# Detecting geographical variability in risk of malaria-attributable morbidity using spatial models

**Lawrence N Kazembe[1], Placid MG Mpeketula[2]**

[1] Mathematical Sciences Department, Chancellor College, University of Malawi, Zomba, Malawi
[2] Biology Department, Chancellor College, University of Malawi, Zomba, Malawi

*Corresponding Author:*
Lawrence Kazembe
Mathematical Sciences Department, University of Malawi, Zomba, Malawi. Tel: (+265) 8668 233 - Fax: (+265) 1524 046
E-mail: lkazembe@yahoo.com

**Summary**

***Background/Objectives.*** Considerable resources and effective tools are in place to control malaria in the sub-Saharan Africa region. The impact of interventions and how it may modify the burden of the disease is critical for national malaria control programmes. In the face of limited resources, quantifying the geographical patterns of the malaria risk is crucial for increased understanding the epidemiology of disease and for spatial targeting of interventions. This study developed disease maps, based on health facility incidence data and used spatial models, to describe geographical variation of malaria risk.

***Materials and Methods.*** We used malaria case data of children under the age of 5 years, collected at sub-district level in northern Malawi. We applied spatial techniques, accounting for heterogeneity and overdispersion in the data, to highlight areas of greatest need.

***Results/Conclusion.*** The disease burden depicted a west-east gradient, with highest risk found on the eastern side of the region. This pattern suggests that people living along the lakeshore region are at highest risk. This study constitutes an important first step for a detailed further study that would investigate risk factors related to the pattern observed in this study.

KEY WORDS: *malaria incidence, Bayesian spatial modelling, standardized incidence ratio, relative risk, under-five children, northern Malawi.*

## Introduction

Considerable resources and effective tools are in place to control malaria in the sub-Saharan Africa region. These tools include insecticide treated nets, indoor residual spraying, effective antimalarials with artemisia combination therapy. The impact of such interventions and how it may have modified the burden of the disease is crucial for national malaria control programmes (1, 2). Demand for data has increased for monitoring and evaluation, to characterize the epidemiological profile of the disease, thus guiding control efforts (3, 4). For those involved in policy development and implementation, knowledge of geographical distribution of the disease burden has become critical because progress can be monitored and the impact of control programmes evaluated through maps, and spatial inequalities in the disease burden highlighted for action and further analysis (4, 5, 6).

The study of the geographical distribution of malaria risk has generated significant interest in recent years (7, 8). Several reasons spurred this growth. First, malaria is a disease characterized by spatial clustering because of the nature of its transmission (8). In fact, geographical epidemiology of the disease has long been recognised (9). Second, advances in spatial methodology and developments in geographical information system have seen immediate applications in

malaria research. Third, there is now abundant source of spatially referenced malaria data amenable to spatial analysis. Recent leading examples in malaria disease mapping include the mapping malaria risk in Africa (MARA) project http://www.mara.org.za. and Malaria Atlas Project (MAP) http://www.map.ox.ac.uk..

This article developed spatial models to detect geographical variability in malaria risk in Malawi. Malaria in Malawi, as in much of the sub-Saharan Africa region, is a historical public health problem. About 3 million new cases of malaria occur every year in the country (10, 11). To assist towards meeting the RBM goals (12), and faced with limited resources, there is need to prioritise interventions to areas of highest need. Very little, however, is known about the disease risk distribution in Malawi. Studies that have attempted to quantify geographical variability of malaria risk include Craig *et al.* (7) and Kazembe *et al.* (13, 14). Craig *et al.* (7) developed a theoretical climatic suitability model, but this has important limitations as it fails to provide insight into the transmission of malaria in Malawi. There is need to define malaria risk based on malaria-specific indicator. Kazembe *et al.* (13), using point-referenced prevalence of infection data, a malaria-specific indicator, predicted and mapped malaria risk in the country. In another study, using hospital register data, Kazembe *et al.* (14) studied spatial patterns of malaria-related hospital admissions and mortality in a single district in southern Malawi.

