Songklanakarin J. Sci. Technol.
30 (2), 207-214, Mar. - Apr. 2008

Songklanakarin Journal of Science and Technology
http://www.sjst.psu.ac.th

Original Article

Malaria in North - Western Thailand

Wattanavadee Sriwattanapongse[1]*, Metta Kuning[2], and Naratip Jansakul[3]

[1]*Department of Statistics, Faculty of Science, Chiang Mai University, Muang, Chiang Mai, 50200 Thailand.*

[2]*Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani, 94000 Thailand.*

[3]*Department of Mathematics, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, 90112 Thailand.*

Received 4 December 2006; Accepted 25 February 2008

Abstract

This study is based on the individual hospital case records of malaria routinely reported from 1999 to 2004 in the North-western area of Thailand, which included Mae Hong Son and Tak provinces. The objective of this study was to model the patterns of hospital-diagnosed malaria incidences by month, district and age-group for the two North-western border provinces in Thailand. The model used linear regression, Poisson regression and negative binomial regression to forecast the districts and age groups in which epidemics are likely to occur in the near future in order to prevent the disease by using suitable measures. Among the models fitted, the best were chosen based on the analysis of deviance and the negative binomial generalized linear model was clearly preferable. The model contains additive effects associated with the season of the year, district, age group and the malaria incidence rates in previous months, and can be used to provide useful short-term forecasts. Having a model that provides such forecasts of disease outbreaks, even if based purely on statistical data analysis, can provide a useful basis for allocation of resources for disease prevention.

Key words: malaria incidence, Mae Hong Son province, Tak province, negative binomial regression model, analysis of deviance

1. Introduction

Over 3 billion people live under the threat of malaria. The disease is a major health problem in the tropics, with at least 300 million clinical cases occurring annually. According to its web page in June 2006, the U.S. Centers for Disease Control and Prevention estimates that 0.7-2.7 million persons die of malaria each year, mostly among children under 5 years of age. The disease is caused by four species of parasites of the genus *Plasmodium* that are transmitted by the bite of infective female mosquitoes of the genus *Anopheles*. The immature stages of the vector's life cycle (egg, larva, and pupa) are aquatic and develop in breeding sites, whereas the aerial adult stage is terrestrial (Zucker 1996). Malaria in Thailand is forest-related and most prevalent along the international borders, especially on the Thai-Myanmar border where young men working in or near forests are at high risk. Although malaria cases and deaths have fallen substantially since 1999, the disease remains a considerable public health problem.

Our study aims are to find a suitable statistical model for predicting monthly incidence rates of reported hospital cases of malaria in districts of the two border provinces with high risk of disease in the north-western region of Thailand, based on routinely collected data available from provincial

*Corresponding author.
Email address: wattanavadees@yahoo.com

health offices. The provinces selected, Mae Hon Son and Tak, both border Myanmar and have high malaria incidence rates. Mae Hong Son occupies an area of 12,700 square kilometers and borders the Myanmar states of Shan, Kayan and Kayin. This province consists of the seven districts Muang Mae Hong Son, Khun Yuam, Pai, Mae Sariang, Mae La Noi, Sop Moei, and Pang Ma Pha. Mae Hong Son is mountainous and thus enjoys a cooler climate than other areas of Thailand. The Salween River forms part of the boundary with Myanmar. Tak occupies an area of 16,400 square kilometers, also borders the Kayin state, and comprises nine districts: Muang Tak, Ban Tak, Sam Ngao, Mae Ramat, Tha Song Yang, Mae Sod, Phop Phra, Umphang, and King Amphoe Wang Chao. The rainy season in each of the two provinces extends from May to October. The average monthly rainfall over the 30 years from 1961 to 1990 varied from 115 mm in October to 252 mm in August in Mae Hong Son and from 92 mm in July to 256 mm in September in Tak (Hong Kong Observatory 2003).

Based on an individual hospital case records routinely reported in each province from 1999-2004, linear regression models of log-transformed incidence rates were used to assess the effects of age, location and season of the year. Autoregressive terms were included to account for time series and spatial correlations. Given that the monthly disease counts in individual cells defined by age group and district were often small numbers with many zero occurrences, Poisson and negative binomial generalized linear models were more appropriate statistical models, and could be used to identify cells with unexpectedly high disease occurrences. Where substantial autocorrelation existed in the time series, such episodes might thus enable public health authorities to establish strategies for preventing outbreaks before they occur.

