Application of a Singular Value Decomposition-based Factorization and Parsimonious Component Model of Mortality to HIV Epidemics in Africa

Samuel J. Clark, Patrick Gerland,
Brian Houle, Sara Hertog, Jonathan Muir

Annual Meeting of the Population Association of America
Session 41: Statistical Modeling and Smoothing in Demography
April 7, 2022

SVD-Comp project

Motivation

- One standard modeling framework for ASMR
- Single year of age model life tables for UNPD
- Incorporate unusual effects of HIV, conflict, and natural disasters via summary indicators
- 'Rotation' at older ages and low overall mortality
- Smoothing, interpolation, (some) extrapolation

Background

- Builds on *long* history of mortality models that utilize dimension reduction approaches, e.g. Bourgeois-Pichat (1962); United Nations, Department of Economic and Social Affairs, Population Division (1982); Bourgeois-Pichat (1990); Wilmoth (1990); Lee and Carter (1992); Clark (2001); Wilmoth et al. (2012); Sharrow et al. (2014); Alexander et al. (2016); Clark (2019), etc.

Plan

We want something like this

$$\text{Model}({}_5q_0, {}_{45}q_{15}, P_{HIV}, \dots) \rightarrow {}_1q_x$$

Overview of plan

- Comprehensive database of empirical single-year ASMR
- ASMR affected by **HIV**, conflict, and natural disaster
- Using data, calibrate and validate SVD-Comp
- Software to implement and recalibrate SVD-Comp in an automated way

SVD-Comp idea: SVD

(Good, 1969; Stewart, 1993; Strang, 2009)

Singular value decomposition of matrix $\mathbf{Q}$ of mortality probabilities, age (rows) $\times$ life table (columns)

$$\mathbf{Q} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}} \tag{1}$$

The age schedule of each life table ℓ is exactly

$$\mathbf{q}_\ell = \sum_{i=1}^{\rho} s_i v_{\ell i} \mathbf{u}_i \tag{2}$$

where i indexes columns of $\mathbf{U}$ and $\mathbf{V}$, and ρ is the rank of $\mathbf{Q}$

The age schedule of each life table ℓ very well approximated as

$$\mathbf{q}_\ell \approx \sum_{i=1}^{4} \underbrace{v_{\ell i}}_{w_i} \cdot \underbrace{s_i \mathbf{u}_i}_{\mathbf{c}_i} \tag{3}$$

SVD-Comp idea: model

(Clark, 2019)

Estimate models $f_i(\cdot)$ relating elements of right singular vectors $\mathbf{V}_i$ to covariate indicators for each life table ℓ

$$v_{\ell i} \approx f_i(\text{indicators}_\ell) \tag{4}$$

For new mortality schedule, predict weights from f using desired indicator values

$$\widehat{w}_i = f_i(\text{indicators}) \tag{5}$$

Combine predicted weights with components from SVD of life table universe to produce predicted life table in single-year age groups

$$\widehat{\mathbf{q}} = \sum_{i=1}^{4} \widehat{w}_i \cdot \mathbf{c}_i \tag{6}$$

Data 1

Data Sources for Vital Registration Life Table Universe

Source	Abridged and/or complete	Number of countries or areas represented	Earliest year	Latest year	Number of life tables
Human Mortality Database (HMD)	Complete	47	1751	2018	10,161
Human Life-Table Database (HLD)	Both	99	1778	2015	6,863
Global Burden of Disease (GBD) life tables	Abridged	112	1950	2016	11,958
Latin American Mortality Database(LAMBdA)	Abridged	17	1908	2008	234
UN Demographic Yearbook (DYB)	Both	181	1908	2018	5,984
WHO Historical Life Tables Databank	Abridged	63	1900	1999	4,066
Total					39,266