Here we extend similar models suitable for area-referenced routine malaria incidence data. Patient data sourced from routine health management information system (HMIS), when reasonably consistent and complete in reporting might offer some indication of the local trend in the malaria burden, and could provide supporting evidence of the impact of the interventions (4). This study, as in Kazembe *et al.* (14), used incidence data of children under the age of 5 years, aggregated at sub-district level in northern Malawi. Our objective was to describe the geographical variation in risk based on these data to guide malaria control efforts, or documenting baseline risk level against which future interventions can be assessed. A small area level was chosen because it is more appropriate for local health planning, monitoring and evaluation. A number of spatial techniques were used and compared to detect the geographical variability in malaria morbidity.

## Materials and Methods

### Study area

The northern region of Malawi is one of the 3 regions of the country. It comprises of 5 districts, which is further divided into 73 sub-districts or wards. Population projections for 2006, based on the 1998 census, was approximately 1.3 million with under-five children constituting 17% of the total population.

### Morbidity data

Cases of malaria were sourced from health facility registers, between January 2004 to December 2006, arising from children under 5 years of age. These cases included both clinically diagnosed and microscopically confirmed cases, and were allocated to 73 sub-districts for spatial analysis. Because of lack acquired immunity in children of under 5 years of age, the incidence of malaria from this group is a better indicator of malaria risk for an endemic area. In addition, the highest burden of malaria, in such endemic areas, is inflicted on children and studying geographical variability is warranted for improved understanding of the disease burden.

The corresponding population at risk estimates for each sub-district were linearly interpolated from the 1998 census data available at the Malawi national statistics office website http://nso.malawi.net.. These were used to calculate crude and standardized incidence ratio (SIR) for each area.

### Statistical methods

#### *Standardized incidence ratio*

The SIR is a common approach to estimate the relative risk (RR) of the disease in an area. This is defined as, $\frac{O_i}{E_i}$, which is an improvement of the crude ratio, $\frac{O_i}{N_i}$, where $O_i$ is the observed number of cases, $E_i$ is expected number of cases and $Ni$ is the population at risk in each area $i, i = 1,\ldots,73$. The expected number was calculated as $E_i = N_i \left(\frac{\sum O_i}{\sum N_i}\right)$ (15). The standard error of SIR is $\sqrt{N_i} / E_i$. The working assumption with SIR is that the observed counts are assumed independent and drawn from a Poisson distribution, $O_i \sim Po(\theta_i E_i)$, with mean $\theta_i E_i$ where $\theta_i$ is the true unknown RR, estimated by SIR.

***Heterogeneity and over-dispersion tests***

A Potthoff-Whittinghill (PW) test (16), was used to test for presence of heterogeneity in the disease rates. This statistic, under the null hypothesis, assumes constant (or homogeneous) risk across the study region, against the alternative that the relative risks are drawn from the Gamma distribution, and is given by,

$$PW = E_{+}\sum_{i}\frac{O_i(O_i-1)}{E_i}$$

where $E_{+}$ is the sum of the expected number of cases for the region.

Further analysis used Dean's test (17) to assess overdispersion in the disease rates. Overdispersion occurs when the variance is significantly higher than their mean, under the assumption that the count data are drawn from the Poisson (Po) model. In the case of overdispersion, a better model must be proposed. A good choice is a Negative Binomial (NB) distribution. Tests for heterogeneity and overdispersion were implemented using DCluster package (18) in R statistical software system (19).

