A very brief overview of the HMD and extensions, USMDB and USFDB

Magali Barbieri

magali@berkeley.edu

University of California, Berkeley, and
French National Institute for Demographic Studies (INED)

Berkeley Formal Demography Workshop, 3 June 2025, Berkeley

Background on the HMD

- **Overall goal**:

  To provide detailed mortality and population data free of charge to all persons interested in the history of human longevity

- **75,000+** registered **users** :

  Researchers, students

  Journalists

  Policy analysts

  Corporations (insurance, reinsurance, pension funds, etc…)

Who is responsible for the HMD?

Three teams of researchers:

- **UC Berkeley** (Dept of Demography)
  led by Magali Barbieri, Co-Director
  (previously John Wimoth, Founding Director)
- Max Plank Institute for Demographic Research (**MPIDR**, in Rostock, Germany)
  led by Dmitri Jdanov, Co-Director
  (previously Vladimir Schkolnikov, Founding Director)
- French Institute for Demographic Studies (**INED**, in Paris, France)

What is in the HMD?

- Detailed historical data and supporting documentation for 41 national populations:
  - Death counts and estimated population exposures (person-years lived) at the finest detail possible
  - Original estimates of age-specific death rates and life tables in various formats (age x time)
- Computed using various forms of input data:
  - Death counts from national statistical offices
  - Birth counts
  - Census counts
  - Official population estimates

https://www.mortality.org

HMD HOME PROJECT PEOPLE METHODS DATA RESEARCH LINKS Registration Login

Human Mortality Database

Reliability and Accuracy Matter

The Human Mortality Database (HMD) is the world´s leading scientific data resource on mortality in developed countries. The HMD provides detailed high-quality harmonized mortality and population estimates to researchers, students, journalists, policy analysts, and others interested in the human longevity. The HMD follows open data principles.

- Short-Term Mortality Fluctuations
- Citing HMD
- Research Team
- Acknowledgements

Data by country or area

Australia	Denmark	Ireland	Norway	Switzerland
Austria	Estonia	Israel	Poland	Taiwan
Belarus	Finland	Italy	Portugal	U.K.
Belgium	France	Japan	Republic of Korea	U.S.A.
Bulgaria	Germany	Latvia	Russia	Ukraine
Canada	Greece	Lithuania	Slovakia	
Chile	Hong Kong	Luxembourg	Slovenia	
Croatia	Hungary	Netherlands	Spain	
Czechia	Iceland	New Zealand	Sweden	

News

Netherlands, Spain and Taiwan revised and updated through 2021.

All News

HMD in zipped data files

To facilitate rapid downloads, the database has been organized into zipped data files.

Zipped Data Files

U.K.

United Kingdom Total Population

Warning: The data given here represent only the civilian population. During the period between pre-war and post-war census (1939-50), population estimates exclude the military while death counts exclude military deaths that occurred abroad. For all other years, the period data in this series represent the total population. For details, please see the Background and Documentation.

Other data series for this country

- England & Wales Total Population
- England & Wales Civilian Population
- Scotland
- Northern Ireland

- Background and documentation
- Data sources
- All data in one zip file

Complete Data Series

Explanatory notes

	Available dates	Age interval x Year interval					
		1x1	1x5	1x10	5x1	5x5	5x10
Period data							
Births	1922 - 2020	1-year					
Deaths	1922 - 2020	1x1	1x5	1x10	5x1	5x5	5x10
Deaths by Lexis triangles	1922 - 2020	Lexis					
Population size	1922 - 2021	1-year			5-year		
Exposure-to-risk	1922 - 2020	1x1	1x5	1x10	5x1	5x5	5x10
Exposure-to-risk by Lexis triangles	1922 - 2020	Lexis					
Death Rates	1922 - 2020	1x1	1x5	1x10	5x1	5x5	5x10
Life tables	1922 - 2020						
Females		1x1	1x5	1x10	5x1	5x5	5x10
Males		1x1	1x5	1x10	5x1	5x5	5x10
Total (both sexes)		1x1	1x5	1x10	5x1	5x5	5x10
Life expectancy at birth	1922 - 2020	1-year	5-year	10-year			
Cohort data							
Exposure-to-risk	1842 - 1990	1x1	1x5	1x10	5x1	5x5	5x10
Death Rates	1842 - 1990	1x1	1x5	1x10	5x1	5x5	5x10

Input Data

Explanatory notes

	Available dates	Data	Lexis map
Births	-	txt	
Births by month	Various	txt	
Deaths	-	txt	html
Population size	-	txt	html

