Bayesian Space-Time Estimation of Under-five Mortality Rates

Jessica Godwin, Statistics & Jon Wakefield, Statistics and Biostatistics

University of Washington

jlg0003@uw.edu

CENTER FOR STUDIES IN DEMOGRAPHY & ECOLOGY

Abstract

Estimation of childhood mortality (U5MR) is difficult in countries without vital registration. DHS surveys provide complete birth histories which can be used. Mercer et al. developed a Bayesian method to smooth survey-based small area estimates of U5MR at administrative level 1 (Admin 1) areas. However, policy and intervention are often implemented at administrative level 2 areas (Admin 2). The finer geographic partition presents many problems in implementing this methodology. Most DHS are geographically stratified at Admin 1, and not Admin 2, leading to smaller sample sizes in most areas and some small areas with no observed deaths. Many earlier DHS do not have GPS information for sampled clusters. This makes it impossible to assign those observations to Admin 2 areas, resulting in large data losses. Our work aims to address these problems.

Introduction

Under-five mortality (U5MR) is an important demographic indicator in its own right; it is also an important indicator of overall mortality. In countries without vital registration, survey data must be used to make estimates of U5MR. However, the countries where childhood mortality may be high, also tend to lack vital registration. The Demographic and Health Surveys (DHS) collect full birth histories from women aged 15-49 in 90 countries in the world. Mercer et al. (2015) develop a method to use retrospective birth histories from women to make estimates small area estimates of U5MR which account for complex survey design, and, also, smooth across space and time. However, we aim to extend this method to make small area estimates at a finer partition of small areas of a country.

Data

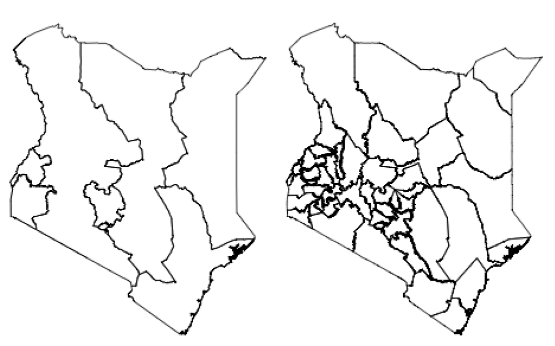

Figure 1: Admin 1 and 2 areas in Kenya.

Survey	Admin 1	GPS	Admin 2	N
1989	×			
1993	×			
1998	×			
2003	✓	✓		**16060**
2008	✓	✓		**17144**
2014	✓	✓	✓	**65772**

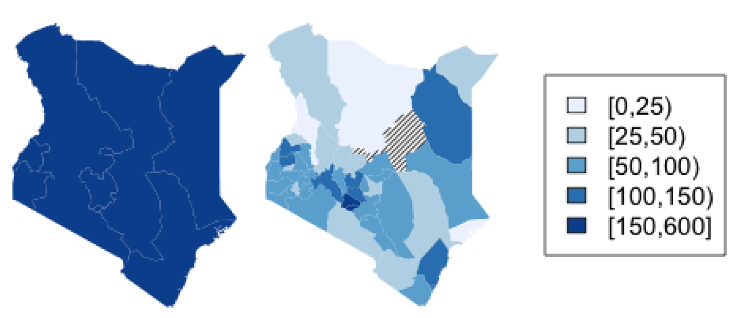

Figure 2: Sample size for KDHS 2003 in 1990-1994 in Admin 1 and Admin 2 areas.

- Admin 1: 8 regions, Admin 2: 47 districts
- KDHS 2003, 2008, 2014 ⇒ 1980-1984 to 2010-2014
- Surveys with GPS cluster locations can be assigned to Admin 2 areas
- Stratified two-stage cluster design, sampling weights
- Full birth histories from women 15-49
- Potential biases:
  - Forget children who were born and died a long time ago
  - Birth transference
  - Fail to mention child who recently died

Discrete-time Survival Analysis

Each child's observation in the full birth histories is converted into person months, i.e. a child contributes 1-59 person months to this stage of analysis. Use design-based regression estimators to get estimates and variances of q_a^m the probability of death in a person month of a child in age band $a \in \{$0-1mo, 1-11mo, 12-23mo, 24-35mo, 36-47mo, 48-59mo$\}$. Note n_a is the number of months in age band a.

$$\text{logit}(q_a^m) = \beta_a$$
$$\hat{q}_a = 1 - (1 - \hat{q}_a^m)^{n_a}$$
$$\widehat{{}_5q_0} = 1 - \prod_{a=1}^{6}(1 - \hat{q}_a)$$

Observed Zeroes

Sometimes there are no observed deaths in an Admin 2 area A_i at time period t. This never happens at the Admin 1 level. We do not trust this estimate, and it leads to unrealistically small standard errors. Every Admin 2 area lies in exactly one Admin 1 area, A_r. Let n_{it} be the number of children in the dataset in A_i at time t, $A_i \subset A_r$, $\widetilde{p_{rt}}$ be a smoothed estimate of U5MR in A_r at time t. We then add another child, who we assume died.

$$\widetilde{p_{rt}} = \widehat{p_{its}} = \frac{\sum_{k=1}^{n_i+1} w_k y_k}{\sum_{k=1}^{n_i+1} w_k}$$

We solve for $w^*_{n_i+1}$ and then use the Hajek estimator $\widehat{p_{its}}$ and its associated design-based standard error in our smoothing model.

