Extending the HMD approach to regional databases - An illustration with the United States Mortality Data Base (USMDB)

Celeste Winant[1,3], Denys Dukhovnov[1,4],
Magali Barbieri[1,2], Carl Boe[1]

1. University of California, Berkeley
2. French National Institute for Demographic Studies (INED)
3. celestewinant@berkeley.edu 4. denys_dukhovnov@berkeley.edu

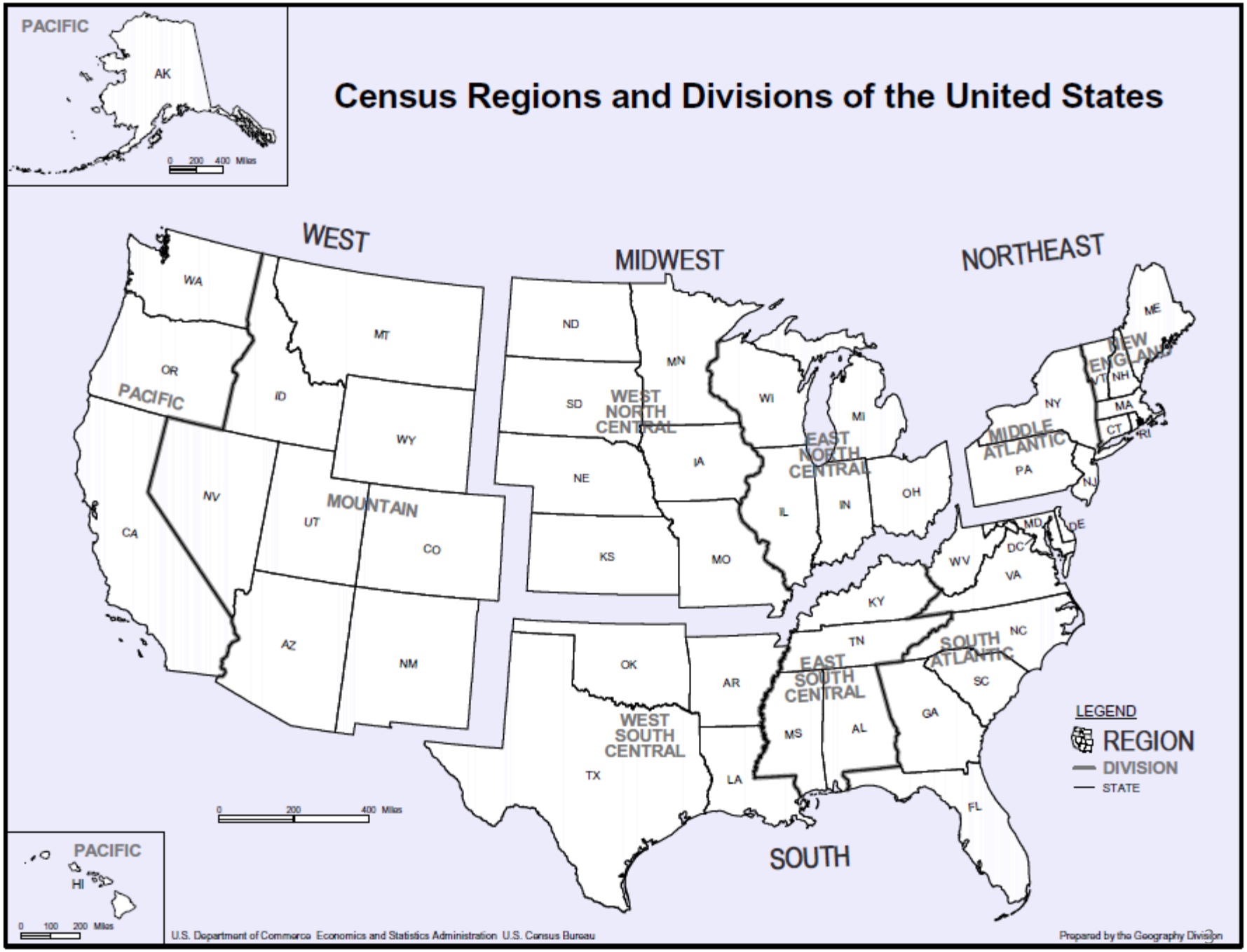
The 5th Human Mortality Database Symposium
13-15 May 2019 - Berlin Germany

Acknowledgement: this presentation is based on the work conducted by many members of the HMD and USMDB teams, at both UC Berkeley, and the Max Planck Institute for Demographic Research (MPIDR), Rostock, with funding provided by the NIH and by the Society of Actuaries (SOA).

USMDB Overview

- What is the USMDB?
  - A database of detailed and abridged period life table series for 1959-2016 and supporting documentation for the United States as a whole and for:
    - 4 Census regions, 9 Census divisions, the 50 states and D.C.
    - Present-day 50+ state-membership established in 1959
- **Goal** of the USMDB:

  To provide detailed mortality data for sub-regional populations of the United States, free of charge, to all persons interested in US geographic variations in longevity.


Census Regions and Divisions of the United States
PACIFIC
AK
0 200 400 Miles
WEST
MIDWEST
NORTHEAST
SOUTH
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
MN
IA
MO
WI
IL
MI
IN
OH
NY
PA
ME
VT
NH
MA
CT
RI
NJ
MD
DE
DC
WV
VA
KY
TN
NC
SC
GA
FL
AL
MS
LA
TX
AR
OK
PACIFIC
MOUNTAIN
WEST NORTH CENTRAL
EAST NORTH CENTRAL
NEW ENGLAND
MIDDLE ATLANTIC
SOUTH ATLANTIC
EAST SOUTH CENTRAL
WEST SOUTH CENTRAL
HI
0 100 200 Miles
0 200 400 Miles
LEGEND
REGION
DIVISION
STATE
U.S. Department of Commerce Economics and Statistics Administration U.S. Census Bureau
Prepared by the Geography Division

Methodology

Faithfully replicates HMD methods:

- Collect inputs for each state (create aggregates for Division, Region, National level)
  - Inputs for each state are from centralized (national-level) organizations
  - Homogeneous data
- Compute Lexis Database for each region, determine period exposure-to-risk
- Compute period life tables from exposures-to-risk and deaths
- Recent (2016) update: HMD method version 5 is used
  - Move to version 6 in next (2017) update by adding (recently obtained) births by month for all states.

