The Annals of Applied Statistics
2015, Vol. 9, No. 4, 1889–1905
DOI: 10.1214/15-AOAS872

SPACE–TIME SMOOTHING OF COMPLEX SURVEY DATA: SMALL AREA ESTIMATION FOR CHILD MORTALITY[1]

BY LAINA D. MERCER[*], JON WAKEFIELD[*], ATHENA PANTAZIS[*], ANGELINA M. LUTAMBI[†], HONORATI MASANJA[†] AND SAMUEL CLARK[*,‡,§,¶,‖,2]

University of Washington[*], Ifakara Health Institute[†], University of Colorado[‡], University of the Witwatersrand[§], INDEPTH Network[¶] and ALPHA Network[‖]

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. The use of conventional hierarchical models requires careful thought since the survey weights may need to be considered to alleviate bias due to nonrandom sampling and nonresponse. The application that motivated this work is an estimation of child mortality rates in five-year time intervals in regions of Tanzania. Data come from Demographic and Health Surveys conducted over the period 1991–2010 and two demographic surveillance system sites. We derive a variance estimator of under five years child mortality that accounts for the complex survey weighting. For our application, the hierarchical models we consider include random effects for area, time and survey and we compare models using a variety of measures including the conditional predictive ordinate (CPO). The method we propose is implemented via the fast and accurate integrated nested Laplace approximation (INLA).

1. Introduction. Over the past fifteen years the United Nations' (UN) Millennium Development Goals (MDGs) [UN (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

Received November 2014; revised September 2015.

[1]Supported in part by USAID through the MEASURE DHS project and by a Shanahan Endowment Fellowship and a Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) training grant, T32 HD007543, to the Center for Studies in Demography & Ecology at the University of Washington and a 2R01CA095994-05A1 from the National Institutes of Health.

[2]Supported in part by Grant K01 HD057246 from the NICHD.

Key words and phrases. Bayesian smoothing, infant mortality, small area estimation, survey sampling.

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. In 2015 the UN and its partners are taking stock of experience with the MDGs and coordinating the establishment of a new set of global goals [UN (2014d)]—the Sustainable Development Goals (SDGs) [UN (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 [UN (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 [UN (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 health and demographic surveillance system (HDSS) sites [INDEPTH Network (2014)].

Reducing child mortality is MDG 4 [UN (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) recently developed a Bayesian B-spline Bias-reduction (B3) method [Alkema and New (2014), Alkema et al. (2014)], and the Institute for Health Metrics and Evaluation (IHME) uses a Gaussian process regression [Wang et al. (2014)]. Both of these methods produce national estimates through time with measures of uncertainty. None are designed to reveal variation in child mortality within countries. A recent paper by Dwyer-Lindgren et al. (2014) compared many Bayesian space–time smoothing models to produced subnational estimates of U5MR for Zambia. The major methodological limitation of this approach is that it does not incorporate area-specific sampling variability at the first stage of analysis, which we show can be quite variable for small areas.

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 HDSS sites. HDSS 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 HDSS 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, for example, using Horvitz–Thompson estimators [Horvitz and Thompson (1952)]. 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–survey 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. Hierarchical Bayesian space–time models are introduced in Section 4. The results of our modeling efforts of under five mortality rates (U5MR) within Tanzania from 1980–2010 are given in Section 5 and discussed in Section 6.

2. Data sources. We focus on child mortality using data from 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. The Ifakara and Rufigi HDSS sites contribute data to the Morogoro and Pwani regions of Tanzania, respectively.

2.2. *Household surveys*. Full TDHS surveys that collected data necessary for child mortality estimates were conducted in Tanzania in 2010, 2004–2005, 1999, 1996 and 1991–1992, in addition to the THMIS that included child mortality which was conducted in 2007–2008. The 2010 TDHS, 2007–2008 THMIS and 2004–2005 TDHS surveys used 2-stage cluster samples. First, enumeration areas were sampled from the 2002 Tanzania census and, second, a systematic sampling of households within each enumeration area was carried out. The 1999 TDHS, 1996 TDHS and 1991–1992 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. For all surveys stratification by urban/rural and region was done at the first stage, with oversampling of Dar es Salaam and other urban areas. The surveys were designed to be nationally representative and to be able to provide estimates of contraceptive prevalence at the regional level. All six household surveys contributed observations to the 21 mainland regions of Tanzania.

