Applying small area models to estimate mortality from birth history data: Under-5 mortality in Zambian districts, 1980-2010

Laura Dwyer-Lindgren[1], Frank Kakungu[2], Peter Hangoma[3], Marie Ng[1], Haidong Wang[1], Abraham Flaxman[1], Felix Masiye[1,3], and Emmanuela Gakidou[1]

[1] *Institute for Health Metrics and Evaluation, University of Washington*
[2] *Central Statistical Office of Zambia*
[3] *University of Zambia*

Abstract

Sub-national estimates of under-5 mortality are useful for evaluating within-country inequality, tracking progress, and identifying areas of greatest need. We estimate under-5 mortality for each of Zambia's 72 districts annually 1980-2010, using summary birth history data from censuses and complete birth history data from Demographic and Health surveys to fit a series of small area models. We consider a variety of generalized linear mixed models that differ in how spatial trends, temporal trends, and spatial-temporal interactions are introduced. All models suggest considerable heterogeneity in levels of under-5 mortality, with the worst off districts experiencing mortality risks 2-3 times as great as those in the best off districts. Distinct spatial trends are also apparent: districts in the northeast and southwest experience noticeably higher mortality than districts in the central part of the country. Progress in decreasing mortality over the past 30 years has also been variable: while there is some evidence of decline in most districts, our models suggest that a subset of districts have experienced decreases in mortality exceeding 50%.

1 Background

Over the period from 1980 to 2010 under-5 mortality (the probability that a child will die before reaching age five) dropped in Zambia by nearly 40% from 164 to 100 deaths per 1,000 live births [1]. Little is known, however, about how levels and trends in child mortality vary within Zambia. Estimation of under-5 mortality in Zambia is difficult as there is no vital registration system to collect information on deaths. National estimates are instead based on survey data in the form of birth histories where women are interviewed about the mortality experience of their children. While birth history data are widely used in countries like Zambia to estimate mortality at a national level, there are relatively few instances where they have been used to estimate mortality at a subnational level due primarily to concerns about small sample sizes. As a consequence, existing subnational estimates derived from birth histories have tended to rely on census data [2–4] where sample size is less of a concern, or have analyzed relatively coarse sub-national units (e.g., provinces or regions) [3,5].

Small area methods are statistical methods which address the issues raised by small sample sizes by explicitly accounting for the large sampling variance and exploiting spatial and temporal relatedness to increase predictive power. Although these methods have seen extensive use in

Table 1: Sample size (number of women) for each data source.

Source	Total	Median (Minimum - Maximum)
1990 Census	1,730,778	26,842 (3,692-204,557)
2000 Census	2,189,309	23,212.5 (3,799-277,404)
2010 Census	3,002,791	28,741.5 (5,253-477,915)
1992 DHS	7,060	89.5 (24-962)
1996-97 DHS	8,021	116 (20-876)
2001-02 DHS	7,658	-
2007 DHS	7,146	76 (12-483)

epidemiology and other fields, they have only infrequently been applied to birth history data on mortality and, even then, the focus has usually been on the relationship between socioeconomic or environmental factors and the risk of death, not prediction of under-5 mortality or other similar childhood mortality indicators[6–12]. In fact, mortality data pose several unique challenges in the context of small area methods. Most small area methods used in epidemiology are designed to work with count data; under-5 mortality, the most common indicator of childhood mortality, is a complex construct, however, not a simple count. Further, in order to derive estimates of under-5 mortality from certain types of birth history data, demographic models must be employed, adding an additional modeling step.

In this paper we present a method of combining birth history methods and small area methods to estimate under-5 mortality from birth history data. We apply this method to districts in Zambia from 1980 to 2010 and report on levels and trends in mortality over that period.

2 Methods

2.1 Data

We used birth history data from three population censuses–1990, 2000, and 2010–and four DHS–1992, 1996-97, 2001-02, and 2007[13–19]. The population censuses include summary birth histories for all women of reproductive age while the DHS include complete and summary birth histories for all women of reproductive age. Our unit of analysis was the district as defined in 2010 for a total of 72 districts grouped into nine provinces. Data collected prior to 2000 used a different set of districts totaling 57. For districts that split in the transition from 57 districts to 72 districts, data from the original district were assigned to all inheriting districts. Unlike the earlier DHS and all of the censuses, the 2001-02 and the 2007 DHS datasets did not contain a district variable. For the 2007 survey the latitude and longitude of each cluster were available and we used this to map clusters to districts. There was no information available in the 2001-02 survey that allowed us to identify districts, so we used this survey only at the province level. Table 1 gives the total sample size as well as the range and median of the sample sizes across districts.