Data 2

Acceptable Life Tables by Region and Source: Reject about 75%

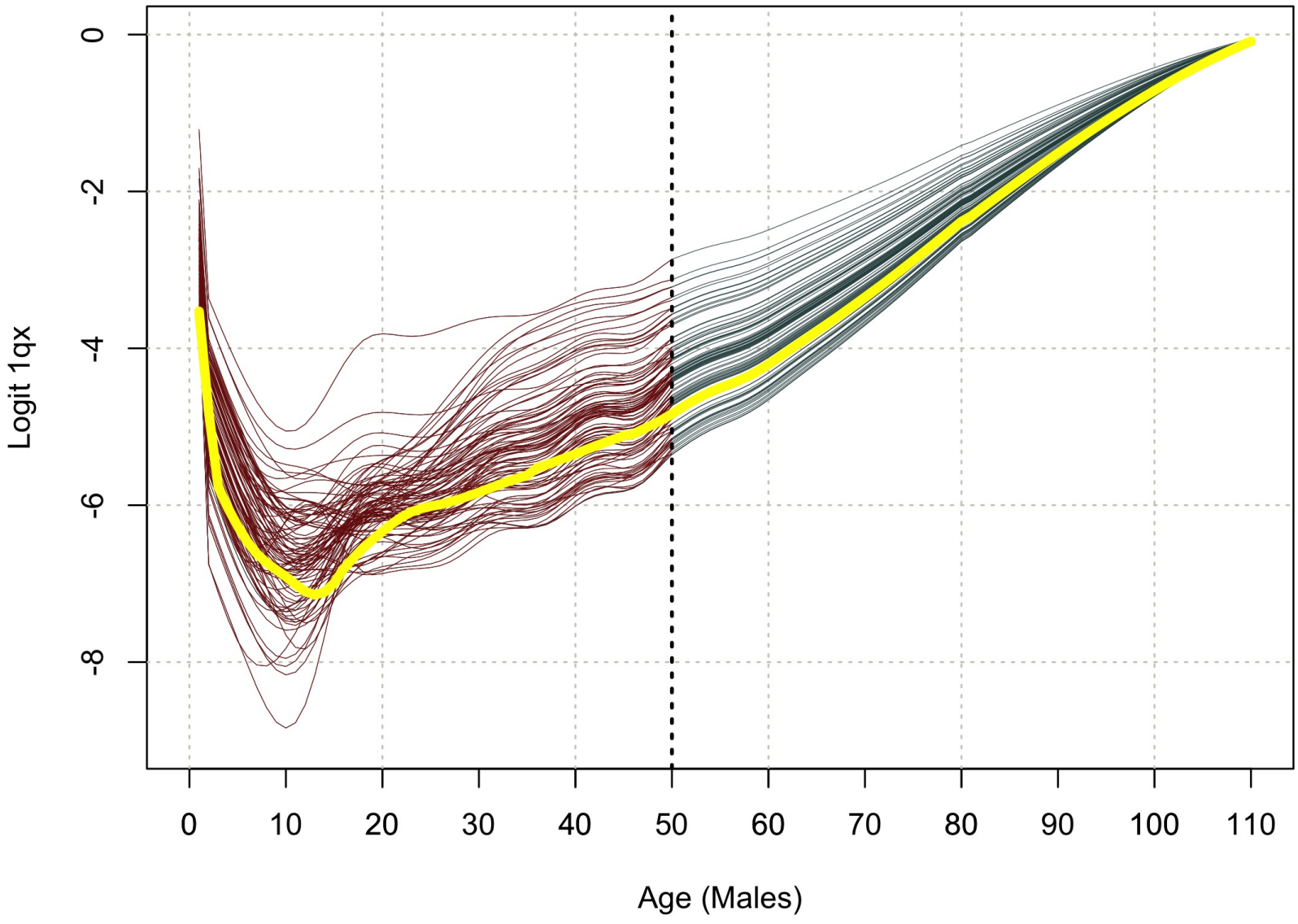
	DYB	GBD	HLD	HMD	LAMBdA	WHO	Total
More developed regions							
Australia/New Zealand	0	0	0	402	0	0	402
Europe and Northern America	86	226	25	7,196	0	77	7,610
Less developed regions							
Central and Southern Asia	62	305	65	0	0	6	438
Eastern and South-Eastern Asia	44	250	28	328	0	9	659
Latin America and the Caribbean	102	1,491	26	34	113	49	1,815
Northern Africa and Western Asia	53	221	20	68	0	6	368
Oceania	0	6	0	0	0	0	6
Sub-Saharan Africa	6	103	12	0	0	6	127
Total	353	2,602	176	8,028	113	153	11,425

