Introduction

Timely and accurate estimates of under-five mortality are important in assessing the effectiveness of intervention programs and public health policy. This is especially relevant as we approach 2015, the year Millennium Development Goal (MDG) 4 aims for a two-thirds reduction in child mortality from 1990. Such estimates can also better guide global action in the post-MDG era.

Trends in under-five mortality are nontrivial to estimate, as most countries do not have a vital registration system that captures all deaths. Mortality data from different sources in the same country often differ. A variety of statistical methods have been applied to create a single time series of point estimates and confidence bands in each country (Wang et al., 2012; Alkema & New, 2013).

In estimating total fertility rate, Alkema et al. (2012) use UN Population Division estimates as an unbiased reference for estimating and correcting for bias in different data sources. This idea of eliminating data bias has potential applicability in mortality estimation as well.

The method presented here, based on the under-five mortality model for the Global Burden of Disease Study 2010 (Wang et al., 2012), incorporates data bias adjustment into the modelling process; it simplifies the first stage regression and corrects for data bias without referencing a previously synthesized time series of point estimates.

Methods

Here we use a nonlinear mixed effects regression to both estimate data bias and provide first stage predictions. A space-time local linear smoother, using bias-adjusted points and first stage predictions, creates a prior for the Gaussian process regression, which produces final estimates and confidence bands.

The nonlinear mixed effects regression equation is

$$_5m_{0cys} = \exp[(\beta_1 + \gamma_{1c}) * \log\left(LDI_{cy}\right) + (\beta_2 + \gamma_{2c}) * education_{cy} + \gamma_c + \gamma_{cs} + \alpha_t] + (\beta_3 + \gamma_{3c}) * HIV_{cy} + \varepsilon_{cys}$$

where *c* is country, *y* is year, *s* is source, and *t* is source type; each source is categorized into one of 16 source types across all countries.

Additionally,

$_5m_0$ is under five mortality rate

LDI is lagged distributed income per capita

education is mean years of education for women of reproductive age (15-49 years)

HIV is under-five crude death rate due to HIV

γ is a random effect

α is a fixed effect on source type β is fixed effect coefficient

ε is the residual

For each country, we rely on expert opinion to choose a source, or combination of sources, which we believe to be the least biased over the time series. If a country has vital registration which we deem to be complete, this is the reference source. If a country does not have complete vital registration, but has Demographic Health Surveys (DHS), direct estimates (from complete birth history questionnaires) are chosen as the reference source. If a country has neither of these types of data or DHS surveys are deemed unreliable, we assign the surveys conducted after 1980, in combination, as the reference (incomplete vital registration data is not included). Additionally, based on expert opinion, in some countries we choose as a reference the combination of data from DHS and another reliable survey.

Each survey has an associated random effect as well as a source type fixed effect. The values of these random and fixed effects for the reference sources are deemed to be the true deviation from the predictions. In cases where there are multiple reference sources, the mean of the random and fixed effects is taken as this true deviation value. We adjust the biased sources by including these reference values for the random and fixed effects values instead of those estimated for each individual source, as shown below.

$$adjusted_5m_{0,cys} = \exp[(\beta_1 + \gamma_{1c}) * \log\left(LDI_{cy}\right) + (\beta_2 + \gamma_{2c}) * education_{cy} + \gamma_c + \gamma_{ref,c} + \alpha_{ref,c}] + (\beta_3 + \gamma_{3c}) * HIV_{cy} + \varepsilon_{cys}$$

The exception to this correction is incomplete vital registration data, which is adjusted upwards using a five year rolling mean of the difference between incomplete vital regression and a LOESS of the already-adjusted survey data.

