Modeling Malaria Incidence in North-Western Thailand

Wattanavadee Sriwattanapongse [a] and Metta Kuning* [b]

[a] Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand.
[b] Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani 94000, Thailand.

*Author for correspondence; e-mail: kmetta@bunga.pn.psu.ac.th

ABSTRACT

The transmission of malaria in Thailand is common, particularly in the North-western region of the country. The objective of this study is to identify the patterns of hospital-diagnosed malaria incidences in districts and quarterly periods in the North-western region of Thailand in 1999-2004. Regression models based on principal components describe these patterns. The models show trends and spatial variations in disease incidence. Graphical displays showing both regional and period effects are presented. The results of this study show that malaria incidence rates decreased substantially in most districts during the study period, but remained very high in border districts with Myanmar.

Keywords: malaria incidence, additive model.

1. INTRODUCTION

Malaria in Thailand is endemic in forest regions and many cases occur along the national borders, particularly on the border with Myanmar to the east. Although malaria cases and deaths have fallen substantially since 1999, the disease remains a considerable public health problem.

Gomez-Elipe et al. [1] developed a model to predict malaria incidence in an area of unstable transmission in Burundi by studying the association between environmental variables and disease dynamics. The model used time series of quarterly notifications of malaria cases from local health facilities, rain and temperature records, and the normalized difference vegetation index. An autoregressive integrated moving average methodology was employed to obtain a model showing the relation between quarterly notifications of malaria cases and the environmental variables. Briet et al. [2] showed that the addition of rainfall as a covariate improved prediction of selected (seasonal) ARIMA models moderately in some districts of Sri Lanka but worsened prediction in others.

Devi and Jauhari [3] investigated the relationship between climate variables and malaria transmission in India. Earlier, Bi et al. [4] also explored the impact of climate on the transmission of malaria in China and suggested that climatic variables should be considered as possible predictors for regions with similar geographic and socio-economic conditions. Hoshen and Morse [5] described a mathematical-biological model of the parasite dynamics in Africa, comprising the...
weather-dependent stages, both within vectors and within hosts. Gagnon et al. [6] found a statistically significant relationship between El Niño and malaria epidemics in South America and thus postulated that global warming will be an important factor in the spatial distribution of infectious disease. Kleinschmidt et al. [7] investigated malaria incidence in children under 10 in South Africa by using logistic regression modeling. The model used climatic, population and topographic variables as potential predictors and described a simple two-stage procedure for producing maps of predicted risk, including environmental factors such as land use. Built-up areas were found to have the highest incidence rates.

Studies like these aim to identify risk factors for the disease, which could provide a basis for health organizations in countries affected by malaria to establish effective prevention programs. However, when resources are limited it is also important to know in which area prevention should be targeted for treatment and control patterns and trends, and this is the focus of the present study.

The objective of our study is thus to identify the spatial patterns and trends of hospital-diagnosed malaria incidences based on case data aggregated by quarterly periods in 65 districts of the North-western region of Thailand. The provinces in our study comprise Lamphun, Phrae, Nan, Chiang Rai, Mae Hong Son and Tak. These provinces were selected for study because complete data were recorded for them, but not for other provinces in the region.

2. METHODS

2.1 Data Management

Data used in the current study were taken from registries of hospital-diagnosed infectious disease cases which were collected routinely in each of Thailand’s 76 provinces by the Ministry of Public Health. For each year after 1998, these data are available in computer files with a record for each case and fields comprising characteristics of the subject and the disease, including dates of sickness and disease diagnosis, the subject’s age, gender, address, severity of the illness, and date of death for mortality cases. Counts for malaria incidence were created for each combination of quarter (24 periods from January, March 1999 to October-December 2004), age group (0-4, 5-14, 15-39 and 40+ years), and district. Incidence rates were computed as the number of cases per 1,000 residents in the district according to the 2,000 Population and Housing Census of Thailand. Since there was little evidence of a gender effect, the data for the two sexes were combined.

