Modeling age-specific mortality by detailed age between 0 and 5 years: Results from a log-quadratic model applied to high-quality vital registration data

Michel Guillot, University of Pennsylvania and French Institute for Demographic Studies (INED)
Julio Romero Prieto, Central Bank of Colombia
Andrea Verhulst, University of Pennsylvania
Patrick Gerland, United Nations

Introduction

The Under-5 Mortality Rate (U5MR) is a key and widely-used indicator of child health, but it conceals important information about how this mortality is distributed by age. For better understanding and monitoring of child health, it is critical to examine how the risk of death varies within the 0-5 age range. This includes age breakdowns beyond the standard cut-off points of 28 days (for neonatal mortality) and 1 year (for infant mortality). In many populations, however, the age pattern of under-5 mortality is not well known. Less-developed countries, in particular, lack the high-quality detailed vital registration information necessary for the analysis of such age patterns. Sample surveys collecting retrospective birth histories do not satisfactorily fill this gap, because they are subject to systematic biases that are particularly consequential for estimating age patterns. This makes the need for high-quality information on age patterns of under-5 mortality even more critical, because regularities in these age patterns can be used as a powerful tool for evaluating and correcting data when sources are deficient.

The goal of this paper is to propose a new model for summarizing regularities about how under-5 mortality (U5M) is distributed by detailed age in human populations. This model is based on a newly-compiled database on under-five mortality by detailed age in countries with high-quality vital registration systems, covering a wide array of mortality levels and patterns. Building on previous work by Wilmoth et al.,[1] this model uses a log-quadratic approach, predicting a full mortality schedule between age 0 and 5 on the basis of only 1 or 2 parameters. We present applications of this model for evaluating and correcting under-5 mortality information by detailed age in countries with deficient vital registration.

This paper builds on the existing literature on model life tables. Model life tables, such as the ones developed by Coale and Demeny[2,3] or the United Nations,[4] represent a useful framework because they allow the estimation of arrays of age-specific mortality rates on the basis of only one or two mortality indicators, chosen as entry parameters.[5] When the model correctly represents the population's age pattern of mortality and the entry parameters are accurate, this approach can produce precise results. This approach is in fact commonly used by international organizations to derive the infant mortality rate (IMR) on the basis of U5MR. Current usage of model life tables for estimating patterns of U5M, however, is affected by several important drawbacks. First, existing model life tables only offer 0 vs. 1-4 as an age breakdown for U5M. This is insufficient for most purposes, including for the estimation of NMR or mortality in non-standard age ranges. (One model that contains additional age details is Bourgeois-Pichat's "biometric" model. This model, however, focuses on the first 12 months of age only and has been shown to poorly fit data in a variety of contexts.[6-9]) Second, existing model life tables rely on rather old data, with the most recent information dating back to the early 1980's. Third, the relationships based on specific regional patterns included in these model life tables are known to

hold only for limited periods of time, and require to be reevaluated as mortality changes over longer time periods.[10] This paper extends existing model life tables by (1) using a newly-collected database that has greater age detail as well as broader geographical and temporal coverage than the ones on which existing model life tables are derived; (2) explicitly expanding the number of age groups in the model, allowing more flexibility than existing models both in terms of entry parameters and model outcomes.

A new database for under-five mortality by detailed age

The model proposed in this paper is based on a newly-collected database for under-five mortality by detailed age. This database includes annual distributions of deaths by sex and by detailed age (days, weeks, months, trimesters, and years) for countries with high-quality vital registration (VR) systems, as defined by the Human Mortality Database (HMD) project. The death distributions for historical periods (prior to 1970) were collected manually from archival sources such as national statistical yearbooks. For periods from 1970 onwards, death distributions were obtained electronically from a database prepared by the United Nations Statistical Division. A total of 1715 country-years were collected for this study, including 1057 country-years collected from archival sources. These death distributions were then combined with exposure terms provided by the HMD for each country-year to obtain age-specific mortality rates (nMx) and corresponding cumulative probabilities of dying q(x) by sex for the following harmonized age groups: 7, 14, 21, 28 days; 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 18, 21 months; 3, 4, 5 years.

Log-quadratic model for age-specific mortality by detailed age between 0 and 5

The model proposed in this paper is adapted from Wilmoth et al.'s log-quadratic model.[1] It is based on the observation of log-quadratic relationships between q(x) and U5MR (=q(5)) for each detailed age x within the under-five age range:

$$\mathbf{ln[q(x)] = a_x + b_x ln[q(5)] + c_x ln[q(5)]^2 + v_x k} \qquad \mathbf{(1)}$$

This model is a two-dimensional model in which q(5) determines the overall level of under-five mortality, while k affects the shape of the age-pattern of mortality between 0 and 5. Depending on the level of k, mortality at a given level of U5MR will be either "early", with higher corresponding values of neonatal and infant mortality, or "late", with lower corresponding values of neonatal and infant mortality.