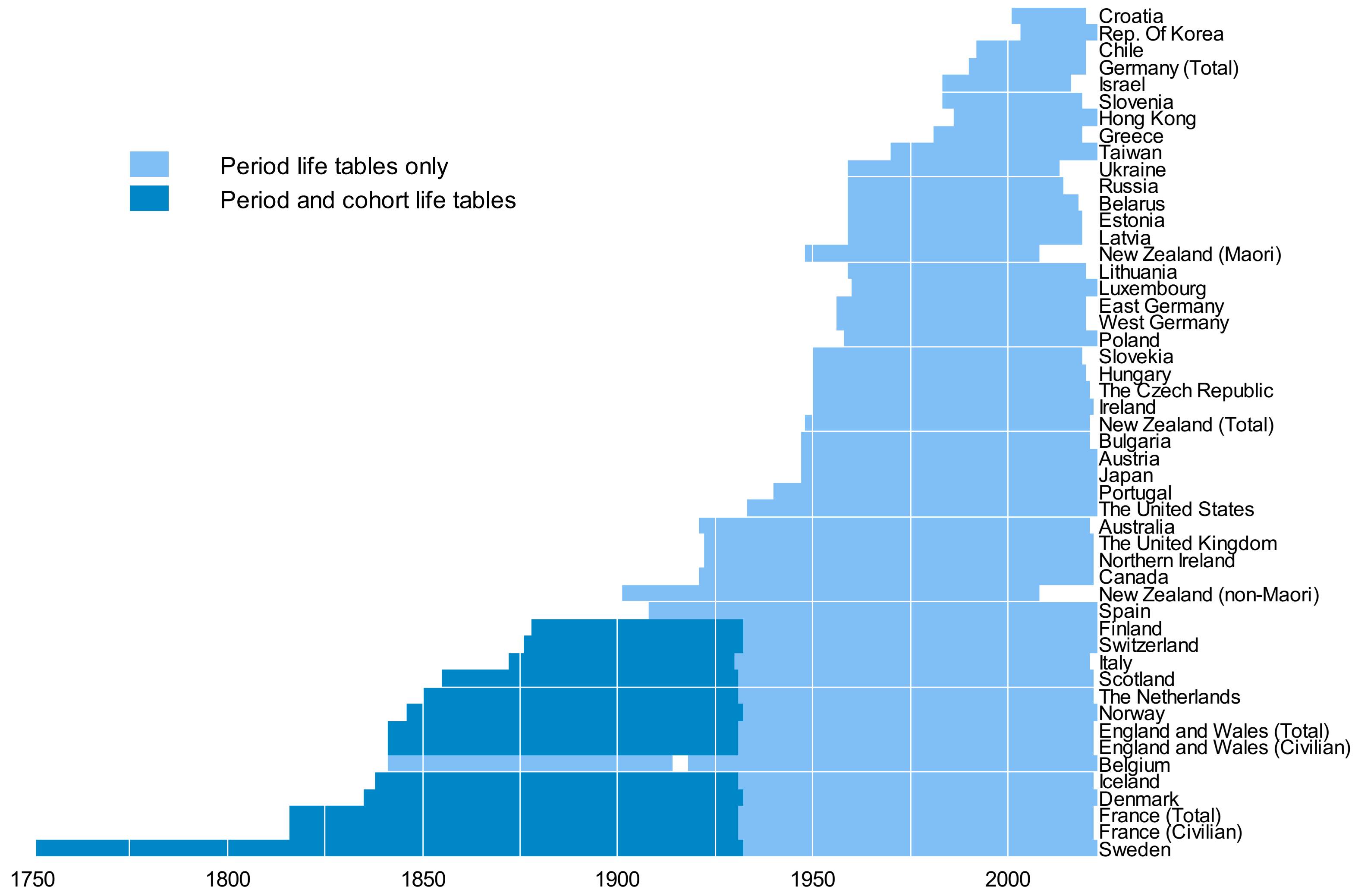
United Kingdom, Life tables (period 1x1), Total Last modified: 27 May 2022; Methods Protocol: v6 (2017)

Year	Age	mx	qx	ax	lx	dx	Lx	Tx	ex
1922	0	0.08150	0.07713	0.31	100000	7713	94643	5702262	57.02
1922	1	0.02613	0.02579	0.50	92287	2380	91096	5607619	60.76
1922	2	0.01257	0.01249	0.50	89906	1123	89345	5516523	61.36
1922	3	0.00620	0.00618	0.50	88784	549	88509	5427178	61.13
1922	4	0.00437	0.00436	0.50	88235	384	88043	5338669	60.51
1922	5	0.00382	0.00381	0.50	87850	335	87683	5250626	59.77
1922	6	0.00306	0.00305	0.50	87515	267	87382	5162943	58.99
1922	7	0.00248	0.00247	0.50	87248	216	87140	5075561	58.17
1922	8	0.00206	0.00206	0.50	87033	179	86943	4988421	57.32
1922	9	0.00185	0.00185	0.50	86853	161	86773	4901478	56.43
1922	10	0.00178	0.00177	0.50	86693	154	86616	4814705	55.54
1922	11	0.00169	0.00169	0.50	86539	146	86466	4728089	54.64
1922	12	0.00173	0.00173	0.50	86392	149	86318	4641624	53.73
1922	13	0.00183	0.00183	0.50	86243	158	86164	4555306	52.82
1922	14	0.00206	0.00206	0.50	86085	177	85997	4469142	51.92
1922	15	0.00219	0.00219	0.50	85908	188	85814	4383145	51.02
1922	16	0.00253	0.00253	0.50	85720	217	85612	4297331	50.13
1922	17	0.00274	0.00273	0.50	85504	234	85387	4211719	49.26
1922	18	0.00295	0.00294	0.50	85270	251	85145	4126332	48.39
1922	19	0.00321	0.00320	0.50	85019	272	84883	4041187	47.53
1922	20	0.00343	0.00343	0.50	84747	290	84602	3956304	46.68
1922	21	0.00342	0.00342	0.50	84456	289	84312	3871703	45.84
1922	22	0.00360	0.00359	0.50	84168	302	84017	3787390	45.00
1922	23	0.00357	0.00356	0.50	83866	299	83716	3703374	44.16
1922	24	0.00385	0.00384	0.50	83567	321	83406	3619658	43.31
1922	25	0.00371	0.00371	0.50	83246	308	83091	3536252	42.48
1922	26	0.00380	0.00379	0.50	82937	314	82780	3453160	41.64
1922	27	0.00398	0.00397	0.50	82623	328	82459	3370380	40.79
1922	28	0.00414	0.00413	0.50	82294	340	82124	3287922	39.95
1922	29	0.00420	0.00419	0.50	81954	343	81783	3205797	39.12
1922	30	0.00416	0.00415	0.50	81611	339	81442	3124015	38.28
1922	31	0.00427	0.00426	0.50	81273	346	81099	3042573	37.44
1922	32	0.00464	0.00463	0.50	80926	374	80739	2961473	36.59
1922	33	0.00474	0.00473	0.50	80552	381	80362	2880734	35.76
1922	34	0.00502	0.00501	0.50	80171	401	79971	2800372	34.93
1922	35	0.00515	0.00514	0.50	79770	410	79565	2720402	34.10

