CAUSE-SPECIFIC SENESCENCE:
CLASSIFYING CAUSES OF DEATH ACCORDING TO THE RATE OF AGING

Carlo Giovanni Camarda
Marketa Pechholdova

INED, Paris, October 2014
Motivation

- Current revision of ICD contains around 12,000 items
Motivation

- Current revision of ICD contains around 12,000 items
- There is need for convenient cause-of-death (COD) shortlist:
  - small in size
  - explicative in its design
Motivation

- Current revision of ICD contains around 12,000 items
- There is need for convenient cause-of-death (COD) shortlist:
  - small in size
  - explicative in its design
- Previous approaches:
  - reclassification of CODs based on their etiology (Meslé, 1999)
  - analytic shortlist from comparability study (Pechholdová, 2012)
  - relationship between aging and COD (Horiuchi et al., 2003)
  - regional similarities of age-cause-specific profile (Brouard and Lopez, 1985; Meslé and Vallin, 2002)
Motivation

- Current revision of ICD contains around 12,000 items
- There is need for convenient cause-of-death (COD) shortlist:
  - small in size
  - explicative in its design
- Previous approaches:
  - reclassification of CODs based on their etiology (Meslé, 1999)
  - analytic shortlist from comparability study (Pechholdová, 2012)
  - relationship between aging and COD (Horiuchi et al., 2003)
  - regional similarities of age-cause-specific profile (Brouard and Lopez, 1985; Meslé and Vallin, 2002)
- We carry out a statistically-oriented reclassification of CODs:
  - based on non-parametric estimation of rate of aging
  - cluster analysis with optimized number of clusters
Motivation

- Current revision of ICD contains around 12,000 items
- There is need for convenient cause-of-death (COD) shortlist:
  - small in size
  - explicative in its design
- Previous approaches:
  - reclassification of CODs based on their etiology (Meslé, 1999)
  - analytic shortlist from comparability study (Pechholdová, 2012)
  - relationship between aging and COD (Horiuchi et al., 2003)
  - regional similarities of age-cause-specific profile (Brouard and Lopez, 1985; Meslé and Vallin, 2002)
- We carry out a statistically-oriented reclassification of CODs:
  - based on non-parametric estimation of rate of aging
  - cluster analysis with optimized number of clusters
- We propose evidence-based COD shortlists
- We isolate cause-specific features in overall mortality
Data

- Individual-level data on deaths in the Czech Republic
Data

- Individual-level data on deaths in the Czech Republic

- We compute death counts:
  - over years 1998-2011
  - by single ages 30-100
  - at the ICD10 3-digit level
  - here for males only
Data

- Individual-level data on deaths in the Czech Republic

- We compute death counts:
  - over years 1998-2011
  - by single ages 30-100
  - at the ICD10 3-digit level
  - here for males only

- We take correspondent exposures from the HMD
Data

- Individual-level data on deaths in the Czech Republic
- We compute death counts:
  - over years 1998-2011
  - by single ages 30-100
  - at the ICD10 3-digit level
  - here for males only
- We take correspondent exposures from the HMD
- We select CODs with 10+ available data-points over ages
Data

- Individual-level data on deaths in the Czech Republic

- We compute death counts:
  - over years 1998-2011
  - by single ages 30-100
  - at the ICD10 3-digit level
  - here for males only

- We take correspondent exposures from the HMD

- We select CODs with 10+ available data-points over ages

- Final datasets for deaths and exposures:
  \[ D = (d_{i,c}) \quad \text{and} \quad E = (e_{i,c}) = e \mathbf{1}_{1,n} \]
  with \( m = 71 \) ages for \( n = 531 \) CODs
Modelling actual data

- Our aim is to classify CODs according to their rate-of-aging:

\[ r(x) = \frac{\partial \mu^c(x)}{\partial x} = \frac{\partial \ln(\mu^c(x))}{\partial x} \]

where \( \mu^c(x) \) is the cause-specific force of mortality
Modelling actual data

