

Use of Space Time Model for Forecasting Mortality due to Malaria: A Case of Ifakara and Rufiji Health and Demographic Surveillance System Sites

Majige Selemani^{1,*}, Sigilbert Mrema²

¹Department of Graduate Studies, Eastern Africa Statistical Training Center, Dar es Salaam, Tanzania

²Ifakara Health Institute (IHI), Dar es Salaam, Tanzania

Email address:

majiges@yahoo.com (M. Selemani), smrema@ihi.or.tz (S. Mrema)

*Corresponding author

To cite this article:

Majige Selemani, Sigilbert Mrema. Use of Space Time Model for Forecasting Mortality due to Malaria: A Case of Ifakara and Rufiji Health and Demographic Surveillance System Sites. *International Journal on Data Science and Technology*. Vol. 4, No. 1, 2018, pp. 24-34.

doi: 10.11648/j.ijdst.20180401.14

Received: May 4, 2017; **Accepted:** June 6, 2017; **Published:** June 7, 2018

Abstract: Malaria is a leading cause of morbidity and mortality in developing countries especially in rural areas where local resources are limited. Accurate disease forecasts can provide information to public and clinical health services to design targeted interventions for malaria control that make effective use of limited resources. Using verbal autopsy data, space-time model was used to forecast mortality due to malaria. The study used longitudinal data which were collected from Rufiji and Ifakara Health Demographic Surveillance System (HDSS) sites for the period of 1999 to 2011 and 2002 to 2012 respectively to assess models. The models included environmental factors and mosquito net ownership as predictor variables for mortality due to malaria. Deviance information criteria (DIC), logarithm score and root mean square error (RMSE) were used to assess the goodness of fit and forecasting accuracy of the models. The results indicate that the model included spatial and temporal random effect terms had small deviance information criteria, logarithm score and root mean square error. This model was the best model for forecasting and prediction of mortality due to malaria in both HDSS sites. In addition, mosquito net ownership and rainfall were significantly associated with mortality due to malaria. The model with spatial and temporal random effect terms is useful tool to provide reasonably reliable forecasts for mortality due to malaria. This might help to design appropriate strategies for targeting malaria control. On the other hand, including spatially and temporal varying random terms in the model is necessary and good strategy for modelling mortality due to malaria.

Keywords: Space Time Model, Mortality due to Malaria, Forecasting

1. Introduction

Resources allocation and malaria control interventions need to be spread in space and time to enhance effective malaria control [1]. Having a forecasting system in place may help to have a more focused approach with positive impact on the resource allocation for malaria control over space and time. However, there are fewer studies which have investigated on forecasting models for mortality due to malaria [2, 3]. The studies indicate the need for improved epidemic early warning by incorporating important predictors such as meteorological factors.

Malaria incidence forecasting models have been developed

in many endemic countries [5-9]. These models mostly used time series models and other use climate related predictors, such as rainfall, temperature and normalized difference vegetation index. However, these models did not include other important predictors for mortality related to malaria such as mosquito nets, availability of anti-malarial and drug resistance. In addition, these models did not include spatial effect which could capture heterogeneity in malaria transmission. The accuracy of forecasts may be weakened due to transmission-reducing predictors were not considered in the models [7].

Health and demographic surveillance system (HDSS) data, however, are becoming increasingly available with the

routinely collecting data and adopting the verbal autopsy approach [8]. Verbal autopsy (VA) approach is an indirect method of determining causes of death from information on symptoms, signs and circumstances preceding death, obtained from the deceased's caretakers. VA has been widely used as a method for determining causes of death in the country where vital registration system is lacking and majority of deaths occur outside the formal health care system [9]. The VA is widely used to provide more information on deaths whose causes are unknown. Several studies have evaluated the validity [10, 11] and accuracy of VA in determining the causes of death and malaria specific cause of death [12, 13]. These studies concluded that VA is the reliable source of data that can be used to estimate specific causes mortality in all age group for public health [13].

This study assess space time models for forecasting mortality due to malaria for resource allocation and optimization of the effects of malaria control interventions over space and time.

2. Methods

2.1. Study Design and Study Area

This was longitudinal study design using secondary data collected from Rufiji and Ifakara Rural Health and Demographic Surveillance System (HDSS) sites [14, 15] in Tanzania. Rufiji HDSS is located in Rufiji District, Coast Region while Ifakara HDSS is located in part of Kilombero and Ulanga District in Morogoro Region. These two HDSS sites were selected because they are among HDSS sites which continuously collect large amount of longitudinal data in well defined geographical areas over time in Tanzania. These HDSS sites are characterized by malaria transmission endemic after the long rain season. The description of the study area is detailed elsewhere [14, 15].

2.2. Data Source

This study utilized secondary data which were prospectively collected from 1999 to 2011 and 2002 to 2012 for Rufiji and Ifakara HDSS sites respectively. The two HDSS sites consistently records pregnancy outcomes, deaths and migrations by visiting households once every four months since 1997 in Ifakara HDSS and 1998 in Rufiji HDSS. In addition, the HDSS sites collect information on social economic status including household ownership of mosquito net annually. The study period covers national scale up of ITNs that occurred in the country [16]. The mosquito nets in this study included both untreated and treated nets (ITNs). Yearly numbers of malaria deaths were extracted from the Rufiji and Ifakara HDSS database covering a period of 1999 to 2011 and 2002 to 2012 respectively. The details on data collection, management and credibility in HDSS have been describe elsewhere [14, 15].

The Verbal autopsy procedures and collect data used in this study are described in detailed elsewhere [8, 17]. In the

HDSS, deaths were captured during rounds of data updates. Then HDSS field interviewers visited the deceased's home after a grieving period to administer a verbal autopsy questionnaire. A face to face interview was administered to relatives or caregivers who were closely associated with the deceased during the period leading to his or her death. The questionnaire was used to explore the identity of the deceased and established the sequence of events leading to death, including symptoms and signs of the illness before death. Verbal autopsy was carried out since 1998 in Rufiji HDSS and 2002 in Ifakara HDSS. The verbal autopsy forms are independently reviewed by two physicians according to a list of causes of death based on the 10th revision of the International Classification of Diseases (ICD-10). Causes of death (main, immediate, and/or contributing) are coded to be consistent with the ICD-10 [18]. Malaria deaths are coded as direct cause of death when malaria is the underlying cause of death or indirect cause of death when malaria is one of the several diseases leading to death but the death is attributed by a different cause) [19].