2. Methods

2.1 Data management

Data used in the current study were 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 were 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, sex, and address, and the severity of the illness including date of death for mortality cases. After cleaning to correct or impute data entry errors the records for the provinces were stored in an SQL database. SQL programs were used to create malaria disease counts by 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 2,000 Population and Housing Census of Thailand. Since there was little evidence of a sex effect the data for the two sexes were combined.

2.2 Linear regression

The simplest model is based on linear regression with the outcome variable defined as the incidence rate in a cell indexed by district, age group, and month, with district, age group and calendar month (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 logarithms. Since monthly disease counts based on small regions are often zero, it is necessary to make some adjustment to avoid taking logarithms of 0: the method we use is to define the outcome as

$$y = \ln\left(1 + K\frac{n}{P}\right), \qquad (1)$$

where n is the number of disease cases in the cell, P is the population at risk, and K is a specified constant. To allow for serial correlations in successive months, lagged incidence rates are included as additional determinants. Such an observation-driven model with m lags could take the form

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \eta_s + \sum_{k=1}^{m} \gamma_k y_{ij,t-k} + \varepsilon_{ijt}, \qquad (2)$$

where N_{ijt} is a random variable denoting the reported number of disease cases in age group i, district j and month t for the region of interest and n_{ijt} is the corresponding number observed, Y_{ijt} is the outcome variable specified in Equation (1) and y_{ijt} the corresponding number observed, ε_{ijt} comprises a set of independent normally distributed random variables with mean 0, and s = mod(t, 12). In this model we constrain the parameters so that $\alpha_1 = 0$, $\beta_1 = 0$ and $\eta_1 = 0$. While linear time trends could be included in the model, they are less useful for short-term forecasting purposes in the presence of high serial correlations, and are not considered in the present study.

To allow for possible spatial correlations between observations on different districts at the same time, and also for correlations between different age groups, additional terms allowing for these effects may be included as determinants in the model. A simple extension of the model (2) incorporating these effects takes the form

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \eta_s + \sum_{k=1}^{m} \gamma_k y_{ij,t-k} + \delta_1 y_{ij,t-1}^{(\alpha)} + \delta_2 y_{ij,t-1}^{(\beta)} + \varepsilon_{ijt}, \qquad (3)$$

where $y_{ijt}^{(\alpha)}$ and $y_{ijt}^{(\beta)}$ denote the observed (transformed) incidence rates in all age groups other than i and in all districts other than j, respectively.

2.3 Generalized linear models

Davis *et al.* (2003) suggested observation-driven models for time series counts N_t based on the Poisson distri-

bution with mean λ_t where $\ln(\lambda_t)$ is expressed as an additive function of determinants and lagged observations on. While these models are not appropriate for disease epidemics because they express the mean of the process at time t as an exponential function of lagged observations on the same process and are thus numerically unstable when substantial variations occur, they become stable when the lagged observation are replaced by logged incidence rates. Thus if p_{ij} is the population in age group i and district j and is the mean of a suitable generalized linear model based on the Poisson distribution could take the form:

$$\ln(\lambda_{ijt}) = \ln(p_{ij}) + \mu + \alpha_i + \beta_j + \eta_s + \sum_{k=1}^{m} \gamma_k y_{ij,t-k} + \delta_1 y_{ij,t-1}^{(\alpha)} + \delta_2 y_{ij,t-1}^{(\beta)}. \quad (4)$$

Poisson models for disease counts are often over-dispersed due to spatial or temporal clustering of cases (Ruru and Barrios, 2003), in which case the negative binomial distribution may be more appropriate. This distribution has an additional parameter and takes the form

$$\text{Prob}\,(N_t = n) = \frac{\Gamma(n+\gamma)}{\Gamma(n+1)\Gamma(\gamma)} \left(\frac{\gamma}{\gamma+\lambda_t}\right)^{\gamma} \left(\frac{\lambda_t}{\gamma+\lambda_t}\right)^{n}. \quad (5)$$

As for the Poisson model is the conditional expected value of, but the conditional variance is $\lambda_t + \lambda_t^2/\gamma$ (see, for example, Jansakul and Hinde, 2004). The parameter is actually inversely related to the over-dispersion, so that the Poisson model arises as the special case in the limit as $\gamma \to \infty$.