Data 3

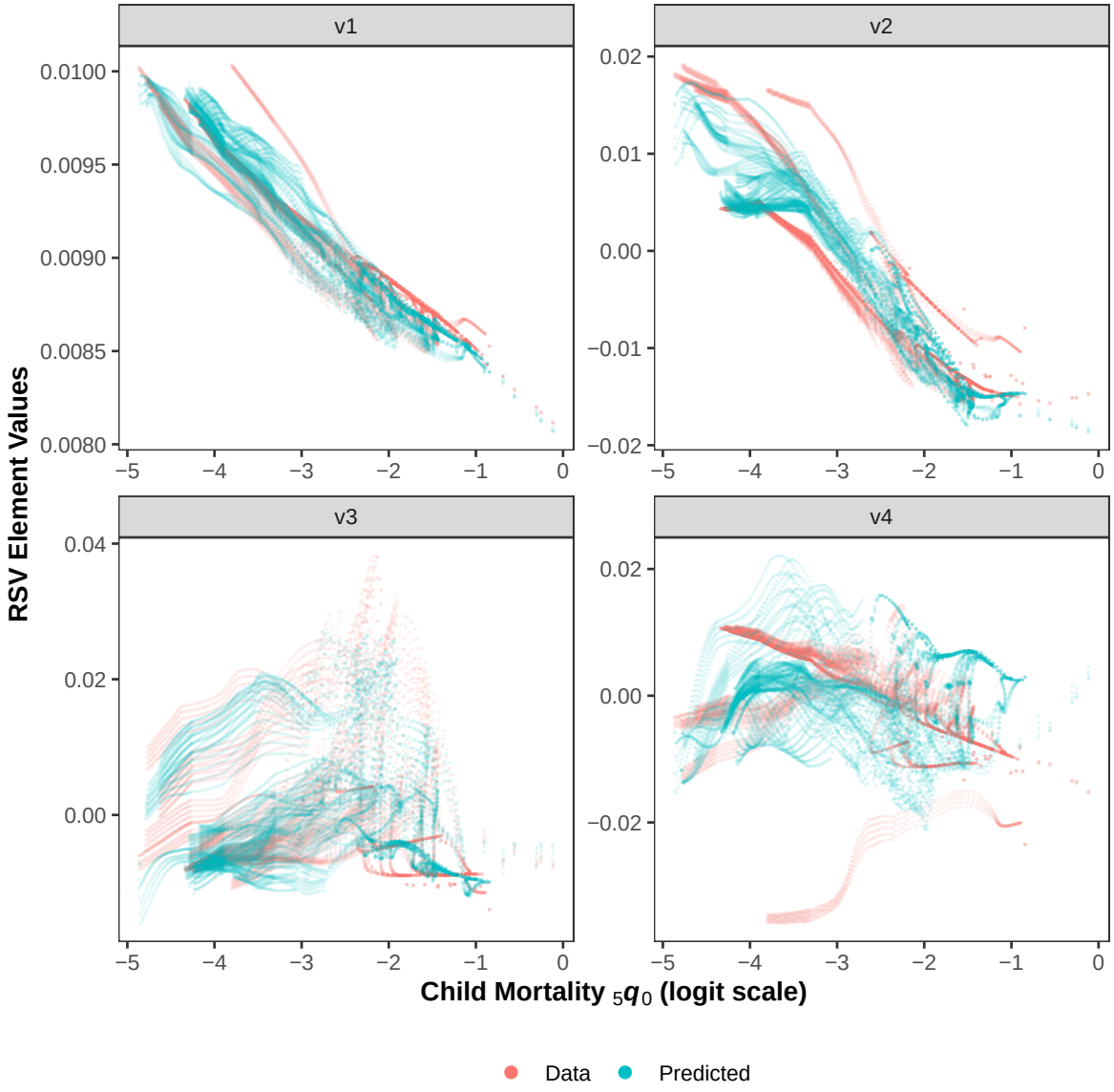
Supplemental Life Tables from Non-VR Sources

Source	Abridged and/or complete	Number of countries or areas represented	Earliest year	Latest year
Demographic and Health Surveys (DHS)	Truncated life tables constructed from: 1) full birth histories for ages 0-14; 2) sibling survival histories for ages 15-49 for the 10-year period preceding each survey	168 surveys from 61 countries	1988	2018
Spectrum simulated life tables	Life tables simulated for each country-period with 5 scenarios for HIV incidence (observed, +/-10% and 20%) and common ART assumptions	18 countries with HIV prevalence $>$ 5% at any point during 1980-2018 per UNAIDS (11,790 LTs)	1970	2100

