Indirect estimation of the neonatal mortality: a model life table approach using prior distributions[1]

Julio E Romero Prieto, London School of Hygiene & Tropical Medicine
Andrea Verhulst, University of Pennsylvania
Michel Guillot, University of Pennsylvania and INED

1. Introduction and background

The under-5 mortality rate has halved in the past three decades, but neonatal mortality has not declined at the same proportion [1, 2]. Since the percentage of neonatal deaths has increased [3], future improvements in child survival will require a considerable reduction in the neonatal mortality rate [4]. Low- and middle-income countries account for the majority of neonatal deaths, but tracking progress is complicated by the lack of reliable statistics [5]. In these contexts, the neonatal mortality is estimated from the full birth histories collected in nationally representative surveys [6, 7]. Although demographic surveys are systematically collected ensuring higher standards of quality, survey estimates are affected by reporting errors. The evidence suggests that heaping at the age of seven days might transfer some early neonatal deaths (first week) to the second week of life [8]. Yet more concern would exist about the neonatal deaths underreported in surveys, inasmuch as the proportion of deaths during the first two days of life has poor fitting to a model of mortality [9]. In fact, stillbirths and early neonatal deaths are misclassified in surveys because the fear of social stigma, or denied when the experience is too painful to be reported [10, 11].

The undercount of neonatal deaths is not exclusively of survey estimates. The same limitation has been detected in vital registration and demographic surveillance systems [12, 13]. On the one hand, some registration systems failed to keep record of the early deaths that they were not previously registered as births. In fact, the underestimation of the neonatal mortality is directly related to the definition of a live birth [14, 15, 16]. While most countries consider any sign of live, preterm and underweight infants dying the first week or life have been classified as stillbirths in some countries [13]. On the other hand, most demographic surveillance systems have a limited scope to record stillbirths and neonatal deaths, unless pregnancies would be registered as relevant events or after linking the records of antenatal clinics. However, these improvements have not been fully implemented yet. As a result, many neonatal deaths remain undercounted and monitoring future interventions would require better estimations.

Models of mortality can help to evaluate and correct both record and survey estimates of neonatal mortality. The classical approach was the Bourgeois-Pichat's "biometric model" [17], used to decompose the neonatal mortality according to endogenous and exogenous causes of death. Since the exogenous component was estimated fitting the monthly distribution of post-neonatal deaths, this component was considered a lower bound to evaluate the completeness of infant deaths in the analysis of historical populations [18, 19]. Modern approaches aim to estimate the proportion of neonatal deaths as a function of the infant or the under-5 mortality rate [1, 20], and the indirect estimation of the neonatal mortality. The indirect estimation takes advantage of the strong correlation between the probability of dying at early ages (the first weeks or months of life) and the probability of dying at post-neonatal ages [21, 13], or the under-5 mortality rate [22, 23]. In the practice, this correlation is estimated from countries with reliable records and the empirical regularities are extrapolated to countries with defective data, nevertheless admitting some degree of flexibility. Although there is consensus in the necessity to adjust some data before estimating the trends and forecasting the future trajectories of the neonatal mortality [22, 23, 4, 24], how to modelling the neonatal mortality is a matter of discussion.

This extended abstract describes a method for the indirect estimation of the neonatal mortality, following the demographic approach of the Model Life Tables [25, 26, 27, 28]. The model was estimated from the most reliable records of the Global Age Patterns of Under-5 Mortality database, a newly collected database of national distributions of deaths by detailed age and calendar year. Given the geographical dispersion (25 countries) and the years of the database (1877-2016), it is a general model able to evaluate and correct both vital record and survey estimates of mortality. The model predicts the mortality rates at early ages, by weeks and months of life, for a given combination of parameters related to the level and pattern of mortality [29]. Hence, the indirect estimation of the neonatal mortality

[1]Acknowledgements: Research reported in this manuscript was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (NIH) under award number R01HD090082. Additional funding was provided by the UNICEF.

is to find the optimal combination of parameters, matching or fitting the model to the most reliable probabilities of dying observed at post-neonatal ages. Preliminary results shows that the neonatal mortality can be estimated from the post-neonatal mortality. This indirect estimation improves significantly if a prior distribution of the level of mortality is available. This prior can be estimated from high-quality Vital Registration (VR) data.

