IUSSP 2025 Submission

Extended Summary

Introduction

Age-specific mortality rates are one of the most foundational demographic indicators in the study of populations and global health. In places with reliable data from civil registration and vital statistics systems, age-specific mortality can be calculated in a straightforward manner. In many low and middle income countries (LMICs) without such data, other sources are available, generally from surveys. Commonly used data include complete birth histories, summary birth histories, sibling survival histories, and household death recall. These data have traditionally been used to estimate mortality in children and adults, separately. To then estimate age-specific mortality, model life table systems are the most common method. Model life table systems take input parameters—usually child and adult mortality—and generate an age pattern from observed life tables in other settings and/or demographic relationships.

The United Nations' World Population Prospects (WPP)[1] and the Institute for Health Metrics and Evaluation's Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) are two of the most popular sources for age-specific mortality estimates, and both sources use different but related model life table systems. These methods generalize age patterns from data in locations with high quality vital registration systems, or from observed demographic relationships in these locations, and apply them to LMICs that are qualitatively different, and thus may not fit the same profile of age-specific mortality that a model life table system would impose. For example, cardiovascular disease and other non-communicable diseases have historically been lower in sub-Saharan Africa compared to high income countries[3], which could lead model life table systems to overestimate older age mortality. Furthermore, model life table systems rely on first producing estimates of child and adult mortality in separate, disconnected models, and then using these estimates in another disconnected model. This approach makes propagating uncertainty through the process difficult and does not allow for appropriate accounting for correlated data among the different models. A single, unified method that models age-specific mortality rates directly from the data is more parsimonious, simpler to implement, and allows for accurate uncertainty estimation.

To address these issues, we propose a Bayesian hierarchical model with correlated splines and residual smoothing via kernel regression to directly model age-specific mortality rates. This approach allows for the nuanced understanding of mortality patterns across different age groups and locations over time, which is pivotal for effective public health planning and intervention.

Data

The model leverages a diverse set of data sources to improve the accuracy and reliability of age-specific mortality estimates. These sources include:

- Vital registration systems, which provide official records of births and deaths.
- Sample registration systems, offering periodic snapshots of demographic events in a sample population.
- Complete birth histories, detailing the births from a sample of mothers, allowing for the estimation of infant and child mortality rates.
- Summary birth histories, which summarize the reproductive history of mothers.
- Sibling survival histories, collected during surveys, providing indirect estimates of adult mortality.
- Household death recall, which can offer additional data on death occurrences within households across all ages.

Data from vital registration systems and sample registration systems are adjusted for completeness via the method in the GBD, and age-specific mortality rates are estimated as deaths divided by either GBD population estimates or sample populations from the sample registration systems. Data from complete birth histories are used to calculate mortality rates in ages under 15 using methods from the GBD, while summary birth histories are used for ages under 5. Sibling survival histories are used to calculate mortality for ages 15-49 using methods from the GBD. Data from household death recall in censuses and surveys are used to calculate age-specific mortality rates using methods from the GBD.

We use all available data from these sources that are in the Global Health Data Exchange[4] which can be accessed via the GBD Sources Tool on the website.

Methods

At the core of the proposed method is a Bayesian hierarchical correlated spline regression model, which is designed to directly model mortality rates across different age groups using predictive covariates, while borrowing strength across the age spectrum, across locations, and over time. This model is essentially a very flexible generalized linear regression model, as it can model continuous, count, and binary data with identity, log, and logistic link functions, and many specifications are available to test and use for modeling. Key features include nested location intercepts, splines across age, and splines across time, each of which can have interactions with each other and the predictive covariates. Furthermore, these location, year, age, and covariate effects can all have correlations specified across all dimensions. For example, we can specify covariate effects that are correlated for adjacent age groups or splines on time that are correlated across neighboring ages and locations. Additionally, the model incorporates a kernel regression component for the smoothing of residuals, further enhancing the model's ability to capture subtle patterns and trends in the data. This is a nonparametric approach that can be tuned with hyperparameters in order to capture similarities across all dimensions of the residuals.

We performed various simulation studies that demonstrate the validity of this modeling approach on mortality data created using data and covariates from the GBD with simulated

residuals. We then fit the best performing model to the full data set and compared it to estimates using previous GBD methodology. We also compared estimates to the UN WPP and compared results in a number of locations with the larges differences. Finally, we used data from Health and Demographic Surveillance Sites (HDSS) in LMICs to compare estimates, in order to demonstrate that our new method produces estimates that are more comparable to these observed data in LMICs than other methods.

