Small area estimation of fertility at the district-level in sub-Saharan Africa

Background

Fertility estimation is core to understanding evolving demographic trends and population structures in sub-Saharan Africa, and is required for the targeting and evaluation of programmatic activities, including family planning services and HIV prevention and treatment, and monitoring progress towards Millennium and Sustainable Development Goals on sexual and reproductive health.

Vital registration systems can provide direct measures of fertility, but these are often underdeveloped or absent in much of sub-Saharan Africa. Instead, the reconstruction of fertility trends from survey- and census-based surveillance sources in these settings is well established, by which biennial national level fertility estimates are published by UN Population Division and the Institute of Health Metrics and Evaluation (IHME). National level estimates, however, give little insight into subnational heterogeneity and how best to target programmatic activities.

Accordingly, policy makers and funders are increasingly interested in subnational population and demographic estimates furnished at the level of local decision making – often the second administrative level. Whilst mortality estimates, particularly those of paediatric mortality, have received more attention at finer geospatial disaggregation, there are fewer sources of subnational fertility estimates.

Estimates at the first administrative level are produced in X countries with the support of the US Census Bureau as part of demographic inputs into the Spectrum modelling tools, and used to inform a range of public health policy making. The WorldPop project produce 1km x 1km pixel-level estimates of births, often derived from first administrative level fertility rates. These are then applied to pixel-level populations. At present there exist no fertility estimates for district-level decision-making estimated from granular data at that level.

Obtaining direct measures of fertility at the district level by single calendar year from surveys is challenging - sample sizes are small, births are few, and uncertainty is large. This is compounded when estimating age-specific fertility rates. Existing literature on under 5 mortality has employed small area methods, leveraging spatiotemporal correlation and smoothing, to produce model-based estimates from survey data.

- Find more subntl fertility
- Expand on SAE
- Cite etc etc.

Here we present an analysis of nationally representative household survey data that furnishes district level estimates of age-specific and total fertility rates in 11 countries in sub-Saharan Africa.

Methods

Data extraction and preparation

Data from X nationally representative household surveys conducted in 11 countries since 1995 were extracted. Geomasked cluster coordinates were available for X surveys (X Demographic Household Surveys (DHS), X Malaria Indicator Surveys (MIS) and X AIDS Indicator Surveys (AIS)). Subnational area boundaries and hierarchies were largely concordant between surveys. Required adjustments to coerce geographic boundaries into a single set of boundaries across all surveys are detailed in Supplementary Table X.

Cluster coordinates were assigned to a given district by overlaying the final district resolution national shapefile. Cluster coordinates are unavailable for Multiple Indicator Cluster Surveys (MICS) and were used at the lowest administrative level included as a survey indicator, normally provincial level. The modelled administrative level and administrative area hierarchy for each of the 11 countries is shown in Supplementary Table X. For ease of reference, the first administrative level is henceforth referred to as "province", and the chosen administrative level for modelled estimates in each country as "district".

Birth histories and century-month codes of date of interview and respondent's date of birth were used to reconstruct numbers of births and survey-weighted observed person years stratified by district (DHS, MIS, AIS) or province (MICS), 5 year age groups, and single calendar years for 15 years preceding DHS surveys and 5 years preceding MIS, AIS, and MICS surveys (Fig X).

Statistical analysis

Let y be the fertility for a given age, $a \in (15 - 19, 20 - 24...45 - 49)$, in a given district, $i \in 1, 2, 3...i$, at a given time, $t \in 1995 : 2018$ and be given by:

$$y_{ait} \sim Po(\lambda_{ait}.E_{ait})$$

where λ is the observed fertility rate, and E survey observed person years. We decompose λ into:

$$log(\lambda_{ait}) = \mu + \alpha_a + \gamma_t + \delta_i + \eta_{a,t} + \eta_{a,i} + \eta_{i,t}$$

where μ is the model intercept, α_a is a structured first order random walk on age group, γ_t is a structured second order random walk on calendar year, and δ_i is the BYM2 spatial model (Riebler et al. 2016). Additional flexibility is provided through three Type IV interaction terms (Knorr-Held 2000) - between time (RW2) and age (RW1) - $\eta_{a,t}$, age (RW1) and space (ICAR) - $\eta_{a,i}$, and space (ICAR) and time (RW2) - $\eta_{i,t}$.

Adjusting for Time Preceding Survey bias in Demographic Health Survey data

Several data quality issues have the potential to impact the quality of fertility estimates derived from DHS survey data. Two sources of bias are addressed here; birth displacement and omission of recent births, both mooted to be driven by survey interviewers seeking to reduce interview length and workload (Schoumaker 2014).