2.2 Birth history methods

There are two different types of birth histories. Complete birth histories (CBH) are where women are asked detailed questions about each child they have given birth to: the date of birth, survival status, and, when applicable, age at death. In contrast, in a summary birth history (SBH) women are asked only for the total number of children they have given birth to and the number of these children that have died. CBH contain sufficient information to calculate under-5 mortality directly, but SBH do not and consequently require demographic models which relate information about total children born and died to under-5 mortality in specific time periods. SBH are more frequently collected, however, because they impose a considerably smaller time burden than CBH.

We analyzed CBH data from the DHS using standard methods described elsewhere[20] to produce estimates of under-5 mortality from the pooled DHS data for each year prior to the most recent survey in each district. We analyzed SBH data from the censuses using the maternal age period (MAP) method described by Rajaratnam et al.[21,22].

2.3 Census correction

Exploratory analyses of the birth history data at the province and national level suggested that the estimates from the census were systematically lower than estimates from the DHS. We posit that this is due to data errors in the census: it is likely that the birth histories in the census, which are somewhat secondary to the primary purpose of a census, are collected with less care than the birth histories in the DHS which are a central component of the survey. We correct census SBH data before fitting the small area model by finding the mean ratio in each province of $q5$ in the DHS to $q5$ in the census matched by year and multiplying the census SBH estimates from all districts within each province by this factor.

2.4 Small area model

The small area model we used can be described as a generalized linear mixed model. We specified a normal likelihood for logit-transformed under-5 mortality (denoted $q5_{i,t}$), which describes the probability of observing the data given two parameters: $\theta_{i,t}$, the mean in district i at time t, and σ^2, the variance:

$$\text{logit}(q5_{i,t})|\theta_{i,t}, \sigma^2 \sim \text{Normal}(\theta_{i,t}, \sigma^2) \tag{1}$$

We employed a logit transformation as this has the desirable property that it restricts predictions of under-5 mortality from all models to between 0 and 1. Our focus in this analysis was on the form of the regression equation for $\theta_{i,t}$ which we specified as a linear combination of spatial effects, temporal effects, and spatial-temporal interaction effects.

We considered a number of different forms for the temporal trend including categorical (i.e., time is modeled by a series of dummies on year), linear, random walk, B-spline, and natural spline. We determined that a natural spline model with 1 interior knot provided the best balance between flexibility and sufficient smoothing over time. Natural splines are incorporated into our model by introducing fixed effects on K spline bases ($S^{(k)}(t)$), where K and $S^{(k)}(t)$ are determined by the type of spline and the number of interior knots[23]:

$$\theta_t = \sum_{k=1}^{K} \beta^{(k)} \cdot S^{(k)}(t) \tag{2}$$

To model spatial trends, we included a series of district-level random effects similar to those proposed by Besag et al.[24]. Typically two random effects are included. The first has an intrinsic conditional autoregressive (ICAR) prior:

$$u_i | u_{j, j \in \delta_j} \sim \text{Normal}\left(\frac{1}{n_i}\sum_{j \in \delta_i} u_j, \frac{\sigma_u^2}{n_i}\right) \tag{3}$$

Under this prior, the effect u_i for each district i is normally distributed around the mean of the effect in neighboring districts (δ_j), indicating a spatially smooth process. We defined neighbors in terms of queen adjacency: districts with borders that share at least one point are considered neighbors. The second random effect has an independent and identically distributed (IID) prior:

$$v_i \sim \text{Normal}(0, \sigma_v^2) \tag{4}$$

Under this prior, the effect v_i for each district i is independent of that for all other districts but the districts share a common variance. We considered models with the IID random effect only (which would be appropriate if there is little or no spatially structured variation), the ICAR effect only (which is appropriate if all variation is spatially structured), and both effects. We found that the model with both effects was most appropriate.

