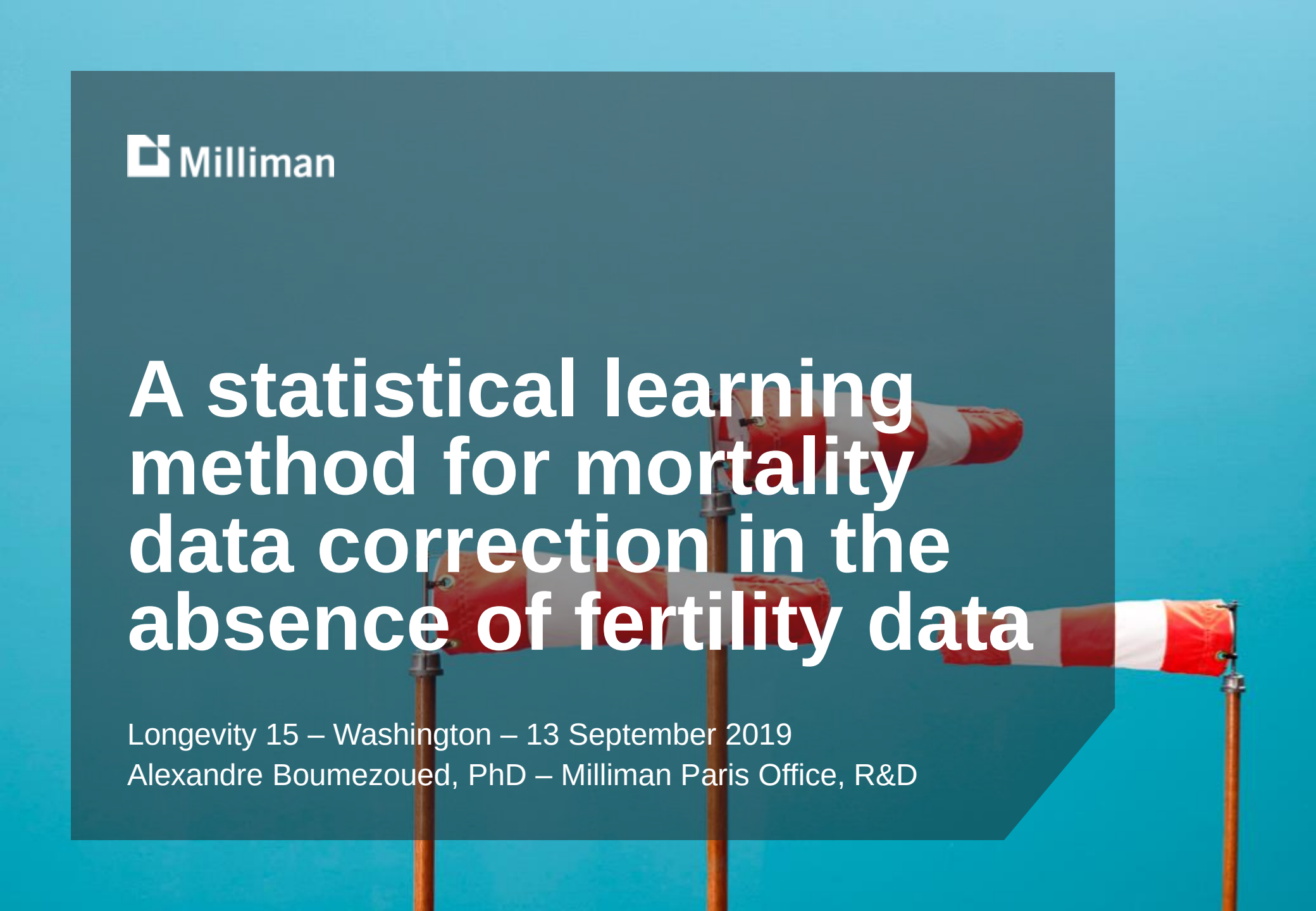


Milliman

A statistical learning method for mortality data correction in the absence of fertility data

Longevity 15 – Washington – 13 September 2019

Alexandre Boumezoued, PhD – Milliman Paris Office, R&D

Agenda

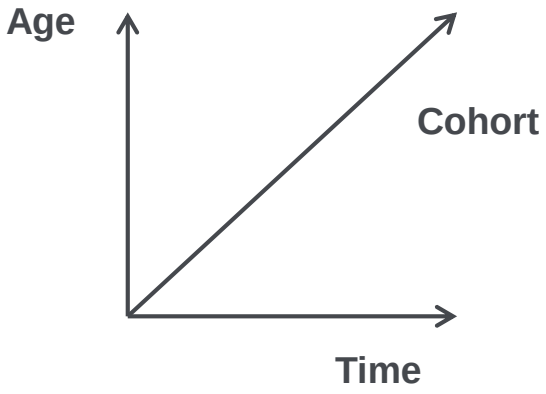
1	**Context**
2	Data correction for countries with insufficient fertility histories
3	Detailed results by country

Context

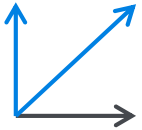
National mortality tables

- How to estimate the **mortality rate** based on national population data?
  - The statistical inference of a death rate with two crossing dimensions (age and time) is an **old (Lexis, 1875) and still challenging estimation problem**
- In practice, individuals are grouped into **age and time blocks**, and the death rate is assumed to be constant on each block
  - This leads to the so-called **Lexis diagram**

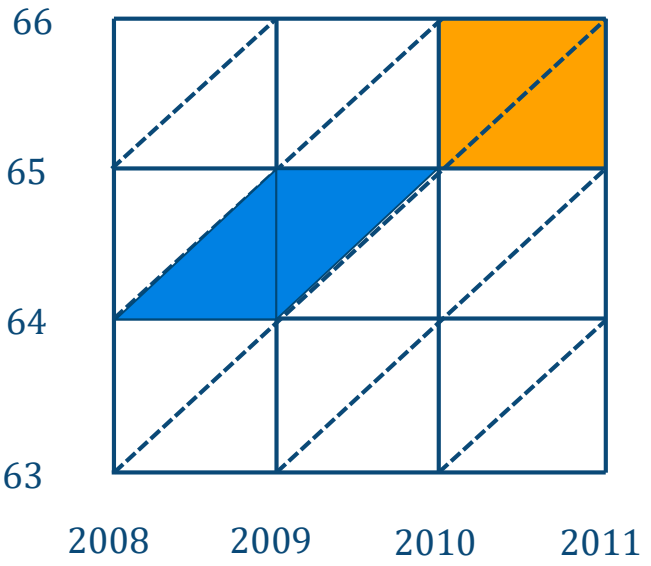
Three directions of analysis

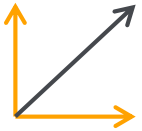


Cohort tables
=
Death rate assumed constant over parallelograms



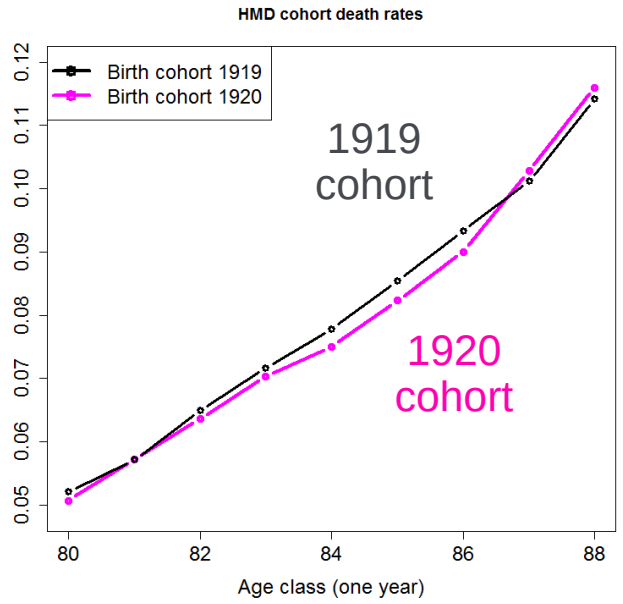
Period tables
=
Death rate assumed constant over squares

