

Small Area Estimation of Child Mortality in the Absence of Vital Registration

Laina D Mercer¹, Jon Wakefield¹², Athena Pantazis³,
Angelina M Lutambi⁴, Honorati Masanja⁴, and Samuel Clark³⁵⁶⁷⁸

Center for Statistics and the Social Sciences
Working Paper Series
Working Paper no. 148
University of Washington

November 14, 2014

¹Department of Statistics University of Washington, USA

²Department of Statistics University of Washington, USA

³Department of Sociology, University of Washington, USA

⁴Ifakara Health Institute, Dar es Salaam, TZA

⁵Institute of Behavioral Science (IBS), University of Colorado at Boulder, Boulder, USA

⁶MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, RSA

⁷INDEPTH Network, Accra, Ghana

⁸ALPHA Network, London, UK

Abstract

Many people living in low and middle-income countries are not covered by civil registration and vital statistics systems. Consequently a wide variety of other types of data including many household sample surveys are used to estimate health and population indicators. In this paper we combine data from sample surveys and demographic surveillance systems to produce small-area estimates of child mortality through time. Small area estimates are necessary to understand geographical heterogeneity in health indicators when full-coverage vital statistics are not available. For this endeavor spatio-temporal smoothing is beneficial to alleviate problems of data sparsity. Conventional hierarchical models are not immediately applicable since one must account for the survey weights in order to alleviate bias due to non-random sampling and non-response. The application that motivated this work is estimation of child mortality rates in five-year time intervals in regions of Tanzania. Data come from Demographic and Health Surveys (DHS) conducted over the period 1980–2010 and two demographic surveillance system sites. We derive a variance estimator that accounts for the complex survey weighting, and a simulation study examines the properties of our estimator, with a comparison to a jackknife alternative. For our application, the hierarchical models we consider include random effects for area, time and survey and we compare models using the conditional predictive ordinate (CPO). The method we propose is implemented via the fast and accurate integrated nested Laplace approximation (INLA).

KEY WORDS: Bayesian smoothing, Infant mortality, Small area estimation,
Survey sampling.

1 Introduction

Over the past fifteen years the United Nations' Millennium Development Goals (MDGs) (United Nations, 2000) have focused the world's attention on improving key indicators of development, health and wellbeing. The requirement to monitor progress toward the MDGs has revealed a stunning absence of data with which to measure and monitor key indicators related to the MDGs in much of the developing world, and this has led to great interest in improving both the data and our ability to use it. As 2015 approaches the United Nations (UN) and its partners are taking stock of experience with the MDGs and coordinating the establishment of a new set of global goals (United Nations, 2014d) – the Sustainable Development Goals (SDGs) (United Nations, 2014e). Even before the SDGs are finalized the UN Secretary General has called for a *Data Revolution for Sustainable Development* and appointed a high level advisory group to define what it should be (United Nations, 2014b). The aim is clear: to rapidly improve the coverage, quality, availability and timeliness of the data used to measure and monitor progress toward the SDGs. Simultaneously there is sustained, strong interest in improving civil registration, vital statistics (CRVS) and the functioning of statistical offices across the developing world (World Bank and World Health Organization, 2014; Paris21, 2014). The key challenges are improving coverage (United Nations, 2014a) and timeliness of reporting.

In this context of far-reaching interest in improving data and methods available to monitor indicators of the SDGs and improve CRVS, in this paper we develop a general approach that combines data from different sources and provides temporal, subnational-specific estimates with uncertainty that accounts for the different designs of the data collection schemes. We demonstrate the method by calculating spatio-temporal estimates of child mortality in Tanzania using data from multiple Demographic and Health Surveys (DHS) (USAID, 2014) and two demographic surveillance system (DSS) sites (INDEPTH Network, 2014).

Reducing child mortality is MDG 4 (United Nations, 2014c), and over the past fifteen years a great deal of effort and resources have been spent in order to meet MDG 4 targets at the national level in many developing nations. This has driven work to develop better methods to estimate trends in child mortality at the national level, and two groups have produced globally comparable trends in child mortality for all nations. The United Nations Inter-agency Group for Child Mortality Estimation (UN IGME) has used a loess-based smoothing approach, and the Institute for Health Metrics and Evaluation (IHME) does something similar using Gaussian process regression (Lozano *et al.*, 2011). Alkema and You (2012) compare and criticize both approaches and then develop an improved Bayesian B-spline Bias-reduction (B3) method (Alkema *et al.*, 2014). All of these methods produce national estimates through time with measures of uncertainty. None are designed to reveal variation in child mortality within countries, and both miss a potentially important source of uncertainty – design effects inherent in the surveys that often contribute the bulk of the data used to construct estimates.

In this paper we combine data from multiple surveys with different sampling designs, and construct subnational estimates through time with uncertainty that reflects the various data collection schemes. Data come from traditional cluster sample surveys (DHS) and two DSS sites. DSS sites intensively monitor everyone within a given area, typically to monitor the effects of health intervention trials of various types. Estimates of child mortality from both

sources of data are useful but potentially flawed in different ways. National cluster sample surveys are generally not able to produce useful subnational estimates, and DSS sites are not designed to be nationally representative, and are also thought to fall prey to the Hawthorne effect by which the communities of these sites have improved health outcomes because they are under observation and, more concretely, because of the trials being conducted.

We construct subnational estimates of Tanzanian child mortality through time with uncertainty intervals. This problem is challenging because in addition to requiring smoothing over space and time, we must also account for the survey design. When sampling is not simple and random and the design variables (upon which sampling was based) are not available, the complex sampling design is accounted for by constructing design weights. Inference is then carried out using design-based inference, e.g. using Horvitz-Thompson estimators (Lumley, 2010). In contrast, a conventional space-time random effects framework, for example, Knorr-Held (2000), is model-based, and requires an explicit likelihood to be specified. In this paper, we marry these two approaches by constructing a working likelihood based on the asymptotic distribution of a design-based estimator and then smooth using a space-time hierarchical prior.

The organization of this paper is as follows. In Section 2 we describe the two data sources upon which estimation will be based. In Section 3 the calculation of child mortality estimates with an appropriate standard error is described using discrete time survival models. In Section 4 we carry out a simulation study to assess the performance of the asymptotic distribution as compared to a currently used method that is based on the jackknife. Hierarchical Bayesian space-time models are introduced in Section 5. The results of our modeling efforts of under five mortality rates (U5MR) within Tanzania from 1980–2010 are given in Section 6 and discussed in Section 7.

2 Data Sources

We focus on child mortality using data from the five Tanzanian Demographic and Health Surveys (TDHS), one Tanzania HIV and Malaria Indicator survey (THMIS), and two health and demographic surveillance system (HDSS) sites in Tanzania, Ifakara and Rufiji. Over the period 1980–2010 estimates of child mortality from the two types of data sources (surveys, surveillance sites) are generally similar but, as described above, different in useful ways. The HDSS estimates are accurate (low bias) and precise (small variance) measurements for comparatively small, geographically-defined populations, and the household survey estimates are less accurate and much less precise but representative of large populations.

2.1 Health and Demographic Surveillance System

The Ifakara Health Institute (IHI), Tanzania runs a number of health and population research projects including two HDSS sites – Ifakara and Rufiji. We collaborated with IHI to estimate child mortality using data from the Ifakara and Rufiji HDSS sites.

The HDSS data are generated through repeated household visits. For the data we use, each household was visited three times per year at regular intervals. During each visit a ‘household roster’ was updated and all new vital and migration events for all members of the

household were recorded. In addition, potentially many other questions were asked as part of both routine and ‘add-on’ studies. For our purposes we require only the basic core HDSS data that include information on dates of birth, death and migration – the information necessary to accurately identify observed person time, categorize that time by calendar period and age, and identify the outcome of interest, death. Because both HDSS sites are long-lived surveillance projects, there are many repeated observations on households and individuals, and all of these must be linked in order to conduct survival analysis.

2.2 Household Surveys

Full TDHS surveys that collected data necessary for child mortality estimates were conducted in Tanzania in 2010, 2004–05, 1999, 1996, and 1991–92, in addition to the THMIS that included child mortality which was conducted in 2007–08. The 2010 TDHS, 2007–08 THMIS and 2004–05 TDHS surveys used 2-stage cluster samples. First, clusters were sampled from enumeration areas from the 2002 Tanzania census and second, a systematic sampling of households within each cluster was carried out. The 1999 TDHS, 1996 TDHS and 1991–92 TDHS used a 3-stage cluster design, first selecting wards and branches using the 1988 Tanzania Census as a sampling frame, second using probability proportional to size sampling to select enumeration areas from each selected ward or branch, and third selecting households from a new list of all households in each selected enumeration area. The same first and second stage units were used for all three of the surveys. Stratification by urban/rural and region was done at the first stage, with oversampling of Dar es Salaam, other urban areas, and Zanzibar. The surveys were designed to be nationally representative and to be able to provide estimates of contraceptive prevalence at the regional level.

All women age 15 to 49 who slept in the household the night before were interviewed in each selected household and response rates were high (above 95% for households in all surveys). TDHS provides sampling (design) weights, assigned to each individual in the dataset. Limited information is provided for each survey concerning the calculation of survey weights, but the general explanation indicates that raw survey weights are the inverse of the product of the 2–3 probabilities of selection from each stage. These raw weights were then adjusted to reflect household response and individual response rates. The 1991–92 Tanzania DHS final report Demographic and Health Surveys (1992) indicates that “final individual weights were calculated by normalizing them for each area so that the total number of weighted cases equals the total number of unweighted cases”, but this normalization is not discussed in later reports (Demographic and Health Surveys, 1997, 2000, 2005, 2010) or the DHS statistics manual (Rutstein and Rojas, 2006). For the purposes of our analysis of child mortality, children identified by the women who were interviewed contributed exposure time and deaths. The data were organized into child-months from birth to either death or date of the mother’s interview.

3 Calculating child mortality with Discrete Time Survival Models

3.1 Estimation

We modeled child mortality using discrete time survival analysis (DTSA) (Allison, 1984; Jenkins, 1995). DTSA does not require an assumption of proportional hazards and allows greater flexibility in modeling time-varying covariates and their interactions. An important aspect of DTSA is that it can incorporate the very different hazard rates within the first 5 years of life. As our main aim is to examine the change in risk as a function of age and historical period, we prefer the DTSA approach that allows us to easily and transparently specify potentially complex temporal relationships without requiring proportionality of risks. The resulting predicted probabilities are easy to interpret and can be used directly in traditional mortality analysis methods such as life tables, in our case to calculate U5MR. Data were organized as child-months where each child was at risk from birth during each month observed up to and including the month of their death. We use logistic regression to estimate the monthly probability of dying conditional on the state of the child at the beginning of the month.

For child-months lived by a generic child our DTSA models are of the general form:

$$\text{logit } p(\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$$

where $p(\mathbf{x})$ is the probability of dying for a child with characteristics (for example, sex, age, time period, region) \mathbf{x} . A more detailed discussion of DTSA method can be found in Clark *et al.* (2013).