Climate data were obtained from Tanzania Meteorological Authority (TMA) head office in Dar es Salaam. These data includes monthly mean rainfall, and maximum and minimum temperature. TMA provides meteorological services, weather forecasts, climate services, and warnings including daily forecast information for each region in Tanzania. The climate data are collected through different gauges located in different stations in each district. In Ifakara HDSS, villages located in Kilombero district used climate data from Kilombero Agricultural Training and Research Institute (KATRIN) weather station. Other villages in Ifakara HDSS used Mahenge weather station data for villages located in Ulanga District. In Rufiji District, Utete and Kibiti are the two weather stations which records climate data. Kibiti weather station is located in Rufiji HDSS site which provided data used in this study.

Remote sensing data were extracted from MODerate-resolution Imaging Spectroradiometer (MODIS) on board NASA's terra satellite. The remote sensing data includes Normalized Difference Vegetation Index (NDVI) and processed from MODIS (MOD13A3) using monthly composite images at a 1 km x 1 km resolution. Vegetation indices are used for global monitoring of vegetation conditions and are used in products displaying land cover and land cover changes. ERDAS Imagine software version 10.1 was used for processing the satellite images, and ArcGIS version 10 was used for spatial analyses. Mean NDVI were calculated for each village in each year within the study area to link with mortality due to malaria and ownership of mosquito net for each village.

2.3. Data Processing and Analysis

Aggregated yearly numbers of malaria deaths were extracted from the HDSS database by each village for the period of 1999 to 2011 and 2002 to 2012 in Rufiji and Ifakara HDSS respectively. The data was aggregated yearly to reduce zero count in each village for appropriate

assumption used in Poisson model. Time at risk (person-years) contributed by each person was calculated from 1st January until exit or 31st December for each year. Exit from the study was due to migration (outside the HDSS area), death or end of the study. In a case where a person migrated to a different household location within the study area, time at risk was computed separately for new location and added to the total person time at risk. The outcome of interest was a total yearly death due to malaria for specific age groups. Age was categorized into two groups i.e. under five and five and above.

The outcome variable was aggregated at village level to capture heterogeneity for mortality in the study area as reported in previous studies [20, 21]. Direct estimates of mortality due to malaria rates were calculated in order to highlight spatial-temporal trends of mortality related malaria in the study areas. The mortality due to malaria rates were calculated by dividing the number of malaria deaths by the person-years of observation and were expressed per 1,000 person-years (py) for each village. Direct estimates of the mortality due to malaria rates were obtained as:

$$\theta_{ij} = \frac{Y_{ij}}{P_{ij}} \quad (1)$$

Where θ_{ij} is the mortality due to malaria rate at j time in i village, Y_{ij} is the number of malaria deaths at time j in village i. represents the number of person-years in i village at time j and represent the exposure to risk.

2.4. Space Time Model

The model was intended to capture both temporal and spatial features of mortality due to malaria trends or along with spatial-temporal interactions. The developed space time model will be useful to explain the patterns of malaria mortality rates in terms of environmental factors and ownership of mosquito nets. Previous studies have proposed hierarchical Bayesian spatial-temporal modelling along with Poisson model [22–24].

Spatial-temporal modelling approach within a hierarchical Bayesian framework was adopted. The spatial aspect of the modelling approach allows for taking into account similarities between values observed at locations across space. Similarly, the temporal aspect of the modelling approach allows for inference concerning temporal trends of changes in malaria mortality/deaths. Finally, the proposed model allows for spatial and temporal trend analysis of the data as well as considering the effect of predictor variables [25].

The Bayesian models requires the data, process and parameters to be modeled [26]. Using data extracted from HDSS database and other source, the covariate information included in this study for village i at year j were mosquito net ownership as exposure variable and Other variables were rainfall, temperature and NDVI. For this study i the number of village in the study, $I = 33$ for Rufiji and 25 for Ifakara HDSS and j for years, $J = 13$ for Rufiji HDSS and $J = 11$ for Ifakara HDSS.

The general specification of the hierarchical space-time model can be represented by the following hierarchical structure, which simultaneously model the spatial effect, temporal effect and space time interaction as described in [27]:

$$\log(\theta_{ij}) = \alpha_0 + \sum_{k=1}^k \beta_k X_{kij} + v_i + \eta_i + \gamma_j + \psi_{ij} \quad (2)$$

Where θ_{ij} denotes the mortality due to malaria rate. Random terms v_i and η_i are spatial terms, γ_j is temporal random effect and ψ_{ij} is space-time random effects. Covariates information for village i at year j are denoted by X_{kij} with β_{kj} being the corresponding regression coefficients of k variables varying over time.

The random effects specifications for the spatial term includes unstructured noise that follows a normal distribution and structure term which modelled with the set of villages that adjacent to other villages by weight of neighbouring village. Weight is considered to be 1 when two villages share the same boundary and 0 otherwise. The prior distribution of the random effect term is a conditional autoregressive prior used to model the spatial dependence.

The temporal effect can be specified in a number of ways, which will determine the prior distribution that was assumed for the term γ_j . In this study time effect as first order autoregression AR (1) given by the partial autocorrelation function. The distribution of the interaction term ψ_{ij} is characterized by a precision matrix obtained as the Kronecker product of the precisions of structure and unstructured noise terms [28].

From the general equation in (2), five sub-models were fitted considering different ways of entering with each of the random terms and covariates in the model. Five models considered with different combination of random effects and covariates as follows:

$$M_1: \log(\theta_{ij}) = \alpha_0 + \sum_{k=1}^k \beta_k X_{kij} \quad (3)$$

$$M_2: \log(\theta_{ij}) = \alpha_0 + \sum_{k=1}^k \beta_k X_{kij} + v_i + \gamma_j \quad (4)$$

$$M_3: \log(\theta_{ij}) = \alpha_0 + \sum_{k=1}^k \beta_k X_{kij} + v_i + \eta_i + \gamma_j \quad (5)$$

$$M_4: \log(\theta_{ij}) = \alpha_0 + \sum_{k=1}^k \beta_k X_{kij} + v_i + \eta_i + \gamma_{1j} + \gamma_{2j} \quad (6)$$

$$M_5: \log(\theta_{ij}) = \alpha_0 + \sum_{k=1}^k \beta_k X_{kij} + v_i + \eta_i + \gamma_j + \psi_{ij} \quad (7)$$

Where M_1 represents a model with covariates only; M_2 is a model with covariates, a spatial random term and a temporal term; M_3 is a model with covariates, spatial component terms and a temporal term; M_4 is a model with covariates, spatial component terms, and two temporal terms (first and second order); M_5 is a model with covariates, spatial, temporal, and an interaction term.

