Is Neonatal Mortality Misreported in the DHS Surveys? A Model-Based Approach

Introduction

A neonatal death happens when a live birth deceases before the 28th day of age. According to the last report of the UN Inter-agency Group for Child Mortality Estimation (UN IGME, 2019), the number of neonatal deaths in 2018 was 2.5 million worldwide. This was almost half of the number of deaths that happened before 5 years (5.3 million). The proportion of under-five deaths occurring during the neonatal period is increasing while under-five mortality is decreasing, and the causes of death are being increasingly related to perinatal conditions. For that reason, the SGDs are now targeting the reduction of neonatal mortality to 12 per 1000 live births along with the reduction under-five mortality to 25 per 1000 live births by 2030 (UN, 2018).

Monitoring these two indicators is challenging because in most of countries the vital registration system is not complete. In this context, the Demographic and Health Surveys (DHS) have been the main alternative data source for estimating mortality at the early ages of life since the 1980s. Indeed, a vast quantity of national estimates produced by the UN IGME and other global monitoring studies (Wang et al., 2014) relies on retrospective birth histories collected in DHS surveys. In these surveys, women in reproductive ages (15-49) are asked to provide the date of birth and the age at death of all their children ever born. Based on that information, direct estimates of the neonatal mortality rate (NMR) and the under-five mortality rate (U5MR) can be computed with a standard event/exposure procedure (Hill, 2013).

Although this is a remarkable and systematically collected source of data, DHS surveys are not perfect and must be used with extra caution. In addition to sampling errors and survival biases, birth histories are subject to reporting errors. Liu et al. (2019) listed four types of reporting errors affecting child mortality estimates in DHS surveys: (1) date displacement happens when the date of birth is misreported. (2) Age errors occur when mothers misreport the age at dead of their deceased children. This might cause the transference of deaths outside the age boundaries targeted for the estimation of mortality, i.e. the neonatal period. (3) Omission of births and deaths are also a risk. Dead children are generally more likely to be omitted due to recall errors (when they happened at early age and long time ago) or on the contrary because they are painful memories. All age intervals are vulnerable to these reporting errors, but the omission of neonatal deaths has been a particular subject of concern (Pullum & Becker, 2014). (4) In addition, there might be difficulties in distinguishing between stillbirths and neonatal deaths. This problem can go in either directions, but in their study, Li Liu et al. found that the misclassification of stillbirths as neonatal deaths (overestimating mortality) was an important source of bias. Overall, with the relative rise of neonatal mortality, these potential reporting issues would also increasingly bias the estimates of the U5MR derived from DHS surveys.

The knowledge on how these reporting errors affect the measurement of neonatal mortality – and hence also the measurement of under-five mortality – is limited. Some case studies compared the data recorded in DHS-type surveys with a variety of alternative source of data (Alam et al., 2017; Bang et al., 2002; Espeut & Becker, 2015; Haws et al., 2010; Liu et al., 2019; Liu et al., 2016). The nature of the employed alternative source of data embraced verbal autopsy, qualitative interview, specific surveys and demographic surveillance sites depending on the research. The results of these studies do not show specific pattern but a large diversity of situations. For example, in Bangladesh, the comparison with the high-quality data from Matlab Health and Demographic Surveillance System showed that the birth histories collected in the 1994 Matlab Demographic and Health Survey missed about 20% of neonatal deaths (Espeut & Becker, 2015). On the other hand, the results of Li Liu et al. showed that, in populations with known distributions of stillbirths, neonatal deaths and postneonatal deaths, reporting errors led to NMR estimates that are too high (Liu et al., 2019).

A more global assessment of the neonatal mortality estimates in DHS surveys was carried out by Hill and Choi (2006). In order to capture potential neonatal omissions, they adopted a modeling approach based historical data from England and Wales. Nonetheless, they found that the ratios of early to late neonatal deaths and of neonatal to infant deaths were generally higher in the birth histories than in the historical record. This is in accordance with the results of Liu et al. (2019). However, the modeling approach of Hill and Choi (2006) was limited to historical data from one single country. It was based on the assumption that the age pattern of mortality at those early ages in DHS surveys is the same than in England and Wales. In this paper, we propose to push further the work of Hill and Choi by building a new modelling approach that takes advantage of a recently-compiled international database for under-five mortality by detailed age. The model's broader empirical basis will allow us to better represent the diversity of age patterns of under-five mortality in human populations and evaluate the quality of neonatal mortality in the DHS with a higher degree of certainty.

