A reliability challenge for the Human Mortality Database

Longevity 12 Conference, Chicago 29-September-2016

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

Agenda

1	**Context**
2	A look at the Human Mortality Database
3	Correcting population exposures with fertility data
4	What can be learned from corrected mortality tables?
5	Conclusion and next steps

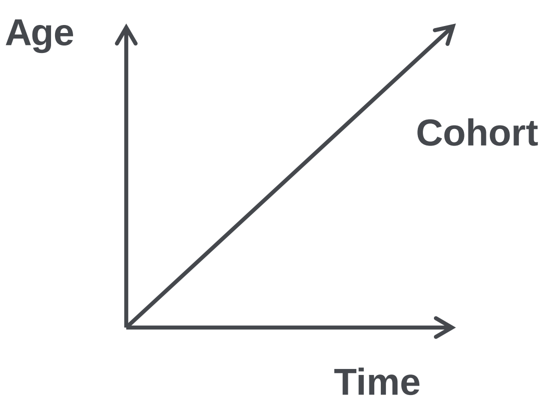
Context

National mortality tables

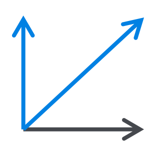
- How to estimate the **mortality rate** based on national population data?
  - The statistical inference of a death rate with two crossing dimensions (age and time) is an **old (Lexis, 1875) and still challenging estimation problem**
- In practice, individuals are grouped into **age and time blocks**,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

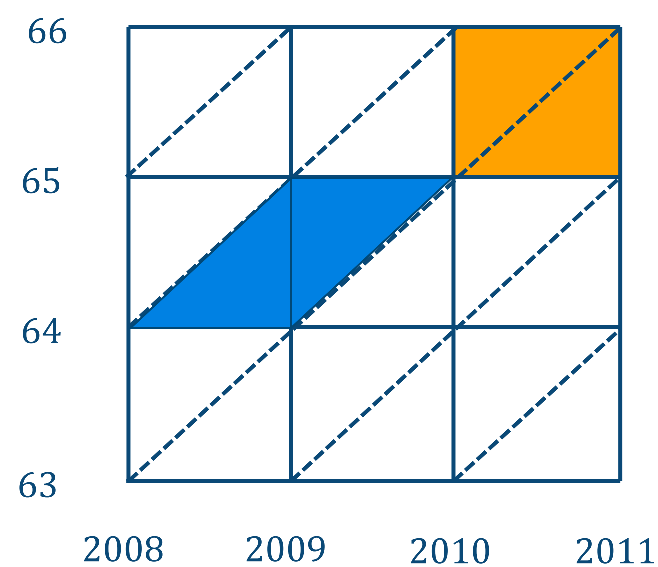
Period tables
=
Death rate assumed constant over squares

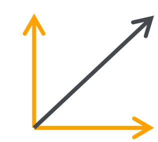
Three directions of analysis



Cohort tables
=
Death rate assumed constant over parallelograms



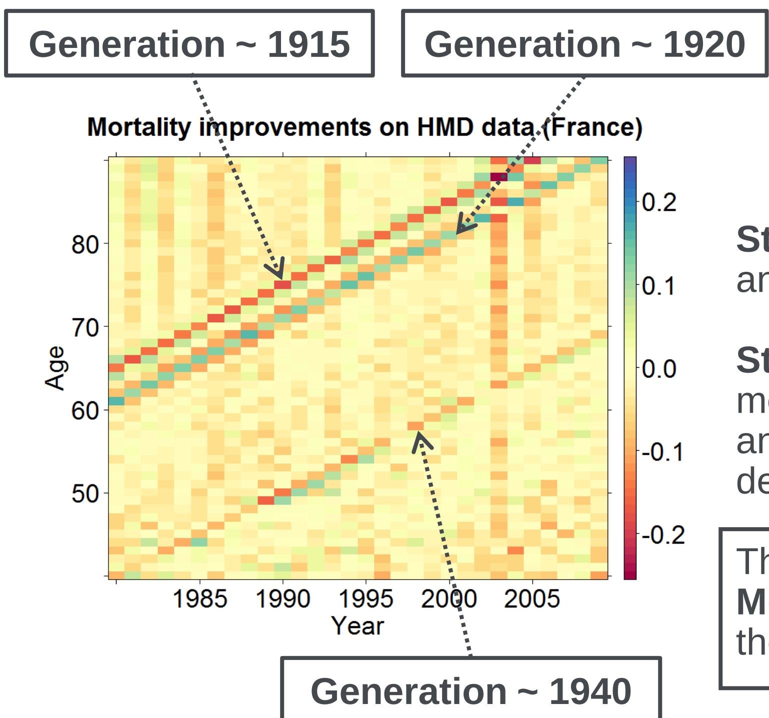




Context

Mortality improvement rates and “cohort effects” (ex: France)

- **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

Step 1: Richards (2008) conjectured that the 1919 cohort effect for England and Wales is an **anomaly** in the mortality table due to **erratic birth patterns**

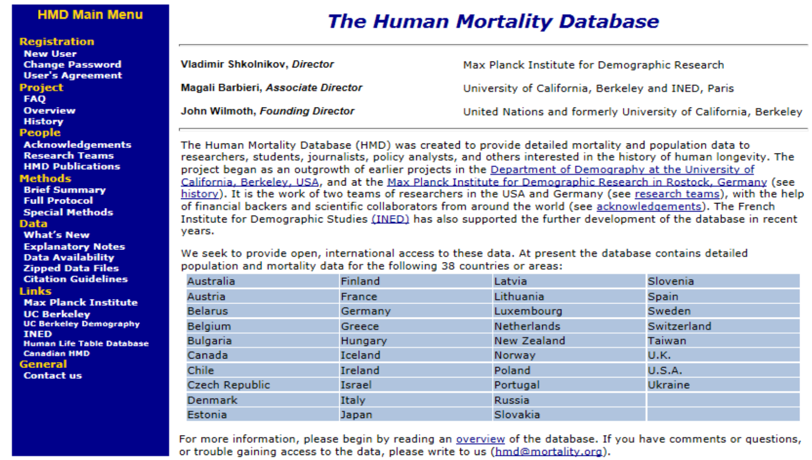
Step 2: Cairns, Blake, Dowd & Kessler (2016) analyzed the ONS methodology and confirmed the conjecture by Richards; they used England and Wales **monthly fertility data** to detect anomalies in the computation of death rates