The analysis in this study used the Integrated Nested Laplace Approximation (INLA) R-package in R software available at <http://www.r-inla.org> to fit five models defined in equation 3 to 7. INLA is a recent computational approach used to perform approximate Bayesian inference based on an efficient combination of Laplace approximations and

numerical integration. The INLA is designed for latent Gaussian models; a very wide and flexible class of models ranging from generalized linear mixed to spatial and spatio-temporal models. The package does not sample from the posterior distribution like Markov Chain Monte Carlo (MCMC) [29]. This approach provides approximates the posterior with a closed form expression and more precise estimates in seconds or minutes. Also, the INLA compute model comparison criteria and various predictive measures so that models can be compared without convergence and mixing problems. The method is best suited to Bayesian hierarchical models for which there are large number of unknown parameters following a Gaussian Markov random field and a small number of hyperparameters, with a specific form of prior covariance on the parameters. This analysis used data for 1999 to 2010 in Rufiji HDSS and 2002 to 2011 in Ifakara HDSS. Data for 2011 and 2012 in Rufiji and Ifakara respectively were used for validation.

The best fitted model was used to predict malaria mortality for 2011 in Rufiji HDSS and 2012 for Ifakara HDSS for validation. Furthermore, the best fitted model was used to forecast malaria mortality for five years 2012 to 2016 in Rufiji HDSS and 2013 to 2017 for Ifakara HDSS. The forecasted mortality due to malaria was multiplied by person-years to represent the number of malaria deaths.

2.5. Model Selection

The model goodness of fit and forecasting accuracy was assessed using commonly known methods for hierarchical Bayesian models called the deviance information criterion (DIC) [30], root mean square error (RMSE) and logarithm score were used to select the best model. Based on the DIC criterion and RMSE, the model with relatively lower DIC values indicate a better fit to the data compared to models with higher DIC values. Similarly the model with small value

of RMSE indicates a better model for prediction and forecasting. The logarithm score also used to assess the predictive quality of the model [27]. A smaller value of the logarithmic score indicates a better prediction quality of the model.

2.6. Ethical Statement

This study was approved by Institutional Review Board of Ifakara Health Institute (IHI/IRB/No: 31-2014). Also, Medical Research Coordinating Committee (MRCC) of the National Institute for Medical Research (NIMR) approved for the establishment of Rufiji and Ifakara Health and Demographic Surveillance System (HDSS) sites. For each household visit, informed verbal consent was obtained from head of the family or eligible adult among the family members aged 18 years and above who was able to provide with more information at household.

3. Results

Figure 1 and 2 presents the trend of mortality due to malaria in Rufiji and Ifakara HDSS. The results indicate that mortality due to malaria decreases from 11.2 per 1000 person-years (py) in 1999 to 5.3 per 1000 py in 2011 for under fives children in Rufiji HDSS (Figure 1). In Ifakara HDSS, the mortality due to malaria decreases from 6.3 per 1000py in 2006 to 4.0 per 100py in 2012 for under five children age (Figure 1). Furthermore, the mortality due to malaria for all ages decreases from 3.5 per 1000py in 1999 to 2.2 per 1000py and 2.3 per 1000py to 1.2 in 2012 in Rufiji and Ifakara HDSS sites respectively. Table1 presents the percentage of households owned at least one mosquito net and environmental factors in Rufiji and Ifakara HDSS. The results show the increase of percentage of households owned at least one mosquito net in both HDSS site.

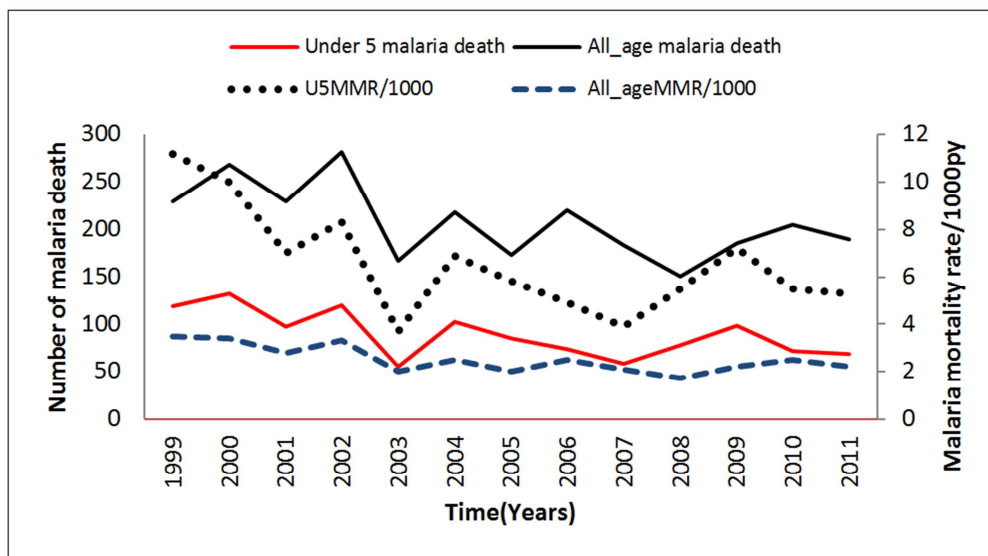


Figure 1. Annual mortality due to malaria in Rufiji HDSS.

