A Biomarker Based Health Measure to Study Cut-Point Shift and Index Shift in Self-Reported Health of Aged Women

Qing Li\textsuperscript{1}  Alan A. Cohen\textsuperscript{2}  Linda P. Fried\textsuperscript{3}

2014

\textsuperscript{1}Université de Sherbrooke, Québec, Canada
\textsuperscript{2}Université de Sherbrooke, Québec, Canada
\textsuperscript{3}Columbia University, New York, USA.
Motivation

- Researchers often rely on subjective health measures
- Subjective measures could be unreliable
- Reporting style may depend on culture - difficult to compare across nations
- Importance of using *objective* health measures in health economics
  - e.g. Bound (1991); Burkhauser and Cawley (2006)
- In some situations it is compulsory to use objective measures
  - e.g. when studying reporting styles of different sub-populations
Motivation

- In this paper we propose a composite health measure based on human biomarker data
  - The health measure is objective and continuous
  - The health measure is comparable across different populations

- We use the health measure to study index shift and cut-point shift in self-reported health
  - Policy implication: for example, disability allowance is often based on self-reported health status (Kerkhofs and Lindeboom 1995)
  - So economists want to know whether "I'm in excellent health" means the same thing across different sub-populations
  - Need to control for objective health
Some existing composite measures of health

- Disability-Adjusted Life Year (World Bank, 1993)
- Disability-Adjusted Life Expectance (e.g. Murray et al., 2002)
- Health and Activity Limitation Index (e.g. Erickson, 1998)
- Meijer et al. (2012)’s internationally comparable health index (combines objective and subjective information)
  - Model health in a system of equations - comparability with Krishnakumar (2007)
  - Calculates $E(\eta|X, Y)$ where $\eta$ is unobservable true health
  - The approach works when health is dependent variable

We take a different approach
Biomarkers are objectively measurable indicators of biological states

Example: Mean Corpuscular Hemoglobin (MCH) is the amount of hemoglobin in an average red blood cell

- Too high or too low values indicate bad health;
- A main reason for too high MCH level is macrocytic anemia
- Too low level of MCH is often due to blood loss over time

There is often a reference range. Within this range, it is not easy to make further judgement with a single biomarker

The change of biomarkers with time is complicated; research on time-trajectories are ongoing

Global profile of multiple biomarkers do give general information on health
Biomarker data have not been widely used in economic studies
- Medical databases often lack socio-economic information
- Household surveys often have very limited health indicators

People have noticed the importance of using biomarker data in various research disciplines (OECD 2011)

In social surveys people started to collect biomarker data
- China Health and Nutrition Survey

Some medical databases do have (limited) socio-economic information
- National Health and Nutrition Examination Survey
- Women’s Health and Aging Study
- Quebec Longitudinal Study on Nutrition and Aging
Health Measure

- Health is multi-dimensional
- Each biomarker measures certain sides of health (with overlaps)
- We consider each individual as a point in a multi-dimensional "biomarker space"
- We measure an individual’s health with multivariate distance between the individual and the centroid of a reference group
  - The reference group is younger and healthier
  - The centroid represents a "perfect" state of health (will come back)
  - There is "one way" to be perfect, but there are millions of ways to go wrong.
Health Measure

- Features of biomarker data
  - Continuous and positive
  - Not necessarily normal
  - Correlation among different biomarkers

**We use the Mahalanobis distance to measure objective health**

- The Mahalanobis distance adjusts for correlation among variables, compared with the Euclidean distance
- When the variables are orthogonal to each other, the Mahalanobis distance reduces to the Euclidean distance
Let $x_i \in \mathbb{R}^p$ and $x_j \in \mathbb{R}^p$ be two points in a p-dimensional biomarker space.

A distance $D(x_i, x_j)$ is a metric which defines a mapping $D : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ over the biomarker space $\mathcal{X} \subseteq \mathbb{R}^p$ for $x_i, x_j \in \mathcal{X}$.

