Exploratory analysis on COVID-19 Mortality Age-patterns

Carlo Giovanni Camarda
sites.google.com/site/carlogiovannicamarda

Institut National d'Études Démographiques, France

TAG Working Group I, Meeting IV

April 30th, 2021

Data and models

- Data:
  - a subset of COVerAGE-DB with 100+ deaths
  - either both sex or males and females separately
  - scaled to JHU total on Dec 31, 2020
- Log-mortality over age x, for population p and sex s:

  C-PCLM: Population-specific, sexes combined

$$\eta(x, p) = \eta^p(x)$$

  S-PCLM: Population-sex-specific

$$\eta(x, p, s) = \eta^{p,s}(x)$$

  A-PCLM: An additive approach:

$$\eta(x, p, s) = \eta^0(x) + s(x) + \delta^p(x)$$

$\eta^0(x)$: reference mortality
$s(x)$: sex factor
$\delta^p(x)$: population-specific deviance

Hierarchical k-means clustering on different outcomes

C-PCLM: Sexes combined rate-of-aging

$$\frac{d\,\eta^{p}(x)}{d\,x}$$

+ available for more populations & more robust
− mixture of eventual different patterns

S-PCLM: Sex-specific rate-of-aging, then combined

$$\left[\frac{d\,\eta^{p,s=F}(x)}{d\,x}; \frac{d\,\eta^{p,s=M}(x)}{d\,x}\right]$$

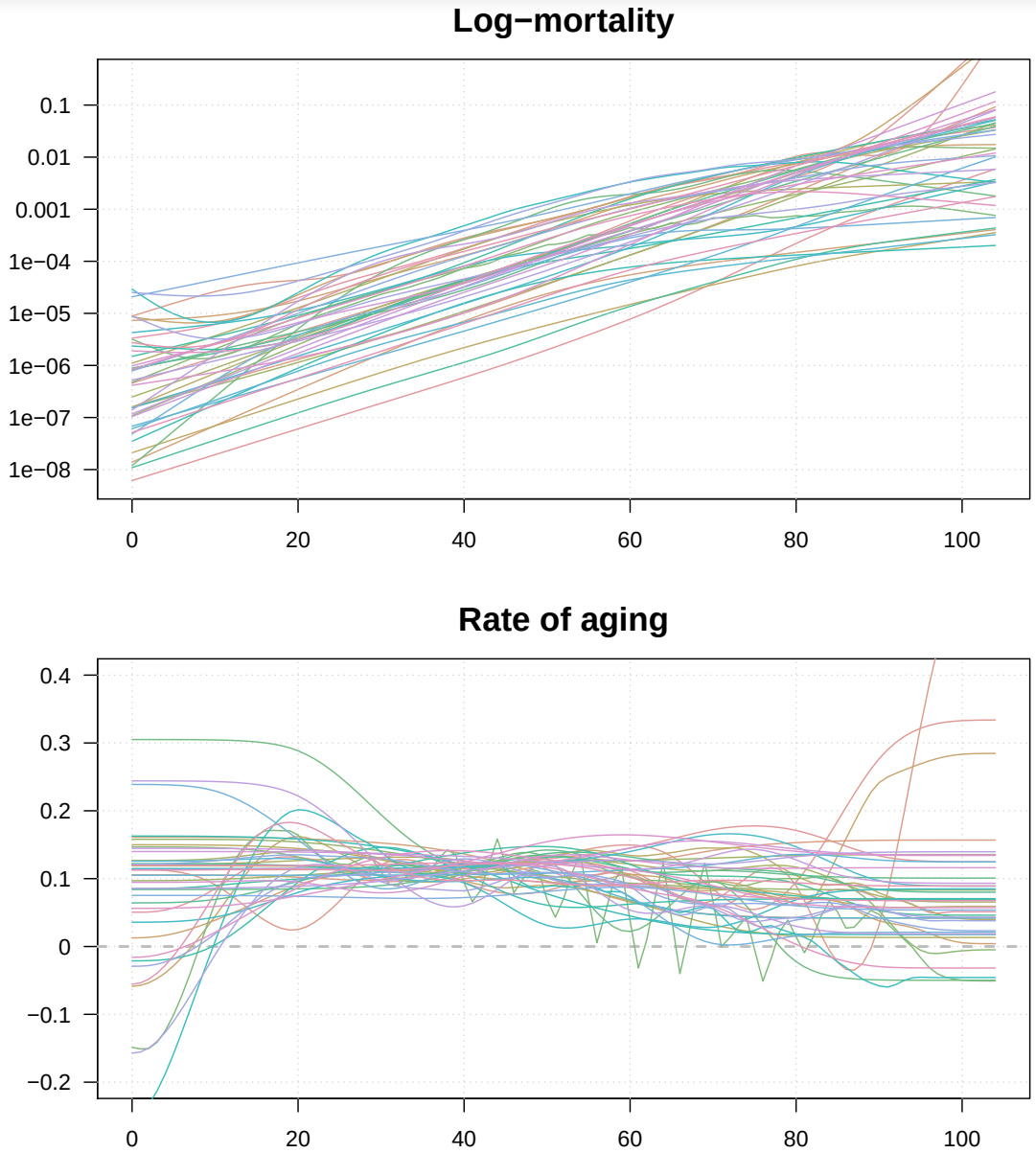
+ enforce both sexes to be included in the same group
− less populations with available data by sex