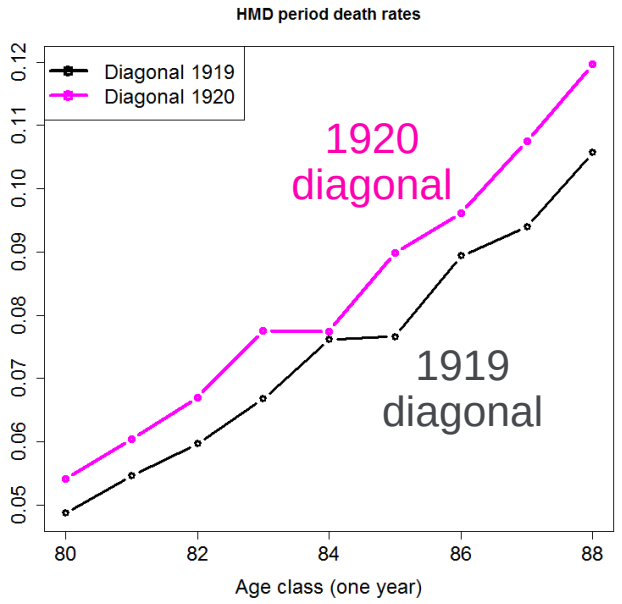




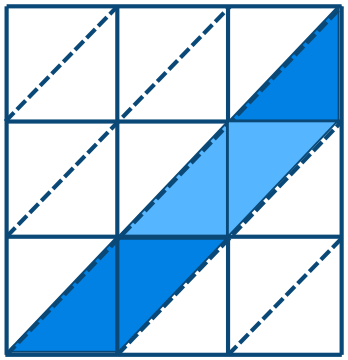
Context

Anomalies in period tables: example for France

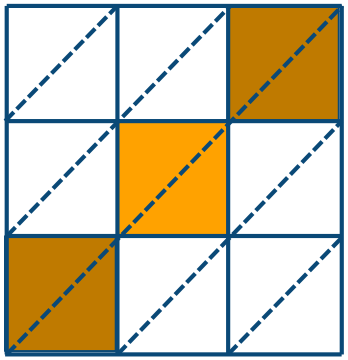




Cohort table

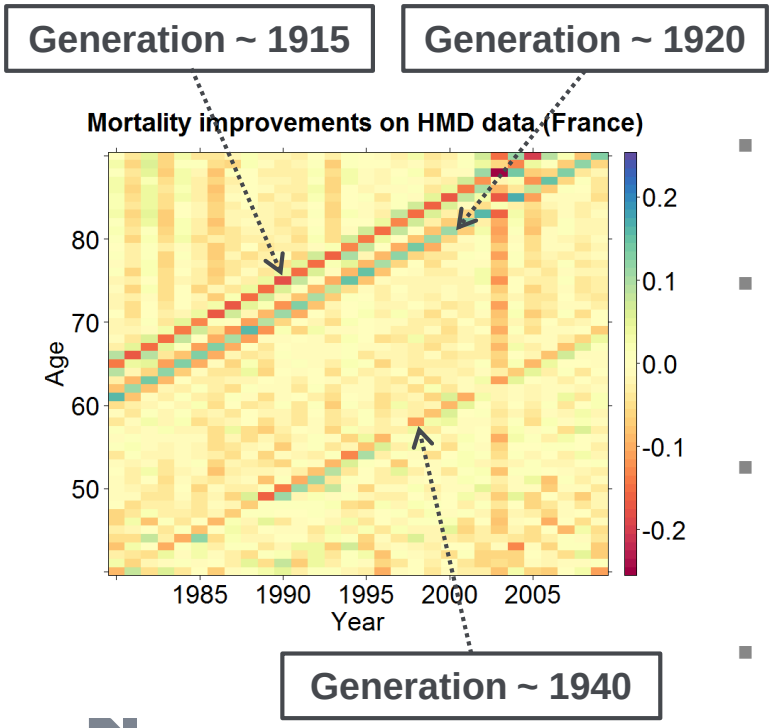


Period table



Context
Reliability issues for period tables

- **Period tables** are useful to study the dynamics of mortality over time
  - Period mortality rate for age x and year t denoted $\mu(x,t)$
  - Improvement rates $r(x,t) = \frac{\mu(x,t+1)-\mu(x,t)}{\mu(x,t)}$ are used to observe particular patterns
  - Clear **« cohort effects »** can be observed for specific generations (born around 1915, 1920 and 1940)



Literature on mortality data reliability

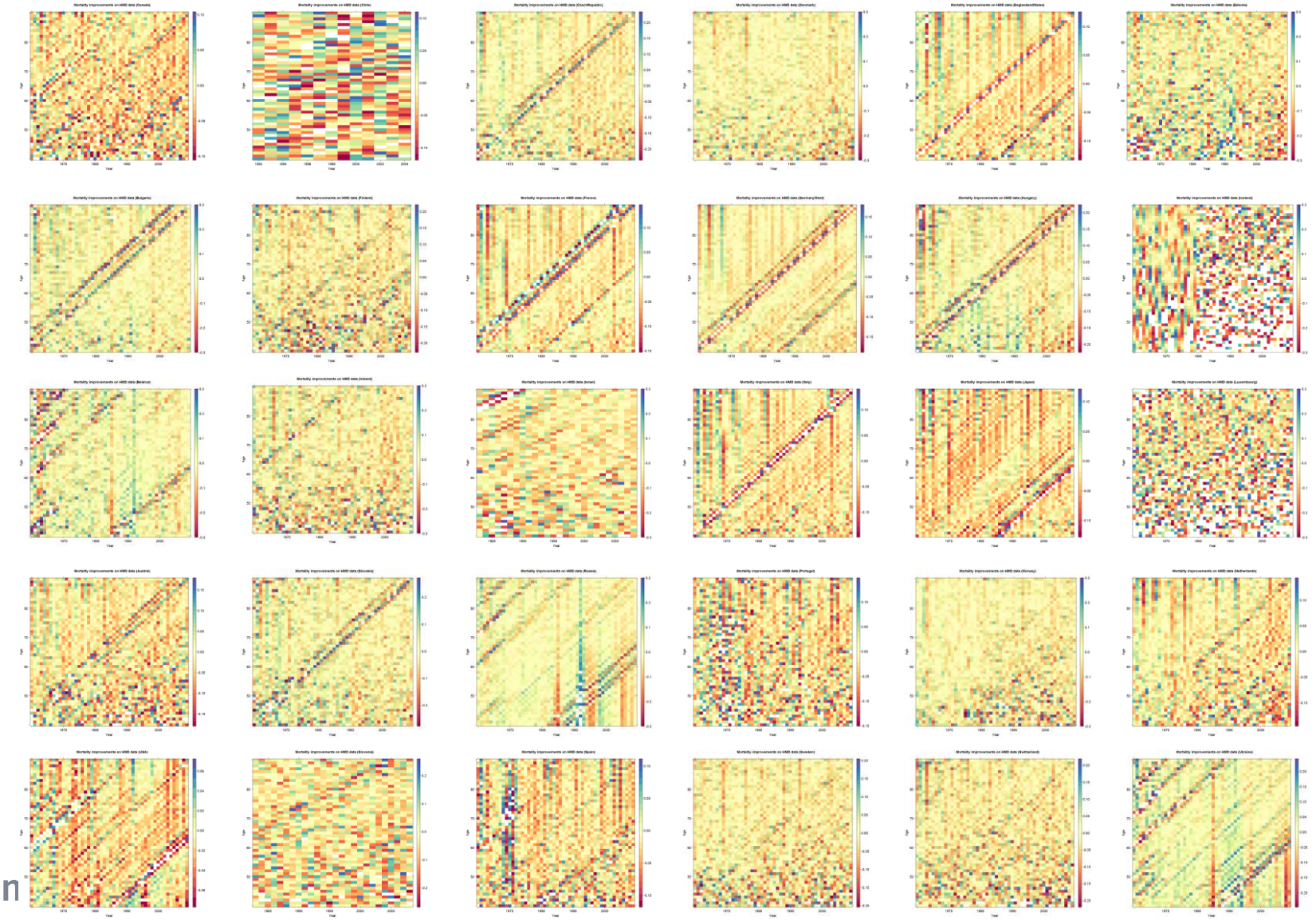
- Richards (2008) conjectured that the 1919 cohort effect for England and Wales is an **anomaly** in the mortality table due to **erratic birth patterns**
- Following Cairns et al. (2016) who studied ONS methodology, Boumezoued (2016) highlighted the **universal nature of these false cohort effects**, which were present in most period tables in the Version 5 of the Human Mortality Database (HMD).
- The HMD has worked on its own approach to this problem and released a **Version 6** update in February 2018, including a revision of exposure calculation based on monthly birth counts when available.
- Further **mathematical developments** have been proposed by Boumezoued et al (2018, 2019), who provide improved estimators and a related theory for death-rate inference.

Context

Version 5 of the Human Mortality Database

Anomalies in the form of isolated cohort effects are quite common, even universal