Ultimately we are interested not only in overall time trends or a single time-invariant spatial pattern, but rather spatial patterns that change over time or, equivalently, temporal patterns that vary by district. We followed the example of several authors[25–27] and combined the temporal and spatial models just described by fitting IID and ICAR random effects for each spline basis in addition to the fixed effects:

$$\theta_{i,t} = \sum_{k=1}^{K} (\beta^{(k)} + u_i^{(k)} + v_i^{(k)}) \cdot S^{(k)}(t) \tag{5}$$

In this model, all districts share a global trend given by $\sum_{k=1}^{K} (\beta^{(k)} \cdot S^{(k)}(t))$. In addition, each district deviates from the trend as given by $\sum_{k=1}^{K} (u_i^{(k)} + v_i^{(k)}) \cdot S^{(k)}(t)$. Because these deviations include both an IID and an ICAR random term, it is possible for the district trends to vary in a spatially structured and spatially unstructured way.

Finally, the birth history data we introduced into these small area models violate one important assumption: that, conditional on $\theta_{i,t}$, each observed $q5_{i,t}$ is independent. Most mothers interviewed have multiple children and in a CBH these children contribute to estimates in multiple time periods, leading to some correlation between the different estimates. The situation is more severe for SBH: since the entire series is modeled, and further since there is a considerable amount of smoothing involved in the demographic models used to generate these estimates, SBH estimates from any given source are highly correlated. We address this by including a district-source level IID random effect in our model to account for the correlation within each source in each district. Thus the final model is

$$\theta_{i,t,s} = \sum_{k=1}^{K} (\beta^{(k)} + u_i^{(k)} + v_i^{(k)}) \cdot S^{(k)}(t) + w_{i,s} \tag{6}$$

where s denotes the source (DHS, 1990 census, 2000 census, or 2010 census) and $w_{i,s}$ is the IID random effect on district-source. When making predictions we set $w_{i,s}$ to 0.

Normal priors with mean 0 and variance 1,000 were used for all fixed effects and gamma (0, 0.00005) priors were specified for the log of the inverse variance (precision) of all random effects. All models were fit using the INLA program[28] in R version 2.15.2[29]. This program provides the median and 2.5th and 97.5th percentile of the posterior distribution for $\theta_{i,t}$ which we then inverse-logit transformed to arrive at our estimates and confidence intervals for under-5 mortality in each district and year. We also calculated confidence intervals for changes between two years in under-5 mortality within each district by simulating 10,000 draws each from the marginal posterior distributions for both years and finding the 2.5th and 97.5th percentile of the differences between these two sets of draws. This procedure ignores the covariance between estimates for different time points, which is not readily available, and thus provides a conservative estimate of the confidence intervals for changes since it likely that the estimates for two different time points within a single district are positively correlated.

3 Results

Figure 1 shows the estimated under-5 mortality for each district every five years between 1980 and 2010. These estimates reflect the national trend of decreasing mortality over this 30 year period: it is evident from these maps that mortality has declined in nearly all districts. In earlier years there is a distinct regional pattern where districts in the north, east, and southwest have the highest levels of mortality, districts in the most urban parts of the country (Lusaka, the capitol, and Copperbelt province) have the lowest levels of mortality, and districts in the rest of the central part of the country have intermediate levels of mortality. This pattern persists relatively unaltered from 1980 to 1995, but starting in the map for 2000 mortality begins to improve in all districts and the highest mortality districts begin to catch up somewhat, while the urban areas appear to lose some of their advantage.

The changes in under-5 mortality over time are shown explicitly in figure 2 which shows the percent change in mortality over each decade and over the entire 30-year period. These maps tell a story of accelerating progress. In the 1980s, most districts made relatively little progress in decreasing child mortality while in 37 districts there is actually some evidence of an increase in mortality. In the 1990s, most districts begin to make moderate progress and in the 2000s there is evidence of progress in nearly all districts, in some cases with decreases exceeding 40% over the decade. Considering the period as a whole, mortality has decreased in nearly all districts, though clearly there is a considerable amount of variation, with some districts experiencing only minimal declines (less than 5% in 2 districts) while others experience large declines (more than 50% in 10 districts). These mortality estimates are associated with some uncertainty, so it is also interesting to look at where there are statistically significant declines (at the 95% level) despite this uncertainty: figure 3 shows changes over the same periods as figure 2, but classifies each change as a significant increase, a non-significant increase, a significant decrease, or a non-significant decrease. There are no significant increases over any period for any district. In the 1980s the districts are fairly evenly split between non-significant increases and non-significant decreases. In the 1990s, all but one district is in the non-significant decrease group. In the 2000s there starts to be evidence of significant declines in some districts, and when considering the entire 30 year period more than half (48) of districts actually experienced a statistically significant decline in under-5 mortality.

