Estimating subnational populations of women of reproductive age in developing countries

Monica Alexander[1] Leontine Alkema[2]

[1]University of California, Berkeley

[2]University of Massachusetts, Amherst

Institute of Disease Modeling Symposium 2018

Introduction

Background

- Important to monitor demographic and health indicators at the subnational level
- Need population counts
  - of particular interest is women of reproductive age (15-49)
  - measures exposure to risk
  - denominator for maternal mortality, fertility rates, contraceptive prevalence..
- However, in many developing countries, population data are limited and/or of poor quality
- Develop statistical methods to overcome missing data and quality issues

Approach

Patrick Leslie

Thomas Bayes

- Use a demographic model for estimation and projections to avoid implausible outcomes
- Estimation in a Bayesian framework to facilitate inclusion of various data sources and pooling of information across populations

Overview of project

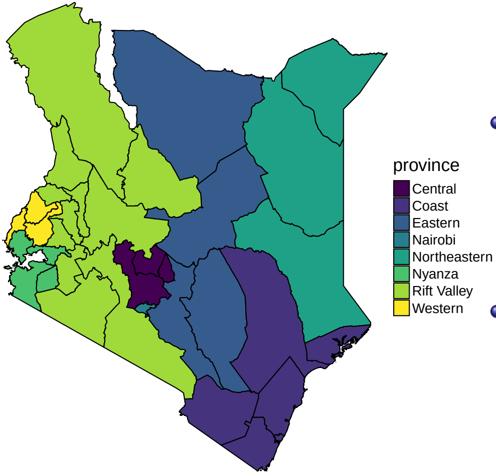
We are developing a Bayesian demographic modeling approach to estimate and project

- female populations aged 15-49 (WRA)
- from $\sim$ 1980 to 2020
- at the subnational level (admin2, i.e. county)
- using data sources that are commonly available in developing countries.

The model is applied to Kenya to estimate and project WRA populations from 1979-2020.

Data

Administrative levels in Kenya



- Nationally, WRA increased from 3.4 million (1980) to 9.4 million (2009, last census year), projected to 13.2 million (2020)
- 8 provinces, 36 districts

Data sources used

Census:

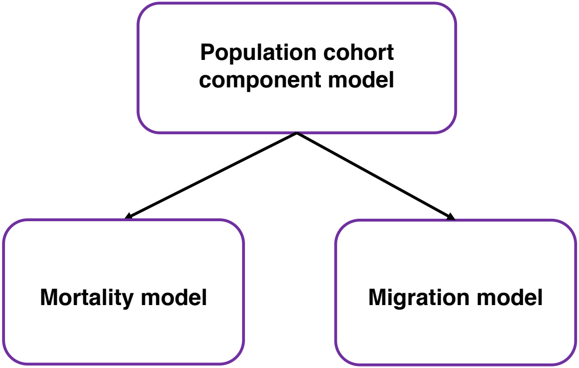
- 10% microsamples are available for years 1979, 1989, 1999, 2009 (IPUMS)
- Subnational population counts (adjusted for age heaping)
- Data on internal migration in year before the census

National estimates from World Population Prospects (WPP):

- National estimates of population counts
- National estimates of mortality rates

Model

Model overview



Cohort component projection

Use a cohort component reconstruction/projection model:

$$\eta_{r,c,a+1} = \eta_{r,c,a} \cdot (1 - \rho_{r,c,a}) + \phi_{r,c,a},$$

with for region r and birth cohort c:

- $\eta_{r,c,a}$ the number of women of age $a = 15, 16, \ldots, 49$,
- $\rho_{r,c,a}$ the expected proportion of women who die between age a and $a + 1$
- $\phi_{r,c,a}$ the expected number of net-migrants into/out of region r at age a

Cohort component projection

$$\eta_{r,c,a+1} = \eta_{r,c,a} \cdot (1 - \rho_{r,c,a}) + \phi_{r,c,a}$$

- Observed census counts

$$y_{r,c,a} \sim N(\eta_{r,c,a}, s^2_{r,c,a}),$$

  where sampling error s is due to 10% microsample.
- We constrain the sum of subnational populations to add up to national WPP estimates within lower and upper bounds (which is approximately 10%)

Mortality

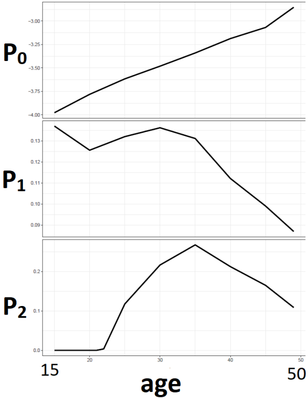
$$\eta_{r,c,a+1} = \eta_{r,c,a} \cdot (1 - \rho_{r,c,a}) + \phi_{r,c,a}$$

Modeling the probability of death ρ by region, cohort and age:

Obtain age-specific 'principal component' vectors:

- mean mortality P_0
- mortality decline P_1
- impact of the AIDS epidemic P_2

using a principal component analysis of logit(national ρ's).