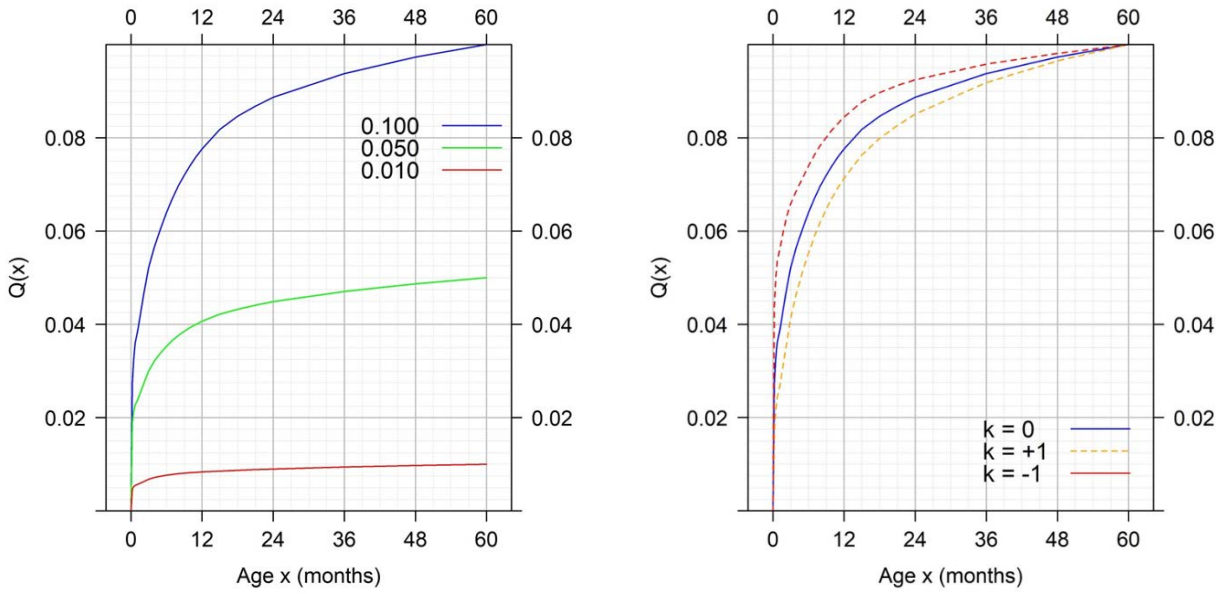
Model parameters a_x, b_x, c_x are estimated by regressing q(x) against q(5) using the log-quadratic portion of Equation (1), and v_x is estimated using Singular Value Decomposition (SVD) applied to the matrix of regression residuals.

Figure 1 shows preliminary results of this modeling approach, summarizing general regularities in the collected VR mortality database. Figure 1(a) shows the effect of varying the level of mortality on predicted q(x) values. It illustrates the fact that as the level of U5MR decreases, a greater portion of mortality takes place early in life. Figure 1(b) shows the effect of varying the shape parameter k on the age pattern of under-five mortality. It shows that given a level of U5MR, negative values of k produce an earlier age pattern while positive values of k produce a later age pattern.

Figure 1: Effect of varying U5MR vs. k on predicted q(x) values in the log-quadratic model

(a) Effect of varying U5MR on predicted q(x)'s (with k=0)

(b) Effect of varying k on predicted q(x)'s (with U5MR=100 p.1000)