Full birth histories are collected within DHS surveys for 15 years preceding the survey. An extended set of questions is asked of the respondent for births occurring in the five years preceding the survey, but only an abbreviated question set for births thereafter. Consequently, survey interviewers are incentivised to age children beyond the five year threshold and ask only the abbreviated question set – "birth displacement". This presents as an underenumeration of births five years preceding the survey, and an excess of births six years preceding the survey. In addition, recent births can be entirely omitted from survey responses by interviewers, particularly in cases where birth displacement would lead to biologically implausible birth intervals. The simultaneous consideration of multiple surveys whose recall periods overlap enable these biases to be adjusted for. Data may exist for a given year that is two years preceding an older survey, exposed to both birth displacement and birth omission biases, and eight years preceding a more recent survey, exposed to neither. Previous analyses (Schoumaker 2014) show that these biases acting in concert can have profound effects, with relative fertility differing by up to 20% either side of the five year threshold.

We adjust for these biases with a dummy variable for before and after the threshold year - $TIPS_d$, and applying a first order random walk to smooth over the coefficients at each year preceding the survey - ω_{TIPS}. Thus we construct an observation model that builds on the linear predictor for λ above to adjust for Time Preceding Survey (TIPS) bias in our observations and predict births in a given age group, district, and year - $\hat{b}_{ait}$.

$$log(\hat{b}_{ait}) = log(\lambda_{ait} \times E_{ait}) + \beta_1 TIPS_d + \omega_{TIPS}$$

$$TIPS_d = \begin{cases} 0, & \text{if TIPS} < 5 \\ 1, & \text{otherwise} \end{cases}$$

Priors and hyperparameters

We specify diffuse normal priors on the model fixed effects ($\mu, \beta_1 \sim N(0,5)$). The BYM2 model, δ_i, can be decomposed into structured (ICAR) and unstructured (iid) components, $\delta_i = \phi_i + \theta_i = \sigma_i(\sqrt{\rho}\phi_i + (1-\sqrt{\rho})\theta_i)$, where ϕ_i is the ICAR model with scaled adjacency matrix, θ_i is the unstructured iid effect with prior $\theta \sim N(0,1)$, and ρ is the apportioning paramater to distribute variance between the structured and unstructured terms with prior $\rho \sim Beta(0.5, 0.5)$, as recommended in (Riebler et al. 2016).

Each of the smoothing random effects over age, time, and TIPS have variances parameters which are estimated from the data - $\alpha_a \sim RW1(\sigma_\alpha^2)$; $\gamma_t \sim RW2(\sigma_\gamma^2)$; and $\omega_{TIPS} \sim RW1(\sigma_\omega^2)$. Gamma priors are specified on the precisions: $\sigma_\alpha^{-2}, \sigma_\gamma^{-2}, \sigma_\omega^{-2} \sim \Gamma(1, 0.0005)$

Computation

Survey datasets extracted with the R package *rdhs* (Watson, FitzJohn, and Eaton 2019) and space-age-time stratified number of births and person years were calculated with the package *demogsurv* (Eaton n.d.). The statistical model was fit separately to each country using Template Model Builder in C++ through the *tmb* R package (Kristensen et al. 2016), and 1000 posterior samples taken. WorldPop pixel-level populations, published for quinquennial periods, were overlaid with district-resolution national shapefiles to calculate district populations, and linearly interpolated to single year estimates. UN Population Division World Population Prospects 2019 national popualtions were interpolated from quinquennial periods to single year, and district populations were then calibrated to interpolated national totals by both age and sex. These district populations were then used to aggregate district fertility rates to higher administrative levels.

Bibliography

Eaton, Jeffrey W. n.d. "mrc-ide/demogsurv: Analysis of demographic indicators from Demographic and Health Surveys (DHS) and other household surveys." Accessed March 31, 2020. https://github.com/mrc-ide/demogsurv.

Knorr-Held, Leonhard. 2000. "Bayesian modelling of inseparable space-time variation in disease risk." *Statistics in Medicine* 19 (17-18): 2555–67. https://doi.org/doi:10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-#.

Kristensen, Kasper, Anders Nielsen, Casper W. Berg, Hans Skaug, and Bradley M. Bell. 2016. "TMB: Automatic differentiation and laplace approximation." *Journal of Statistical Software* 70 (1). American Statistical Association: 1–21. https://doi.org/10.18637/jss.v070.i05.

Riebler, Andrea, Sigrunn H. Sørbye, Daniel Simpson, Håvard Rue, Andrew B. Lawson, Duncan Lee, and Ying MacNab. 2016. "An intuitive Bayesian spatial model for disease mapping that accounts for scaling." In *Statistical Methods in Medical Research*, 25:1145–65. 4. SAGE Publications Ltd. https://doi.org/10.1177/0962280216660421.

Schoumaker, Bruno. 2014. "Quality and consistency of DHS fertility estimates."

Watson, Oliver J., Rich FitzJohn, and Jeffrey W. Eaton. 2019. "rdhs: an R package to interact with The Demographic and Health Surveys (DHS) Program datasets." *Wellcome Open Research* 4 (June). F1000 Research Ltd: 103. https://doi.org/10.12688/wellcomeopenres.15311.1.