2. Data

The main source of data is the Global Age Patterns of Under-5 Mortality database (GAPU5M). Specifically, a selection of 1,235 life tables for both sexes combined allocated in two random samples. The 60 per cent of the dataset (741 life tables) was used to estimate the coefficients of the model and for drawing prior distributions of the parameters. The remaining 40 per cent (494 life tables) was used to evaluate the accuracy and precision of the indirect estimates of the neonatal mortality. This random selection was repeated several times, in order to avoid dependency on the selection and to calculate confidence intervals. Secondary datasets included 262 life tables estimated from the Demographic and Health Survey (DHS) and some Incomplete VR (not part of this extended abstract). These secondary datasets were used for testing purposes.

3. Methods

The model

We propose a Model Life Table (MLT) that produces a mortality schedule by detailed age between 0 and 5 years for a given set of parameters h and k, related to the level and the shape of the mortality at early ages [29], following the approach of Wilmoth *et al*. for adult ages [28]. The model assumes that mortality rates are log-quadratic functions of the level of mortality, which is the standard approach in the analysis of the neonatal mortality [22, 23, 4]. The MLT is a system of 22 equations, given the following exact-age cut-off points for age-specific mortality rates: 0, 7, 14, 21, 28 days; 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 18, 21 months; and 2, 3, 4, 5 years. Since the neonatal mortality rate is calculated as the probability of dying within the first 28 days of life, the first four equations of the MLT are necessary to estimate its value. This property adds flexibility and precision to the model, compared to the most conventional approaches depending on just one equation. From the perspective of a MLT, the neonatal mortality rate can be estimated indirectly given the probabilities of dying at post-neonatal ages.

$$\ln\left[{}_nM_x \right] = \mathrm{a}_x + \mathrm{b}_x \cdot h + \mathrm{c}_x \cdot h^2 + \mathrm{v}_x \cdot k \qquad (1)$$

As shown in equation (1), the model includes a set of age-specific coefficients $\{\mathrm{a}_x, \mathrm{b}_x, \mathrm{c}_x, \mathrm{v}_x\}$, that have been estimated in two steps. The first three coefficients were estimated by Ordinary Least Squares (OLS), assuming the probability of dying below the age of five years as a proxy of the level of mortality, henceforth $h = \ln[q(60m)]$. The last coefficient is an element of the first-orthonormal eigenvector estimated from the Singular Value Decomposition of the covariance matrix of the residuals, following a standard approach in demographic estimation [30, 31, 28, 32]. The intuition is that given the orthogonality condition of OLS, residuals are unrelated to the level of mortality. Hence, the age-covariance of the residuals would inform systematic deviations from the general pattern of mortality. From this perspective, the last coefficient would help to adjust the mortality rate at some age x, given a change in the mortality rate at some other ages. Given the coefficients of the model, Figure A1 shows the neonatal mortality as a function of the post-neonatal mortality for different values of k (see Appendix). The figure shows that keeping constant the probability of dying after 28 days, DHS estimates of neonatal mortality are lower than any prediction based on the vital records of GAPU5M.

The indirect estimation of the neonatal mortality (one parameter and one entry value: $k = 0$)

If neonatal deaths are underreported, the infant and the under-5 mortality rates cannot provide a satisfactory correction of the neonatal mortality, as they are also underestimated. Therefore, conventional approaches using just one equation and depending on these probabilities of dying have a null application as methods for demographic estimation. Conversely, MLTs are built on the necessity of indirect estimations. The indirect estimation of the neonatal mortality is to find the values of h and k, matching or fitting some probabilities of dying at post-neonatal ages that are more reliable. At least one probability of dying will be necessary to estimate the value of h, and two probabilities to estimate h and k. However, some demographic or statistical restrictions can replace one of these probabilities of dying for the model to be identified.

