JOINT ESTIMATION OF NEONATAL MORTALITY AND VITAL REGISTRATION COMPLETENESS ACROSS MEXICO

PREPRINT

Nathaniel Henry

1 Introduction

Proper surveillance of vital events and disease burden across a country is critical for national health planning, making it an important component of the right to proper medical care enshrined in the Universal Declaration of Human Rights (1,2). Many countries implement a system of civil registration that requires medical professionals to produce legal documents for vital events such as births and deaths; these records are compiled in nationwide vital statistics that provide the foundation for national strategic planning (1). A growing set of countries have also implemented electronic health information systems, such as the District Health Information System 2, that capture records from medical institutions and reports from field workers in a unified web platform (3). However, despite their importance in setting national priorities for health care, local variation in the quality and completeness of these Civil Registration and Vital Statistics systems (CRVS) remains poorly understood.

Many geospatial studies investigating local variation in disease burden derive their estimates from cluster-level observations in household surveys and censuses (4,5). Routine health surveillance includes features that make it an appealing alternative or supplement to traditional geospatial data sources: most notably, the sample sizes associated with CRVS datasets are typically orders of magnitude larger than those collected in any household survey. While years may pass between two household surveys, functioning health surveillance systems provide an unbroken series of observations over time. Many surveillance systems already report health status at the administrative level that is most relevant to country-level financing and planning. More broadly, global health researchers have the opportunity to invest in, and advocate for, data sources that are fundamentally tied to the success of national health systems in the countries where our research is focused.

However, critical issues must be resolved before CRVS data can be incorporated into geospatial analyses of health. The most pressing of these is the question of varying incompleteness in health surveillance in space and time. Previous analyses have shown that the completeness of CRVS data varies across countries, across states or provinces within countries, and over time; completeness is also generally lower for the registration of child deaths (6,7). Figure 1, below, shows the completeness of CRVS death registration among children under 5 in the years 2000 and 2015, derived from estimated produced by the Global Burden of Disease 2017 study (8). While completeness of death registration has improved in many countries during this time period, it remains low in many low- and middle-income countries where the global burden of child mortality is concentrated. While methods have been developed to account for incompleteness in health surveillance data at the national level (6,9), these methods cannot be directly applied at more local levels due to the general geospatial problem of small sample sizes.

Geospatial researchers must also grapple with the issue that health surveillance data is naturally reported across areal units, while observations from household survey data are typically reported at precise point locations. Any model that incorporates both points and polygons must simultaneously accommodate multiple spatial scales of analysis, an issue known as the change of support problem (10). The administrative divisions within a country may also change over time, complicating typical statistical methods for analyzing areal data. Any attempt to estimate local variation in health based on health surveillance data must also account for these multiple, changing spatial scales of analysis.

In the following manuscript, I present an extensible method for formatting subnational CRVS data that accounts for changing administrative units, multiple potential scales of analysis, and improper cause of death diagnoses. I then demonstrate how household survey data and CRVS data can be incorporated into a novel geostatistical model that simultaneously estimates child mortality and CRVS incompleteness at a local level. The data preparation methods have been designed to accommodate both CRVS data, while the geostatistical model could be extended to estimate cause-specific mortality or disease incidence. I discuss how these methods have been applied to subnational CRVS data. Finally, I develop a timeline towards submitting these methods for publication.

1.1 Neonatal mortality estimation in middle-income countries

PAR:

PAR:

PAR:

PAR:

1.2 History of vital statistics in Mexico

PAR:

PAR:

PAR:

1.3 Estimating local variation in neonatal mortality across Mexico

PAR:

PAR:

PAR:

PAR:

2 Methods

Two classes of data have historically informed estimates of child mortality: birth history (BH) data, which retrospectively lists the life histories of all children born to a mother, and CRVS data. At the national level, previous methods have combined estimates from these two data types to estimate both mortality trends and CRVS completeness (6,15), but these methods are problematic for subnational estimation due to the smaller sample sizes available at each observation. Most subnational estimates of child mortality have relied solely on birth history data from household surveys, informed by spatial covariates (4,5); however, this paradigm is not well-suited for countries with high-quality CRVS data and where little or no BH data is available. A previous study in Brazil found that high-quality CRVS data could be incorporated into a geospatial model of child mortality by using strong priors of CRVS completeness from field audits conducted by local public health researchers (16). Although this method successfully incorporated CRVS data at a local level, it cannot be reproduced in countries where these audits have not been conducted. To successfully incorporate CRVS data at the local level across a wider range of countries, a new method must be developed that uses more widely-available data.