Exceptions to HMD methods

- Upper/Lower triangle probability assigned to death records w/ missing cohort.
  - Same method used for HMD-USA

The USMDB is computed entirely with in-house-developed R software and has been used to verify the translation/migration of legacy MATLAB HMD routines to R (Tim Riffe, Carl Boe)

Input data

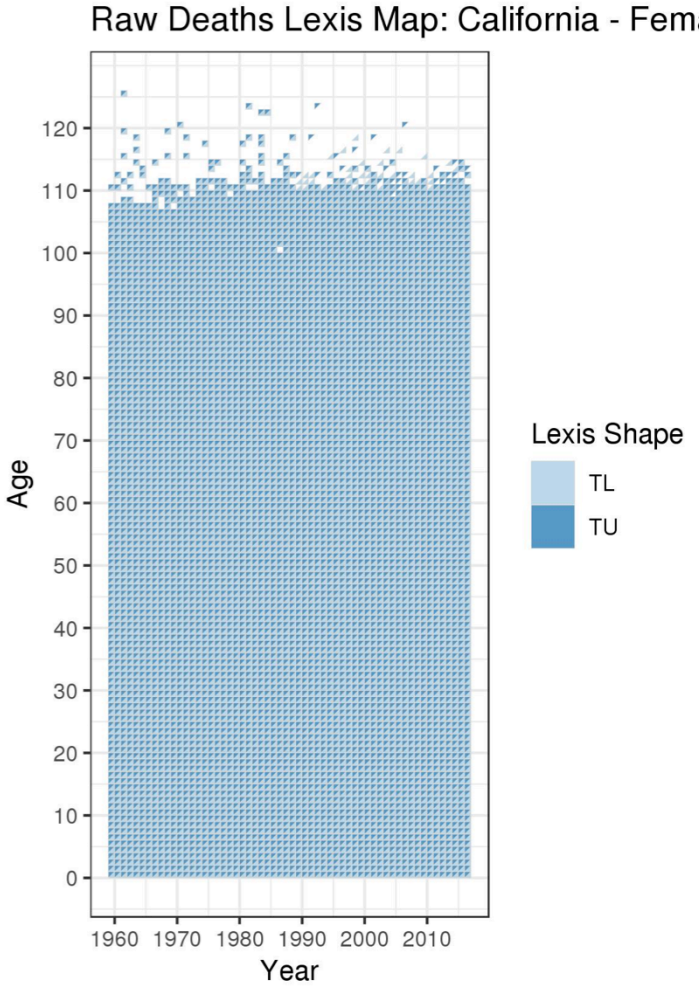
A. From the National Center for Health Statistics (NCHS):

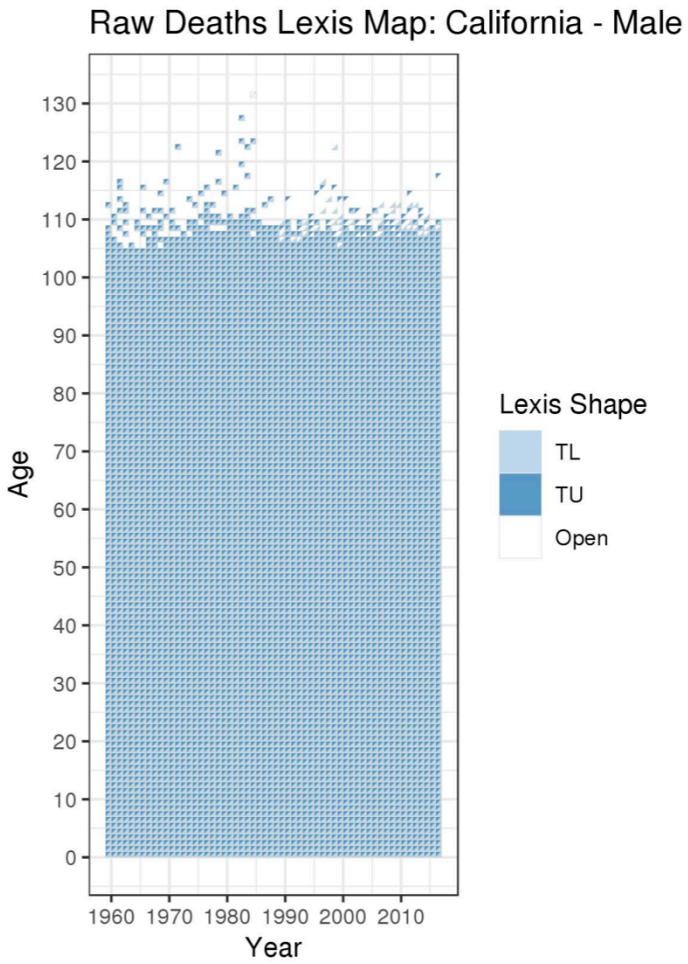
1. Natality files => births by area, year, and sex
2. Mortality files => deaths by area, year, sex, and Lexis triangle (full cohort detail from 1989 - present, single year of age before 1989)

B. From the Census Bureau:

1. Census populations (1960 on) by area, single year of age and sex
2. July 1st population estimates (1970 on) by area, single year of age and sex - only available age 85+