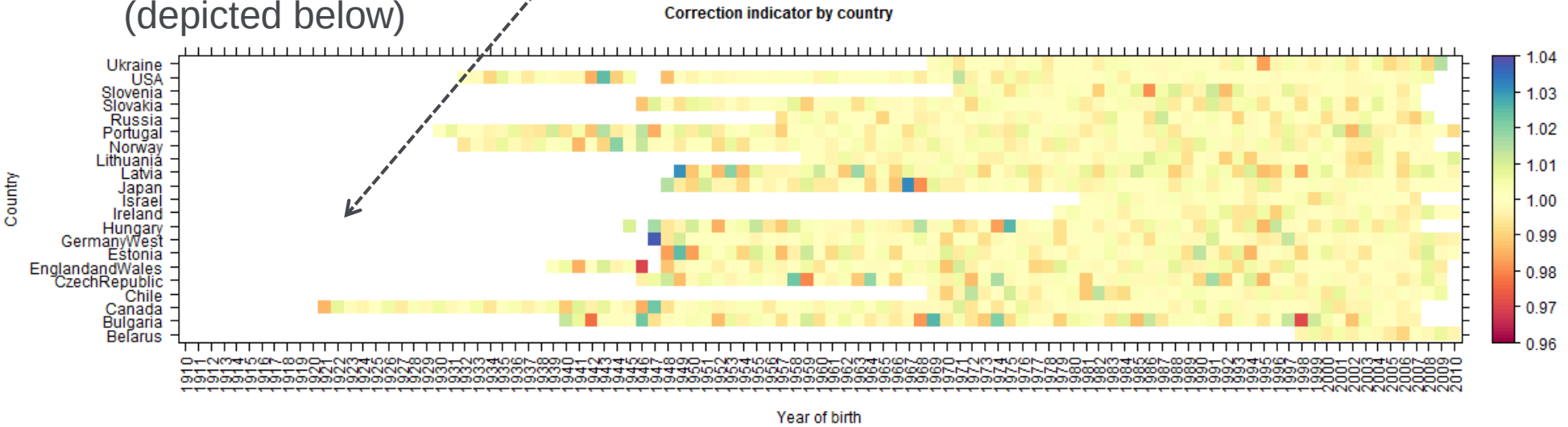
Depicted: mortality improvement rates

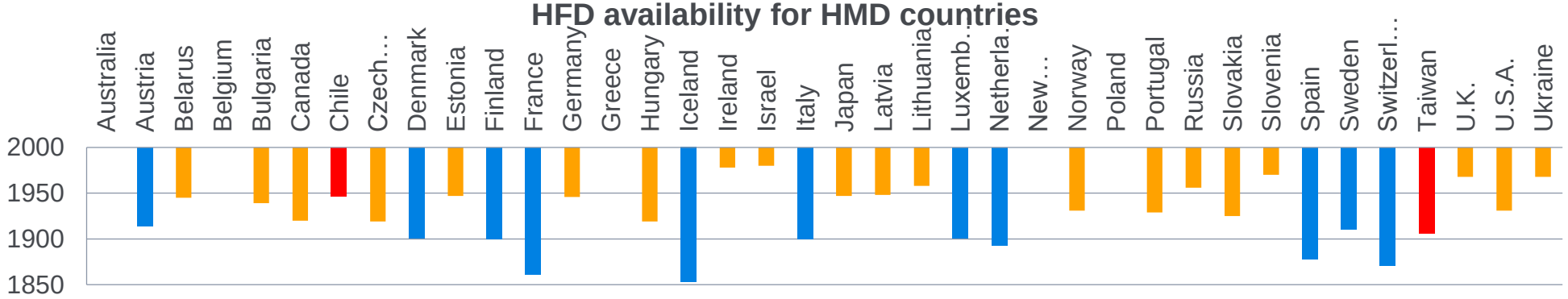


Context

Availability issue of the indicator for selected countries

- **Aim:** correct cohorts in the **white area** where monthly birth counts are not available, especially for the **countries for which the crucial 1920 effect is not captured** (depicted below)





Agenda

1	Context
2	**Data correction for countries with insufficient fertility histories**
3	Detailed results by country

A new correction method for more countries

Preliminary remarks

- We developed and compared **several statistical methodologies** to reconstruct the quality indicator in the period of unavailable fertility data, and to build the corrected mortality table
- The methods aim at expressing the correction indicator as a function of explanatory variables in a penalized-regression procedure based on
  i. the **fertility experience** of other countries in blue in previous slide,
  ii. and/or the **smoothness of the corrected mortality table**.
- The key ideas underlying components i) and ii) are the following:
  i. what are the other countries (in blue) showing **similar fertility patterns** in the available period, which can then be used to predict the quality indicator?
  ii. what is the quality indicator which best reduces, after correction, **the cohort irregularities** of the crude mortality table?

A new correction method for more countries

Statistical method – Insight (1/2)

- **Idea:** use the smoothness of the age x time corrected mortality surface as penalty -> this makes use of external mortality data as well
- This is inspired by **classical roughness penalties** for continuous functions f (see e.g. Meinguet (1979), Titterington (1985), Inoue (1986) and Ogata & Katsura (1987)) as

$$\phi_2(f) = \int \left\{ \left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right\} \mathrm{d}x\mathrm{d}y$$

- An “expert knowledge” perspective is embedded assuming that the resulting **mortality surface is smooth**
- How to define the “smoothness criteria”?
  - **Idea:** the mortality is expected to decrease nearly regularly over time, which is equivalent to the first order derivative being nearly constant, in other words the **second order derivative being close to zero**

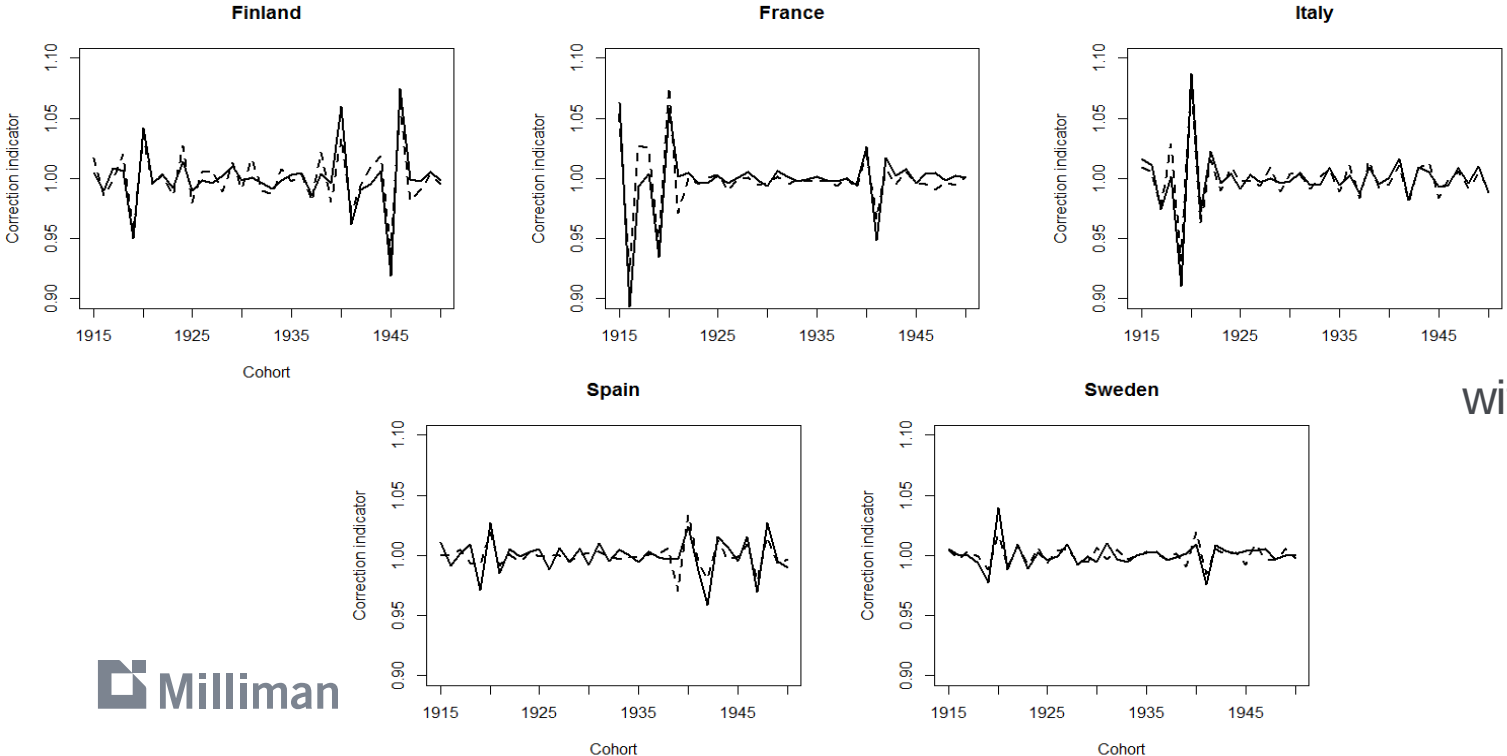
A new correction method for more countries

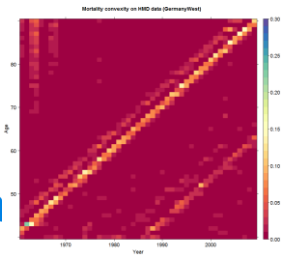
Statistical method – Insight (2/2)

- **The convexity matrix itself is a good predictor of the correction indica**
- We use a **convexity matrix** which does not exhibit a trend over time
  - It is the first order derivative of mortality improvements:

$$\Phi(x,t) = \frac{\partial r(x,t)}{\partial t} = r(x,t+1) - r(x,t) = \frac{\mu(x,t+2)}{\mu(x,t+1)} - \frac{\mu(x,t+1)}{\mu(x,t)}$$

- We directly test the **"predictive power" of the convexity matrix** on countries below:





The convexity matrix highlights the possible anomalies

Regression formula:

$$I(c) = \alpha_1 + \alpha_2 V(c),$$

with $V(c) = \frac{1}{n}\sum_{t-x=c} \Phi(x,t)$

n is the number of ages in the cohort

Milliman

A new correction method for more countries

Statistical method - Description

- The method aims at finding optimal coefficients which **maximize the smoothness of the corrected table**
  - **Constraint:** the indicator can not deviate from one in average (ensures a pure redistribution of the population exposures, as observed in practice with the standard correction approach)
    - This second constraint is included as a penalty in the optimization program
- Details of the optimization procedure:
  - Recall that the cohort vector is $V(c) = \frac{1}{n}\sum_{t-x=c}\Phi(x,t)$ with $\Phi(x,t) = \frac{\partial r(x,t)}{\partial t}$
  - The penalized smoothness criterion to be minimized is:

Tuning parameter

Sum on the cohorts to be extrapolated

Sum on the age x period region to be favored by the modeler

$$F(\alpha_1,\alpha_2) = \beta\left(\frac{1}{c_{max}-c_{min}+1}\sum_{c=c_{min}}^{c_{max}} I^{\alpha_1,\alpha_2}(c) - 1\right)^2 + \sum_{x=x_{min}}^{x_{max}}\sum_{t=t_{min}}^{t_{max}} \tilde{\Phi}^{\alpha_1,\alpha_2}(x,t)^2$$

Convexity matrix of the corrected table

$$\tilde{\Phi}^{\alpha_1,\alpha_2}(x,t) = \frac{\tilde{\mu}^{\alpha_1,\alpha_2}(x,t+2)}{\tilde{\mu}^{\alpha_1,\alpha_2}(x,t+1)} - \frac{\tilde{\mu}^{\alpha_1,\alpha_2}(x,t+1)}{\tilde{\mu}^{\alpha_1,\alpha_2}(x,t)}$$

Correction indicator

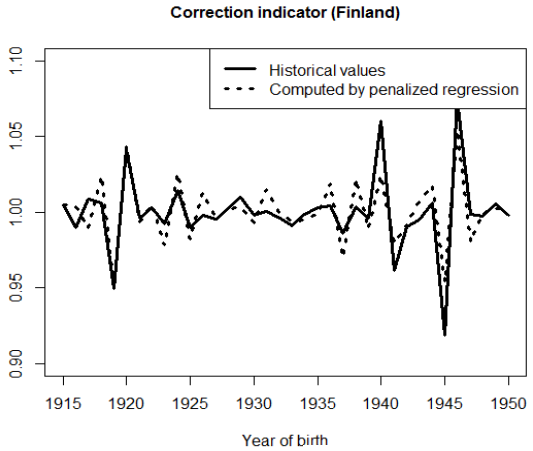
$$I^{\alpha_1,\alpha_2}(c) = \alpha_1 + \alpha_2 V(c)$$

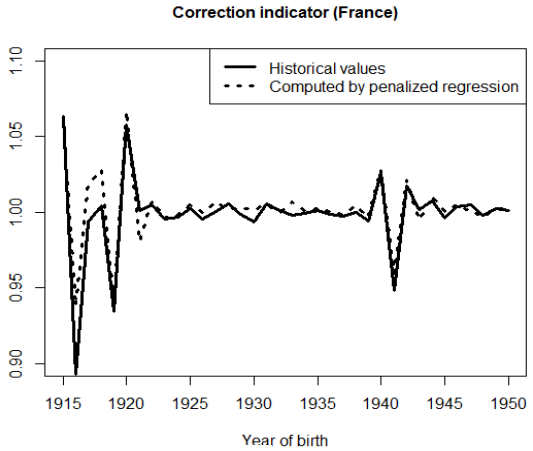
Corrected death rates

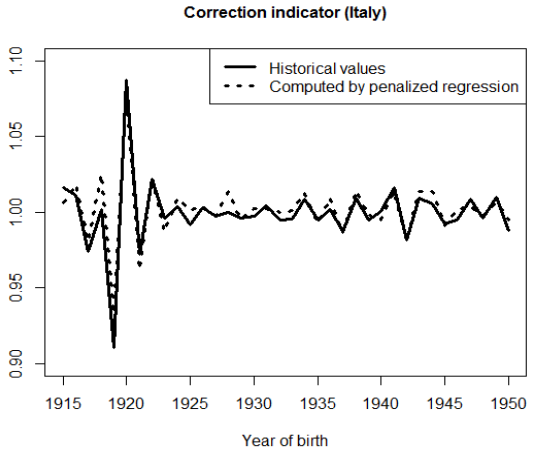
$$\tilde{\mu}^{\alpha_1,\alpha_2}(x,t) = \frac{\mu(x,t)}{I^{\alpha_1,\alpha_2}(t-x)}$$

A new correction method for more countries
Backtesting

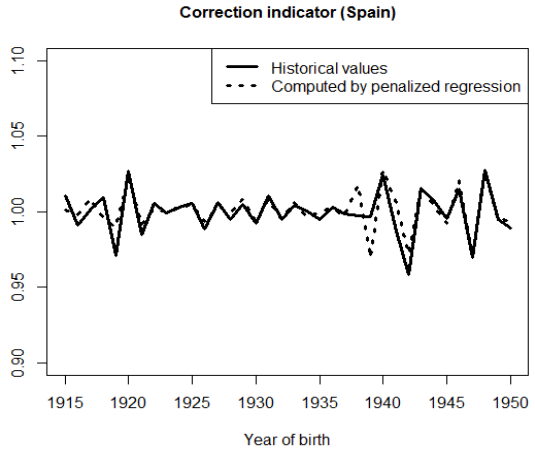
- This method shows **good backtesting results** for the prediction of the correction indicator:

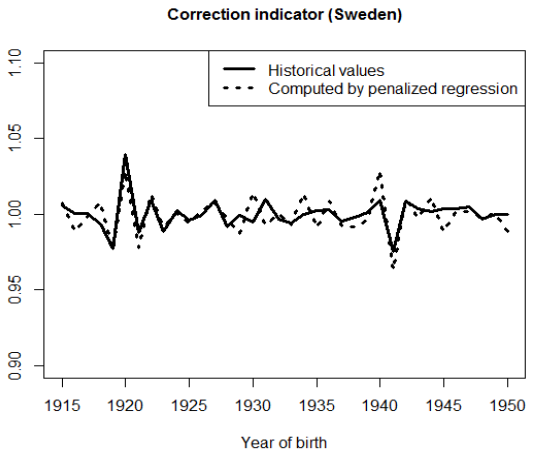






The tuning parameter β is first set to 1000 for all countries – in next slides it is proposed a cross-validation procedure to compute the β parameter automatically





A new correction method for more countries
Cross validation (1/2)

- A cross-validation procedure can be performed in order to **compute the tuning parameter β**
- The cross-validation algorithm works as follows:
  - Find optimal $\hat{\beta}$ which **minimizes the function**

$$\beta \longrightarrow \sum_{c=c_{min}}^{c_{max}} \left(I^{\widehat{\alpha_1(\beta)},\widehat{\alpha_2(\beta)}}(c) - I(c) \right)^2$$

where for each β, the $\widehat{\alpha_1(\beta)}, \widehat{\alpha_2(\beta)}$ are found to **minimize the map**

$$(\alpha_1, \alpha_2) \longrightarrow \beta \left(\frac{1}{c_{max} - c_{min} + 1} \sum_{c=c_{min}}^{c_{max}} I^{\alpha_1,\alpha_2}(c) - 1 \right)^2 + \sum_{x=x_{min}}^{x_{max}} \sum_{t=t_{min}}^{t_{max}} \tilde{\Phi}^{\alpha_1,\alpha_2}(x,t)^2$$

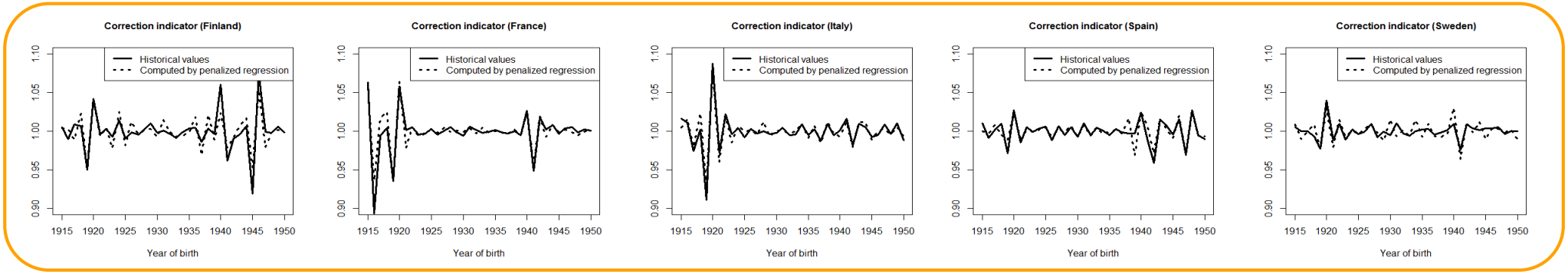
A new correction method for more countries

Cross validation (2/2)

- Results of the cross-validation procedure to compute the tuning parameter β
  - The tuning parameters for each country are of the same **order of magnitude**

	Finland	France	Italy	Spain	Sweden
tuning parameter β	59507	88969	96081	61420	73771

  - The final indicator prediction results are given below and compared to the original indicator
    - Compared to the previous slides with $\beta = 1000$, this shows that the **tuning parameter variations in this region have a low impact**



- To correct the mortality tables for other countries, we take a rough estimate $\boldsymbol{\beta = 75000}$