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 data set. 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 1991–1992 Tanzania DHS final report [Demographic and Health Surveys (1992)] states 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. We modeled child mortality using discrete time survival analysis (DTSA) [Allison (1984), Jenkins (1995)]. Our main aim is to examine the change in risk as a function of age and historical period. DTSA allows us to easily estimate the predicted probabilities which can be used directly in traditional mortality analysis methods such as life tables, in our case to calculate U5MR. We wish to estimate the U5MR and define ${}_nq_x = \Pr(\text{dying before } x+n | \text{lived until } x)$ and the discrete hazards model splits the $[0,5)$ period into J intervals $[x_1, x_2), [x_2, x_3), \ldots, [x_J, x_{J+1})$, where $x_{j+1} = x_j + n_j$ so that n_j is the length of the interval beginning at x_j, $j = 1, \ldots, J$. Then U5MR is calculated as

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

For our purposes, ${}_5q_0$ is calculated by dividing the first 60 months into six intervals ($J = 6$), $[0,1), [1,12), [12,24), [24,36), [36,48), [48,60)$ with $(x_1, \ldots, x_6) = (0, 1, 12, 24, 36, 48)$ and $(n_1, \ldots, n_6) = (1, 11, 12, 12, 12, 12)$. Data were organized as child-months where each child was at risk during each month observed from birth up to and including the month of their death. The observed data consist of, for each birth, a binary sequence up to length 60 with 0/1 corresponding to survival/death. For example, a child that died in their fourth month would contribute one child-month to the first age category and three to the second age category. The first three child-months would be assigned a 0 outcome and the final month would be assigned a 1.

We use logistic regression to estimate the monthly probability of dying conditional on the state of the child at the beginning of the month. The monthly probability of death for each interval, ${}_1q_x$, is the probability of dying in $[x, x+1)$ for $x \in [x_j, x_j + n_j)$ and can be estimated using a logistic generalized linear model (GLM) with J factors for age intervals, $\text{logit}({}_1q_x) = \beta_j$ for $x \in [x_j, x_j + n_j)$. A more detailed discussion of the DTSA method can be found in Clark et al. (2013).

In the complex survey context that is relevant for the Tanzanian household surveys, an important consideration is that the design weights must be acknowledged. This is achieved by solving a (design) weighted score statistic [Binder (1983)], resulting in estimates of the finite population parameter $\mathbf{B} = [B_1, \ldots, B_J]$; see details in the supplementary material [Mercer et al. (2015)]. Once $\widehat{B}_j$ are estimated, we can calculate $\widehat{{}_1q_x} = \exp(\widehat{B}_j)/[1 + \exp(\widehat{B}_j)]$ for $x \in [x_j, x_j + n_j)$. The complement of surviving each month of the interval $[x_j, x_j + n_j)$ is used to calculate ${}_{n_j}\widehat{q}_{x_j} = 1 - (1 - \widehat{{}_1q_x})^{n_j}$, which may be substituted into (3.1) to give ${}_5\widehat{q}_0$ (for additional details see the supplementary material).

In Section 4 we will construct, for a generic U5MR, a working likelihood based on the asymptotic distribution

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

where $\eta = \log[{}_5q_0/(1-{}_5q_0)]$ and $\widehat{V}_{\text{DES}}$ is the estimated asymptotic (design-based) variance estimate of $\text{logit}({}_5\widehat{q}_0)$, which is obtained via the delta method; the supplementary material contain details of this calculation and a simulation study which investigates the asymptotic properties of the variance estimate compared with a jackknife variance estimate [Lohr (2010), Chapter 9] that is often used in the in the context of child mortality estimates [Pedersen and Liu (2012)].

