HealthPaths Dynamics II: Using Functional Health Trajectories to Quantify Impacts on Health-Adjusted Life Expectancy (HALE) in Canada

- core concepts – life course trajectories, functional health
- estimating multiple co-evolving dynamic relationships
- drawing out implications using microsimulation

Geoff Rowe & Michael Wolfson, uOttawa
REVES: Austin, May 2013
General Plan of Analysis

- use one major longitudinal data set – Statistics Canada’s National Population Health Survey
- characterize statistically multiple co-evolving individual health and health-related characteristics
- incorporate all estimated statistical relationships into a computerized microsimulation model
- simulate health-adjusted life expectancy (HALE)
- attribute $\Delta$HALE to selected health determinants via comparative simulations
- today: progress towards HealthPaths II
Statistics Canada’s National Population Health Survey (NPHS)

- developed and fielded by Canada’s national statistical bureau
- started in 1994; interviews every 2 years; includes institutionalized; includes mortality follow-up
- $n = \sim 20,000$ individuals initially; now $\sim 14,000$
- all responses self-report
- mostly conventional health survey content, e.g. socio-demographics, chronic disease check list, major risk factors, health care utilization
- some content more exploratory, e.g. Antonovsky’s Sense of Coherence, McMaster Health Utilities Index (HUI)
Focus of Analysis – Functional Health

- using NPHS Health Utility Index (HUI): a generic index of functional health status.
  - 1 ⇒ full health
  - 0 ⇒ as good as dead
  - < 0 ⇒ worse than dead

- based on eight separately assessed attributes: vision, hearing, speech, mobility, dexterity, cognition, emotion, and pain

- aggregated into a summary numerical index based on an empirical “weighting function”
Focus of Analysis: Health-Adjusted Life Expectancy

- extension of widely used concept of life expectancy (LE)
- combine length of life with “healthiness” of life, or “capacity to function” while alive, using HUI
- original approach – Sullivan method
- but here – complete lifecycle trajectories, using microsimulation
Basic Definitions

- LE = area under survival curve
- HALE = “weighted” area under survival curve
  - where “weights” are levels of individual health status, ranging between zero (dead) and one (fully healthy)
Risk factors & events included

**Ordinal Variables**
- Vision
- Hearing
- Speech
- Mobility
- Dexterity
- Emotion
- Cognition
- Pain
- Income Deciles
- Leisure Activity
- Daily Activity
- **Smoking Status**

**Binary Variables**
- Employed this Year
- Family Member
- Institutional Resident
- High School Graduation
- Community College
- University Graduation
- Mortality

**Quantitative Variables**
- Body Mass Index
- Sense of Mastery
- Sense of Coherence
- Years of Daily Smoking

Functional Health summarized via HUI
Modeling Health & Risk Factor transitions

Transitions from the 'current' level of a health variable or risk factor to a higher or a lower level.

Separate binary or ordinal logistic regressions for each row of each transition matrix.

Each health variable or risk factor plays both the role of dependent and of independent variable (everything effects everything over time).
Dynamic Structure of Equations

Lagged States

\[
\begin{bmatrix}
Y_i \\
\vdots
\end{bmatrix}
\]

\[t-4\]

Current States

\[
\begin{bmatrix}
Y_1 \\
\vdots
\end{bmatrix}
\]

\[t-2\]

Transition States

\[
\begin{bmatrix}
\vdots \\
Y_{18} \\
\vdots
\end{bmatrix}
\]

\[t\]