We wish to estimate the U5MR denoted by ${}_5q_0$. We define ${}_nq_x = \Pr(\text{dying before } x + n \mid \text{lived until } x)$ and the discrete hazards model splits the $[0,5)$ period into $J + 1$ intervals $[x_0, x_1), [x_1, x_2), \dots, [x_J, x_{J+1})$, where $x_{j+1} = x_j + n_j$ so that n_j is the length of the interval beginning at x_j , $j = 0, \dots, J$. The probability of death in $[x_j, x_{j+1})$, given survival until x_j is ${}_n_j q_{x_j}$. Then U5MR is calculated as

$${}_5q_0 = 1 - \prod_{j=0}^J (1 - {}_{n_j} q_{x_j}). \quad (3.1)$$

For our purposes, ${}_5q_0$ is calculated by dividing the first 60 months into six intervals ($J = 5$), $[0, 1), [1, 12), [12, 24), [24, 36), [36, 48), [48, 60)$. The observed data consist of, for each birth, a binary sequence up to length 60 with 0/1 corresponding to survival/death. The monthly probability of death for each interval, $p_j = \Pr(\text{dying in any month in the interval } x_j + n_j \mid \text{lived until } x_j)$, may be estimated using a logistic generalized linear model (GLM):

$$\text{logit } (p_j) = \sum_{j=0}^J \beta_j I_j,$$

where I_j is the indicator for the $[x_j, x_{j+1})$ time interval. In the complex survey context, that is relevant for the Tanzania example, an important consideration is that the design

weights must be acknowledged. This is achieved by solving a (design) weighted score statistic (Binder, 1983). Once estimates $\widehat{\beta}_j$ are estimated we can calculate $\widehat{p}_j = \exp(\widehat{\beta}_j)/[1+\exp(\widehat{\beta}_j)]$. The complement of surviving each month of the interval $[x_j, x_j + n_j]$ is used to calculate ${}_{n_j} \widehat{q}_{x_j} = 1 - (1 - \widehat{p}_j)^{n_j}$ which may be substituted into (A.1) to give ${}_5 \widehat{q}_0$.

In Section 5 we will construct, for a generic U5MR, a working likelihood based on the asymptotic distribution

$$y = \text{logit}({}_5 \widehat{q}_0) \sim N(\eta, \widehat{V}_{\text{DES}})$$

where $\eta = \log[{}_5 q_0 / (1 - {}_5 q_0)]$ and \widehat{V}_{DES} is the asymptotic (design-based) variance estimate of $\text{logit}({}_{n_j} \widehat{q}_{x_j})$, which is obtained via the delta method, the Supplementary Materials contain details of this calculation.

4 Simulation to test coverage performance of derived SE

Pedersen and Liu (2012) discuss the difficulties associated with deriving a variance estimate in the context of child mortality estimates. The TDHS typically uses a jackknife estimator, $V_{\text{JACK}}({}_5 \widehat{q}_0)$ of ${}_5 \widehat{q}_0$, which for cluster sampling is

$$\widehat{V}_{\text{JACK}} = \frac{1}{k} \frac{1}{k-1} \sum_{i=1}^k ({}_{5 \widehat{q}_{0i}} - {}_5 \widehat{q}_0)^2 \quad (4.1)$$

where k is the number of clusters, ${}_{5 \widehat{q}_{0i}} = k {}_5 \widehat{q}_0 - (k-1) {}_5 \widehat{q}_{0(i)}$ and ${}_{5 \widehat{q}_{0(i)}}$ is the estimate based on all of the data while holding out the i -th cluster. A 95% confidence interval for ${}_5 q_0$ is based on

$$\left[{}_5 \widehat{q}_0 - 1.96 \times \sqrt{\widehat{V}_{\text{JACK}}} , {}_5 \widehat{q}_0 + 1.96 \times \sqrt{\widehat{V}_{\text{JACK}}} \right]. \quad (4.2)$$

4.1 Data Generation

A simulated dataset was created to asses the coverage properties of interval based on V_{DES} and V_{JACK} . To simulate the estimation process within one region 100,000 women were assigned to 500 enumeration areas. The number of births for each woman were generated from a Poisson distribution with rate of 3. Each birth was assigned a calendar month between 0 and 239 (a 20 year period).

Death within the first 60 months of life were assigned based on a multinomial distribution with the following discrete hazards: 0.04 for the first month, 0.004 for months 2–12 and 0.001 for months 13–60. As we are only interested in death within the first 5 years, the remaining probability was assigned to a 61st category for death after the age of 5. The distribution of probability represents the highest risk in the first month of life, an elevated risk for the remainder of the first year, and a fairly low risk for the following 4 years. These probabilities result in an expected U5MR of 124.5 deaths per 1,000, which is similar to the late 1990s U5MR in Tanzania.

When creating estimates for a particular 5 year calendar period from the household survey and HDSS data births before or during this period contribute person-time from children who

are born or die or are censored in other 5 year time periods. To mimic this aspect of the real data, births were simulated from a wider time period and then subsetted to include only observations within the relevant time period. Children who survive past 60 months only contribute person time for months 0–59. This subset included 140,024 births from 75,385 mothers and the U5MR in this population was 131.0 per 1,000.

4.2 Simulation Procedure

We employed a two stage cluster sampling design. At the first stage n_E enumeration areas were randomly selected from the N_E available. At the second stage, suppose cluster i is selected, then n_{wi} women were randomly selected from the N_{wi} total women within the selected enumeration area. The resulting sampling weights for a mother selected in cluster i is

$$w_{Ei} = \frac{N_E}{n_E} \times \frac{N_{wi}}{n_{wi}}.$$

The number of enumeration areas was set at $N_E = 500$ for all simulations and the number selected at the first stage was one of $n_E = 10, 20, 30$. Sample sizes within enumeration areas varied by 5 within 10–50 and $n_w = 10, 15, \dots, 45, 50$. For each combination of n_E and n_w we draw 1,000 samples and the corresponding delta method and jackknife confidence intervals were created based on \hat{V}_{DES} and \hat{V}_{JACK} , respectively. The sample coverage of each interval type was calculated as the average number of intervals that contained the true population U5MR.

4.3 Results

The coverage of the delta method and jackknife intervals by number of clusters and within sample size cluster are shown in Figure 1. Results are much as one would expect from clustered sampling, coverage improves when there are more clusters and within a given number of clusters there is not much gain in increasing the sample size. Generally the performance of the delta method and jackknife intervals is very similar. The TDHS sampling scheme is generally around 20 clusters with a within sample size of approximately 20 women, which corresponds to the red line at a sample size of 20. This suggest that TDHS intervals may be slightly anti-conservative. The supplementary materials contain summaries of the interval widths, and again the two methods perform similarly. We prefer the delta method as it is generally applicable (i.e., to a variety of designs) and has a far smaller computational burden. We conclude that the asymptotic normal sampling distribution and the delta method variance result in sufficiently accurate confidence interval coverage for the cluster and sample sizes considered. Consequently, we will use the asymptotic distribution with the delta method variance as a working likelihood.

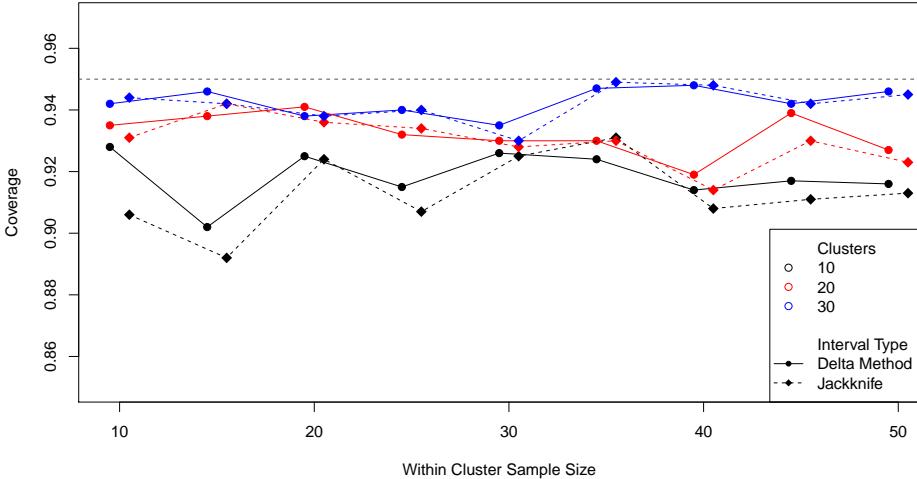


Figure 1: Coverage of jackknife and delta method 95% confidence intervals for a variety of numbers of clusters and cluster sizes. Sample size values are offset horizontally to improve the visibility.

5 Combining Data Sources in Hierarchical Bayesian Space-Time Model

5.1 The First Stage

Let ${}_5\hat{q}_{0its}$ represent the estimate of U5MR from survey s in region i and in period t . A model-based approach to inference with survey data may be carried out if the variables upon which sampling were based (the design variables) are available Gelman (2007). Unfortunately, these variables are not available for the Tanzania surveys. As an alternative we summarize the data in area i at time point t from survey s via the asymptotic distribution of the estimator of the empirical logit:

$$y_{its} = \log \left[\frac{{}_5\hat{q}_{0its}}{1 - {}_5\hat{q}_{0its}} \right].$$

We define the area, period and survey summary as $\eta_{its} = \log [{}_5q_{0its}/(1 - {}_5q_{0its})]$. We take as working likelihood the asymptotic distribution

$$y_{its} \mid \eta_{its} \sim N \left(\eta_{its}, \widehat{V}_{DES,its} \right) \quad (5.1)$$

which has been shown to perform well in the context of small area estimation from complex surveys (Mercer *et al.*, 2014).

5.2 Second Stage Smoothing Models

We wish to smooth over time period, region and surveys, but would like as parsimonious a model as possible, to avoid overfitting. At the second stage of our model we adopt a model

similar to the ‘type I’ inseparable space-time model of Knorr-Held (2000). However, unlike Knorr-Held (2000) our data provides multiple observations for each area i and time point t through the THMIS, five TDHS and two HDSS, denoted as surveys s . Thus we consider models that allow the option of survey-specific effects. The survey effects could be constant over time and space, could vary with time, vary with space, or vary by time and space. The six candidate models we consider are given in Table 1, with the caption containing the random effects specification. There are two temporal terms with α_t being independent and identically distributed random effects that pick up short-term fluctuations with no structure, and γ_t being given an (intrinsic) random walk prior of order 1, to pick up local temporal smooth fluctuations, for $t = 1, \dots, T = 6$ time periods. Similarly, there are two spatial terms, corresponding to the convolution model of Besag *et al.* (1991). The independent random effects are denoted θ_i and the intrinsic conditional autoregressive (ICAR) terms are ϕ_i for $i = 1, \dots, I = 21$ regions of Tanzania. The latter perform local geographical smoothing. The space-time interaction terms δ_{it} are taken to be independent, which corresponds to the Type I interaction model of Knorr-Held (2000). There are $S = 8$ different surveys that are carried out over the various time periods (since mothers are surveyed on their complete birth history and so report on births from previous time periods) and the five TDHS and THMIS surveys cover multiple regions over the different time periods they were administered. The independent random effects ν_s allow for these surveys to have a systematic displacement from the true logit of U5MR. The interactions ν_{ts} and ν_{is} allow these displacements to vary with period and space, respectively, while ν_{its} allow the complete interaction between survey, period and area. Model I contains crossed random effects only, since each area is represented in each of the time periods. Models II–VI contain a combination of nested and crossed random effects. The random walk and ICAR models are described in detail in Rue and Held (2005) and more details are given in the Supplementary Materials.