Simulation results were much as one would expect from clustered sampling; coverage improves when there are more clusters and within a given number of clusters there is little gain in precision when increasing the sample size. Generally the performance of the delta method and jackknife intervals is very similar. 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 in our application. Consequently, we will use the asymptotic distribution with the delta method variance as a working likelihood.

4. Combining data sources in the hierarchical Bayesian space–time model.

4.1. *The first stage.* Let ${}_5\widehat{q}_{0its}$ 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 design variables upon which sampling were based, and associated population totals, 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 pseudo-maximum likelihood estimator (MLE):

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

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

$$y_{its}|\eta_{its} \sim \text{N}(\eta_{its}, \widehat{V}_{\text{DES},its}), \tag{4.1}$$

which has been shown to perform well in the context of small area estimation from complex surveys [Mercer et al. (2014)]. Dwyer-Lindgren et al. (2014) also used the pseudo-MLE, but did not incorporate design effects and instead assumed a common variance across all observations. However, Figure 13 from the supplementary material shows that the variance of the five-year direct estimates can vary significantly by survey and region.

TABLE 1

Random effects models for time period t, region i and survey s. In all models μ is the intercept and $\alpha_t \sim_{\text{i.i.d.}} \text{N}(0, \sigma_\alpha^2)$, $\theta_i \sim_{\text{i.i.d.}} \text{N}(0, \sigma_\theta^2)$, $\phi_i \sim \text{ICAR}(\sigma_\phi^2)$, $\delta_{it} \sim_{\text{i.i.d.}} \text{N}(0, \sigma_\delta^2)$. Specific models contain random effects with distributions $\nu_s \sim_{\text{i.i.d.}} \text{N}(0, \sigma_{\nu 1}^2)$, $\nu_{is} \sim_{\text{i.i.d.}} \text{N}(0, \sigma_{\nu 2}^2)$, $\nu_{ts} \sim_{\text{i.i.d.}} \text{N}(0, \sigma_{\nu 3}^2)$, $\nu_{its} \sim_{\text{i.i.d.}} \text{N}(0, \sigma_{\nu 4}^2)$. In the "a" models $\gamma_t \sim \text{RW1}(\sigma_\gamma^2)$ and in the "b" models $\gamma_t \sim \text{RW2}(\sigma_\gamma^2)$

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	$\mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it} + \nu_s + \nu_{ts} + \nu_{is} + \nu_{its}$

4.2. *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 or 2 (models type "a" or "b"), to pick up local temporal smooth fluctuations, for $t = 1, \ldots, T = 6$ time periods. Five-year time periods were chosen because survey-specific regional sample sizes can be quite small. The UN IGME has only recently moved to annual estimates at the national level because the sample size of recent DHS has increased [Pedersen and Liu (2012)]. We are combining recent and older DHS at a regional level, and thus sample sizes are not sufficiently large to produce reliable annual estimates.

There are also two spatial terms, corresponding to the convolution model of Besag, York and Mollié (1991). The independent random effects are denoted θ_i and the intrinsic conditional autoregressive (ICAR) terms are ϕ_i for $i = 1, \ldots, I = 21$ 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). Type II–IV interaction models were considered, which include spatial and/or temporal structure on the prior for δ_{it}, but these models did not substantially modify estimates, so Type I was selected for parsimony.

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), the five TDHS and THMIS surveys cover all 21 regions over the different time periods they were administered and the HDSS sites contribute data for one region each in the last three time periods. 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 Rue and Held (2005).