Mortality

Use principal components as the basis of a regression model for $\rho_{r,c,a}$

$$logit(\rho_{r,c,a}) = P_{0,a} + \beta_{r,c,1} \cdot P_{1,a} + \beta_{r,c,2} \cdot P_{2,a}$$

Many different shapes of $\rho_{r,c,a}$ can be represented by different combinations of the P's.

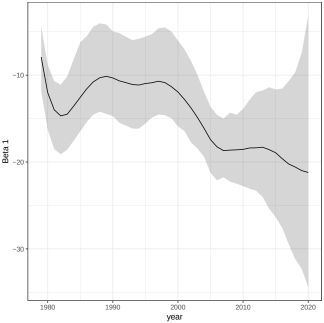
Mortality

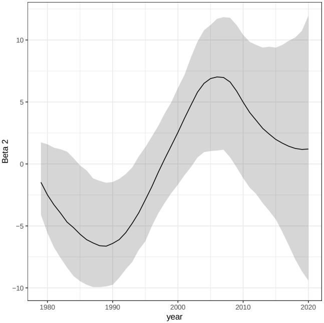
$$logit(\rho_{r,c,a}) = P_{0,a} + \beta_{r,c,1} \cdot P_{1,a} + \beta_{r,c,2} \cdot P_{2,a}$$

Use hierarchical time series model on the βs:

$$\beta_{r,c,1} = \mu_c + \delta_{r,c}$$

with across-regions cohort mean μ_c and "cohort-specific deviations within a region" $\delta_{r,c}$ (modeled by an AR(1) process).

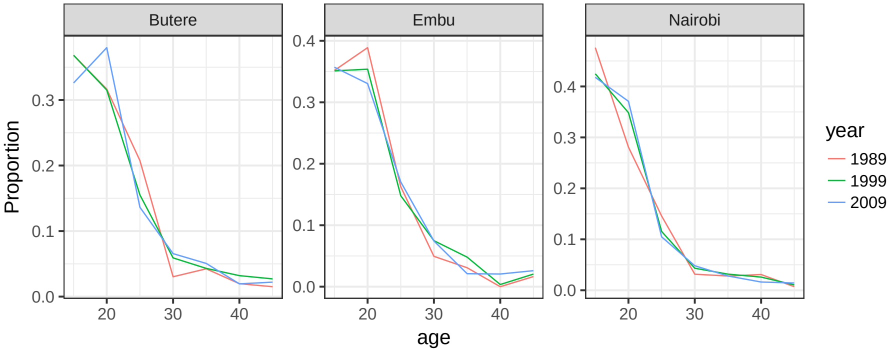




Net migration

$$\eta_{r,c,a+1} = \eta_{r,c,a} \cdot (1 - \rho_{r,c,a}) + \phi_{r,c,a}$$

Data suggest constant age patterns in net-migration:



Net migration

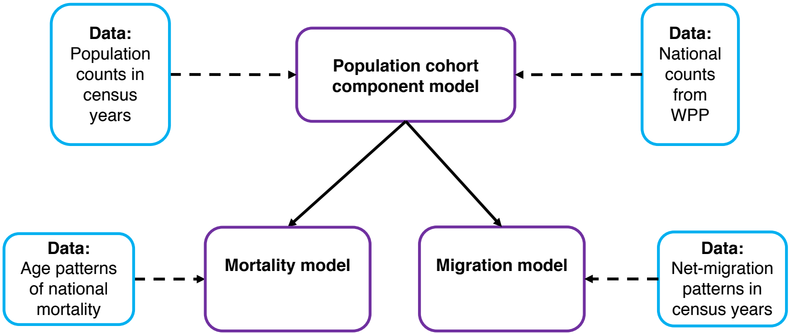
We assume for net-migration numbers ϕ:

$$\phi_{r,c,a} = \eta_{r,c} \cdot x_{r,a} \cdot \pi_{r,c},$$

where

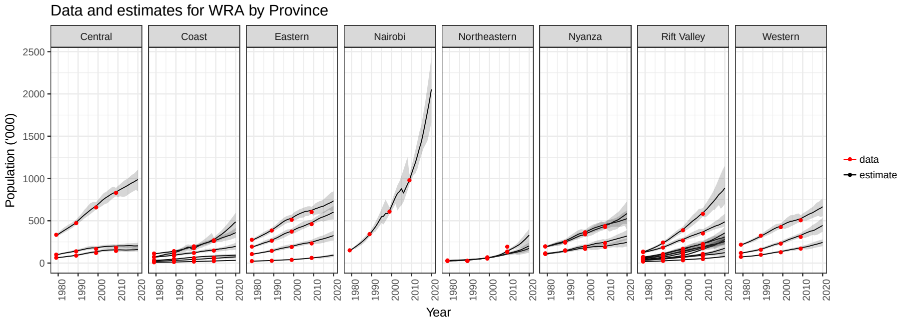
- $\eta_{r,c} = \sum_a \eta_{r,c,a}$ the total population of WRA,
- $x_{r,a}$ = proportion of net-migration at age a, age pattern assumed constant across cohorts, obtained from census data,
- $\pi_{r,c}$ = proportion of WRA that is net-migration:
  - Modeled with a random walk, $\pi_{r,c} \sim N(\pi_{r,c-1}, \sigma_\pi^2) T(-0.2, 0.2)$,
  - Informed by data from census, $p_{r,c} \sim N(\pi_{r,c}, \sigma_p^2)$.

Model overview



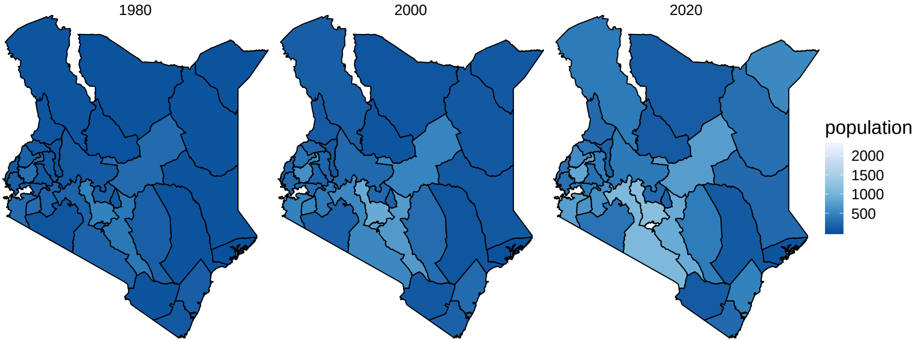
Results

Population by district

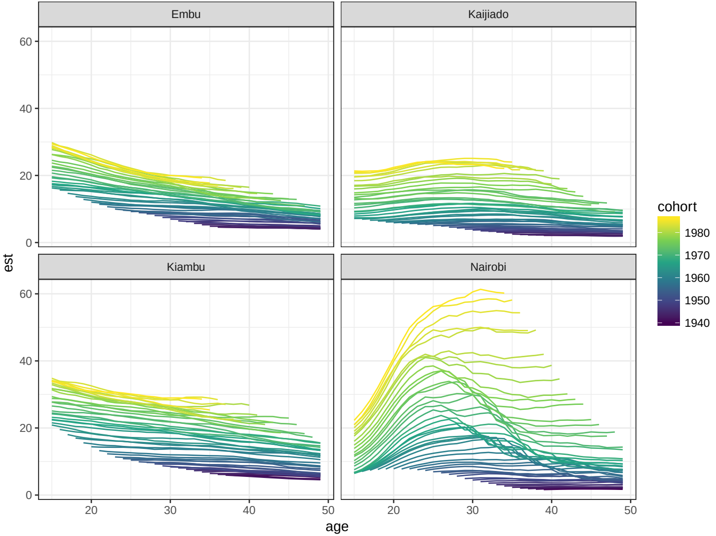


Population by district

Population (· 1,000)



Age patterns by cohort



Summary

Summary

- Cohort component projection model embedded with a Bayesian hierarchical framework
- Robust estimates of WRA populations at the subnational level
- Incorporating different data sources that are widely available
- Future work will focus on investigating how to use other data sources
    - e.g. sibling death information; census household deaths

Thanks!

monicaalexander.com

: MJAlexander

: @monjalexander

UC Berkeley Demography