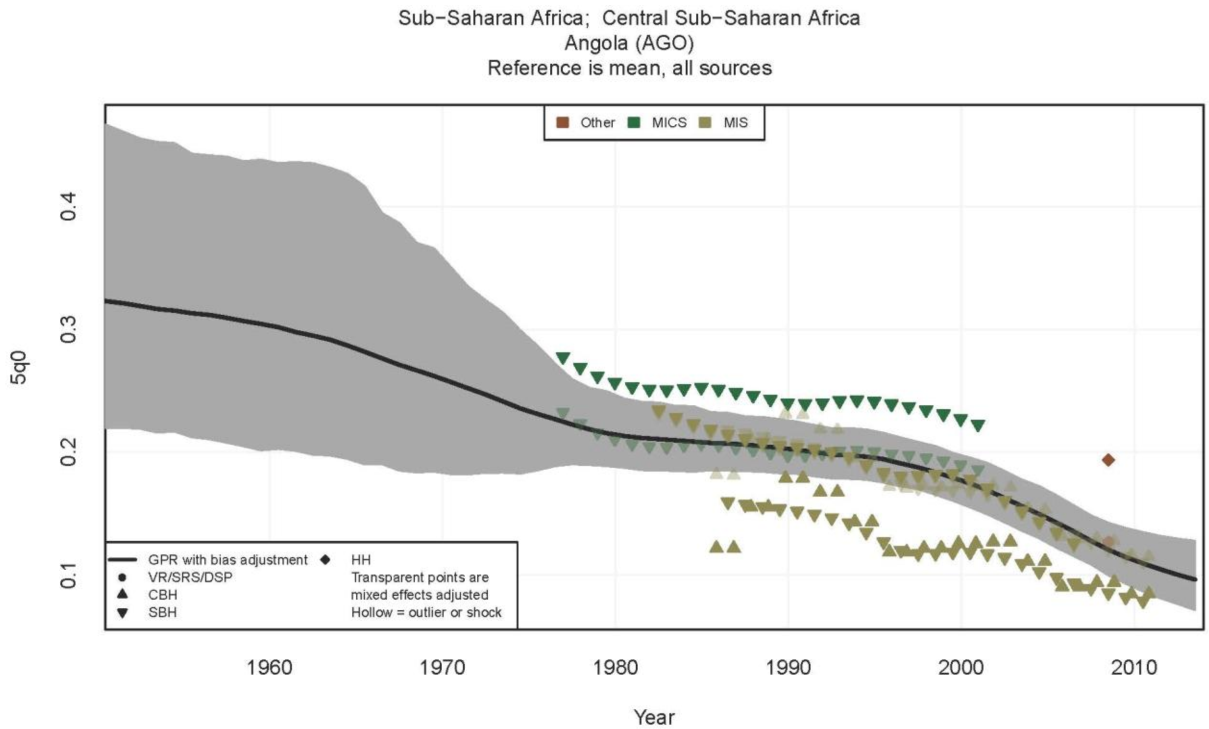
The adjusted data are then used as the input data for the space-time local smoother. The predicted time series for this smoother is obtained from the equation below; no random effects or survey type fixed effects are included.

$$predicted_5m_{0,cy} = \exp[\beta_1 * \log\left(LDI_{cy}\right) + \beta_2 * education_{cy} + \alpha_{intercept}] + \beta_3 * HIV_{cy}$$

Finally, the output of the space-time local smoother is used as a prior for the Gaussian process regression, which produces a final time series of point estimates, as well as confidence bounds. The space-time local smoother and Gaussian process regression are further described in The Lancet, December 15, 2012 (Wang, et al.).