Assuming the simplest case of observing just one probability of dying at post-neonatal ages $q[28d, 60m)$ and considering that the MLS is a system of nonlinear equations, the optimal value of h is calculated by numerical methods making the error of prediction equal to zero, $r(\quad) = 0$.

$$r(h) = \ln[q[28d, 60m)]^{Observed} - \ln[q[28d, 60m)|h]^{Model} \quad (2)$$

Given an initial value of h, assuming that $k = 0$, and following the Newton's method, the optimal value of h^* is iteratively updated by equation (3), as the relative error of prediction $r(\quad)$ approaches to zero.

$$h^* = h - \left[\frac{\partial r(\quad)}{\partial h}\right]^{-1} \cdot r(h) \quad (3)$$

The indirect estimation of the neonatal mortality (two parameters, one entry value and prior distributions)

If two probabilities were observed, then the same numerical method can be applied to calculate the optimal values of h and k, matching the model to the observed data and taking advantage of a Jacobian matrix of the first-order partial derivatives of the errors. We argue that this approach (matching two probabilities of dying at post-neonatal ages) does not necessarily add more information to improve the indirect estimation of the neonatal mortality. However, having a prior knowledge of the level of mortality h, allows calculating the optimal value of k, matching the value of post-neonatal mortality. Assuming that the prior is closely related to the neonatal mortality, the two parameters would improve the precision of the model. In this regard, we aim to have a prior distribution of h, given $q(28d)$ and $q[28d, 60m)$, when the post-neonatal mortality is observed.

Prior distributions of the level of mortality were drawn from the estimation sample. Given that we observe $q(28d)$ and $q[28d, 60m)$, we calculated the optimal values of h and k, matching these two probabilities of dying. Figure A2 shows the value of h conditional to matching these probabilities in the context of high-quality VR data. Applying local regression, we predicted the value of h for a given value of $q[28d, 60m)$ for each life table of the evaluation sample. Nearest errors around the value of $q[28d, 60m)$ were used to account for the dispersion of the predicted value of h. The distribution h was smoothed and evaluated at 5,000 random points. Then, the optimal value of k was calculated keeping constant the value of h, and matching the probability of dying after the first month of life $q[28d, 60m)$. Finally, we assessed the accuracy and precision of the indirect estimation comparing the actual value of the neonatal mortality with a model's prediction using the evaluation sample.

4. Preliminary results

Having a prior distribution of h improves the indirect estimation of the neonatal mortality. Compared to the model of one parameter $k = 0$, two parameters increase the goodness-of-fit and reduce the variance of errors. We extended the approach in order to consider more than one way to estimate a prior distribution of h. This is possible when more than one probability of dying is observed. Hence, the distribution of the level of mortality h was investigated as the optimal value resulting of matching the new probability of dying and the neonatal mortality. From this perspective, each prior provides relevant information for the indirect estimation of the neonatal mortality. The resulting distribution reduces the range of possible values of h, as shown in Figure A3 for the life table of Finland 1950. As shown in Table A1, having more than one probability of dying to estimate a prior distribution of h increases the goodness-of-fit and reduces the variance of errors even more.

5. Further directions

This extended abstract describes an indirect method to estimate the neonatal mortality from observed probabilities at post-neonatal ages. Indirect estimation improves when more than one probability of dying is observed. Further research will be concentrated in improving the precision of the estimations. Hence, we aim to identify the groups of ages that are more relevant for the indirect estimation of the neonatal mortality. Defining the relevant ages, the model will be ready to evaluate survey estimates from the DHS and incomplete VR. Further research will dedicate more analysis to the problem of bias, when the model fails to reproduce the value used to predict the mortality at all ages $h \neq \ln[q(60m \mid h, k)]$. Figure A5 shows that optimal solutions when the neonatal mortality is observed are located at regions of minimal bias. Therefore, several values drew in the distribution of h can be discharged.

6. Appendix

Figure A1:The neonatal mortality as a function of the probability of dying after 28 days of life

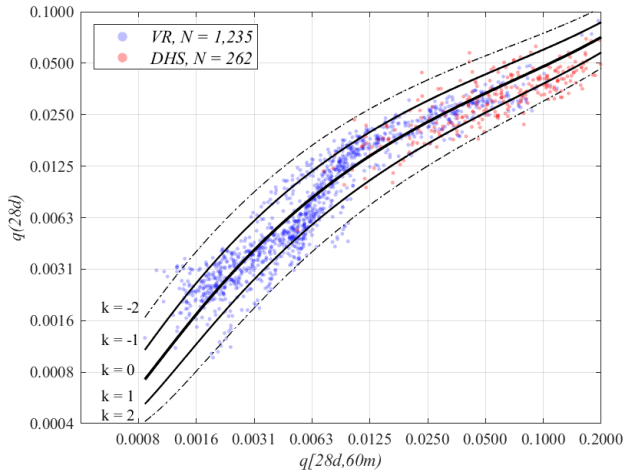


Figure A2: The post-neonatal mortality and the optimal value of h, matching the neonatal mortality