4.3. *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 Gamma($a_{\text{MARG}}, b_{\text{MARG}}$) priors for precisions [Wakefield (2009)] with $a_{\text{MARG}} = 0.5$, $b_{\text{MARG}} = 0.001488$. For the RW1, RW2 and ICAR models the precisions have *conditional* rather than *marginal* interpretations. Let $\mathbf{z}$ represent a random effect from an improper GMRF with "mean" $\mathbf{0}$ and "precision" $\tau^\star\mathbf{Q}$. Following the supplementary material of Fong, Rue and Wakefield (2010), we gain compatibility by calculating an approximate measure of the average marginal "variance" of $\mathbf{z}$ in the situation with $\tau^\star = 1$; 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, and 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, Rue and Wakefield (2010). This method is closely related to that later described by Sørbye and Rue (2014). The supplementary material contain `R` code for reproducing these prior specifications. For the Tanzania data this leads to gamma priors for the RW1 of $\tau_\gamma \sim \text{Gamma}(0.5, 0.00153)$, for the RW2 of $\tau_\gamma \sim \text{Gamma}(0.5, 0.00286)$, and for the ICAR of $\tau_\phi \sim \text{Gamma}(0.5, 0.00360)$.

4.4. *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 (see supplementary material for further details). The hierarchical Bayesian space–time models were fitted using the Integrated Nested Laplace Approximation (INLA) [Rue, Martino and Chopin (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, Rue and Wakefield (2010), Held, Schrödle and Rue (2010) and Schrödle and Held (2011).

4.5. *Model selection.* In Table 1 we describe twelve plausible random effects specifications (allowing for RW1 or RW2 models). 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 twelve 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

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

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

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

and good models will have relatively high values of LCPO. Held, Schrödle and Rue (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

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

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

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

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).

5. Applying methods to household surveys and HDSS sites in Tanzania. We fit models Ia–VIb (as summarized in Table 1) to the Tanzania survey data and Table 2 provides the summaries of various model comparison summaries. Model Vb is the favored model according to both the DIC, LCPO, and log of the normalizing constant criterion. Results for models Vb and VIb are very similar, but we see from the effective number of parameters that even though the number of 3-way

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*, $\overline{D}$; *LCPO is defined as* $\sum_{its} \log(\text{CPO}_{its})$. *In the "a" models* $\gamma_t \sim \text{RW1}(\sigma_\gamma^2)$ *and in the "b" models* $\gamma_t \sim \text{RW2}(\sigma_\gamma^2)$

Model	No pars	log p(y)	p_D	$\overline{D}$	DIC	LCPO
Ia	181	−297.3	74.5	409.3	483.8	−294.5
IIa	189	−291.0	80.1	384.2	464.3	−287.3
IIIa	313	−244.1	118.9	221.8	340.7	−193.5
IVa	223	−288.6	88.6	367.5	456.2	−283.4
Va	347	−241.2	121.8	210.1	332.0	−183.1
VIa	920	−241.4	134.5	199.4	334.0	−183.9
Ib	181	−293.3	74.2	409.1	483.3	−293.7
IIb	189	−287.0	79.8	383.9	463.7	−286.4
IIIb	313	−239.9	118.6	221.7	340.3	−192.9
IVb	223	−284.5	88.2	367.4	455.6	−282.5
Vb	347	**−236.9**	121.6	209.9	**331.5**	**−183.1**
VIb	920	−237.6	133.3	200.2	333.4	−183.4

interaction random effects is 573, there are only 13 effective parameters due to the closeness of the interactions to zero. Hence, from this point onward we shall report summaries with respect to model Vb. We begin by summarizing the posterior distribution, and then describe regional trends.

5.1. *Summarizing the posterior distribution*. Table 3 provides numerical summaries and the proportion of total variation explained by each random effect. 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 RW2 and ICAR models. The structured temporal and unstructured spatial random effects explain 77% 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.