2.4 Assessment of model based on residual plots

In conventional linear regression models where errors are assumed to be independent and normally distributed, the adequacy of the model can be assessed by plotting residuals, obtained from the observations simply by subtracting their conditional means, against corresponding normal scores. For count data where a generalized linear model is fitted, (Pearson) residuals are defined by subtracting the conditional means from the counts and then dividing by the conditional standard deviations (see for example, Kedem and Fokianos 2002). These residuals can be plotted against scores based on the appropriate asymptotic distribution. For Poisson-distributed models, this asymptotic distribution is the normal distribution so normal scores can still be used.

For negative binomial time series models, the asymptotic distribution of the standardized residuals can be derived using moment generating functions. The moment generating function for the negative binomial distribution given by Equation (5) is (Wackerly *et al.*, 1996)

$$E[\exp(\theta N_t)] = \left(1 + \frac{\lambda_t}{\gamma} - \frac{\lambda_t}{\gamma} e^{\theta}\right)^{-\gamma}, \quad (6)$$

so the moment generating function for the standardized residuals is

$$E\left[\exp\left(\frac{\theta(N_t - \lambda_t)}{\sqrt{\lambda_t + \lambda_t^2/\gamma}}\right)\right] = \left(1 + \frac{\lambda_t}{\gamma} - \frac{\lambda_t}{\gamma}\exp\left(\theta/\sqrt{\lambda_t + \lambda_t^2/\gamma}\right)\right)^{-\gamma} \exp\left(-\theta\sqrt{\lambda_t}/\sqrt{1+\lambda_t/\gamma}\right) \quad (7)$$

Using a Taylor expansion, the right-hand side of Equation (7) reduces to

$$\left(1 - \frac{\theta}{\gamma\sqrt{1/\lambda_t + 1/\gamma}} - \frac{\theta^2}{2\gamma(1+\lambda_t/\gamma)}\right)^{-\gamma} \exp\left(-\theta\sqrt{\gamma}\right) + o(1/\lambda_t),$$

where $o(x)$ is any function that tends to zero faster than x. Since the limit of $(1+x/n)^{-n}$ as $n \to \infty$ is e^{-x}, it follows that the limiting distribution of the Pearson residuals has moment generating function

$$(1 - \theta/\sqrt{\gamma})^{-\gamma} \exp(-\theta\sqrt{\gamma}). \quad (8)$$

This corresponds to a gamma distribution with shape parameter γ and scale parameter $1/\sqrt{\gamma}$, shifted to the left by $\sqrt{\gamma}$.

3. Results

3.1 Distributions of incidence rates

During the study period from January 1999 to December 2004, 29,498 hospital cases of malaria were reported in Mae Hong Son province and 31,658 in Tak province. The number of cases in a month for a particular age group and district varied from zero to 460 in Mae Hong Son and from 0 to 177 in Tak, and the corresponding maximum disease rates were 24.7 cases per 1,000 in Mae Hong Son and 19.0 cases per 1000 in Tak.The time series of average monthly age-specific malaria rates per 1,000 population at risk for all districts in each province are plotted in Figure 1. In Mae Hong Son the rates show a marked seasonal periodicity, and decreased from very high levels in 1999 to around 1 case per 1,000 per month in recent years, whereas in Tak the reported malaria rate shows a less pronounced seasonal pattern and decreased relatively slightly over the same period. The age patterns are similar in each province.

Figure 2 shows that how the average malaria incidence rate varied by district in the two provinces. The lowest rates occurred in the four non-border districts, and the highest occurred in the southernmost district of Tak (Um Phang: 49.5 cases/1,000/year) and Mae Hong Son city in the north (31.3 cases/1,000/year).