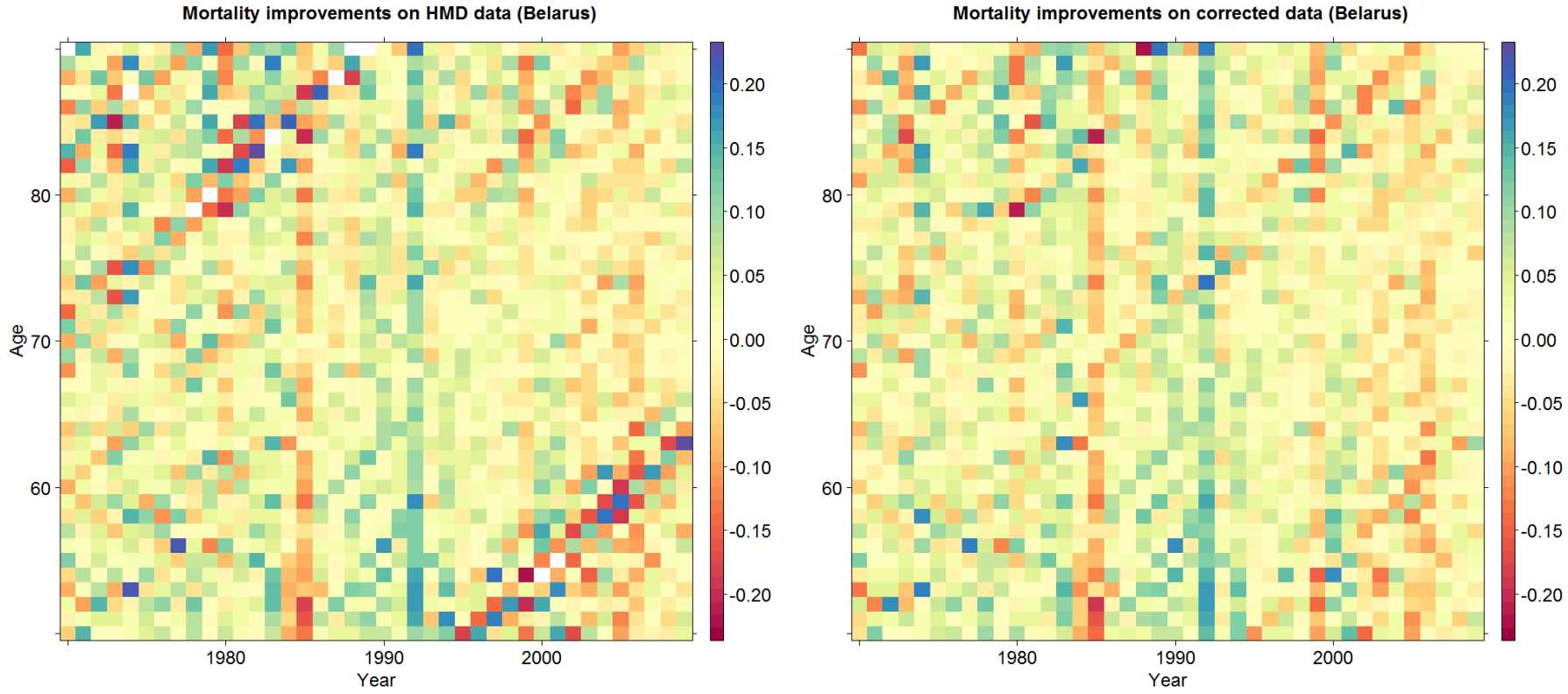
This correction method is the basis of the detailed results presented in the next section

Agenda

1	Context
2	Data correction for countries with insufficient fertility histories
3	**Detailed results by country**

Detailed results per country
Belarus

- Crude (left) and corrected (right) mortality improvements



Detailed results per country
Bulgaria

- Crude (left) and corrected (right) mortality improvements

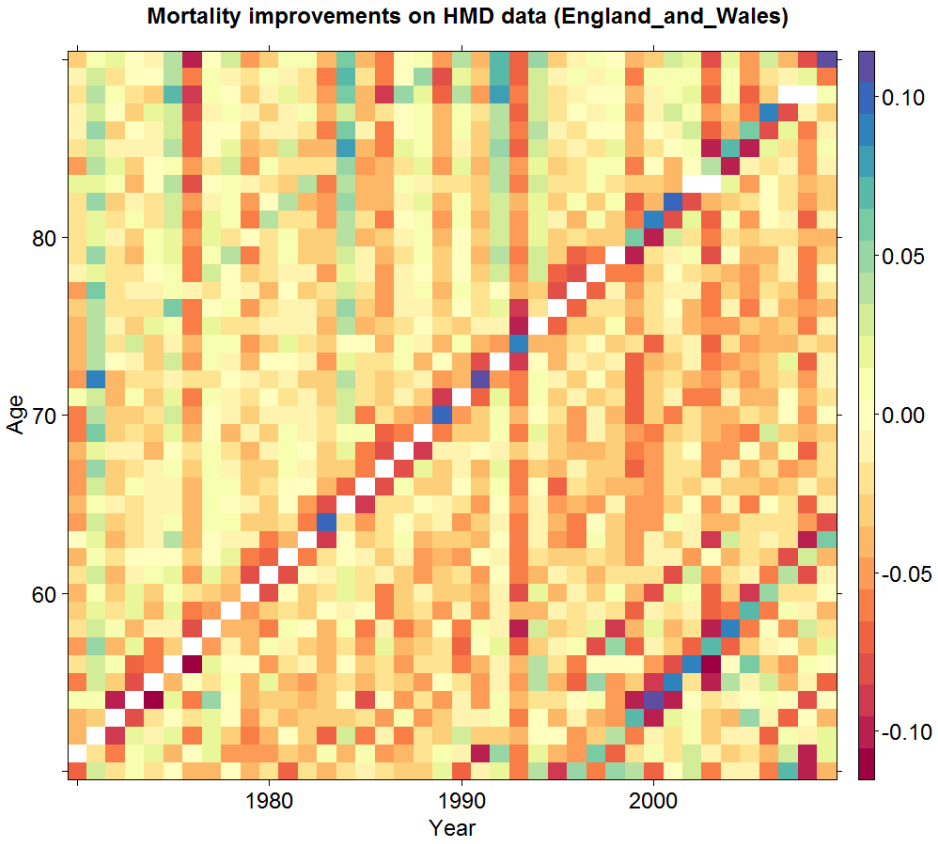
Detailed results per country
Czech Republic

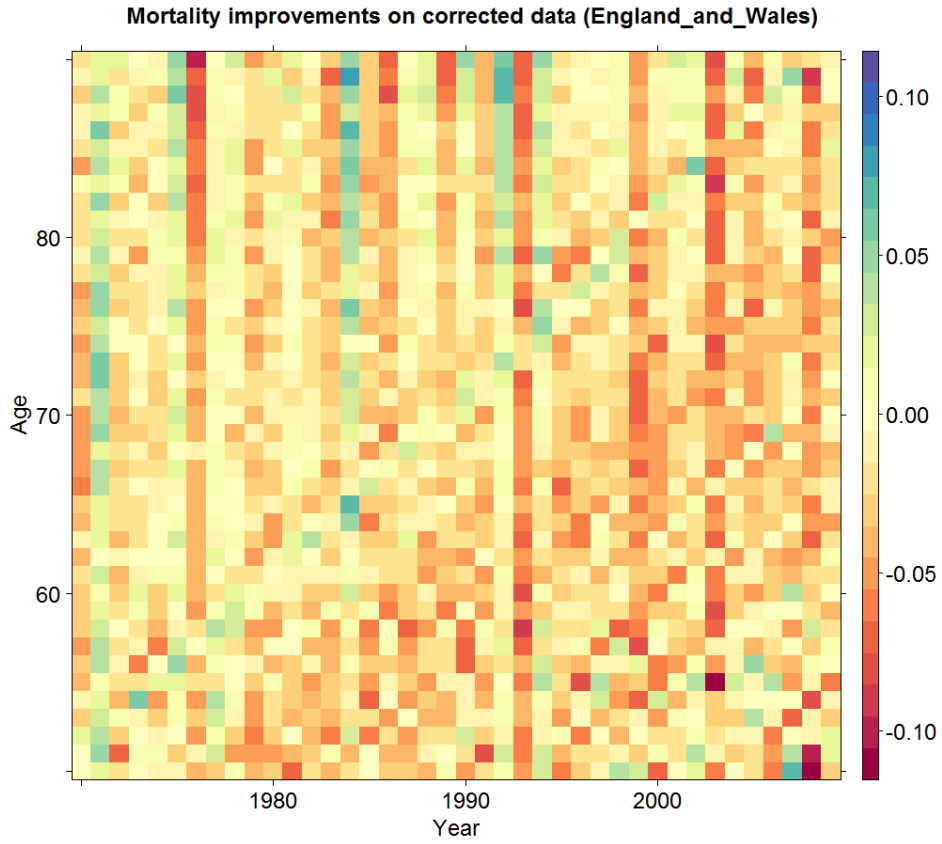
- Crude (left) and corrected (right) mortality improvements

