# Modeling Age Patterns of Under-5 Mortality by Detailed Age in Low- and Middle-Income Countries using Demographic and Health Surveys

## 1. Introduction

In this paper, we propose a model for summarizing regularities about how under-5 mortality is distributed by detailed age groups –including weeks, months, and trimesters– in Sub-Saharan Africa and South Asia. This model is based on estimates from Demographic and Health Surveys (DHS). The model fills an important gap in the existing literature on model life tables by focusing on countries that depart from the historical mortality experience of high-income countries.

The Under-5 Mortality Rate –denoted here $q(5\text{y})$– 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. This is particularly the case in low- and middle-income countries (LMCIs) where Vital Registration (VR) systems are not complete.

In a recent work, we developed a model that summarizes the distribution of under-5 mortality within detailed age groups observed in high-quality VR data (Guillot et al.). This VR model is based on the Under-5 Mortality Database. This database gathers yearly deaths distribution that covers the historical experience of 25 high-income countries. The VR model aimed to overcome the drawbacks of the traditional Model Life Tables (MLTs) initially developed by Coale et al. (1966) and still limited nowadays to 0 vs. 1-4 as an age breakdown for the under-5 mortality. However, the VR model has not expanded the geographical scope of the traditional MLTs which were derived from VR data collected in high-income countries as well. Nevertheless, using a wide range of data sources, we showed that the VR model was able to represent correctly the age pattern of under-5 mortality of many LMICs (Verhulst et al., 2021a). This allowed the applications of the model for evaluating and correcting under-5 mortality information by detailed age in countries with deficient vital registration (Verhulst et al., 2021b).

However, two findings justify the need to expand the geographical scope of this model using DHS surveys. First, the VR model does not represent correctly the age patterns of under-5 mortality in two world regions: Sub-Saharan Africa and South Asia. We found that, in comparison to the rest of the world, the age patterns of under-5 mortality observed in these two regions were characterized by a very distinct double burden with excess mortality at both early and later ages. The need to include these divergent age patterns in a new model is particularly critical since these two regions still rely widely on the traditional MLTs to monitor under-5 mortality. Note that despite the UN

(1982) update of the Coale and Demeny MLTs with data from LMICs, the scope of these models has remained similar (Guillot et al., 2012), and thus does not entail the divergent age patterns of under-5 mortality that we found in Sub-Saharan Africa and South Asia.

Second, the use of DHS surveys is justified by broad comparisons that we carried out with Health and Demographic Surveillance Sites (HDSS) (Verhulst et al., 2021a). Contrary to expectations, we found that the HDSS sites of Sub-Saharan Africa were largely affected by strong omissions of early deaths. We came to the conclusion that the large variability in the quality of data makes HDSS sites unsuitable for a global modelling of the age pattern of under-5 mortality. The difficulty to find objective indicators to establish the quality of the data collection in HDSS sites was already stressed by Eilerts et al. (2021). Although certain authors pointed out the possible overestimation of neonatal mortality in DHS surveys (Helleringer et al., 2020; Liu et al., 2016), we gathered evidence that the excess neonatal mortality that we found DHS surveys was also commonly observed in other sources of information such as Cohort Studies, Sample Register Systems, and certain HDSS sites. On the other hand, the omission of early deaths in DHS surveys is also a concern that we cannot discard (Pullum et al., 2014). In addition, age heaping is another usual concern associated with the retrospective nature of DHS data, in particular for its underestimating impact on the infant mortality rate. However, we showed that this impact was much less severe than previously thought (Romero Prieto et al.).

In the current abstract, we used large pool of 135 DHS surveys indiscriminately in relation to data quality. However, in the final paper we will select only DHS surveys for which evidence of early omission of deaths is absent. For this purpose, we will rely principally on the comparison with other sources of data from the same country gathered in the publicly available database of the UN Inter-agency Group for Child Mortality Estimation.[1] Nonetheless, despite the inevitable risk of data issue associated with DHS surveys, we think that this new model will be a strong contribution to the literature on MLTs given the geographical limitation of the exiting models. Following our previous work, the model will be shared publicly as a R package.[2]

## 2. Method

The model proposed in this paper has the same characteristic that the model we developed using the VR data (Guillot et al.). Both models are built on a previous work by Wilmoth et al. (2012), that is a log-quadratic approach predicting a full mortality schedule on the basis of only 1 or 2 parameters. Both models are based on the observation of log-quadratic relationships between probabilities of dying $q(\mathrm{x})$ and $q(5\mathrm{y})$ for each detailed age $x$ within the under-5 age range:

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

---

[1] https://childmortality.org.

[2] https://github.com/verhulsta.

Where $x$ takes the following values: 7, 14, 21, 28 days; 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 18, 21 months; 3, 4, 5 years.