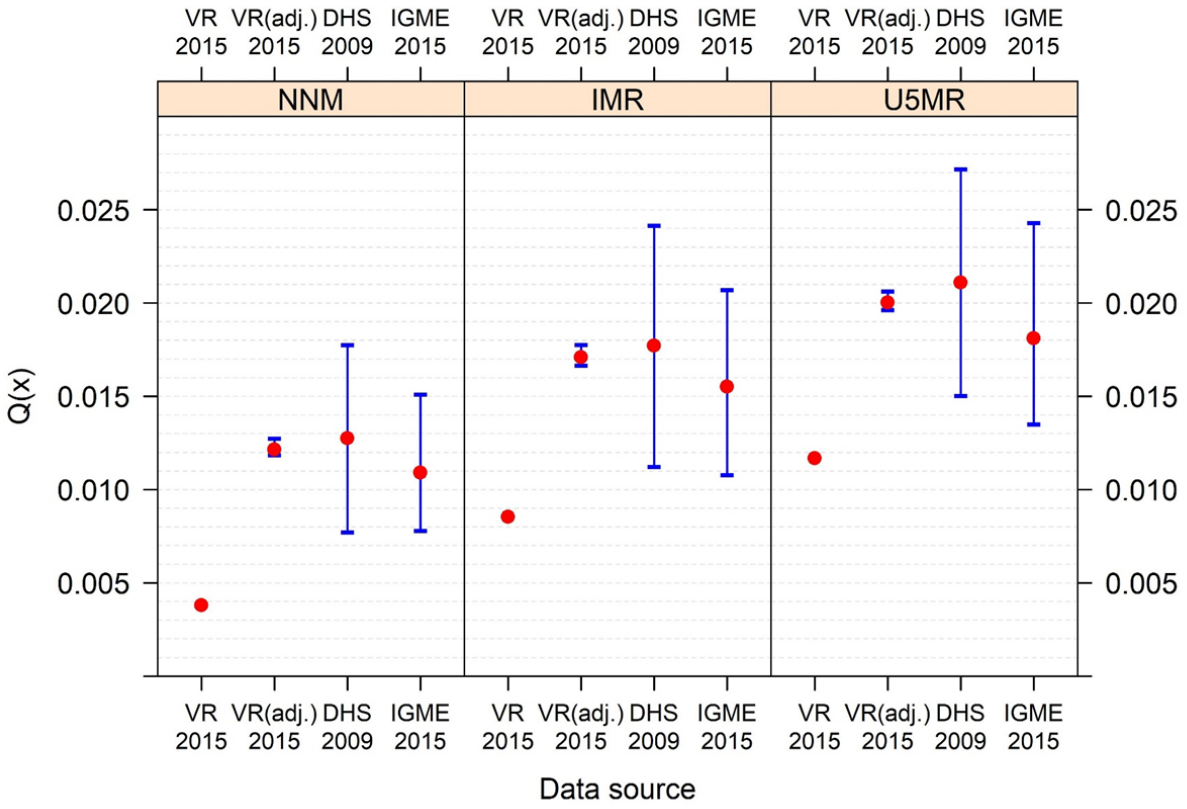
Varying both U5MR and k in this model allows us to fit the VR data remarkably well. Using U5MR and q(5m) as entry parameters, the root mean square errors (RMSE) in predicted q(x)'s is as low as 1.72%. Using all observed data points q(x)'s to solve for the shape parameter k, RMSE decreases to 1.48%. An outsampling test (estimating the model coefficients on 80% of the data and using these coefficients to predict q(x)'s on the remaining 20% of the data) confirms the remarkable fit of our model.

Using the model for adjusting under-five mortality in countries with deficient VR systems

This new model has many practical applications. It can be used to: (1) smooth noisy age schedules; (2) correct mortality estimates in the presence of age heaping; or (3) adjust mortality data for under-reporting in specific age ranges. In this paper, we propose an application of the model to evaluate and adjust vital registration data in countries where the quality of the mortality data for the neonatal period is highly questionable. The logic of the approach is to fit the model on the observed data excluding the neonatal portion the mortality curve, and using the estimated level and shape parameters from the log-quadratic model to predict an adjusted q(x) curve (including adjusted levels of neonatal, infant and under-five mortality).

Preliminary results of this approach applied to recent (2015) VR data from Jordan (Figure 2) shows that the model adjusts the VR mortality estimates upwards by a factor of almost 2, producing mortality levels that are consistent with DHS estimates as well as estimates produced by the Inter-Agency Group for Mortality Estimation (IGME). Taking the IGME estimate as a reference, our results suggest that the log-quadratic model is a promising tool for adjusting VR data in countries where the quality of reporting for the neonatal period is highly questionable.

Figure 2: Comparison of neonatal, infant and under-five mortality estimates for Jordan in 2015 using various sources



Next steps

In the final paper, we will provide a comprehensive approach for estimating uncertainty around predicted q(x) values using the log-quadratic model. We will also provide additional practical applications of our model for evaluating and adjusting under-five mortality data from various sources.

References

1. Wilmoth J, Zureick S, Canudas-Romo V, Inoue M, Sawyer C. A flexible two-dimensional mortality model for use in indirect estimation. *Popul Stud (Camb)* 2012; **66**(1): 1-28.
2. Coale AJ, Demeny PG. Regional model life tables and stable populations. Princeton, N.J.,: Princeton University Press; 1966.
3. Coale AJ, Demeny PG, Vaughan B. Regional model life tables and stable populations. 2nd ed. New York: Academic Press; 1983.
4. United Nations. Model life tables for developing countries. New York: United Nations; 1982.
5. United Nations. MortPak-Lite -- the United Nations software package for mortality measurement : interactive software for the IBM-PC and compatibles. New York: United Nations; 1988.
6. Galley C, Woods R. Reflections on the distribution of deaths in the first year of life. *Population* 1998; **53**(5): 921-46.
7. Lantoine C, Pressat R. New Aspects of Infant-Mortality. *Population* 1984; **39**(2): 253-64.
8. Lynch KA, Greenhouse JB, Brändström A. Biometric Modeling n the Study of Infant Mortality: Evidence from Nineteenth-Century Sweden. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 1998; **31**(2): 53-64.
9. Manfredini M. The bourgeois-pichat's biometric method and the influence of climate: New evidences from late 19th-century Italy. *Biodemography and Social Biology* 2004; **51**(1-2): 24-36.
10. Guillot M, Gerland P, Pelletier F, Saabneh A. Child Mortality Estimation: A Global Overview of Infant and Child Mortality Age Patterns in Light of New Empirical Data. *PLoS medicine* 2012; **9**(8).