**Smoothing incidence relative risk**

Relatively small population totals per area result in large random variability, and the standardized incidence ratios are unstable. Bayesian methods for disease mapping (15, 20, 21), were carried out to smooth the relative risk estimates in each sub-district. Three approaches were proposed. The first approach assumed spatially unstructured variability to produce globally smoothed estimates for all sub-districts. The second approach assumed spatially structured variability, leading to locally smoothed risk estimates, which is achieved by pooling information from neighbouring areas. The conditional autoregressive (CAR) model was used to smooth the data (20). Two areas were assumed neighbours if they have a common boundary. The last approach combined the spatially structured variation and unstructured heterogeneity (20). The fitted three models are given as:

$$\log(O_i) = \log(E_i) + U_i, \qquad (1)$$

$$\log(O_i) = \log(E_i) + S_i \qquad (2)$$

$$\log(O_i) = \log(E_i) + U_i + S_i \qquad (3)$$

where $\log(E_i)$ is the offset. The unstructured spatial effects are captured through $(U_i)$, while the spatial structured are modelled by $S_i$. Smoothing was carried out in BayesX 1.14 (22), using Markov Chain Monte Carlo simulation techniques. Detailed methodological aspects of the Bayesian analysis that apply in spatial mapping are given
elsewhere (15, 20, 21). For each approach, both the Poisson and Negative Binomial models were fitted, giving six models. Model comparison was based on the deviance information criteria (DIC). Models with a smaller DIC were preferred as best fitting (23).

For the best model, maps were plotted accompanied by measures of uncertainty. For the SIR map we also plotted the standard error map. Bayesian maps were accompanied by maps of posterior probabilities. These maps assessed areas of significantly lower or greater risk compared to the overall mean of the whole study area. For the Bayesian maps, we subdivided the posterior probabilities into <20%, 20-80% and >80%. Values of greater than 80% strongly indicate that the risk was higher than the mean of the whole area, and values of less than 20% indicate that the area had risk lower than the mean risk of entire area. Intermediate values (20-80%) suggest that there was not enough evidence to differentiate from the overall risk (5).

## Results

A total of 24 022 malaria cases and 31 325 population at risk among under 5 years old children were recorded in the northern region of Malawi between 2004 and 2006. The overall raw incidence rate (crude ratio) was 1.44 (range: 0.008, 6.33) episodes per person per year. The expected number of cases ranged from 10.33 to 32 346.25. Summaries are given in Table 1.

Figure 1 shows the maps of unsmoothed SIR with the corresponding standard errors. The maps have to be interpreted by considering that different shades of grey were proportional to risk values. The darker the area, the higher the risk. The mean unsmoothed risk estimate was 1.15 (95% confidence interval (CI): 0.14, 1.42). Figure 1a portrays an inhomogeneous map, indicating that the pattern was strongly influenced by random variability in the distribution of malaria cases. A slight shrinkage in rate values can be seen compared to the crude ratio. The standard error of SIRs were high for isolated areas reflecting extreme values on which these estimates were based (Figure 1b).

Table 1. Observed, expected cases, raw rates and unsmoothed standardised incidence ratio (SIR) at sub-district level in northern Malawi.

| Statistics | Observed Cases | Expected Cases | Raw rate | Unsmoothed SIR* |
|---|---|---|---|---|
| Mean | 6 135 | 6 135.06 | 144.29 | 110.77 |
| Median | 3 747 | 4 526.93 | 107.40 | 59.96 |
| 95%CI | 1 000- 11 135 | 2 420.42-8 183.37 | 0.08-628.39 | 13.89-122.32 |