Table 1: Random effects models for time period t , region i and survey s . In all models μ is the intercept and $\alpha_t \sim_{iid} N(0, \sigma_\alpha^2)$, $\gamma_t \sim RW1(\sigma_\gamma^2)$, $\theta_i \sim_{iid} N(0, \sigma_\theta^2)$, $\phi_i \sim ICAR(\sigma_\phi^2)$, $\delta_{it} \sim_{iid} N(0, \sigma_\delta^2)$. Specific models contain random effects with distributions $\nu_s \sim_{iid} N(0, \sigma_{\nu_1}^2)$, $\nu_{is} \sim_{iid} N(0, \sigma_{\nu_2}^2)$, $\nu_{ts} \sim_{iid} N(0, \sigma_{\nu_3}^2)$, $\nu_{its} \sim_{iid} N(0, \sigma_{\nu_4}^2)$.

Model	Linear Predictor η_{its}
I	$\mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it}$
II	$\mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it} + \nu_s$
III	$\mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it} + \nu_s + \nu_{is}$
IV	$\mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it} + \nu_s + \nu_{ts}$
V	$\mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it} + \nu_s + \nu_{ts} + \nu_{is}$
VI	$\mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it} + \nu_s + \nu_{ts} + \nu_{is} + \nu_{its}$

5.3 Hyperpriors

For a generic set of independent random effects we specify priors on the precision τ such that a 95% prior interval for the residual odds ratios lies in the interval [0.5,2] which

leads to $\text{Gamma}(a_{\text{MARG}}, b_{\text{MARG}})$ priors for precisions (Wakefield, 2009) with $a_{\text{MARG}} = 0.5$, $b_{\text{MARG}} = 0.001488$. For both the RW1 and ICAR models the precisions have *conditional* rather than *marginal* interpretations. Let \boldsymbol{z} represent a random effect from a improper GMRF with “mean” $\mathbf{0}$ and “precision” $\tau^* \mathbf{Q}$. Following the supplementary materials of Fong *et al.* (2010), we gain compatibility by calculating an approximate measure of the average marginal “variance” of \boldsymbol{z} in the situation with $\tau^* = 1$; call this average c . Then to put on the same scale we take $a_{\text{COND}} = a_{\text{MARG}}$ and $b_{\text{COND}} = b_{\text{MARG}}/c$. In the above description, the words mean, precision, variance are written in italics to acknowledge that strictly speaking these quantities do not exist since the distribution is improper. However, one may calculate a generalized inverse using the equation given at the end of Section 4.4 of Fong *et al.* (2010). This method is closely related to that later described by Sørbye and Rue (2014). The supplementary materials contain R code for reproducing these prior specifications. For the Tanzania data this leads to gamma priors for the RW1 of $\tau_\gamma \sim \text{Gamma}(0.5, 0.00153)$ and for the ICAR of $\tau_\phi \sim \text{Gamma}(0.5, 0.00360)$.

5.4 Computation

Model fitting was carried out within the R computing environment. Weighted logistic regressions were fit using the `svyglm()` function from the `survey` package (Lumley, 2004) from which the design-based variance was extracted. The hierarchical Bayesian space-time models were fitted using the Integrated Nested Laplace Approximation (INLA) (Rue *et al.*, 2009) as implemented in the `INLA` package. INLA provides a fast alternative to MCMC for approximating the marginal posterior distributions of Markov random field (MRF) models. There is now extensive evidence that the approximations are accurate for space-time modeling, see for example Fong *et al.* (2010); Held *et al.* (2010).

5.5 Model Selection

In Table 1 we describe six plausible random effects specifications. A number of approaches have been described for comparing models, including the conditional predictive ordinate (CPO), the deviance information criteria (DIC) as introduced by Spiegelhalter *et al.* (2002) and the normalizing constants $p(\mathbf{y}|M)$ for the six models indexed by M . Let \mathbf{y}_{-its} represent the vector of data with the observation from region i , time period t and survey s removed. The idea behind the CPO is to predict the density ordinate of the left-out observation, based on those that remain. Specifically, the CPO for observation i, t, s is defined as:

$$\text{CPO}_{its} = p(y_i|\mathbf{y}_{-its}) = \int p(y_{its}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{-its}) d\boldsymbol{\theta} = E_{\boldsymbol{\theta}|y_{-its}} [p(y_{its}|\boldsymbol{\theta})]$$

where $\boldsymbol{\theta}$ represents the totality of parameters and in the U5MR setting the distribution of $y_{its}|\boldsymbol{\theta}$ is $N(\eta_{its}, \widehat{V}_{\text{DES},its})$. The CPOs can be used to look at local fit, or one can define an overall score for each model:

$$\text{LCPO} = \log(\text{CPO}) = \sum_{i=1}^I \sum_{t=1}^T \sum_{s=1}^S \log \text{CPO}_{its},$$

Table 2: Model comparison: p_D is the effective degrees of freedom, as defined for the calculation of the deviance information criteria (DIC), which also uses the deviance evaluated at the posterior mean, \bar{D} ; LCPO is defined as $\sum_{its} \log(CPO_{its})$.

Model	No Pars	$\log p(\mathbf{y})$	p_D	\bar{D}	DIC	LCPO
I	181	-311	75	409	484	-294
II	189	-305	80	384	464	-287
III	313	-258	119	222	341	-194
IV	223	-302	89	368	456	-283
V	347	-255	122	210	332	-183
VI	920	-255	135	199	334	-184

and good models will have relatively high values of LCPO. Held *et al.* (2010) discuss shortcuts for computation (i.e. avoidance of fitting the model $I \times T \times S$ times) using INLA.

We also calculate another widely-used model comparison measure, the deviance information criteria, or DIC (Spiegelhalter *et al.*, 2002). To define the DIC with respect to a generic set of parameters $\boldsymbol{\theta}$, first define an “effective number of parameters” as

$$p_D = E_{\boldsymbol{\theta}|\mathbf{y}} \{-2 \log[p(\mathbf{y}|\boldsymbol{\theta})]\} + 2 \log[p(\mathbf{y}|\bar{\boldsymbol{\theta}})] = \bar{D} + D(\bar{\boldsymbol{\theta}})$$

where D is the deviance, $\bar{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\mathbf{y}]$ is the posterior mean, $D(\bar{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior mean and $\bar{D} = E[D|\mathbf{y}]$. The DIC is given by

$$\text{DIC} = D(\bar{\boldsymbol{\theta}}) + 2p_D = \bar{D} + p_D,$$

so that we have the sum of a measure of goodness of fit and model complexity. We are wary of interpretation of DIC in our setting, since Plummer (2008) has shown that DIC is prone to inappropriately under-penalize large models such as the ones we are fitting, see also Spiegelhalter *et al.* (2014).

6 Applying Methods to household surveys and HDSS sites in Tanzania

We fit Models I–VI (as summarized in Table 1) to the Tanzania survey data and Table 2 provides the summaries of various model comparison summaries. Model V is the favored model according to both the DIC and LCPO criterion. The log of the normalizing constant for Models V and VI are very similar, but we see from the effective number of parameters that even though the number of interaction random effects is 373, there are only 13 effective parameters due to the closeness of the interactions to zero. Hence, from this point onwards we shall report summaries with respect to Model V. We begin by summarizing the posterior distribution, and then describe regional trends before moving to national estimates.

6.1 Summarizing the Posterior Distribution

Graphical summaries of the posteriors and priors for the variance components are given in Figure 7 while Table 3 gives numerical summaries and the proportion of total variation

explained. The total variance is

$$\sigma_\alpha^2 + s_\gamma^2 + \sigma_\theta^2 + s_\phi^2 + \sigma_\delta^2 + \sigma_{\nu_s}^2 + \sigma_{\nu_{si}}^2 + \sigma_{\nu_{st}}^2$$

where s_γ^2 and s_ϕ^2 are empirical estimates of the marginal variances in the RW1 and ICAR models. The structured temporal and unstructured spatial random effects explain 75% of the total variation. Hence, there is strong temporal structure and large spatial heterogeneity, which we shall discuss subsequently. The third largest contribution to the variation is 11% from the survey-space interaction. Different survey teams are sent to different regions which explains to some extent this relatively large contribution.

Table 3: Summaries of variance components. The proportion of variation is calculated as the contribution the relevant set of random effects makes to the total variation. In the case of the RW1 and ICAR models, the relevant contribution is evaluated empirically, since the variance parameter is conditional rather than marginal.

Variance	Interpretation	Median (95% Interval)	Percentage Variation
σ_α^2	Indept Time	0.003 (0.001, 0.035)	2.5
σ_γ^2	RW1 Time	0.038 (0.012, 0.146)	43.3
σ_θ^2	Indept Space	0.067 (0.033, 0.131)	32.0
σ_ϕ^2	ICAR Space	0.016 (0.002, 0.342)	5.0
σ_δ^2	Indept Space-Time Interaction	0.005 (0.001, 0.013)	2.4
$\sigma_{\nu_s}^2$	Indept Survey	0.002 (0.001, 0.013)	1.5
$\sigma_{\nu_{st}}^2$	Indept Survey-Time Interaction	0.004 (0.001, 0.012)	2.1
$\sigma_{\nu_{si}}^2$	Indept Survey-Space Interaction	0.024 (0.015, 0.038)	11.2

6.2 Regional Estimates

For regional i and 5 year period t , estimates and credible intervals of U5MR are based on

$${}_5q_{0,it} = \text{expit}(\mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it}).$$

Figure 2 shows maps of the posterior median estimates of child mortality (per 1000 births) by region for the six 5 year time periods. Child mortality has decreased markedly over the 30 year period considered but overall around 9% of infants still die before they turn 5, and there are strong regional differences. Inverse-variance weighted averages of the six household surveys and two surveillance site estimates by region and time point are displayed in Section 6 of the supplementary materials. Figure 3 shows the direct comparison of smoothed and weighted 5 year regional estimates. There is some deviation from the line of equality, with some attenuation of estimates from the most recent period in particular.

6.3 National Estimates

We examine national estimates based on the survey data collapsed across space, in order to make a comparison with UN estimates using the spatially aggregated version of our model.

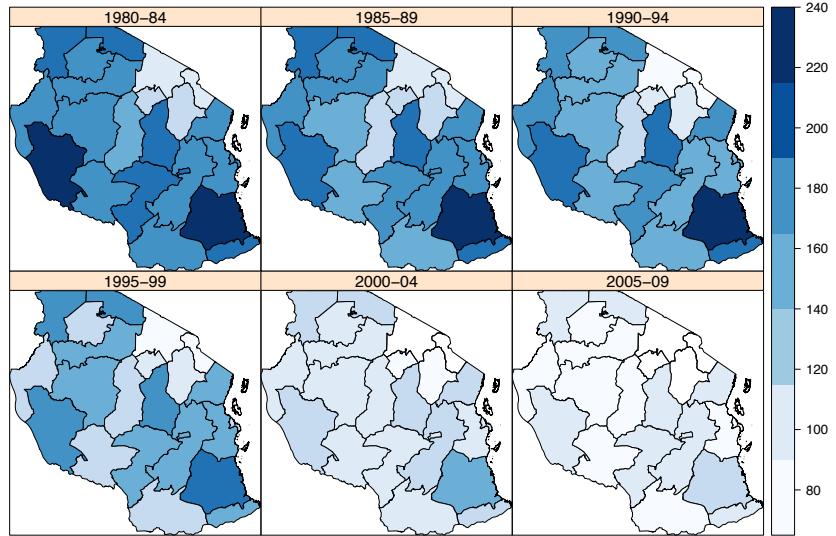


Figure 2: Smoothed regional estimates of child mortality (per 1000 births).

