Adult mortality estimates from censuses and surveys: ongoing work on youth mortality and possible extensions

Masquelier Bruno

UCLouvain, Belgium

EGM WPP 2021, April 2020

Mortality among adolescents and young people aged 15-24 (${}_{10}q_{15}$): data sources

More than 6500 country-year data points from:

1. Vital registration data (covering 27% of the population aged 15-24),
2. Estimates from sample vital registration systems (e.g. India) and Rapid Mortality Surveillance in South Africa.
3. Nationally-representative sample surveys (DHS, MICS, WHS) providing sibling histories
   - ▷ Direct estimation based on ages at survey/at death and the timing of deaths
   - ▷ Stratified jackknife approach used for variance estimation, with reference periods optimized based on the coefficient of variation of the estimates ($< 20\%$).
4. Rates from reports on recent household deaths in censuses and large-scale surveys (e.g. in China, Pakistan).

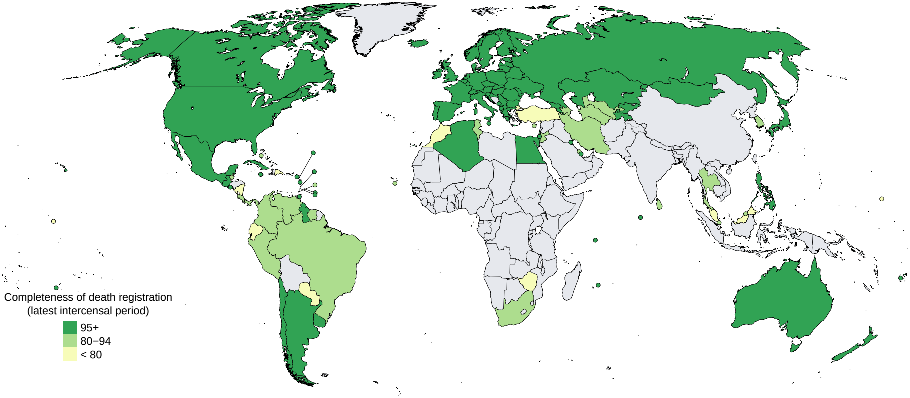
A database similar to UN IGME databases for U5MR and ${}_{10}q_5$, except that we use sibling histories from surveys (instead of birth histories).

Completeness assessment

VR data available for 145 countries from WHO, including 38 HMD countries. For other countries, death distribution methods (DDMs) are used to estimate completeness when possible (73 countries, 116 intercensal periods).

- Estimates are based on the hybrid method (GGBSEG), which consists in using first the GGB to obtain the coverage of census 1 relative to census 2 and adjusting populations accordingly prior to using the SEG method.
- No consensus on age trimming (Murray et al., 2010; Hill et al., 2009): range selected automatically here, as implemented in the DDM package in R (Riffe et al., 2017).
- VR data excluded when completeness is below 80%, adjusted if between 80% and 95%, included without adjustment if higher than 95%. Estimates truncted at 100%.

A set of completeness estimates based on an entirely **reproducible** assessment, using the WHO mortality database and the UN Demographic Yearbook census data.

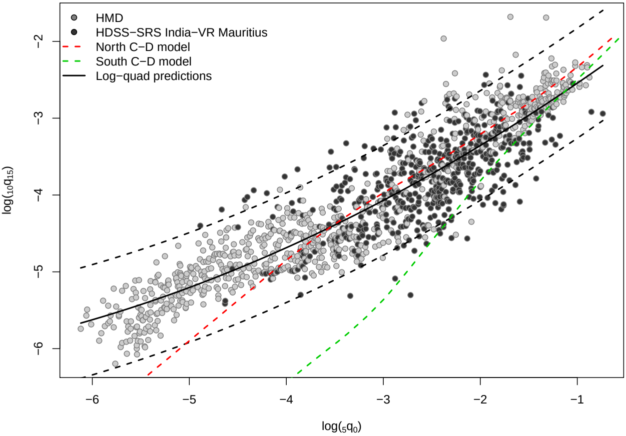


Some deviations from GBD or WHO smoothed estimates of completeness › more work is needed to understand these differences.

