Modeling and Forecasting Malaria and Dengue Hemorrhagic Fever Incidence and Prevalence in Northern Thailand

Wattanavadee Sriwattanapongse and Surin Khanabsakdi
Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

Received: May 5, 2011 / Accepted: May 14, 2011 / Published: June 25, 2011.

Abstract: Malaria and dengue hemorrhagic fever (DHF) are infectious diseases prevalent in many tropical countries, including Thailand. Thailand is located geographically in a tropical zone and the transmission of malaria and DHF is common, particularly in the upper Northern region of the country. The objective of this study is to identify the patterns of hospital-diagnosed Malaria and DHF incidences by using the previous monthly or quarterly periods of incidences occurring in the upper Northern region of Thailand. The authors use additive plus multiplicative regression models to describe these patterns. The models can be used to forecast malaria and DHF incidences, thus predicting where epidemics are likely to occur. This information can be used to prevent disease outbreaks occurring. Graphical displays showing district and period effects are presented. The results of this study show that historical malaria and DHF incidence rates can be used to provide a useful model for forecasting future epidemics. The graphical display shows the improvement of risk prediction brought about by model. The model, even if based purely on statistical data analysis, can provide a useful basis for allocation of resources for disease prevention.

Key words: Additive model, dengue hemorrhagic fever incidence, malaria incidence.

1. Introduction

Malaria is a biological phenomenon where all the three elements of the infection system, namely man, mosquito and parasite are influenced by various environmental variables. Approximately 40% of population of the world, mostly those living in the world’s poorest countries, is at risk of malaria. Every year, more than 500 million people become severely ill with malaria [1].

Malaria in Thailand is endemic in forest regions and most prevalent 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.

Dengue hemorrhagic fever (DHF) is a severe disease, particularly among children age under 15 years [2]. It is estimated that there are over 2.5 billion people, or 40% of the world population, living in dengue prevalent areas of the world [3]. According to current estimates, at least 100 countries are endemic for DHF, with over 50 million infections and about 400,000 cases of DHF being reported annually. It is a leading cause of childhood mortality in several Asian countries. A recent study of Poovorawan et al. [4] found DHF to be a major cause of acute hepatic failure in Thai children.

In Thailand, DHF has been and remains a major public health concern. In 1998, it was a leading cause of hospitalization and death, with 29,954 cases diagnosed nationally [5]. This outbreak was the second largest DHF epidemic in Thailand after a similar one that occurred in 1987.

The objective of this study is to identify the patterns
of hospital-diagnosed malaria and DHF incidences by using previous monthly or quarterly data of incidences occurring in the Northern region of Thailand. The provinces in this region consist of Lumphun, Phrae, Nan, Chiang Rai, Mae Hong Son and Tak. The number of districts within these six provinces totals 65.

2. Experimental Section

2.1 Data Management

Individual hospital case records routinely reported in each province from 1999-2004, and linear regression models of log-transformed incidence rates were used to assess the effects of age, period and district. Given that the monthly disease counts in individual cells defined by age group, district and period were often low, with many zero, we omitted districts with cases of malaria numbering less than 60 (22 districts) and cases of DHF numbering less than 50 (29 districts). Both the additive and additive plus multiplicative statistical models were used to identify trends and high disease occurrences.

Data used in the current study are taken from a registry of hospital-diagnosed infectious disease cases collected routinely in each of Thailand’s 76 provinces by the Ministry of 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, and the severity of the illness including date of death for mortality cases. Malaria and DHF disease counts were created for each combination of month (72 months from January 1999 to 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 2000 Population and Housing Census of Thailand. Since there was little evidence of a gender effect, the data for the two sexes were not combined.