A-PCLM: population-specific deviance

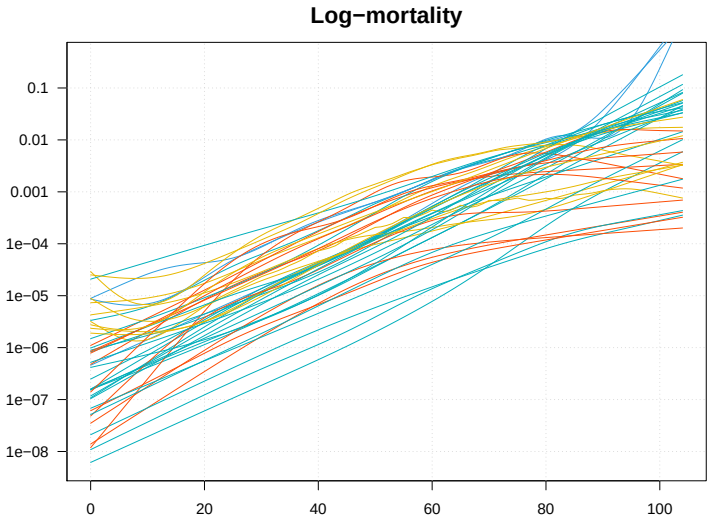
$$\delta^{p}(x)$$

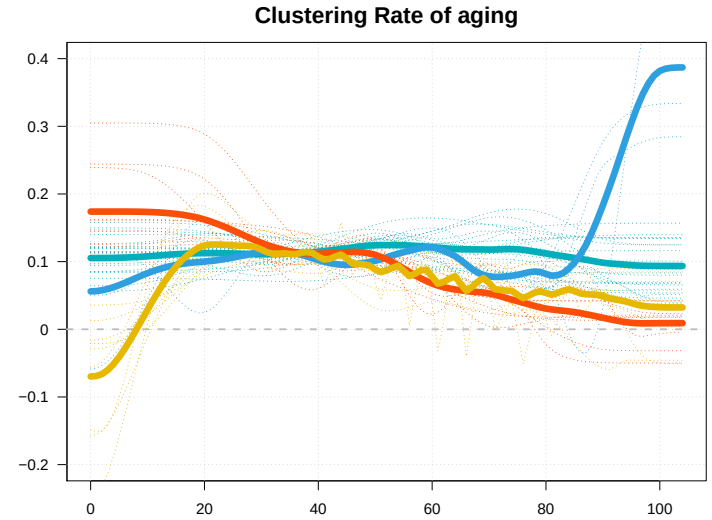
+ all data estimated within a single model & control for sex
− slightly more complex interpretation

C-PCLM

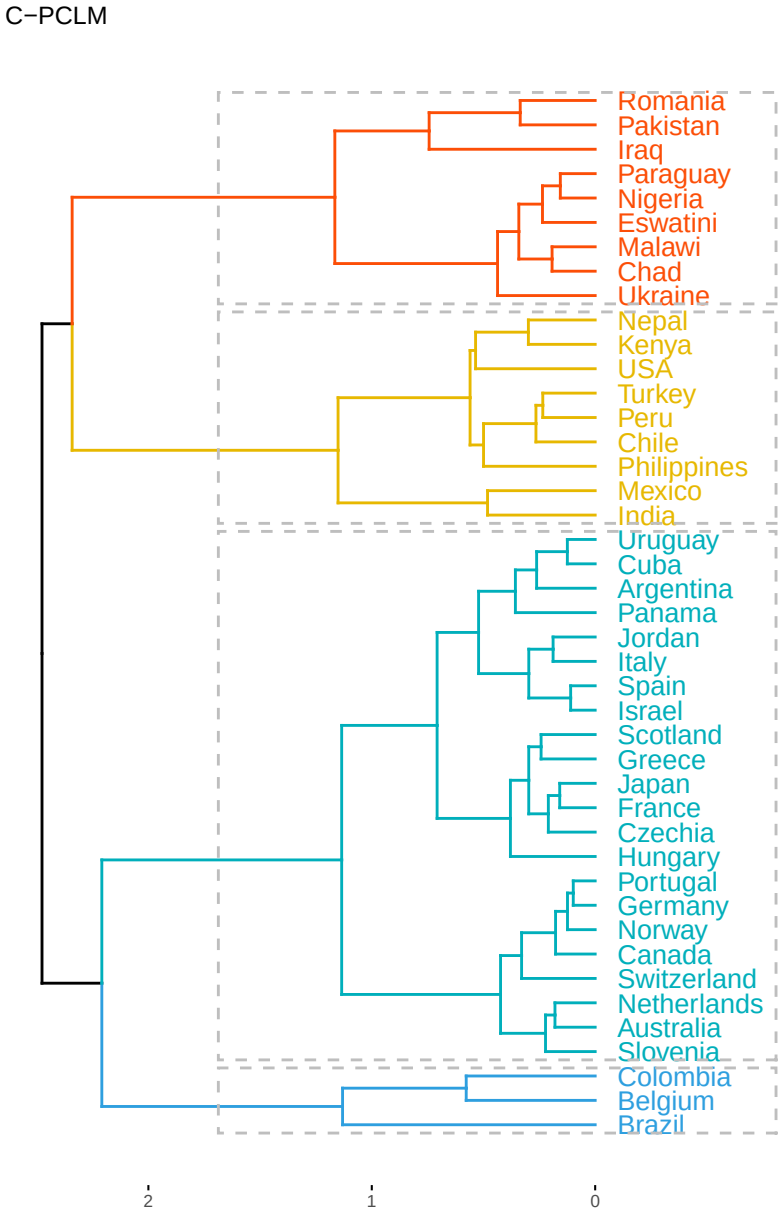

Log−mortality
0.1
0.01
0.001
1e−04
1e−05
1e−06
1e−07
1e−08
0
20
40
60
80
100
Rate of aging
0.4
0.3
0.2
0.1
0
−0.1
−0.2
0
20
40
60
80
100