2.2 Statistical Analysis

We consider two alternative statistical models for describing the relation between malaria incidence and the age group, district and period factors, namely a negative binomial and log-normal distribution, respectively. In each case the mean function for the selected distribution is a specified combination of demographic factors. In the simplest case this combination is additive, taking the form \( \alpha_i + \beta_j + \gamma_t \) where the indexes \( i, j \) and \( t \) denote age-group, district, and period, respectively.

The negative binomial model is an extension of the Poisson model for incidence rates that allows for the overdispersion that commonly occurs for disease counts in regions. It takes the form (see, for example, Venables and Ripley [8]),

\[
E[n_{ijt}] = \mu_{ijt} = \mu_{ij} \exp(\alpha_i + \beta_j + \gamma_t),
\]

\[
\text{var}[n_{ijt}] = E[n_{ijt}](1 + E[n_{ijt}]/\theta).
\]
where $y_{ijt}$ is the natural logarithm of the incidence rate, defined as $n_{ijt}/p_{ijt}$, the number of cases per 1,000 population in the age-group and district.

We also consider an alternative model with normally distributed errors for the log-transformed incidence rate, taking the form

$$E \left[ \log(y_{ijt}) \right] = \alpha_i + \beta_j + \gamma_t. \quad (3)$$

To allow for zero counts in this model, we replaced them by a specified constant between 0 and 1 before log-transforming.

Since there is evidence of very much higher malaria incidence rates in districts bordering Myanmar [9], and preliminary analysis of our data indicates that the age distribution of incidence rates for border and non-border districts differs substantially, we extend the model (3) to

$$E \left[ \log(y_{ijt}) \right] = \alpha_{ik} + \beta_j + \gamma_t. \quad (4)$$

where $k = 1$ for border districts, 2 otherwise.

Since districts are nested within border location, one of the $\alpha_{ik}$ parameters in this model is redundant, requiring a constant to be specified. For definiteness we assume $\alpha_{i1} = \alpha_{i2}$. Since we also wish to investigate the possibility that the trend $\gamma_t$ varies from district to district, we consider a further extension of model (4) to include such a multiplicative interaction by replacing $\gamma_t$ by $\delta \gamma_t$, that is

$$E \left[ \log(y_{ijt}) \right] = \alpha_{ik} + \beta_j + \delta \gamma_t. \quad (5)$$

This model, though non-linear, can be fitted by regression analysis after first computing by scaling the first principal component of the covariance matrix of residuals after fitting the model $\alpha_{ik} + \beta_j$, giving a further district-specific predictor factor $\delta$. This technique appears to have been first used by Fisher and Mackenzie [10]. It was later used by McNeil and Tukey [11] to model socio-demographic data in regions and has been applied to spectroscopy, chromatography and other fields to model additive mixtures of overlapping curves [11-14]. The method has also been used more recently for modeling age-specific mortality rates [15-17]. Graphical plots against normal quantiles of deviance residuals (for the negative binomial model) and studentized residuals (for the log-normal model) are used to assess model distribution assumptions. The 95% confidence intervals for incidence rates associated with each factor in the model after adjusting for other factors are computed using sum contrasts.

R software [18] was used for graphical displays and statistical analysis.

3. RESULTS

Distributions of Incidence Rates

Between January 1999 and December 2004, number of 67,347 malaria hospital cases were reported in the 6 provinces (65 districts) considered in the study. The number of cases in any quarter period for a particular age group and district varied from 0 to 1,023 and the mean annual malaria incidence rate was 3.5 cases per 1,000.

Additive models of the form (1) and (3) were fitted to the data with residuals plots shown in the top panel of Figure 1. We need to transform the data (observation: $y$) by using log-transformation for all observations. Not only replace zero counts by 0.25 but also add 0.25 for non-zero counts. In other word we use $\log(y+0.25)$ to replace $y$. 