**RR is multiplied by 100*

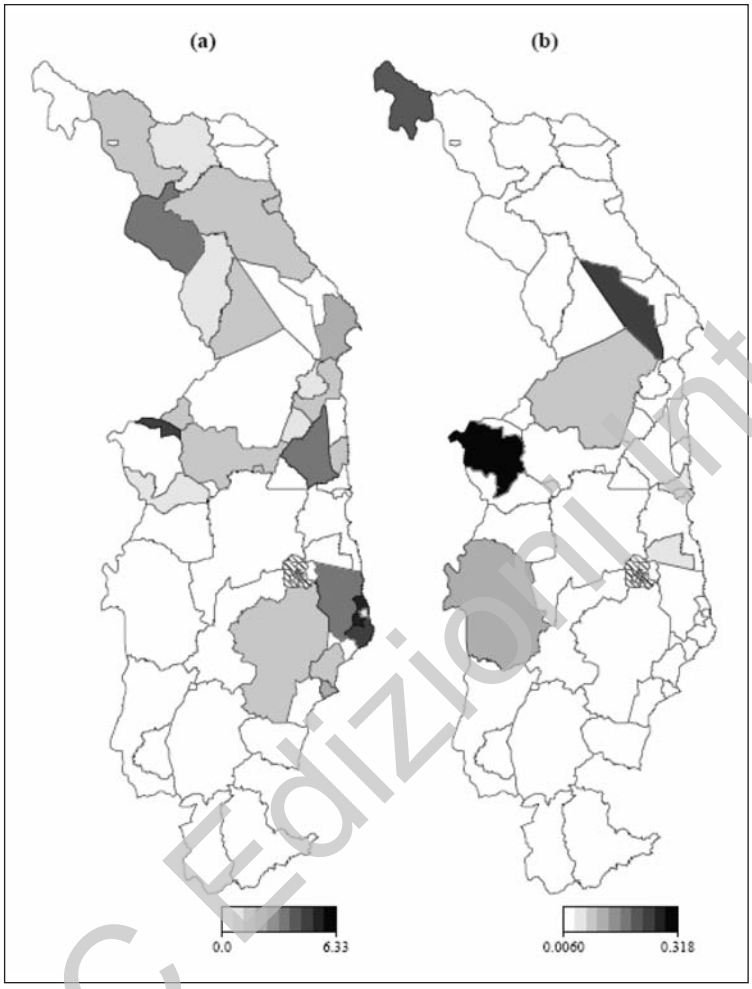


Figure 1. Choropleth maps of sub-district level standardized incidence ratios for malaria for under-five years old children in four districts in northern region of Malawi from January 2004 to December 2006: (a) the SIR and (b) the corresponding standard error. Areas with no or insufficient data are marked with diagonal solid lines.

There was significant heterogeneity (PW test, $p < 0.001$) in risk across sub-districts in northern Malawi. In other words, the variation in RR was more than by chance. This agreed with the unsmoothed SIR (Figure 1). Part of this heterogeneity was due to overdispersion in the data (Dean test, $p < 0.001$). This means that the assumption of data drawn from the Poisson model might be wrong. As an alternative the NB model was fitted. Table 2 gives the DIC for the all the six models fitted. Clearly the NB models provided a better fit than Poisson model, and the model with both spatially structured variation and unstructured heterogeneity terms fitted the data adequately (Model 5 in Table 2). Therefore the SIR were adequately smoothed using a NB model that allowed for both spatially structured variation (local smoothing) and unstructured random effects (global smoothing).

Table 3 summarises the smoothed relative risks for the northern region of Malawi, using model 5, but also compared with the other NB model that allowed for only spatially structured variation or unstructured variation. Figure 2 gives a plot of the total spatial effects based

Table 2. Model comparison measures for various spatial models fitted in the study. See text for explanation.

| Model | Type* | Spatial effect | DIC |
|---|---|---|---|
| 1 | NB | Structured | 63.30 |
| 2 | Po | Structured | 102.31 |
| 3 | NB | Random | 56.81 |
| 4 | Po | Random | 95.98 |
| 5 | NB | Structured+ Random | 51.44 |
| 6 | Po | Structured+ Random | 95.95 |

**NB-Negative Binomial model, Po-Poisson model*

on the best model. The relative risk ranged from 0.01 to 1.25, with mean RR=1.10 (plot a), with relatively smaller RR for most areas. The corresponding map gives posterior probabilities (plot b). These were subdivided into three intervals: <20%, 20-80% and >80%. Areas with probability values of equal or higher that 80% indicates that RR was significantly higher than the overall RR, and those with probabilities <20% suggest that RR in the area was significantly lower than the overall RR. In between values implied that RR were not significant. From the map we have areas of both positive excess risk (black areas), and significantly lower risk (white areas).