- Our aim is to classify CODs according to their rate-of-aging:

\[ r(x) = \frac{\partial \mu^c(x)}{\partial x} = \frac{\partial \ln(\mu^c(x))}{\partial x} \]

where \( \mu^c(x) \) is the cause-specific force of mortality

- We model data nonparametrically:

\[ \ln(\hat{\mu}^c) = B^q \hat{\beta}^c. \]

where \( B^q \) is a matrix of \( r \) B-splines of degree \( q \) equally spaced by a distance \( h \). \( \hat{\beta}^c \) are cause-specific penalized coefficients
Modelling actual data

- Our aim is to classify CODs according to their rate-of-aging:

  \[ r(x) = \frac{\partial \mu^c(x)}{\partial x} = \frac{\partial \ln(\mu^c(x))}{\partial x} \]

  where \( \mu^c(x) \) is the cause-specific force of mortality

- We model data nonparametrically:

  \[ \ln(\hat{\mu}^c) = B^q \hat{\beta}^c. \]

  where \( B^q \) is a matrix of \( r \) \( B \)-splines of degree \( q \) equally spaced by a distance \( h \). \( \hat{\beta}^c \) are cause-specific penalized coefficients

- Advantages:
  - not influenced by noise
  - not rigid structure
  - reduction of dimensionality: \( (m = 71) \rightarrow (r = 17) \)
  - easy computation of (instantaneous) relative derivatives
Smooth data

Actual and smooth death rates in log-scale for 20 CODs.
Czech Republic, 1998-2011, ages 30-100, males.
Estimating rate-of-aging

- To compute cause-specific rate-of-aging we can either
  1. derive the linear combination of $B$-splines
  2. use difference operator of the coefficients

$$r^c(x) = C \hat{\beta}^c = \frac{1}{h} B^{q-1} \hat{\alpha}^c$$

where $C$ is a matrix incorporating first order difference of $\hat{\beta}^c$ and $\hat{\alpha}^c$ denotes the first difference of $\hat{\beta}^c$
Estimating rate-of-aging

- To compute cause-specific rate-of-aging we can either
  1. derive the linear combination of $B$-splines
  2. use difference operator of the coefficients

$$r^c(x) = C \hat{\beta}^c = \frac{1}{h} B^{q-1} \hat{\alpha}^c$$

where $C$ is a matrix incorporating first order difference of $\hat{\beta}^c$ and $\hat{\alpha}^c$ denotes the first difference of $\hat{\beta}^c$

- We obtain identical $r^c(x)$ either using the estimated coefficients or applying difference operator on them
Estimating rate-of-aging

- To compute cause-specific rate-of-aging we can either
  1. derive the linear combination of B-splines
  2. use difference operator of the coefficients

\[ r_c(x) = C \hat{\beta}_c = \frac{1}{h} B^{q-1} \hat{\alpha}_c \]

where \( C \) is a matrix incorporating first order difference of \( \hat{\beta}_c \) and \( \hat{\alpha}_c \) denotes the first difference of \( \hat{\beta}_c \)

- We obtain identical \( r_c(x) \) either using the estimated coefficients or applying difference operator on them

- This allows us to cluster directly of the first difference of estimated coefficients \( \hat{\alpha}_c \)
  - without loosing smoothness behavior of age-patterns
  - keeping relationship between \( r(x) \) for each COD and the associated age profiles
Smooth derivatives

Actual and smooth relative derivatives of \( \hat{\mu}^c \) for 20 CODs.
Czech Republic, 1998-2011, ages 30-100, males.
Smooth derivatives

Estimated rate-of-aging, $r(x)$, for 531 CODs.
Czech Republic, 1998-2011, ages 30-100, males.
Clustering cause-specific rate-of-aging

- We aim to classify all \( n \) CODs by their \( r^c(x) \) which we reduced to a vector of lagged-coefficients for each COD \( c \).

\[
[\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^c, \ldots, \hat{\alpha}^n]
\]
Clustering cause-specific rate-of-aging

- We aim to classify all \( n \) CODs by their \( r^c(x) \) which we reduced to a vector of lagged-coefficients for each COD \( c \).

\[
[\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^c, \ldots, \hat{\alpha}^n]
\]

- We use a \( k \)-means clustering which allows us to
  - cluster our observations
  - extract a center for each cluster, i.e. prototypes for the different rates-of-aging
Clustering cause-specific rate-of-aging

- We aim to classify all \( n \) CODs by their \( r^c(x) \) which we reduced to a vector of lagged-coefficients for each COD \( c \).