3.2 Linear regression

Table 1 shows the results obtained from fitting the linear regression model given by Equation (3) to the log-transformed incidence rates for each of the two provinces. In each case we took K = 10,000, giving reasonably linear normal scores plots. The number of zero counts is 239 (12.4%) for Mae Hong Son province, compared with 605 (24.4%) for Tak province where the disease incidence rates are

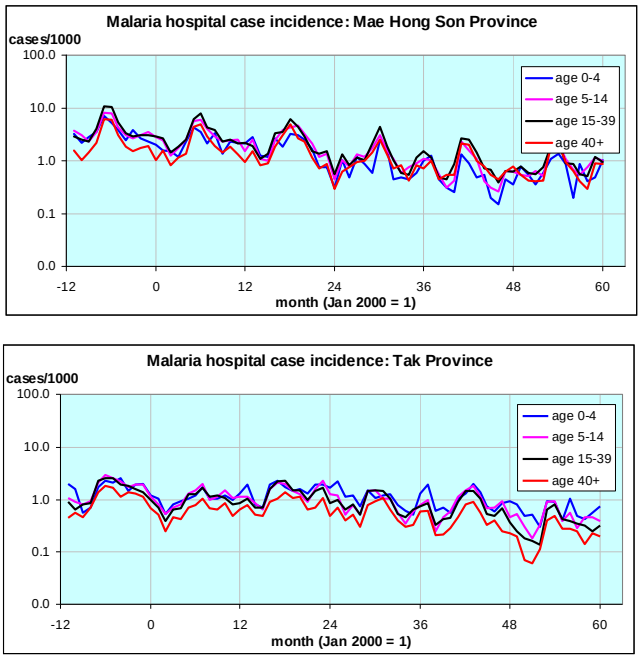

Figure 1. Time series of monthly disease rates for each age group in the two provinces

much lower. Choosing the much smaller value K = 2,000 increases the r-squared for Mae Hong Son from 0.6121 to 0.642, and also reduces the gap between the zero-valued incidence rates and the smallest non-zero incidence rate from 0.56 to 0.14. However, it is questionable whether choosing the smaller value of K provides a better model for identifying unusually high disease incidence rates in future months. This question is considered further in relation to the generalized linear models. While all components in the model for each province are statistically significant, the lagged incidence rates account for the largest single contribution to the r-squared statistic (45.6% for Mae Hong Son and 74.9% for Tak), and the coefficients incorporating the further correlations between age groups and districts are also quite substantial. The largest residual obtained for Mae Hong Son is 2.58, corresponding to 3 cases reported among infants below 5 in Khum Yuam district in September 2002 (incidence rate 1.7 per 1,000). However, this residual does not show up as an outlier on the normal scores plot, and the numbers of cases reported in the same district and age group in the following months were small (0, 1, 1, 2 and 1, respectively). But in Tak province the highest residual of 2.46, corresponding to 16 cases that occurred among infants below 5 years of age in Phop Pra district in February 2001 (incidence rate 2.7 per 1,000), heralded a small epidemic comprising 5, 6, 17, 16 and 13 cases in the following five months.

3.3 Poisson and negative binomial generalized linear models

Turning to the Poisson and negative binomial regression models given by Equations (4) and (5), Figure 3 shows plots of Pearson residuals versus corresponding (normal or gamma) scores. The Poisson model gives residual deviances of 8,135.9 for Mae Hong Son and 8,898.4 for Tak. The negative binomial model gives residual deviances of 2,212.7 and 2,834.2, respectively, so the negative binomial model is clearly preferable. Some care is needed with the choice of the constant K, because with lagged incidence rates needed to account for substantial correlations between adjacent cells, values for K outside a certain range make these models numerically unstable. For the Poisson model the sum of the predicted disease counts equals the sum of the observed counts. However, when the negative binomial model is fitted using maximum likelihood rather than moment estimators this constraint is not necessarily satisfied, and satisfying this requirement could govern the choice of K. We found that choosing K = 10,000 gives sums of predicted disease counts that are reasonably close to the observed sums.

Table 2 gives the results obtained from fitting the negative regression model given by Equations (4) and (5) to the malaria disease counts for each of the two provinces. The dispersion parameter estimates are 3.521 for Mae Hong Son and 3.725 for Tak. The largest Pearson residual obtained for Mae Hong Son is 8.09, corresponding to 52 cases reported among young adults in Mae Sariang district in January 2002. Since 385 further cases were reported in the same district and age group in the following six months (32, 36, 18, 78, 168 and 53, respectively), and a further 203 cases were reported in the six months after that, this particular outlier