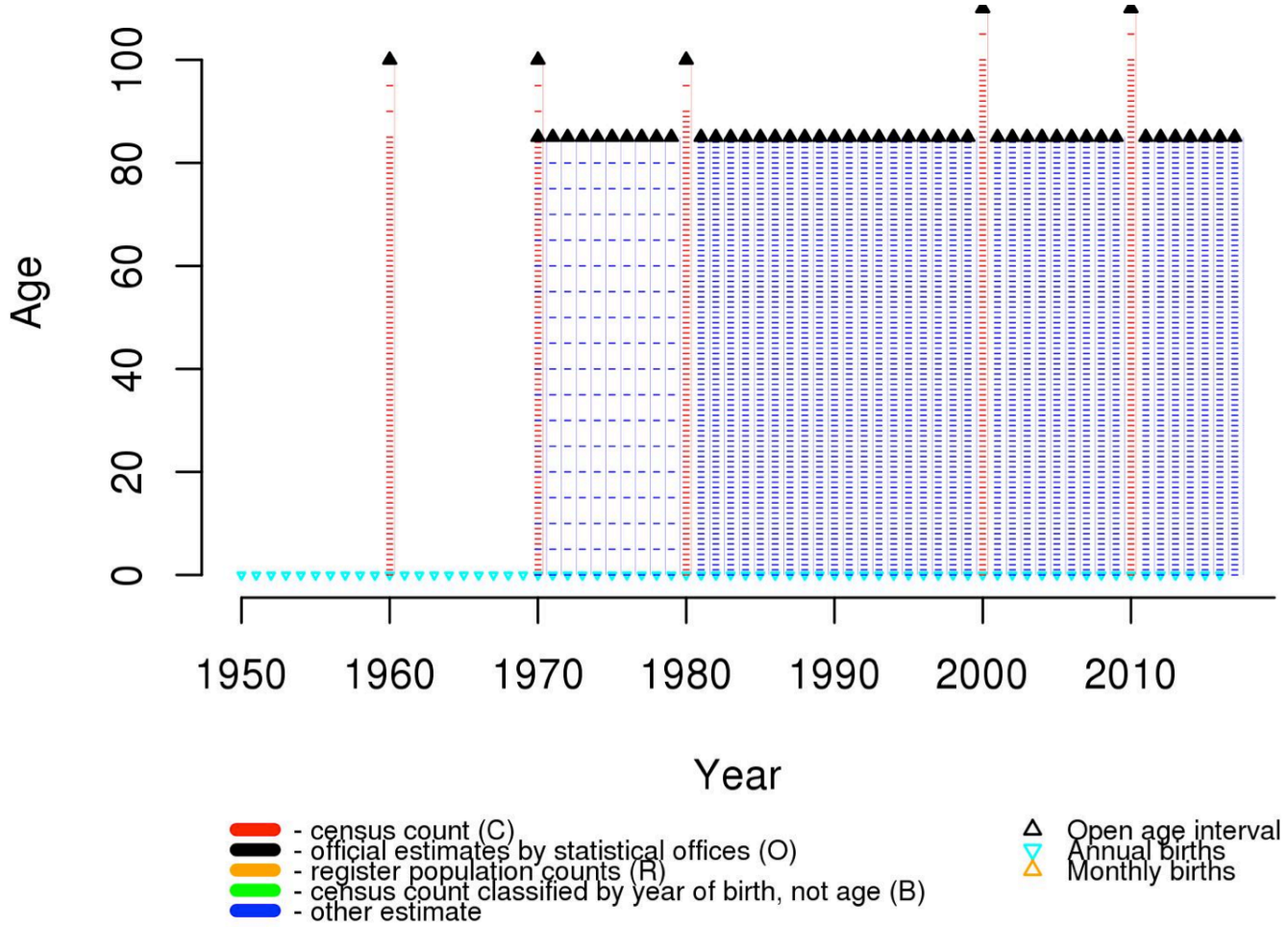
Coverage of Available Mortality Data





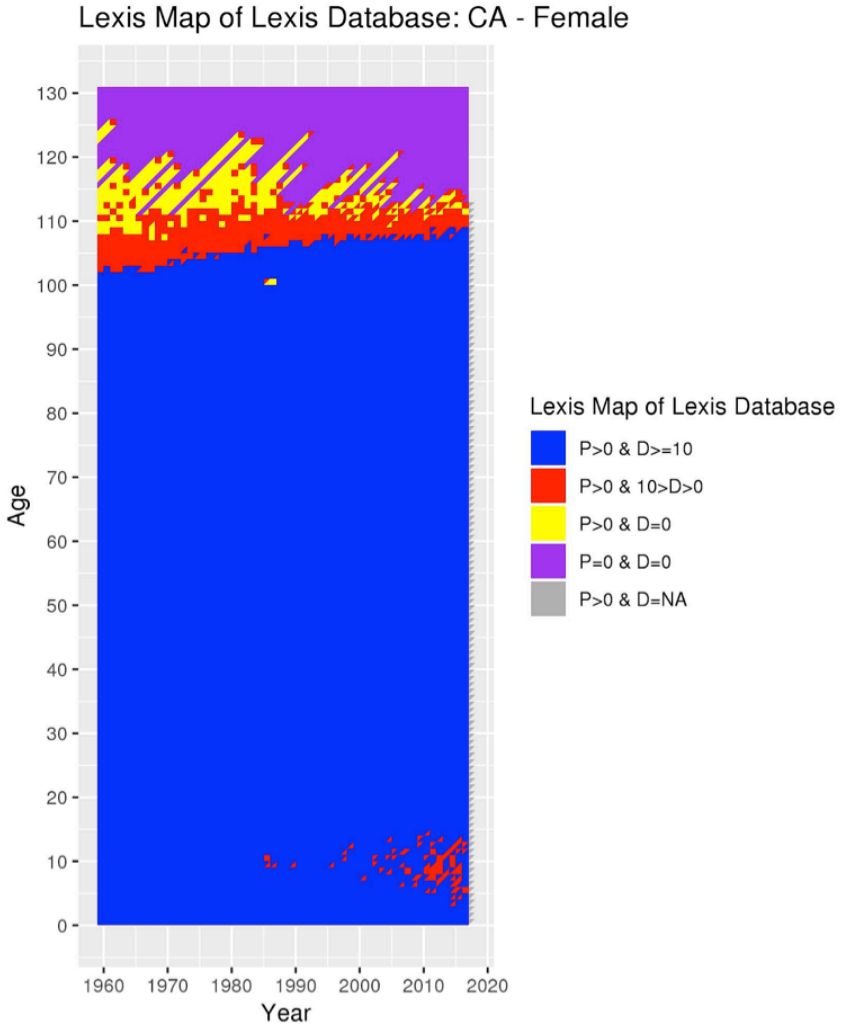
Restricted access to detailed mortality data via US Census Bureau Federal Statistics Research Data Center.

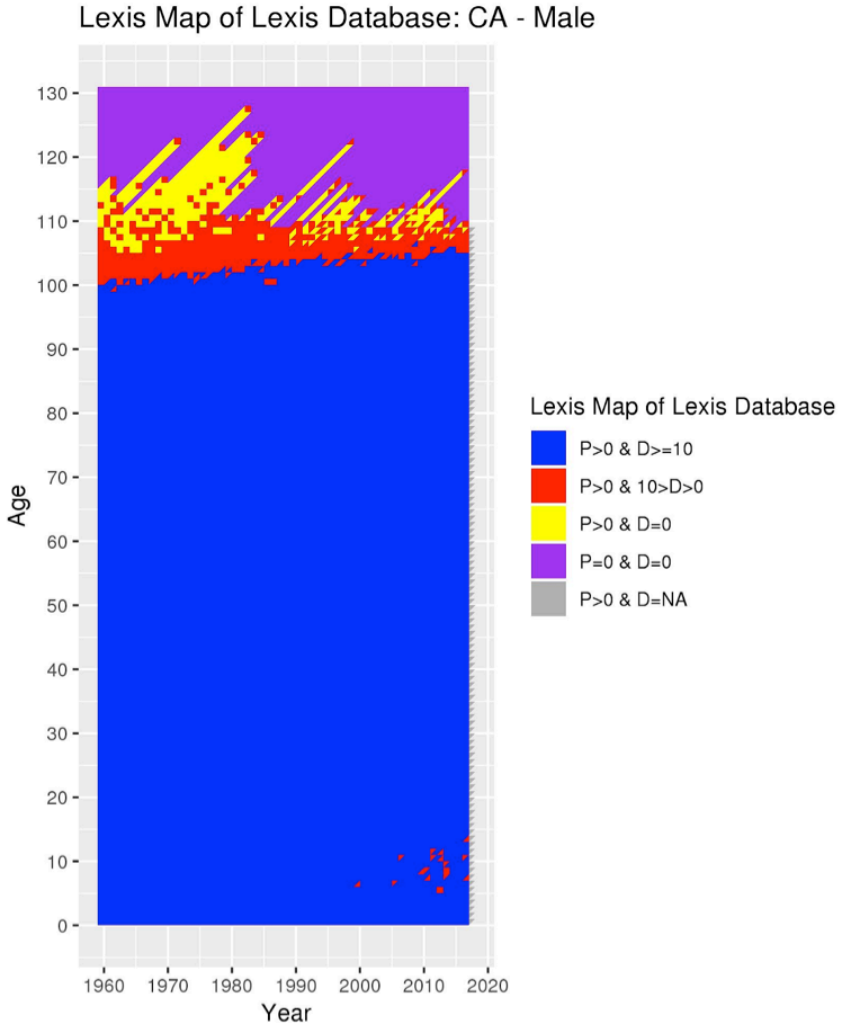
Coverage of Available Population & Natality Data