Detailed results per country

England & Wales

- Crude (left) and corrected (right) mortality improvements

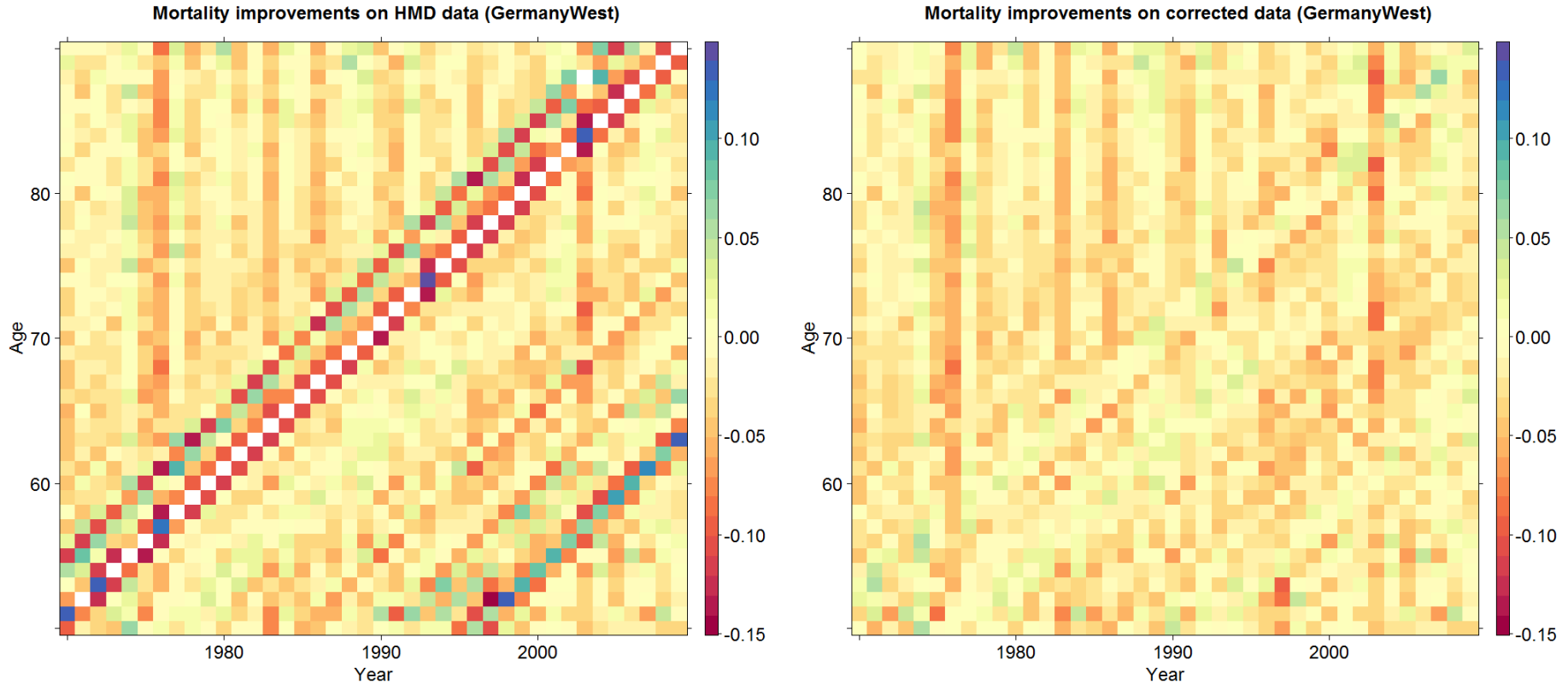




Detailed results per country

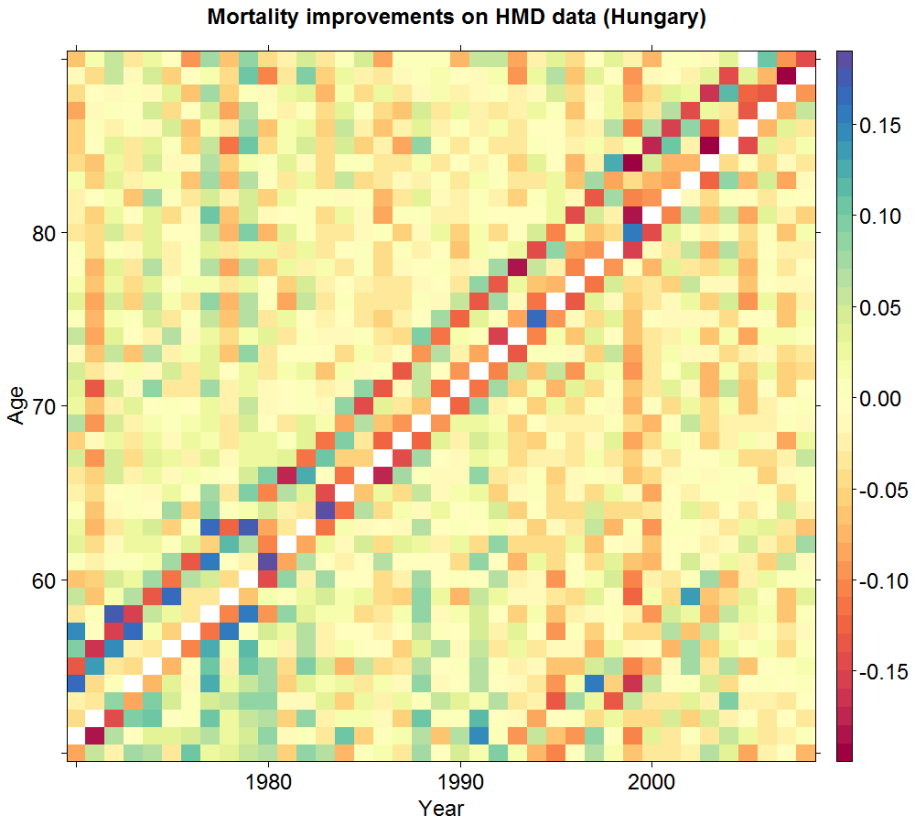
Germany West

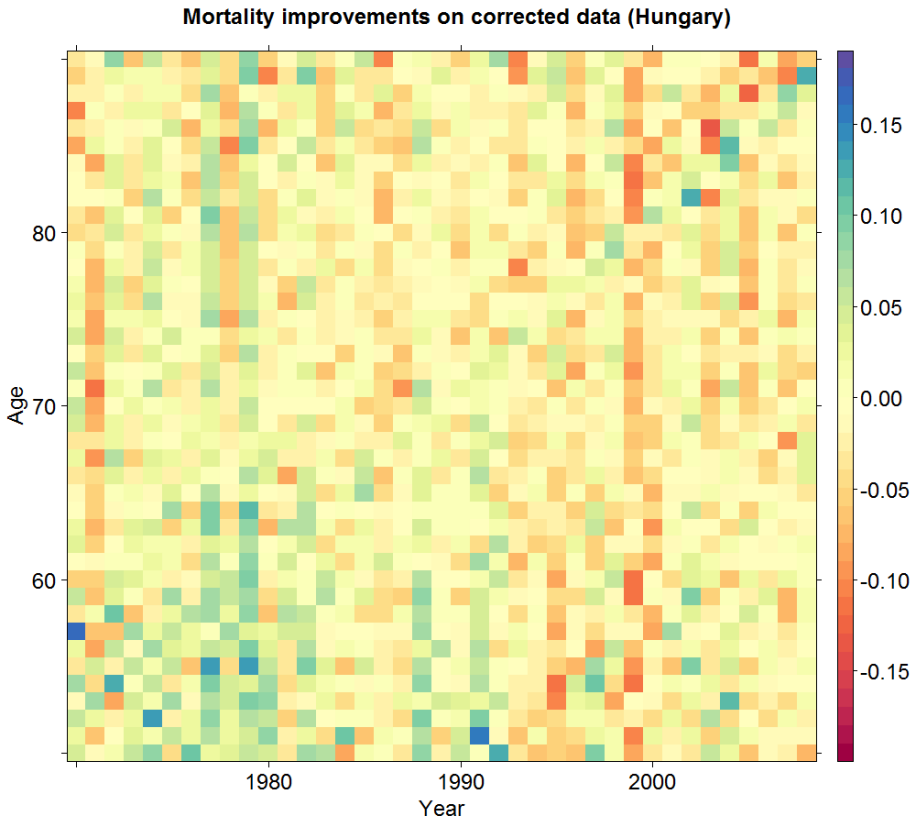
- Crude (left) and corrected (right) mortality improvements