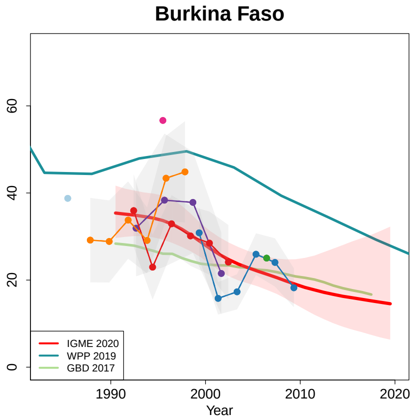
Evaluation of survey and census data based on age patterns

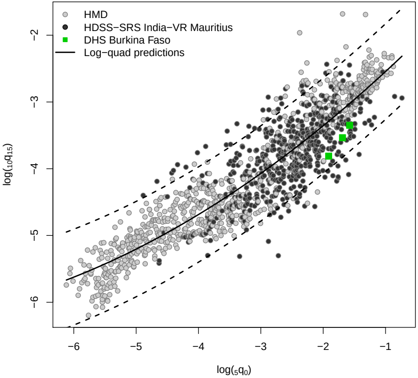
1. HMD 5x5 life tables
2. HDSS life tables (50 sites)
3. Indian SRS
4. Mauritius VR

$$log({}_{10}q_{15}) = \beta_0 + \beta_1 \log(\text{U5MR}) + \beta_2 \log(\text{U5MR})^2$$



This automated approach results in the exclusion of 13 DHS (out of 151), 3 MICS (out of 9), 17 WHS out of 42. It also helps in assessing the plausibility of the estimates for some countries with deviations from WPP.

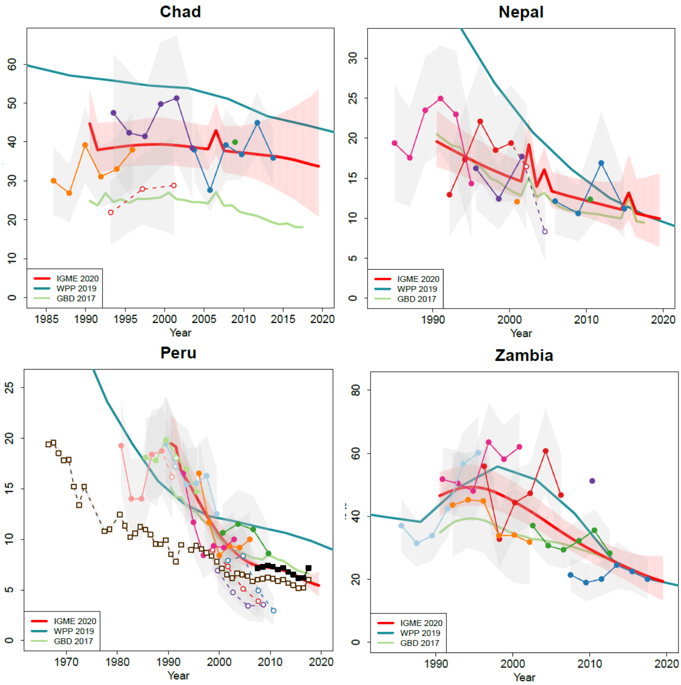




Statistical model for ${}_{10}q_{15}$

Point estimates of ${}_{10}q_{15}$ and 90% uncertainty intervals are obtained from the Bayesian B-spline bias-reduction (B3) model developed by Alkema and New (2014) (already used for U5MR and ${}_{10}q_5$).

- The model accounts for stochastic standard errors for VR and SVR, potential bias in trends and levels in survey estimates and possible outliers.
- Estimates are produced for 195 countries, from 1990 to 2020. No covariates.
- A penalized spline regression model is used to estimate country-specific trends in the logit transformation of r ($log(r/(1-r))$), where r is the ratio of the probability ${}_5q_{15}$ to the median B3 estimate of ${}_{10}q_{15}$. The probability ${}_5q_{15}$ is constrained to be lower than ${}_{10}q_{15}$.
- Adjustments are introduced for conflicts and disasters, based on GHE estimates.