Incomplete population inputs by year (missing intercensal estimates for 1959-1969), by age-detail (only 5yr age groups for 1971-1979) and for all ages (Open age interval starts at age 85 for intercensal estimates): Also, 1990 Census discarded because of known age-reporting problems)

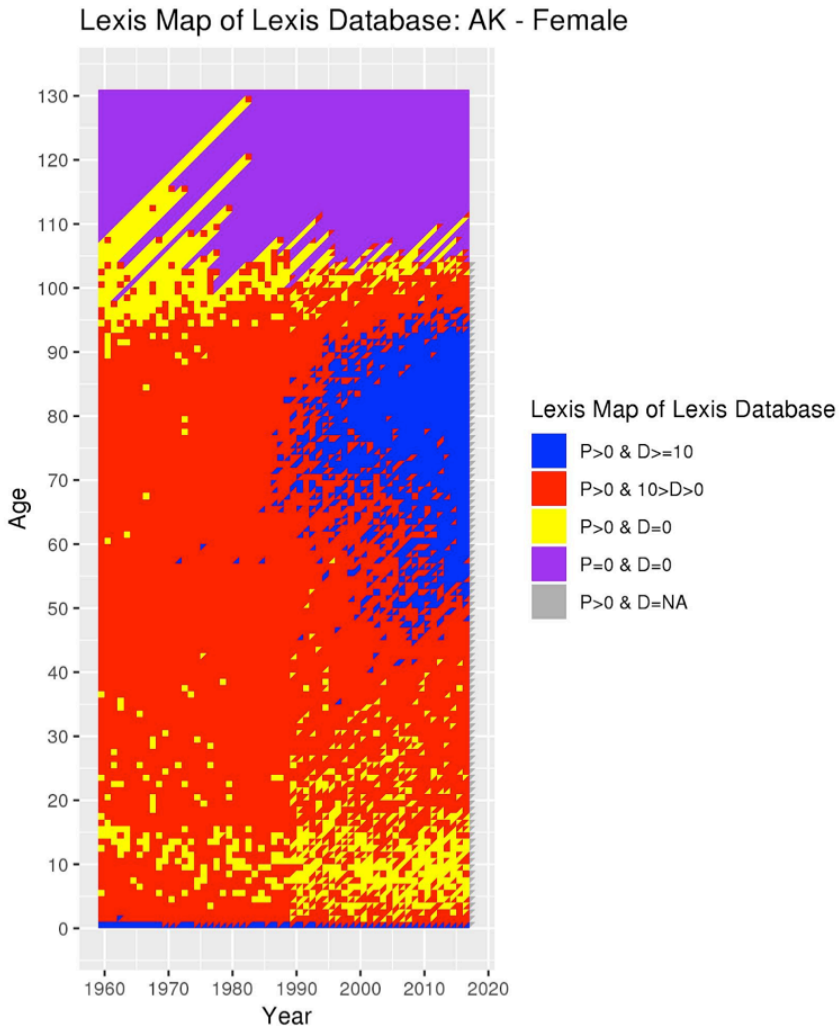
Lexis surface sampling in high-population state: California

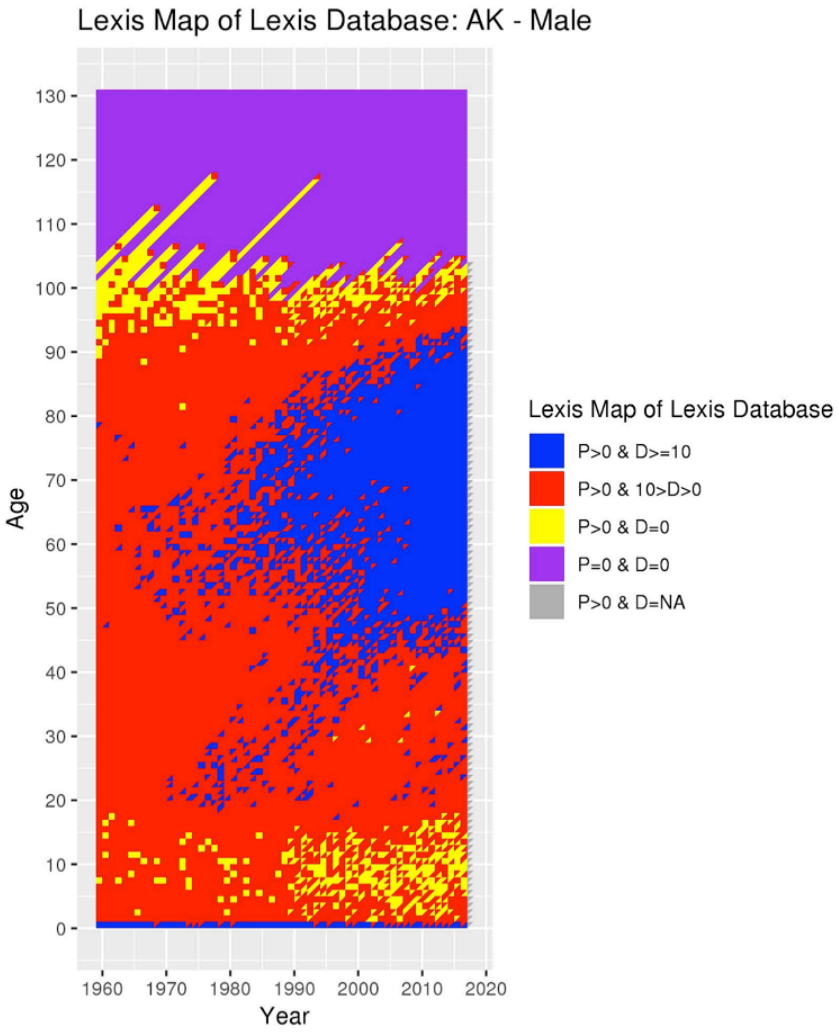




Total 2018 CA population: 39.56 million

Lexis surface sampling in low-population state: Alaska





Total 2018 Alaska population: 734,458 - compare w/ Luxembourg population: 599,499 (2017)