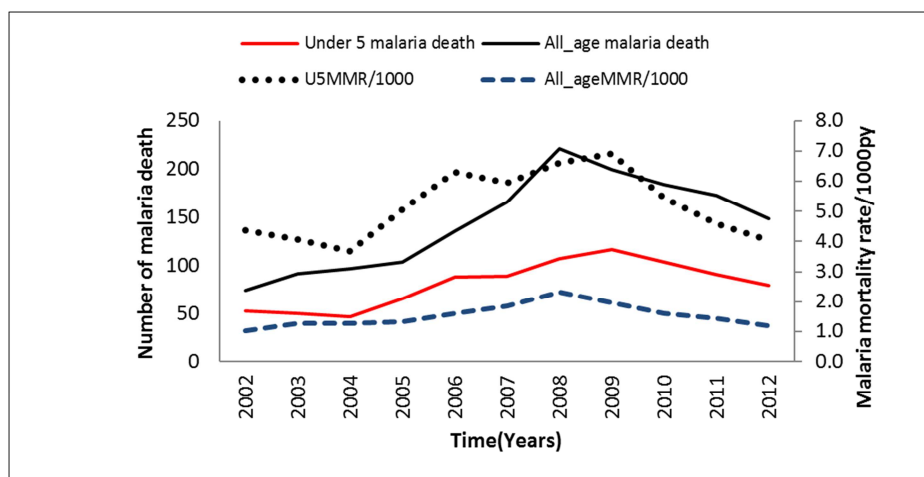


Figure 2. Annual mortality due to malaria in Ifakara HDSS.

Figure 3 and 4 presents the annual mortality due to malaria for selected villages for under-five children age in Rufiji and Ifakara HDSS for the study period respectively. The results indicate that there was heterogeneity in mortality due to malaria across villages. Some villages reported low mortality due to malaria throughout the entire study period while others had high mortality due to malaria rate. Furthermore, the results show that all villages experienced an overall decreasing trend in annual mortality due to malaria in Rufiji HDSS site.

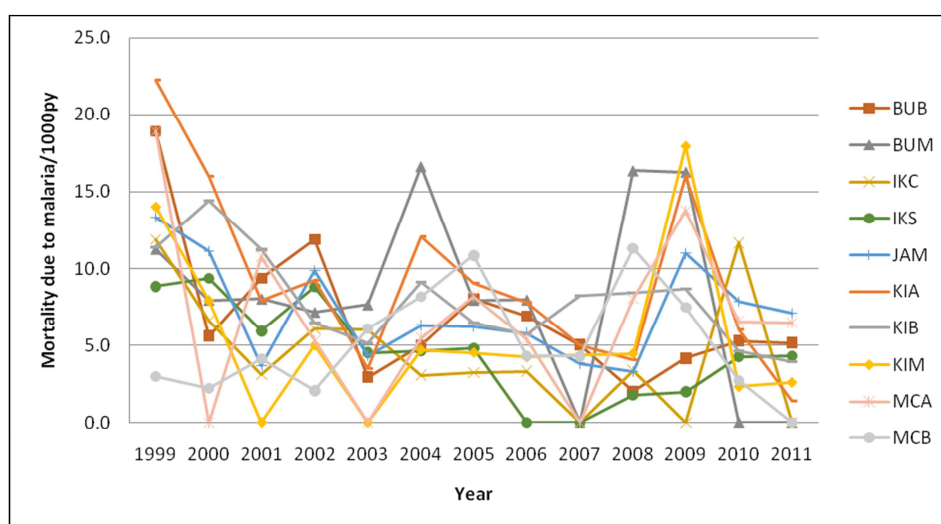


Figure 3. Annual mortality due to malaria by village in Rufiji HDSS.

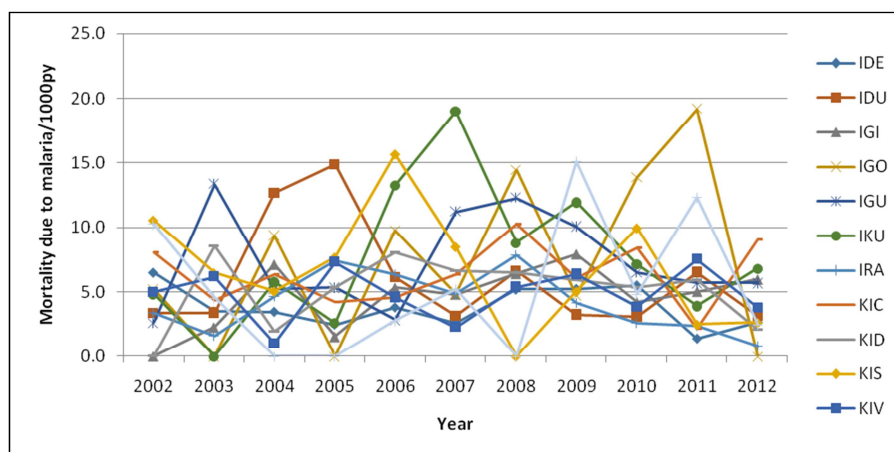


Figure 4. Annual mortality due to malaria by village in Ifakara HDSS.

Table 1. Ownership of mosquito net and environmental factors in Rufiji and Ifakara HDSS.

Rufiji HDSS				
Year	% mosquito net ownership	Annual mean Rainfall	Annual mean Temperature	Annual mean NDVI
1999	15.3	1032.2	312.1	0.58
2000	21.3	1020.8	314.3	0.55
2001	26.3	806.5	316.9	0.55
2002	31.9	1210.0	319.0	0.54
2003	37.3	664.8	328.4	0.46
2004	42.3	1171.1	321.1	0.43
2005	42.1	547.6	318.7	0.60
2006	60.0	935.3	315.4	0.61
2007	77.5	732.5	319.4	0.54
2008	77.7	898.5	315.4	0.51
2009	74.5	730.5	321.9	0.47
2010	68.1	601.7	323.1	0.54
2011	92.5	652.6	324.8	0.54
Ifakara HDSS				
2002	82.6	2033.7	244.9	0.49
2003	84.9	1194.4	256.0	0.47
2004	84.2	1826.9	253.3	0.47
2005	91.9	1378.1	258.1	0.58
2006	92.2	1862.7	253.9	0.49
2007	91.8	1580.1	256.2	0.56
2008	90.2	1721.5	251.7	0.55
2009	90.0	1519.5	258.7	0.52
2010	92.8	1565.7	245.4	0.59
2011	95.4	1842.8	256.6	0.58
2012	95.6	1217.0	253.6	0.50

3.1. Model Comparison

Table 2. Variance inflation factors for covariates.