\[
[\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^c, \ldots, \hat{\alpha}^n]
\]

- We use a \( k \)-means clustering which allows us to
  - cluster our observations
  - extract a center for each cluster, i.e. prototypes for the different rates-of-aging

- We aim to partition the \( n = 531 \) observations into \( k \) sets
  \( S = \{S_1, S_2, \ldots, S_k\} \) such that

\[
\arg \min_S \sum_{i=1}^k \sum_{\alpha \in S_i} \| \alpha - \nu_i \|^2
\]

where \( \nu_i \) is the mean of points in \( S_i \).
Clustering cause-specific rate-of-aging

- We aim to classify all $n$ CODs by their $r^c(x)$ which we reduced to a vector of lagged-coefficients for each COD $c$.

$$[\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_c, \ldots, \hat{\alpha}_n]$$

- We use a $k$-means clustering which allows us to:
  - cluster our observations
  - extract a center for each cluster, i.e. prototypes for the different rates-of-aging

- We aim to partition the $n = 531$ observations into $k$ sets $S = \{S_1, S_2, \ldots, S_k\}$ such that

$$\arg\min_S \sum_{i=1}^{k} \sum_{\alpha \in S_i} ||\alpha - \nu_i||^2$$

where $\nu_i$ is the mean of points in $S_i$

- The procedure highly depends upon the number of clusters $k$
Optimal number of clusters

- Various criteria are available for optimize the number of clusters
Optimal number of clusters

- Various criteria are available for optimize the number of clusters
- Options:
  1. subjectively pick a given number of clusters
  2. choose a selection criterion
  3. run several indices for determining the number of clusters and pick the modal value
Optimal number of clusters

- Various criteria are available for optimize the number of clusters
- Options:
  1. subjectively pick a given number of clusters
  2. choose a selection criterion
  3. run several indices for determining the number of clusters and pick the modal value
Optimal number of clusters

- Various criteria are available for optimize the number of clusters

- Options:
  1. subjectively pick a given number of clusters
  2. choose a selection criterion
  3. run several indices for determining the number of clusters and pick the modal value

Distribution of optimal number of clusters for $r^c(x)$ for 24 different methods. Czech Republic, 1998-2011, ages 30-100, males.
Optimal number of clusters

- Various criteria are available for optimize the number of clusters
- Options:
  1. subjectively pick a given number of clusters
  2. choose a selection criterion
  3. run several indices for determining the number of clusters and pick the modal value

With $k = 3$ we obtain the means for each cluster $\nu_i$ and we evaluate the centers of the 3 different instantaneous rate-of-aging

$$\gamma_i = \frac{1}{h} B^{q-1} \nu_i \quad \text{for} \quad i = 1, 2, 3$$

and the associated age-profiles
Clusters’ centers

Instantaneous rate-of-aging (left) and log-mortality (right) for the centers of the 3 clusters formed by $k$-means algorithm. The width of the lines is proportional to the number of deaths within each cluster. Czech Republic, 1998-2011, ages 30-100, males.
Looking within clusters

Cluster 1, red lines

- Features (~6% of all deaths):
  - decreasing rate-of-aging
  - unusual age pattern reaching maximum around age 50
Looking within clusters

Cluster 1, red lines

- Features (~6% of all deaths):
  - decreasing rate-of-aging
  - unusual age pattern reaching maximum around age 50

- CODs:
  - genetically-conditioned diseases (rarer types of cancer, epilepsy, lupus, ulcerative colitis)
  - accidental deaths (traffic accidents, suicide, homicide, drowning)
  - alcohol related mortality
Looking within clusters

Cluster 1, red lines: **Premature**

- Features (~6% of all deaths):
  - decreasing rate-of-aging
  - unusual age pattern reaching maximum around age 50

- CODs:
  - genetically-conditioned diseases (rarer types of cancer, epilepsy, lupus, ulcerative colitis)
  - accidental deaths (traffic accidents, suicide, homicide, drowning)
  - alcohol related mortality
Looking within clusters

Cluster 2, blue lines

- Features (∼12%):
  - a quasi-constant rate of mortality change
  - log-linear of the force of mortality
Looking within clusters

Cluster 2, blue lines

- Features (~12%):
  - a quasi-constant rate of mortality change
  - log-linear of the force of mortality