C-PCLM

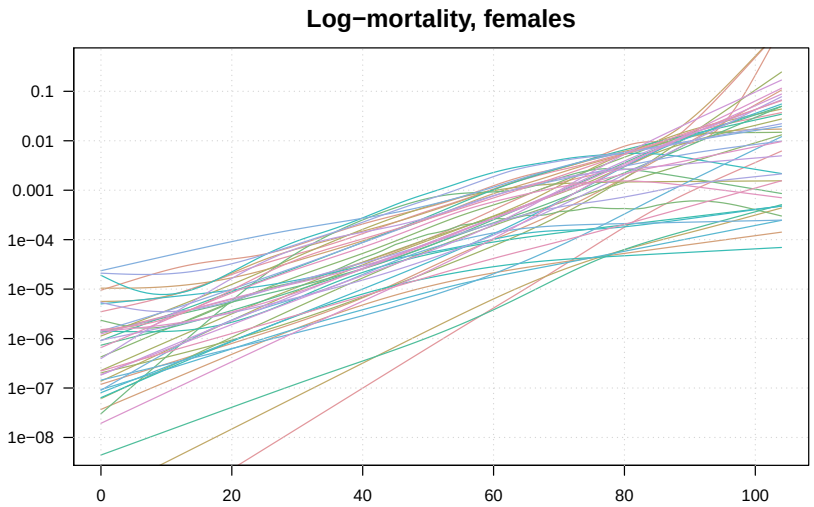


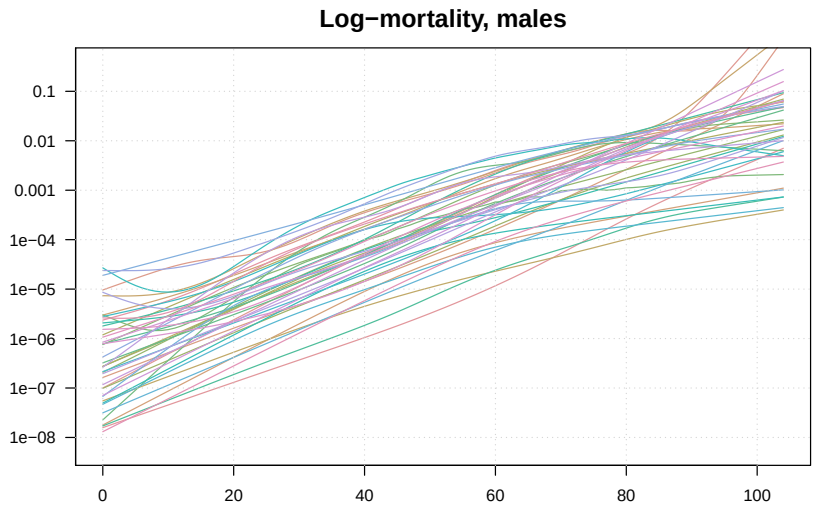


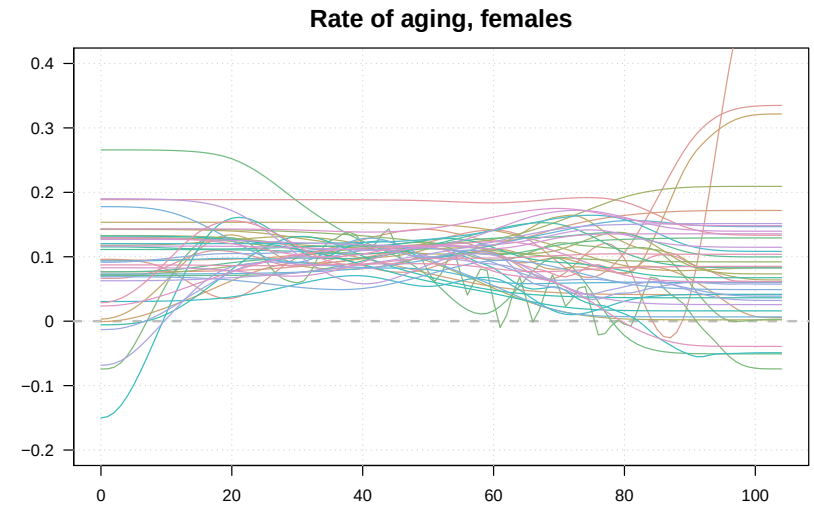
Explained variance PC1+PC2: 40.1% + 22.1%