This talk is based on Boumezoued (2016) and focuses on the **Human Mortality Database**, showing that **these anomalies are universal** and that the **Human Fertility Database** can be processed to correct such errors

Context

A reference: the Human Mortality Database*

- Since its launch in 2002, the Human Mortality Database has become the **reference provider** of mortality estimates (both period and cohort tables) given in an homogenous format for several countries



- Possible anomalies in **period mortality tables** are already suggested in the HMD technical note *(Wilmoth et al. 2007):*

" The assumption [of uniform distribution of births] is violated most severely in situations where there are rapid changes in the size of successive cohorts, owing to **fluctuations in the birth series many years before**. The worst situation is when a sharp discontinuity in births occurs in the middle of one calendar year, creating a cohort that is "heavy" at one end and "light" at the other. **We have not attempted to correct our mortality estimates for the error introduced by such occurrences, which may result in artificially elevated or depressed levels of mortality along a diagonal of the Lexis diagram that follows the cohort(s) in question. The user should be aware of this possibility and not misinterpret the data.**"

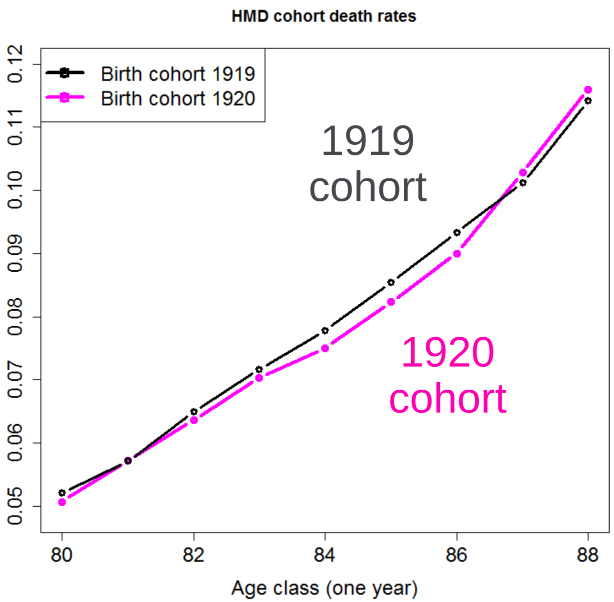
**Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on October 2015).*

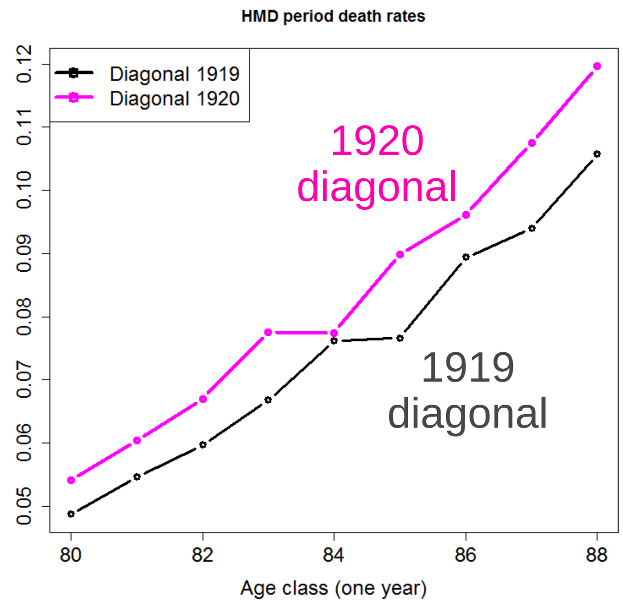
Agenda

1	Context
2	**A look at the Human Mortality Database**
3	Correcting population exposures with fertility data
4	What can be learned from corrected mortality tables?
5	Conclusion and next steps

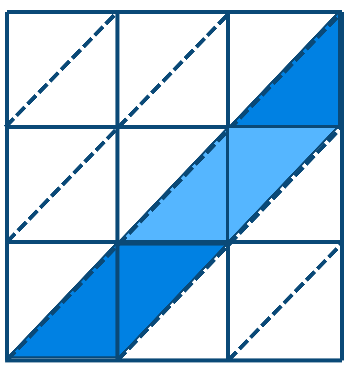
A look at the Human Mortality Database

Inversion of cohort and period tables: example for France (1/2)