Data

The Global Age Patterns of Under-5 Mortality (GAPU5M) database is a newly-collected database for under-five mortality by detailed age. This database includes annual distributions of deaths by sex and by detailed age (7, 14, 21, 28 days; 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 18, 21 months; 3, 4, 5 years) for 25 Western countries with high-quality vital registration (VR) systems, as defined by the Human Mortality Database (HMD) project[1]. 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. These death distributions were 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. A total of 1175 life tables are underlying the model adopted in this paper.

We used the same age intervals to estimate age-specific mortality rates and cumulative probabilities of dying from the birth histories collected in 261 DHS surveys.

Methods

The model used in this paper (Guillot et al., 2019) is adapted from log-quadratic model of (Wilmoth et al., 2012). It is based on the observation of log-quadratic relationships between nMx and U5MR (=q(5)) for each detailed age x within the under-five age range in the GAPU5M database:

$$\ln[nMx] = a_x + b_x \ln[q(5)] + c_x \ln[q(5)]^2 + v_x k$$

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 nMx 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.

Although the model was developed using q(5) as the main predictor, the model parameters can be estimated on the basis of any portion of the 0-5 age range, as long as at least two mortality indicators within that age range are available as predictors. We used this characteristic to predict neonatal mortality in the DHS based on observed mortality between 28 days and 5 years, which we assumed to be less subject to underreporting than mortality between 0 and 28 days. Then, we compared the value of the NMR predicted by our model with the one observed in the DHS.

[1] https://www.mortality.org

Preliminary results and further work

In our preliminary results, we set the parameter k of the model to 0 and predicted the NMR based solely on the value of probability of dying between 28 days and 5 years q(28d,5y). Fixing k to 0 means that our predicted values of the NMR are based on the average shape of the age pattern of under-five mortality observed in the GAPU5M database.

However, contrary to Hill and Choi (2006), we did not apply this average age pattern indiscriminately to all DHS surveys. In order to generate robust results, we decided to first evaluate if the age pattern used for predictions was consistent with the observed data. For that reason, we selected only the surveys that presented a good fit of the cumulative mortality between 28 days and 5 years. Surveys with poor fit were removed from the analysis. For illustration, Figure 1 shows one example of good fit and one of poor fit. As a selection criterion, we established maximal values for the mean relative error of the prediction between 28 days and 5 years. These boundaries were derived from the distribution of the mean relative error over that age range obtained when applying the model to the GAPU5M database. We retained mean relative error boundaries corresponding to the 95% and the 99% percent best fits in the GAPU5M database, respectively (-9%, 6%) and (-13%, 8%). Table 1 shows that between 49% and 61% of DHS surveys could be fitted with the model and were thus retained for the study. Most of the Sub-Saharan surveys show a different shape in their age pattern and were excluded from the analysis.

Figure 2 shows the relative difference between the observed and predicted value of the NMR by world region for the surveys selected according to the less conservative fit (-13%, 8%). These preliminary results reflect well the diversity of situations found in the previous case studies. A majority of observations fit within +/-20% relative error intervals. But selected DHS surveys in Latin American and Sub-Saharan countries seem to be more affected than the other regions by the underestimation of the NMR. In contrast, in Asia, observed values of the NMR tend to be higher than predicted. These preliminary results suggest that the quality of NMR is highly variable and that the bias does not always go in the same direction. Overall, these results suggest that DHS-based estimates of NMR (and, consequently, U5MR) can be affected by a large amount of reporting errors and thus will often need to be adjusted prior to analysis and interpretation.

Our future work will focus on the improvement of the predictions by fine-tuning our modeling approach. This includes mainly exploring the best entry points for the estimation of the age pattern parameter k. Since omissions and misclassifications are more likely to occur for very early deaths, we will reproduce the full analysis for the early neonatal mortality rate, that is the probability of dying for the 7th day of life. In our final paper, we will explore the results in detail, in particular by level of mortality, and using a variety of indicators including neonatal vs infant and early vs late neonatal mortality ratios.