Variables	VIF	Tolerance
Rufiji HDSS		
Annual rainfall	1.56	0.64
Average temperature	1.84	0.55
Mean NDVI	1.38	0.73
Ifakara HDSS		
Annual rainfall	1.08	0.92
Average temperature	1.08	0.92
Mean NDVI	1.00	1.00

Mortality due to malaria rate in 33 villages and 25 villages for Rufiji and Ifakara HDSS respectively were incorporated in the model fitting. Furthermore, annual mean rainfall, average temperature, mean NDVI and annual percentage of ownership of mosquito net in each village were included in the model fitting as covariates. Table 2 shows the results on

the variance inflation factors values and indicates less than 10 between covariates (rainfall=1.56, temperature =1.84 and NDVI=1.38) in Rufiji HDSS. Also, the variance inflation factors were less than 10 (rainfall=1.08, temperature=1.08 and NDVI=1.00) in Ifakara HDSS (Table 2). This indicates that there was no evidence for collinearity between covariates in the regression model.

Table 3 presents a summary for the deviance information criteria, logarithm score, mean square error and root mean square error for five models. Based on DIC values, Model M_2 with covariates, spatial and temporal terms gave a better fit than other models in both HDSS for all age and under five children. Model M_4 had the worst performance with highest DIC value in Rufiji HDSS followed by model M_1 . In Ifakara HDSS, model M_2 performed better than other models while model M_1 was having highest DIC value followed by model M_4 . Based on these results, it is clear that including spatially and temporal random terms in the model is necessary.

Table 3. Deviance information criterion and mean square error for the five models.

	M_1	M_2	M_3	M_4	M_5
Rufiji HDSS					
All age					
DIC	1794.61	1744.07	1766.24	1799.04	1766.07
MSE	9.09	6.56	6.70	8.26	6.67
RMSE	3.01	2.56	2.59	2.87	2.58
Log score	2.27	2.21	2.58	2.61	2.58
Under five					
DIC	1390.95	1375.87	1399.16	1399.17	1399.17
MSE	3.40	3.09	3.10	3.10	3.09
RMSE	1.84	1.76	1.76	1.76	1.76
Log score	1.72	1.69	2.08	2.08	2.09
Ifakara HDSS					

	M_1	M_2	M_3	M_4	M_5
All age					
DIC	1187.76	1111.59	1126.07	1142.55	1126.10
MSE	8.35	5.20	5.31	6.19	5.30
RMSE	2.89	2.28	2.30	2.49	2.30
Log score	2.38	2.23	2.26	2.31	2.26
under five					
DIC	969.04	953.93	968.2711	968.5318	968.3714
MSE	3.58	3.23	3.29	3.28	3.27
RMSE	1.89	1.80	1.81	1.81	1.81
Log score	1.93	1.91	1.94	1.94	1.94

Note: M_1 a model with covariates only, M_2 a model with covariates, a spatial random term and a temporal term, M_3 a model with covariates, spatial component terms and a temporal term, M_4 a model with covariates, spatial component terms, and two temporal terms (first and second order) M_5 a model with covariates, spatial, temporal, and an interaction term.

3.2. Prediction and Forecasting

Furthermore, Table 3 indicates that model M_2 had small value for logarithm score compared to other models for Rufiji and Ifakara HDSS respectively. The results also show that model M_2 had small value for RMSE compared to other models for both Rufiji and Ifakara HDSS based on one year validation. According to the DIC, logarithm score and RMSE values, model M_2 was considered as the best model for goodness of fit, prediction and forecasting.

Table 4 shows coefficients for the covariates incorporated in the selected best model and the results show that mosquito net ownership and rainfall were statistically significant associated with mortality due to malaria in both HDSS sites.

In the other hand, mosquito nets ownership was negatively associated with mortality due to malaria while rainfall was positively associated with mortality due to malaria in the study areas. Temperature and NDVI were not statistically associated with mortality due to malaria. The results also showed that there was a precision for temporal and spatial effect terms (Table 4). Furthermore, the results show a significant difference in temporal and spatial effect for mortality due to malaria for all age in Rufiji and Ifakara HDSS respectively. There was no spatial difference with respect to mortality due to malaria for under-five children in Rufiji HDSS as compared to Ifakara HDSS.

Table 4. Posterior mean, standard deviation and 95% credible interval (CI) for the best model (M_2) for forecasting malaria mortality and death.

	Rufiji HDSS			Ifakara HDSS		
Parameter	Mean	SD	95%CI	Mean	SD	95%CI
All age						
Intercept	-8.86	3.76	-16.66 -1.67	-3.90	2.25	-7.99-0.87
Mosquito net	-0.05	0.02	-0.09- -0.02	-0.13	0.04	-0.22--0.04
Rainfall	0.07	0.02	0.03-0.10	0.01	0.01	0.00-0.03
Average Temp	0.01	0.01	-0.01-0.03	-0.01	0.01	-0.03-0.01
Mean NDVI	0.05	0.04	-0.03-0.14	0.03	0.03	-0.03-0.09
Precision for temporal	74.97	64.55	10.98-245.06	25.69	18.03	5.30-72.40
Rho for temporal	0.63	0.24	0.05-0.94	0.66	0.20	0.16-0.93
Precision for spatial	21.10	12.82	6.74-54.64	87.45	101.04	11.69-347.50
Under five						
Intercept	-0.97	5.67	-12.58- 9.95	-4.59	2.32	-8.98-0.19
Mosquito net	-0.06	0.02	-0.10-0.014	-0.11	0.05	-0.20--0.01
Rainfall	0.07	0.03	0.02-0.13	0.01	0.01	0.00-0.04
Average Temp	-0.01	0.02	-0.05-0.02	-0.00	0.01	-0.02-0.02
Mean NDVI	0.04	0.05	-0.07-0.14	0.03	0.04	-0.05-0.10
Precision for temporal	29.21	21.00	5.84-83.97	74.97	64.55	10.98-245.06
Rho for temporal	0.59	0.24	0.01-0.93	0.63	0.24	0.05-0.94
Precision for spatial	28570.50	23350.8	3457.69-89189.08	21.10	12.82	6.74- 54.64

Table 4 and 5 presents forecasted mortality due to malaria for selected villages using best model M_2 for Rufiji and Ifakara HDSS respectively. Table 4 shows the selected villages with forecasted mortality due to malaria using identified best model M_2 . The forecasted mortality due to malaria for Nyamwimbe village in Rufiji HDSS was higher

in 2012 to 2016 (2.62 per 1000 person-years and 3.01 per 1000 person-years respectively) as compared to Kibiti A village (2.54 per 1000 person-year and 2.39 per 1000 person-years) respectively. The lowest mortality due to malaria forecasted was 1.54 and 1.42 per 1000 person years in Kinyanya village for 2012 and 2016 respectively (Table 4).