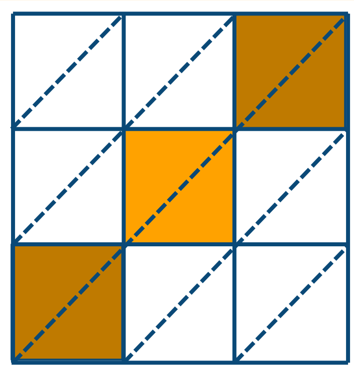




Cohort table

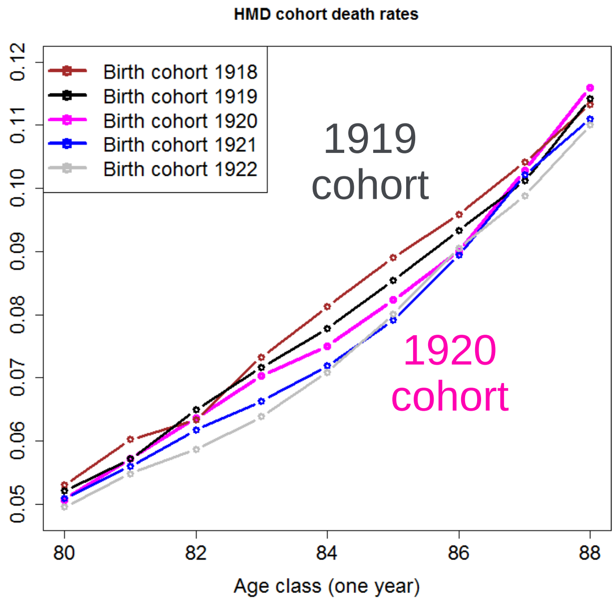


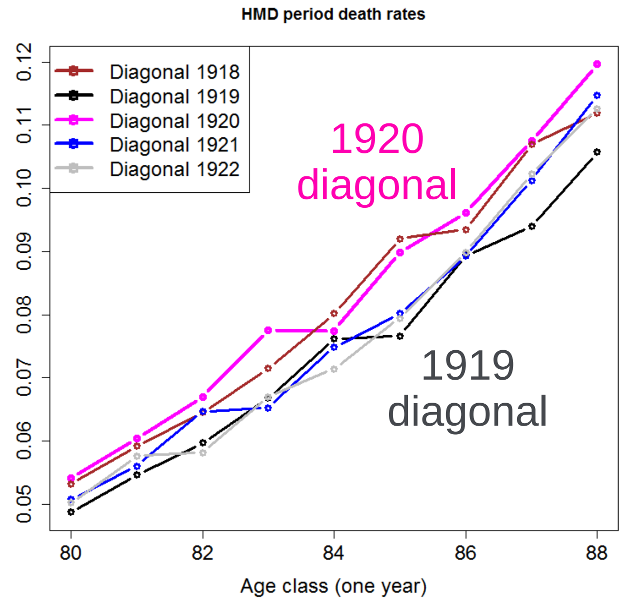
Period table



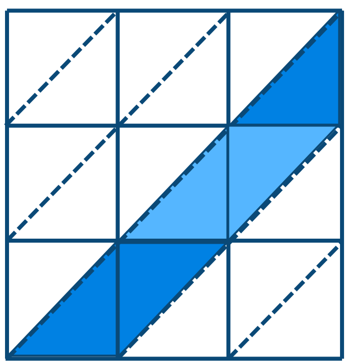
A look at the Human Mortality Database

Inversion of cohort and period tables: example for France (2/2)

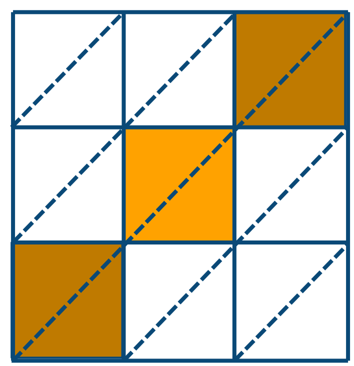




Cohort table



Period table



A look at the Human Mortality Database

How to properly estimate the period death rate?

- Quantity of individuals with **exact age a at exact time s: $g(a,s)$**
- The **death rate $\mu(a,s)$** drives the evolution of each cohort over time
  - Let $g(0,s)$ be given (the number of newborns at time s)
  - The number of survivors at age a in the cohort born at time s is given as:

$$g(a,s+a) = g(0,s)\exp\left\{-\int_0^a \mu(u,s+u)du\right\}$$

  - Differentiation by age and time leads to the **aging term** of the McKendrick-Von Foerster equation (1926)

$$(\partial_a + \partial_s)g(a,s) = -\mu(a,s)g(a,s)$$

= $d(a,s)$ = Number of deaths at (exact) age a and time s

- **Statistical estimation:** the period death rate is assumed to be constant on squares
  - That is $\mu(a,s) = \mu(x,t)$ for each $a \in [x,x+1)$ and $s \in [t,t+1)$
  - **Under this assumption, one recovers the classical formula of the estimated death rate, as the number of deaths divided by the so-called exposure-to-risk**

$$D(x,t) = \int_t^{t+1}\int_x^{x+1} d(a,s)\,\mathrm{dads} = \mu(x,t)\int_t^{t+1}\int_x^{x+1} g(a,s)\,\mathrm{dads}$$

$$= \mu(x,t)\int_t^{t+1} \underbrace{P(x,s)}\,\mathrm{ds}$$

Exposure-to-risk = total time lived in year t by individuals aged x last birthday

Population at time s aged x last birthday

Milliman

A look at the Human Mortality Database

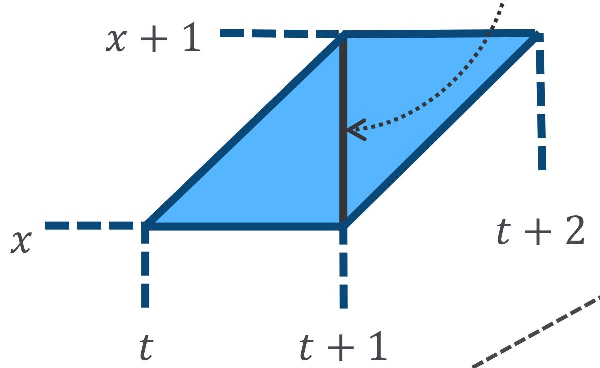
How is the exposure-to-risk estimated in the HMD?