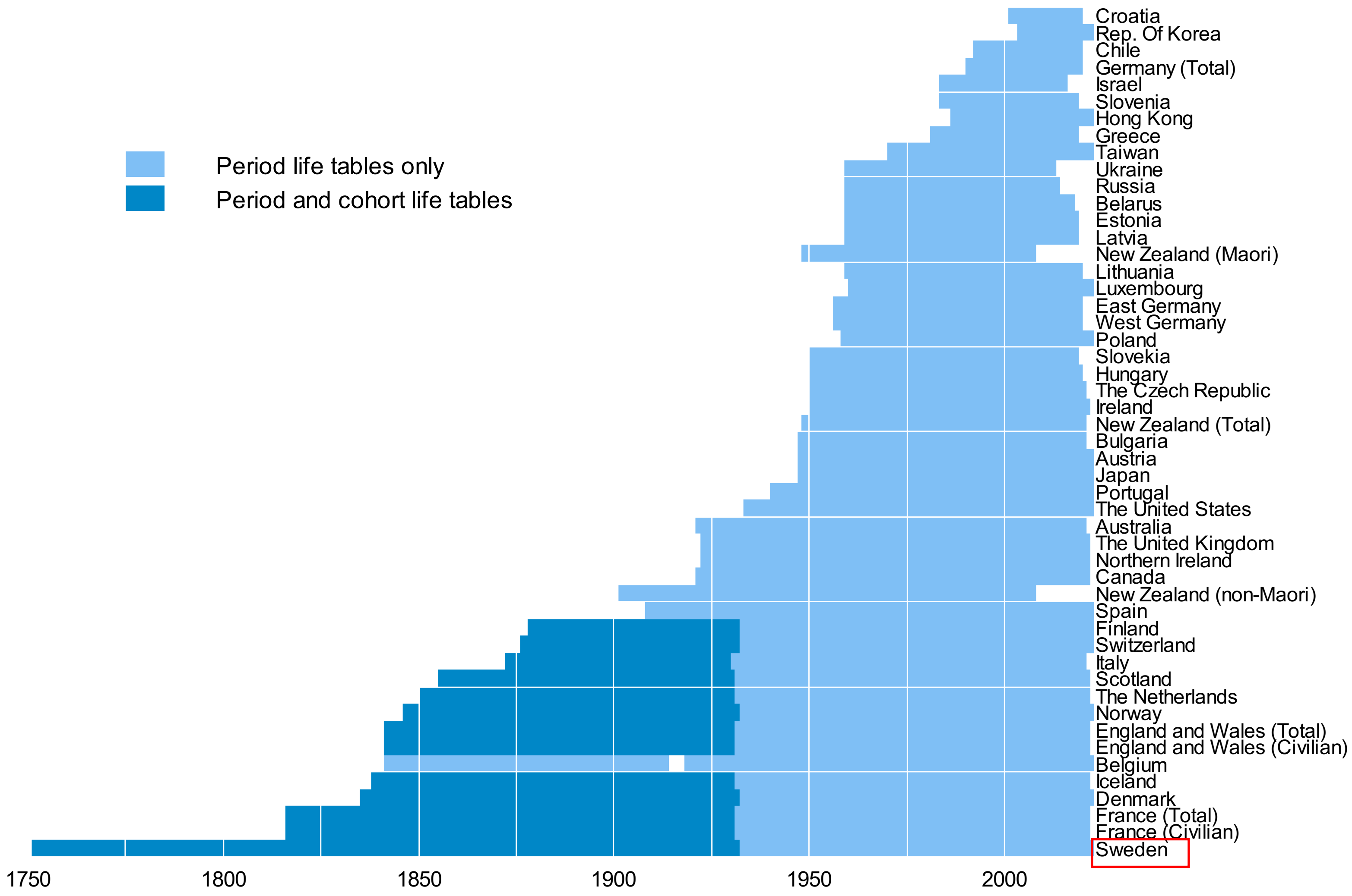
Guiding principles

- Comparability
    - Over time (from 1751 to 2024)
    - Across countries (41 mostly high-income)
- Accessibility (free and relatively painless)
- Flexibility (estimates in multiple formats)
- Reproducibility (all input data included, full documentation provided)
- Quality control (procedures to identify and correct errors in data or calculations)