Table 5. Selected Villages with high and low forecasted malaria mortality using best model (M_2) for 2012-2016 in Rufiji HDSS.

Village	Year	Under five			All age		
		MMR/1000py	Malaria Deaths	95%CI	MMR/1000py	Malaria Deaths	95%CI
KibitiA	2012	4.59	3	2-5	2.54	12	8-16
Nyamwimbe	2012	2.32	1	0-2	2.62	3	1-4
Kinyanya	2012	4.70	2	1-3	1.54	4	2-6
KibitiA	2013	4.88	3	2-5	2.61	12	8-17
Nyamwimbe	2013	3.88	2	2-6	2.61	4	2-6
Kinyanya	2013	4.95	2	1-3	1.57	4	2-6
Machipi	2014	4.12	2	1-4	2.47	7	3-12
Mang'wi	2014	4.09	3	2-4	2.47	7	4-12
Nyamwimbe	2014	4.02	3	1-4	2.75	7	4-12
Mlanzi	2014	4.48	3	2-4	1.31	4	2-6
Mang'wi	2015	4.01	2	0-3	2.83	7	4-12
Nyamwimbe	2015	4.21	2	1-3	3.15	8	4-14
Kinyanya	2015	5.11	2	1-3	1.47	4	2-6
Nyamwimbe	2016	0.00	3	1-4	3.05	8	4-13
Machipi	2016	4.78	2	1-3	2.74	7	3-13
Mang'wi	2016	4.14	3	1-4	2.74	7	4-12
Kinyanya	2016	4.83	2	1-3	1.42	4	2-6

MMR= malaria mortality rate

In Ifakara HDSS, Lukolongo (1.49 per 1000person-years), Ikule (1.59 per 1000person-years) and Mpofu (1.95 per 1000person-years) villages with highest mortality due to malaria in Kilombero district for 2013 2014 and 2017 respectively (Table 6). The results also show that Idete village in Kilombero has lowest forecasted mortality due to malaria throughout the forecasted years. In Ulanga district, Kidungalo

and Igumbiro were forecasted with highest mortality due to malaria compared to other villages in 2013. Igumbiro village appeared to have high mortality due to malaria in all forecasted years. On the other hand Milola village was having with lowest malaria mortality forecasted in 2013 and Idunda in 2014. Iragua village appeared to have lowest forecasted mortality due to malaria in 2015, 2016 and 2017.

Table 6. Selected Villages with high and low forecasted malaria mortality using best model (M_2) for 2013-2017 in Ifakara HDSS.

		Under five			all age		
Village	Year	MMR/1000py	Malaria Deaths	95%CI	MMR/1000py	Malaria Deaths	95%CI
Kilombero District							
Idete	2013	4.44	4	2-5	0.93	6	4-10
Ikule	2013	3.05	2	1-3	1.49	7	4-10
Lukolongo	2013	3.85	2	1-3	1.49	9	5-13
Idete	2014	6.72	3	1-4	0.94	7	3-11
Ikule	2014	5.46	3	2-4	1.59	7	4-12
Lukolongo	2014	5.01	3	2-4	1.59	10	6-15
Idete	2015	4.22	3	2-5	0.87	6	4-9
Ikule	2015	4.24	3	2-4	1.79	6	4-9
Lukolongo	2015	4.25	3	1-3	1.55	11	7-15
Idete	2016	4.53	4	2-5	0.84	6	3-9
Ikule	2016	4.59	3	2-5	1.81	6	4-10
Mpofu	2016	4.18	2	1-3	1.99	8	5-13
Idete	2017	4.41	4	2-5	0.86	6	3-10
Mpofu	2017	4.07	2	1-3	1.95	8	4-13
Ulanga District							
Igumbiro	2013	4.79	3	2-5	1.43	5	2-7
Kidungalo	2013	4.84	3	2-5	1.43	10	6-15
Milola	2013	4.56	3	2-4	1.11	3	1-5
Igumbiro	2014	5.46	3	2-4	1.51	5	3-8
Idunda	2014	6.52	3	2-4	1.17	4	2-6
Igumbiro	2015	4.28	3	2-4	1.71	6	4-9
Iragua	2015	4.28	3	2-4	1.16	9	6-13
Igumbiro	2016	4.58	3	2-5	1.66	7	4-10
Iragua	2016	4.59	3	2-5	1.13	8	5-12
Igumbiro	2017	4.44	3	2-5	1.70	7	4-11
Iragua	2017	4.45	3	2-5	1.15	8	5-13

MMR= malaria mortality rate

4. Discussion

In this study, different models were compared for forecasting mortality due to malaria from historical malaria deaths data in the study areas. The findings indicate that model with spatial and temporal random terms was the best model for goodness of fit, prediction and forecasting malaria mortality. The finding is consistent with previous studies [31–34] which found potential use of spatial and temporal terms in the model.

Some studies which used Bayesian hierarchical framework for diseases mapping and ecological studies of health environment association [22, 33, 35, 36] found that spatial and temporal terms in the model are necessary if data are collected over space and time. The possible explanation for this could be due to the complex dependence patterns over space and over time of the occurrence of malaria deaths, and the inherent large stochastic variability due to rare events. In other hand, it may be due to inclusion of spatial and temporal random terms in the model to account other unmeasured confounding variables.

Estimating separately time trends in each area will not be efficient because it will be difficult to establish a baseline pattern separately for each area. In Bayesian approach, the study used the power of hierarchical modelling to borrow information over space and time in order to estimate typical predictable patterns for each area. Some previous studies also have indicated that the statistically advanced techniques using time series models may produce very good fit to the data but in post-sample forecast, they would not be robust enough to handle a possible change in behaviour of the series [37, 38].