- CODs:
  - bacterial infections
  - malignant melanoma of skin
  - acute respiratory diseases
  - most of digestive and genitourinary diseases
  - accidental falls
Looking within clusters

Cluster 2, blue lines: **Gompertzian**

- **Features (~12%)**:  
  - A quasi-constant rate of mortality change  
  - Log-linear of the force of mortality

- **CODs**:
  - Bacterial infections  
  - Malignant melanoma of skin  
  - Acute respiratory diseases  
  - Most of digestive and genitourinary diseases  
  - Accidental falls

**Looking within clusters**

**Cluster 2, blue lines**: **Gompertzian**

- **Features (~12%)**:
  - A quasi-constant rate of mortality change
  - Log-linear of the force of mortality

- **CODs**:
  - Bacterial infections
  - Malignant melanoma of skin
  - Acute respiratory diseases
  - Most of digestive and genitourinary diseases
  - Accidental falls
Looking within clusters

Cluster 3, green lines

- Features (~82%):
  - decreasing rate-of-aging
  - rapid increase of $\mu(x)$ until age 60 and deceleration afterward
  - lowest mortality at age 30 and highest mortality at 50+
Looking within clusters

Cluster 3, green lines

- Features (~82%):
  - decreasing rate-of-aging
  - rapid increase of $\mu(x)$ until age 60 and deceleration afterward
  - lowest mortality at age 30 and highest mortality at 50+

- CODs:
  - tuberculosis
  - colorectal and stomach cancers
  - most of the smoking-related CODs
  - man-made respiratory infections
  - diabetes, dementias and Alzheimer disease
  - circulatory diseases
  - no accidental deaths
Looking within clusters

Cluster 3, green lines: **Degenerative**

- **Features (~82%)**:
  - decreasing rate-of-aging
  - rapid increase of $\mu(x)$ until age 60 and deceleration afterward
  - lowest mortality at age 30 and highest mortality at 50+

- **CODs**:
  - tuberculosis
  - colorectal and stomach cancers
  - most of the smoking-related CODs
  - man-made respiratory infections
  - diabetes, dementias and Alzheimer disease
  - circulatory diseases
  - no accidental deaths
Summary

- We aimed to classify CODs based on their age profile, i.e. rate-of-aging.
Summary

- We aimed to classify CODs based on their age profile, i.e. rate-of-aging.

- By means of nonparametric techniques, we estimated cause-specific instantaneous rate-of-aging, \( r^c(x) \).
Summary

▶ We aimed to classify CODs based on their age profile, i.e. rate-of-aging

▶ By means of nonparametric techniques, we estimated cause-specific instantaneous rate-of-aging, $r_c(x)$

▶ We carry about a cluster analysis on $r_c(x)$ minimizing the within-cluster sum of squares
Summary

▶ We aimed to classify CODs based on their age profile, i.e. rate-of-aging

▶ By means of nonparametric techniques, we estimated cause-specific instantaneous rate-of-aging, $r_c(x)$

▶ We carry about a cluster analysis on $r_c(x)$ minimizing the within-cluster sum of squares

▶ For the Czech data, we identify three predominant age patterns of human diseases:
  1. age-independent
  2. rising constantly with age
  3. decelerating mortality at older ages
Outlook

- Explore other levels of clustering to define more clusters within the main three groups

Dendogram of the cluster analysis carried on $r^c(x)$.
Czech Republic, 1998-2011, ages 30-100, males.
Outlook

- Explore other levels of clustering to define more clusters within the main three groups

- Pre-group CODs entering into the analysis based on common medical definitions
Outlook

- Explore other levels of clustering to define more clusters within the main three groups

- Pre-group CODs entering into the analysis based on common medical definitions

- Tested this approach on other countries to assess whether there are universal patterns
Outlook

▶ Explore other levels of clustering to define more clusters within the main three groups

▶ Pre-group CODs entering into the analysis based on common medical definitions

▶ Tested this approach on other countries to assess whether there are universal patterns

▶ Suggestions from you side?
Thanks for your attention.

Comments and/or questions?

Giancarlo: carlo-giovanni.camarda@ined.fr
Marketa: Marketa.Pechholdova@seznam.cz