Assuming $x_i, x_j \sim N(\mu, \Sigma)$, the Mahalanobis distance between $x_i$ and $x_j$ is $D_M(x_i, x_j) := \| T(x_i) - T(x_j) \|$

- $T(x) = \Sigma^{-\frac{1}{2}}(x - \mu)$
- $\| \cdot \|$ the Euclidean norm

The $D_M$ is a composition of the Euclidean distance and the transformation $T(x)$ which transforms the variables into standard normal distribution.

Nonetheless, existence and uniqueness of the Mahalanobis distance still hold under non-normality (with proper transformation).
The Mahalanobis distance in 2-D case
Intuitive example of the Mahalanobis distance in 2-D case

- We see A) people with 2m and 130kg
- We see B) people with 1.5m and 40kg
- We *rarely* see C) people with 2m and 40kg or D) 1.5m and 130 kg
- A), B), C) and D) have similar Euclidean distance to the centroid
- C) and D) have much bigger Mahalanobis distance than A) and B)
Cut-point shift and index shift of self-reported health

We are interested in knowing whether people from different sub-population (race, gender, education etc.) report health in different manners. For example, when a disabled person and an undisabled person both say "I’m in bad health", does it mean the same thing? If there is a difference, whether the reporting manner shifts parallely for all health categories (index shift) or shifts differently for different health categories (cut-point shift)

Need to control for objective health
We take the approach of Lindeboom and van Doorslaer (Journal of Health Economics 2004)

- \( H^S = f_1(H^*, X_1, \varepsilon_1; \beta_1) \) subjective health \( H^S \) depends on true health \( H^* \) and the covariates \( X_1 \)

- \( H^* = f_2(H^o, X_2, \varepsilon_2; \beta_2) \) true health \( H^* \) depends on objectively measured health \( H^o \) and the covariates \( X_2 \)
  - If \( H^o \) were a perfect measure, there would be no need for \( X_2 \)
  - But this is almost impossible, so \( H^o \) should be viewed as a conditioning variable
Application

- Usually $H^S$ is ordered response (ordered logit/probit model)
- $H^S = i \Leftrightarrow c_{i-1} < H^* < c_i, \quad i = 1, \ldots, n$
- $X_1$ affects reporting behavior, e.g. $c_i = g_i(X_1, \beta_{1i})$
- $X_2$ affects true health, e.g. $H^* = f(H^o, \alpha) + X'_2 \beta_2 + \varepsilon_2$
- Hence
  \[ H^S = i \Leftrightarrow g_{i-1}(X_1, \beta_{1i-1}) - X'_2 \beta_2 < f(H^o, \alpha) + \varepsilon_2 < g_i(X_1, \beta_{1i}) - X'_2 \beta_2 \]
Application

- **Difficulty:** How to distinguish between $X_1$ and $X_2$?

- **Strategy**
  - stratify the sample according to the variables in $X_1$ and $X_2$
  - $H^S_k = i$ (for stratified subsample $k$)
  - or equivalently $\delta_{i-1}^k < f(H_0, \alpha^k) + \varepsilon^k_2 < \delta_i^k$
  - recall the restricted model $g_{i-1}(X_1, \beta_{1i-1}) < f(H_0, \alpha) + X_2'\beta_2 + \varepsilon_2 < g_i(X_1, \beta_{1i})$
  - stratification gets rid of the necessity to distinguish between $X_1$ and $X_2$
  - Then we can test for the difference between restricted and unrestricted model
• LR test for index shift and cut-point shift

• First step
  • $\chi^2$ test for overall shift $-2^*(\Lambda^R - \Lambda^U)$
  • Difference between the likelihood of the restricted model and (sum of) the likelihood of the unrestricted models
  • If the null is rejected, we can further test whether the overall shift is due to cut-point shift or index shift
  • Under the null, things are more ambiguous and we cannot proceed

• Second step
  • Estimate a model with the indexes of the restricted model, but leaving the cut-points free
  • $\chi^2$ test for cut-point shift $-2^*(\Lambda^R - \Lambda^{CP})$
  • $\chi^2$ test for index shift $-2^*(\Lambda^{CP} - \Lambda^U)$
A remark