Extending truncated DHS life tables



Spectrum simulated data RVS models

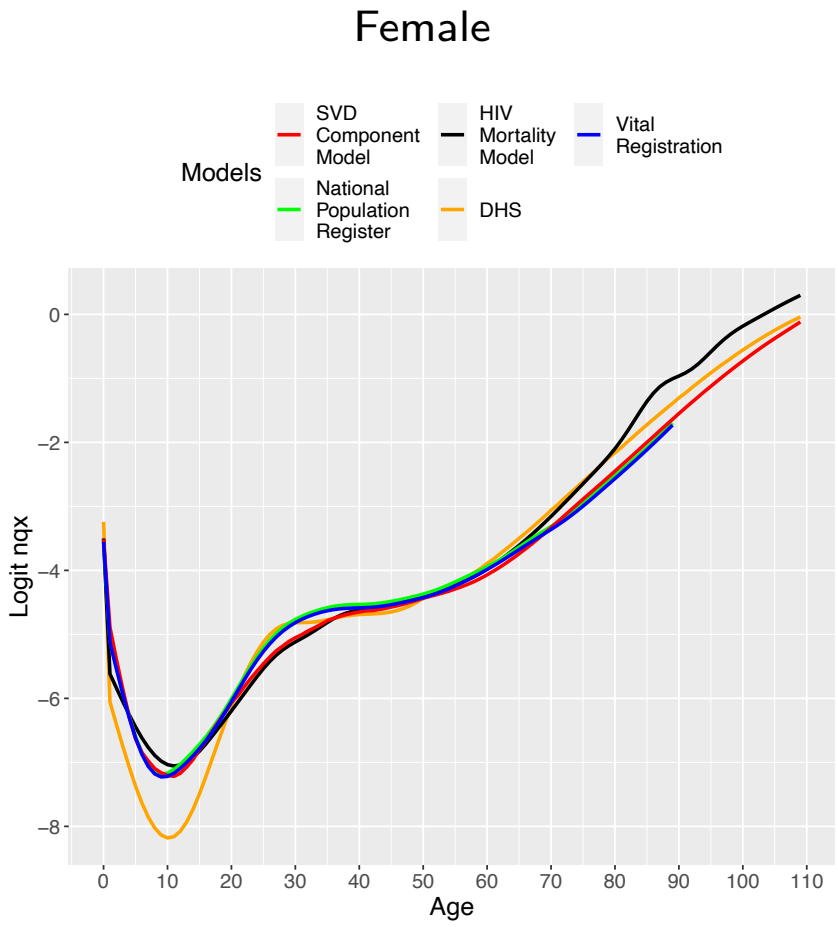


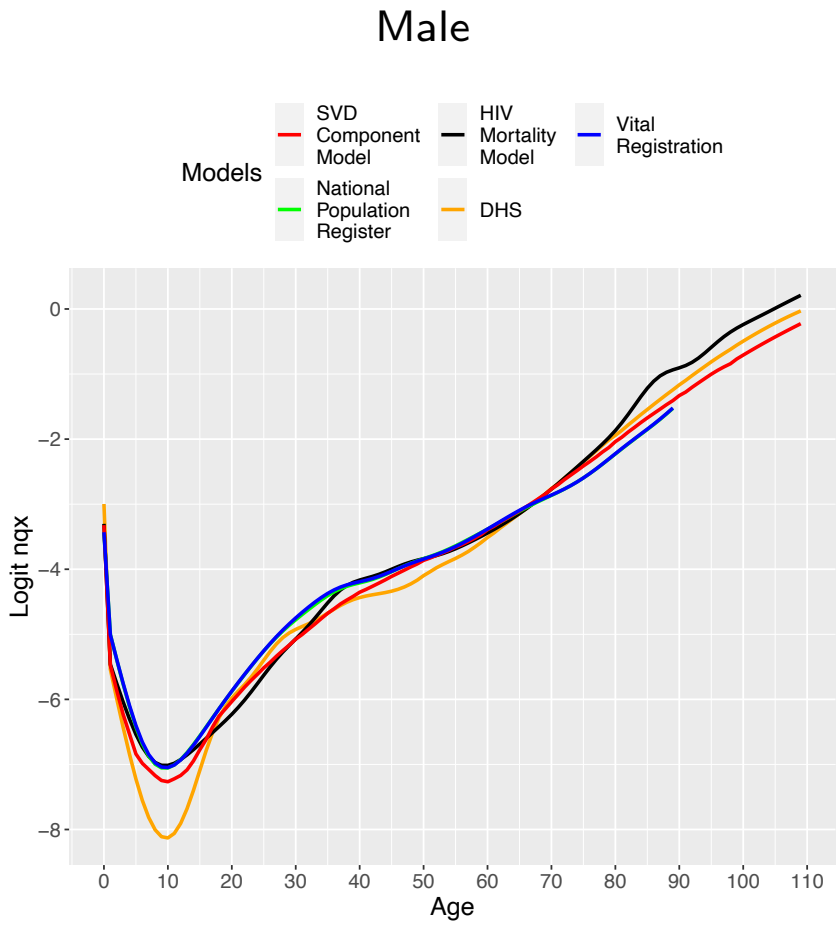