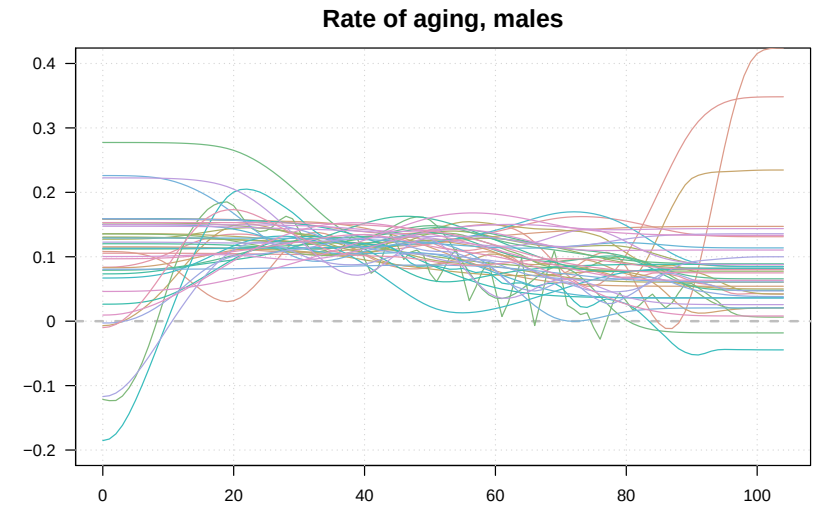


S-PCLM

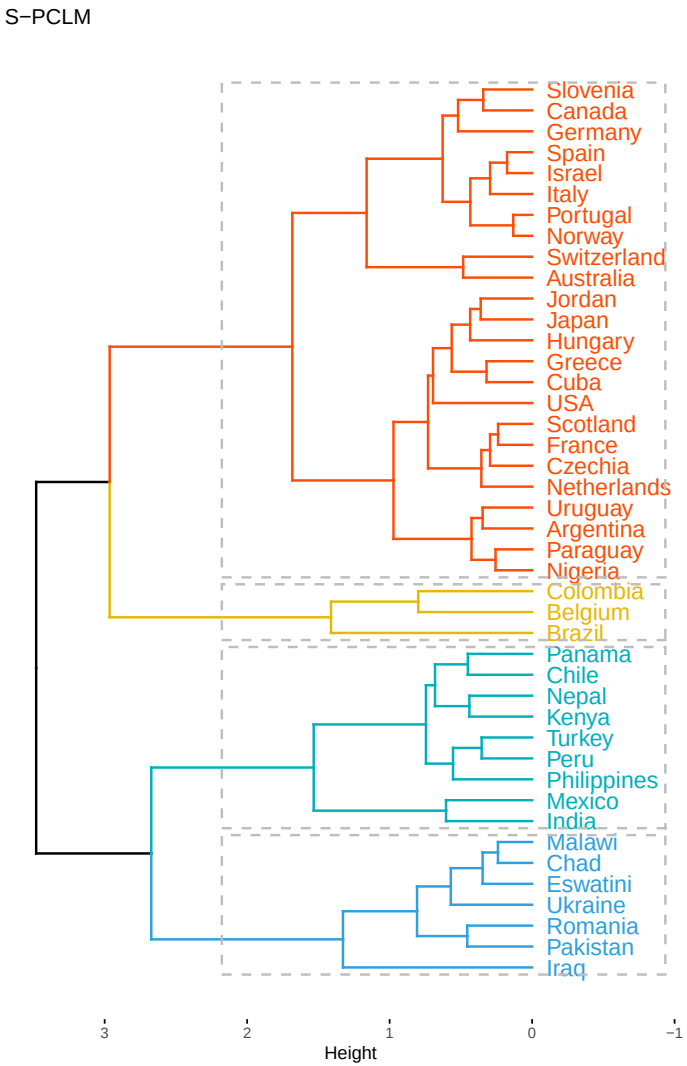

Log−mortality, females
0.1
0.01
0.001
1e−04
1e−05
1e−06
1e−07
1e−08
0
20
40
60
80
100


Log−mortality, males
0.1
0.01
0.001
1e−04
1e−05
1e−06
1e−07
1e−08
0
20
40
60
80
100


