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

Background

Fertility estimation is core to understanding evolving demographic trends and population structures in sub-Saharan Africa. In high-income settings, long-standing vital registration systems provide direct measures of fertility, but these are often underdeveloped or absent in low and middle income countries. Instead, the reconstruction of fertility trends from nationally representative household survey- and census-based surveillance sources in these settings is well established, by which national-level estimates are published by UN Population Division as part of the World Population Prospects (UN Population Division n.d.) and by the Institute of Health Metrics and Evalation as part of the Global Burden of Disease study (Murray, Callender, and Kulikoff 2018). National population estimates underpinned by these reconstructed trends, in concert with estimates of mortality and migration, are utilised throughout domestic policymaking.

While quinquennial national-level estimates of total fertility from the World Population Prospects, widely used as gold standard estimates, are desirable for policymaking with long projection horizons, they are less well suited for use in public health target setting and monitoring which require higher resolution estimates, both temporally and spatially. Accordingly, public health 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 - permitting targeted and efficient resource allocation. Estimates of total fertility rate 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 ("Avenir Health" 2020), using fixed national-level age-specific fertility rates (UNAIDS Reference Group on Estimates Modelling and Projections 2019). The WorldPop project produce 1km x 1km pixel-level estimates of births, using survey-derived age-specific fertility rates at the first administrative level applied to pixel-level age-specific populations (Tatem et al. 2014).

The application of fertility estimates from higher administrative levels to all nested districts is undesirable as it overstates homogeneity which, due to socioeconomic and cultural differences, can be substantial between neighbouring districts. This is problematic for policy making at granular scales, for example programmatic planning for need of prevention of mother-to-child HIV transmission at the district level. Obtaining direct district-level measures of fertility by single calendar year from surveys is, however, challenging: nationally representative household surveys are typically powered to produce credible estimates no lower than the first administrative level, and age-disaggregated strata at the district level have few births, with estimates suffering from stochastic noise and large associated uncertainty. Instead, small-area model-based approaches, leveraging spatiotemporal correlation and smoothing to share information between data-sparse areal units, have been used to estimate demographic indicators at the district level, particularly those of under-5 mortality (Wakefield et al. 2019; Dwyer-Lindgren et al. 2018; Z. Li et al. 2019), and to produce estimates pertaining to sexual and reproductive health, including family planning need and contraceptive prevalence (Mercer, Lu, and Proctor 2019; Q. Li et al. 2019).

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

Methods

Data extraction and preparation

Data from X nationally representative household surveys conducted in X 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 the first administrative level. The modelled administrative level and administrative area hierarchy for each of the X 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 in a given country, $x \in 1, 2, 3, ..., X$, 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:

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

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

$$log(\lambda_{xait}) = \mu + \alpha_a + \gamma_t + \delta_i + \theta_x + \eta_{1,x,a,t} + \eta_{2,a,i} + \eta_{3,i,t} + \eta_{4,x,t} + \eta_{5,x,a}$$

where μ is the model intercept, α_a is a structured first order random walk on age group - $\alpha_a \sim RW1(\sigma_\alpha)$, γ_t is a structured second order random walk on calendar year - $\gamma_t \sim RW2(\sigma_\gamma)$, δ_i is an ICAR spatial model - $\delta_i \sim ICAR(\sigma_\delta)$, and θ_x is an i.i.d effect by country - $\theta_x \sim N(0, \sigma_\theta)$.

Additional flexibility is provided through five Type IV interaction terms (Knorr-Held 2000) - between age, time, and country (1), age and district (2), district and time (3), country and time (4), and country and age (5):

$$\eta_{1,x,a,t} \sim i.i.d \otimes AR1 \otimes AR1(\sigma^2_{\eta_1}, \rho_{1,a}, \rho_{1,t}) \tag{1}$$

$$\eta_{2,i,a} \sim AR1 \otimes ICAR(\sigma^2_{\eta_2}, \rho_{2,a}) \tag{2}$$

$$\sum_{i=1}^{I} \eta_{2,i,a} = 0 \quad a \in (15-19, 20-24, ..., 45-49)$$

$$\eta_{3,i,t} \sim ICAR \otimes AR1(\sigma^2_{\eta_3}, \rho_{3,t}) \tag{3}$$

$$\sum_{i=1}^{I} \eta_{2,i,t} = 0 \quad t \in 1995 : 2018$$

$$\eta_{4,x,t} \sim i.i.d \otimes AR1(\sigma^2_{\eta_4}, \rho_{4,t}) \tag{4}$$

$$\eta_{5,x,a} \sim i.i.d \otimes AR1(\sigma^2_{\eta_5}, \rho_{5,a}) \tag{5}$$

Adjusting for Time Preceding Survey bias in Demographic Health Survey data

Several sources of non-sampling bias 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. Overlapping recall periods between multiple surveys permit the magnitude of theses biases to be estimated and 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 fixed effect 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 from DHS surveys and predict births in a given age group, district, and year - $\hat{b}_{ait}$.