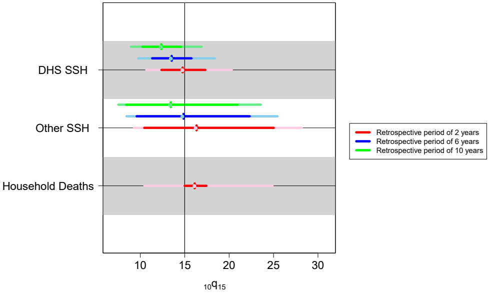


Chad
Nepal
Peru
Zambia
IGME 2020
WPP 2019
GBD 2017
Year

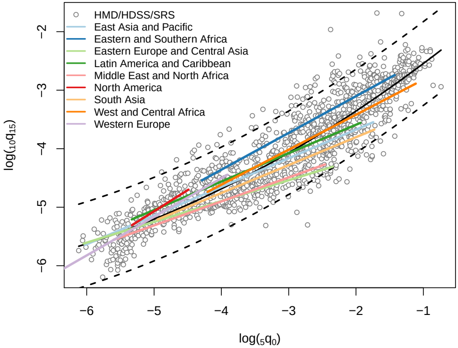
Possible underestimation of mortality in SSH, but little evidence on the 15-24 age group.

- Helleringer et al. (2014): SSH lead to an underestimation of mortality, but especially from the age of 50 onwards. Prior to that, SSH estimates track underlying mortality level in the HDSS.
- This was due to a lower proportion of omissions for younger age groups (compared to sisters aged 45 and over) and a tendency to underestimate ages at death, raising mortality estimates for younger age groups.

Biases captured by the B3 model in terms of levels and trends:

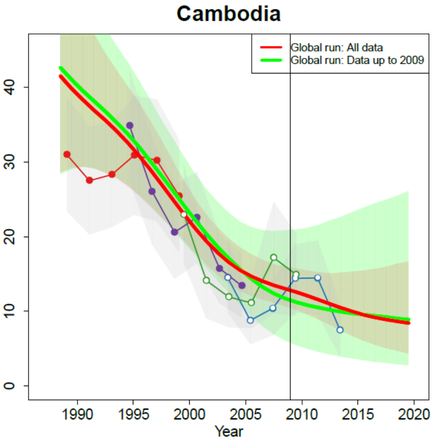


In 38 countries, there are not enough data inputs to estimate the probability ${}_{10}q_{15}$ from VR, surveys or censuses.



- This represents about 5% of the population aged 15-24.
- A linear model with random intercept and slope for each region is used to regress $\log({}_{10}q_{15})$ against log(U5MR).
- There is room for improvement here. Ex: modelling the ${}_{10}q_{15}$-to-${}_{5}q_{0}$ relationship based on data in a structured model would allow including the limited data available for some countries.

Out-of-sample validation for ${}_{10}q_{15}$

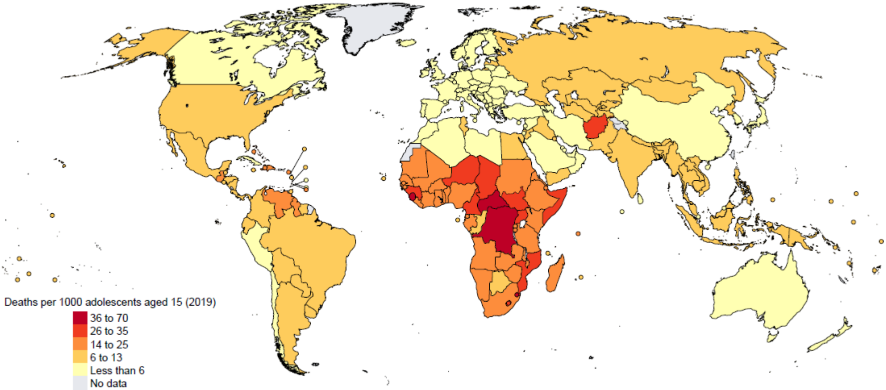


A training set is constructed by removing all data collected in or after 2009 (i.e. about 20% of observations are left out).