Table 1: Number of surveys included and excluded in the preliminary analysis

	]-9%,6%[			**]-13%,8%[**		
World Region	incl.	excl.	incl.(%)	incl.	excl.	incl.(%)
Asia	34	16	68%	47	3	94%
Europe	1	2	33%	2	1	67%
Latin America and the Caribbean	43	7	86%	47	3	94%
Middle-East and Northern Africa	22	6	79%	24	4	86%
Southern and Eastern Sub-Saharan Africa	27	40	40%	33	34	49%
Western and Central Sub-Saharan Africa	2	63	3%	8	57	12%
Total	129	134	49%	161	102	61%

Figure 1: Examples of included and excluded DHS surveys in the preliminary analysis

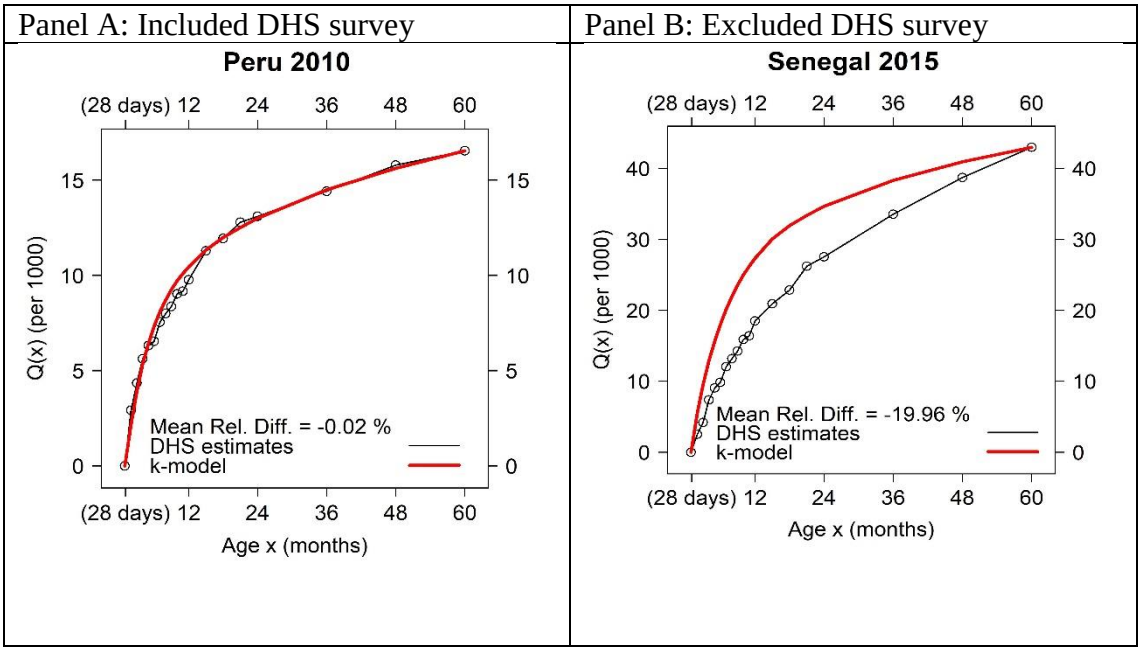
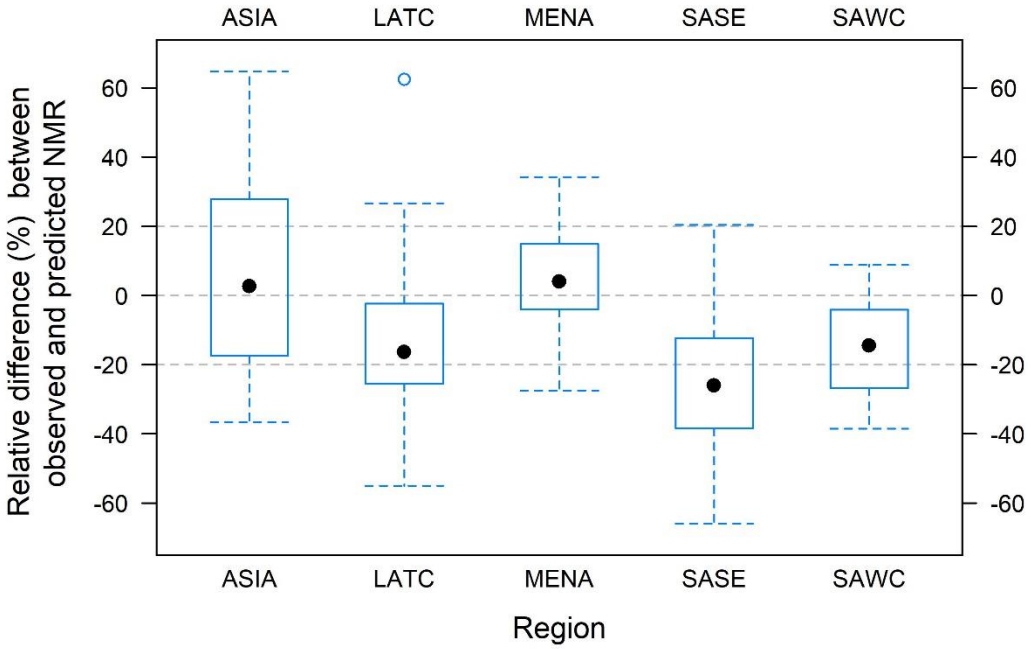