An extended set of questions is also asked in MICS surveys for births within two years of the survey, and an abbreviated set thereafter. Reflecting the difference in survey design, the observation model for MICS data includes a MICS-specific first order random walk applied over TIPS, but no MICS-specific TIPS fixed effect is used as no evidence for birth displacement in MICS surveys was found in this analysis.

DHS data:

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

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

MICS data:

$$log(\hat{b}_{xait}) = log(\lambda_{xait} \times E_{xait}) + \omega_{MICS-TIPS}$$

Priors and hyperparameters

Diffuse normal priors are specified on the model fixed effects - $\mu, \beta_1 \sim N(0, 5)$.

The random effects over age, time, district, country, and TIPS have variances parameters which are estimated from the data as specified above - $\alpha_a \sim RW1(\sigma_\alpha^2)$; $\gamma_t \sim RW2(\sigma_\gamma^2)$; $\delta_i \sim RW2(\sigma_\delta^2)$; and $\omega_{DHS-TIPS}, \omega_{MICS-TIPS} \sim RW1(\sigma_\omega^2)$. Gamma priors are specified on all precisions: $\sigma_\alpha^{-2}, \sigma_\gamma^{-2}, \sigma_\omega^{-2}, \sigma_{\eta_{1:5}}^{-2} \sim \Gamma(1, 2E5)$

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 to all countries simultaneously 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

"Avenir Health." 2020. https://www.avenirhealth.org/software-spectrum.php.

Dwyer-Lindgren, Laura, Ellen R. Squires, Stephanie Teeple, Gloria Ikilezi, D. Allen Roberts, Danny V. Colombara, Sarah Katherine Allen, et al. 2018. "Small area estimation of under-5 mortality in Bangladesh, Cameroon, Chad, Mozambique, Uganda, and Zambia using spatially misaligned data." *Population Health Metrics* 16 (1): 13. https://doi.org/10.1186/s12963-018-0171-7.

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%3C2555::AID-SIM587%3E3.0.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): 1–21. https://doi.org/10.18637/jss.v070.i05.

Li, Qingfeng, Thomas A. Louis, Li Liu, Chenguang Wang, and Amy O. Tsui. 2019. "Subnational estimation of modern contraceptive prevalence in five sub-Saharan African countries: a Bayesian hierarchical approach." *BMC Public Health* 19 (1): 216. https://doi.org/10.1186/s12889-019-6545-3.

Li, Zehang, Yuan Hsiao, Jessica Godwin, Bryan D Martin, Jon Wakefield, Samuel J Clark, and with support from the United Nations Inter-agency Group for Child Mortality Estimation and its technical advisory with support from the United Nations Inter-agency Group for Child Mortality Estimation and its technical advisory group. 2019. "Changes in the spatial distribution of the under-five mortality rate: Small-area analysis of 122 DHS surveys in 262 subregions of 35 countries in Africa." *PloS One* 14 (1): e0210645. https://doi.org/10.1371/journal.pone.0210645.

Mercer, Laina D., Fred Lu, and Joshua L. Proctor. 2019. "Sub-national levels and trends in contraceptive prevalence, unmet need, and demand for family planning in Nigeria with survey uncertainty." *BMC Public Health* 19 (1): 1–9. https://doi.org/10.1186/s12889-019-8043-z.

Murray, Christopher J. L., Charlton S. K. H. Callender, and Xie Rachel Kulikoff. 2018. "Population and fertility by age and sex for 195 countries and territories, 1950–2017: a systematic analysis for the Global Burden of Disease Study 2017." *The Lancet* 392 (10159): 1995–2051. https://doi.org/10.1016/S0140-6736(18)32278-5.

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

Tatem, Andrew J, James Campbell, Maria Guerra-Arias, Luc de Bernis, Allisyn Moran, and Zoë Matthews. 2014. "Mapping for maternal and newborn health: the distributions of women of childbearing age, pregnancies and births." *International Journal of Health Geographics* 13 (1): 2. https://doi.org/10.1186/1476-072X-13-2.

UNAIDS Reference Group on Estimates Modelling and Projections. 2019. "Methods and assumption for subnational demographic inputs to Spectrum in sub-Saharan Africa."

UN Population Division. n.d. "World Population Prospects - Population Division - United Nations." Accessed April 3, 2020. https://population.un.org/wpp/.

Wakefield, Jon, Geir Arne Fuglstad, Andrea Riebler, Jessica Godwin, Katie Wilson, and Samuel J. Clark. 2019. "Estimating under-five mortality in space and time in a developing world context." *Statistical Methods in Medical Research* 28 (9): 2614–34. https://doi.org/10.1177/0962280218767988.

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): 103. https://doi.org/10.12688/wellcomeopenres.15311.1.