USMDB Results: Snapshot of life table indicators

Low- and high- mortality cases

2016: Female

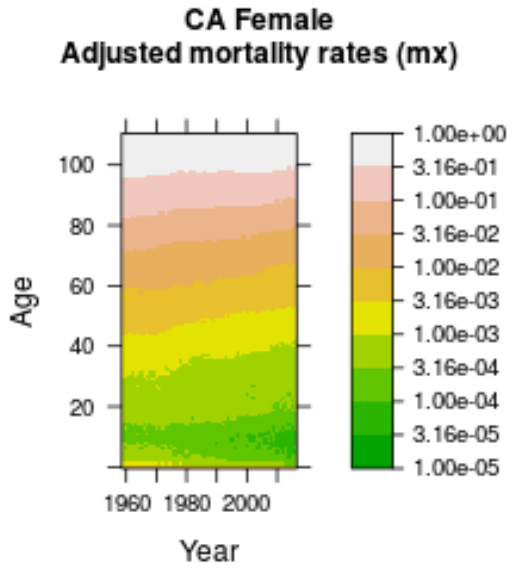
Region	q0	e0	e65	e80	l65*
California	0.00375	83.49	22.06	10.80	90,290
Mississippi	0.00845	77.78	19.25	9.31	81,625
USA (USMDB)	0.00529	81.38	20.93	10.15	87,569

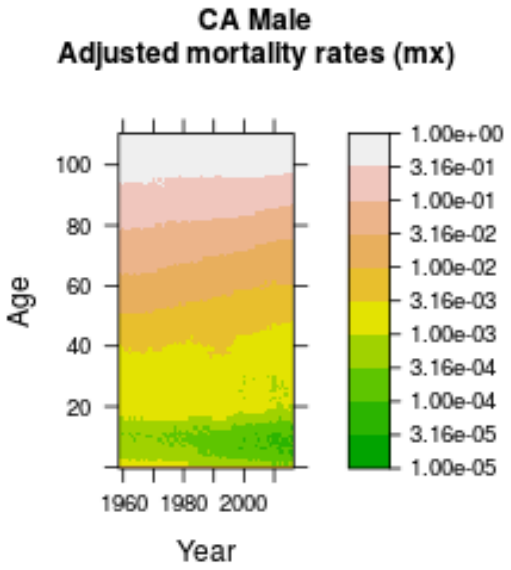
2016: Male

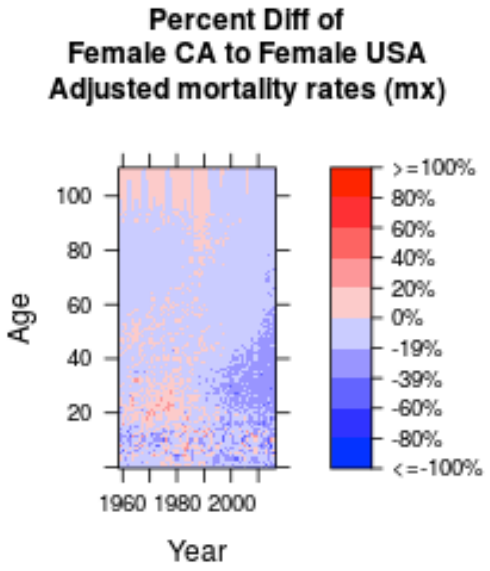
Region	q0	e0	e65	e80	l65*
California	0.00458	78.76	19.50	9.45	83,559
Mississippi	0.00873	71.64	16.23	7.81	70,839
USA (USMDB)	0.00631	76.34	18.34	8.78	79,821

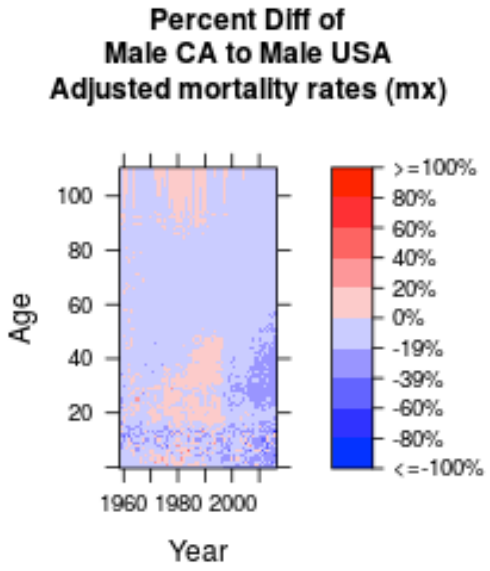
*Radix = 100,000

Mortality rates: Low mortality case - California (CA)