Figures 1 and 2 both suggest that the districts that have experienced the greatest declines are those that started at the highest levels of mortality. Similarly, figure 3 shows that districts with statistically significant declines over the entire 30 year period are concentrated in those areas that had the highest levels of mortality in 1980. Figure 4 plots the percent change in under-5 mortality between 1980 and 2010 against the level of mortality in 1980 for all 72 districts and also shows the ordinary least squares regression line for this relationship. This emphasizes the observation from the earlier figures that, in general, districts with higher mortality in 1980 have experienced greater declines: the ordinary least squares regression suggests that the percent decline was 2.5 percentage points greater for every 10 per 1,000 live birth higher the level of under-5 mortality in 1980. Nonetheless, at any given level of mortality in 1980, and particularly at the lower levels, there is still considerable variation in the amount of progress made over the next 30 years. Districts that experienced especially large declines given their starting level of mortality in 1980 have been explicitly labeled on this plot.

Given that districts with the highest level of mortality have generally experienced greater declines in mortality it is interesting to consider how within-country inequality has changed. Figure 5 shows the range (light gray box), interquartile range (dark gray box) and median (black line) for mortality at the district level every five years from 1980 to 2010. This figure shows that not only does the median of the distribution shift downward, indicating declining mortality overall, but the amount of variation in the distribution, as measured by the range and interquartile range, also decreases, indicating declining absolute inequality. Relative inequality has also decreased. In 1980 a child born in the Lundazi district (Eastern province), the district with the highest mortality, was 2.4 times more likely to die before age five than one born in Mufulira district (Copperbelt province), where mortality was lowest. In contrast, in 2010 a child born in Chiengi district (Luapula province), where mortality is highest, is 1.8 times more likely to die before age five than one born in Mongu district (Western province), where mortality is lowest. Although both the overall level of child mortality and the degree of inequality within country remain high, both improved markedly between 1980 and 2010.

4 Discussion

Combining birth history methods and small area methods is useful for estimating under-5 mortality in a country like Zambia where more traditional data sources are unavailable. The model described here borrows strength in space and time in order to improve predictions despite small sample sizes at the district level. Further, we are able to incorporate data from both summary and complete birth histories, allowing us to make use of all available data sources. Finally, these methods allow us not only to predict under-5 mortality in each district, but also to quantify the uncertainty associated with all predictions.

District level estimates of under-5 mortality in Zambia reveal sub-national trends that are masked by national level estimates. While mortality has declined in nearly all districts, as at the national level, the magnitude of this decline varies considerably within Zambia. This has had implications for the amount of inequality between districts: because districts with the highest initial levels of mortality were also the most likely to experience larger declines inequality has decreased over the past 30 years in Zambia. Nonetheless, under-5 mortality remains high in all districts and there is still a nearly two-fold difference between the best performing and worst performing districts. These findings have important implications for future efforts to continue to reduce child mortality in Zambia. First, districts that have experienced the greatest declines merit study: what were the drivers of these declines? Can the lessons learned from these districts be applied to other districts where progress has been less impressive? Second, districts that have experienced little progress merit renewed attention: what is keeping these districts from making the kind of progress made in other parts of the country? Further, the methods described in this paper for combining birth history data and methods with small area models are easy to update–as new survey data become available these methods could be used to track progress in the immediate past, providing an important feedback mechanism with regards to efforts to improve child health and survival.