- We could control for some covariates in sub-sample regressions even after stratifying the samples with respect to such variables.
- For example, after stratifying the whole sample with respect to income (into several income groups), we could still control for income for each individual in the sub-sample regressions.
- This is because the objective health measure cannot be perfect in reality.
- This is not necessary, and not key to the results.
- Nevertheless, the coefficients on such regressors should be explained with caution (two effects).
Women’s Health and Aging Study (WHAS)
- 1438 community-dwelling women aged above 65 from Baltimore City and Baltimore County
- mainly middle to upper-class
- detailed biomarker information
- information on self-reported health
- information on disability, education, race, marital status

The sample can be viewed as a (stratified) sub-sample
National Health and Nutrition Examination Survey (NHANES)
- Cross-sectional study based on representative samples of the US population
- Conducted in various waves since the 1970s
- Information on hundreds of biomarkers from all ages
- The data are publicly available
- Females aged between 20 and 40 in NHANES as reference group for the aged women in WHAS
- Results not sensitive to the above thresholds
- We use the biomarkers available in both datasets that allows a final sample size of 1,000 aged women in WHAS
- This results in 41 biomarkers
- For each woman in the study group (WHAS), we calculate the Mahalanobis distance with respect to the centroid of the reference group (NHANES)
### Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>$H^S$</th>
<th>$H^O_N$</th>
<th>Disabled</th>
<th>White</th>
<th>Edu</th>
<th>Married</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.080</td>
<td>10.77</td>
<td>0.63</td>
<td>0.75</td>
<td>10.72</td>
<td>0.28</td>
</tr>
<tr>
<td>Sd</td>
<td>1.095</td>
<td>6.033</td>
<td>0.484</td>
<td>0.432</td>
<td>3.807</td>
<td>0.451</td>
</tr>
<tr>
<td>Min</td>
<td>1</td>
<td>6.37</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>5</td>
<td>191.76</td>
<td>1</td>
<td>1</td>
<td>18</td>
<td>1</td>
</tr>
</tbody>
</table>

$H^S$: self-reported health (from 1 "Excellent" to 5 "Very bad")

$H^O_N$: objective health with NHANES as reference group (bigger value indicates worse health)
Regression results from the restricted model

<table>
<thead>
<tr>
<th>$H_N^O$</th>
<th>$(H_N^O)^2$</th>
<th>Disabled</th>
<th>White</th>
<th>Edu</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>-0.02</td>
<td>1.62</td>
<td>-0.56</td>
<td>-0.09</td>
</tr>
<tr>
<td>(0.071)</td>
<td>(0.008)</td>
<td>(0.139)</td>
<td>(0.141)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Married</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0.33</td>
<td>-3.13</td>
<td>-1.39</td>
<td>0.41</td>
<td>2.23</td>
</tr>
<tr>
<td>(0.130)</td>
<td>(0.266)</td>
<td>(0.246)</td>
<td>(0.244)</td>
<td>(0.254)</td>
</tr>
</tbody>
</table>

The coefficients of the covariates should be interpreted with caution
They include "true health effects" and "reporting habits"
Application

- Index shift and cut-point shift

<table>
<thead>
<tr>
<th>Disability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood disabled</td>
<td>-869.4</td>
</tr>
<tr>
<td>Log likelihood undisabled</td>
<td>-464.23</td>
</tr>
<tr>
<td>Log likelihood sum (unrestricted model)</td>
<td>-1333.63</td>
</tr>
<tr>
<td>Log likelihood restricted model</td>
<td>-1349.749</td>
</tr>
<tr>
<td>test for overall shift</td>
<td>32.238</td>
</tr>
<tr>
<td>P-value (8 d.f.)</td>
<td>0</td>
</tr>
<tr>
<td>test for cut-point shift</td>
<td>15.751</td>
</tr>
<tr>
<td>P-value (3 d.f.)</td>
<td>0.001</td>
</tr>
<tr>
<td>test for index shift</td>
<td>16.487</td>
</tr>
<tr>
<td>P-value (5 d.f.)</td>
<td>0.006</td>
</tr>
</tbody>
</table>
**Application**