Number of deaths → $D(x,t)$; Exposure-to-risk → $E(x,t)$

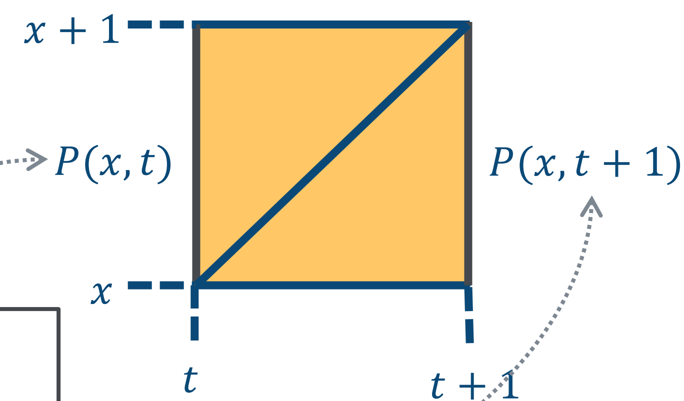
Death rate estimation : $\hat{\mu}(x,t) = D(x,t)/E(x,t)$

COHORT TABLE

HMD Cohort Exposure-to-risk = Population estimate at a given time + Small correction based on number of deaths in each triangle



PERIOD TABLE



HMD Period Exposure-to-risk = Approximation under the **assumption of uniform distribution of births** on consecutive years $\frac{1}{2}[P(x,t)+P(x,t+1)]$ + Small correction based on number of deaths in each triangle

⚠ $\frac{1}{2}[P(x,t)+P(x,t+1)] \approx \int_t^{t+1} P(x,s)ds$

Agenda

1	Context
2	A look at the Human Mortality Database
3	**Correcting population exposures with fertility data**
4	What can be learned from corrected mortality tables?
5	Conclusion and next steps

Correcting population exposures with fertility data

Processing the HFD to correct the HMD (1/3)

- **Aim:** use **monthly fertility records** from the Human Fertility Database* (HFD) to refine exposure-to-risk computation
- A **quality indicator** for each cohort is constructed in three steps:
  1. **Extract monthly fertility estimates** $P(0,s)$ from the HFD: this is the number of individuals born in month s
  2. **Refine (annual) exposure-to-risk computation** by using monthly estimates instead of annual:
  $$E(0,t) = \int_t^{t+1} P(0,s)\,\mathrm{d}s \approx \sum_{i=0}^{12} w_i\, P\left(0, t + \frac{i}{12}\right)$$
  3. **Compare it to the annual approximation** as performed in the HMD:

An indicator $\boldsymbol{I(t)}$ **far from 1** shows that period mortality rates for the cohort born in year t is **not reliable**

$$I(t) = \frac{E(0,t)}{\frac{1}{2}\left[P(0,t) + P(0,t+1)\right]}$$

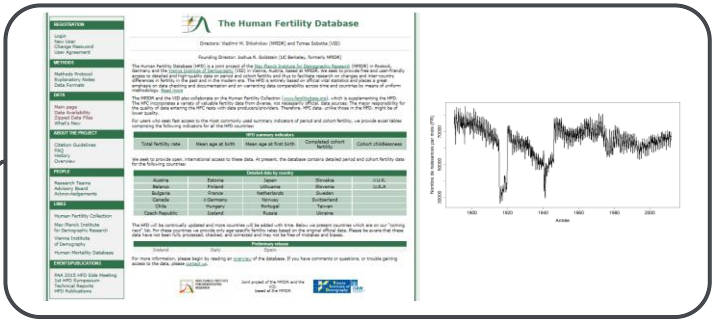
The methodology focuses on age zero, then adjusts the mortality table along the diagonal (cohort) – see the analogy in Cairns, Blake, Dowd & Kessler (2016) for the detection of anomalies

**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).*

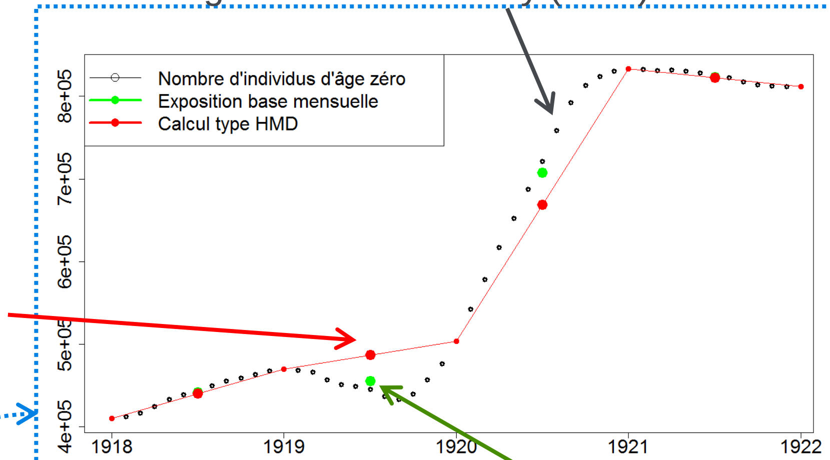
Correcting population exposures with fertility data

Processing the HFD to correct the HMD (2/3)

*Extract monthly fertility records from the **Human Fertility Database***

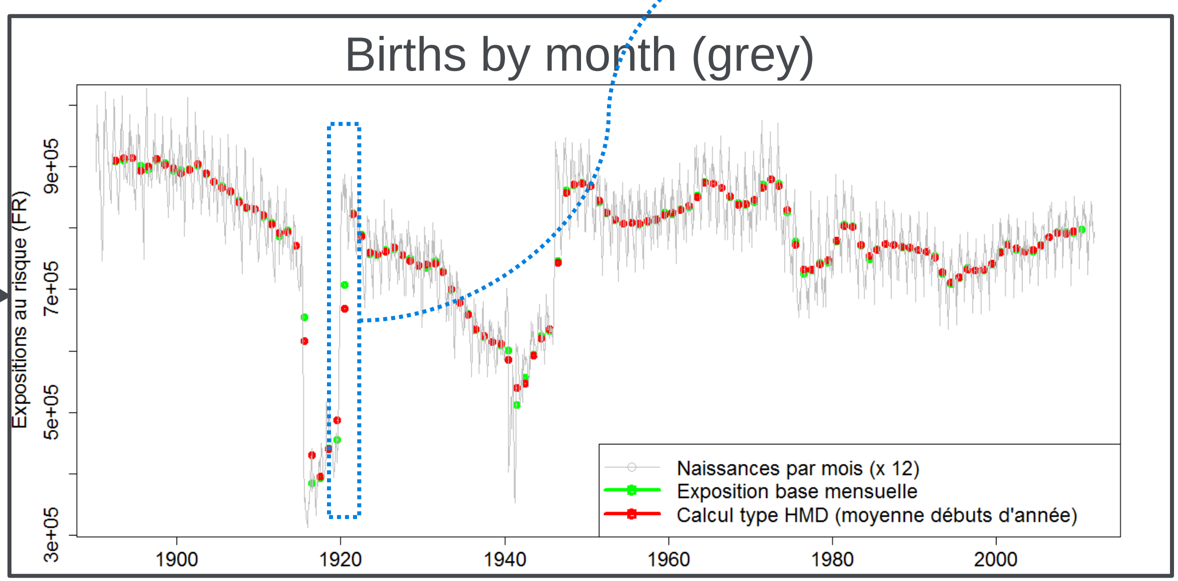


Monthly population size with age zero last birthday (HFD)