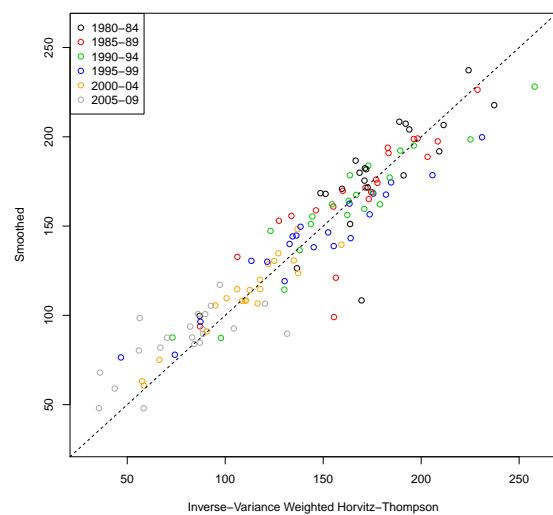


Figure 3: Smoothed and inverse-variance weighted Horvitz-Thompson regional estimates of child mortality (per 1000 births).

Table 4: Comparison of smoothed U5MR estimates with UN ${}_5q_{0t}$ estimates over period t . Rates are per 1000 births.

Period	Smoothed		UN	
	Median	95% Interval	Median	95% Interval
1980–1984	171.3	(158.8, 184.7)	175.4	(164.0, 189.3)
1985–1989	163.1	(152.0, 174.9)	173.1	(163.0, 183.6)
1990–1994	153.6	(143.4, 164.3)	164.0	(154.5, 174.1)
1995–1999	137.9	(129.1, 147.0)	152.2	(143.2, 161.6)
2000–2004	109.3	(101.7, 117.4)	113.8	(105.6, 122.0)
2005–2009	85.9	(79.0, 93.6)	77.2	(68.7, 86.4)

The model is

$$\eta_{ts} = \mu + \alpha_t + \gamma_t + \nu_s + \nu_{st}, \quad (6.1)$$

with the random effects described in the caption of Table 1. With respect to this model we define smoothed national U5MR estimates as

$${}_5q_{0,t} = \text{expit}(\mu + \alpha_t + \gamma_t)$$

which is intuitively reasonable since the left out regional and survey random effects are either zero mean (the independent random effects) or contain a sum-to-zero constraint (the ICAR random effects). Table 4 displays the smoothed median 5 year national estimates, from the relevant time periods, as well as the midpoint estimates from each 5 year period provided by the UN (<http://childmortality.org>). Figure 4 displays the national point estimate and 95% confidence intervals from the six household surveys and two surveillance sites with the smoothed estimates. Broadly speaking our results agree with the UN results, but are slightly smaller until the most recent time interval.

Figure 4 shows the smoothed estimates (with credible intervals) from the model given in (6.1), along with the survey design-based estimates (with design-based confidence intervals). The decline over time is evident, though the final national rate is still 86 deaths per 1000 births.

7 Discussion

We have described a general method for spatiotemporal smoothing of a health outcome, with the data arising from complex surveys and surveillance. The method was illustrated with child mortality in regions of Tanzania over 1980–2009 using data from DHS surveys and demographic surveillance. A great advantage of the model is that there is a fast implementation within the R computing environment using the existing `survey` and `INLA` packages. The supplementary materials contain example code. As an example, fitting the most complex model for the Tanzania data took just 18.7 seconds on a Macbook Pro¹. Another use of our model is for prediction, with the RW1 terms drawn from the relevant conditional distribution.

¹processor: 2.9 GHz Intel Core i7; memory: 8 GB 1,600 MHz DDR3.

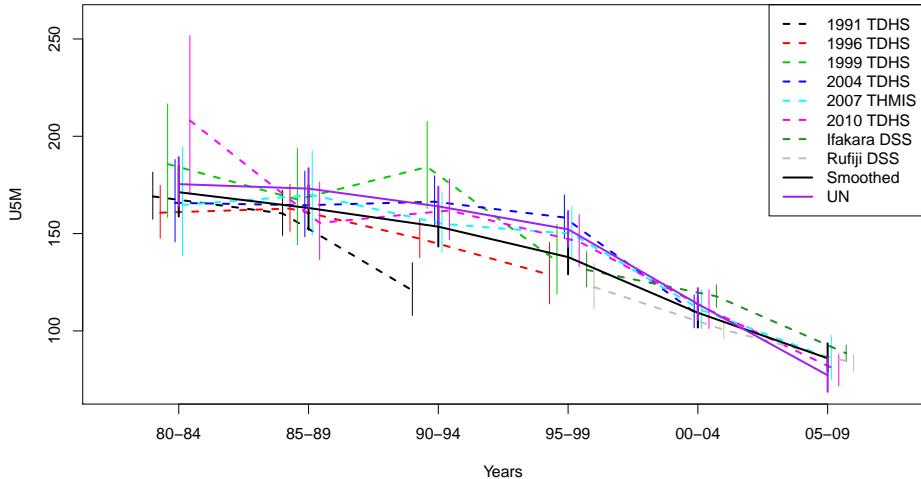


Figure 4: National estimates of $5q_{0,t}$ from 1980–2009 in Tanzania with design-based confidence intervals and smoothed estimates from Table 4.

Our model contains a relatively complex combination of nested and crossed random effects and we described a particular approach to hyperprior selection. As with any such suggestion, it is beneficial to examine prior sensitivity, and the supplementary materials contain details of a sensitivity study that we carried out. In the analysis we presented we did not consider clustering due to births on the same mother. We carried out an analysis in which this aspect was examined, by including in the design weights this component of clustering, but the results were very similar to those reported.

The substantive results for Tanzania confirm that our method produces results similar to both the UN IGME, Figure 4 and IHME (Alkema and You, 2012). However in contrast to those, we provide consistent subnational estimates (by region), which is not the aim of either the UN IGME or IHME approaches. An integral part of our method involves calculating and pooling estimates of child mortality from the DHS surveys and demographic surveillance sites and allowing both to inform our overall estimates by region and for the country as a whole. A byproduct of this procedure is an ability to carefully compare the DHS-based and demographic surveillance-based estimates of child mortality in the regions that include DSS sites. As Figure 4 makes clear, the central estimates from the two different data collection schemes are very similar. This adds more weight to similar findings by others (Byass *et al.*, 2007; Fottrell *et al.*, 2009; Hammer *et al.*, 2006) and reduces concerns about the Hawthorne effect preventing measures of child mortality from DSS sites from being more widely relevant, i.e. similar to surrounding populations.

Although we have demonstrated our method with a single country and outcome, it is sufficiently general to be applied to produce spatiotemporal estimates of a variety of indicators. Because this approach provides consistent, precise estimates across both time and space utilizing data from a variety of sources, including complex sample surveys, accounting for study designs, it should be considered as an additional approach for producing national and subnational estimates of child mortality and other key health, demographic and development

indicators.

The world's rapidly growing appetite for timely, subnational estimates of key development indicators will continue to motivate innovative new developments in both data collection and analysis. In addition to providing a means to improve indicator estimates using different sources of data, our results also hint at the possibility of eventually creating integrated data collection and analysis schemes that build on existing infrastructure to yield some of the functionality of full-coverage CRVS. Clark *et al.* (2012) and Ye *et al.* (2012) begin to discuss ideas in this vein, e.g. how one might utilize both sample surveys and demographic surveillance to continuously provide indicators equivalent to what is normally produced by vital registration. The method and results we present in this paper encourage future development of those ideas.

Acknowledgements

The first and second authors were supported by 2R01CA095994-05A1 from the National Institutes of Health. This work was partially supported by USAID through the MEASURE DHS project. Samuel Clark was partially supported by grant K01 HD057246 from the National Institute of Child Health and Human Development (NICHD). We would like to acknowledge the hard work and commitment of the HDSSs field teams and the data team and thank the residents in the surveillance area for offering their time for interviews and sharing their valuable personal information over the years.

References

- Alkema, L. and You, D. (2012). Child mortality estimation: a comparison of un igme and ihme estimates of levels and trends in under-five mortality rates and deaths. *PLoS medicine*, **9**(8), e1001288.
- Alkema, L., New, J. R., Pedersen, J., You, D., *et al.* (2014). Child mortality estimation 2013: An overview of updates in estimation methods by the united nations inter-agency group for child mortality estimation. *PloS one*, **9**(7), e101112.
- Allison, P. (1984). *Event History Analysis: Regression for Longitudinal Event Data*. Number 46. Sage Publications Inc.
- Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics*, **43**, 1–59.
- Binder, D. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279–292.
- Byass, P., Worku, A., Emmelin, A., and Berhane, Y. (2007). Dss and dhs: longitudinal and cross-sectional viewpoints on child and adolescent mortality in ethiopia. *Population Health Metrics*, **5**(1), 12.

- Clark, S. J., Wakefield, J., McCormick, T., and Michelle, R. (2012). Hyak mortality monitoring system innovative sampling and estimation methods: Proof of concept by simulation. Technical Report 118, Center for Statistics and the Social Sciences (CSSS), University of Washington, <https://www.csss.washington.edu/Papers/wp118.pdf>.
- Clark, S. J., Kahn, K., Houle, B., Arteche, A., Collinson, M. A., Tollman, S. M., and Stein, A. (2013). Young children's probability of dying before and after their mother's death: A rural South African population-based surveillance study. *PLoS medicine*, **10**(3), e1001409.
- Demographic and Health Surveys (1992). *Demographic Health Survey 1991/1992*. Bureau of Statistics Planning Commission.
- Demographic and Health Surveys (1997). *Tanzania Demographic and Health Survey 1996*. Bureau of Statistics [Tanzania] and Macro International Inc.
- Demographic and Health Surveys (2000). *Tanzania Demographic and Health Survey 1999*. Bureau of Statistics [Tanzania] and Macro International Inc.
- Demographic and Health Surveys (2005). *Tanzania Demographic and Health Survey 2004-05*. National Bureau of Statistics (NBS) [Tanzania] and ORC Macro.
- Demographic and Health Surveys (2010). *Tanzania Demographic and Health Survey 2010*. National Bureau of Statistics (NBS) [Tanzania] and ICF Macro.
- Fong, Y., Rue, H., and Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, **11**, 397–412.
- Fottrell, E., Enquselassie, F., and Byass, P. (2009). The distribution and effects of child mortality risk factors in ethiopia: a comparison of estimates from dss and dhs. *Ethiopian Journal of Health Development*, **23**(2).
- Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statistical Science*, **22**, 153–164.
- Hammer, G. P., Kouyaté, B., Ramroth, H., and Becher, H. (2006). Risk factors for childhood mortality in sub-saharan africa: a comparison of data from a demographic and health survey and from a demographic surveillance system. *Acta tropica*, **98**(3), 212–218.
- Held, L., Schrödle, B., and Rue, H. (2010). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA. In T. Kneib and G. Tutz, editors, *Statistical Modeling and Regression Structures – Festschrift in Honour of Ludwig Fahrmeir*, pages 91–110. Physica-Verlag.
- INDEPTH Network (2014). *Health and Demographic Surveillance Systems*. http://www.indepth-network.org/index.php?option=com_content&task=view&id=1798&Itemid=501.
- Jenkins, S. P. (1995). Easy estimation methods for discrete-time duration models. *Oxford bulletin of economics and statistics*, **57**(1), 129–136.

- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, **19**, 2555–2567.
- Lozano, R., Wang, H., Foreman, K. J., Rajaratnam, J. K., Naghavi, M., Marcus, J. R., Dwyer-Lindgren, L., Lofgren, K. T., Phillips, D., Atkinson, C., *et al.* (2011). Progress towards millennium development goals 4 and 5 on maternal and child mortality: an updated systematic analysis. *The Lancet*, **378**(9797), 1139–1165.
- Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, **9**.
- Lumley, T. (2010). *Complex Surveys: A Guide to Analysis using R*. John Wiley and Sons, Hoboken, Jersey.
- Mercer, L., Wakefield, J., Chen, C., and Lumley, T. (2014). A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spatial Statistics*, **8**, 69–85.
- Paris21 (2014). *Paris21: Partnership for Statistics in Development in the 21st Century*. <http://www.paris21.org>.
- Pedersen, J. and Liu, J. (2012). Child mortality estimation: Appropriate time periods for child mortality estimates from full birth histories. *PLoS Med*, **9**(8).
- Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, **9**, 523–539.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Application*. Chapman and Hall/CRC Press, Boca Raton.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.
- Rutstein, S. O. and Rojas, G. (2006). Tanzania demographic and health survey 1996. *Calverton, Maryland: ORC Macro*.
- Sørbye, S. and Rue, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, **8**, 39–51.
- Spiegelhalter, D., Best, N., Carlin, B., and Linde, A. V. D. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, **64**, 583–639.
- Spiegelhalter, D., Best, N., Carlin, B., and Linde, A. V. D. (2014). The deviance information criterion: 12 years on (with discussion). *Journal of the Royal Statistical Society: Series B*, **64**, 485–493.
- United Nations (2000). *Millennium Development Goals*. <http://www.un.org/millenniumgoals/>.
- United Nations (2014a). *Civil Registration and Vital Statistics Coverage*. http://unstats.un.org/unsd/demographic/CRVS/CR_coverage.htm.

United Nations (2014b). *Data Revolution for Sustainable Development*. <http://www.un.org/apps/news/story.asp?NewsID=48594#.VEVQpoctuvJ>.

United Nations (2014c). *Millennium Development Goal number 4: Reduce by two thirds, between 1990 and 2015, the under-five mortality rate*. <http://www.un.org/millenniumgoals/childhealth.shtml>.

United Nations (2014d). *The Post-2015 Development Agenda*. <http://www.post2015hlp.org/the-report/>.

United Nations (2014e). *Sustainable Development Goals*. <http://sustainabledevelopment.un.org/owg.html>.

USAID (2014). *Demographic and Health Surveys*. United States Agency for International Development, <http://www.dhsprogram.com>.

Wakefield, J. (2009). Multi-level modelling, the ecologic fallacy, and hybrid study designs. *International Journal of Epidemiology*, **38**, 330–336.

World Bank and World Health Organization (2014). *Global Civil Registration and Vital Statistics Scaling Up Investment Plan 2015-2024*. <http://www.worldbank.org/en/topic/health/publication/global-civil-registration-vital-statistics-scaling-up-investment>.

Ye, Y., Wamukoya, M., Ezeh, A., Emina, J. B., and Sankoh, O. (2012). Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-sahara africa? *BMC Public Health*, **12**(1), 741.

A Details of Discrete Survival Model

We wish to estimate the under 5 mortality rate (U5MR) denoted by ${}_5q_0$. The U5MR is calculated as

$${}_5q_0 = 1 - \prod_{j=0}^J (1 - {}_{n_j} q_{x_j}). \quad (\text{A.1})$$

Where we define ${}_nq_x = \Pr(\text{dying before } x + n \mid \text{lived until } x)$ The monthly probability of death for each interval,

$$p_j = \Pr(\text{dying in any month in the interval } x_j + n_j \mid \text{lived until } x_j),$$

may be estimated using a logistic generalized linear model (GLM):

$$\text{logit}(p_j) = \sum_{j=0}^J \beta_j I_j,$$

where I_j is the indicator for the $[x_j, x_{j+1})$ time interval. Once estimates $\hat{\beta}_j$ are estimated we can calculate $\hat{p}_j = \exp(\hat{\beta}_j)/[1 + \exp(\hat{\beta}_j)] = \text{expit}(\hat{\beta}_j)$. The complement of surviving each month of the interval $[x_j, x_j + n_j)$ is used to calculate ${}_{n_j}\hat{q}_{x_j} = 1 - (1 - \text{expit}(\hat{\beta}_j))^{n_j}$. Since

$$\begin{aligned} 1 - {}_{n_j}\hat{q}_{x_j} &= (1 - \text{expit}(\hat{\beta}_j))^{n_j} = \left(\frac{e^{\hat{\beta}_j} + 1}{e^{\hat{\beta}_j} + 1} - \frac{e^{\hat{\beta}_j}}{e^{\hat{\beta}_j} + 1} \right)^{n_j} \\ &= \left(\frac{1}{e^{\hat{\beta}_j} + 1} \right)^{n_j} = (e^{\hat{\beta}_j} + 1)^{-n_j} \end{aligned}$$

(A.1) can be written in terms of the estimated $\hat{\beta}$ as

$${}_5\hat{q}_0 = 1 - \prod_{j=0}^J (1 - {}_{n_j}\hat{q}_{x_j}) = 1 - \prod_{j=0}^J (e^{\hat{\beta}_j} + 1)^{-n_j}.$$

B Derivation of Standard Error for U5M

We choose to model the parameter $\eta = \text{logit}({}_5q_0)$. Ultimately we will use a Gaussian distribution for the first stage of our hierarchical model, so we would like a parameter that can take values along the whole real line. We have

$$\begin{aligned} \eta &= \text{logit}({}_5q_0) = \log \left(\frac{{}_5q_0}{1 - {}_5q_0} \right) \\ &= \log \left(\frac{1 - \prod_{j=0}^J (e^{\beta_j} + 1)^{-n_j}}{\prod_{j=0}^J (e^{\beta_j} + 1)^{-n_j}} \right) \\ &= \log \left(\prod_{j=0}^J (e^{\beta_j} + 1)^{n_j} - 1 \right) \end{aligned} \tag{B.1}$$

The asymptotic distribution of the MLE is $\hat{\beta} \sim N(\beta, \Sigma)$ and from the delta method we can find the asymptotic distribution of $\hat{\eta}$:

$$\hat{\eta} \sim N(\text{logit}{}_5q_0, \widehat{\text{var}}(\hat{\eta})) \tag{B.2}$$

where

$$\widehat{\text{var}}(\hat{\eta}) = \frac{\partial \boldsymbol{\eta}^T}{\partial \beta} \widehat{\Sigma} \frac{\partial \boldsymbol{\eta}}{\partial \beta}. \tag{B.3}$$

The form of the covariance matrix $\widehat{\Sigma}$, depends on the survey design and can be extracted from the `svyglm()` function within the `survey` package. If we define:

$$\gamma = \prod_{j=0}^J (e^{\beta_j} + 1)^{n_j}$$

We take $\eta = \log(\gamma - 1)$ and, after some algebra,

$$\frac{\partial\eta}{\partial\beta_j} = \frac{\gamma}{\gamma-1} \times [n_j \cdot \text{expit}(\beta_j)]$$

The value of $\frac{\partial\eta}{\partial\beta}$ is used in (B.3) to asymptotic distribution, (B.2) which is used to derive $(1 - \alpha)\%$ confidence intervals for ${}_5q_0$ as

$$[\text{expit}(\hat{\eta} + \sqrt{\text{var}(\hat{\eta})} \times z_{\alpha/2}), \text{expit}(\hat{\eta} + \sqrt{\text{var}(\hat{\eta})} \times z_{1-\alpha/2})]. \quad (\text{B.4})$$

C Simulation Study

Deaths within the first 60 months of life were assigned based on the probabilities shown in Figure 5 using a multinomial distribution. As we are only interested in death within the first 5 years, the remaining probability was assigned to a 61st category for death after the age of 5.

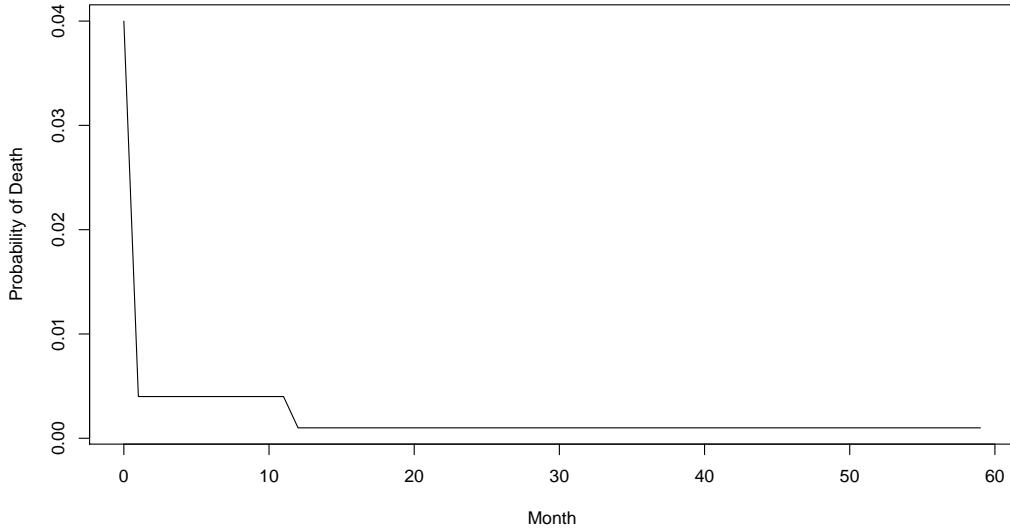


Figure 5: Monthly probability of death for the first 60 months.

Figure 6 displays the interval widths by number of clusters and sample size within cluster. As expected, the intervals narrow as either the number of first stage clusters or the number of samples within each cluster increases. The delta method and jackknife variances are quite similar in their performance.

D Hyperprior Specification

Figure 7 illustrates the sensitivity of point and interval estimates with three different prior distributions corresponding to 95% ranges for the residuals odds ratios of [0.5,2], [0.2,5],

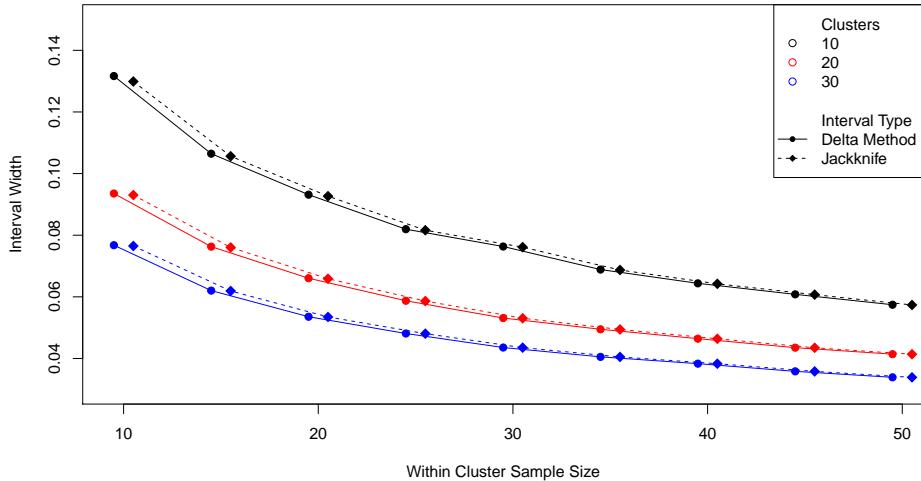


Figure 6: Width of jackknife and delta method 95% confidence intervals over a variety of clusters and cluster sizes.