Expected Findings

Preliminary results show accurate fits of this model to vital registration data, along with improved estimates in locations with sparse data compared to previous methods. Figures 1 and 2 show estimates for the United States and Kenya, respectively, compared to previous GBD and WPP estimates. Figure 3 shows our estimates and previous GBD estimates compared to HDSS data and available vital registration data in sub-Saharan Africa in terms of older adult mortality, 30q50, and younger adult mortality, 35q15. Analyses of simulated data and model validation using held out data are forthcoming.

This modeling approach is expected to yield several key findings:

- Improved accuracy and precision of age-specific mortality estimates, particularly in age groups and locations where data are traditionally sparse or unreliable. Notably, this will produce updated estimates of all-cause mortality during the COVID-19 pandemic period of 2020-2023.
- Enhanced understanding of mortality patterns and trends across different populations, including the identification of atypical mortality profiles and potential demographic and health-related determinants.
- Greater flexibility and adaptability of the model in incorporating various data sources and adjusting to different contexts, making it a valuable tool for demographic research and public health decision-making worldwide.

By directly modeling age-specific mortality rates and incorporating a wide array of data sources, the proposed method offers a significant improvement over traditional model life table systems. It has the potential to provide more accurate, nuanced insights into mortality patterns, which are crucial for developing effective health interventions and policies.

References

1. United Nations Department of Economic and Social Affairs. World Population Prospects (2024 Revision). (UN DESA/POP/2024/TR/NO. 9)
2. Schumacher AE et al. Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950–2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the Global Burden of Disease Study 2021. The Lancet. 2024 May 18;403(10440):1989-2056.

3. Yuyun MF, Sliwa K, Kengne AP, Mocumbi AO, Bukhman G. Cardiovascular diseases in sub-Saharan Africa compared to high-income countries: an epidemiological perspective. Global heart. 2020;15(1).
4. Institute for Health Metrics and Evaluation. Global Health Data Exchange. ghdx.healthdata.org. Accessed 2024 September 1.

