Estimating under-five mortality with incomplete vital registration data

Andrea Verhulst[1], José Henrique Costa Monteiro da Silva[2], Julio Romero Prieto[3], and Michel Guillot[4]

Introduction

The under-five mortality rate, denoted here q(5y), is the probability of dying between birth and age 5 year. As a key indicator of the social progress and health status of a population, the q(5y) is routinely monitored worldwide. Currently the Sustainable Development Goals Target 3.2 aims at the reduction of under-5 mortality to at least 25 per 1000 livebirths in all countries. To carry out a robust real-time monitoring of this target, a well-functioning Vital Registration (VR) system is indispensable. However, in most Low- and Middle-Income Countries (LMICs), the coverage of deaths is insufficient for that purpose (UNSD, 2022).

To monitor q(5y), worldwide, the UN IGME has developed a model that summarizes and smooths all the sources of information existing in each country (UN IGME, 2023). In the absence of a well-functioning VR system, the alternative sources are principally censuses and nationally representative surveys, in particular *Demographic and Health Survey* (DHS). Despite the important contribution of these two sources of information, their limitations are well-known. Census estimates are rough approximations obtained from simple counts of deaths and surviving children (Verhulst, 2016). Although census data allow a fine geographic disaggregation, they are only collected every 10 years or so in the best-case scenario. Survey estimates derived from birth histories are considered more robust, but their collection is also scattered over time (in most cases by 5 years or more) which does not allow a real-time monitoring of q(5y) either. Moreover, the surveys also do a poor job detecting short-term mortality fluctuations because of recall errors, selective biases, and limited sample size (Hill, 2013; Silva, 2012).

Against this backdrop, there has been over the last decade a new international impetus to strengthen VR systems (AbouZahr et al., 2015; Oomman et al., 2013). More than a technical issue, it is an empowering objective for LMICs to produce robust mortality estimates that are based on their own VR data (even after adjustments) rather than on international survey programs such as the DHS one. Finding new solutions for evaluating and correcting incomplete VR information is thus an important goal.

In this paper, we propose a model for estimating under-5 mortality using incomplete VR data. This model relies on age regularities between age 0 and 5 observed in high-quality VR data. Specifically, the model was derived from the *Under-Five Mortality Database* (U5MD) providing distributions of deaths by detailed age, including daily, weekly, and monthly breakdowns. The premise of this method is that underregistration of deaths does not affect all ages equally. At some ages, deaths are more likely to be underreported in the first weeks and months after birth. Our strategy

[1] Center for Demographic Studies (Barcelona)
[2] University of Pennsylvania
[3] London School of Hygiene & Tropical Medicine
[4] University of Pennsylvania and Institut National d'Etudes Démographiques (INED)

consists then in correcting the mortality for very specific age (such as neonatal age) based on the remaining age ranges considered more robust. This makes possible the use of large unexploited series of data as well as the monitoring of the coverage of deaths. In this paper, we test this method with data from 25 countries. The goal is to use this set of countries to provide guidelines with global applicability.

Data and method

We used the Demographic Yearbook (DYB) System of the United Nations Statistics Division (UNSD). The DYB System provides (1) yearly death counts from VR data by detailed age between 0 and 1 and by single year of age between 1 and 5, and (2) population counts from censuses by single year of age (that we interpolated and used as exposure to the risk of dying following a method developed elsewhere). We selected countries where the coverage of deaths was considered insufficient by the UN IGME for direct estimation between 2000-2020 (including countries that have reached completeness during that period), and we selected country-years for which at least one age breakdown was available between age 0 and 1. We excluded countries from South Asia and sub-Saharan Africa given that they are not well represented by the model used in this paper. These criteria allowed us to analyse a pool 25 LMICs.

For these selected countries, we computed several mortality indicators that exclude the earliest periods of life prone to underreporting of deaths. Specifically, we computed cumulative probabilities of dying from age x to 5 years q(x,5y), where x can be (when available) 28d, 3m, 6m, 9m, 12m. We used these probabilities of dying that are potentially not affected by early underreporting of death to predict mortality between age 0 and 5.

We generated the prediction using a model developed elsewhere(Guillot et al., 2022). In this model, the cumulative probability of dying between birth and age x, q(x), is assumed to be a log-quadratic function that depends on two entry values: i) q(5y) which determines the overall level of mortality; and ii) a parameter k which determines the age pattern of mortality:

$$ln[q(x)] = ax + bxln[q(x)] \; + cxln[q(x)]^2 + vxk$$

The set of coefficients [ax,bx,cx,vx] was derived from 1275 country-years selected among Under-Five Mortality Database (U5MD, covering the historical experience of industrialized countries from the early 20th century to the most recent years. Age groups are defined following 22 exact-age cut-off points: 0, 7, 14, 21, 28 days; 2,3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 18, 21 months; 2, 3, 4, 5 years. The model allows to use any age group defined by these cut-off points to predict a full series of 22 q(x) from 0 to 5.