- Index shift and cut-point shift

<table>
<thead>
<tr>
<th>Education</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood low education</td>
<td>-541.617</td>
</tr>
<tr>
<td>Log likelihood high education</td>
<td>-796.937</td>
</tr>
<tr>
<td>Log likelihood sum (unrestricted model)</td>
<td>-1338.553</td>
</tr>
<tr>
<td>Log likelihood restricted model</td>
<td>-1349.749</td>
</tr>
<tr>
<td>test for overall shift</td>
<td>22.391</td>
</tr>
<tr>
<td>P-value (10 d.f.)</td>
<td>0.013</td>
</tr>
<tr>
<td>test for cut-point shift</td>
<td>5.338</td>
</tr>
<tr>
<td>P-value (4 d.f.)</td>
<td>0.254</td>
</tr>
<tr>
<td>test for index shift</td>
<td>17.053</td>
</tr>
<tr>
<td>P-value (6 d.f.)</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Li et al. (2014)
Index shift and cut-point shift

<table>
<thead>
<tr>
<th>Marital Status</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood married</td>
<td>-347.934</td>
</tr>
<tr>
<td>Log likelihood unmarried</td>
<td>-990.589</td>
</tr>
<tr>
<td>Log likelihood sum (unrestricted model)</td>
<td>-1338.523</td>
</tr>
<tr>
<td>Log likelihood restricted model</td>
<td>-1349.749</td>
</tr>
<tr>
<td>test for overall shift</td>
<td>22.451</td>
</tr>
<tr>
<td>P-value (8 d.f.)</td>
<td>0.004</td>
</tr>
<tr>
<td>test for cut-point shift</td>
<td>1.724</td>
</tr>
<tr>
<td>P-value (3 d.f.)</td>
<td>0.632</td>
</tr>
<tr>
<td>test for index shift</td>
<td>20.727</td>
</tr>
<tr>
<td>P-value (5 d.f.)</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Index shift and cut-point shift

White

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood Caucasian</td>
<td>-1008.187</td>
</tr>
<tr>
<td>Log likelihood other races</td>
<td>-339.633</td>
</tr>
<tr>
<td>Log likelihood sum (unrestricted model)</td>
<td>-1347.821</td>
</tr>
<tr>
<td>Log likelihood restricted model</td>
<td>-1349.749</td>
</tr>
<tr>
<td>test for overall shift</td>
<td>3.856</td>
</tr>
<tr>
<td>P-value (8 d.f.)</td>
<td>0.87</td>
</tr>
<tr>
<td>test for cut-point shift</td>
<td>2.175</td>
</tr>
<tr>
<td>P-value (3 d.f.)</td>
<td>0.537</td>
</tr>
<tr>
<td>test for index shift</td>
<td>1.681</td>
</tr>
<tr>
<td>P-value (5 d.f.)</td>
<td>0.891</td>
</tr>
</tbody>
</table>

Note: When the null cannot be rejected in the test for overall shift, the tests for cut-point shift and index shift are meaningless.
Issues and Discussions

- Why not just use the biomarkers as regressors?
  - Trajectories of single biomarkers are complex and difficult to capture
  - Linear or quadratic specifications not sufficient
  - Higher order polynomial or other non-parametric specifications require large sample size
  - But even with large samples there are problems like multicollinearity, curse of dimensionality, etc.
  - So, we search for "higher order signal" rather than signal from raw biomarkers
Issues and Discussions

- How many biomarkers are necessary for a signal?
- It is not always the case to have many biomarkers?
- Is it possible to get a (even weak) signal from fewer biomarkers?
- The figure below shows results with different number of biomarkers
  - Random choice
  - Alphabetical choice (adding one biomarker each time)
Issues and Discussions

- Linear coefficient (random choice)
- Quadratic coefficient (random choice)
- Linear coefficient (alphabetical choice)
- Quadratic coefficient (alphabetical choice)
Issues and Discussions

- Choice of reference group and centroid
- There may be different points in the biomarker space representing good health
  - Results not sensitive (qualitatively) to changes in reference group
  - The measure should be viewed as error-ridden proxy for true health in any case
- There may be natural adaptations in physiological systems when people age, so part of the distance may not be bad
  - We use WHAS as its own reference group in order to compare
  - No qualitative change
Recall the summary statistics