The two parameters are (1) $q(5\mathrm{y})$ which determines the overall level of under-5 mortality and (2) $k$ affecting the shape of the age-pattern of mortality between 0 and 5. Depending on the value of $k$, mortality at a given level of $q(5\mathrm{y})$ will be either "earlier" or "later" than the average outcome (when $k = 0$). Earlier or later means with higher values of mortality at early or later ages between 0 and 5 years.

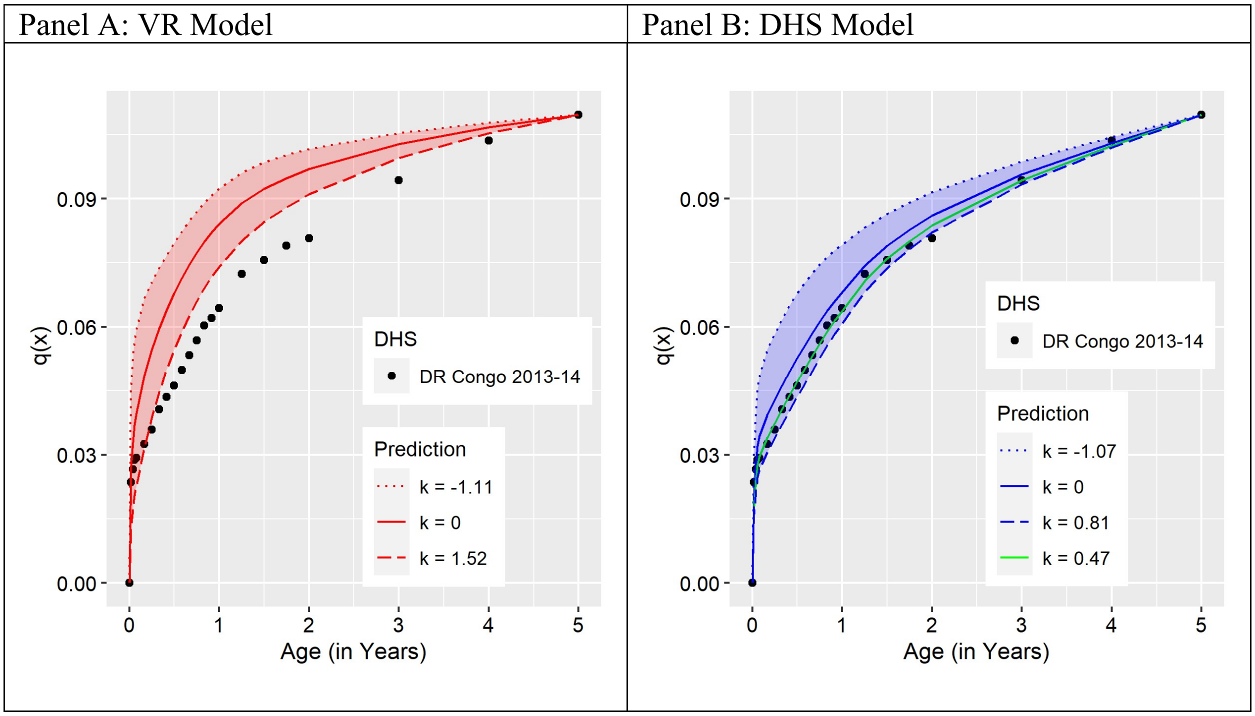
Model parameters $a_x$, $b_x$, $c_x$ were estimated by regressing $q(\mathrm{x})$ against $q(5\mathrm{y})$ using the log-quadratic portion of Equation (1), and $v_x$ is estimated using Singular Value Decomposition (SVD) applied to the matrix of regression residuals. In the case of the VR model, the parameters were derived from 1219 country-years of the Under-5 Mortality Database. For the present model, we used 135 DHS surveys for which the age pattern of under-5 mortality departed from the VR model (Verhulst et al., 2021a). Among these divergent surveys, 118 were collected in Sub-Saharan Africa and 12 in South Asia. Respectively, they represent 88% and 55% of the surveys collected in each region.

## 3. Preliminary Results

Figure 1 shows the effect of varying the shape parameter $k$ on the $q(\mathrm{x})$ between 0 and 5 years compared to the $q(\mathrm{x})$ observed in the DHS RD Congo 2013-14. Panels A and B compared the results of the VR model vs the DHS model showing the minimum and maximum value that $k$ can take respectively. In the first case, the DHS estimates are out of the scope of the model, while the new model captures perfectly the observed age pattern.