5.2. *Model validation*. To validate the model, we removed all of the observations in area i for time point t and then generated 95% intervals around the posterior mean ${}_5\tilde{q}_{0,it}$ using the variance of the observed response, defined as $\tilde{S}_{its}^2 = \tilde{\sigma}_{it}^2 + \widehat{V}_{\text{DES,its}}$, where $\tilde{\sigma}_{it}^2$ is the variance of the posterior distribution of $\text{logit}({}_5\tilde{q}_{0,it})$ and $\widehat{V}_{\text{DES,its}}$ is the design-based variance described in Section 3. This

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 RW2 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	0.002 (0.001, 0.012)	1.3
σ_γ^2	RW2 time	0.009 (0.002, 0.054)	46.0
σ_θ^2	Indept space	0.068 (0.033, 0.133)	31.3
σ_ϕ^2	ICAR space	0.017 (0.002, 0.378)	4.9
σ_δ^2	Indept space–time interaction	0.005 (0.001, 0.013)	2.3
$\sigma_{v_s}^2$	Indept survey	0.002 (0.001, 0.013)	1.4
$\sigma_{v_{st}}^2$	Indept survey–time interaction	0.004 (0.001, 0.011)	2.0
$\sigma_{v_{si}}^2$	Indept survey–space interaction	0.024 (0.015, 0.038)	10.9

was completed for the 21 regions and 6 time points (figures shown in the supplementary material). Intervals contained the design-based estimates 92.5% of the time overall. Time/area-specific coverages range from 89.9–96.9% and the coverage for the final time point is 93.2%.

5.3. *Regional estimates and projections*. For region i and 5-year period t, estimates, projections and credible intervals of U5MR are taken from posterior draws of

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

Figure 1 shows maps of the posterior median estimates of child mortality (per 1000 births) by region for the six observed 5-year time periods. Child mortality has decreased markedly over the 30-year period considered, but overall more than 5% of infants still die before they turn 5, and there are strong regional differences. Figures 2 and 3 display the observed direct estimates and smoothed results for the Morgoro and Pwani regions, respectively. Additionally, each plot shows the projected U5MR for the 2010–2014 time period. The direct estimates have a great deal of variability between surveys, especially for the first four time points, and design-based intervals are very wide. Smoothed rates and projections for all 21 regions are located in the supplementary material.

6. 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 Tan-

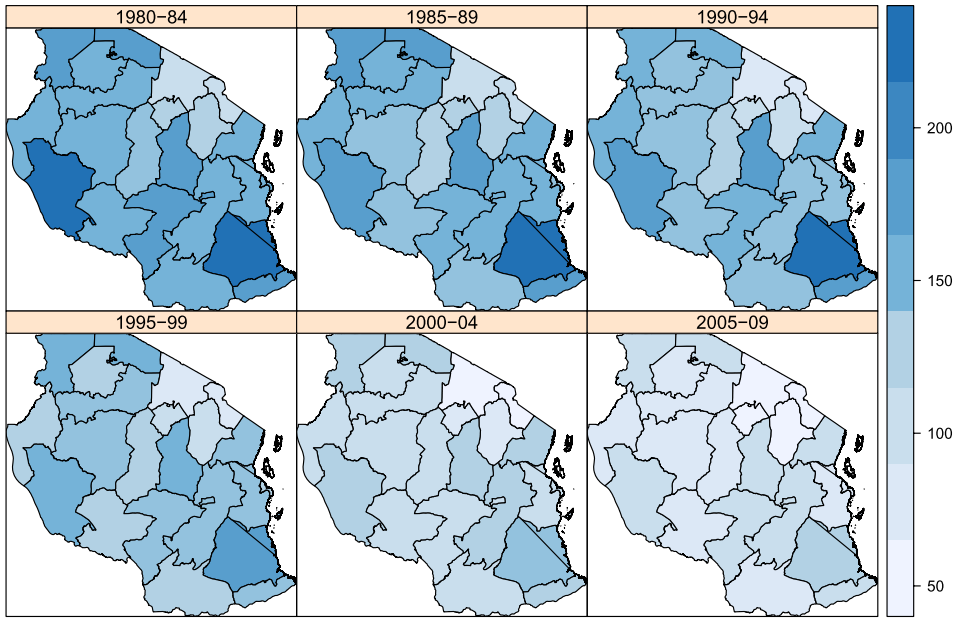

FIG. 1. *The solid line represents five-year model-smoothed estimates of* $_5q_0$ *in Pwani region, TZA, with* 95% *confidence intervals as vertical lines. The dashed lines display the five-year direct estimates from the region by household survey and surveillance site, with* 95% *confidence intervals as vertical lines.*

