<th>All sample</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fraction of people</strong></td>
<td>1</td>
<td>0.57</td>
<td>0.28</td>
<td>0.02</td>
<td>0.10</td>
<td>0.03</td>
</tr>
</tbody>
</table>

**Health outcomes during clustering period**

<table>
<thead>
<tr>
<th></th>
<th>All sample</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average frailty</td>
<td>0.17</td>
<td>0.06</td>
<td>0.20</td>
<td>0.43</td>
<td>0.44</td>
<td>0.77</td>
</tr>
<tr>
<td>Average health deficits</td>
<td>6.0</td>
<td>2.1</td>
<td>7.0</td>
<td>15.1</td>
<td>15.4</td>
<td>27.0</td>
</tr>
<tr>
<td>Fraction dead by 60</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.94</td>
<td>0</td>
<td>0.89</td>
</tr>
</tbody>
</table>

**Health at 52**

<table>
<thead>
<tr>
<th></th>
<th>All sample</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average frailty</td>
<td>0.13</td>
<td>0.05</td>
<td>0.17</td>
<td>0.15</td>
<td>0.40</td>
<td>0.36</td>
</tr>
<tr>
<td>Average health deficits</td>
<td>4.6</td>
<td>1.8</td>
<td>5.9</td>
<td>5.1</td>
<td>13.9</td>
<td>12.5</td>
</tr>
<tr>
<td>Average SRHS</td>
<td>2.64</td>
<td>2.12</td>
<td>3.01</td>
<td>3.15</td>
<td>4.03</td>
<td>3.95</td>
</tr>
<tr>
<td>Std. Dev. of frailty</td>
<td>0.14</td>
<td>0.04</td>
<td>0.08</td>
<td>0.12</td>
<td>0.13</td>
<td>0.23</td>
</tr>
</tbody>
</table>
### Health types and observable characteristics