Detailed results per country
Hungary

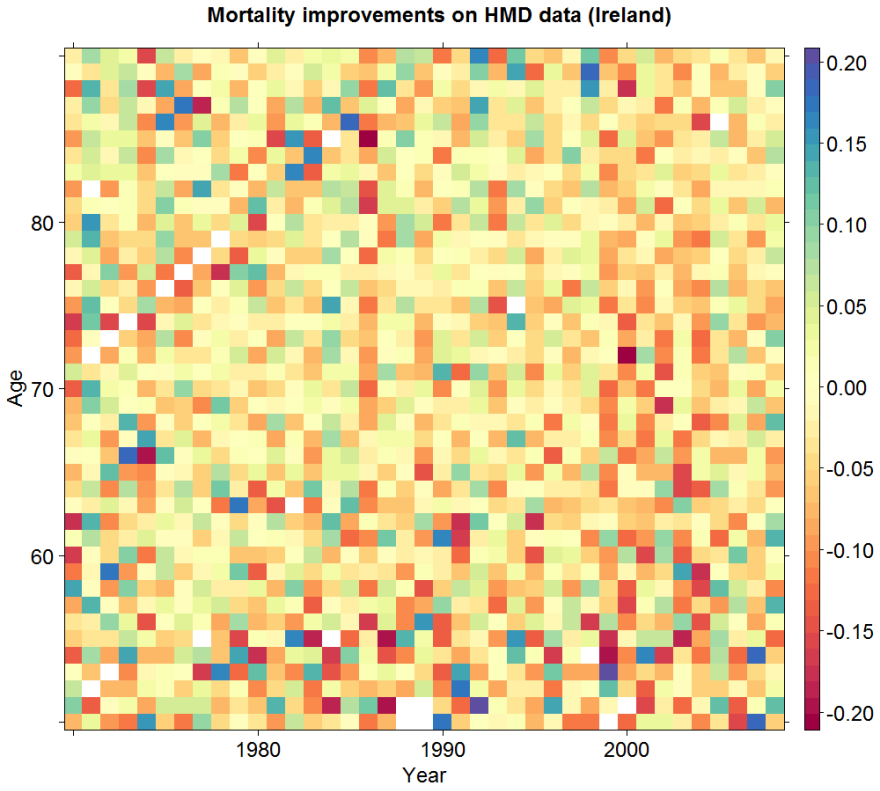
- Crude (left) and corrected (right) mortality improvements





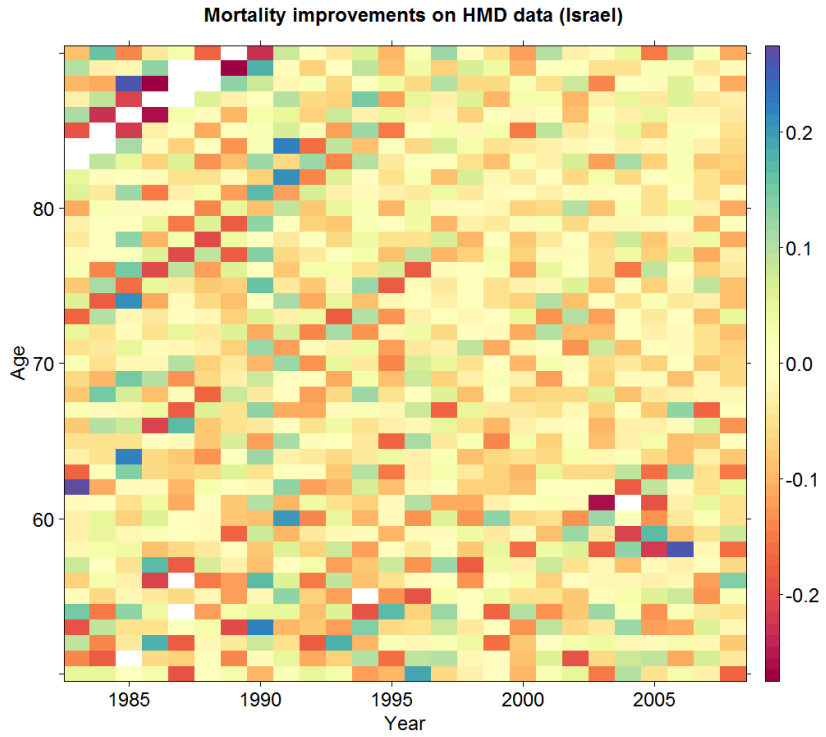
Detailed results per country
Ireland

- Crude (left) and corrected (right) mortality improvements



Detailed results per country
Israël

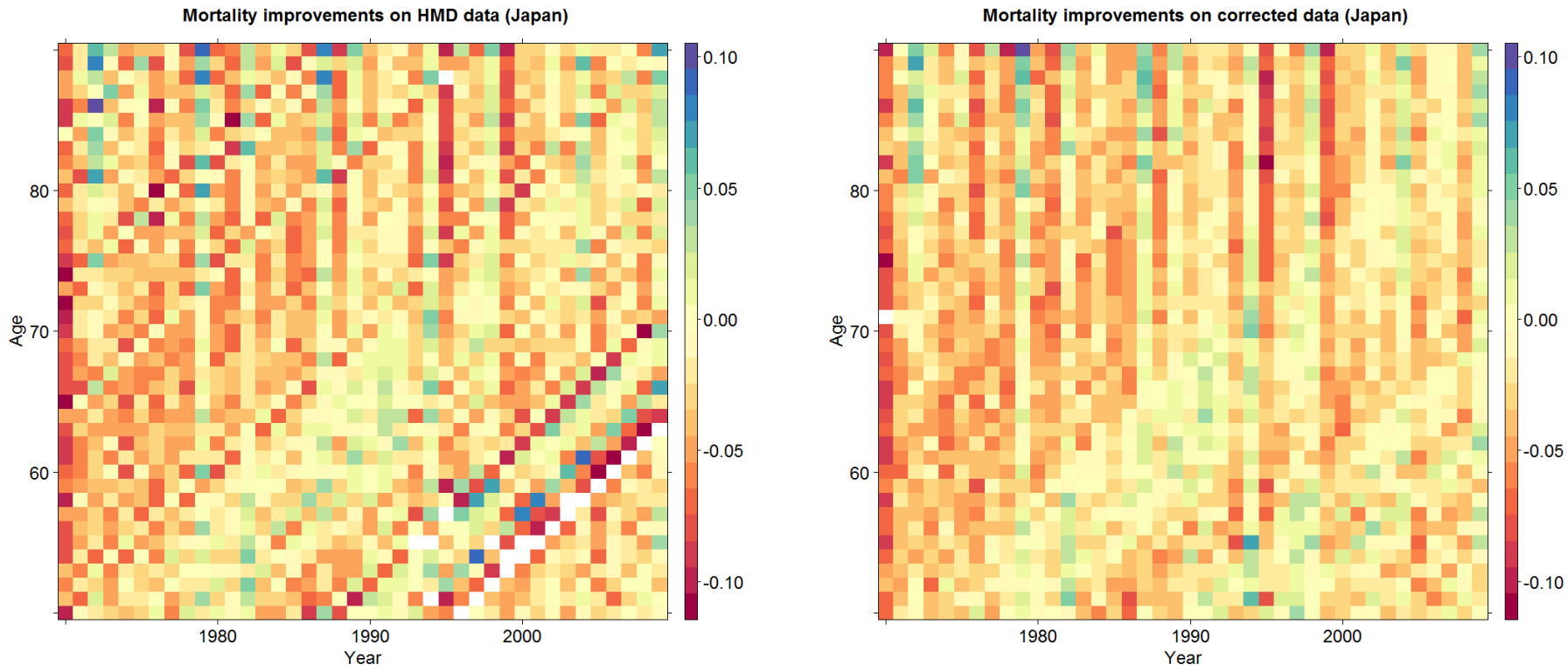
- Crude (left) and corrected (right) mortality improvements