HMD-type exposure-to-risk = linear interpolation

Refined exposure-to-risk (based on HFD)



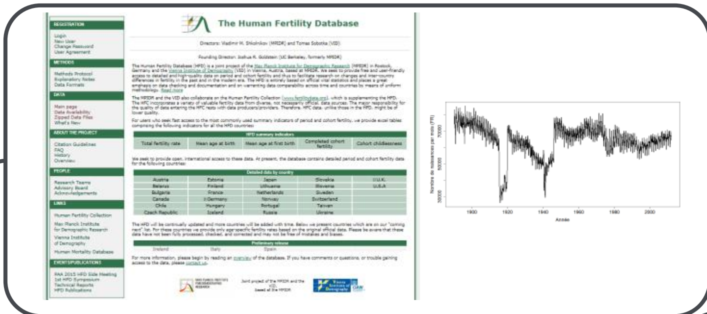
The **bias** in exposure-to-risk computation is high in periods in which births are fluctuating

Explanation: shocks in birth patterns create **convexity** in population numbers => HMD linear approximation is no longer valid

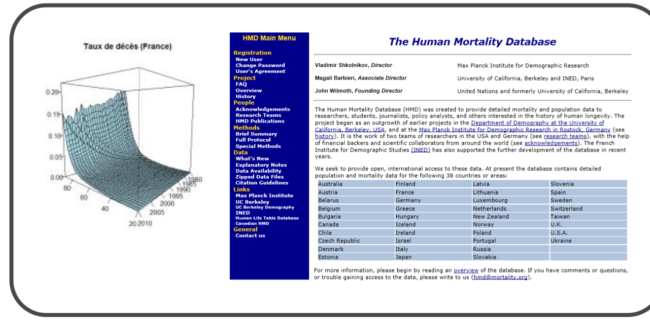
Correcting population exposures with fertility data

Processing the HFD to correct the HMD (3/3)

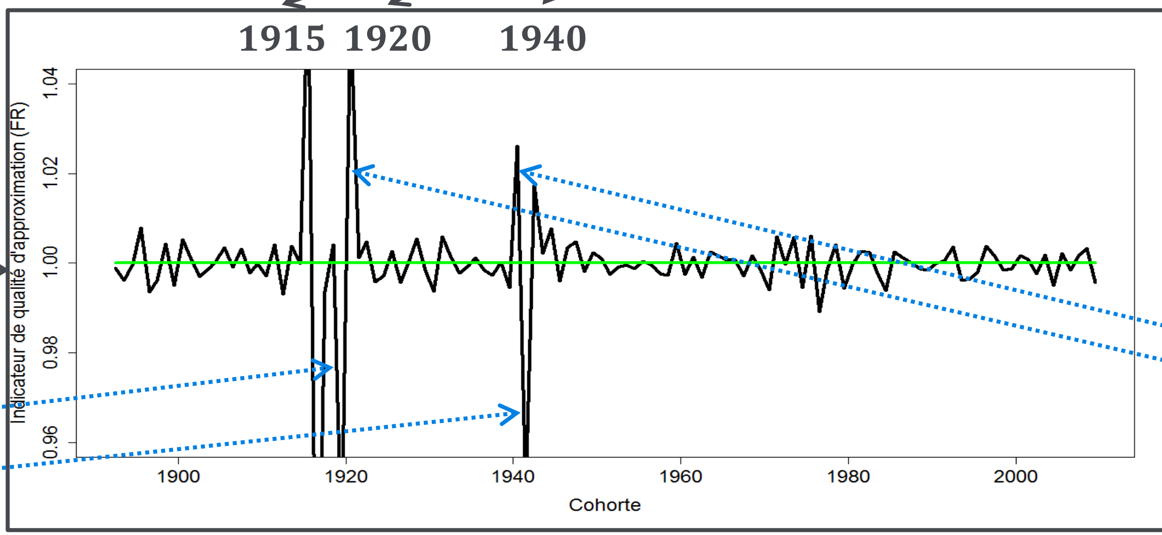
*Extract monthly fertility records from the **Human Fertility Database***



*Correction of period mortality tables in the **Human Mortality Database***



Cohorts with major anomalies (born around)



HMD Period death rates are **under-estimated** for the cohorts (as 1919)

HMD Period death rates are **over-estimated** for the cohorts (as 1920)

Quality indicator for each year of birth

Milliman

Agenda

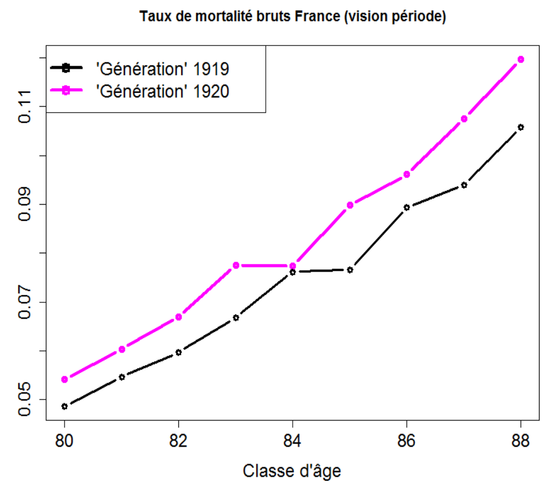
1	Context
2	A look at the Human Mortality Database
3	Correcting population exposures with fertility data
4	**What can be learned from corrected mortality tables?**
5	Conclusion and next steps

What can be learned from corrected tables?