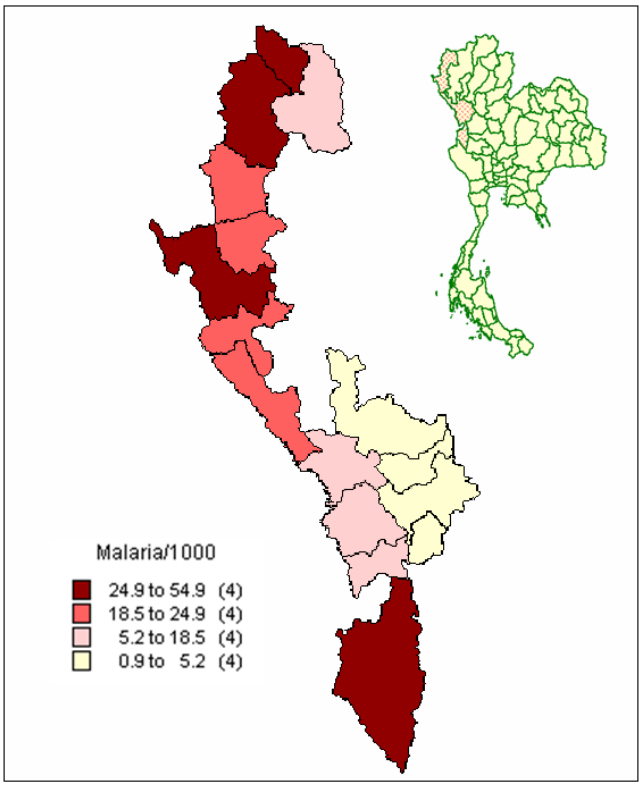

Figure 2. Average annual incidence rates for malaria in Mae Hong Son and Tak provinces, 1999-2004

Table 1. Results from fitting linear model to log-transformed incidence rates

Determinant		Mae Hong Son		Tak	
		Coefficient	St. Error	Coefficient	St. Error
Constant		-0.349	0.108	-0.498	0.094
Age Group:	0-4	0	-	0	-
	5-14	0.226	0.051	0.086	0.034
	15-39	0.402	0.056	0.186	0.036
	40+	0.201	0.050	0.112	0.039
District:	1	0	-	0	-
	2	-0.239	0.068	0.022	0.051
	3	-0.474	0.080	-0.018	0.051
	4	-0.056	0.066	0.609	0.073
	5	-0.230	0.068	0.704	0.081
	6	-0.032	0.066	0.248	0.056
	7	-0.194	0.067	0.325	0.060
	8			0.913	0.094
	9			0.113	0.052
Month:	January	0	-	0	-
	February	0.586	0.089	0.256	0.060
	March	-0.214	0.088	-0.162	0.060
	April	0.306	0.090	0.202	0.062
	May	0.570	0.085	0.186	0.058
	June	1.107	0.084	0.735	0.058
	July	0.675	0.089	0.574	0.059
	August	0.021	0.090	0.338	0.060
	September	-0.088	0.087	0.081	0.058
	October	-0.090	0.087	-0.134	0.058
	November	0.136	0.086	0.060	0.058
	December	0.508	0.085	0.122	0.057
Autoregressive Lag:	1	0.224	0.026	0.291	0.024
	2	0.091	0.024	0.146	0.021
	3	0.075	0.022	0.113	0.020
Other Age Groups		0.341	0.034	0.172	0.028
Other Districts		0.211	0.036	0.148	0.031
R-squared Statistic		0.6121		0.8099	

heralded a large epidemic. The next largest two residuals also correspond to clusters of cases followed by substantial outbreaks in the following six months (14 cases followed by another 107, and 26 cases followed by another 97, respectively). In Tak province the highest Pearson residual of 10.99, corresponds to the 16 cases occurring among infants below 5 years of age in Phop Pra district in February 2001, already noted in the linear regression modeling of the log-transformed incidence rates. In contrast, the second highest Pearson residual for Tak, 8.60, corresponds to an isolated outbreak of 21 cases among older adults in Tak city in July 2003, with no further cases recorded in this district and age group in the following six months.