## Discussion

Unlike our previous estimates based on historical prevalence data (13), this study relied on routine incidence data (3), and therefore provided more contemporary epidemiological characterisation of malaria risk in the region. This paper also highlights the importance of mapping the extent of malaria burden at a smaller level, i.e., sub-national scale. Because malaria transmission can be being influenced by geographical environmental factors, estimates at such level are more robust and re-

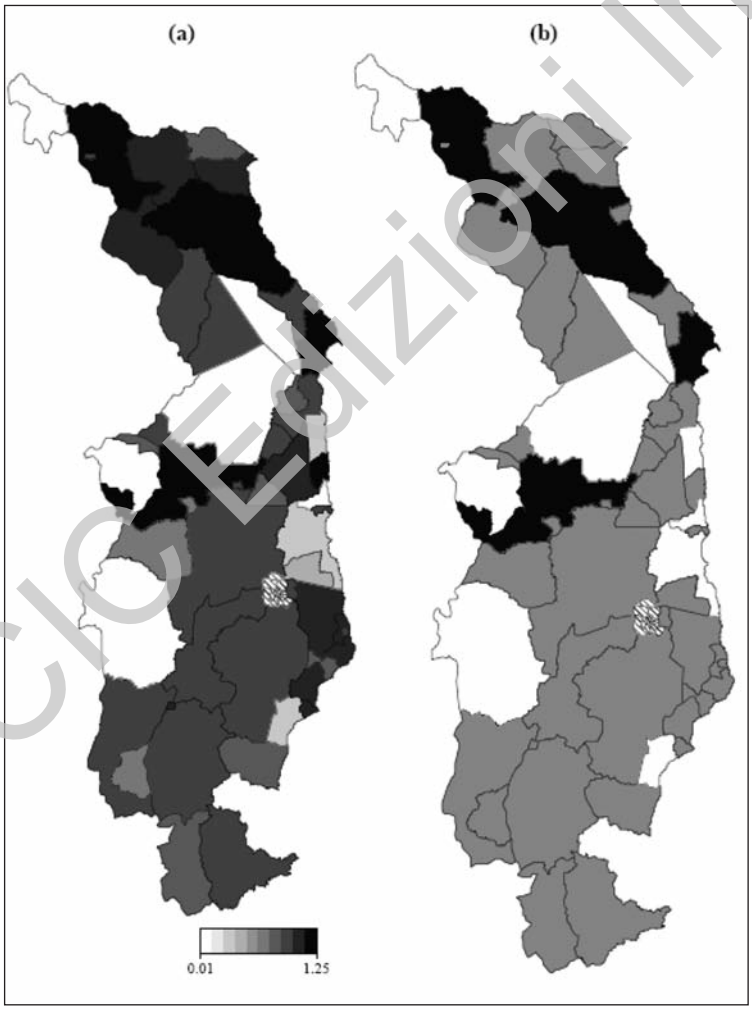


Figure 2. (a) Smoothed estimates (relative risk: RR) of malaria and (b) the corresponding significance map of RR: negative significance (white areas), not significant (grey areas), and positive significance (black areas).

Table 3. Smoothed estimates of relative risk of malaria incidence of under-five years children in northern Malawi, using the negative binomial model.

| Statistics | Model 1 | Model 3 | Model 5 |
|---|---|---|---|
| Mean | 130.43 | 115.99 | 113.35 |
| Median | 79.34 | 81.33 | 82.78 |
| 95%CI | 38.72-146.58 | 42.55-136.92 | 45.02-134.51 |

*RR is multiplied by 100*

flective of underlying drivers of disease risk (24, 25). Such disease burden estimates are crucial for informing malaria control programmes. These are important for planning, monitoring and advocacy (2, 4).

We applied Bayesian hierarchical models to describe the distribution of risk of malaria in children under the age of five. This case study is but an example of many conceivable applications of such models in public health research. It is evident that the risk estimates based on the raw or standardized mortality ratio estimators can be misleading (Figure 1a). This seems to suggest that spatial correlation was strong in the data, and failure to account for these may lead to erroneous conclusions. The smoothed estimates, through the Bayesian hierarchical model, are easy to interpret, despite the fact that the posterior estimates are conservative, with high specificity and low sensitivity (5). This has its own advantage as it avoids false positives thereby producing true clusters in the maps. Put differently, the smoothed malaria risk estimates give a more stable pattern of the underlying risk of disease than that provided by the raw estimates (5, 21).