The parameter k takes a continuous range of values between -1 and 1. When k=0, the model predicts a series of q(x) values corresponding to the average age pattern observed in the U5MD. This parameter determines a higher-than-average concentration of deaths either at older (k >1) or at younger age (k <1) ages. The parameter k is adjustable by fitting observed data and can be optimally calculated when q(5y) is related to one or more independent probabilities of dying. Preliminary results are based on the average age pattern (k=0).

Preliminary results and Further work

Figure 1 shows the results for selected cases that represent four types of situations that we identified among the 25 studied countries.

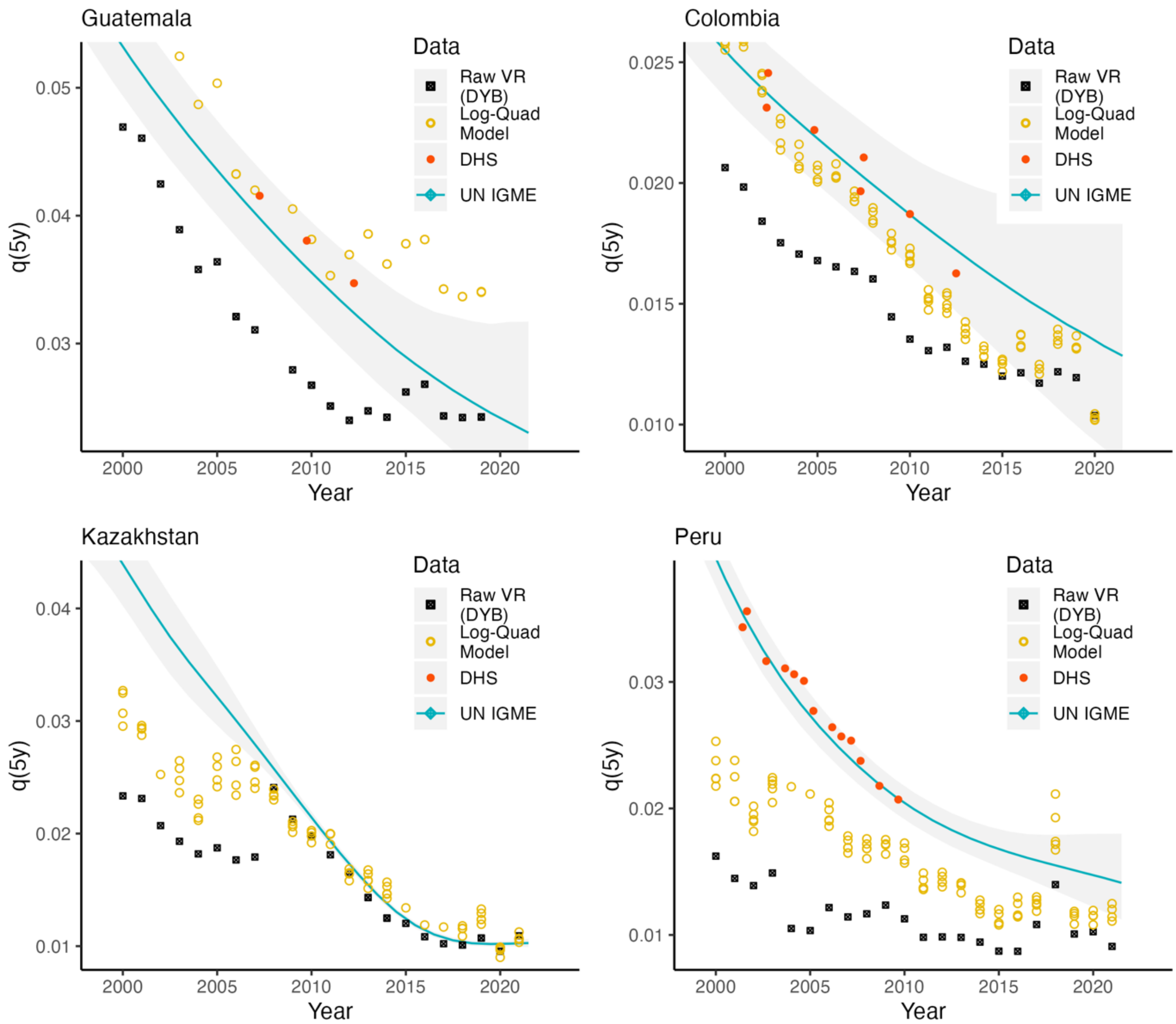
First, the situation of Guatemala corresponds to the case scenario for which the model predictions are well aligned with the levels of mortality derived from existing DHS estimates. In this case, predictions were based only on q(28d,5y) due to the limited availability of age breakdown in the DYB System for this country. Results show that using that age group, our model was able to correct for the underreporting of about 25% deaths between 0 and 28 days (taking thus the DHS estimates as reference). The great advantage of this method in this type of situation is the ability to track at very low-cost annual variations—including increases in mortality—that the DHS is not able to capture. Note that the UN IGME curve relies mostly on the DHS for recent estimates. Therefore, this curve only extrapolates the trend and ignores the annual variation in the raw VR data. In contrast, our model takes advantage of the existing information between 28 days and 5 years and enable the real-time monitoring of under-5 mortality.