Figure 1

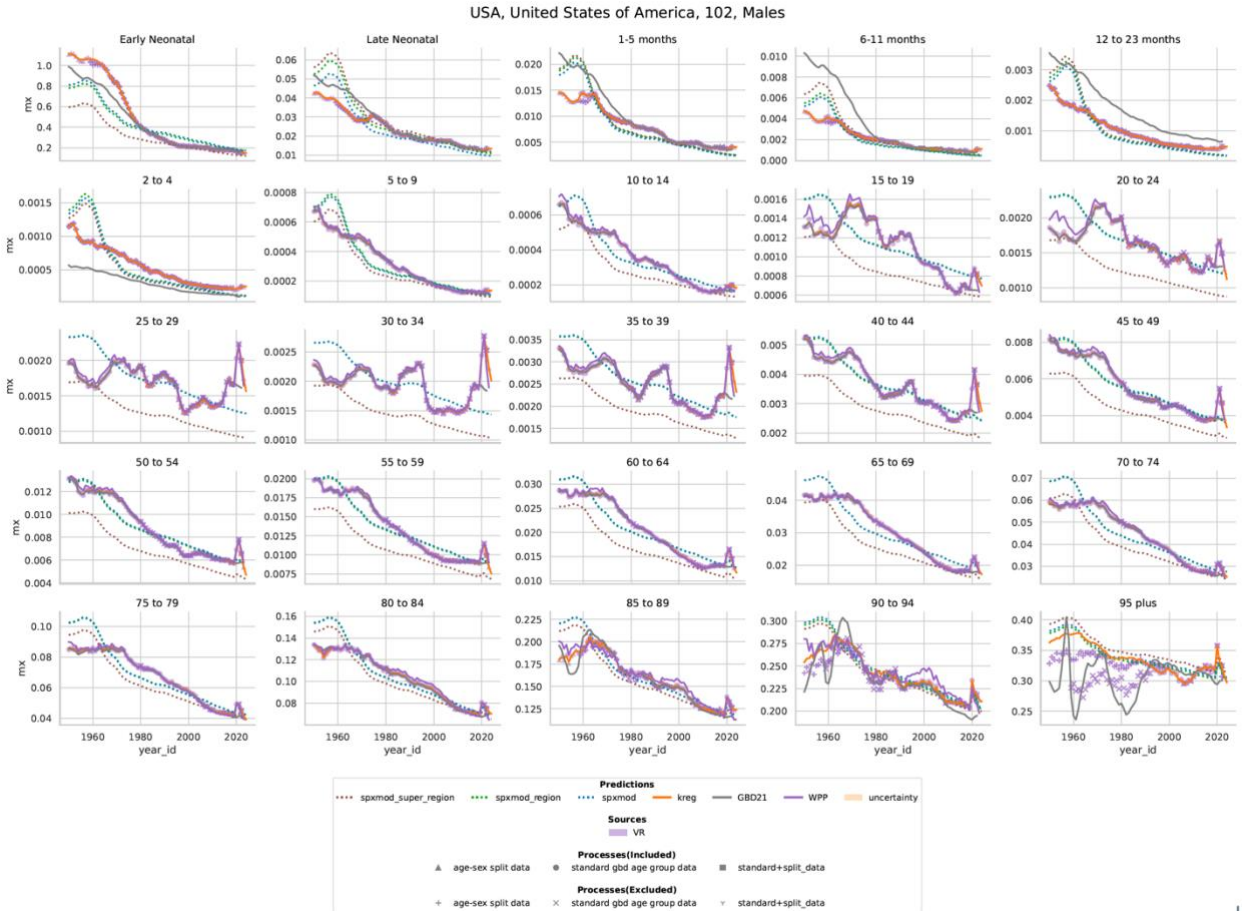

Figure 2

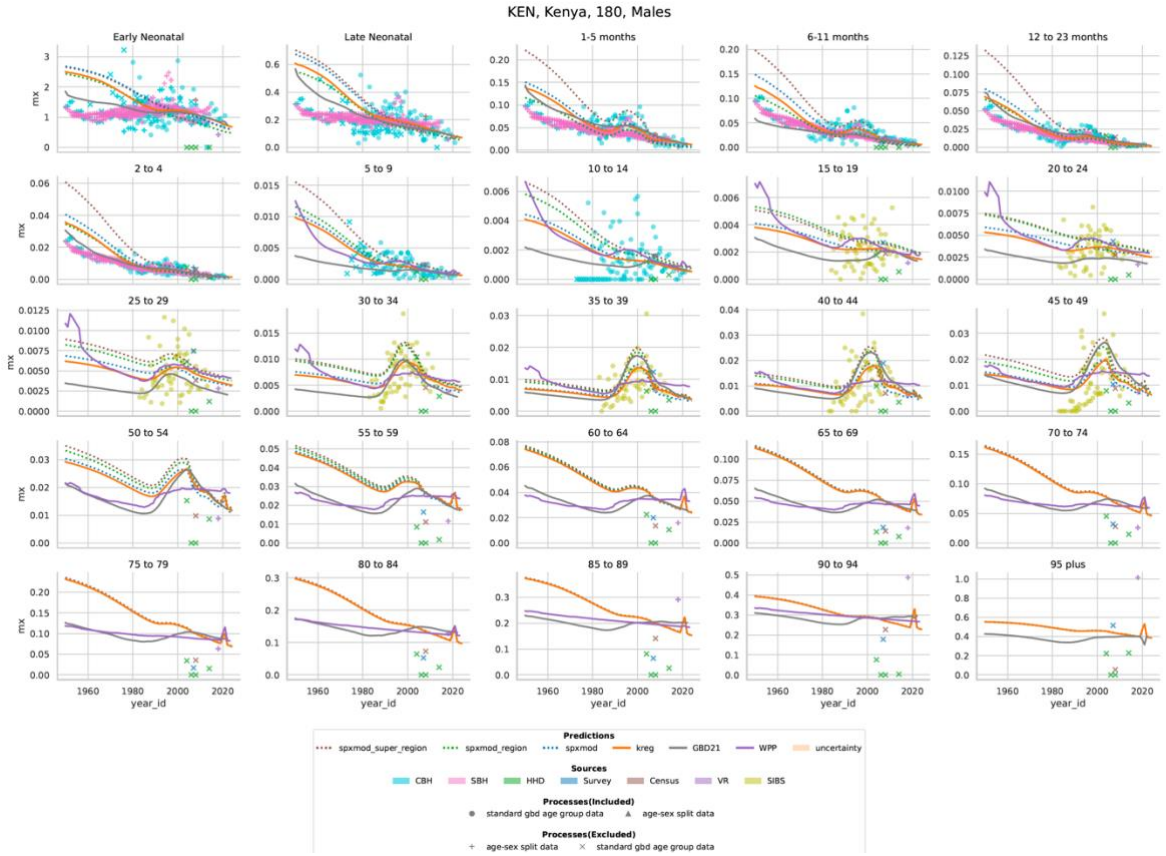

Figure 3