In this analysis, we saw that the risk of malaria is varied between sub-districts, and even among areas that neighbour each other. Our results is a critical first step, which as in any disease mapping, should generate relevant aetiological hypotheses with regards possible risk factors and covariates of the disease patterns. Factors contributing to this pattern are a matter of conjecture. It is likely that environmental factors including altitude, nearness to water bodies, climatic factors, soil type may influence malaria transmission (26, 27, 28). Unmeasured socio-economic differences may also contribute to this pattern. Rural masses are at increased risk of malaria infection and death because they are not able to pay for effective malaria drugs nor afford transport to a health facility for prompt effective treatment. Rurality is, therefore, one of the factors worth considering in future research (29). The geographical pattern can also be explained by variability in health seeking behaviour. Health seeking behaviour plays a critical role in accessing prompt and effective care. Home based care or traditional medicines are often the first sources of care in most African communities because of traditional beliefs, difficulties in accessing and unavailability of formal health services. Only when the disease is perceived to be severe or near fatal, do people seek modern biomedical care at health facilities (30).

This analysis depended on the HMIS data and, as with all routinely collected data, there are known limitations on data quality in terms of completeness, correctness and consistency. Health facility data under-report malaria cases that occur in the community because most people resort to home or community-based care (3, 31). It is therefore reasonable to interpret the risk pattern realized in this study as representing the risk of dying from severe malaria (1, 4). When considered necessary, under-reporting can be adjusted for in the analysis (3).

In conclusion, this study has filled a significant gap in the knowledge of geographical distribution of morbidity attributed to malaria in Malawi. The maps identified sub-districts that were of high risk of malaria morbidity, and demonstrates the significance of using these disease mapping techniques in a surveillance system. This has important potential implications for research and health policy planning purposes. First, the maps can generate leads for in-depth epidemiologic or geographic studies that may shed light on factors contributing to malaria risk. Secondly, findings may help policy decision makers to pinpoint high-risk areas with specific health problems. Thirdly, the maps may contribute to developing and prioritizing health targets at the district/ area level. Fourthly, results of this study could form basis for distributing and targeting interventions across geographical zones (6). For instance, the national malaria control programme may want to explore why the risk of malaria is higher in some areas, and is significantly lower than other sub-districts.

**Acknowledgment**

This research was supported by a training grant received from UNDP/World Bank/WHO
Special programme for Research and Training in Tropical Diseases (TDR).