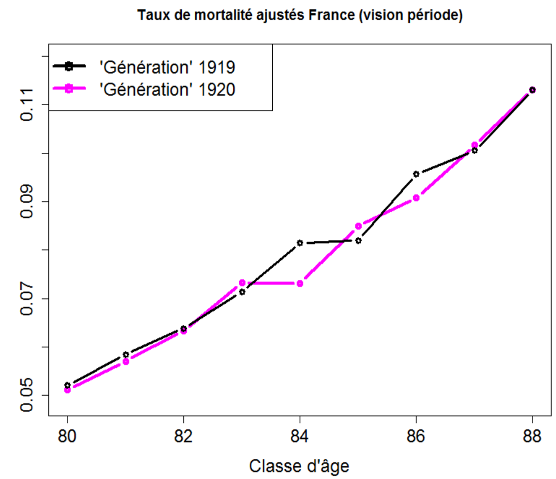
Corrected period mortality tables – focus on 1919-1920 cohorts

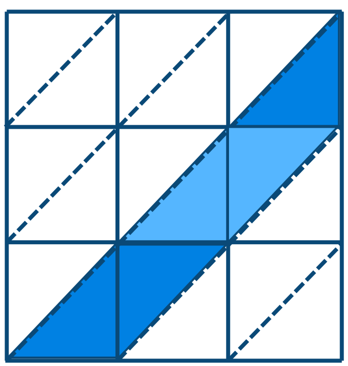
HMD cohort table

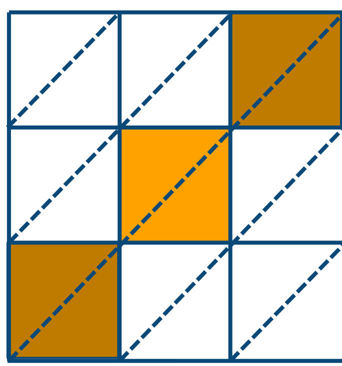
HMD period table



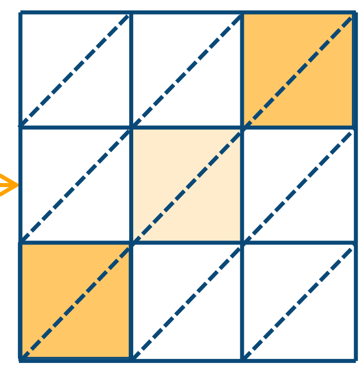
Corrected period table







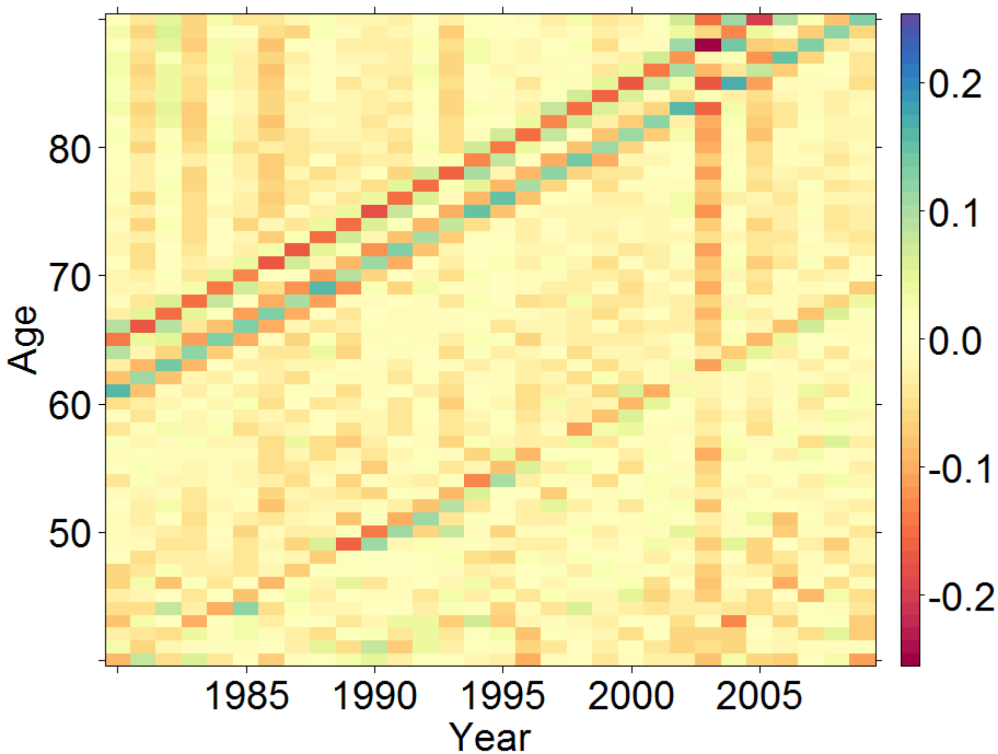
Correction based on the quality indicator $I(t)$