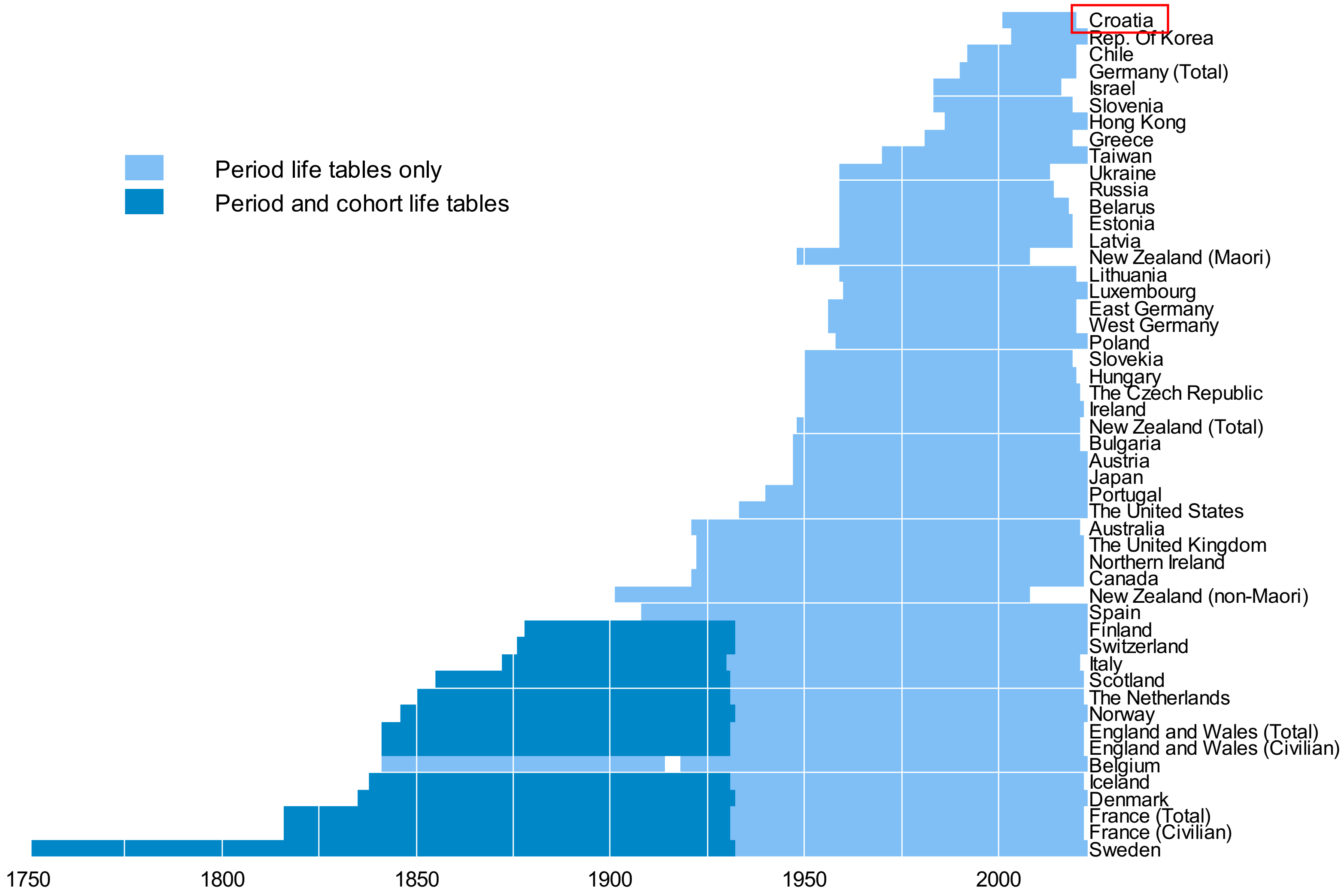
HMD series by country and time period



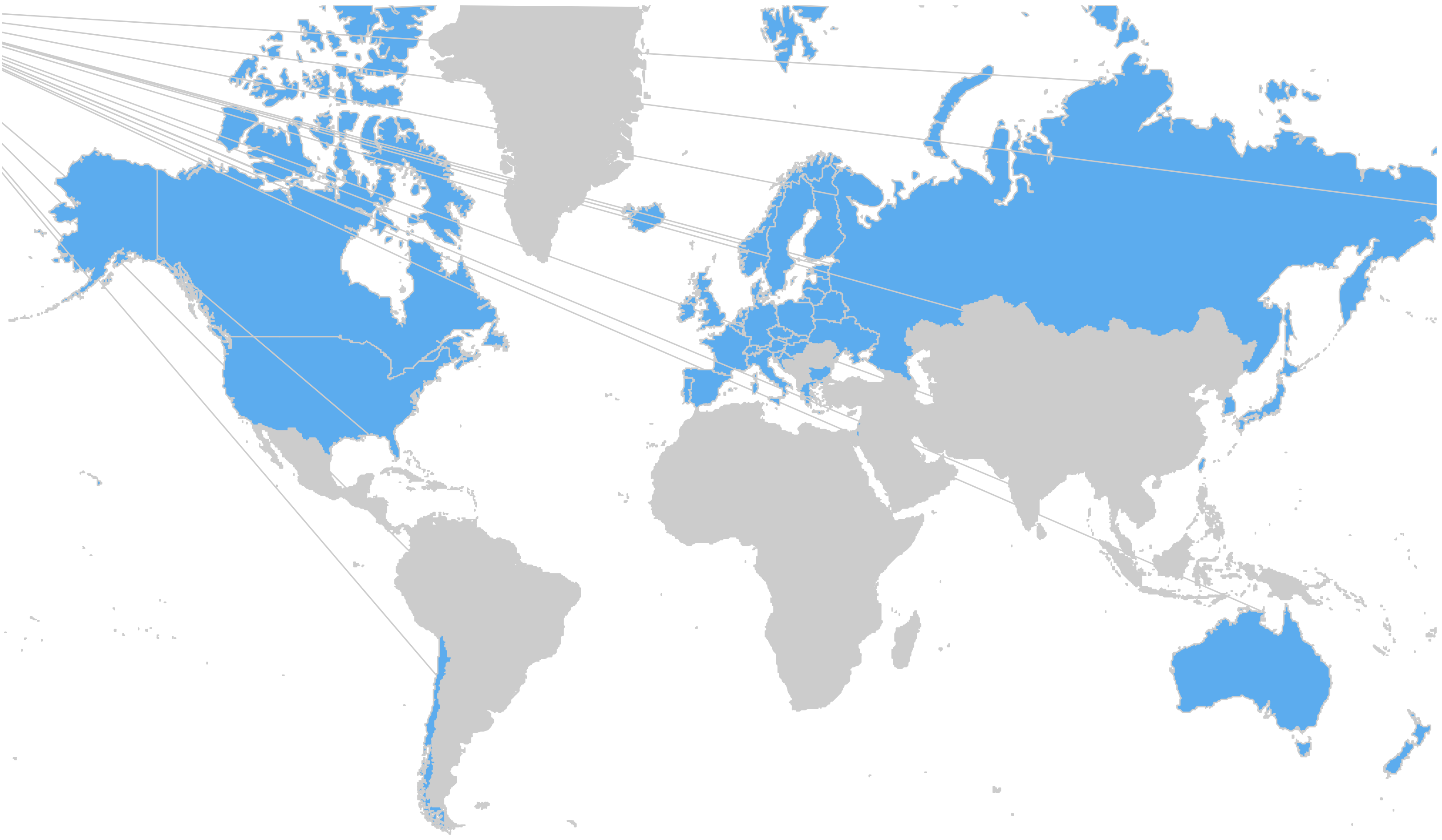
HMD series by country and time period



HMD series by country and time period



Geographic location of HMD countries



HMD Methods

Full HMD Methods Protocol available online at:
http://www.mortality.org/Public/Docs/MethodsProtocol.pdf


Last Revised: November 27, 2017 (Version 6)

Methods Protocol for the Human Mortality Database
J.R. Wilmoth, K. Andreev, D. Jdanov, D.A. Glei and T. Riffe with the assistance of C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, P. Vachon, C. Winant, M. Barbieri[1]

Table of Contents

- 2 -


Last Revised: November 27, 2017 (Version 6)

List of Figures

- 3 -

Additional information in

Int. J. Epidemiol. Advance Access published June 23, 2015

International Journal of Epidemiology, 2015, 1–8
doi: 10.1093/ije/dyv105
Data Resource Profile

Data Resource Profile

Data Resource Profile: The Human Mortality Database (HMD)

Magali Barbieri,[1,2]* John R Wilmoth,[1,3] Vladimir M Shkolnikov,[4,5] Dana Glei,[6] Domantas Jasilionis,[4] Dmitri Jdanov,[4,5] Carl Boe,[1] Timothy Riffe,[1] Pavel Grigoriev[4] and Celeste Winant[1]

[1]Department of Demography, University of California, Berkeley, CA, USA, [2]French Institute for Demographic Studies, Paris, France, [3]Population Division, United Nations, Department of Economic and Social Affairs, New York, NY, USA, [4]Max Planck Institute for Demographic Research, Rostock, Germany, [5]New Economic School, Moscow, Russia and [6]Georgetown University, Washington, DC, USA

*Corresponding author. Department of Demography, University of California, 2232 Piedmont Avenue, Berkeley CA 94720-2120, USA. E-mail: magali@demog.berkeley.edu