## References

1. Roll Back Malaria Partnership. Monitoring and Evaluation Reference Group (MERG), Morbidity Task For-

ce Meeting on Proposed method of estimating malaria incidence at country level (October 2004) minutes available at http:// www.rbm.who.int/ partnership/ wg/ wg-monitoring/, 2004.
2. United Nations Millennium Project. Coming to grips with malaria in the new millennium. Task force on HIV/AIDS, Malaria, TB, and Access to Essentail Medicines, Working Group on Malaria; 2005.
3. Chaulagai CN, Moyo CM, Koot J, Moyo HB, Sambakunsi TC, Khunga FM, Naphini PD. Design and implementation of a health management information system in Malawi: issues, innovations and results. Health Policy Plan 2005; 20:375-84.
4. WHO/Roll Back Malaria. Framework for monitoring progress and evaluating outcomes and impact. Roll Back Malaria, World Health Organisation: WHO/CDS/RBM/ 2000.25; 2000.
5. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. Environ Health Perspect 2004; 112:1016-25.
6. Carter R, Mendis KN, Roberts D. Spatial targeting of interventions against malaria. Bull World Health Organ 2000; 78:1401-1411.
7. Craig MH, Snow RW, le Sueur D. A climate-based distribution model of malaria transmission in sub-Saharan Africa. Parasitol Today 1999; 15:105-11.
8. Kleinschmidt I, Sharp BL, Clarke GP, Curtis B, Fraser C. Use of generalized linear mixed models in the spatial analysis of small area malaria incidence rates in KwaZulu Natal, South Africa. Am J Epidemiol 2001; 153:1213-21.
9. Le Sueur D, Binka F, Lengeler C, De Savigny D, Snow B, Teuscher T, Toure Y. An atlas of malaria in Africa. Africa Health 1997; 19:23-24.
10. Kabuluzi S, Ali D, Chiphwanya J, Nsona H, Zoya J, Bizuneh K. Malawi Roll Back Malaria Consultative Mission: Essential Actions to Support the Attainment of the Abuja Targerts, 19-23 April 2004. Malawi RBM Country Report, 2004.
11. Government of Malawi. Health Information Annual Bulletin 2002-2003. Ministry of Health and Population. Lilongwe, Malawi, 2003.
12. Government of Malawi. Malaria Policy. National Malaria Control Programme. Community Health Sciences Unit. Government of Malawi, Lilongwe, 2002.
13. Kazembe LN, Kleinschmidt I, Holtz TH, Sharp BL. Spatial analysis and mapping of malaria risk in Malawi using point referenced prevalence of infection data. Int J Health Geogr 2006; 5:41.
14. Kazembe LN, Kleinschmidt I, Sharp BL. Patterns of malaria-related hospital admissions and mortality among Malawian children: an example of spatial modeling of hospital data. Malaria J 2006; 5:93.
15. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987; 43:671-681.
16. Potthoff RF, Whittinghill M. Testing for homogeneity: the Poisson distribution. Biometrika 1996; 53:183-190.
17. Dean CB. Testing for overdispersion in Poisson and binomial regression models. J Am Statist Assoc 1992; 87:451-457.
18. Gmez-Rubio V, Ferrndiz-Ferragud J, Lpez-Qulez A. Detecting clusters of disease with R. J Geographical Systems 2005; 7:189-206.
19. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available at http://www.r-project.org, 2005.
20. Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics (with discussion). Ann Instit Stat Math 1991; 43:1-59.
21. Ugarte MD, Ibanez B, Militino AF. Modelling risks in disease mapping. Stat Meth Med Res 2006; 15:21-35.
22. Brezger A, Kneib T, Lang S. BayesX: Analyzing Bayesian structured additive regression models. J Statist Soft 2005; 14:11.
23. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and ¯t (with discussion). J R Statist Soc B 2002; 64:583-639.
24. Bell DR, Jorgensen P, Christophel EM, Palmer KL. Estimation of the malaria burden. Nature 2005; 437:E3-E4.
25. Giles HM. Epidemiology of Malaria. *In* Giles HM, Warrell DA (Eds). *Bruce-Chwatt's Essential Malariology*. 3rd Edn, pp.124-162. Boston: Little and Brown: Boston, 1993.
26. Hay SI, Noor AM, Simba M. Clinical epidemiology of malaria in the highlands of western Kenya. Emerg Infect Dis 2002; 8:543-548.
27. Shaman J, Day JF. Achieving operational hydrologic monitoring of mosquito borne disease. Emerg Infect Dis 2005; 11:1343-50.
28. Srivastava A, Nagpal BN, Saxena R, Sharma VP. Geographic information system as a tool to study malaria receptivity in Nadiad Taluka, Kheda district, Gujarat, India. Southeast Asian J Trop Med Public Health 1999; 30:650-6.
29. Kazembe LN, Appleton CC, Kleinschmidt I. Choices of treatment for malarial fever: examining spatial patterns. Malaria J 2007; 6:40.
30. De Savigny D, Mayombana C, Mwageni E, Masanja H. Care-seeking patterns for fatal malaria in Tanzania. Malar J 2004; 3:27.
31. De Savigny D, Binka F. Monitoring future impact of malaria burden in sub-Saharan Africa. Am J Trop Med Hyg 2004; 71S:224-231.