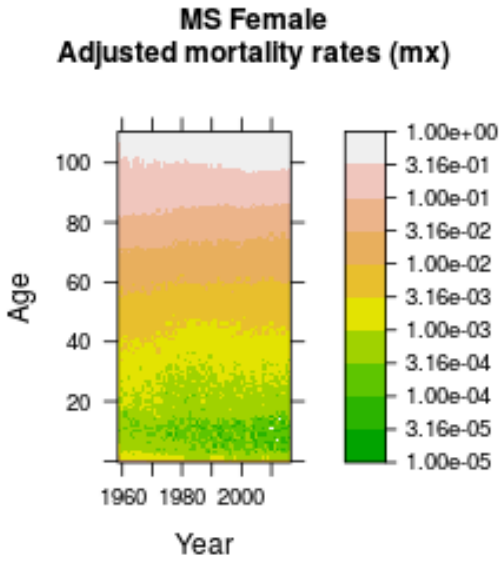


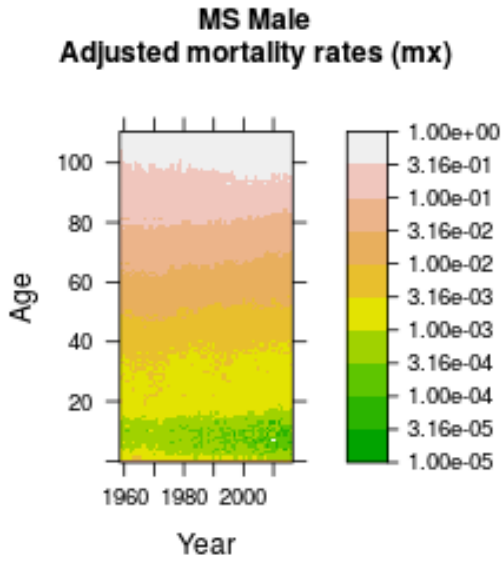


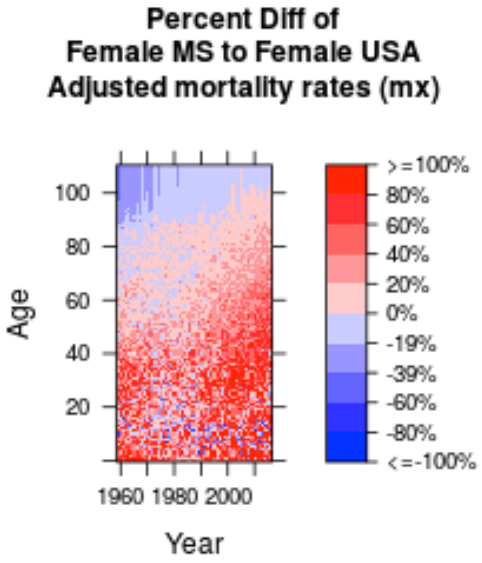


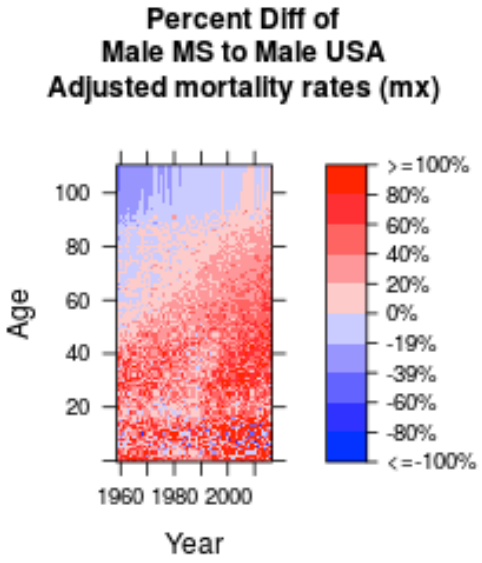


Mortality rates: High mortality case - Mississippi (MI)