[0.1,10]. The first choice was used for the results presented in the paper. We see sensitivity for the spatial random effects, though the total spatial random effects contributions remain relatively constant since as the structured random effects increase, the unstructured random effects decrease. The structured temporal random effects are robust, which is reassuring since these provide the largest contribution to the overall variability; these are well-estimated, however, since the trend is strong. Similarly, the unstructured survey-area random effects are robust, but all of the standard deviations of the remaining independent random effects show modest increases as the prior moves further from zero.

Examples of the code used to find the prior parameters for the ICAR and RW1 models are included in Section 7.

E Summary of Random Effects

In this section we present graphical summaries of the various random effects present in Model V of the Tanzania U5MR model. Figure 8 presents the posterior medians of the spatial (ICAR) and unstructured random effects (note that the scales on the different plots differ).

Figure 9 plots the unstructured temporal random effects versus period, along with the survey by period random effects. The unstructured random effects are relatively small in magnitude compared with the survey by period random effects.

Figure 10 displays the unstructured time random effects (α_t) compared with the structured time random effects (γ_t). The structured random effects have a much larger range than the unstructured effects.

Figure 11 displays structured time random effects (γ_t) by time. There is a noticeable negative trend in the random effects.

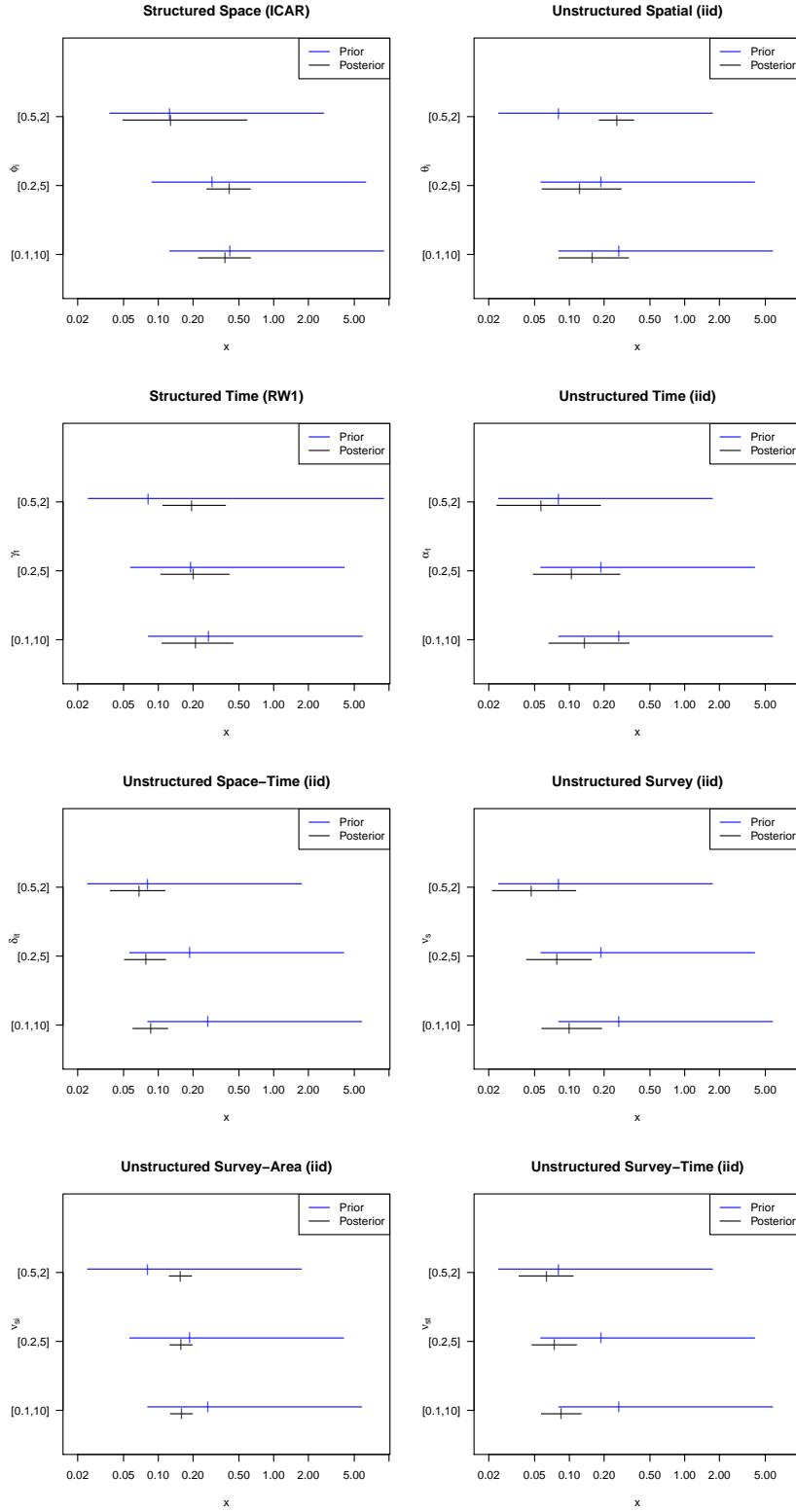


Figure 7: Prior sensitivity of the standard deviations of the eight random effects in the model. The three priors are based on 95% prior intervals on the residual odds ratios of [0.5,2], [0.2,5], [0.1,10].

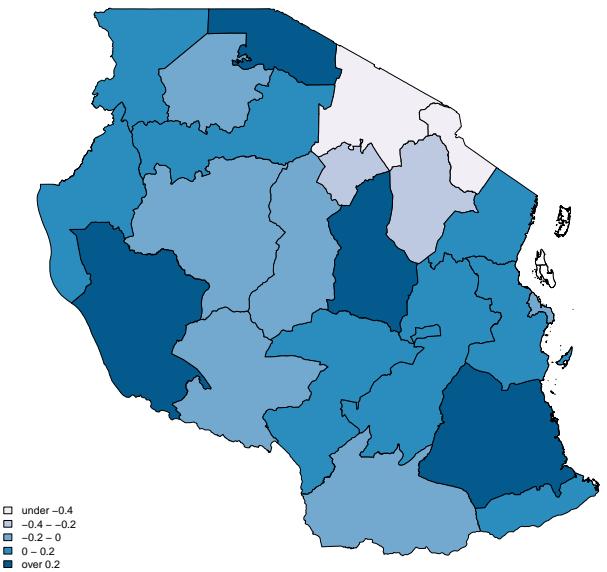
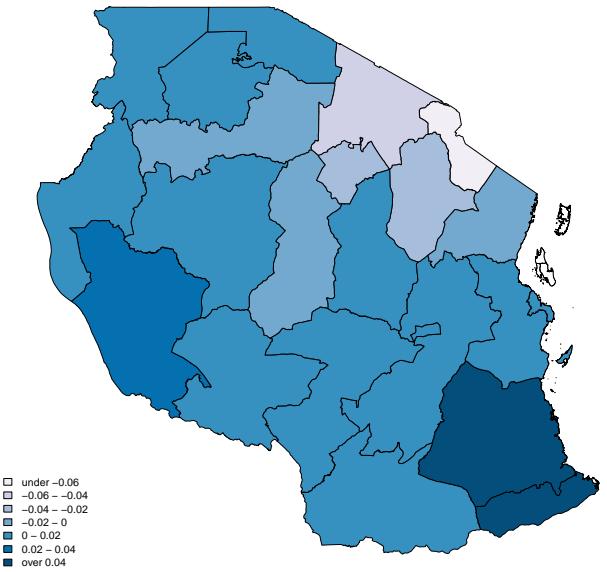


Figure 8: ICAR random effects, ϕ_i (top) and unstructured spatial random effects, θ_i (bottom).

Figure 12 provides maps of the unstructured space-time random effects (δ_{it}). There is no clear spatial pattern to the high-valued and low-valued random effects by time period. So, based on this plot, there is little evidence of space-time interaction (which is consistent with the model comparison statistics given in the paper). Similarly, the plots in Figure 13 which display the survey-area random effects (ν_{is}) along with the magnitude of the survey-specific (ν_s) random effect, do not show similar spatial patterns over the different time periods.

Figure 14 provides maps of the inverse-variance weighted Horvitz-Thompson regional estimates of child mortality. Unlike the smoothed maps provided in the main text these maps display some unlikely temporal trends.

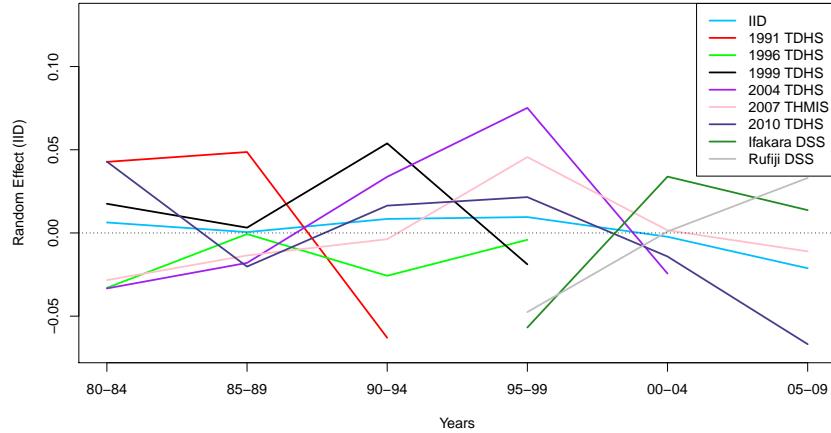


Figure 9: Unstructured time (α_t) and survey-time (ν_{st}) random effects.

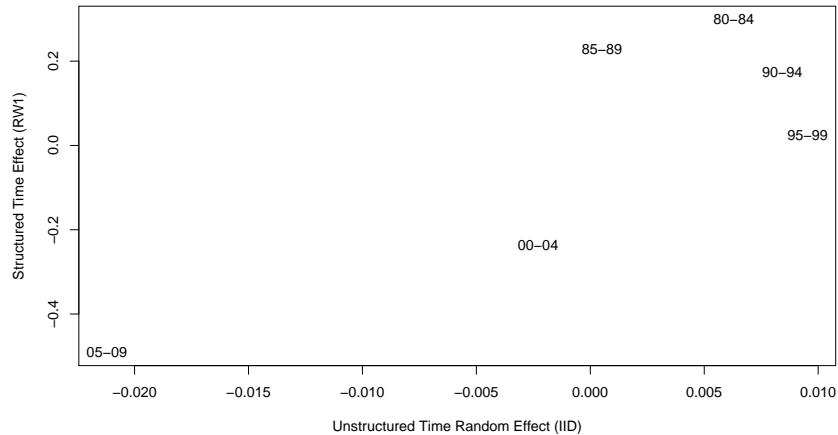


Figure 10: Unstructured time (α_t) and structured time (γ_t) random effects.