Rate of aging, females
0.4
0.3
0.2
0.1
0
−0.1
−0.2
0
20
40
60
80
100

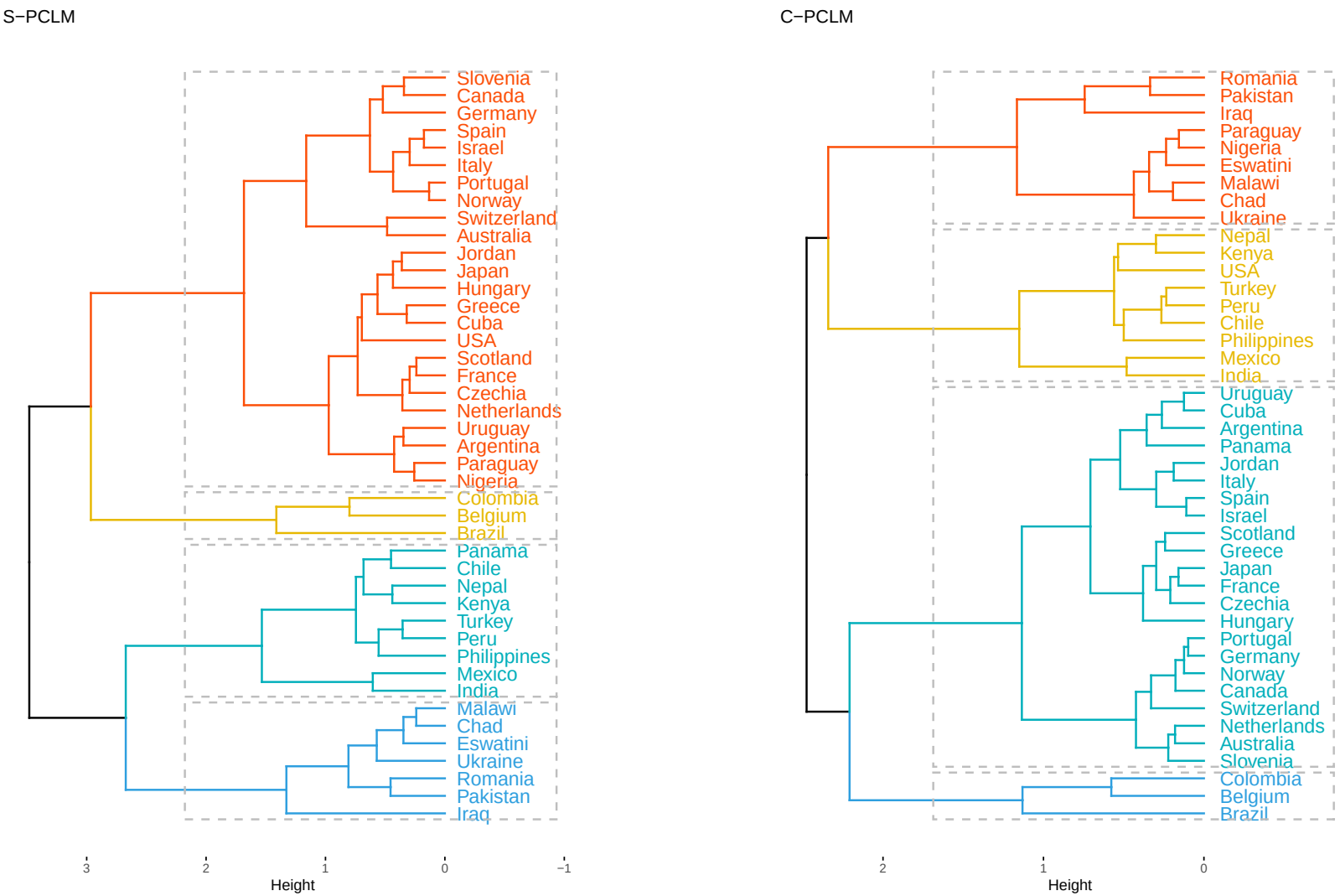

Rate of aging, males
0.4
0.3
0.2
0.1
0
−0.1
−0.2
0
20
40
60
80
100

S-PCLM

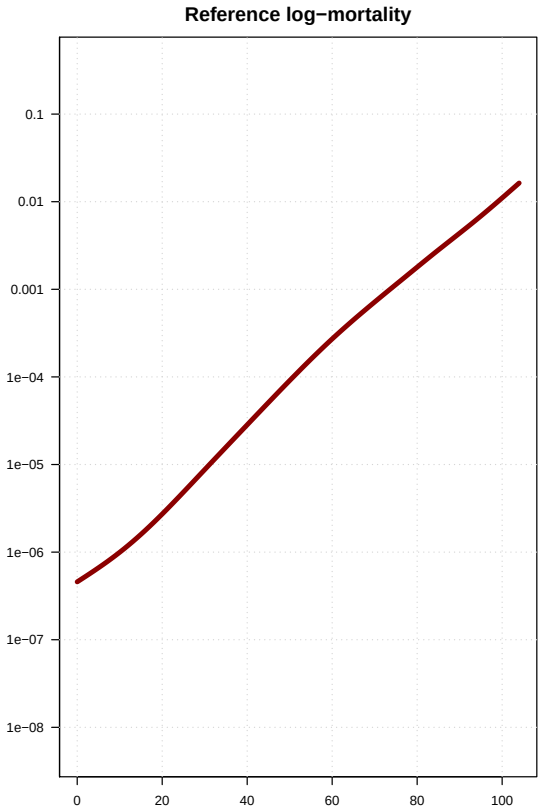


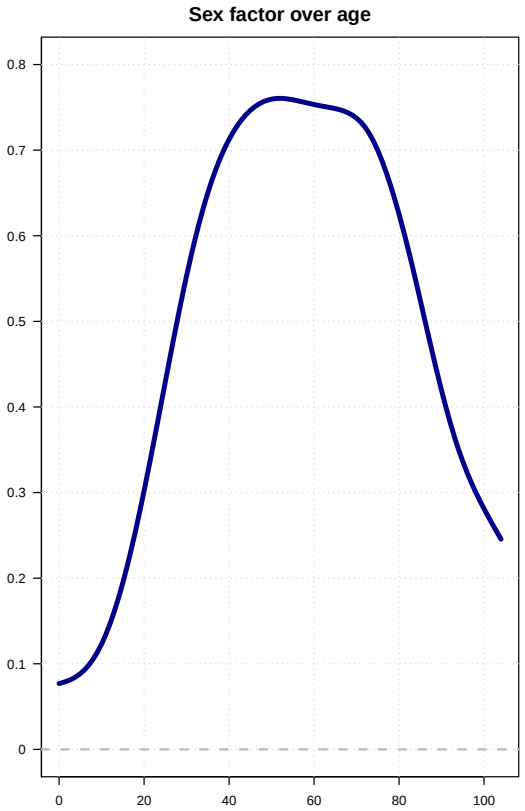
Explained variance PC1+PC2: 37.6% + 22.3%