PAR:

PAR:

PAR:

PAR:

2.1 Data preparation

PAR:

PAR:

PAR:

2.2 Joint estimation of neonatal mortality and CRVS data completeness

Here, I present a new geospatial model that simultaneously estimates child mortality and CRVS completeness in space and time, using data from both BH and CRVS sources in a single country. The two outcomes of interest are the probability of death before reaching age 5 (5q0 in demographic notation, which I will refer to as Q in the definitions below), defined on a continuous spatial surface; and the proportion of under-5 deaths captured by the CRVS system (which I will refer to as π in the definitions below), defined for each polygon and year on a set of stable administrative boundaries. The two space-time surfaces are defined as follows:

TODO

Here, both Q and π are logit-linear surfaces indexed in space and time, varying according to an assigned set of covariate fixed effects that also vary in space and time, $\beta X_s, t$, although the sets of covariates can differ for Q and π. Variation not captured through covariates is fitted through a space-time random effect ω, where $\omega_Q(s,t)$ has spatial correlation corresponding to a two-dimensional Gaussian process with Matern covariance (17), and $\omega_\pi(s,t)$ corresponds to a discrete polygon surface with spatial defined according to the Leroux conditional autoregressive (LCAR) formulation (18). Both surfaces have time correlation corresponding to an order-1 autoregressive process.

Sample sizes N and number of deaths D from each birth history observation are incorporated as binomial observations of the underlying mortality probability (5q0) surface. The number of deaths from each CRVS observation in a given polygon-year is treated as a Poisson process centered at the population multiplied by the mortality rate (5m0) and the CRVS completeness π. The related demographic processes of mortality rate (5m0 or M) and mortality probability (5q0 or Q) are linked using a known demographic formula:

$$D_{BH}(i) = Binomial(N_{BH}(i), Q(s_i, t_i))$$

$$D_{CRVS}(i) = Poisson(N_{CRVS}(i) M(s_i, t_i) \pi(s_i, t_i))$$

$$M(s,t) = \frac{Q(s,t)}{n - nQ(s,t) + A(t)Q(s,t)}, \;\; n = 5, \;\; A(t) \simeq 0.4$$

To account for the different spatial scales of analysis, following Wilson and Wakefield (10), the continuous spatial estimated are converted from 5q0 to 5m0, then aggregated from a raster grid to polygons using a population-weighted mean.

I developed this model using the Template Model Builder package in R version 3.5.0, and tested the model with BH and CRVS data simulated using the RandomFields and ar.matrix packages (19–22).

PAR:

2.2.1 Prior specification for CRVS completeness

PAR:

PAR:

2.3 Simulation model

I tested the model using simulated BH and CRVS data, where the underlying surfaces for both child mortality and CRVS completeness were known. I sampled data from the known underlying surfaces, then fit a model using these generated data to reconstruct the underlying surfaces. An underlying grid measuring one decimal degree square was chosen in central Mexico, and covariate and population surfaces were extracted at that grid. Four space-time covariates were used to fit the child mortality surface: the year, nighttime light intensity, DPT3 vaccination coverage, and maternal education. Two space-time covariates were used to fit CRVS completeness: population density and travel time to the nearest city.

Because the simulated data also included the true number of deaths in each polygon and year, I also created three VR field audit observations, simulating a situation where field workers used active surveillance to determine the true number of underlying deaths in a region.

PAR:

3 Results

PAR:

3.1 Predictive validity from simulation

The results from one simulation, mimicking a scenario with high child mortality (100 per 1,000 live births) and moderate CRvS completeness (70%) is shown in Figure 1. Panel A demonstrates how this combined model can identify key features of the underlying variation in child mortaliy, including the

decreasing time trend from 2009 to 2010; the general distribution of deaths across the surface; and the locations of the highest-mortality focal regions.

These simulation results suggest that the model can reconstruct underlying surfaces for child mortality and CRVS completeness when given unbiased BH input data. However, Panel B demonstrates that wide uncertainty can surround estimates of subnational CRVS completeness even when field audit data is available.

PAR:

3.2 Case study: neonatal mortality across Mexico

PAR:

PAR:

PAR:

PAR:

PAR:

PAR:

PAR:

4 Discussion

PAR:

PAR:

PAR:

4.1 Performance of the simulation model

The wide applicability of the joint BH and CRVS model makes it an appealing starting point for future research. Most major household surveys include BH questions as part of their standard questionnaire, making this method potentially usable in most low and middle income countries with a functioning CRVS system. This estimation approach also overcomes some limitations of BH-only geospatial modeling strategies, acknowledged in Burstein et al. (4): in countries with high-quality CRVS data, high estimates from CRVS can push child mortality estimates upwards when BH estimates are uncertain or biased downwards.

Because this model generates estimates both of child mortality and of CRVS completeness, it can be used programmatically to target multiple aspects of health system performance. Finally, the Bayesian modeling framework captures uncertainty in both surfaces, allowing for appropriately cautious interpretations of the results.

PAR:

PAR:

PAR:

4.2 Local variation in CRVS completeness across Mexico

PAR:

PAR:

PAR:

PAR:

PAR:

4.3 Relationship between neonatal mortality and vital statistics performance

PAR:

PAR:

PAR:

4.4 Model limitations

Decause CRVS completeness is only estimated directly through relatively uncommon field audits, it can suffer from wide uncertainty even when CRVS sample sizes are relatively high. This suggests that the CRVS completeness estimates should be treated with great caution, and this aspect of the results may need further refinement before it can be used to inform policy decisions. Other predictive factors could be also be included to estimate the CRVS completeness surface: for example, a suspected inverse relationship between child mortality rates and CRVS completeness could be explicitly incorporated into the formulation for the CRVS completeness surface (20).