18 Binary & Ordinal Variables plus Mortality

3 Quantitative Variables

\[
\begin{bmatrix}
X_1 \\
\vdots \\
X_3
\end{bmatrix}
\]

\[t-4\]

\[t-2\]
## Covariates in Binary/Ordinal equations

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Logistic Equation Terms</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immigrant Status</td>
<td>Immigrant &amp; Non-European Immigrant</td>
<td>2</td>
</tr>
<tr>
<td>Institutional Status</td>
<td>Non-Institutional, Institutionalized</td>
<td>2</td>
</tr>
<tr>
<td>Vision</td>
<td>5 levels of health deficit</td>
<td>10</td>
</tr>
<tr>
<td>Hearing</td>
<td>5 levels of health deficit</td>
<td>10</td>
</tr>
<tr>
<td>Speech</td>
<td>4 levels of health deficit</td>
<td>8</td>
</tr>
<tr>
<td>Mobility</td>
<td>5 levels of health deficit</td>
<td>10</td>
</tr>
<tr>
<td>Dexterity</td>
<td>5 levels of health deficit</td>
<td>10</td>
</tr>
<tr>
<td>Emotion</td>
<td>4 levels of health deficit</td>
<td>8</td>
</tr>
<tr>
<td>Cognition</td>
<td>5 levels of health deficit</td>
<td>10</td>
</tr>
<tr>
<td>Pain</td>
<td>4 levels of health deficit</td>
<td>8</td>
</tr>
<tr>
<td>Education</td>
<td>Secondary School, College, &amp; University</td>
<td>6</td>
</tr>
<tr>
<td>Leisure Activity</td>
<td>Moderate &amp; Active</td>
<td>4</td>
</tr>
<tr>
<td>Daily Activity</td>
<td>Walking, Light Work, &amp; Heavy Work</td>
<td>6</td>
</tr>
<tr>
<td>Smoking Status</td>
<td>Occasional, Former Daily, &amp; Daily</td>
<td>6</td>
</tr>
<tr>
<td>Years of Smoking</td>
<td>linear slope &amp; spline coefficient</td>
<td>4</td>
</tr>
<tr>
<td>Employment</td>
<td>Employed in the past 12 months</td>
<td>2</td>
</tr>
<tr>
<td>Family Membership</td>
<td>Family Member/Non-Member</td>
<td>2</td>
</tr>
<tr>
<td>Household Income</td>
<td>Household Income Deciles</td>
<td>18</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td>linear slope &amp; spline coefficient</td>
<td>4</td>
</tr>
<tr>
<td>Sense of Mastery</td>
<td>28 increments</td>
<td>56</td>
</tr>
<tr>
<td>Sense of Coherence</td>
<td>linear slope &amp; spline coefficient</td>
<td>4</td>
</tr>
<tr>
<td>Time Interval</td>
<td>slope ( log(ΔTime) ) = 1</td>
<td>0</td>
</tr>
</tbody>
</table>

Total 190
Age-varying estimates

Parameters are estimated using separate local weights for each target age from 20 to 100.

If respondent age = target age:
   local weight = survey weight

If respondent age ≠ target age:
   local weight = survey weight  \times  f( (age-target)^2 )

Local weights shrink relative to survey weights. After age 75, the rate of shrinkage increases to compensate for sparse data.
Estimating large/complex equations

Model selection using penalized logistic regression: 

**elastic net** - a compromise between

**ridge regression** and the **lasso** minimizing:

\[-2 \text{loglikelihood} / N + \lambda \sum [ \alpha |\beta| + (1- \alpha) \beta^2 ]\]

\(\lambda\) is chosen by cross-validation with \(\alpha = 0.5\)

Zhou & Hastie (2005), Friedman, et al. (2010)

R package **glmnet**
HealthPaths II: a detailed model => a large model

2       Sexes
81      Ages
40      Bootstrap samples
190+    Coefficients per equation
22      Binary/ordinal dependent variables
2+      Transition matrix rows

=> more than 46 million coefficient estimates, or about 7,000 age-sex profiles of odds ratios like the “Age-specific Mobility effects on Mortality“
Age-specific Mobility effect on Mortality

Women: Lagged Mobility Effect

Men: Lagged Mobility Effect

Women: Current Mobility Effect

Men: Current Mobility Effect
Average coefficient effect sizes: strong effects are rare