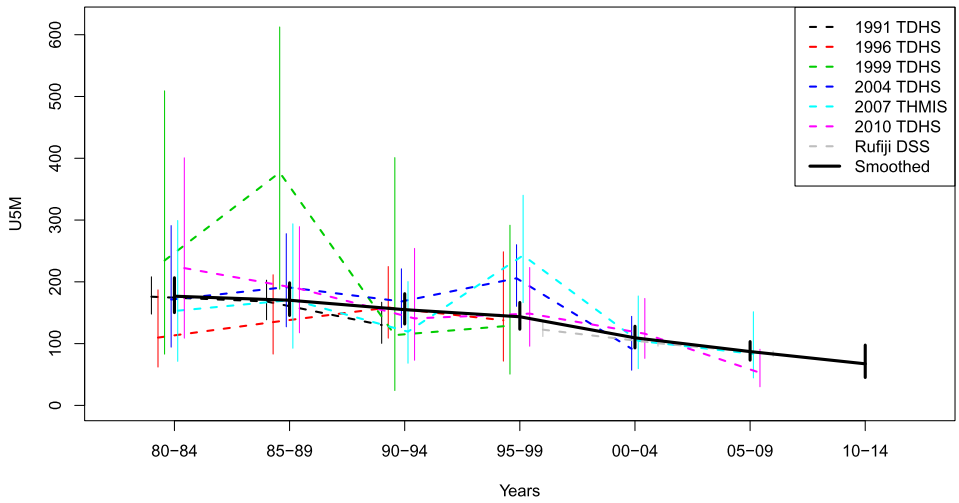

FIG. 2. *Regional five-year direct and model-based smoothed of* $_5q_0$ *in Pwani, TZA, with* 95% *confidence intervals.*

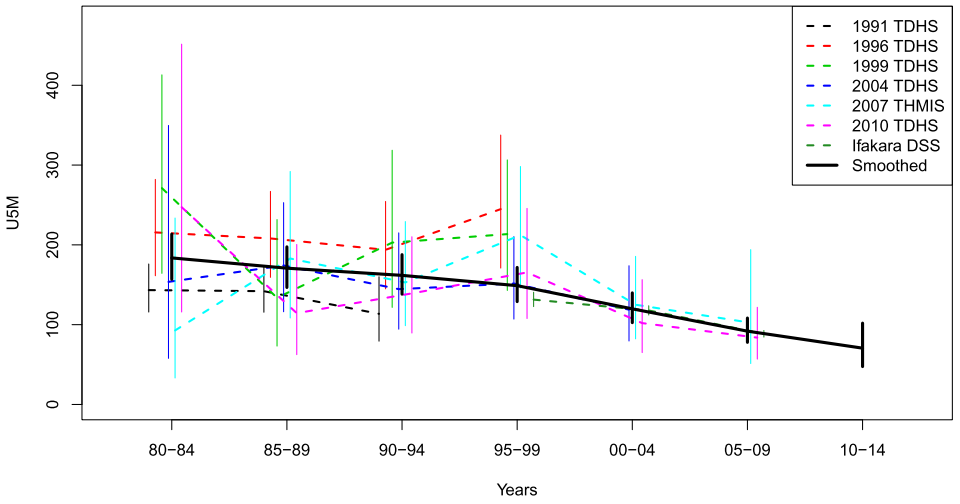

FIG. 3. *The solid line represents five-year model-smoothed estimates of* $_5q_0$ *in Morogoro region, TZA, with* 95% *confidence intervals as vertical lines. The dashed lines display the five-year direct estimates from the region by household survey and surveillance site, with* 95% *confidence intervals as vertical lines.*

zania over 1980–2009 using data from household surveys and surveillance sites. 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 material contain example code. As an example, fitting the most complex model for the Tanzania data took just 18.7 seconds on a Macbook Pro.[3]

In our hierarchical modeling approach, we explicitly acknowledge the weights by taking as (pseudo-)likelihood the (design-based) sampling distribution of the estimator. In the supplementary material we illustrate the effect of the weights on both the estimates and the standard errors. Another use of our model is for prediction, with the RW2 terms drawn from the relevant conditional distribution.