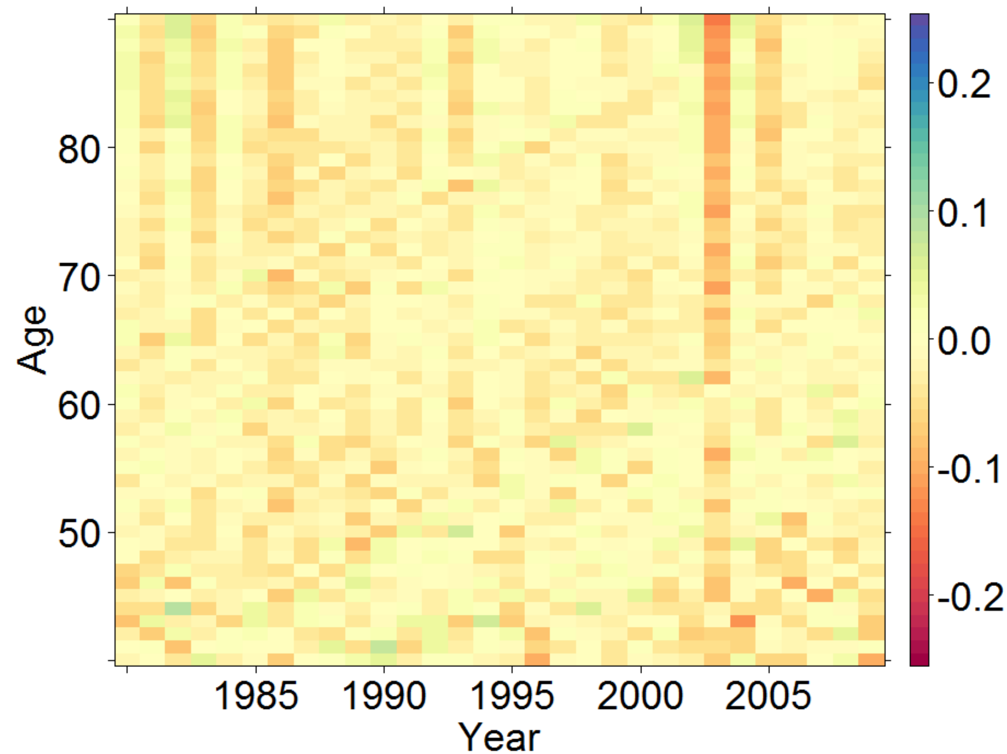
What can be learned from corrected tables?

Isolated cohort effects as data anomalies

Mortality improvements on HMD data (France)



Mortality improvements on corrected data (France)

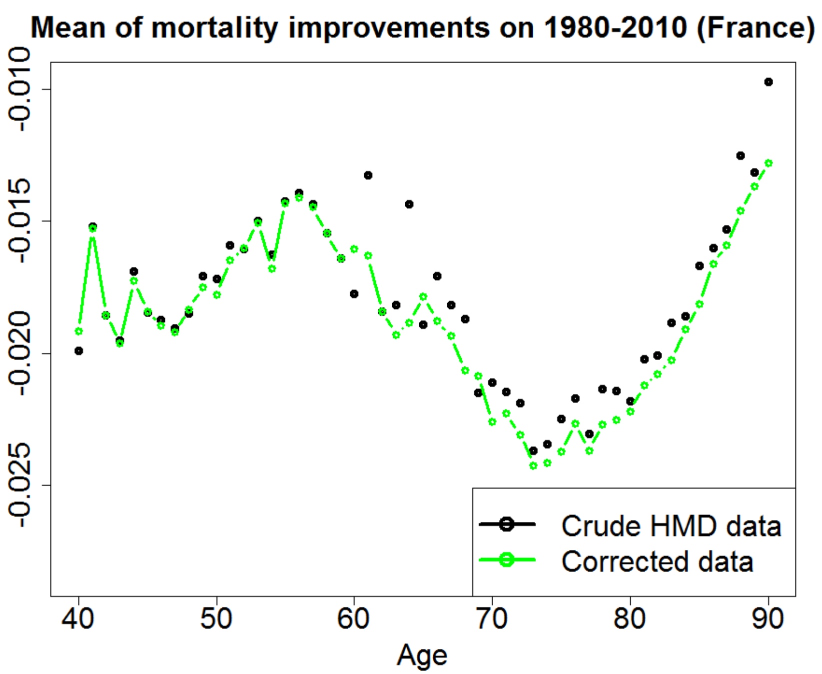


$$r(x,t) = \frac{\mu(x,t+1) - \mu(x,t)}{\mu(x,t)}$$

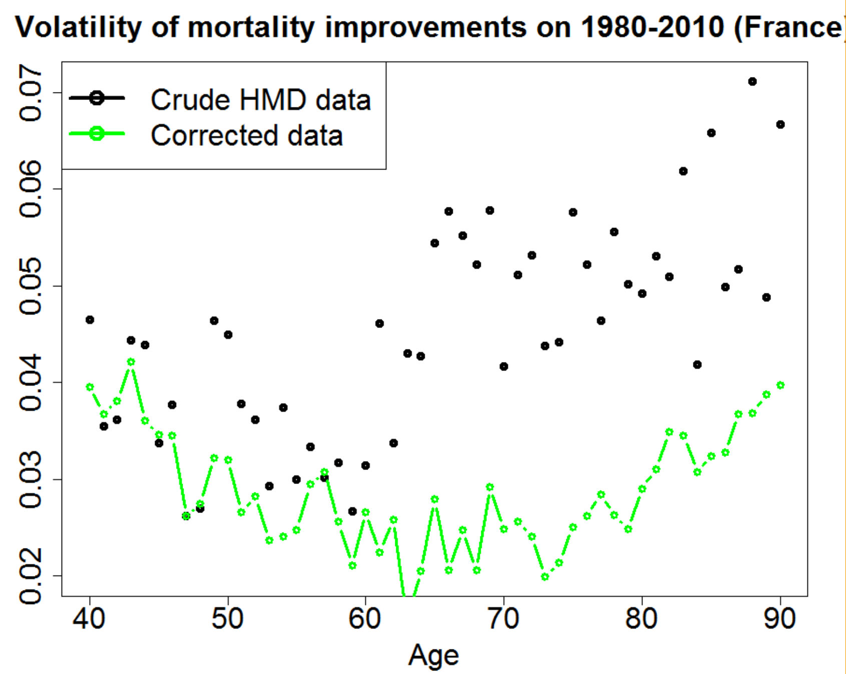
What can be learned from corrected tables?

Improvement rates properties after data correction

Average mortality improvements by age are smoothed after data correction



Volatility of mortality improvement is reduced at high ages after correction



What can be learned from corrected tables?