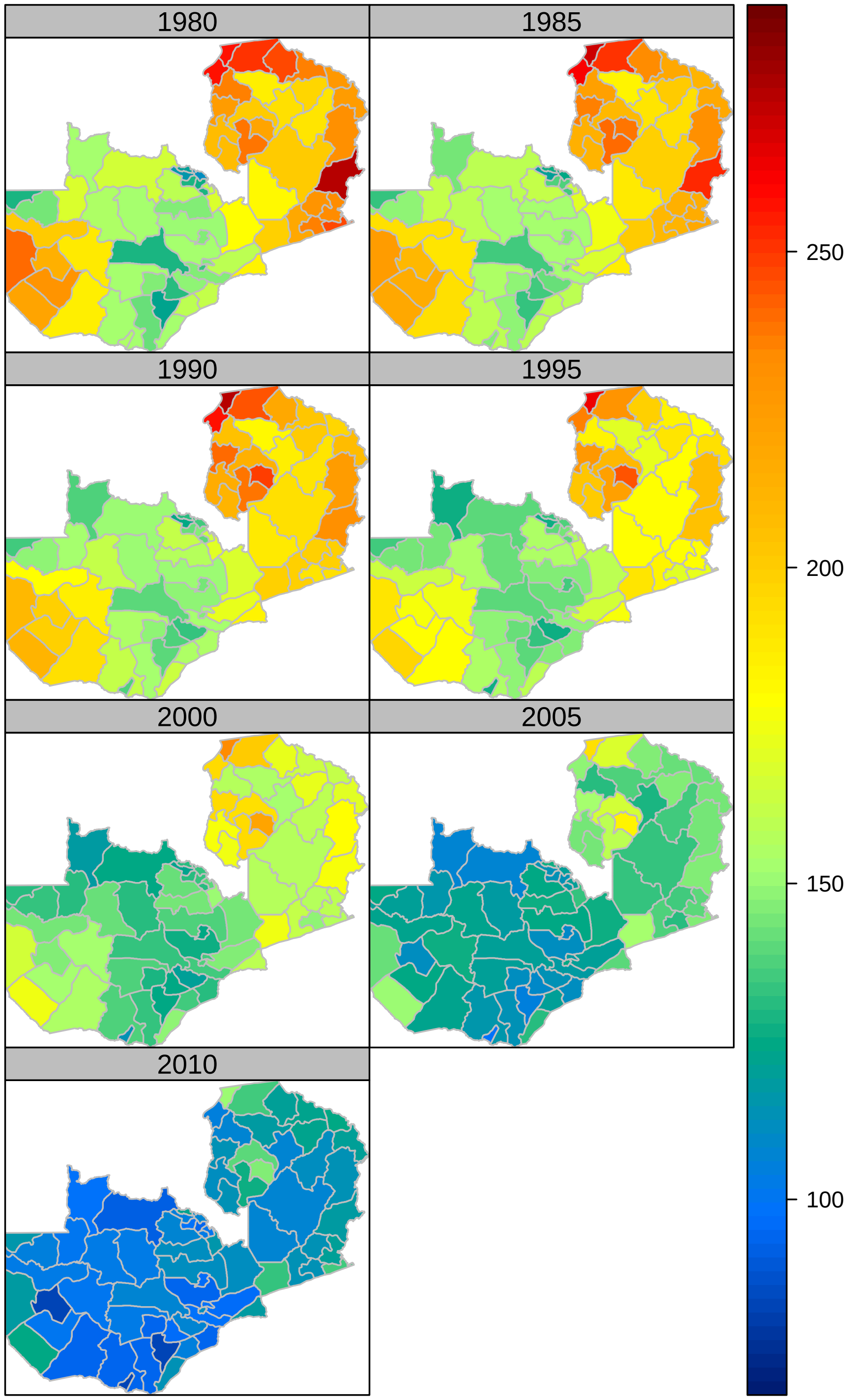

Figure 1: Under-5 mortality by district from 1980 to 2010 (deaths per 1,000 live births).