This study included covariates information in the model for forecasting mortality due to malaria. The covariates information may have contributed significantly to variations in mortality due to malaria. In this study, the models included covariates information such as rainfall, temperature and NDVI which are important factors for malaria transmission [42], and mosquito net which reduces malaria transmission. The inclusion of these covariates in the model may not weaken the accuracy of forecasts because malaria transmission-reducing interventions are considered in the models [7]. Rainfall and ownership of mosquito net were independently associated with mortality due to malaria in the study areas. The result is in agreement with previous studies [39, 40] which found an association between malaria epidemics with changes in meteorological factors. The possible explanation for this may be due to the fact that rainfall provides breeding sites for mosquitoes and increases the humidity which enhances their survival and therefore, increases the spread of the disease [41].

In the other hand, better forecasts were obtained for shorter term forecasts indicating that there was some contribution of the inherent pattern in the historical mortality data that may be considered in multivariate models. The best fitted model performed quite well in hind casting for a single previously

removed one to two year or forecasting for one to two years in advance using INLA. However, as more years are removed, the space time model using INLA tends to predict closer and closer to the mean of the remaining data [33]. Forecasting five to ten years depends on the pattern of historical data. For example, when five or more years are removed in observed data for validation, the model only has 8 years for Rufiji and 6 years for Ifakara HDSS of data left to build from and the prediction became very close to the mean of the malaria mortality for remaining years than when you have long time period.

The Abuja Declaration noted the importance of accurate disease prediction for targeting and evaluating control measures [42]. For forecasting models to be useful for malaria control interventions and public health decision-making, models must produce accurate forecasts. The study examined various predictors and models across two different settings in rural Tanzania and consistently found that in both sites, models with spatial and temporal random term effect were necessary to achieve the highest possible forecasting accuracy. This is the first study, to the best of our knowledge, which incorporate mosquito nets predictor in combination with environmental predictors for forecasting malaria mortality in space and time.

Furthermore, the study focused on forecasting in space and over time. Fitting such complicated models could potentially be quite difficult and we used INLA, a relatively new method for Bayesian analysis. The INLA method is a new approach for Bayesian inference which was introduced by Rue and Martino [43, 44]. INLA provides a faster computation compared to Markov Chain Monte Carlo (MCMC) method. The method has been applied in different disease modelling and forecasting in which it performed quite well in forecasting for one to two years in advance [33, 45].

Strengths and limitations

This study used large datasets from Health and Demographic Surveillance System sites which are continuously registered vital demographic events in a geographical defined area. This has provided better estimates with high precision to allow generalization of the findings to a large population. The findings from Health and Demographic Surveillance Systems data provide information to policy makers and program managers which can be translated into policy and practice for planning.

This study used VA data collected in HDSS sites; Verbal autopsy has great potential for countries like Tanzania where more people die outside of health facility care where no records are available. Also, VA has been shown to provide the best results to obtain the specific causes of death in most of SSA [46] and widely used [12].

This study has some limitations that need to be considered in interpreting the findings. First, ownership of mosquito net was considered as a proxy for use of bed nets in the house, as information about use of bed nets in households and use were not collected during the study. Secondly, there was a possibility of misclassification bias

with regard to the causes of death when the sensitivity and specificity of the VA technique is relatively low for assessing cause of death. However, if this happened in this study it may lead to underestimate or overestimation of the reported malaria deaths. Hence, the direction of bias may be non-differential misclassification. Thirdly, there are other possible factors associated with malaria deaths that were not assessed in this study because are not captured in HDSS database, such as anti-malaria availability. The effect of these variables in mortality due to malaria estimates remains not quantified.

5. Conclusion

This study used Bayesian framework approach for developing space time model for forecasting malaria mortality. Although the model is reasonably reliable, especially with regard to the magnitude of forecasting one to five years, the model needs further evaluation to determine its accuracy. The INLA method performed quite well in hind casting for a single previously removed year or forecasting for one to two years in advance.

The model with spatial and temporal random effect terms predicted mortality due to malaria with combination of environmental factors and mosquito net which improved the forecasting accuracy. This model may be a useful tool for producing reasonably reliable forecasts of the malaria mortality for targeted intervention strategies. Moreover, including spatially and temporal varying random terms in the model may be necessary and good strategy for modelling malaria counts.