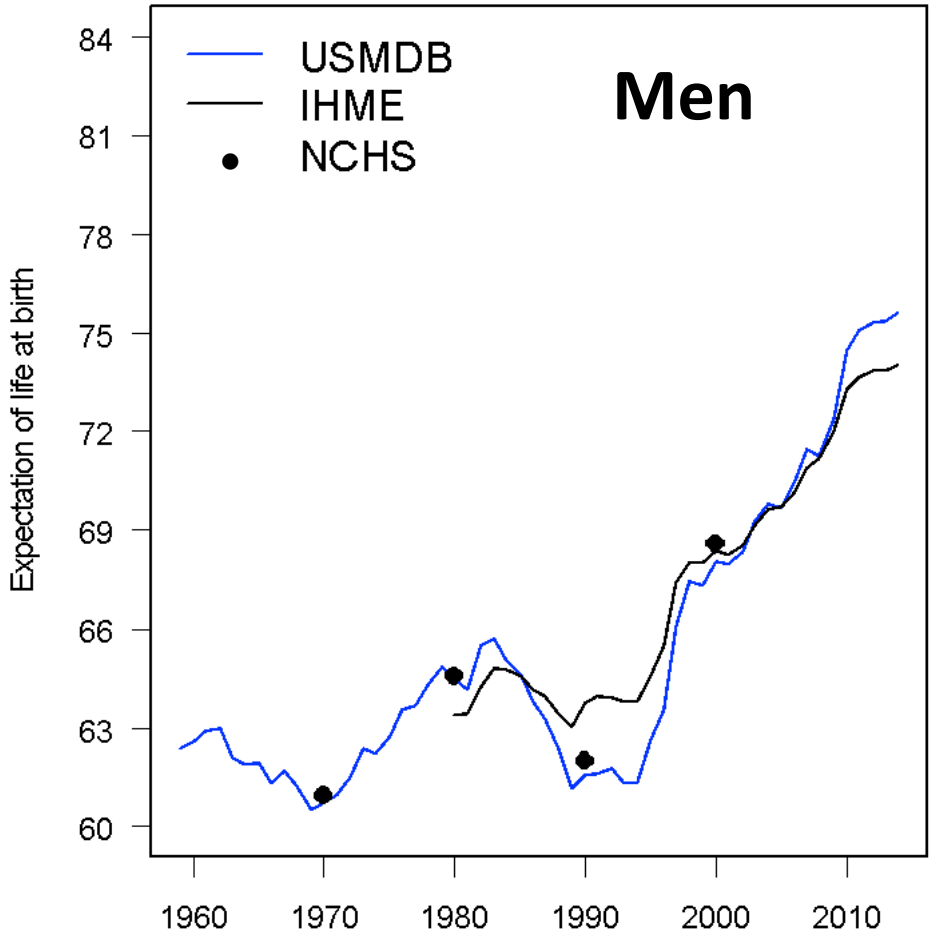


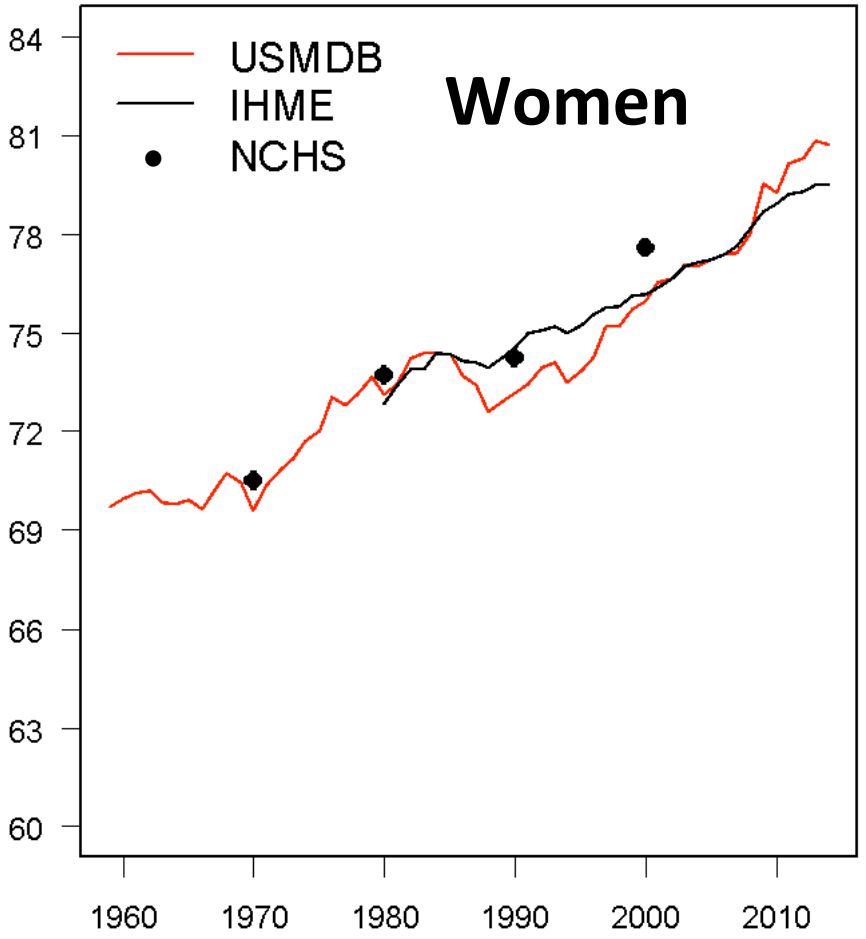


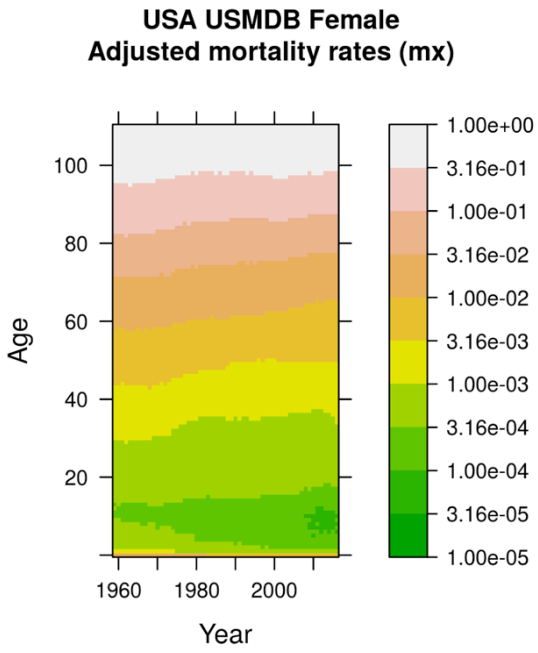
Verifications

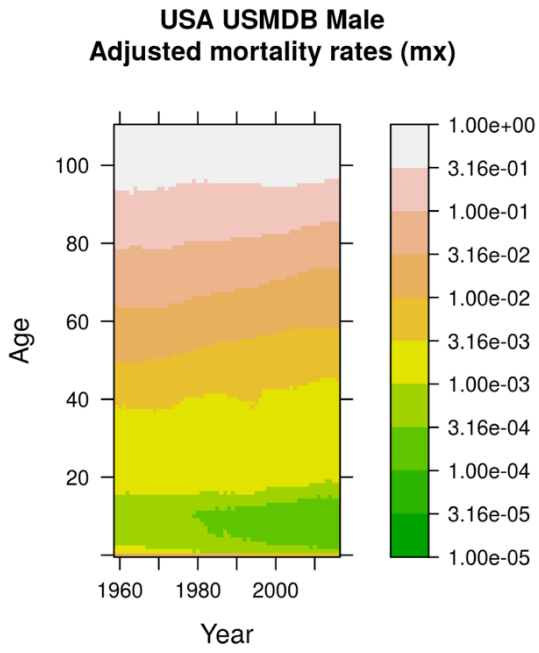
- Internal checks (data quality checks on age reporting; consistency of implied migration; age structure of mortality; etc…)
- Comparison with NCHS Decennial life tables (1959-1961 through 1999-2001)
- Comparison with estimates from the Institute for Health Metrics and Evaluation (IHME)

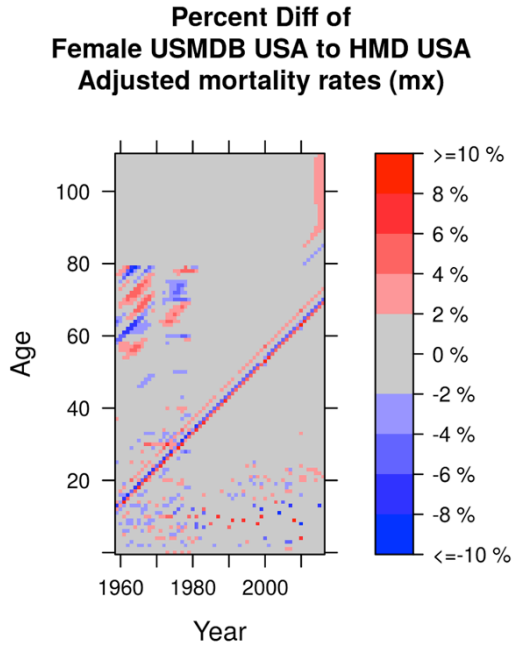
Life expectancy at birth, District of Columbia, USMDB, NCHS, and IHME, 1959-2014