The best fitting value of $k$ can be found using any single or associated age groups between 0 and 5 with remarkable outcome (see Table 1). Using $q(5\mathrm{y})$ and $q(6\mathrm{m})$ as entry parameters, the root mean square errors (RMSE) in predicted q(x)'s is as low as 3.9%. Using all observed data points $q(\mathrm{x})$'s to solve for the shape parameter $k$, RMSE decreases to 3.5%. We expect that the outsampling test (estimating the model coefficients on 80% of the data and using these coefficients to predict $q(\mathrm{x})$'s on the remaining 20% of the data) will be similarly good as it was the case with the VR model (Guillot et al.). We will include this outsampling test in the final paper as well as further applications showing how these regularities in the age pattern can be used as a powerful tool for evaluating and correcting data in Sub-Saharan Africa and South Asia. As discussed elsewhere (Guillot et al.), we think that the most powerful application relates to the adjustment of neonatal mortality, for example in the HDSS sites for which we identified strong omissions of early deaths. In these cases, the modeling approach we propose is able to use an indicator that is assumed unbiased such as the probability of dying between 28 days and 5 years to correct the neonatal mortality.

Figure 1: Observed and predicted $q(\mathrm{x})$'s for the DHS RD Congo 2013-2014 using the VR and DHS log-quadratic models


Panel A: VR Model

Panel B: DHS Model

Note: The shaded areas represent the scope of each model. The green line represents the best fit reached by the DHS model when minimizing the Root Mean Square Error of the predicted of $q(\mathrm{x})$'s.

Table 1: Root Mean Square Error (RMSE) of predicted $q(\mathrm{x})$'s using the DHS log-quadratic model applied to the underlying 135 DHS surveys with various combinations of outcomes and entry points for estimating the parameter $k$, both sexes combined

| | | RMSE for the following outcomes: | | | |
|---|---|---|---|---|---|
| *Entry point(s)* | | *all q(x)* | *q(28d)* | *q(12m)* | *q(5y)* |
| *q*(*60m*) only, *k* = 0 | | 0.0687 | 0.1707 | 0.0925 | 0.0000 |
| *q*(*60m*) and | *q*(*7d*) | 0.0850 | 0.2109 | 0.1277 | 0.0000 |
| | *q*(*28d*) | 0.0467 | 0.0000 | 0.0782 | 0.0000 |
| | *q*(*3m*) | 0.0412 | 0.0465 | 0.0672 | 0.0000 |
| | *q*(*6m*) | **0.0385** | 0.0896 | 0.0481 | 0.0000 |
| | *q*(*12m*) | 0.0510 | 0.1765 | 0.0000 | 0.0000 |
| | all *q*(*x*) | **0.0359** | 0.0849 | 0.0455 | 0.0000 |

## 4. References

Coale, A. J., & Demeny, P. (1966). *Regional Model Life Tables and Stable Populations*. Princeton: Princeton University Press.

Eilerts, H., Romero Prieto, J., Eaton, J. W., & Reniers, G. (2021). Age patterns of under-5 mortality in sub-Saharan Africa during 1990–2018: A comparison of estimates from demographic surveillance with full birth histories and the historic record. *Demographic Research, 44*(18), 415-442.

Guillot, M., Gerland, P., Pelletier, F., & Saabneh, A. (2012). A Global Overview of Infant and Child Mortality Age Patterns in Light of New Empirical Data. *PLoS Medicine, 9*(8), e1001299. doi:10.1371/journal.pmed.1001299

Guillot, M., Romero Prieto, J., Verhulst, A., & Gerland, P. Modeling Age Patterns of Under-5 Mortality: Results From a Log-Quadratic Model Applied to High-Quality Vital Registration Data. *Demography (forthcoming)*.

Helleringer, S., Liu, L., Chu, Y., Rodrigues, A., & Fisker, A. B. (2020). Biases in survey estimates of neonatal mortality: Results from a validation study in urban areas of Guinea-Bissau. *Demography, 57*(5).

Liu, L., Kalter, H. D., Chu, Y., et al. (2016). Understanding misclassification between neonatal deaths and stillbirths: Empirical evidence from Malawi. *PLOS ONE, 11*(12).

Pullum, T. W., & Becker, S. (2014). *Evidence of omission and displacement in DHS birth histories. DHS Methodological Reports No. 11*. Rockville: ICF International.

Romero Prieto, J., Verhulst, A., & Guillot, M. Estimating Infant Mortality Rate From DHS Full Birth Histories in the Presence of Age Heaping. *Plos ONE (forthcoming)*.

UN. (1982). *Model Life Tables for Developing Countries* New York: United Nations.

Verhulst, A., Romero Prieto, J., Alam, N., et al. (2021a). *Identifying New Age Patterns of Under-5 Mortality Using the Under-5 Mortality Database*. Paper presented at the PAA Annual Meeting, Online.

Verhulst, A., Romero Prieto, J., & Guillot, M. (2021b). *Using Vital Registration Data for Estimating Under-5 Mortality in Low- and Middle-Income Countries* Paper presented at the PAA Annual Meeting, Online.

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