HMD-associated projects

- The Human Cause of death Data series (HCD)
- The Short-Term Mortality Fluctuation Database (STMF)
- The Lifetable Database
- Subnational Mortality Databases (USMDB, CHMD, JHMP, AHMD…)

The United States Mortality DataBase

USMDB @ usa.mortality.org

usa.mortality.org

Return to main USMDB page

▼ Registration

New User
Change Password
User's Agreement

▼ USMDB Menu

Acknowledgments
Overview
Frequently Asked Questions (FAQ)
Research Team
Explanatory Notes
Methods
Documentation
Citation Guidelines
USMDB Citation List
Contact us

▼ Links

UC Berkeley Department of Demography
INED
HMD

United States Mortality DataBase

Magali Barbieri, Director
University of California, Berkeley & INED, France

John Wilmoth, Founding Director
United Nations & formerly University of California, Berkeley

** **Disclaimer** ** The USMDB life tables for years 2010-2023 are preliminary as the US Census Bureau inter-censal (2010-2019) and post-censal (2020-2023) population estimates are provisional. The final 2010-2019 intercensal estimates are expected to be published for all geographic levels (national, state, and county) in the fall of 2025. We will update and extend the USMDB data series as soon as the US Census Bureau has released new population estimates. We expect these revised data to create noticeable changes in all lifetable indicators for years 2010 onward.

Updated February 18, 2025 US State life table series for years **1941-2022** (link here). We have extended the life table series to include 1941-1958 for all states (except Alaska and Hawaii), DC, all Census Divisions (except Div9), Regions (except Reg4). Life tables for geographical units composed or comprised of Alaska and Hawaii begin with the year 1959.
Updated Sep 13, 2023 US County life table series for years 1982-2019 (link here).

Link to map-based data visualization tool for comparing US Mortality Database lifetable indicators across- and by state: https://usa.mortality.org/uploads/Private/ShinyInfo.php **Note, USMDB user name and password required for access.**

General Project Description: The United States Mortality DataBase (USMDB) brings you the first clearly documented historical set of complete state-level life tables, updated to the latest year of available data, designed to foster research on geographic variations in mortality across the United States and to monitor trends in health inequalities. This data set currently includes complete and abridged life tables by sex for each of the US 9 Census Divisions, 4 Census Regions, 50 States and the District of Columbia, for each year since 1941 (1959 for Alaska and Hawaii) with mortality values up to age 110. The data are available for free to all interested upon registration.

The USMDB is an outgrowth of the Human Mortality Database (HMD) project. It was developed to provide detailed mortality

Metrics 100 557 782 Downloads

Contact Share

United States Mortality Database Advanced Search

Add Data

- Dataverses (742)
- Datasets (83 772)
- Files (54 868)

Dataverse Category

Research Project (335)
Researcher (170)
Organization or Institution (73)
Research Group (60)
Journal (14)

More...

Metadata Source

Harvard Dataverse (103 400)
Harvested (35 982)

Publication Year

2025 (4 515)
2024 (15 103)

1 to 10 of 139 382 Results

Sort

The United States Fertility and Mortality Database (University of California, Berkeley)
14 mai 2025
Welcome to the landing page for the United States Mortality DataBase! (formerly hosted at usa.mortality.org) The United States Mortality DataBase (USMDB) and forthcoming United States Fertility DataBase (USFDB) will soon be offered together on a new webpage. In the meantime, we provide zipped files of USMDB state-level and county-level life tables,...

USMDB Documentation
14 mai 2025 - The United States Fertility and Mortality Database
Barbieri, Magali; Winant, Celeste, 2025, "USMDB Documentation", https://doi.org/10.7910/DVN/Z9FT4I, Harvard Dataverse, V2
... • USMDB county-level methods • We also link to the companion reference, A Flexible Bayesian Model for Estimating Subnational **Mortality** (Alexander, Zagheni, Barbieri, Demography 2017) ...

U.S **State** Life Tables CSV + TXT 1941-2022
14 mai 2025 - The United States Fertility and Mortality Database
Barbieri, Magali; Winant, Celeste, 2025, "U.S State Life Tables CSV + TXT 1941-2022", https://doi.org/10.7910/DVN/19WYUX, Harvard Dataverse, V1

Feedback

U.S. State/County Mortality Database

- Uses restricted data from the US vital statistics system
- HMD-like mortality series with same basic principles (comparability, [accessibility], flexibility, quality control)
- Except for reproducibility due to privacy issues (no public access to the raw mortality data)
- Back to 1941 (State) or 1982 (counties)
- Updated every year
- Main issue: large random fluctuations => small numerators and denominators (state COD, counties…); require additional methodological developments.

Demography
DOI 10.1007/s13524-017-0618-7

CrossMark

A Flexible Bayesian Model for Estimating Subnational Mortality

Monica Alexander[1] · **Emilio Zagheni**[2] · **Magali Barbieri**[1,3]

Abstract Reliable subnational mortality estimates are essential in the study of health inequalities within a country. One of the difficulties in producing such estimates is the presence of small populations among which the stochastic variation in death counts is relatively high, and thus the underlying mortality levels are unclear. We present a Bayesian hierarchical model to estimate mortality at the subnational level. The model builds on characteristic age patterns in mortality curves, which are constructed using principal components from a set of reference mortality curves. Information on mortality rates are pooled across geographic space and are smoothed over time. Testing of the model shows reasonable estimates and uncertainty levels when it is applied both to simulated data that mimic U.S. counties and to real data for French *départements*. The model estimates have direct applications to the study of subregional health patterns and disparities.