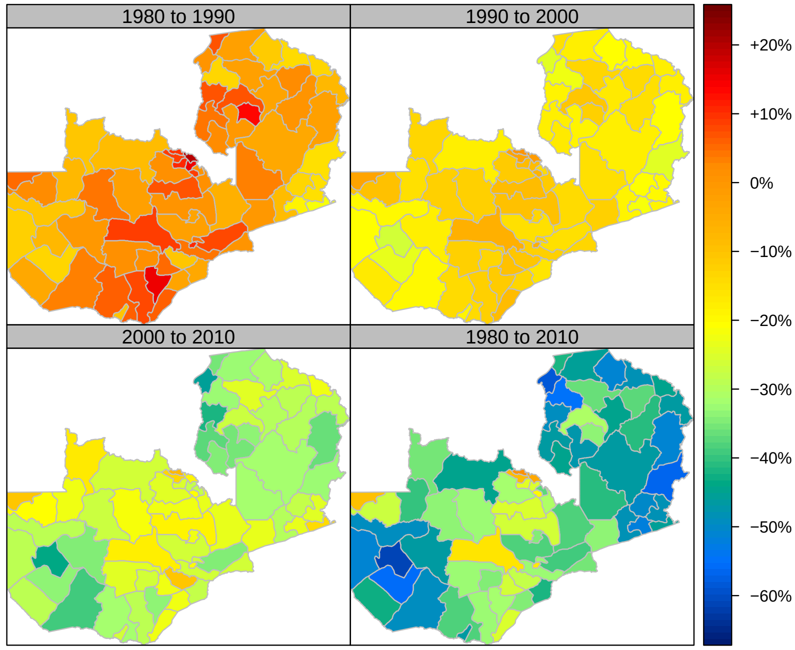

Figure 2: Percent change in under-5 morality by district for each decade and from 1980 to 2010.

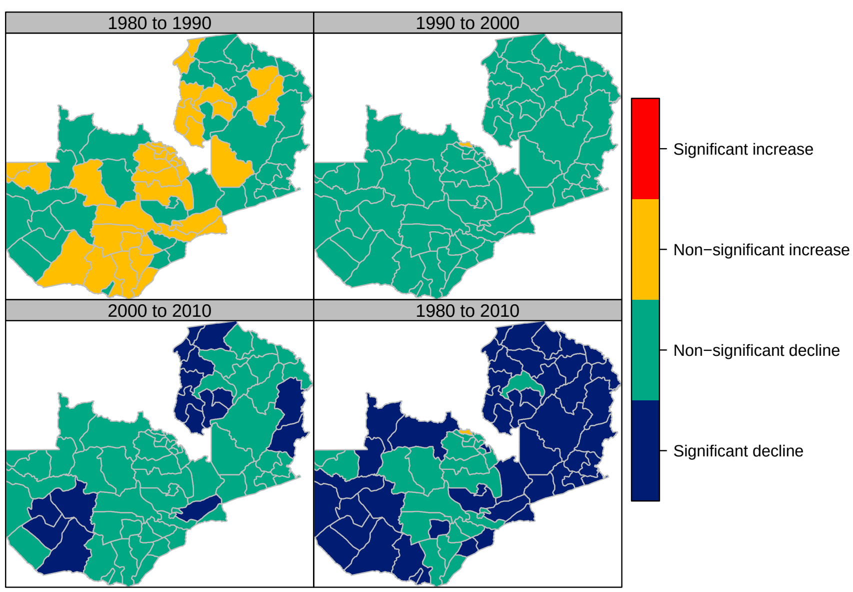

Figure 3: Statistical significance and direction of change in under-5 mortality by district for each decade and from 1980 to 2010.