4. Summary

We have shown that malaria is a serious health problem in Mae Hon Son and Tak. The negative binomial generalized linear model provides a perfect fit to age-group, districts, and month. The probability plot in this study was shown as an instrument for verifying the distributional assumption of a model and an effective way in capturing any unexpected increase in malaria counts. Also it provides information for malaria prevention. We found the prevalence of malaria was high among aged 15 to 39 year olds. In contrast, Kaewsompak *et al.* (2005) used the negative binomial distribution to model the incidence rate of commonly

Table 2. Results from fitting negative binomial models to malaria disease counts

Determinant		Mae Hong Son		Tak	
		Coefficient	St. Error	Coefficient	St. Error
Constant	0.098	0.101	-1.357	0.120	
Age Group:	0-4	0	0	0	0
	5-14	0.043	0.049	0.010	0.044
	15-39	0.149	0.050	0.092	0.043
	40+	-0.039	0.048	0.084	0.048
District:	1	0	0	0	0
	2	-0.141	0.060	0.225	0.080
	3	-0.430	0.074	0.173	0.087
	4	-0.069	0.053	1.036	0.091
	5	-0.161	0.057	1.140	0.101
	6	-0.029	0.054	0.592	0.072
	7	-0.092	0.058	0.833	0.080
	8			1.312	0.117
	9			0.603	0.079
Month:	January	0	0	0	0
	February	0.530	0.080	0.239	0.074
	March	-0.287	0.084	-0.400	0.078
	April	0.225	0.085	0.142	0.079
	May	0.548	0.076	0.246	0.073
	June	1.153	0.073	0.914	0.069
	July	0.661	0.077	0.627	0.070
	August	-0.049	0.080	0.372	0.071
	September	-0.128	0.078	0.008	0.070
	October	-0.140	0.080	-0.228	0.071
	November	0.041	0.080	0.032	0.071
	December	0.428	0.077	0.081	0.071
Autoregressive Lag:	1	0.292	0.027	0.412	0.032
	2	0.090	0.024	0.168	0.028
	3	0.069	0.022	0.145	0.026
Other Age Groups		0.284	0.033	0.121	0.036
Other Districts		0.161	0.032	0.247	0.038
Dispersion: *g*		3.521	0.181	3.725	0.188
Residual Deviance		2212.7		2834.2	

occurring acute febrile illnesses in subdistricts of Yala province between 2002 and 2003 and found the relationship between geographic locations, age, time effect, and malaria incidence. The prevalence of malaria was high among people aged 39 years old or more between April and December. It is also different from Kleinschmidt and Sharp (2001),who suggested that the malaria was more prevalent among children under 15 years old. The differences may be due to (a) the transmission patterns, the climate, living condition and living density in which may be very different from other tropical areas, (b) the immune status and genetic background, and (c) the range of age groups. Such as age shift is supported by our study, and also clearly evidenced in several other recent studies. From 1990 to 1996, malaria rates were the highest in the 15 to 34 year olds.

According to this study, the malaria transmission rate in Mae Hong Son and Tak during the study period was relatively high from May to September. That period largely overlaps with the rainy season, suggesting that this may be associated with a high risk of malaria. A study by Gagnon *et al.* (2002) also reported a statistically significant relationship between El Niño and malaria epidemics in Colombia, Guyana, Peru, and Venezuela.In most of these countries, the prevalence was the highest in the wet season. However, they also suggested that on an inter-annual scale, malaria was also associated with drought. The high malaria incidence in Mae