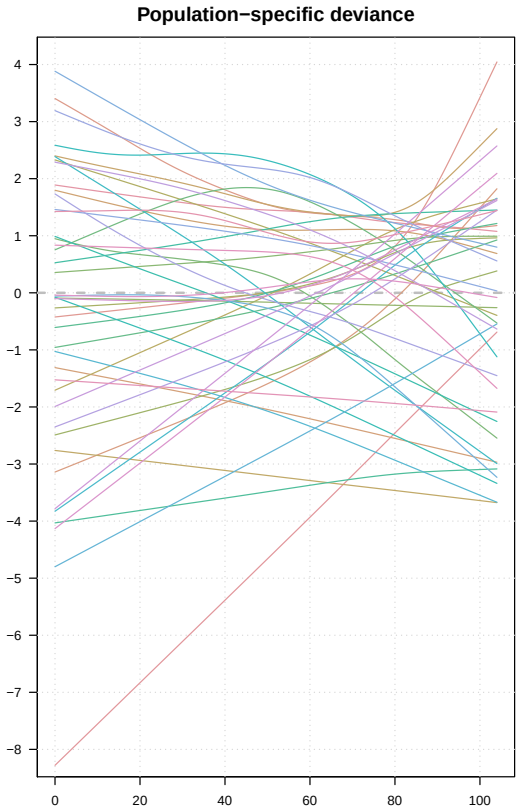
S-PCLM vs. *C-PCLM*



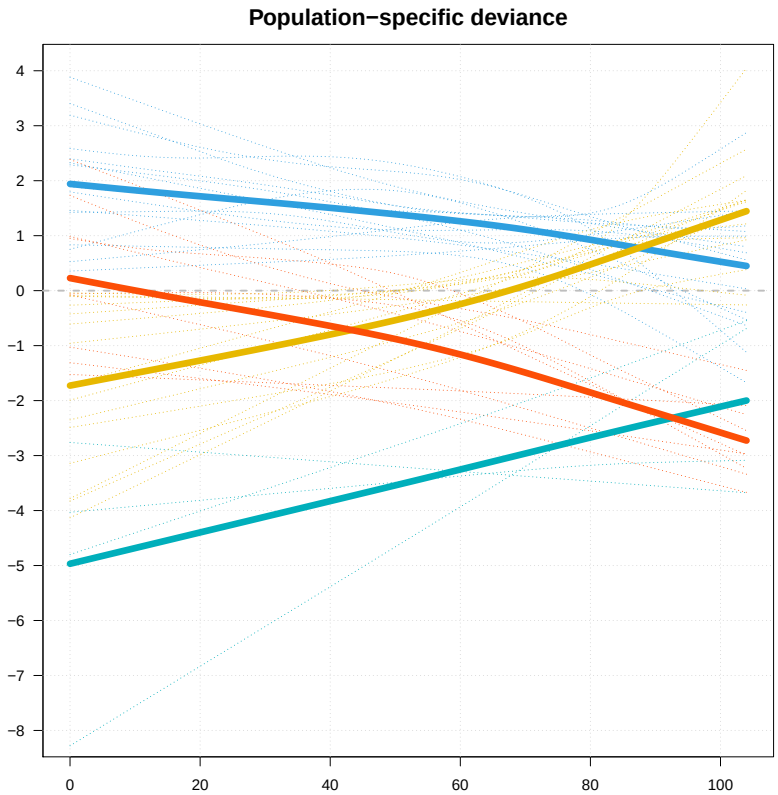
A-PCLM


Reference log−mortality
0.1
0.01
0.001
1e−04
1e−05
1e−06
1e−07
1e−08
0
20
40
60
80
100