Note: There are 987 Admin 2, time period, survey combinations. 287 have contribute no data, 51 are imputed.

Bayesian Hierarchical Model

Notation

- i: Admin 2 area
- t: five-year time period
- s: survey
- η_{its}: logit$({}_5q_{0its})$
- y_{its}: logit$(\widehat{{}_5q_{0its}})$
- $\widehat{V}_{DES,its}$: design-based variance of y_{its}
- μ: fixed intercept
- α_t, θ_i, ν_s, δ_{it}, ν_{is}, ν_{ts}: iid random effects
- γ_t: RW1 random effect
- ϕ_i ICAR random effect

Model

$$y_{its}|\eta_{its} \sim N(\eta_{its}, \widehat{V}_{DES,its})$$
$$\eta_{its} = \mu + \alpha_t + \theta_i + \nu_s + \gamma_t + \phi_i + \delta_{it} + \nu_{is} + \nu_{ts}$$

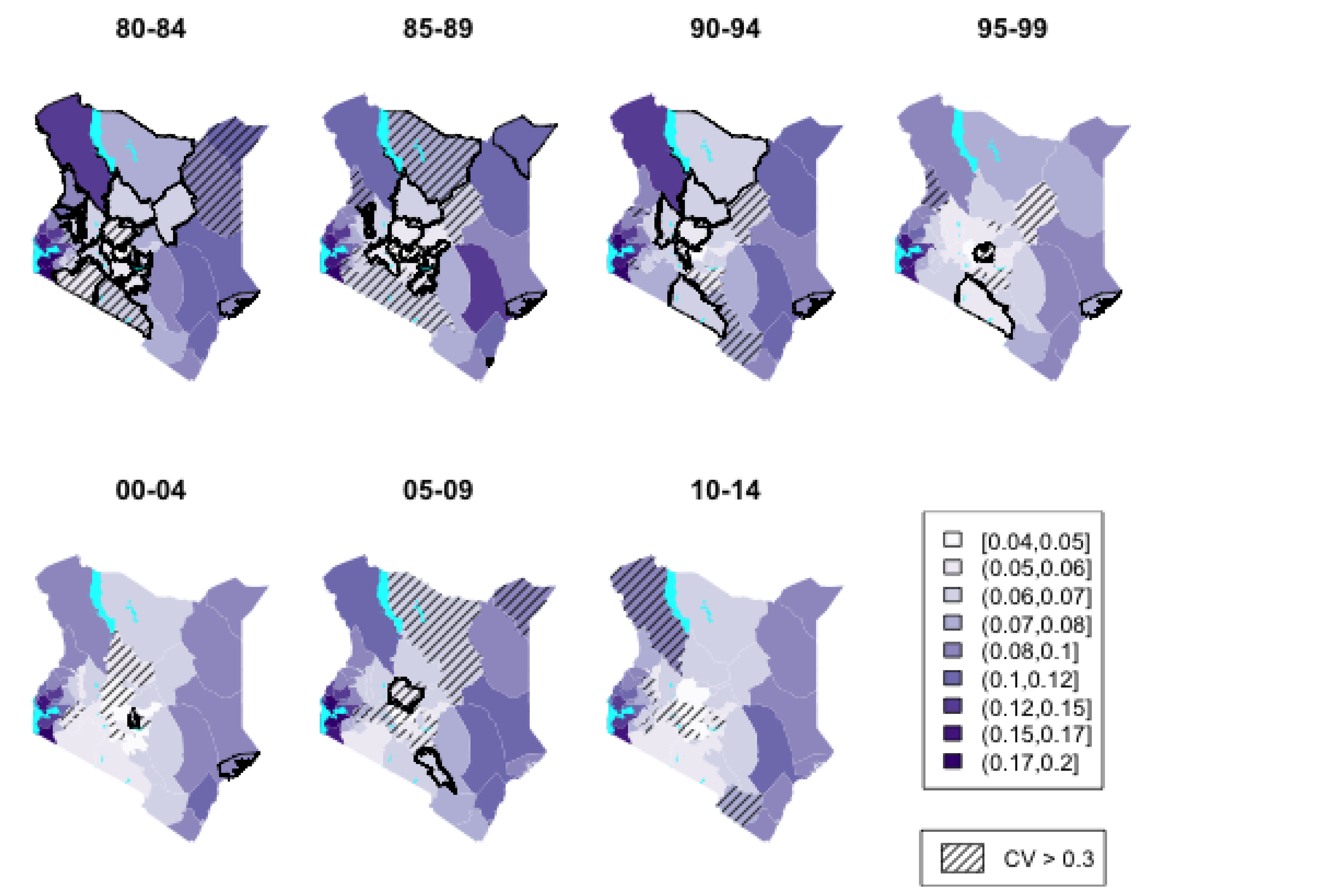

Figure 3: Smoothed Admin 2 estimates by time period. Black outline are adjusted areas, hatching indicates a higher level of uncertainty.