Second, the case of Colombia displays similar features to that of Guatemala, that is a strong correction of the level of mortality and a better monitoring than the DHS and the UN IGME estimates. However, in this case, the level of the DHS is not completely reached and the different age groups (not distinguished on the plot) generate a range of predictions. This indicates that the age groups for prediction might be affected themselves by underreporting of deaths, although to a small extend. On the other hand, the gap between the model prediction and the DHS estimates is also potentially due to the uncertainty on the DHS estimates and the uncertainty in the value of k. In the final paper, the results for all countries will include a procedure to determine the best age group to be used as predictor and the analysis of the uncertainty of all estimates.

Third, the case of Kazakhstan represents a group of countries that have reached completeness in the coverage of under-5 deaths over the last 20 years. In this case, no DHS survey is available for comparison. Nonetheless, this example shows how the model can be used to track the critical moment when a VR system is reaching completeness. In the 2008, Kazakhstan has benefited from a reform of its VR system that has radically increased the reported number of deaths. As visible, this change is fully captured by the model and followed by a strong convergence of the different predictions based on different age groups. Therefore, the method that we propose in this paper constitutes tool that can help validate the completeness of a VR system in low- and middle-income countries. In the final paper, we will show this model's ability in other countries, in particular those that have experienced a slower progress in the coverage of deaths.

Fourth, the case of Peru shows the situation in which to the underregistration of deaths is too high and affects all ages groups between age 0 and 5. In this situation, the model is unable to leverage the exiting VR data and predictions remain strongly underestimated compared to DHS. In the final paper, we will determine a threshold of death coverage from which the model is expected to perform well. All the insight that we will derive from the three case scenarios described above will be key for the countries reaching this threshold in the future.

Figure 1. Trends in the under-five mortality rates (q(5y)). Four case scenarios of VR data correction



References

AbouZahr, C., de Savigny, D., Mikkelsen, L., Setel, P. W., Lozano, R., Nichols, E., . . . Lopez, A. D. (2015). Civil registration and vital statistics: progress in the data revolution for counting and accountability. The Lancet, 386(10001), 1373-1385. doi:10.1016/S0140-6736(15)60173-8

Guillot, M., Romero Prieto, J., Verhulst, A., & Gerland, P. (2022). Modeling Age Patterns of Under-5 Mortality: Results From a Log-Quadratic Model Applied to High-Quality Vital Registration Data. Demography, 59(1), 321-347. doi:10.1215/00703370-9709538

Hill, K. (2013). Introduction to child mortality analysis. In T. Moultrie, R. Dorrington, A. Hill, K. Hill, I. Timæus, & B. Zaba (Eds.), Tools for Demographic Estimation. Paris: International Union for the Scientific Study of Population.

Oomman, N., Mehl, G., Berg, M., & Silverman, R. (2013). Modernising vital registration systems: why now? The Lancet, 381(9875), 1336-1337. doi:10.1016/S0140-6736(13)60847-8

Silva, R. (2012). Child Mortality Estimation: Consistency of Under-Five Mortality Rate Estimates Using Full Birth Histories and Summary Birth Histories. PLOS Medicine, 9(8), e1001296. doi:10.1371/journal.pmed.1001296

UN IGME. (2023). Levels and Trends in Child Mortality. Report 2022. New York: UNICEF.

UNSD. (2022). Population and vital statistics report: Series A, Volume LXXIV. New York: United Nations.

Verhulst, A. (2016). Child mortality estimation: An assessment of summary birth history methods using microsimulation. Demographic Research, 34(39), 1075-1128.