Keywords Mortality · Subnational estimation · Bayesian hierarchical model · Principal components · France

Introduction

To effectively study health disparities within a country, one must obtain reliable subnational mortality estimates to quantify geographic differences accurately. There

✉ Monica Alexander
monicaalexander@berkeley.edu

1 Department of Demography, University of California, Berkeley, 2232 Piedmont Avenue, Berkeley, CA 94720-2120, USA

2 Department of Sociology, University of Washington, Seattle, 211 Savery Hall, Box 353340, Seattle, WA 98195-3340, USA

3 Institut National d'Études Démographiques, 133 Boulevard Dabout, 75020 Paris Cedex, France

Published online: 10 October 2017

Springer

DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 52, ARTICLE 3, PAGES 71–110
PUBLISHED 10 JANUARY 2025
http://www.demographic-research.org/Volumes/Vol52/3/
DOI: 10.4054/DemRes.2025.52.3

Research Article

Jointly estimating subnational mortality for multiple populations

Ameer Dharamshi

Monica Alexander

Celeste Winant

Magali Barbieri

An NSF-funded project 10/1/2024 – 9/30/2027

The United States Fertility Database (USFDB) Monitoring and Understanding Change in Local Fertility Patterns

PI: Magali Barbieri (UCB), Co-PI: Alison Gemmill (Johns Hopkins)
Other team members:
Celeste Winant (UCB) and Emma Lonstrup (UCB)
Ameer Dharamshi (UW) and Monica Alexander (Toronto)
Carl Schmertmann (Florida State)

Goals

- **Overall** purpose : to facilitate research into the contextual factors driving the recent US fertility decline
- **Specific** purpose : construct an open-access database of fertility indicators for US geographies
  - Since 1941 for states, US Census regions and divisions
  - Since 1982 for counties

Filling a gap

- **Longer time series** to understand why fertility remained high in the US compared to other high-income countries and why it started declining in 2008
- **Finer geographic level** than the state (NCHS) as preliminary research in the US and in Europe demonstrated the role of area-level social, economic, cultural and policy characteristics
- **Larger range of indicators** than simply the period TFR or ASFRs to evaluate the role of postponement and changes in family sizes, including childlessness

Indicators (1)

- **At the state level (1941-2023)**
  - **Period measures**
    - Total birth count
    - Crude birth rate
    - Age-specific-fertility rates by single and five-year age group
    - Cumulative fertility rates
    - Total fertility rate
    - Gross reproduction rate
    - Tempo-adjusted total fertility rate
    - Mean age at birth
    - Parity-progression ratios

Indicators (2)

- **At the state level (1941-2023)**
  - **Cohort measures** (estimated from single-year ASFRs)
    - Age specific fertility rates
    - Completed (and incompleted) cohort fertility rate
    - Mean age at birth

Indicators (3)

- **At the county level (1982-2023)**
  - **Period measures**
    - Total birth count
    - Crude birth rate
    - Age-specific-fertility rates by five-year age group
    - Cumulative fertility rates
    - Total fertility rate
    - Gross reproduction rate
    - Tempo-adjusted total fertility rate
    - Mean age at birth
  - **Cohort measures?**

Data

1. **Vital statistics data (1941/1982-2023)**
   - Birth counts by state/county, calendar year and maternal age (no information on mothers' year of birth)
2. **Census Bureau population estimates/census data**
   - 1940, 1950, 1960 and 1970 population census => population exposures estimated using HMD methods
   - Annual July 1st population estimates by state/county, sex and age for years since 1970
3. **Current Population Surveys**
   - Women by age and number of children ever born

Methods (1)

- **At the state level =>**
  - **Classic demographic methods** with some adjustments
    - To redistribute ASFRs from 5-year age groups to single-year of age for 1941-1967
    - To compute cohort-fertility from single-year of age period ASFRs

- **At the County level =>**
  - **Statistical modelling** using Bayesian inference necessary

    => Large year-to-year random fluctuations and many zero-count cells due to small populations (median county-size = 25,000 people)

Methods (2)

- **Overall approach for the Bayesian model**: takes advantages of
  - The relative regularity in the fertility curve over ages
  - Relatively gradual changes in fertility behavior over time
  - The similarity in fertility patterns across geographically proximate areas

$\Rightarrow$The model borrows information across

- Ages
- Time
- Areas (states => Principal Component Analysis of the ASFRs)

Modeling Goals

- Aim to estimate the age-specific fertility rates for each county and year between 1982 and 2022
- We estimate five-year rates for reproductive life course [10-14, 15-19, …, 45-49]
- We designed a statistical model that
  - Accounts for high stochasticity / zero birth counts in small areas
  - Models patterns over time (sharing of information over time, smoothing, forecasting)
  - Shares information across geographic space (smaller counties are partially informed by other counties that have geographic similarities)

Notation

- $B_{a,c,t}$ are the number of births to women in age group *a,* county *c,* year *t*
- $P_{a,c,t}$ is the population of women in age group *a*, county *c,* year *t*
- $F_{a,c,t} = \frac{B_{a,c,t}}{P_{a,c,t}}$ is the (observed) age-specific fertility rate in age group *a*, county *c,* year *t*
- Let $\lambda_{a,c,t}$ be the ‘true’ / latent set of age-specific fertility rates. This is the what we are interested in estimating.