2.2 Statistical Analysis

Various statistical models are available for forecasting malaria and DHF incidence. Gomez-Elipe developed a model of time series of monthly notifications of malaria cases to predict malaria incidence in an area of unstable transmission by studying the association between environmental variables and disease dynamics. This model is a useful tool for producing reasonably reliable forecasts of the malaria incidence rate [6]. Briet shows that the addition of rainfall as a covariate improved prediction of selected (seasonal) ARIMA models moderately in some districts but worsened prediction in other districts [7]. In another study, geographical weighted regression (GWR) analysis was used to reveal a strong ($R^2 = 0.87$) positive spatial association between dengue fever prevalence and population distribution in the Johor State of Malaysia [8]. Fourier analysis and autoregressive modeling was used to remove the confounding effect of seasonal patterns. Some statistically significant but slight associations were found, suggesting that year–round seasonal climatic factors do not independently substantially affect the incidence of DHF [9]. Wongkoon developed a forecasting time series analysis model on the Dengue Haemorrhagic Fever (DHF) incidence in Northern Thailand [10].

The simplest model is based on linear regression with the outcome variable defined as the malaria and DHF incidence rate in a cell indexed by district effect, age group effect and period effect with half year (allowing for a seasonal effect) as categorical determinants. Such incidence rates generally have positively skewed distributions, so it is conventional to transform them by taking natural logarithms. Since period disease counts in small regions are often zero, it was necessary to omit districts with no cases of malaria or DHF in at least one age group. The method we used is to define the outcome as

$$y_{ijt} = \ln(\text{delta} + \frac{\text{no.cases(Dist,ageGrp,period)}}{\text{pop(Dist,ageGrp,period)/1,000}})$$

$$y_{ijt} = \ln\left(1 + \frac{n_i}{P}\right)$$

where $n_i$ is the number of disease cases in the cell, $P$ is
Modeling and Forecasting Malaria and Dengue Hemorrhagic Fever Incidence and Prevalence in Northern Thailand

the population at risk, and $K$ is a specified constant.

Malaria and DHF incidence rates were obtained from persons in four-year age groups for the period 1999 to 2004 in the North of Thailand. The model used is an extension of two-way analysis of variance (ANOVA) incorporating additional terms based on a principal components analysis in order to characterize these incidence rate patterns [11-12]. This technique is well known and has been applied to spectroscopy, chromatography and many other fields to model additive mixtures of overlapping curves [13-14]. The two-way ANOVA method is the simplest regression model that allows for district, age group and period, and is expressed as the additive combination.

In order to examine districts, age group and the period effects in a data array $y_{ijt}$ indexed by districts $i$, age group $j$ and periods $t$, respectively, a simple additive model can be used. This model can be written as:

$$y_{ijt} = \mu_i + \alpha_j + \beta_t + z_{ijt}$$

where $\mu_i$ represents the mean incidence in district $i$, $\alpha_j$ represents the age group effect $j$, $\beta_t$ represents the period effect $t$, and $z_{ijt}$ are the residuals. The principal components are also of interest because they describe the predominant temporal (seasonal and trend) patterns present in the data. Even where trends are relatively small compared to seasonal patterns, they can provide a useful basis for allocation of resources for disease prevention. The method of fitting the model is based on linear least squares regression, which assumes that the errors are normally distributed with common standard deviation [15-16]. Graphical methods [17] are used to check these assumptions.

The data may need to be transformed, using logarithms or power transformations. $R$ software [18] was used for graphical displays and statistical analysis.

3. Results

Between January 1999 and December 2004, 67,347 hospital cases of malaria and 6,577 hospital cases of DHF were reported in the 6 provinces (65 districts) previously mentioned. Districts with only a few cases were omitted from the analysis leaving 66,733 hospital cases of malaria (43 districts) and 5,980 hospital cases of DHF (36 districts). The number of cases in any half year period for a particular age group and district varied from 0 to 1,139 for malaria and from 0 to 175 for DHF. The maximum disease rates were 4.125 cases per 1,000 for malaria and 2.395 cases per 1,000 for DHF.
Initial additive models for the prevalence of malaria and DHF were fitted to the data with resulting R-squared values of 0.8055 and 0.4277, respectively. An additive plus multiplicative model yielded higher coefficients of determination to forecast the prevalence of malaria and DHF. The additive plus multiplicative model of malaria and DHF prevalence are shown in Fig. 1.