$R$ software was used for graphical displays and statistical analysis.
Figure 1. Diagnostic residual plots for negative binomial (top left) and log-linear (top right) models, and plots of counts and incidence rates for the log-normal model (lower panels) for malaria incidence rates in North-western Thai districts.

The residuals plot for the negative binomial distribution clearly indicates that this model does not fit the data, so in further analysis we used models (4) and (5).

Figure 2 shows confidence intervals for adjusted incidence rates by age group, trend and district after fitting model (4). The dotted line denotes the overall annual mean malaria incidence.
Figure 3 shows malaria incidence rates for each district based on model (5) after adjusting for age group and trend. The district specific incidence rates (denoted by the parameter $\beta$ in the model) are plotted against the parameter $\delta$, which indicate the extent to which each district follows the prevailing trend and is thus labeled the “district trend amplitude”. This model gave an r-squared of 0.817.

Figure 3. District components of malaria incidence in North-western Thailand 1999-2004: additive model.
Districts with a confidence interval lower bound higher than the mean in Figure 3 were categorized as having a higher than average incidence, while districts with a confidence interval upper bound less than the mean were categorized as having a lower than average incidence. Figure 4 shows a thematic map based on this classification, with the district specific trend effects also shown on the map. Note that no districts have confidence intervals crossing the mean, so the districts are classified into just two groups: (1) higher than average and (2) lower than average.

Figure 4 shows a substantial difference in annual malaria incidence between districts bordering Myanmar and other districts. While most districts show a downward trend in malaria incidence over the 6 year period, and for two districts this trend is very high (Thung Hua Chang in Lamphun and Wang Chao in Tak), two districts (Lamphun Mueang and Ban Hong in Lamphun) show no trend, and one district (Li in Lamphun) shows an increasing trend.

**Figure 4.** Malaria incidence in North-western border provinces of Thailand: 1999-2004.
4. DISCUSSION

The additive plus multiplicative linear model provides an appropriate fit to the malaria incidence rates in the study region classified by districts, age-group and period. The probability plots of residuals against normal quantiles indicate that the negative binomial model does not provide a satisfactory fit but a linear model for log-transformed incidence rates does, provided the zero counts are replaced by 0.25. However, it should be noted that for the models given by equations (3), (4) and (5) the estimated standard errors for the parameters within each factor are the same and thus do not depend on the number of cases or the population at risk within each cell. This is because the data to which the models are fitted only comprise incidence rates within each cell, in contrast to the alternative negative binomial model, which takes into account both the number of cases and the population at risk in each cell. It is possible to extend the log-normal model to take these factors into account by using a weighted linear model [19], but this is beyond the scope of the present study.

According to this study, the incidence of malaria was highest among persons aged between 5 and 39 years in border districts, but in contrast, children aged 0-4 had highest incidence rates among those in other districts. However, these differences are relatively small compared to the differences between incidence rates in border and non-border districts.

From this study, malaria trends for most districts in the six provinces showed a consistent decreasing trend, with a seasonal pattern peaking in the April-June quarter (Figure 2). As remarked by Khasnis and Nettleman [20], the ability of mankind to adapt is dependent upon the magnitude and speed of change, and also depends on recognizing epidemics early, containing them effectively, providing appropriate treatment, and committing resources to prevention and research.

The maps of malaria prediction risk use regression modeling to determine approximate risk on a larger scale and employ geo-statistical approaches to improve prediction at a local level. Malaria incidence was predicted by means of a model containing district, age group and period variables as factors. The results are illustrated by a thematic map showing both the districts with high incidence rates and the trend in each district. Such maps can be used by public health authorities for applying preventive measures to control malaria outbreaks by focusing preventive measures according to priority in high, average and low zones.

ACKNOWLEDGEMENTS

We would like express our gratitude to Don McNeil, Phattrawan Tongkumchum, Chhamnein Choonpradub and Supreeya Wongtra-ngan for their invaluable assistance, encouragement and helpful guidance.

REFERENCES


[4] Bi P., Tong S., Donald K., Parton K. and Ni J., Climatic variables and transmission...