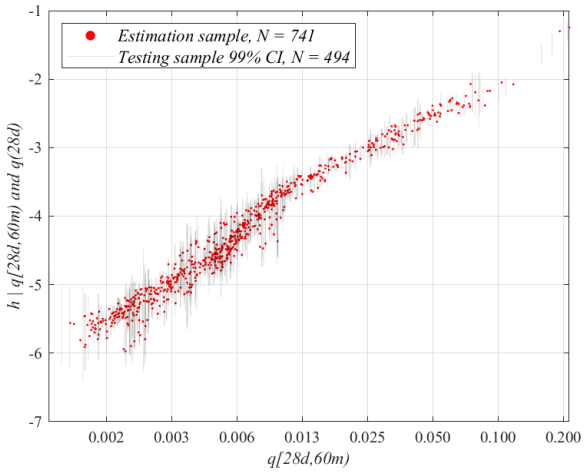


Figure A3: Empirical distribution of h, given the probabilities of dying at different ages, Finland 1950

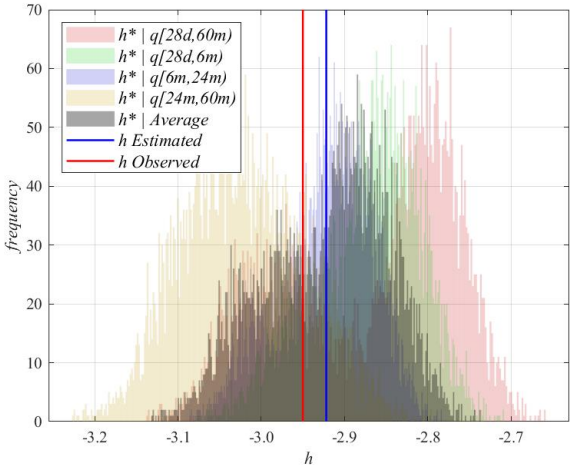


Table A1: Precision and accuracy of the indirect estimation of the neonatal mortality

RMSE	*p50*	*p2.5*	*p97.5*
h and q[28d,60m) --- k = 0	**0.30**	0.28	0.32
h, k, and q[28d,60m)	**0.28**	0.26	0.32
h, k, and four probabilities	**0.25**	0.24	0.27

Goodness-of-fit	*p50*	*p2.5*	*p97.5*
h and q[28d,60m) --- k = 0	**0.89**	0.87	0.90
h, k, and q[28d,60m)	**0.90**	0.87	0.91
h, k, and four probabilities	**0.92**	0.91	0.93

1-var(error)/var(observed)
Values calculated from 200 random samples. 60 per cent of life tables for estimation and 40 per cent for evaluation.

Figure A4: Empirical distribution of $q(28d)$, given the probabilities of dying at different ages, Finland 1950

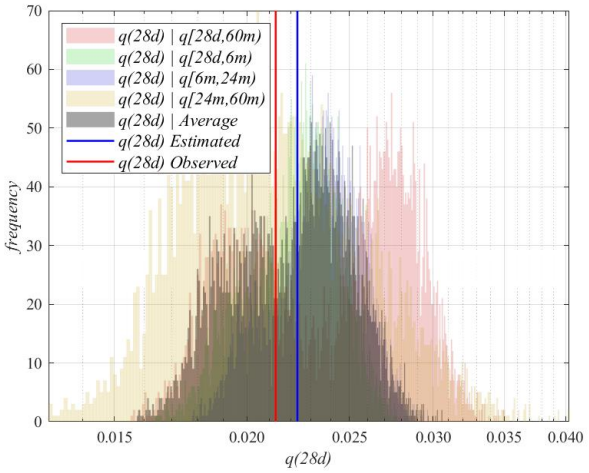
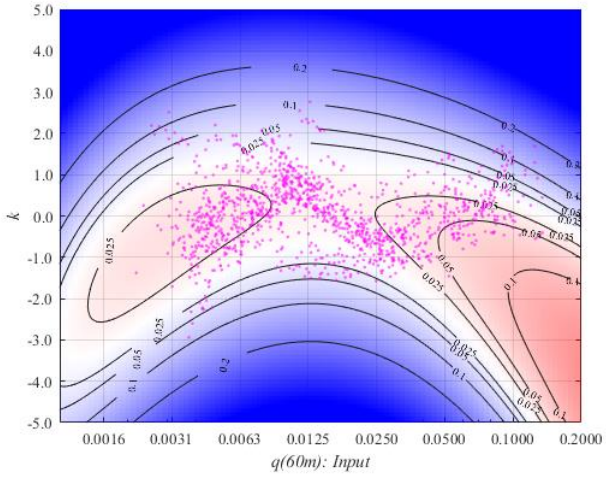