The model for malaria prevalence (number of cases per district per age group and period per 1,000 population) gave an R-squared of 0.8402, RSE of 0.906 indicating quite a good fit but it is not normality distributed whereas for DHF, the fit is not that good, but it is the normality assumption (R-squared = 0.462, RSE=0.9315).

Fig. 2 shows the age group, trend and district effects of malaria after adjusting for the overall mean. The malaria prevalence was highest in the 15-39 year age group (category 3). The malaria prevalence trend in the 6-month period decreased dramatically and was similar among the years 1999 and 2004. Malaria prevalence was high in Mae Hong Son and Tak provinces. Fig. 3 demonstrates the age group, trend and district effects of DHF after adjusting for the overall mean. The DHF prevalence was highest in the 5-14 year age group (category 2). The prevalence of DHF in each 6-month period was not stable with noticeable fluctuations over the five years. The prevalence of DHF was highest in Mae Sot district, Tak province.

Fig. 4 demonstrates the malaria trends for each district among 1999-2004 by average age group. The prevalence was high in Mae Hong Son and Tak provinces and slightly decreased during the 5-year period. The prevalence was highest in Umpang district, Tak province.

The prevalence of DHF is shown in Fig. 5.

DHF peaked in the 6-month period from January to June, 2001 in Lumphun city; between January and June 2002 in the districts of Song, Phrae province, Pua, Nan province, Wiang Pa Pao, Chiang Rai province and Mae Ramat, Tak province; and during the period from July to October 2003 in Mae Sot, Tak province.
to December, 2002 in Mae Hong Son city. Fig. 6 shows maps of malaria and DHF prevalence in the study region for each district.

To allow comparison between these, we used the count of malaria transmission to convert the malaria prevalence data at each of the 43 locations sampled and 36 for the DHF locations sampled to estimate transmission. We then fitted an additive plus multiplicative model to this data. The figure is the first detailed empirical map of variations in malaria and DHF transmission. There is considerable geographical variation in the precision of the model estimates. The

<table>
<thead>
<tr>
<th>Location</th>
<th>Prevalence/1000</th>
<th>Year</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lamphun</td>
<td>0.005</td>
<td>99</td>
<td>01</td>
</tr>
<tr>
<td>Phrae</td>
<td>0.05</td>
<td>99</td>
<td>02</td>
</tr>
<tr>
<td>Nan</td>
<td>0.2</td>
<td>99</td>
<td>03</td>
</tr>
<tr>
<td>Chiang Rai</td>
<td>0.5</td>
<td>99</td>
<td>04</td>
</tr>
<tr>
<td>Mae Hong Son</td>
<td>1</td>
<td>99</td>
<td>01</td>
</tr>
<tr>
<td>Tak</td>
<td>5</td>
<td>99</td>
<td>02</td>
</tr>
</tbody>
</table>

Fig. 4 Malaria trends for districts in northern Thailand 1999-2004: average of four age group.

<table>
<thead>
<tr>
<th>Location</th>
<th>Prevalence/1000</th>
<th>Year</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lamphun</td>
<td>0.005</td>
<td>99</td>
<td>01</td>
</tr>
<tr>
<td>Phrae</td>
<td>0.05</td>
<td>99</td>
<td>02</td>
</tr>
<tr>
<td>Nan</td>
<td>0.2</td>
<td>99</td>
<td>03</td>
</tr>
<tr>
<td>Chiang Rai</td>
<td>0.5</td>
<td>99</td>
<td>04</td>
</tr>
<tr>
<td>Mae Hong Son</td>
<td>1</td>
<td>99</td>
<td>01</td>
</tr>
<tr>
<td>Tak</td>
<td>5</td>
<td>99</td>
<td>02</td>
</tr>
</tbody>
</table>

Fig. 5 DHF trends for districts in northern Thailand 1999-2004: Average of four age groups.
4. Conclusions

We have shown that malaria and DHF are a serious health problem in the Northern region of Thailand. The additive plus multiplicative linear model provides an appropriate fit to age-group, districts, and period. The probability plot in this study was shown as an instrument for verifying the distributional assumption of a model and an effective way of capturing any unexpected increase in malaria and DHF counts. Also it provides information for malaria and DHF prevention.