- Validation results are based on left-out observations and on the comparison between estimates obtained from the training and full data set.
- The validation suggests that the B3 model performs reasonably well for estimating mortality in youth (comparable to when applied to children aged 5-14), but there is still room for improvement.

2019 estimates: adolescents and young adults (15-24)

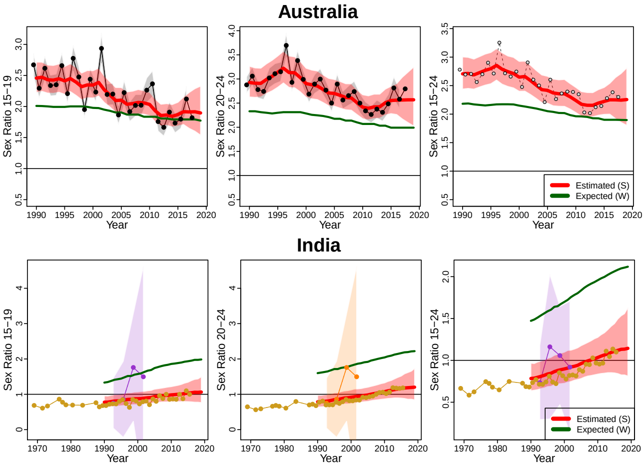
Globally, the probability of dying between the ages 15 and 25 was 10.9 deaths per 1000 adolescents aged 15 (90% confidence interval 10.4-12.7) in 2019.



$\mapsto$ This is higher than the global estimate of child (1-4) mortality in 2018 (10.0 per 1000 (9.4-11.3))

Quantifying the excess male mortality

Fengqing Chao is currently exploring the use of the Bayesian Hierarchical linear model developed by Alkema et al. (2014) to capture the relation between the total mortality levels and the sex ratios of mortality in the age groups 5-14 and 15-24.



Potential extension

The B3 model could be used to estimate adult mortality from age 15 to 50 in future revisions of the WPP (but not from 1950). In combination with U5MR trends **and** the log-quad/SVD models, complete life tables (including annual) can be derived.

- Estimating ${}_{35}q_{15}$ and using log-quad or SVD directly with U5MR?
- Modelling age-specific probabilities separately?
  - There is a vast amount of orphanhood data that can be used to estimate ${}_{10}q_{25}$ (not ${}_{10}q_{15}$) and cover the 1970s and 1980s.
  - Age patterns and sex ratios within the 15-49 age range would be data-driven
  - Biases in censuses and surveys could also vary by age
- Or a mixed approach, whereby ${}_{10}q_{15}$, ${}_{15}q_{25}$ and ${}_{10}q_{40}$ are estimated jointly from a single database, in a structured model reflecting the age patterns.

Acknowledgements

Thanks are due to Lucia Hug, David Sharrow, Danzhen You, Patrick Gerland, Benjamin-Samuel Schlüter, Fengqing Chao, Leontine Alkema.

References

L. Alkema, F. Chao, D. You, J. Pedersen, and C. C. Sawyer. National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: a systematic assessment. *Lancet Glob Health*, 2(9):e521–e530, Sep 2014.

Leontine Alkema and Jin Rou New. Global estimation of child mortality using a bayesian b-spline bias-reduction model. *The Annals of Applied Statistics*, 8(4):2122–2149, 2014.

S. Helleringer, G. Pison, M. Kanté, G. Duthé, and A. Andro. Reporting errors in survey data on adult mortality: results from a record linkage study in Senegal. *Demography*, 51(2): 387–411, 2014.

K. Hill, D. You, and Y. Choi. Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data error. *Demographic Research*, 21:235–254, 2009.

Ch. Murray, J.l Rajaratnam, J. Marcus, Th. Laakso, and A. Lopez. What Can We Conclude from Death Registration? Improved Methods for Evaluating Completeness. *PLoS Med*, 7(4): e1000262, 04 2010.

Tim Riffe, Everton Lima, and Bernardo Queiroz. *DDM: Death Registration Coverage Estimation*, 2017. URL `https://CRAN.R-project.org/package=DDM`. R package version 1.0-0.