References

- [1] Winskill P, Rowland M, Mtove G, Malima RC, Kirby MJ. Malaria risk factors in north-east Tanzania. *Malar J. BioMed Central Ltd*; 2011;10: 98.
- [2] Dan ED, Jude O IO. Modelling and Forecasting Malaria Mortality Rate using SARIMA Models (A Case Study of Aboh Mbaize General Hospital, Imo State Nigeria). *Sci J Appl Math Stat*. 2014;2: 31.
- [3] Sriwattanapongse W, Prasitwattanaseree S. Malaria mortality modeling in Thailand. *Southeast-Asian J Sci*. 2013;2: 108–116.
- [4] Villalta D, Guenni L, Rubio-Palis Y ARR. Bayesian Space-Time modelling of malaria incidence in Sucre state-Venezuela. *AstA Adv Stat Anal*. 2013;97: 151–171.
- [5] Nkurunziza H, Gebhardt A, Pilz J. Forecasting Malaria Cases in Bujumbura. *World Acad Sci Eng Technol*. 2010; 253–258.
- [6] Gaudart J, Touré O, Dessay N, Dickoa L, Ranque S, Forest L, et al. Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali. *Malar J*. 2009;8: 61.
- [7] Zinszer K, Verma AD, Charland K, Brewer TF, Brownstein JS, Sun Z, et al. A scoping review of malaria forecasting : past work and future directions. *BMJ Open*. 2012; 1–12.
- [8] INDEPTH Network. INDEPTH Network:INDEPTH Standardized Verbal Autopsy Questionnaire [Internet].
- [9] Sibai AM, Fletcher A, Hills M, Campbell O. Non-communicable disease mortality rates using the verbal autopsy in a cohort of middle aged and older populations in Beirut during wartime, 1983 – 93. *J Epidemiol Community Heal*. 2001;55: 271–276.
- [10] Chandramohan D, Quigley M. Effect of misclassification of causes of death in verbal autopsy : can it be adjusted ? *Int J Epidemiol*. 2001;30: 509–514.
- [11] Yang G, Rao C, Ma J, Wang L, Wan X, Dubrovsky G, et al. Validation of verbal autopsy procedures for adult deaths in China. *Int J Epidemiol*. 2006;35: 741–748.
- [12] Mpimbaza A, Filler S, Katureebe A, Kinara SO, Nzabandora E, Quick L, et al. Validity of verbal autopsy procedures for determining malaria deaths in different epidemiological settings in uganda. *PLoS One*. 2011;6: e26892.
- [13] Setel PW, Whiting DR, Hemed Y, Chandramohan D, Wolfson LJ, Alberti KG LA. Validity of verbal autopsy procedures for determining cause of death in Tanzania. *Trop Med Int Heal*. 2006;11: 681–96.
- [14] Geubbels E, Amri S, Levira F, Schellenberg J, Masanja H, Nathan R. Health & Demographic Surveillance System Profile Health & Demographic Surveillance System Profile : The Ifakara Rural and Urban Health and Demographic Surveillance System (Ifakara HDSS). 2015;0: 1–14.
- [15] Mrema S, Kante AM, Levira F, Mono A, Irema K, Savigny D De. Health & Demographic Surveillance System Profile Health & Demographic Surveillance System Profile : The Rufiji Health and Demographic Surveillance System (Rufiji HDSS). *Int J Epidemiol*. 2015;0: 1–12.
- [16] Hanson K, Marchant T, Nathan R, Mponda H, Jones C, Bruce J, et al. Household ownership and use of insecticide treated nets among target groups after implementation of a national voucher programme in the United Republic of Tanzania: plausibility study using three annual cross sectional household surveys. *BMJ*. 2009;339.
- [17] Selemani M, Msengwa AS, Mrema S, Shamte A, Mahande MJ, Yeates K, et al. Assessing the effects of mosquito nets on malaria mortality using a space time model : a case study of Rufiji and Ifakara Health... Assessing the effects of mosquito nets on malaria mortality using a space time model : a case study of Rufiji and Ifakara. *Malar J. BioMed Central*; 2016.
- [18] WHO. International Statistical Classification of Diseases and Related Health Problems (ICD-10). Geneva, Switzerland; 1999.
- [19] Rowe AK, Steketee RW. Predictions of the impact of malaria control efforts on all-cause child mortality in sub-Saharan Africa. *Am J Trop Med Hyg*. 2007;77: 48–55. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18165474>
- [20] Lutambi AM, Alexander M, Charles J, Mahutanga C, Nathan R. Under-five mortality: spatial-temporal clusters in Ifakara HDSS in South-eastern Tanzania. *Glob Health Action*. 2010;3: 32–41. doi:10.3402/gha.v3i0.5254
- [21] Shabani J, Lutambi AM, Mwakalinga V, Masanja H. Rufiji Health and Demographic Surveillance System in rural Tanzania. *Glob Health Action*. 2010.

- [22] Waller LA, Carlin BP, Xia H GA. Hierarchical Spatiotemporal Rates, Mapping of Disease. *J Am Stat Assoc.* 1997;92: 607–617.
- [23] Kim H, Lim H. Comparison of Bayesian Spatio-Temporal Models. *J Data Sci.* 2010;8: 189–211.
- [24] Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Stat Med.* 2000;19: 2555–2567.
- [25] Arab A, Jackson MC, Kongoli C. Modelling the effects of weather and climate on malaria distributions in West Africa. *Malar J.* 2014;13: 126.
- [26] Berliner L. Hierarchical Bayesian time series models. In *Maximum entropy and Bayesian methods*. Netherlands: Springer; 1996.
- [27] Dimaggio C. Notes and Code for Small Area NYC Pedestrian Injury Spatiotemporal Analyses With INLA. 2014.
- [28] Knorr-Held L RG. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics.* 2000;56: 13–21.
- [29] Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol.* 2009;71: 319–392.
- [30] Zhu L CB. Comparing hierarchical models for spatio-temporally misaligned data using deviance information criterion. *Stat Med.* 2000;19: 2265–2278.
- [31] Branscum J, Perez M, Johnson WO, Thurmond MC. Bayesian spatiotemporal analysis of foot-and-mouth disease data from the Republic of Turkey. *Epidemiol Infect.* 2008;136: 833–842.
- [32] Abellan JJ, Richardson S, Best N. Use of space time models to investigate the stability of patterns of disease. *Environ Health Perspect.* 2008;116: 1111–1119.
- [33] Popoff E. An Approximate Spatio-Temporal Bayesian Model for Alberta Wheat Yield. The University of British Columbia. 2014.
- [34] Ensoy C. Forecasting and prediction models for demography and animal infectious diseases. Hasselt University, Belgium. 2014.
- [35] Bayesian Estimation and Prediction of Stochastic Volatility Models via INLA. 2012.
- [36] Wikle, Christopher K BI, Cressie N. Hierarchical Bayesian Space-Time Models. *Environ Ecol Stat.* 1998;5: 117–154.
- [37] Abeku TA, De Vlas SJ, Borsboom G, Teklehaimanot A, Kebede A, Olana D, et al. Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: A simple seasonal adjustment method performs best. *Trop Med Int Heal.* 2002;7: 851–857.
- [38] Makridakis S, Wheelwright SC HR. *Forecasting: Methods and Applications*. Edition 3rd, editor. New York.: John Wiley&Sons, Inc.; 1998.
- [39] Kilian AHD, Langi P, Talisuna A KG. Rainfall pattern, El Niño and malaria in Uganda. *Trans R Soc Trop Med Hyg.* 1999;93: 22–23.
- [40] Gomez-Elipse A, Otero A, van Herp M, Aguirre-Jaime A. Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997-2003. *Malar J.* 2007;6: 129.
- [41] Reiter P. Climate change and mosquito borne diseases. *Environ Health Perspect.* 2001;109: 223–234.
- [42] Secretariat. RBMP. The Abuja declaration and the plan of action. Roll Back Malaria. 2003.
- [43] Rue H MS. Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *J Stat Plan Inference.* 2006;137: 3177–3192.
- [44] Rue H, Martino S CN. Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations. *Statistics Report No. 1*. Trondheim, Norway; 2007.
- [45] Cosandey-Godin A, Krainski ET, Worm B, Flemming JM. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. *Can J Fish Aquat Sci.* 2014;72: 186–197.
- [46] Kaufman JS, Asuzu MC, Rotimi CN, Johnson OO, Owoaje EE, Cooper RS. The absence of adult mortality data for sub-Saharan Africa: A practical solution. *Bull World Health Organ.* 1997;75: 389–395.