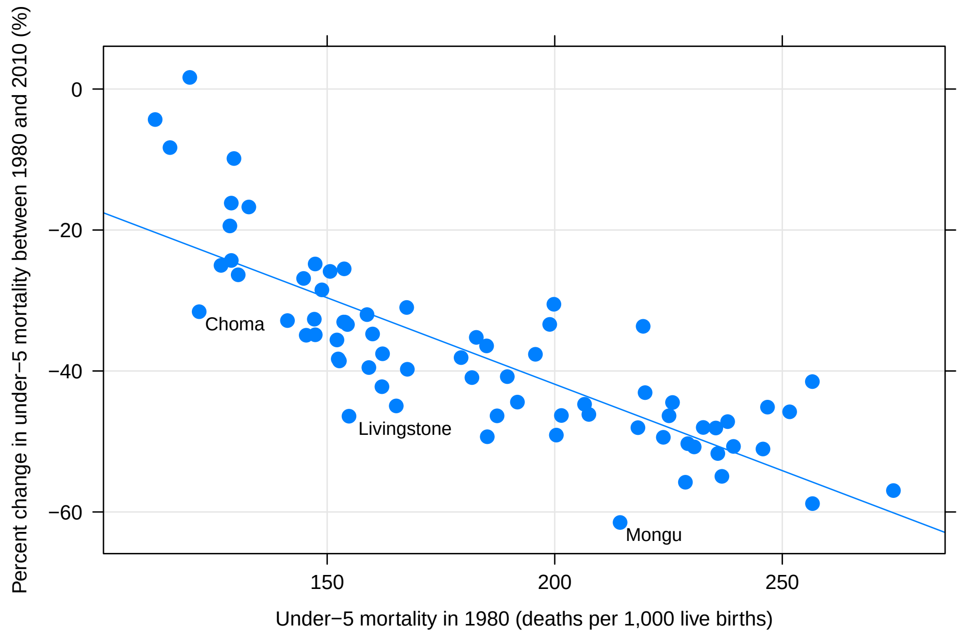

Figure 4: Percent change in under-5 morality from 1980 to 2010 and level of under-5 mortality in 1980.

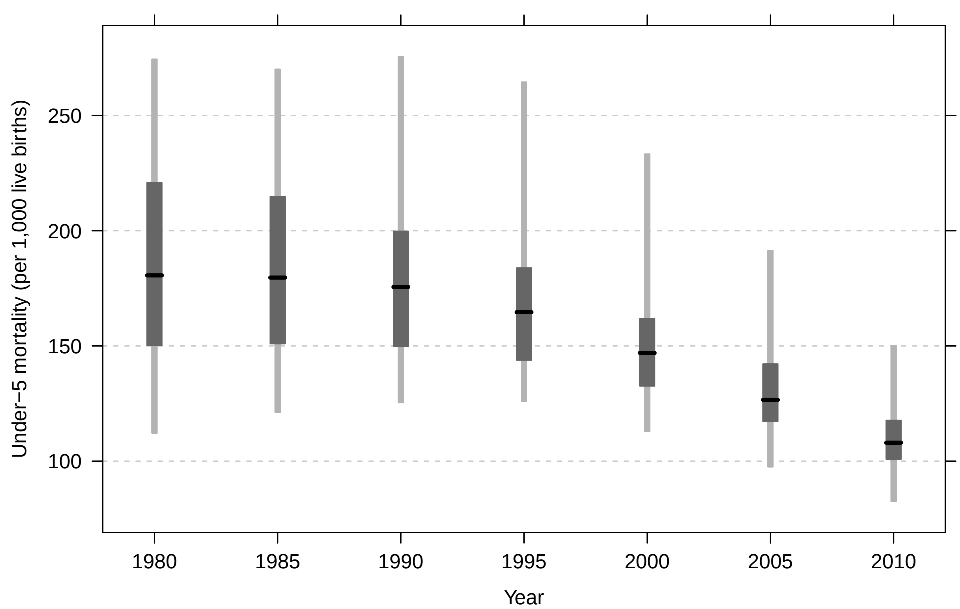

Figure 5: Distribution of district-level under-5 mortality by year from 1980 to 2010. Light gray box gives the range, dark gray boxes the interquartile range, and black circles the median across all districts in each year.

References

[1] Wang H, Dwyer-Lindgren L, Lofgren KT, Rajaratnam JK, Marcus JR, Levin-Rector A, et al. Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012;380(9859):2071–2094.

[2] Bangha M, Simelane S. Spatial Differentials in Childhood Mortality in South Africa: Evidence from the 2001 Census. African Population Studies. 2007;22(2):3–21.

[3] Storeygard A, Balk D, Levy M, Deane G. The global distribution of infant mortality: a subnational spatial view. Population, Space and Place. 2008;14(3):209–229.

[4] Bauze AE, Tran LN, Nguyen KH, Firth S, Jimenez-Soto E, Dwyer-Lindgren L, et al. Equity and Geography: The Case of Child Mortality in Papua New Guinea. PLoS ONE. 2012;7(5):e37861.