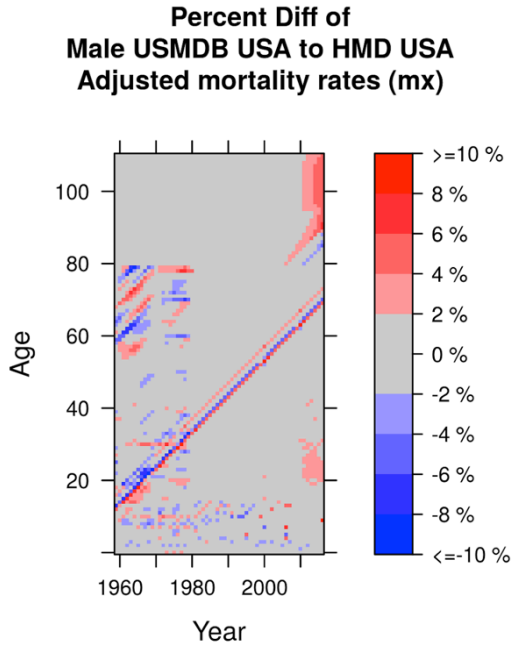












USMDB vs. HMD

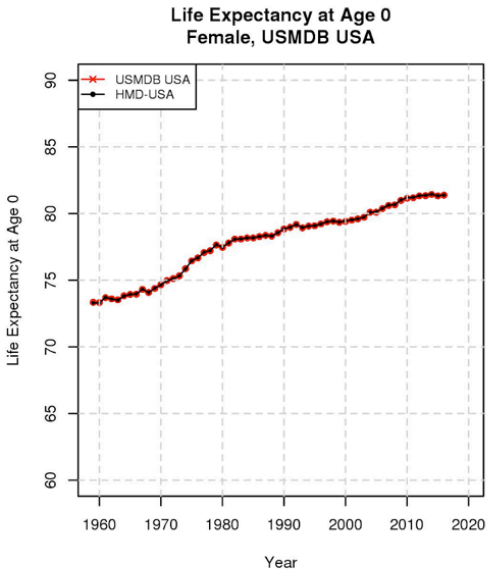
Verify USMDB against HMD

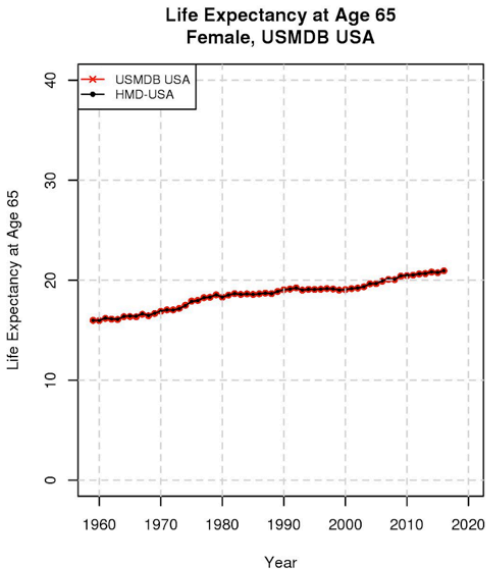
- Create USA-level aggregate from USMDB states
- Compare lifetable indicators against HMD

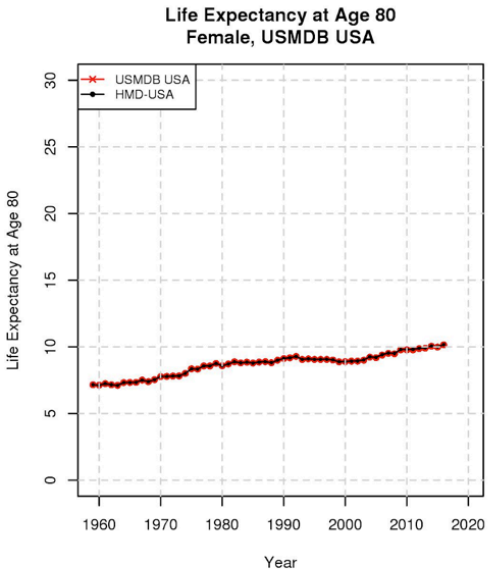
Discrepancies arise from

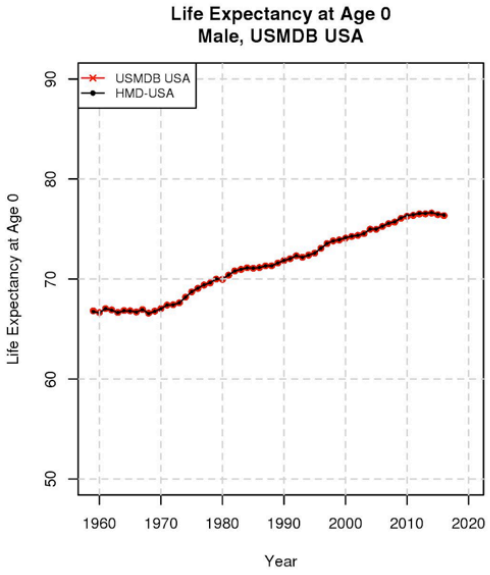
- Intercensal estimation and age-group fitting: 1959-1979
- Annually revised State pop. estimates 2010-2017
- V.5 vs. V.6 (1945 cohort)

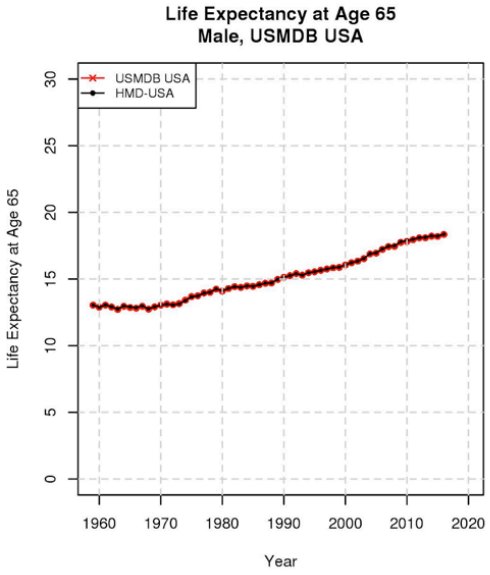
USMDB vs. HMD estimates of e_x

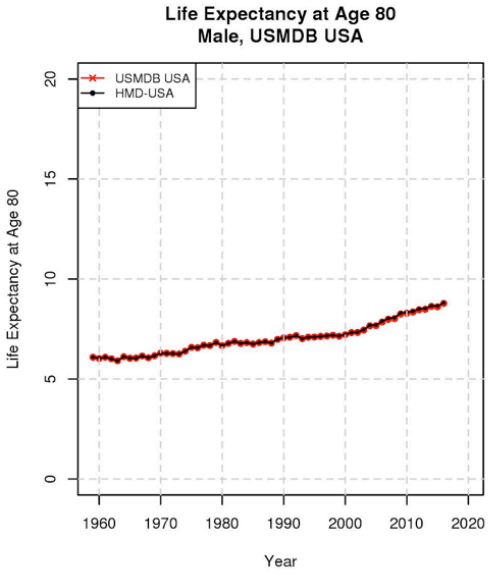












Website & Data-products

- http://usa.mortality.org
  - Anyone can visit the site.
  - Users must register (ID + password) to access data-products
- Data available: complete (1x1) & abridged (1x5,1x10,5x1,5x5,5x10) sex-specific lifetables, ages 0-110+, by state, division, region,USA for years 1959-2016
  - CSV and tab-delimited formats
  - No raw inputs, Lexis database, Exposures, raw mortality rates
- Also available, visually interactive map-based diagnostic tool for glancing at life-expectancy (e_x) and mortality rates (m_x) by state
  - http://shiny.demog.berkeley.edu/hmd/USHMD_MapApp/
  - developed with R-shiny by D. Dukhovnov

Questions

Is it better to use Census population estimates or official intercensal estimates, when available?

- More age detail in Census measurements, but more measurement error?

Validity of HMD methods for high-age mortality patterns in low population states?

How to account for interstate migration?

Future Directions

- Develop more user-driven data visualization for website
- Estimate mortality rates at the US-county level using Bayesian estimation
  - ~3,150 counties, ranging in population from <100 to >1e6
  - Work in progress - funded by NIH grant

Acknowledgments

Funding provided by the U.S. National Institute on Aging (grant R01-AG040245) and the Society of Actuaries (REX-Pool Fund) with additional support from the Center on the Economics and Demography of Aging (CEDA) at UC Berkeley and from the French National Institute for Demographic Studies.

Disclaimer: The author is solely responsible for the content of this presentation, which does not necessarily represent the official views of the National Institutes of Health and other sponsors.