Figure 2: Relative difference between observed and predicted value of the NMR by world region



Note: ASIA = Asia, LATC = Latin America and the Caribbean, MENA = Middle-East and Northern Africa, SASE = Southern and Eastern Sub-Saharan Africa, SAWC = Western and Central Sub-Saharan Africa.

References

Alam, N., Ali, T., Razzaque, A., et al. (2017). Health and Demographic Surveillance System (HDSS) in Matlab, Bangladesh. *International Journal of Epidemiology, 46*(3), 809-816. doi:10.1093/ije/dyx076

Bang, A., Reddy, M., & Deshmukh, M. (2002). Child mortality in Maharashtra. *Economic and Political Weekly, 37*(49), 4947-4965.

Espeut, D., & Becker, S. (2015). The validity of birth and pregnancy histories in rural Bangladesh. *Journal of health, population, and nutrition, 33*, 17-17. doi:10.1186/s41043-015-0027-8

Guillot, M., Romero Prieto, J., Verhulst, A., & Gerland, P. (2019). *Modeling age-specific mortality by detailed age between 0 and 5 years: Results from a logquadratic model applied to high-quality vital registration data*. Paper presented at the 2019 Annual Meeting of the PAA, Austin, TX.

Haws, R. A., Mashasi, I., Mrisho, M., et al. (2010). "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, 71*(10), 1764-1772. doi:https://doi.org/10.1016/j.socscimed.2010.03.051

Hill, K. (2013). Direct estimation of child mortality from birth histories. 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.

Hill, K., & Choi, Y. (2006). Neonatal mortality in the developing world. *Demographic Research, 14*(18), 429-452.

Liu, L., Chu, Y., Rodrigues, A., Fisker, A., & Helleringer, S. (2019). *Understanding Reporting Errors of Population Survey–Based Neonatal Mortality: A Validation Study From Guinea-Bissau*. Paper presented at the 2019 Annual Meeting of the PAA, Austin, TX.

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), e0168743. doi:10.1371/journal.pone.0168743

Pullum, T. W., & Becker, S. (2014). *Evidence of Omission and Displacement in DHS Birth Histories*. Retrieved from Rockville, Maryland, USA:

UN. (2018). *Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development*. Retrieved from New York:

UN IGME. (2019). *Levels and Trends in Child Mortality. Report 2019*. Retrieved from New York:

Wang, H., Liddell, C. A., Coates, M. M., et al. (2014). 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. *The Lancet, 384*(9947), 957-979. doi:10.1016/S0140-6736(14)60497-9

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. doi:10.1080/00324728.2011.611411