Figure A5: The bias of the model and the optimal values of h and k, matching the neonatal and post-neonatal mortality



7. References

[1] K. Hill and Y. Choi, "Neonatal mortality in the developing world," *Demographic Research,* vol. 14, no. 18, pp. 429-452, 2006.

[2] L. Liu, S. Oza, D. Hogan, Y. Chu, J. Perin, J. Zhu, J. E. Lawn, S. Cousens, C. Mathers and R. E. Black, "Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals," *Lancet,* vol. 388, p. 3027–35, 2016A.

[3] H. Wang and et al., "Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," *Lancet,* vol. 384, p. 957–79, 2014.

[4] UN-IGME, Levels & Trends in Child Mortality, United Nations Children's Fund, 2019.

[5] J. E. Lawn, S. Cousens, J. Zupan and the Lancet Neonatal Survival Steering Team, "4 million neonatal deaths: When? Where? Why?," *Lancet,* vol. 365, p. 891–900, 2005.

[6] K. Hill, "Approaches to the measurement of childhood mortality: a comparative review," *Population Index,* vol. 57, no. 3, pp. 368-382, 1991.

[7] K. Hill, "Direct estimation of child mortality from birth histories," in *Tools for demographic estimation,* T. Moultrie, R. Dorrington, A. Hill, K. Hill, I. Tim and B. Zaba, Eds., Paris, International Union for the Scientific Study of Population, 2013, p. 166–177.

[8] T. W. Pullum and S. Becker, "Evidence of Omission and Displacement in DHS Birth Histories," ICF International, Rockville, MD, 2014.

[9] S. Oza, S. N. Cousens and J. E. Lawn, "Estimation of daily risk of neonatal death, including the day of birth, in 186 countries in 2013: a vital-registration and modelling-based study," *Lancet Global Health,* vol. 2, p. e635–44, 2014.

[10] R. A. Haws, I. Mashasi, M. Mrisho, J. A. Schellenberg, G. L. Darmstadt and P. J. Winch, ""These are not good things for other people to know": How rural Tanzanian women's experiences of pregnancy loss and early neonatal death may impact survey data quality," *Social Science & Medicine,* vol. 71, pp. 1764-1772, 2010.

[11] L. Liu, H. D. Kalter, Y. Chu, N. Kazmi, A. K. Koffi, A. Amouzou, O. Joos, M. Munos and R. E. Black, "Understanding misclassification between neonatal deaths and stillbirths: empirical evidence from Malawi," *PLoS ONE,* vol. 11, no. 12, p. e0168743, 2016B.

[12] J. E. Lawn, D. Osrin, A. Adler and S. Cousens, "Four million neonatal deaths: counting and attribution of cause of death," *Paediatric and Perinatal Epidemiology,* vol. 22, no. 5, p. 410–416, 2008.

[13] M. Guillot, S.-j. Lim, L. Torgasheva and M. Denisenko, "Infant mortality in Kyrgyzstan before and after the break-up of the Soviet Union," *Population Studies,* vol. 67, no. 3, pp. 335-352, 2013.

[14] K. Joseph, S. Liu, J. Rouleau, S. Lisonkova, J. A. Hutcheon, R. Sauve, A. C. Allen, M. S. Kramer and C. P. S. S. Fetal and Infant Health Study Group, "Influence of definition based versus pragmatic birth registration on international comparisons of perinatal and infant mortality: population based retrospective study," *BMJ,* vol. 344, p. e746, 2012.

[15] K. Joseph, N. Razaz, G. M. Muraca and S. Lisonkova, "Methodological Challenges in International Comparisons of Perinatal Mortality," *Current Epidemiology Reports,* vol. 4, p. 73–82, 2017.