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 material contain details of a sensitivity study that we performed for the Tanzania data.

An integral part of our method involves calculating and pooling estimates of child mortality from household 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 HDSS sites. As Figures 2 and 3 make clear, the central estimates

[3]Processor: 2.9 GHz Intel Core i7; memory: 8 GB 1600 MHz DDR3.

from the two different data collection schemes are very similar. This adds more weight to similar findings by others [Byass et al. (2007), Fottrell, Enquselassie and Byass (2009), Hammer et al. (2006)] and reduces concerns about the Hawthorne effect preventing measures of child mortality from HDSS sites from being more widely relevant, that is, 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 approach for producing subnational estimates of child mortality and other key health, demographic and development indicators. However, countries with a substantial HIV/AIDs burden may suffer from underreporting biases. The UN IGME preprocesses data in a number of countries, including Tanzania, to take account of underreporting biases because of HIV/AIDS. We base our analysis on direct subnational estimates of U5MR, and so do not adjust for this bias, but our smoothed results do not differ substantially from the UN results at the national level and so we believe that any bias from this source will be small.

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, for example, 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.

Acknowledgments. We would like to acknowledge the hard work and commitment of the HDSS field teams and the data team, and thank the residents in the surveillance areas for offering their time for interviews and sharing their valuable personal information over the years.

SUPPLEMENTARY MATERIAL

Supplement to "Space–time smoothing models for complex survey data: Small area estimation for child mortality" (DOI: 10.1214/15-AOAS872SUPP; .pdf). The organization of the supplementary material is as follows. In Section 1 we provide the details of the discrete survival model. In Section 2 we provide the derivation of the standard error for U5M. Section 3 describes a simulation study aimed to test the coverage performance of the derived standard error against the

jackknife standard error used by DHS. In Section 4 we describe the hyperprior specifications for the Bayesian hierarchical model. Section 5 provides a summary of the posterior distribution of the random effects. In Section 6 we provide a comparison of weighted and unweighted direct estimates of U5M. In Section 7 we have included some exploratory analysis looking at the rates and magnitude of regional decreases in U5M and how they relate to the fourth millennium development goal of two thirds reduction in child mortality by 2015. The results of our model validation are presented in Section 8. Lastly, Section 9 includes example R code for the analyses.

REFERENCES

ALKEMA, L. and NEW, J. R. (2014). Global estimation of child mortality using a Bayesian B-spline Bias-reduction model. *Ann. Appl. Stat.* **8** 2122–2149. MR3292491

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** e101112.

ALLISON, P. (1984). *Event History Analysis*: *Regression for Longitudinal Event Data*. Number 46. Sage, Thousand Oaks, CA.

BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–59. MR1105822

BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51** 279–292. MR0731144

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** 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), Univ. Washington.

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 Med.* **10** 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.

DWYER-LINDGREN, L., KAKUNGU, F., HANGOMA, P., NG, M., WANG, H., FLAXMAN, A. D., MASIYE, F. and GAKIDOU, E. (2014). Estimation of district-level under-5 mortality in Zambia using birth history data, 1980–2010. *Spat. Spatiotemporal Epidemiol.* **11** 89–107.

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** 163–168.

GELMAN, A. (2007). Struggles with survey weighting and regression modeling. *Statist. Sci.* **22** 153–164. MR2408951

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 Trop.* **98** 212–218.

HELD, L., SCHRÖDLE, B. and RUE, H. (2010). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA. In *Statistical Modelling and Regression Structures* 91–110. Physica-Verlag/Springer, Heidelberg. MR2664630

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

INDEPTH NETWORK (2014). Health and demographic surveillance systems. Available at 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** 129–136.