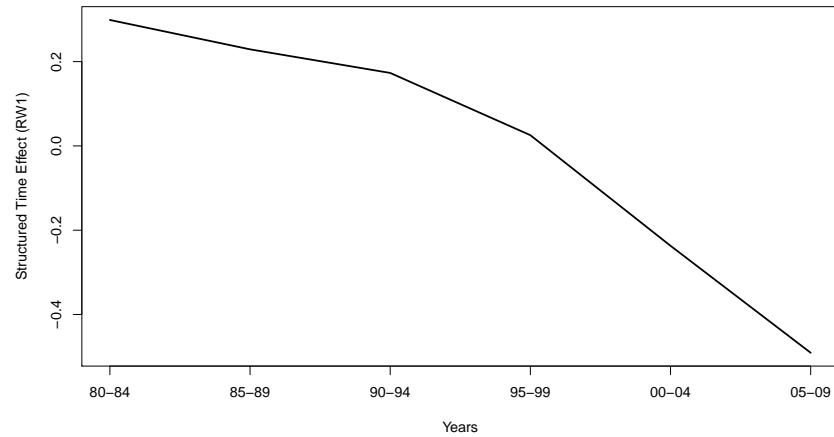


Figure 11: Structured time (γ_t) random effects.

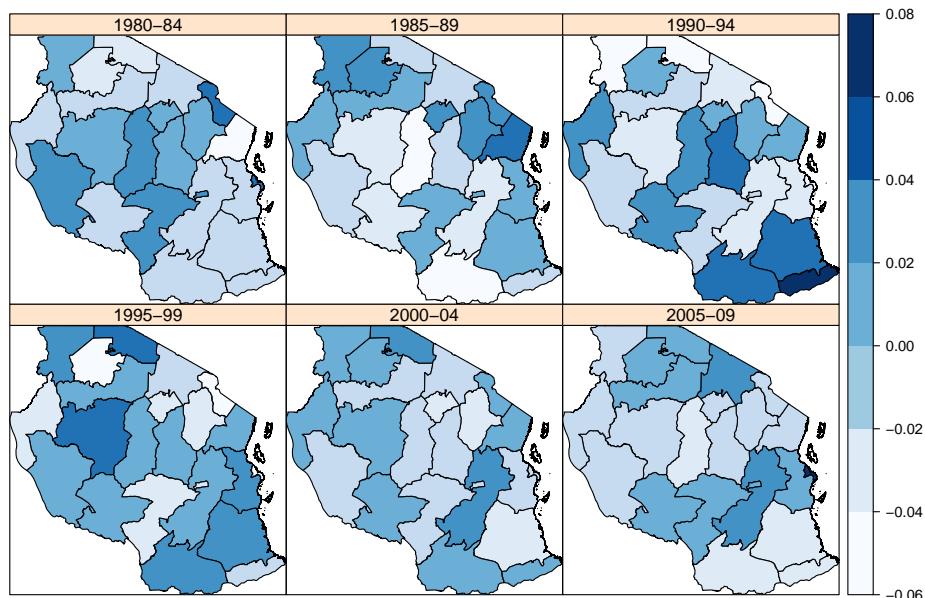


Figure 12: Unstructured space-time random effects (δ_{it}).

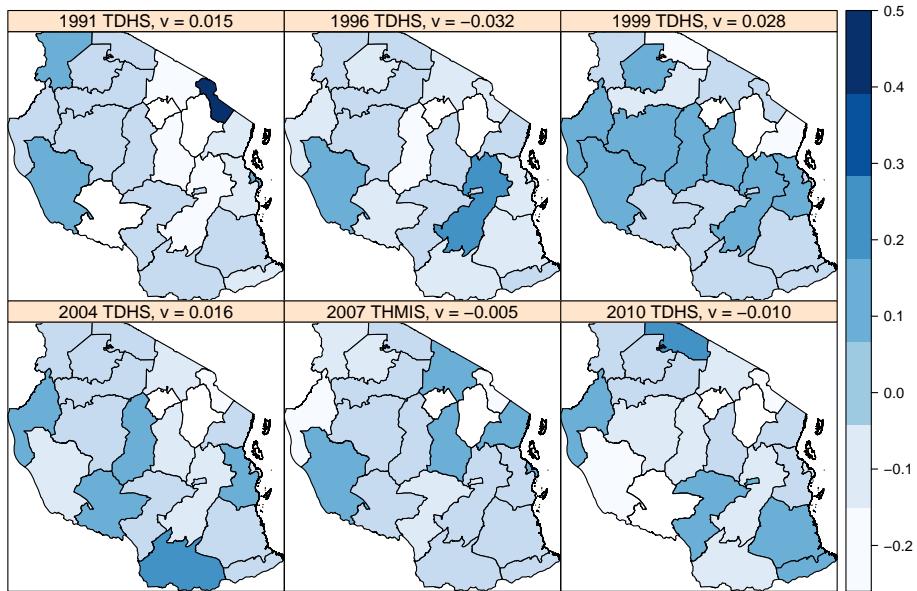


Figure 13: Survey (ν_s) and survey-area (ν_{si}) random effects. The median random effect (ν_s) is given in the heading of each plot. There are five Demographic and Health Surveys (DHS) and one Tanzania HIV and Malaria Indicator survey (THMIS).

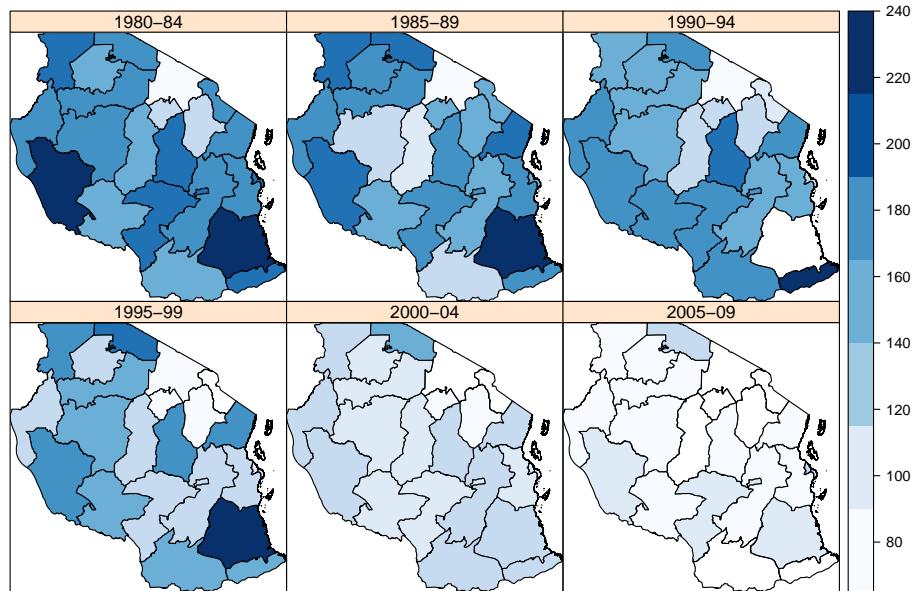


Figure 14: Inverse-variance weighted Horvitz-Thompson regional estimates of child mortality (per 1000 births).

F Regional Analyses

F.1 Ifakara and Morogoro

Table 5 compares the DHS, DSS, and smoothed ${}_5q_0$ estimates for the Morogoro Region. Figure 15 displays the Morogoro point estimates and 95% confidence intervals from the six household surveys and the Ifakara surveillance site with the smoothed estimates.

Table 5: Comparison of standard estimates from Morogoro Region based on 2010 DHS, Ifakara DSS, and the smoothed combined estimates using all DHS and DSS.

Model	2010 DHS		DSS		Smoothed	
	Est.	95% CI	Est.	95% CI	Est.	95% CI
1980-1984	247.2	(115.8, 451.5)			182.4	(155.6, 212.1)
1985-1989	114.4	(62.4, 200.5)			169.7	(145.4, 196.6)
1990-1994	139.2	(89.5, 210.1)			162.2	(138.4, 188.5)
1995-1999	165.4	(107.7, 245.7)	131.3	(122.6, 140.7)	149.6	(129.4, 172.8)
2000-2004	102.0	(65.0, 156.4)	117.8	(112.1, 123.7)	119.9	(102.4, 140.2)
2005-2009	83.8	(56.8, 121.8)	88.5	(84.4, 92.8)	92.4	(78.2, 109.1)

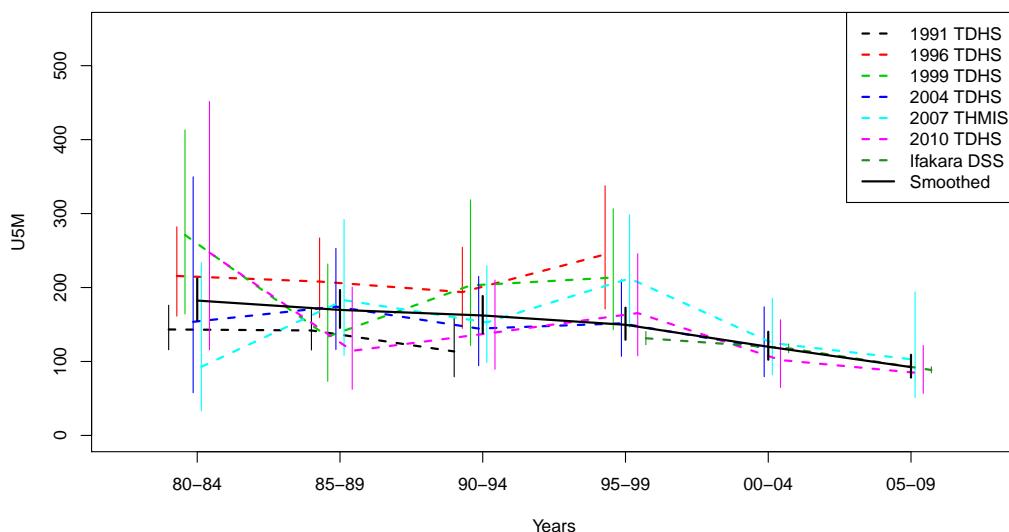


Figure 15: Regional estimates of ${}_5q_0$ in Morogoro, TZA with confidence intervals.

F.2 Rufiji and Pwani

Table 6 compares the DHS, DSS, and smoothed ${}_5q_0$ estimates for the Pwani Region. Figure 16 displays the Pwani point estimates and 95% confidence intervals from the six household surveys and the Rufiji surveillance site with the smoothed estimates.

Table 6: Comparison of standard estimates from Pwani Region based on 2010 DHS, Rufiji DSS, and the smoothed combined estimates using all DHS and DSS..