[16] P. Deb-Rinker, J. León, N. Gilbert, J. Rouleau, A. Andersen, R. Bjarnadóttir, M. Gissler, L. Mortensen, R. Skjærven, S. Vollset, X. Zhang, P. Shah, R. Sauve, M. Kramer, K. Joseph and the Canadian Perinatal Surveillance System, "Differences in perinatal and infant mortality in high-income countries: artifacts of birth registration or evidence of true differences?," *BMC Pediatrics,* vol. 15, no. 112, pp. 1-15, 2015.

[17] J. Bourgeois-Pichat, "Analyse de la mortalité infantile," *Revue de l'Institut International de Statistique/Review of the International Statistical Institute,* vol. 18, no. 1-2, pp. 45-68, 1950.

[18] London, A, "The impact of advances in medicine on the biometric analysis of infant mortality," *Biodemography and Social Biology,* vol. 40, no. 3-4, pp. 260-282, 1993.

[19] C. Galley and R. Woods, "On the distribution of deaths during the first year of life," *Population: An English Selection,* vol. 11, pp. 35-59, 1999.

[20] M. Alexander and L. Alkema, "Global estimation of neonatal mortality using a Bayesian hierarchical splines regression model," *Demographic Research,* vol. 38, no. 15, p. 335–372, 2018.

[21] W. W. Kingkade and E. E. Arriaga, "Mortality in the new independent sates: Patterns and impacts," in *Premature death in the New Independent States,* J. L. Bobadilla, C. Costello and F. Mitchell., Eds., Washington, DC, National Academy Press, 1997, pp. 156-183.

[22] M. Z. Oestergaard, M. Inoue, S. Yoshida, W. R. Mahanani, F. M. Gore, S. Cousens, J. E. Lawn, C. D. Mathers, UN IGME and Child Health Epidemiology Reference Group, "Neonatal Mortality Levels for 193 Countries in 2009 with Trends since 1990: A Systematic Analysis of Progress, Projections, and Priorities," *PLoS Medicine,* vol. 8, no. 8, p. e1001080, 2011.

[23] K. Hill, D. You, M. Inoue, M. Z. Oestergaard and Technical Advisory Group of the UN IGME, "Child mortality estimation: accelerated progress in reducing global child mortality, 1990–2010," *PLoS Medicine,* vol. 9, no. 8, p. e1001303, 2012.

[24] I. Mejia-Guevara, W. Zuo, E. Bendavid, N. Li and S. Tuljapurkar, "Age distribution, trends, and forecasts of under-5 mortality in 31 sub-Saharan African countries: A modeling study," *PLoS Medicine,* vol. 16, no. 3, p. e1002757, 2019.

[25] United Nations, "Age and Sex Patterns of Mortality: Model Life-Tables for Under-Developed Countries," United Nations, Department of Social Affairs - Population Branch, New York, NY, 1955.

[26] A. J. Coale and P. Demeny, Regional Model Life Tables and Stable Populations, Princeton: Princeton University Press, 1966.

[27] United Nations, "Model Life Tables for Developing Countries," United Nations, Department of International Economic and Social Affairs, New York, NY, 1982.

[28] J. Wilmoth, S. Zureick, V. Canudas-Romo, M. Inoue and C. Sawyer, "A flexible two-dimensional mortality model for use in indirect estimation," *Population Studies,* vol. 66, no. 1, pp. 1-28, 2012.

[29] M. Guillot, J. E. Romero Prieto, A. Verhulst and P. Gerland, "Modeling age-specific mortality by detailed age between 0 and 5 years: Results from a logquadratic model applied to high-quality vital registration data," in *Annual Meeting of the Population Association of America*, Austin, TX, 2019.

[30] J. Wilmoth, "Variation in vital rates by age, period, and cohort," *Sociological Methodology,* vol. 20, pp. 295-335, 1990.

[31] R. D. Lee and L. R. Carter, "Modeling and forecasting US mortality," *Journal of the American Statistical Association,* vol. 87, no. 419, pp. 659-671, 1992.

[32] S. J. Clark, "A general age-specific mortality model with an example indexed by child mortality or both child and adult mortality," *Demography,* vol. 56, no. 3, pp. 1131-1159, 2019.