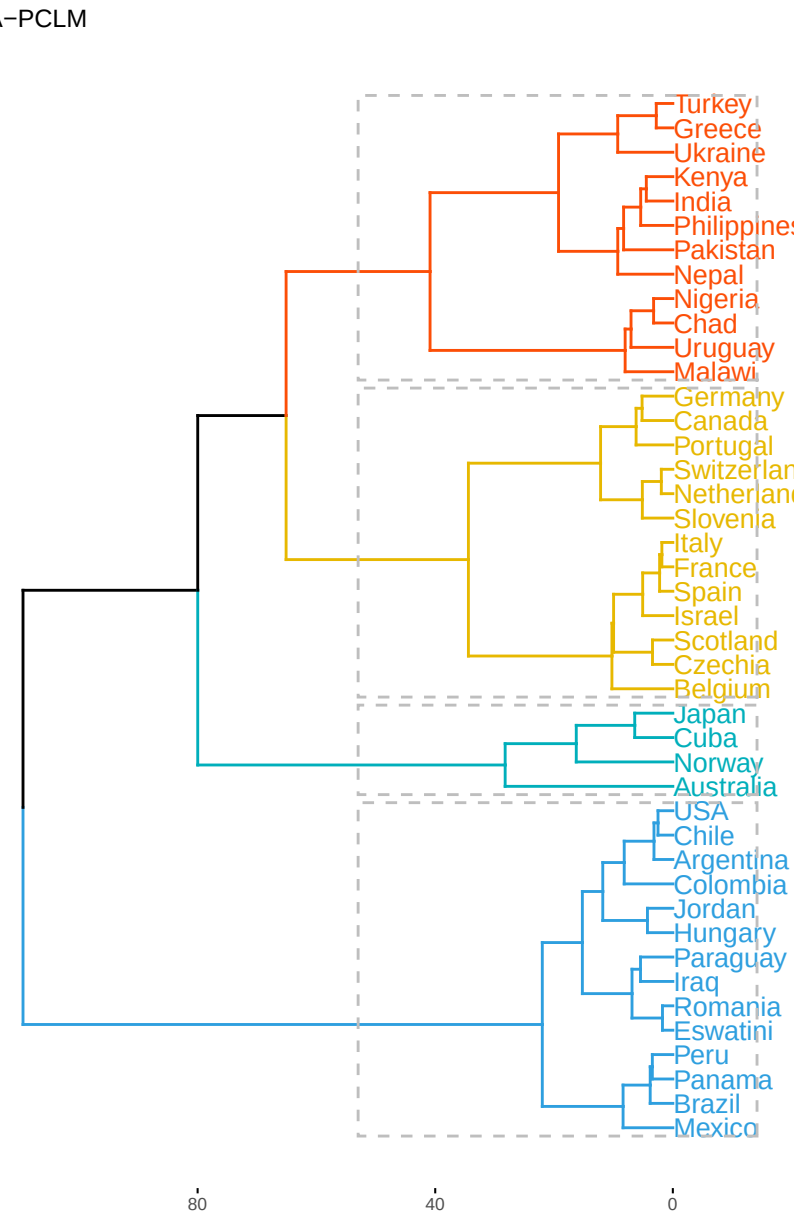

Sex factor over age
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100


Population−specific deviance
4
3
2
1
0
−1
−2
−3
−4
−5
−6
−7
−8
0
20
40
60
80
100

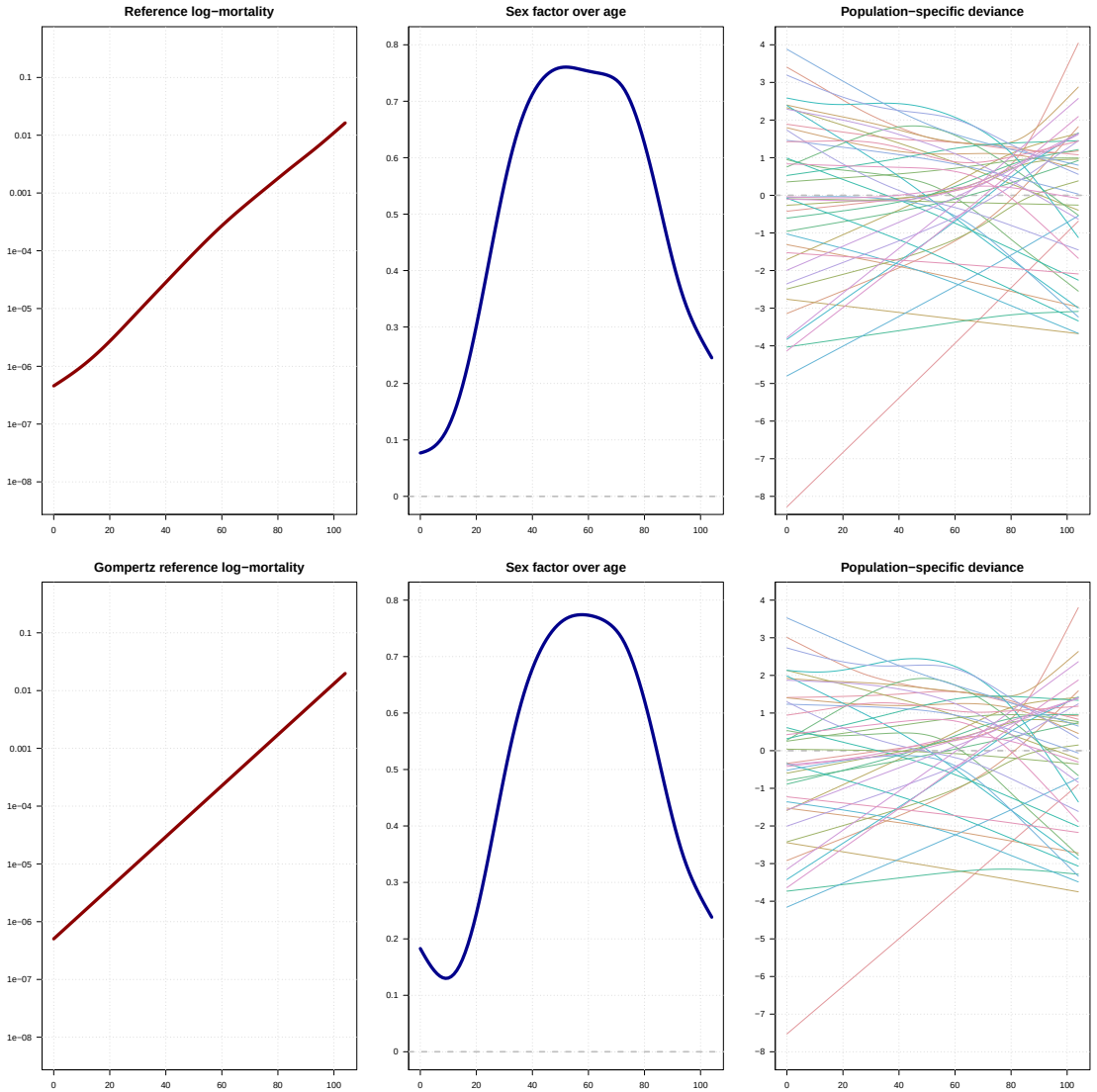
A-PCLM



Explained variance PC1+PC2: 77.3% + 21.5%

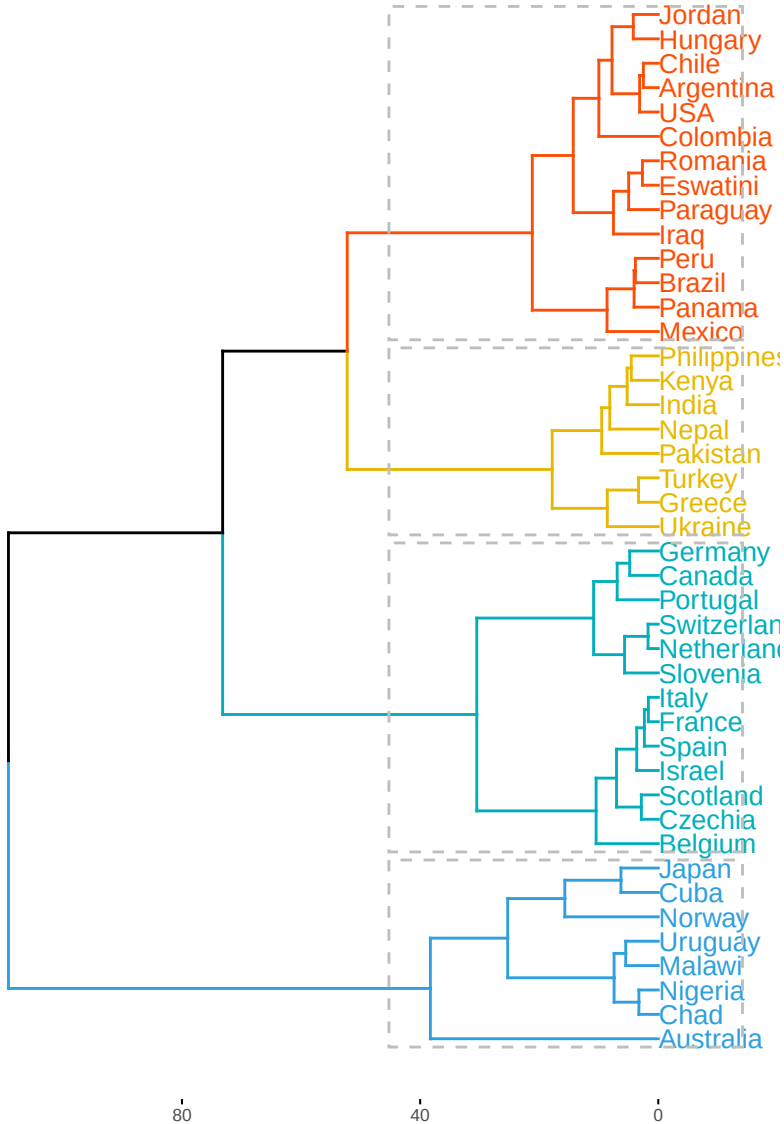


A-PCLM

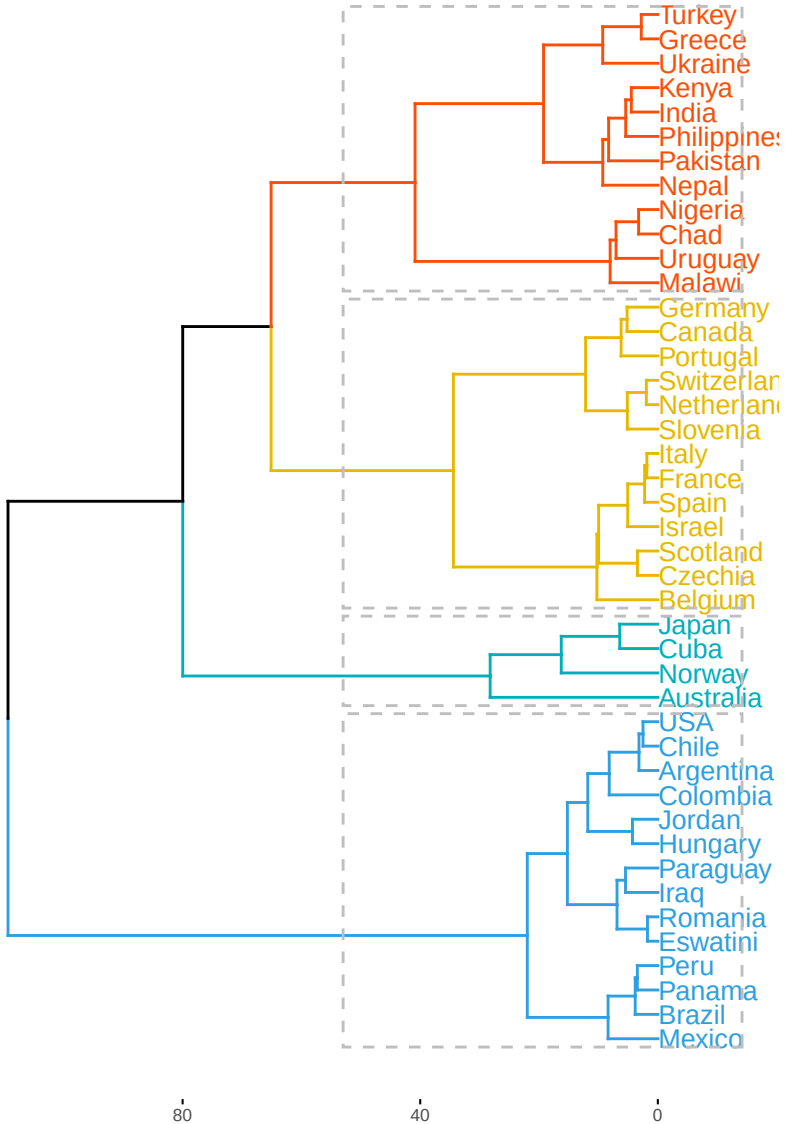

Reference log−mortality
Sex factor over age
Population−specific deviance
Gompertz reference log−mortality
Sex factor over age
Population−specific deviance

A-PCLM

A−PCLM with Gompertz reference



A−PCLM



Still to discuss/solve

- **Main question**: What do we do when difference outcomes/models lead to different classification?
- Unclear optimal number of clusters (fuzzy clustering?)
- Restricting the analysis to adult mortality?
- Mapping clusters to identify eventual (geographical) regions
- "Spatial borrowing" to estimate when no data is available
- Can we do the same exercise on excess mortality? Issue: data availability
- Computing confidence intervals of the estimates, optimizing smoothing parameters