<table>
<thead>
<tr>
<th>Health types and observable characteristics</th>
<th>All sample</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fraction of people</strong></td>
<td>1</td>
<td>0.57</td>
<td>0.28</td>
<td>0.02</td>
<td>0.10</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Health outcomes during clustering period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average frailty</td>
<td>0.17</td>
<td>0.06</td>
<td>0.20</td>
<td>0.43</td>
<td>0.44</td>
<td>0.77</td>
</tr>
<tr>
<td>Average health deficits</td>
<td>6.0</td>
<td>2.1</td>
<td>7.0</td>
<td>15.1</td>
<td>15.4</td>
<td>27.0</td>
</tr>
<tr>
<td>Fraction dead by 60</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.94</td>
<td>0</td>
<td>0.89</td>
</tr>
<tr>
<td><strong>Health at 52</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average frailty</td>
<td>0.13</td>
<td>0.05</td>
<td>0.17</td>
<td>0.15</td>
<td>0.40</td>
<td>0.36</td>
</tr>
<tr>
<td>Average health deficits</td>
<td>4.6</td>
<td>1.8</td>
<td>5.9</td>
<td>5.1</td>
<td>13.9</td>
<td>12.5</td>
</tr>
<tr>
<td>Average SRHS</td>
<td>2.64</td>
<td>2.12</td>
<td>3.01</td>
<td>3.15</td>
<td>4.03</td>
<td>3.95</td>
</tr>
<tr>
<td>Std. Dev. of frailty</td>
<td>0.14</td>
<td>0.04</td>
<td>0.08</td>
<td>0.12</td>
<td>0.13</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction women</td>
<td>0.63</td>
<td>0.59</td>
<td>0.69</td>
<td>0.57</td>
<td>0.73</td>
<td>0.55</td>
</tr>
<tr>
<td>Fraction black people</td>
<td>0.17</td>
<td>0.13</td>
<td>0.20</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Mean years of education</td>
<td>13.01</td>
<td>13.60</td>
<td>12.46</td>
<td>12.72</td>
<td>11.52</td>
<td>12.27</td>
</tr>
<tr>
<td>Fraction partnered at 52</td>
<td>0.78</td>
<td>0.82</td>
<td>0.77</td>
<td>0.66</td>
<td>0.64</td>
<td>0.63</td>
</tr>
<tr>
<td>Mean individual income at 52</td>
<td>30,828</td>
<td>39,303</td>
<td>24,239</td>
<td>18,177</td>
<td>10,818</td>
<td>9,941</td>
</tr>
<tr>
<td>Mean household income at 52</td>
<td>56,322</td>
<td>70,156</td>
<td>45,660</td>
<td>34,925</td>
<td>22,211</td>
<td>26,710</td>
</tr>
<tr>
<td><strong>Healthy behaviours</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction ever smoked</td>
<td>0.56</td>
<td>0.49</td>
<td>0.64</td>
<td>0.72</td>
<td>0.67</td>
<td>0.76</td>
</tr>
<tr>
<td>Fraction vigorous activity at 52</td>
<td>0.50</td>
<td>0.61</td>
<td>0.44</td>
<td>0.46</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td><strong>Health insurance status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private health insurance at 52</td>
<td>0.76</td>
<td>0.85</td>
<td>0.74</td>
<td>0.61</td>
<td>0.42</td>
<td>0.41</td>
</tr>
<tr>
<td>Public health insurance at 52</td>
<td>0.13</td>
<td>0.04</td>
<td>0.13</td>
<td>0.19</td>
<td>0.45</td>
<td>0.49</td>
</tr>
<tr>
<td>Medicaid</td>
<td>0.06</td>
<td>0.01</td>
<td>0.06</td>
<td>0.07</td>
<td>0.24</td>
<td>0.29</td>
</tr>
<tr>
<td>Medicare</td>
<td>0.06</td>
<td>0.01</td>
<td>0.06</td>
<td>0.12</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td>Uninsured at 52</td>
<td>0.14</td>
<td>0.12</td>
<td>0.16</td>
<td>0.22</td>
<td>0.20</td>
<td>0.17</td>
</tr>
<tr>
<td>Health Types</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td><strong>Initial Frailty</strong></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><strong>Healthy behaviours</strong></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><strong>Health insurance</strong></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Prob of living up to 75</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Pseudo R2: 0.133 0.434 0.032 0.451 0.147 0.456

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

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

- 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?" ⇒ including frailty composition as observable characteristics does not help.
- Frailty composition is not key in explaining health types.
Health type and observable characteristics: Frailty composition

<table>
<thead>
<tr>
<th></th>
<th>Health Types</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Initial Frailty</strong></td>
<td>x</td>
</tr>
<tr>
<td><strong>Initial Frailty composition</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Health behaviours</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Health insurance</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Pseudo R2</strong></td>
<td>0.434</td>
</tr>
</tbody>
</table>

Multinomial Regression details

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

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 (\( w_i \))

\[ X_{it} = (1, t_i, t_i^2, DHS_{it}, EA_i, c_{it}, w_i, w_i t_i, w_i t_i^2, w_i DHS_{it}, w_i EA_i, w_i c_{it}) \]
Additional Material - Not for presentation
Cluster Assignments: $K=4$ and $K=5$

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type 1</strong></td>
<td>2837</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2837</td>
</tr>
<tr>
<td><strong>Type 2</strong></td>
<td>64</td>
<td>1310</td>
<td>1</td>
<td>0</td>
<td>1375</td>
</tr>
<tr>
<td><strong>Type 3</strong></td>
<td>0</td>
<td>20</td>
<td>42</td>
<td>61</td>
<td>123</td>
</tr>
<tr>
<td><strong>Type 4</strong></td>
<td>0</td>
<td>12</td>
<td>494</td>
<td>0</td>
<td>506</td>
</tr>
<tr>
<td><strong>Type 5</strong></td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>152</td>
<td>155</td>
</tr>
<tr>
<td><strong>Column Total</strong></td>
<td>2901</td>
<td>1342</td>
<td>540</td>
<td>213</td>
<td></td>
</tr>
</tbody>
</table>
### Cluster Assignments: K=4 and K=5

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