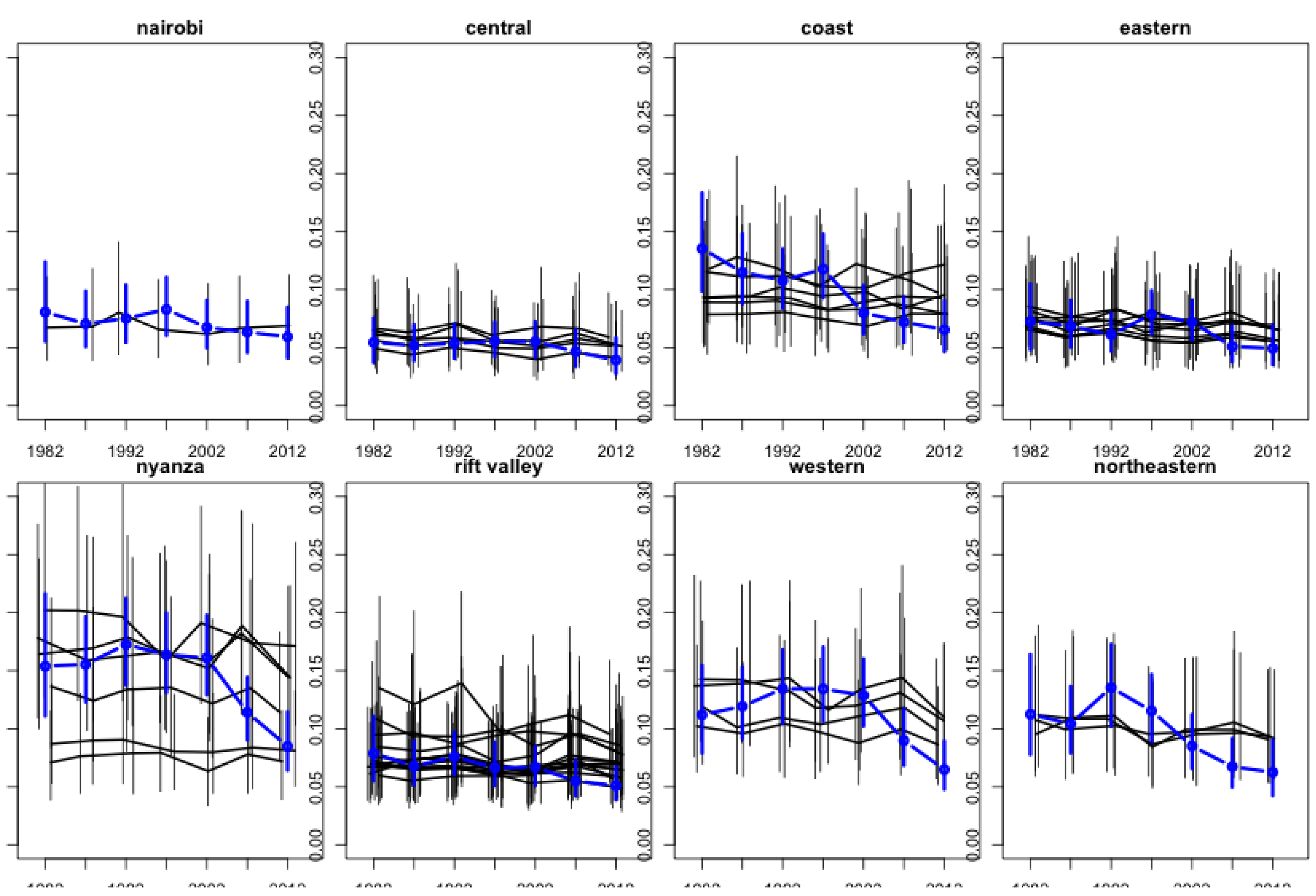

Figure 4: Comparison of point and interval estimates of Admin 1 areas (in blue), and the Admin 2 areas they contain (in black).

Conclusions

- Smoothed estimates at the Admin 2 area have much larger uncertainty than at the Admin 1 area.
- Can we add covariates at either modeling stage to improve precision?
  - Spatial: annual average rainfall
  - Individual: SES index
- Can we combine point and areal data to get more precise estimates?