Model	2010 DHS		DSS		Smoothed	
	Est.	95% CI	Est.	95% CI	Est.	95% CI
1980–1984	222.2	(108.8, 400.8)			175.5	(149.0, 205.3)
1985–1989	188.8	(117.4, 289.4)			169.1	(144.4, 197.5)
1990–1994	140.8	(73.1, 254.1)			155.4	(131.9, 181.6)
1995–1999	148.5	(95.7, 223.3)	122.6	(111.5, 134.6)	144.2	(123.9, 167.7)
2000–2004	116.0	(76.1, 173.1)	100.6	(95.9, 105.4)	109.5	(93.1, 128.6)
2005–2009	52.8	(30.1, 90.9)	83.4	(79.1, 88.0)	87.5	(73.6, 103.9)

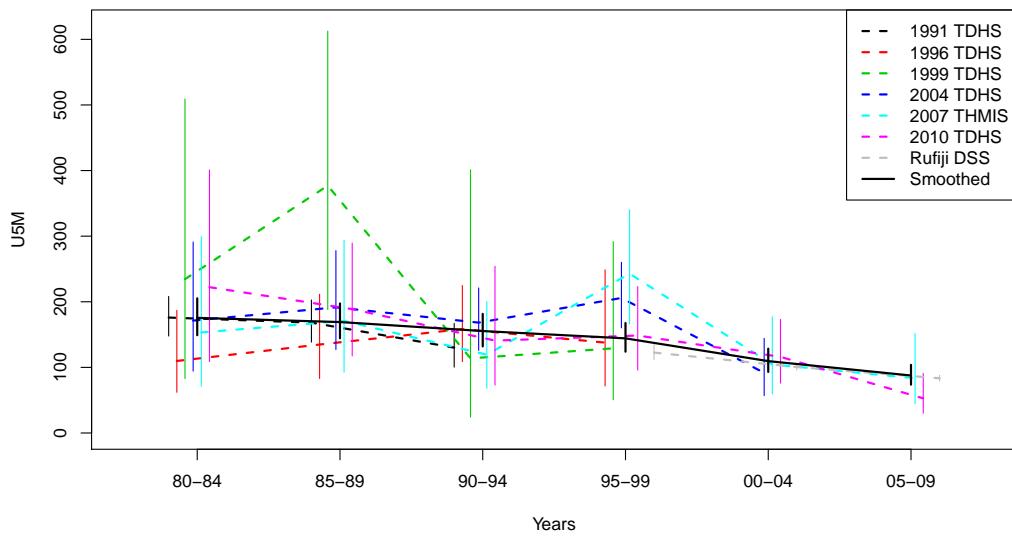


Figure 16: Regional estimates of ${}_{5}q_0$ in Pwani, TZA with confidence intervals.

G R Code

Section 7.1 provides examples of using the `survey` package in R to calculate the U5MR as well as a function to provide the intervals described in Section 2. Section 7.2 provides code for selecting the hyperprior for the Random Walk 1 random effect.s Section 7.3 provides code for selecting the hyper prior for the ICAR random effects. Section 7.4 fits the space-time model selected (in the main text) using the `inla()` function in R.

G.1 Using the survey package

```
# read in child month data #
births<-read.dta("dhs2010ChildMonths.dta")

# create survey design object #
my.svydesign <- svydesign(id= ~v002+caseid,strata=~v024,nest=T,
  weights= ~v005, data=births)

# subsetting design object #
design80<-subset(my.svydesign,per5=="80-84")

# fitting the logistic model #
glm80<-svyglm(died~factor(ageGrpD),design=design80,family=binomial)

# a function to calculate logit(U5M) and Variance #
get.est<-function(glm1){
  V<-vcov(glm1)
  betas<-summary(glm1)$coef[,1]
  ns<-c(1,11,12,12,12,12)
  probs<-expit(c(0,rep(betas[1],5))+betas)

  u5m.est<-(1-probs)^ns,na.rm=T))##1000
  ## finding the se of the logit(5q0) ##

  ## partial derivatives ##
  ebets<-exp(c(0,rep(betas[1],5))+betas)
  expitbets<-expit(c(0,rep(betas[1],5))+betas)

  gamma<-prod((1+ebets)^ns)-1

  derivatives<-(gamma+1)/gamma*ns*expitbets
  derivatives[1]<-(gamma+1)/gamma* ns%*%expitbets

  # results #
  var.est<-t(derivatives)%*%V%*%derivatives
  lims<-logit(u5m.est)+qnorm(c(0.025,0.975))*sqrt(var.est)
```

```

    return(c(u5m.est,expit(lims),logit(u5m.est),var.est))
}

```

```

# getting the estimates #
get.est(glm80)

```

G.2 Choosing Prior for RW1 Random Effect

```

#
# (R,1/R) is the range of the residual odds ratios
# gives a=d/2 where d = degrees of freedom of marginal Student's t

R <- 0.5; d <- 1; a <- d/2; p <- 0.025; b <-(log(R))^2*d/(2*qt(p,df=d)^2)

# Gives a=0.5 and b=0.001488
c(a,b)
1/sqrt(qgamma(p=c(0.025,0.975),a,b))
1/sqrt(qgamma(p=c(0.9,0.1),a,b)) # 0.03316518 0.4341181

# Check range using simulation #

nsamp <- 100000
tausamp <- rgamma(nsamp,a,b)
Usamp <- rnorm(nsamp,mean=0,sd=1/sqrt(tausamp))
quantile(exp(Usamp),p=c(0.025,0.5,0.975))

# The adjacency matrix for 6 years
m2 = c(1,2,2,2,2,1)
# create the adjacency #
adj2<-c(2,
       1,3,
       2,4,
       3,5,
       4,6,
       5)

make.Q <- function(num.neighbors, neighbors, omega.sq = 1){
  n <- length(num.neighbors)
  mat <- matrix(0, ncol = n, nrow = n)
  diag(mat) <- num.neighbors
  mat[cbind(rep(1:n, num.neighbors), neighbors)] <- -1
  mat/omega.sq
}
vars.Q <- function(eigenvalues,eigenvectors){


```

```

margsum <- 0
nloop <- length(eigenvalues)-1
for (i in 1:nloop){
  ev <- eigenvectors[,i]
  margsum <- margsum + ev %*% t(ev)/ eigenvalues[i]
}
margvars <- diag(margsum)
margvars
}
#
sim.Q <- function(Q){
  eigenQ <- eigen(Q)
  rankQ <- qr(Q)$rank
  sim <- as.vector(eigenQ$vectors[,1:rankQ] %*%
    matrix(
      rnorm(rep(1, rankQ), rep(0, rankQ),
      1/sqrt(eigenQ$values[1:rankQ])),
      ncol = 1))
  sim
}
#
Q <- make.Q(m2, adj2, 1)
eigentemp <- eigen(Q)
eigenvaluesQ <- eigentemp$values
eigenvectorsQ <- eigentemp$vectors
rankQ <- qr(Q)$rank # 5
margy <- mean(vars.Q(eigenvaluesQ,eigenvectorsQ))
#
nsamp <- 5000
astar <- a; bstar <- b/margy
c(astar,bstar)
taustarsamp <- rgamma(nsamp,astar,bstar)
Ustarsamp <- matrix(nrow=nsamp,ncol=6)
for (i in 1:nsamp){
  Qstar <- Q*taustarsamp[i]
  Ustarsamp[i,] <- sim.Q(Qstar)
}
quantile(exp(Ustarsamp),p=c(0.025,0.5,0.975)) # 0.5285916 0.9999869 1.8503786

```

G.3 Choosing Prior for ICAR Random Effect

```

#
# (R,1/R) is the range of the residual odds ratios
# gives a=d/2 where d = degrees of freedom of marginal Student's t

```

```

R <- 0.5; d <- 1; a <- d/2; p <- 0.025; b <-(log(R))^2*d/(2*qt(p,df=d)^2)
# Gives a=0.5 and b=0.001488

1/sqrt(qgamma(p=c(0.9,0.1),a,b)) # 0.03316518 0.4341181

# Check range using simulation #

nsamp <- 100000
tausamp <- rgamma(nsamp,a,b)
Usamp <- rnorm(nsamp,mean=0,sd=1/sqrt(tausamp))
quantile(exp(Usamp),p=c(0.025,0.5,0.975))

# The adjacency matrix for Tanzania without Zanzibar
Amat<-as.matrix(read.table("adj_regions_nozanz.txt"))
m2 = apply(Amat,1,sum)

# create the adjacency list #
nums<-c(1:21)

adj2<-NULL

for(i in 1:21){

  adj2<-c(adj2,nums [as.numeric(Amat[i,])==1])
}

make.Q <- function(num.neighbors, neighbors, omega.sq = 1){
  n <- length(num.neighbors)
  mat <- matrix(0, ncol = n, nrow = n)
  diag(mat) <- num.neighbors
  mat[cbind(rep(1:n, num.neighbors), neighbors)] <- -1
  mat/omega.sq
}
vars.Q <- function(eigenvalues,eigenvectors){
  margsum <- 0
  nloop <- length(eigenvalues)-1
  for (i in 1:nloop){
    ev <- eigenvectors[,i]
    margsum <- margsum + ev %*% t(ev)/ eigenvalues[i]
  }
  margvars <- diag(margsum)
  margvars
}
#

```

```

sim.Q <- function(Q){
  eigenQ <- eigen(Q)
  rankQ <- qr(Q)$rank
  sim <- as.vector(eigenQ$vectors[,1:rankQ] %*%
    matrix(
      rnorm(rep(1, rankQ), rep(0, rankQ),
      1/sqrt(eigenQ$values[1:rankQ])),
      ncol = 1))
  sim
}
#
Q <- make.Q(m2, adj2, 1)
eigentemp <- eigen(Q)
eigenvaluesQ <- eigentemp$values
eigenvectorsQ <- eigentemp$vectors
rankQ <- qr(Q)$rank # 20
margy <- mean(vars.Q(eigenvaluesQ,eigenvectorsQ))
#
nsamp <- 5000
astar <- a; bstar <- b/margy
c(astar,bstar)
taustarsamp <- rgamma(nsamp,astar,bstar)
Ustarsamp <- matrix(nrow=nsamp,ncol=21)
for (i in 1:nsamp){
  Qstar <- Q*taustarsamp[i]
  Ustarsamp[i,] <- sim.Q(Qstar)
}
quantile(exp(Ustarsamp),p=c(0.025,0.5,0.975)) # 0.5152361 0.9999245 1.9420017

```

G.4 Using INLA to fit the models

```

##### --- Setting priors --- #####
mu0<-log(0.1/0.9)
sig2.0<-10000

# range of [0.5,2] #
a.iid<-0.5
b.iid<-0.001488

a.rw1<-0.5
b.rw1<-0.001530466

a.icar<-0.5
b.icar<-0.003602143

```

```

# the formula #
mod5<-logit.est~f(survey,model="iid", param=c(a.iid,b.iid))
+f(survey.area,model="iid", param=c(a.iid,b.iid))
+f(survey.time,model="iid", param=c(a.iid,b.iid))
+f(region.unstruct,model="iid",param=c(a.iid,b.iid))
+f(region.struct, graph=Amat,model="besag",param=c(a.icar,b.icar))
+f(time.struct,model="rw1",param=c(a.rw1,b.rw1))
+f(time.unstruct,model="iid",param=c(a.iid,b.iid))
+f(time.area,model="iid", param=c(a.iid,b.iid))

inla.fit <- inla(mod5,
  family = "gaussian", control.compute=list(dic=T,mlik=T,cpo=T),
  data =exdat,
  control.predictor=list(compute=TRUE),
  control.family=list(hyper=list(prec=list(initial=log(1),fixed=TRUE))),
  scale=logit.prec)

summary(inla.fit)

```