There are other methodological limitations to this model that should be considered before it is widely implemented. Because mortality estimates are primarily grounded by BH data, this model is not applicable to countries where death registration is complete but recent BH surveys have not been conducted. The current model also does not account for biases in particular BH surveys, although a

larger regional model incorporating many surveys might be able to include a survey-specific random effect to account for bias. The conversion from the mortality probability (5q0) to a mortality rate (5m0) relies on a national-level input from the Global Burden of Disease project. While this estimate is relatively stable across countries, there might be in-country variation that is not captured. Finally, common to all geospatial work, the uncertainty associated with local estimates must be explicitly accounted for and described in any presentation of the results.

4.5 Conclusions

PAR:

5 References

1. Mikkelsen L, Lopez AD. Improving the quality and use of birth, death and cause-of-death information [Internet]. 2010. Available from: http://www.who.int/healthinfo/tool_cod_2010.pdf

2. UN General Assembly. Universal declaration of human rights. Vol. 2, UN General Assembly. 1948.

3. Dehnavieh R, Haghdoost AA, Khosravi A, Hoseinabadi F, Rahimi H, Poursheikhali A, et al. The District Health Information System (DHIS2): A literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. Health Inf Manag. 2019;48(2):62–75.

4. Burstein R, Henry NJ, Collison ML, Marczak LB, Sligar A, Watson S, et al. Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. Nature [Internet]. 2019 Oct 16;574(7778):353–8. Available from: http://www.nature.com/articles/s41586-019-1545-0

5. Reiner RC, Graetz N, Casey DC, Troeger C, Garcia GM, Mosser JF, et al. Variation in childhood diarrheal morbidity and mortality in Africa, 2000-2015. N Engl J Med. 2018;379(12):1128–38.

6. Adair T, Lopez AD. Estimating the completeness of death registration: An empirical method. PLoS One. 2018;13(5):1–19.

7. Målqvist M, Eriksson L, Nga NT, Fagerland LI, Hoa DP, Wallin L, et al. Unreported births and deaths, a severe obstacle for improved neonatal survival in low-income countries; a population based study. BMC Int Health Hum Rights. 2008;8(4):1–7.

8. Dicker D, Nguyen G, Abate D, Abate KH, Abay SM, Abbafati C, et al. Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for

the Global Burden of Disease Study 2017. Lancet [Internet]. 2018 Nov;392(10159):1684–735. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673618318919

9. Preston SH, Lahiri S. A short-cut method for estimating death registration completeness in destabilized populations. Math Popul Stud. 1991;3(1):39–51.

10. Wilson K, Wakefield J. Pointless spatial modeling. Biostatistics [Internet]. 2018;1–16. Available from: https://doi.org/10.1093/biostatistics/kxy041

11. Ross JM, Henry NJ, Dwyer-Lindgren LA, de Paula Lobo A, Marinho de Souza F, Biehl MH, et al. Progress toward eliminating TB and HIV deaths in Brazil, 2001–2015: a spatial assessment. BMC Med [Internet]. 2018 Dec 6;16(1):144. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1131-6

12. Osgood-Zimmerman A, Millear AI, Stubbs RW, Shields C, Pickering B V., Earl L, et al. Mapping child growth failure in Africa between 2000 and 2015. Nature [Internet]. 2018 Mar 1;555(7694):41–7. Available from: http://www.nature.com/articles/nature25760

13. Naghavi M, Makela S, Foreman K, O'Brien J, Pourmalek F, Lozano R. Algorithms for enhancing public health utility of national causes-of-death data. Popul Health Metr. 2010;8(1):1–14.

14. Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet [Internet]. 2017 Sep;390(10100):1151–210. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673617321529

15. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet [Internet]. 2018 Nov;392(10159):1736–88. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673618322037

16. Schmertmann CP, Gonzaga MR. Bayesian Estimation of Age-Specific Mortality and Life Expectancy for Small Areas With Defective Vital Records. Demography. 2018;55(4):1363–88.

17. Lindgren F, Rue H. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc Ser B. 2011;73(4):423–98.

18. Riebler A, Sørbye SH, Simpson D, Rue H, Lawson AB, Lee D, et al. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. Stat Methods Med Res. 2016;25(4):1145–65.

19. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2018. Available from: https://www.r-project.org/

20. Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM. TMB: Automatic Differentiation and Laplace Approximation. J Stat Softw. 2016;70(5).

21. Schlather M, Malinowski A, Oesting M, Boecker D, Strokorb K, Engelke S, et al. {RandomFields}: Simulation and Analysis of Random Fields [Internet]. 2019. Available from: https://cran.r-project.org/package=RandomFields

22. Marquez N. ar.matrix: Simulate Auto Regressive Data from Precision Matricies [Internet]. 2018. Available from: https://cran.r-project.org/package=ar.matrix

23. Krainski E, Gómez-Rubio V, Bakka H, Lenzi A, Castro-Camilo D, Simpson D, et al. Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. 2018.