<table>
<thead>
<tr>
<th></th>
<th>$H^S$</th>
<th>$H^O_N$</th>
<th>$H^O_W$</th>
<th>Disabled</th>
<th>White</th>
<th>Edu</th>
<th>Married</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.080</td>
<td>10.77</td>
<td><strong>6.20</strong></td>
<td>0.63</td>
<td>0.75</td>
<td>10.72</td>
<td>0.28</td>
</tr>
<tr>
<td>Sd</td>
<td>1.095</td>
<td>6.033</td>
<td><strong>1.589</strong></td>
<td>0.484</td>
<td>0.432</td>
<td>3.807</td>
<td>0.451</td>
</tr>
<tr>
<td>Min</td>
<td>1</td>
<td>6.37</td>
<td><strong>2.61</strong></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>5</td>
<td>191.76</td>
<td><strong>31.33</strong></td>
<td>1</td>
<td>1</td>
<td>18</td>
<td>1</td>
</tr>
</tbody>
</table>

$H^S$: self-reported health

$H^O_N$: objective health with NHANES as reference group

$H^O_W$: **objective health with WHAS as its own reference group**
Issues and Discussions

- Which biomarkers provide "stronger signal"?
- How good is the signal with less biomarkers?
- The figure below shows the distribution of the estimated linear coefficient with all possibilities of choosing 5 biomarkers (from the 41 biomarkers)
  - \(C_{41}^5 = 749,398\) possibilities of choosing 5 markers
  - 749,398 regressions to obtain the distribution
Issues and Discussions

Distributions and Means

linear coefficients
Are there biomarkers that provide "stronger signal"?

Percentages of some biomarkers appearing in the five-marker suites that lead to "the 500, 2500 and 5000 regressions with strongest signals"

<table>
<thead>
<tr>
<th></th>
<th>N 500</th>
<th>N 2500</th>
<th>N 5000</th>
<th>W 500</th>
<th>W 2500</th>
<th>W 5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALT</td>
<td>0.088</td>
<td>0.106</td>
<td>0.111</td>
<td>0.250</td>
<td>0.231</td>
<td>0.215</td>
</tr>
<tr>
<td>AST</td>
<td>0.168</td>
<td>0.165</td>
<td>0.150</td>
<td>0.292</td>
<td>0.267</td>
<td>0.257</td>
</tr>
<tr>
<td>Blood urea nitrogen</td>
<td>0.334</td>
<td>0.352</td>
<td>0.326</td>
<td>0.390</td>
<td>0.318</td>
<td>0.288</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>0.030</td>
<td>0.038</td>
<td>0.046</td>
<td>0.158</td>
<td>0.192</td>
<td>0.186</td>
</tr>
<tr>
<td>Chloride</td>
<td>0.308</td>
<td>0.286</td>
<td>0.268</td>
<td>0.682</td>
<td>0.511</td>
<td>0.440</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>0.064</td>
<td>0.105</td>
<td>0.115</td>
<td>0.184</td>
<td>0.193</td>
<td>0.198</td>
</tr>
<tr>
<td>Glucose</td>
<td>1.000</td>
<td>0.998</td>
<td>0.995</td>
<td>0.972</td>
<td>0.927</td>
<td>0.888</td>
</tr>
<tr>
<td>Uric acid</td>
<td>0.700</td>
<td>0.585</td>
<td>0.528</td>
<td>0.068</td>
<td>0.089</td>
<td>0.104</td>
</tr>
</tbody>
</table>
Conclusion

- We provide a composite measure of objective health based on biomarker data.
- We measure health with the Mahalanobis distance to the centroid of a reference group.
- It is possible to get a signal even with a small number of biomarkers.
- The measure is internationally comparable if the populations are not too different racially or physiologically.
- We apply this measure to study cut-point shift and index shift in self-reported health of aged women.
- We find evidence of both cut-point shift and index shift for disability, index shift for education and marital status, and no shift for race.
Thank you very much for your attention!

Questions and comments are welcome!