According to this study, the prevalence of malaria was high among 15 to 39 year olds whereas DHF was high among 5 to 14 year olds. We found that the prevalence of malaria had decreased in the study period whereas DHF fluctuated radically. The prevalence of malaria and DHF has correlations with trend effects peaking at 0.9952 and 0.9818, respectively. In contrast, Devi investigated Pearson’s correlation analysis that relationship between climate variables and malaria transmission [19]. In addition, Bi was to explore the impact of climate variable on the transmission of malaria and indicate that climatic variables should be considered as possible predictors for regions with similar geographic, climatic, and socio-economic conditions to those of Shuchen County [20]. Hoshen developed a mathematical-biological model that may be suitable for the simulation of malaria forecasts based on seasonal weather forecasts [21]. Gagnon et al. found that there was a statistically significant relationship between El Niño and malaria epidemics [22]. Global warming will certainly affect the abundance and distribution of infectious disease and will cause changes in the epidemiology of infectious diseases. The ability of mankind to react or adapt is dependent upon the magnitude and speed of the change. The outcome will also depend on our ability to recognize epidemics early, to contain them effectively, to provide appropriate treatment, and to commit resources to prevention and research [23].

Whereas Promprou indicated that mean temperature, rainfall, and relative humidity were associated with
DHF incidence in the areas bordering the Andaman Sea [24]. Furthermore, Wiwanitkit showed that the correlation between the rainfall and the prevalence of dengue in a central region of Thailand, the highly endemic area, was investigated in 2004 [25]. Gagnon shows that there is a statistically significant correlation at the 95% confidence level between El Niño and dengue epidemics in French Guiana and Indonesia and at the 90% confidence level in Colombia and Surinam [26]. From this supporting paper, malaria and DHF trends for the districts in all six provinces in Northern Thailand 1999-2004 depend on seasonal variation but have different patterns. Malaria trends gradually decreased in this study period whereas DHF trends were highest and the lowest in over the same period.

Finally, we present maps of malaria and DHF prediction risk by using of additive plus multiplicative regression modeling to determine approximate risk on a larger scale and we employ geo-statistical approaches to improve prediction at a local level. Malaria and DHF prevalence was predicted by means of a model which used district, age group and period variables as potential predictors. After the regression analysis, spatial dependence of model local variation in malaria and DHF risk over and above that which is predicted by the regression model was used. The method is illustrated by a map showing the improvement of risk prediction brought about by the model. The advantage of this approach is shown in the context of development of methodology and R-software. Kleinschmit investigated malaria prevalence in children under 10 years 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 [27]. Bobra and Andrianasolo showed highlights of the statistical and spatial model development based on the analysis of socio-cultural practices adopted by dengue affected samples (DAS) and unaffected samples (UAS) and the application of GIS [28]. Nakhapakorn and Tripathi explore the potential of remotely sensed data and GIS technology to analyse the spatial factors affecting DF/DHF epidemic. An analysis of physic-environmental factors such as land use was carried out. Influence of these factors was obtained in quantitative terms using Information Value method in the GIS environment. It was found that built-up areas have the highest incidence and constitute the highest risk zones. Forest areas have no affect on DF/DHF epidemics. Agricultural areas present a moderate risk in DF/DHF incidences [5].

The map of malaria and DHF prevalence predicted risk in Northern border provinces of Thailand can be used by the Public Health Department as a base map for applying preventive measures to control the malaria and DHF outbreak. This will help in focusing the preventive measures being according to priority in high, average and low zones and help in saving time and money.

Acknowledgments

The authors thank McNeil D., Werasophon S. and Wongtra-ngan S. for encouragement and helpful guidance.

References


[18] V. Chongsuvivatwong, Analysis of epidemiological data using R and Epicalc, Epidemiology Unit, Prince of Songkla University, Thailand, 2008.