Detailed results per country

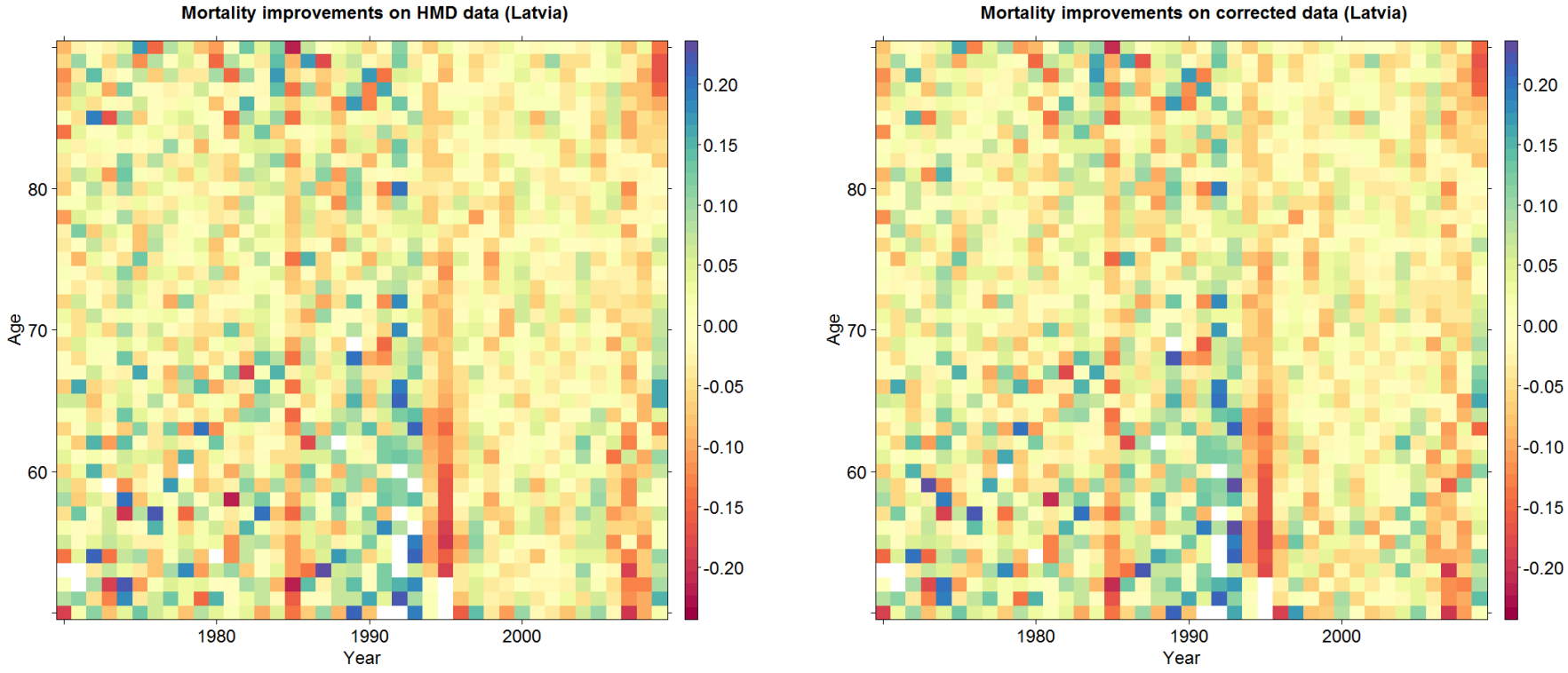
Japan

- Crude (left) and corrected (right) mortality improvements



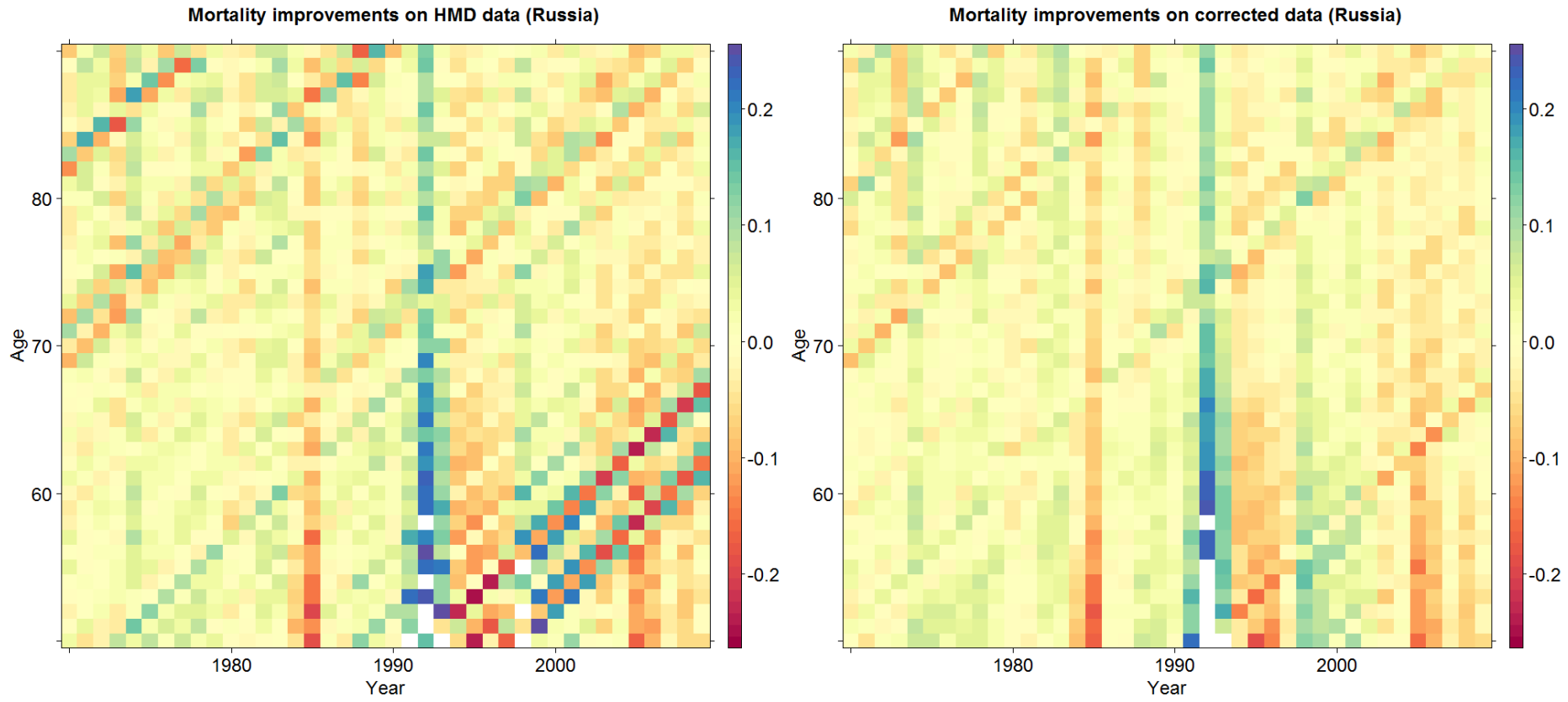
Detailed results per country
Latvia

- Crude (left) and corrected (right) mortality improvements



Detailed results per country
Russia

- Crude (left) and corrected (right) mortality improvements



Detailed results per country
Slovakia

- Crude (left) and corrected (right) mortality improvements

Detailed results per country
Slovenia

- Crude (left) and corrected (right) mortality improvements

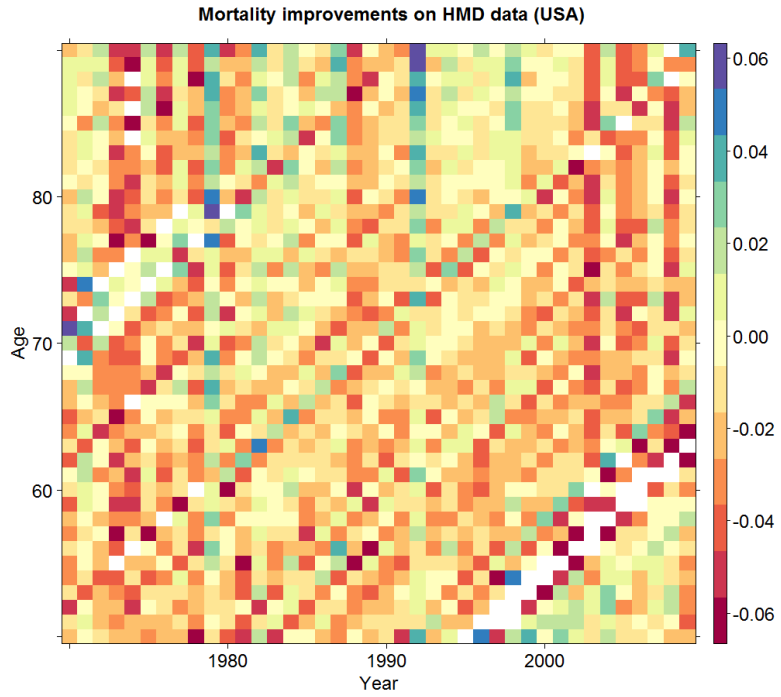
Detailed results per country

Ukraine

- Crude (left) and corrected (right) mortality improvements

Detailed results per country
USA

- Crude (left) and corrected (right) mortality improvements



References

- http://fr.milliman.com/insight/2017/Reliability-issues-in-the-construction-of-national-mortality-tables-for-the-general-population-What-you-should-know/
- http://www.milliman.com/insight/2017/New-developments-in-the-computation-of-mortality-rates-An-actuarys-bread-and-butter/
- A. Boumezoued, 2016. Improving HMD mortality estimates with HFD fertility data. HAL preprint: https://hal.archives-ouvertes.fr/hal-01270565v1 (North American Actuarial Journal)
- F. Balland, A. Boumezoued, L. Devineau, M. Habart, T. Popa. 2018. Mortality data reliability in an internal model. In revision. https://arxiv.org/abs/1803.00464
- A. Boumezoued, M. Hoffmann, P. Jeunesse. 2018. A new inference strategy for general population mortality tables. In revision. https://hal.archives-ouvertes.fr/hal-01773665
- A. Boumezoued, M. Hoffmann, P. Jeunesse. 2018. Non-parametric adaptative inference of birth and death models in a large population limit. Submitted. http://arxiv.org/abs/1903.00673Forthcoming
- A.J.G. Cairns, D. Blake, K. Dowd and A.R. Kessler. 2016. Phantoms Never Die: Living with Unreliable Population Data. To appear in Journal of the Royal Statistical Society, Series A.
- Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on October 2015).
- Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available at www.humanfertility.org (data downloaded on October 2015).
- S.J. Richards. 2008. Detecting year-of-birth mortality patterns with limited data. Journal of the Royal Statistical Society: Series A (Statistics in Society) , 171(1): 279-298.
- J.R. Wilmoth, K. Andreev, D. Jdanov and D.A. Glei. 2007. Methods Protocol for the Human Mortality Database. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality.org [version 31/05/2007].

Milliman

Thank you

alexandre.boumezoued@milliman.com