Medium & strong coefficients tend to reflect 'auto-persistence': i.e. lagged logistic regression effects of $Y_{i,t-2}$ and $Y_{i,t-4}$ on $Y_{i,t}$
Validation: Mortality Hazards
NPHS Rates & Simulated

Mortality Hazards (Simulated & NPHS): Baseline Scenario

Women: Simulated

Men: Simulated

NPHS Rates
Validation: Average HUI
NPHS & Simulated

Average HUI (Simulated & NPHS): Baseline

Women: Simulated
NPHS Averages

NPHS Cohorts born: 1900-1976

Men: Simulated
Simulated: 1976 cohort

Age =>
Baseline LEs and Dynamic & Sullivan HALEs

* 1975 birth cohort projections (CPP actuaries)

LE: life expectancy at 20

Dynamic HALE

Sullivan HALE

Women

Men
Constructing “what-if” scenarios

**Baseline:** everything influences everything else

- $\text{Vision}$
- $\text{Emotion}$
- $\text{Vision}$
- $\text{Emotion}$
- $\text{Vision}$
- $\text{Emotion}$

[Diagram showing connections between Vision and Emotion at time steps $t-4$, $t-2$, and $t$.]
Cutting pathways to selected variables

Perfect Vision Scenario: over-ride transitions, and assign a 'Perfect' score at each step

\[
\begin{bmatrix}
\text{Vision} \\
\text{Emotion}
\end{bmatrix}_{t-4} \rightarrow
\begin{bmatrix}
\text{Perfect} \\
\text{Emotion}
\end{bmatrix}_t
\]

\[
\begin{bmatrix}
\text{Vision} \\
\text{Emotion}
\end{bmatrix}_{t-2} \rightarrow
\begin{bmatrix}
\text{Vision} \\
\text{Emotion}
\end{bmatrix}_{t-4}
\]
Assessing effect size

\[ \text{HALE}_{\text{perfect vision}} - \text{HALE}_{\text{baseline}} \approx \text{healthy years lost due to imperfect vision} \]

Sensitive to:

- Initial states (at age 20)
- Subsequent frequency of transitions
Functional Health Effect Sizes: scenario minus baseline at age 20

Effects ordered by medians

Low

High

LE: Women

HALE: Women

LE: Men

HALE: Men
Composite Risk Factors I

- **Socio-Economic Status** = Education + Income
- **Physical Activity** = Leisure + Daily Non-leisure
- **Coping Skills** = Sense of Coherence + Sense of Mastery

Composite scenarios fix a set of variables at 'optimal' scores (eg. everyone is a university graduate with income in the top decile)
Risk Factor Effect Sizes: scenario minus baseline at age 20

**Coping Skills: high median & high missing value imputation variance**
Composite Risk Factors II

- **Physical Function** = Physical Activity + Mobility + Dexterity

- **Mental Condition** = Coping Skills + Emotion + Cognition

- **Sensory Function** = Vision + Hearing + Speech + Pain
Grouped Risk Factor Effect Sizes: both sexes, ranked sets

LE

Family Membership
Sensory Function
Smoking Status
Employment
Mental Condition
BMI
Socio-Economic Status
Physical Function

Low Rank
Mid-High

Years

HALE

Family Membership
BMI
Smoking Status
Employment
Socio-Economic Status
Physical Function
Mental Condition
Sensory Function

Low Rank
Mid
High

Weighted Years
Concluding Comments

Effect Sizes: many changes/improvements in HealthPaths II

- HALE rankings markedly different from LE rankings
- **Sensory Function** tops the list for HALE
- **Smoking** effect weak in this cohort because there are many life-long non-smokers (the war is won!)
- **BMI & Hearing** effects point to need for more complex counterfactuals: healthy ageing can include change

Effect Variance

Multiple sources of variability: sampling, non-response (missing value imputation), response error, variable choice, equation specification, coefficient estimation, ...

Unambiguous ranking of risk factor impacts on HALE is not yet possible given the limitations of data as it is generally available.