Data model

- (Data model = the model that relates what we observe to what we are trying to estimate)
- We assume the observed rates $F_{a,c,t}$ are Normally distributed

$$F_{a,c,t} \sim N(\lambda_{a,c,t}, \sigma^2_{a,c,t})$$

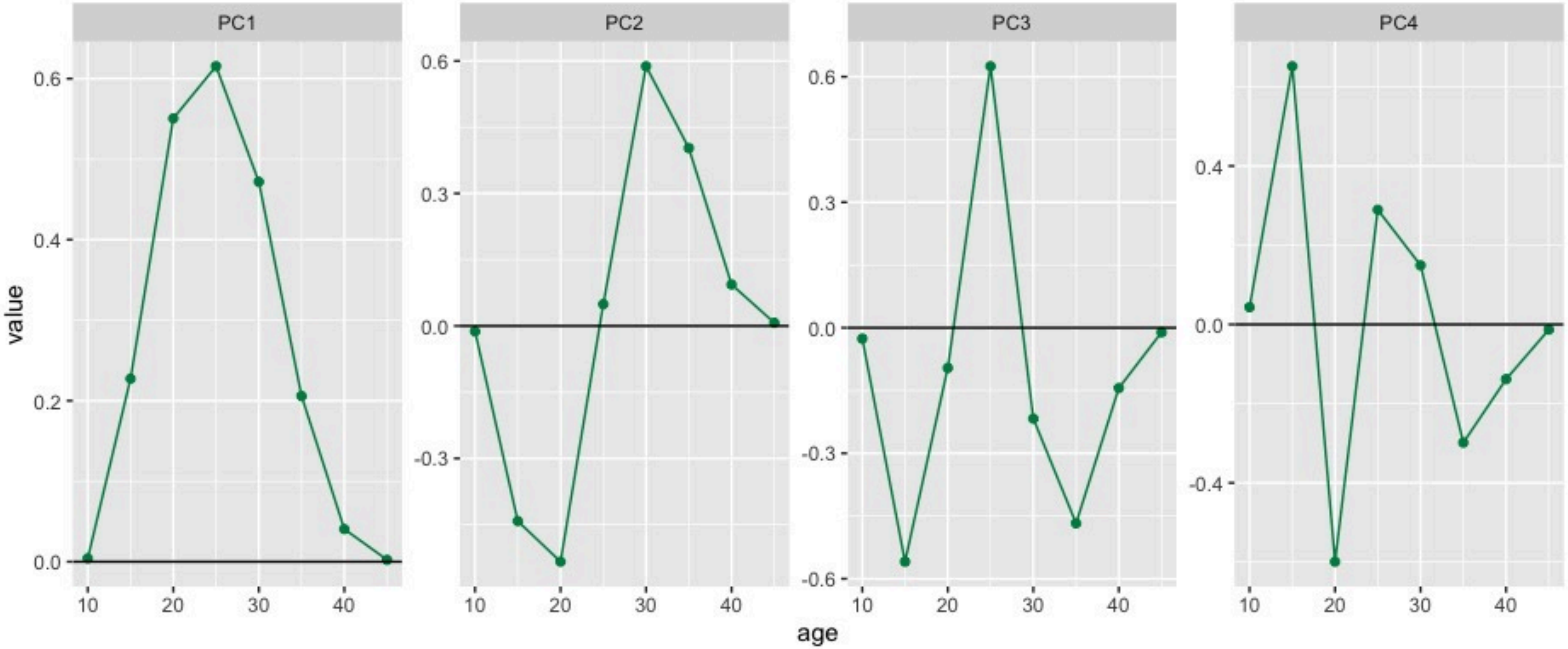
- Variance is adjusted to allow for multiple births
- (also investigating different forms, specifically the Poisson distribution on birth counts)

Model for latent rates $\boldsymbol{\lambda}$

$$\lambda_{a,c,t} = \beta_{1,c,t}V_{1,a} + \beta_{2,c,t}V_{2,a} + \beta_{3,c,t}V_{3,a} + \beta_{4,c,t}V_{4,a} + \varepsilon_{a,c,t}$$

- The $V_{j,a}$ are ‘principal components’ that capture main patterns of variation in fertility rates over age
- These are derived from state-level fertility schedules

Principal components



Spatial-temporal model on PC coefficients

- The principal component coefficients $\beta_{j,c,t}$ are modeled as a combination of spatial and temporal effects:

$$\beta_{j,c,t} = \mu_j + \alpha_j(t) + \omega_j(c) + \delta_j(c,t)$$

Temporal component

Spatial component

Spatial temporal interaction