[5] Singh A, Pathak PK, Chauhan RK, Pan W. Infant and Child Mortality in India in the Last Two Decades: A Geospatial Analysis. PLoS ONE. 2011;6(11):e26856.

[6] Gemperli A, Vounatsou P, Kleinschmidt I, Bagayoko M, Lengeler C, Smith T. Spatial Patterns of Infant Mortality in Mali: The Effect of Malaria Endemicity. American Journal of Epidemiology. 2004;159(1):64–72.

[7] Adebayo SB, Fahrmeir L, Klasen S. Analyzing infant mortality with geoadditive categorical regression models: a case study for Nigeria. Economics & Human Biology. 2004;2(2):229–244.

[8] Kandala NB, Ghilagaber G. A Geo-Additive Bayesian Discrete-Time Survival Model and its Application to Spatial Analysis of Childhood Mortality in Malawi. Quality and Quantity. 2006;40(6):935–957.

[9] Kazembe LN, Appleton CC, Kleinschmidt I. Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi. Geospatial Health. 2007;2(1):41–50.

[10] Kazembe LN, Mpeketula PMG. Quantifying Spatial Disparities in Neonatal Mortality Using a Structured Additive Regression Model. PLoS ONE. 2010;5(6):e11180.

[11] Asiimwe J, Jehopio P, Atuhaire L, Mbonye A. Examining small area estimation techniques for public health intervention: lessons from application to under-5 mortality data in Uganda. Journal of Public Health Policy. 2011;32(1):1–14.

[12] Chin B, Montana L, Basagana X. Spatial Modeling of Geographic Inequalities in Infant and Child Mortality Across Nepal. Health & Place. 2011;17(4):929–936.

[13] Central Statistical Office (Zambia). Zambia Census of Population, Housing and Agriculture 1990; 1990. Lusaka, Zambia: Central Statistical Office.

[14] Central Statistical Office (Zambia). Zambia Census of Population and Housing 2000; 2000. Lusaka, Zambia: Central Statistical Office.

[15] Central Statistical Office (Zambia). Zambia Census of Population and Housing 2010; 2010. Lusaka, Zambia: Central Statistical Office.

[16] University of Zambia, Central Statistical Office (Zambia), Macro International, Inc. Zambia Demographic and Health Survey 1992; 1992. Calverton, MD: Macro International, Inc.

[17] Central Statistical Office (Zambia), Macro International, Inc, Ministry of Health (Zambia). Zambia Demographic and Health Survey 1996-1997; 1997. Calverton, MD: Macro International, Inc.

[18] Central Statistical Office (Zambia), Macro International, Inc, Central Board of Health (Zambia). Zambia Demographic and Health Survey 2001-2002; 2002. Calverton, MD: Macro International, Inc.

[19] Central Statistical Office (Zambia), Macro International, Inc. Zambia Demographic and Health Survey 2007; 2007. Calverton, MD: Macro International, Inc.

[20] Dwyer-Lindgren L, Gakidou E, Flaxman A, Wang H. Error and bias in under-5 mortality estimates derived from birth histories with small sample sizes. Population Health Metrics. 2013;11(1).

[21] Rajaratnam J, Tran L, Lopez A, Murray C. Measuring Under-Five Mortality: Validation of New Low-Cost Methods. PLoS Med. 2010;7(4):e1000253.

[22] Lozano R, Wang H, Foreman K, Rajaratnam J, Naghavi M, Marcus J, et al. Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: an updated systematic analysis. The Lancet. 2011;378(9797):1139–1165.

[23] Hastie T, Ribshirani R, Friedman J. Elements of Statistical Learning: data mining, inference, and prediction. 2nd ed. Springer-Verlag; 2009.

[24] Besag J, York J, Molli A. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics. 1991;43(1):1–20.

[25] MacNab YC, Dean CB. Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates. Biometrics. 2001;57(3):949–956.

[26] MacNab YC, Dean CB. Spatio-temporal modelling of rates for the construction of disease maps. Statistics in Medicine. 2002;21(3):347–358.

[27] Silva GL, Dean CB, Niyonsenga T, Vanasse A. Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines. Statistics in Medicine. 2008;27(13):2381–2401.

[28] Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2009;71(2):319–392.

[29] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2013. Available: www.R-project.org.