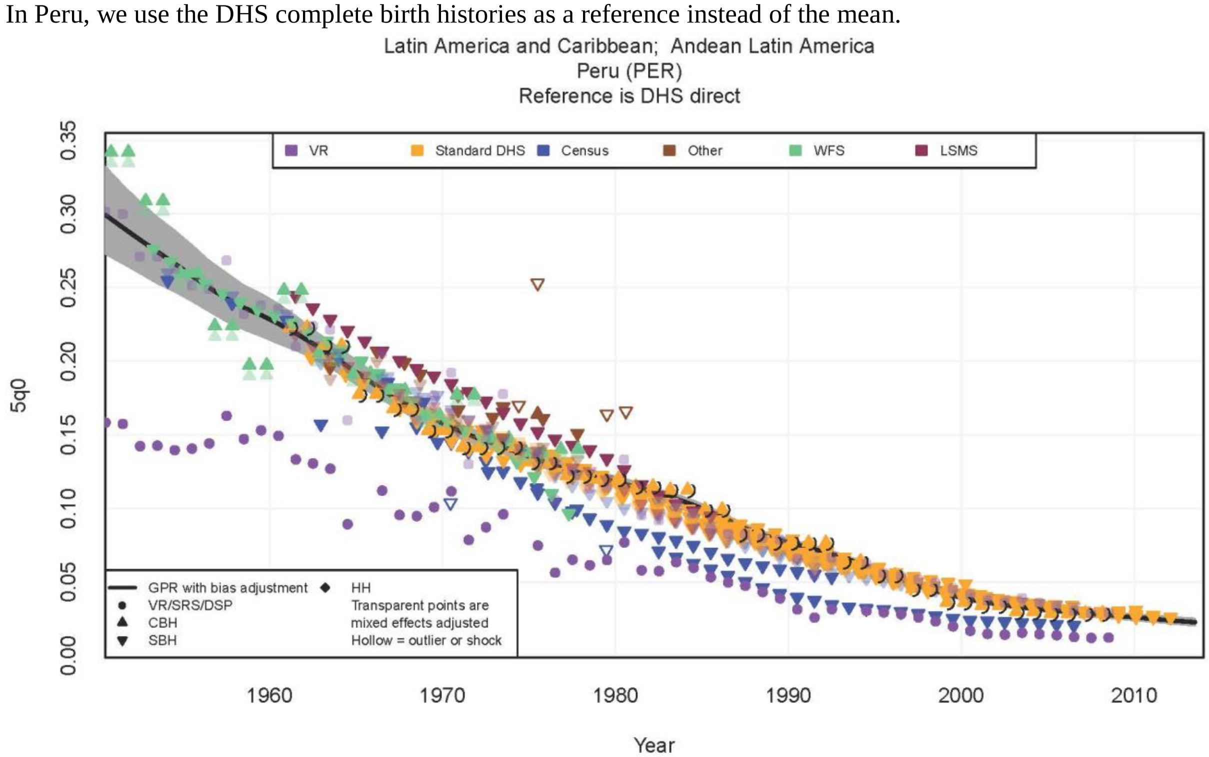
Results and Conclusions

The methods above give a time series of under-five mortality estimates for 1950-2013 for each country, using an adjusted and unbiased dataset. This allows us to use more data to inform the trend of mortality, and creates a more cohesive dataset. One country in which we see this is Angola.



In Angola, we see that the MICS and MIS are both potentially biased; the MICS being biased upwards and the MIS downwards. Our method assumes that combination of all points is unbiased, and so adjusts both data series to somewhere in between. Once the data is adjusted, it creates a much more cohesive time series. Additionally, we see that the confidence bands become wider, reflecting more accurate the uncertainty in the estimates.

In Peru, we use the DHS complete birth histories as a reference instead of the mean.



In addition, Peru differs from Angola in that it is data rich and has a vital registration system (though incomplete). While we have outliered a few points in Peru, most data series have been included in the analysis, and we clearly see that they are adjusted to the level of the DHS complete birth history estimates (black circles), giving a much more cohesive data series.

Finally, this is an automated algorithm, with the additional option for expert opinion in choosing unbiased surveys. This estimation technique requires, as do any estimates using expert opinion, strong judgment on the part of researchers and full engagement with debate in the scientific community to ensure impartiality. This will be an ongoing process as future iterations of the Global Burden of Disease study are published and new data is made available.

This new modelling process is an important step forward in mortality estimation from multiple sources with potentially strong and varied biases. Adjustment for data source biases, in combination with spatial-temporal smoothing and Gaussian Process regression, will greatly improve our estimates of child mortality in countries without complete civil registration systems.

References

Alkema, L., & New, J. R. (2013, September 9). Global estimation of child mortality using a Bayesian B-spline bias-reduction model. Retrieved from http://arxiv.org/pdf/1309.1602v1.pdf

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F. (2012, April 25). Estimating trends in the total fertility rate with uncertainty using imperfect data: Examples from West Africa. *Demographic Research, 25*, 331-362. doi: 10.4054/DemRes.2012.26.15

Wang, H., Dwyer-Lindgren, L., Lofgren, K. T., Rajaratnam, J. K. , Marcus, J. R., Levin-Rector, A., . . . Murray, C. J. L. (2012, December 15). Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2071-2094.