Model includes

- $_5q_0$
- $_{45}q_{15}$
- HIV prevalence
- ART coverage

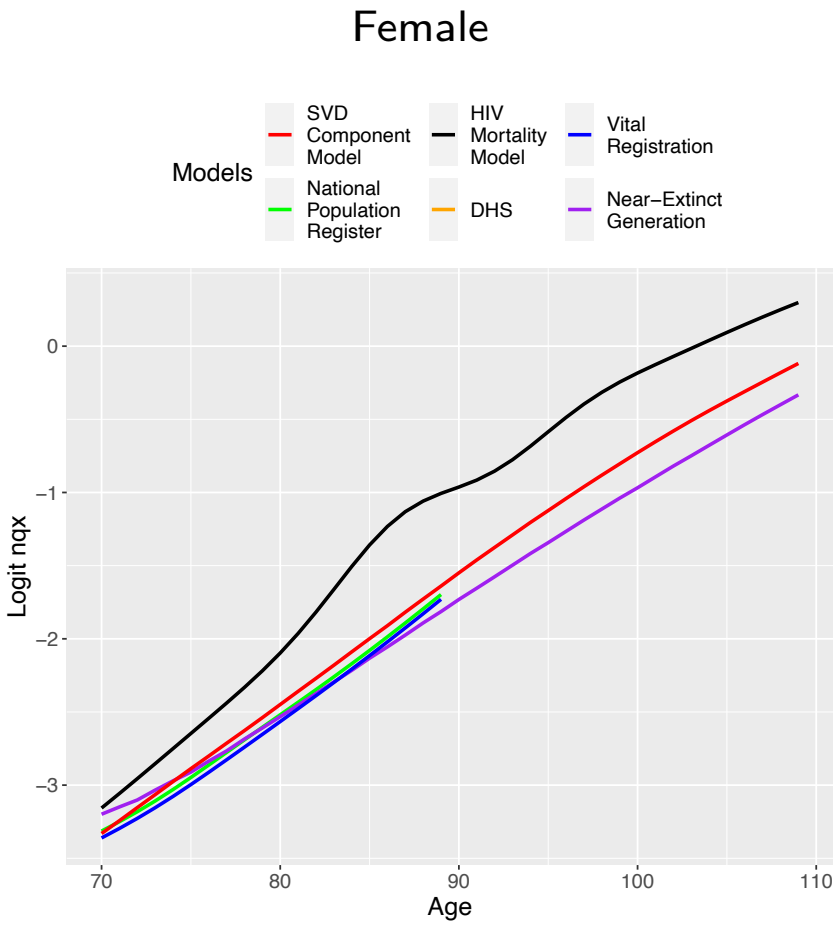
Estimated using restricted cubic splines with 7 knots