Choice and fitting of classical mortality models*

Conclusion 1

The estimated parameters are close

Conclusion 2

The residuals do not embed clear cohort trends anymore

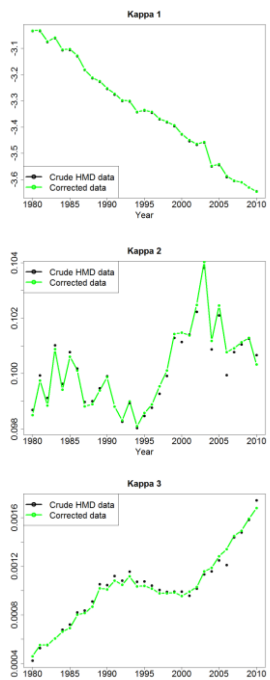
Conclusion 3

The historical volatility is better reproduced

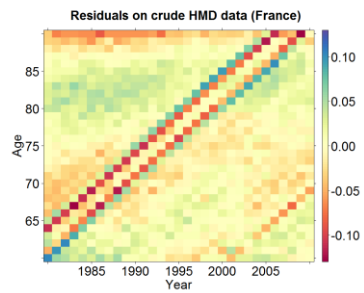
$\kappa_1(t)$

$\kappa_2(t)$

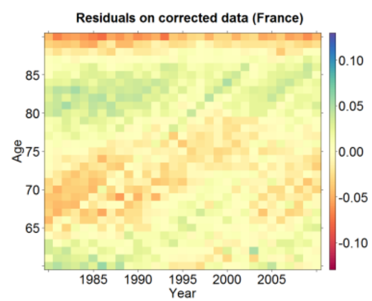
$\kappa_3(t)$



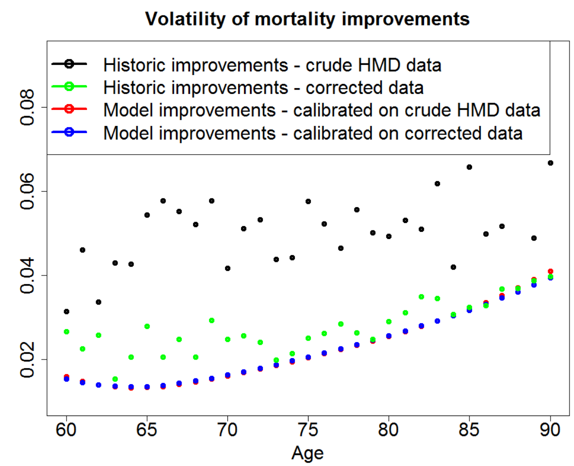
– **Residuals on crude data:**



– **Residuals on corrected data:**



– **The historical mortality volatility is better reproduced by the classical model:**



Milliman

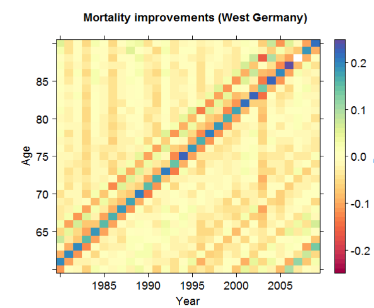
*Example based on the model: $\ln(\mu(x,t)) = \kappa_1(t) + \kappa_2(t)(x-\bar{x}) + \kappa_3(t)\left((x-\bar{x})^2 - \widehat{\sigma_x}^2\right) + \epsilon(x,t)$

Agenda

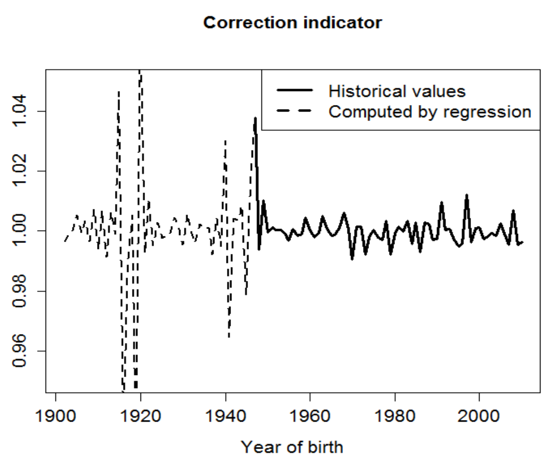
1	Context
2	A look at the Human Mortality Database
3	Correcting population exposures with fertility data
4	What can be learned from corrected mortality tables?
5	**Conclusion and next steps**

Conclusion and next steps
Extending the scope of countries

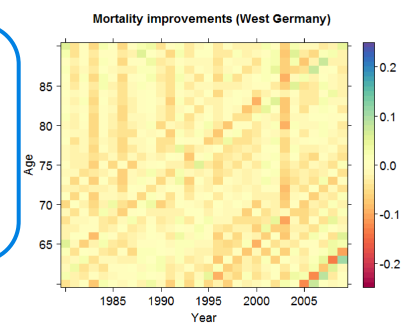
- The present methodology allows to correct a given cohort if the associated date of birth is included in the monthly fertility history from the **Human Fertility Database**
  - A **reasonable scope of countries** for which the 1919-1920 anomaly can be corrected includes: Austria, Finland, France, Italy, Sweden, Switzerland,…
- **Issue**: for many countries, the fertility historical depth is **not sufficient** to correct crucial anomalies as those for generations 1919-1920
- **Idea**: develop a method to **reconstruct the quality indicator** for those countries
  - Illustration below: first insights for the example of Germany



Step 1: Key countries are used to reconstruct the adjustment ratio history



Step 2: The original method then allows to correct the abnormal 1919-1920 effect



References

- A. Boumezoued, 2016. **Improving HMD mortality estimates with HFD fertility data.** HAL preprint: https://hal.archives-ouvertes.fr/hal-01270565v1
- H. Bensusan, A. Boumezoued, N. El Karoui, and S. Loisel. 2015. **Bridging the gap from microsimulation practice to population models: a survey.** Preliminary version as Chapter 2 of A. Boumezoued PhD thesis available at https://tel.archives-ouvertes.fr/tel-01307921/document.
- 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