KNORR-HELD, L. (2000). Bayesian modelling of inseparable space–time variation in disease risk. *Stat. Med.* **19** 2555–2567.

LOHR, S. L. (2010). *Sampling*: *Design and Analysis*, 2nd ed. Brooks/Cole, Cengage Learning, Boston, MA. MR3057878

LUMLEY, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software* **9** 1–19.

MERCER, L., WAKEFIELD, J., CHEN, C. and LUMLEY, T. (2014). A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spat. Stat.* **8** 69–85. MR3326822

MERCER, L. D., WAKEFIELD, J., PANTAZIS, A., LUTAMBI, A., MASANJA, H. and CLARK, S. (2015). Supplement to "Space–time smoothing of complex survey data: Small area estimation for child mortality." DOI:10.1214/15-AOAS872SUPP.

PARIS21 (2014). Paris21: Partnership for statistics in development in the 21st century. Available at 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** e1001289.

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 Applications*. *Monographs on Statistics and Applied Probability* **104**. Chapman & Hall/CRC, Boca Raton, FL. MR2130347

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602

RUTSTEIN, S. O. and ROJAS, G. (2006). *Tanzania Demographic and Health Survey* 1996. ORC Macro, Calverton, MD.

SCHRÖDLE, B. and HELD, L. (2011). Spatio-temporal disease mapping using INLA. *Environmetrics* **22** 725–734. MR2843139

SØRBYE, S. H. and RUE, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spat. Stat.* **8** 39–51. MR3326820

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 583–639. MR1979380

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2014). The deviance information criterion: 12 years on. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 485–493. MR3210727

UN (2000). Millennium development goals. Available at http://www.un.org/millenniumgoals/.

UN (2014a). Civil registration and vital statistics coverage. Available at http://unstats.un.org/unsd/demographic/CRVS/CR_coverage.htm.

UN (2014b). Data revolution for sustainable development. Available at http://www.un.org/apps/news/story.asp?NewsID=48594#.VEVQpoctuvJ.

UN (2014c). Millennium development goal number 4: Reduce by two thirds, between 1990 and 2015, the under-five mortality rate. Available at http://www.un.org/millenniumgoals/childhealth.shtml.

UN (2014d). The post-2015 development agenda. Available at http://www.post2015hlp.org/the-report/.

UN (2014e). Sustainable development goals. Available at http://sustainabledevelopment.un.org/owg.html.

USAID (2014). Demographic and health surveys. United States Agency for International Development. Available at http://www.dhsprogram.com.

WAKEFIELD, J. (2009). Multi-level modelling, the ecologic fallacy, and hybrid study designs. *Int. J. Epidemiol.* **38** 330–336.

WANG, H., LIDDELL, C. A., COATES, M. M., MOONEY, M. D., LEVITZ, C. E., SCHUMACHER, A. E., APFEL, H., IANNARONE, M., PHILLIPS, B., LOFGREN, K. T. et al. (2014). Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: A systematic analysis for the global burden of disease study 2013. *The Lancet* **384** 957–979.

WORLD BANK AND WORLD HEALTH ORGANIZATION (2014). Global civil registration and vital statistics scaling up investment plan 2015–2024. Available at 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** 741.

L. D. MERCER
DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
USA
E-MAIL: mercel@uw.edu

J. WAKEFIELD
DEPARTMENT OF STATISTICS
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
USA
E-MAIL: jonno@uw.edu

A. PANTAZIS
S. CLARK
DEPARTMENT OF SOCIOLOGY
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
USA
E-MAIL: apantazi@u.washington.edu
samclark@u.washington.edu

A. M. LUTAMBI
H. MASANJA
IFAKARA HEALTH INSTITUTE
DAR ES SALAAM
TANZANIA
E-MAIL: alutambi@ihi.or.tz
hmasanja@ihi.or.tz