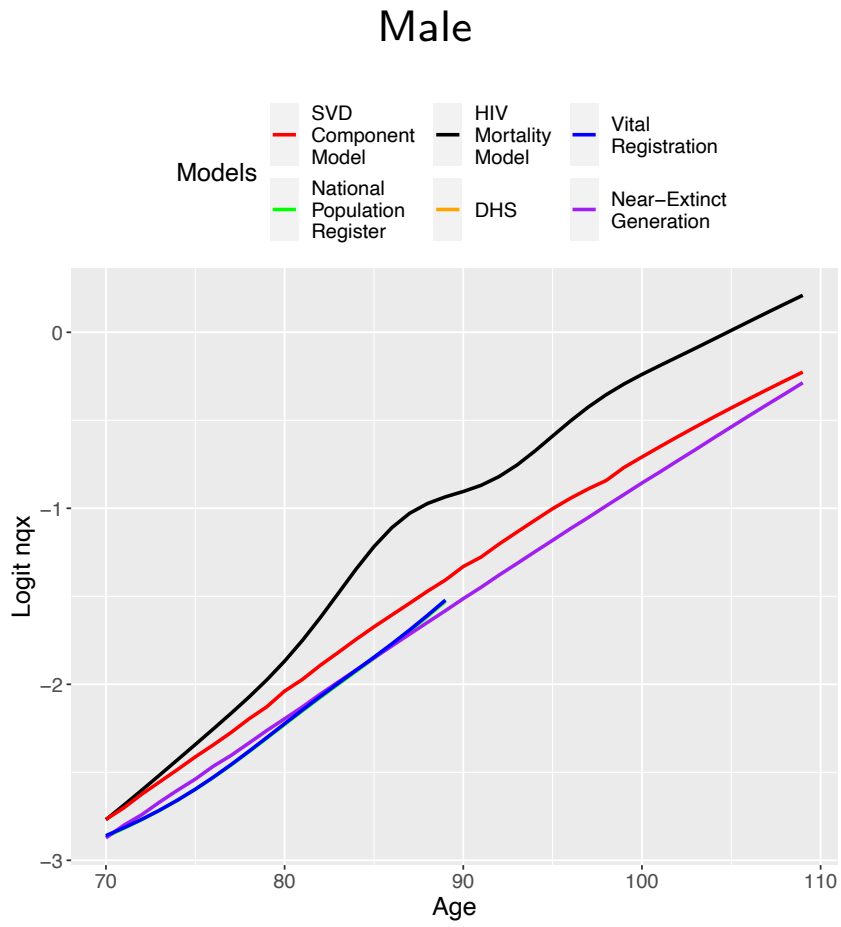
Example prediction: HIV country





Example prediction ages 70+: HIV country





Plan

Now

- Finish compiling empirical life table database
- Include detailed 1-year child mortality data, similar to truncated DHS
- Refine right singular vector models
- Build rotation model
- Characterize performance and validate overall SVD-Comp model
- Create R packages and R Shiny GUIs to implement in a user-friendly way

Medium term

- **Replace/augment simulated Spectrum HIV mortality schedules with data from ALPHA Network of HIV surveillance HDSS sites**
- Add conflict and natural disaster data
- Refine R packages to allow semi-automated recalibration of overall model as empirical life table database grows

Publication fully describing this approach

(Clark, 2019)

Demography (2019) 56:1131–1159
https://doi.org/10.1007/s13524-019-00785-3

A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality

Samuel J. Clark[1,2]

Published online: 28 May 2019

- Full reproducibility materials at https://github.com/sinafala/svd-comp
- `R` package that implements model
  - https://github.com/sinafala/svd-comp/tree/master/package
  - Install using *devtools*: `install_github(repo='sinafala/svdComp5q0')`

References I

Alexander, M., E. Zagheni, and M. Barbieri (2016). A flexible bayesian model for estimating subnational mortality. *arXiv preprint arXiv:1607.03534*.

Bourgeois-Pichat, J. (1962). Factor analysis and sex-age-specific death rates: a contribution to the study of the dimensions of mortality. *United Nations Population Bulletin* (6), 147–201.

Bourgeois-Pichat, J. (1990). Application de l'analyse factorielle à l'étude de la mortalité. *Population (french edition) 45*(4-5), 773–802.

Clark, S. J. (2001). *An Investigation into the Impact of HIV on Population Dynamics in Africa*. Ph. D. thesis, University of Pennsylvania, Philadelphia.

Clark, S. J. (2019). A General Age-Specific Mortality Model with An Example Indexed by Child or Child/Adult Mortality. *Demography (Reproducibility code) 56*(3), 1131–1159.

Good, I. J. (1969). Some applications of the singular decomposition of a matrix. *Technometrics 11*(4), 823–831.

Lee, R. D. and L. R. Carter (1992). Modeling and forecasting US mortality. *Journal of the American statistical association 87*(419), 659–671.

Sharrow, D. J., S. J. Clark, and A. E. Raftery (2014). Modeling age-specific mortality for countries with generalized hiv epidemics. *PloS ONE 9*(5), e96447.

References II

Stewart, G. W. (1993). On the early history of the singular value decomposition. *SIAM review 35*(4), 551–566.

Strang, G. (2009). *Introduction to Linear Algebra 4e*. Wellesley-Cambridge Press.

United Nations, Department of Economic and Social Affairs, Population Division (1982). *Model life tables for developing countries*. Number 77. New York: United Nations Department of International Economic and Social Affairs Population Division.

Wilmoth, J., S. Zureick, V. Canudas-Romo, M. Inoue, and C. Sawyer (2012). A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies 66*(1), 1–28.

Wilmoth, J. R. (1990). Variation in vital rates by age, period, and cohort. *Sociological Methodology 20*, 295–335.