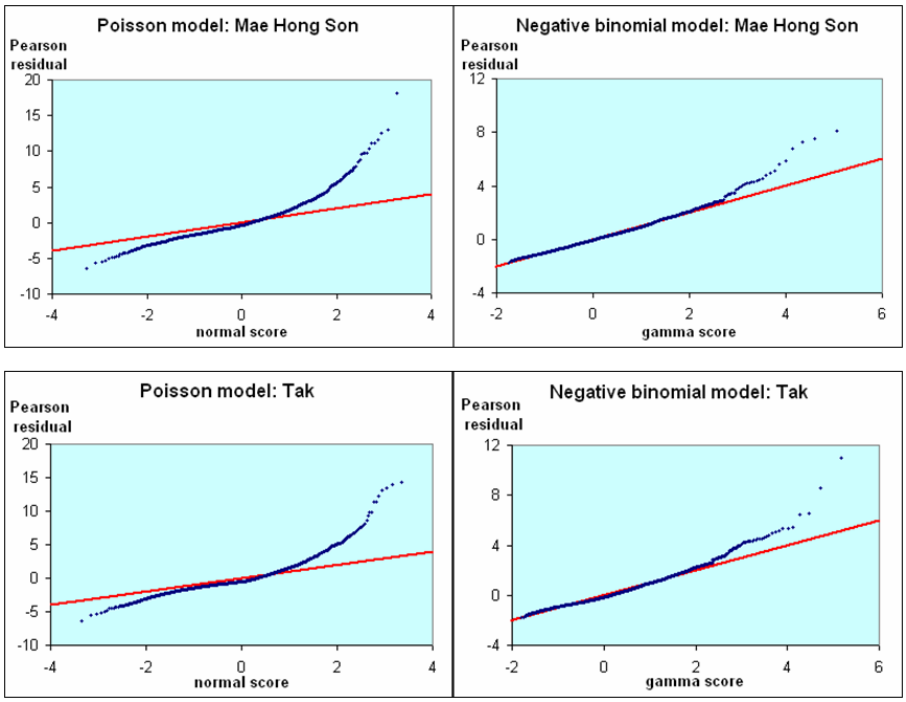

Figure 3. Plots of Pearson residuals versus asymptotic scores after fitting Poisson and negative binomial models for malaria disease counts in the two provinces

Hong Son is in Muang Mae Hon Son and Mae La Noi which is explained by these geographical location and significant social economic status. For Tak is the high malaria incidences are in Um Phang and Tha Song Yang due to their location. Um Phang is extremely vulnerable to natural deep forest area such as Thu Yai Nareasuan (Srivastava *et al.*, 2001) because of (a) the mass migration of refugees in response to civil disturbances; (b) the influx of workers to areas undergoing rapid urbanization and development (Chaveepojnkamjorn and Pichainarong, 2005), and (c) the development of the tourist industry. To gain more understanding factors associated with the risk of malaria, further studies should be undertaken to examine the relationships between climate variables including rainfall, humidity, temperature and geographical location and malaria on both regional and global scales (Zhou, 2004).

Acknowledgements

We would like express our gratitude to Don McNeil, Phattrawan Tongkumchum, Chamnien Choonpradub and Supreeya Wongtrangan for their invaluable assistance, encouragement and helpful guidance.

References

Chaveepojnkamjorn, W. and Pichainarong, N. 2005. Behavioral factors and malaria infection among the migrant population, Chiang Rai Province, Journal of The Medical Association of Thailand ., 88(9): 1293-301.

Davis, R.A., Dunsmuir, W.T.M. and Streett, S.B. 2003. Observation-driven models for Poisson count, Biometrika., 90(4): 777-790.

Gagnon, A.S., Smoyer-Tomic, K.E. and Bush, A.B. 2002. The El Ni□o Southern Oscillation and malaria epidemics in South America, International Journal of Biometeorology., 46(2): 81-89.

Hong Kong Observatory. 2003. Hong Kong Observatory: The government of the Hong Kong special administrative region. http://www.hko.gov.hk/contente.htm.

Jansakul, N. and Hinde, J.P. 2004. Linear mean-variance negative binomial models for analysis of orange tissue-culture data, Songklanakarin Journal of Science and Technology., 26(5): 683-696.

Kaewsompak, S., Boonpradit, S., Choonpradub, C. and Chaisuksant, Y. 2005. Mapping acute febrile illness incidence in Yala province, Songklanakarin Medical Journal., 23(6): 455-462

Kedem, B. and Fokianos, K. 2002. Regression Models for Time Series Analysis, John Wiley & Son, UK.

Kleinschmidt, I. and Sharp, B. 2001. Patterns in age-specific malaria incidence in a population exposed to low levels of malaria transmission intensity, Tropical Medicine & International Health., 6(12): 986.

Ruru, Y. and Barrios, E.B. 2003. Poisson regression models of malaria incidence in Jayapura: Indonesia, The Philippine Statistician., 52(1-4): 27-38.

Srivastava, A., Nagpal, B.N., Saxena, R. and Subbarao, S.K. 2001. Predictive habitat modelling for forest malaria

vector species *An. dirus* in India - A GIS-based approach, Current Science., 80(9): 1129-34.

Wackerly, D.D., Mendenhall, W. and Schaeffer, R.L. 1996. Mathematical statistics with applications, 5th edition, Duxbury Press.

Zhou, G., Minakawa, N., Githeko, A.K. and Yan, G. 2004. Association between climate variability and malaria epidemics in the East African highlands, Proceedings of the National Academy of Sciences of the United State of America, 2004 Sep 14;101(37): 2375-80.

Zucker, J. 1996. Changing patterns of autochthonous malaria transmission in the United States: A review of recent outbreaks, Emerging Infectious Diseases., 2(1): 37-43.