# A MIXTURE MODEL APPLICATION IN DISEASE MAPPING OF MALARIA

Sasivimol Rattanasiri[1], Dankmar Böhning[2], Piangchan Rojanavipart[3] and Suthi Athipanyakom[3]

[1]Clinical Epidemiology Unit, Office of the Dean, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok; [2]Biometry and Epidemiology, Institute of International Health, Free University of Berlin, Berlin, Germany; [3]Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok, Thailand

**Abstract.** Disease mapping, a method for displaying the geographical distribution of disease occurrence, has received attention for more than 2 decades. Because traditional approaches to disease mapping have some deficiencies and disadvantages in presenting the geographical distribution of disease, the mixture model-as an alternative approach-overcomes some of these deficiencies and provides a clearer picture of the spatial risk structure. The purpose of this study was twofold: (1) to investigate the geographical distribution of malaria in Thailand during 1995, 1996, and 1997 by applying the mixture model to disease mapping, and (2) to investigate the dynamic nature of malaria in Thailand during the 3-year time frame by applying the space-time mixture model. Non-parametric maximum likelihood estimation was employed to estimate the parameters of both the mixture model and the space-time mixture model. Applying Bayes' theorem, the 76 provinces of Thailand were classified into component risk levels by the rate of malaria for each province. Malaria intensively occurred in 4 provinces on the Thai-Myanmar border and in 2 provinces on the Thai-Cambodian border. Of the 76 provinces studied, 10 showed an increasing trend over the 3-year period. A comparison of the map based on the mixture model with the map based on the traditional percentiles method indicates that the non-parametric mixture model removes random variability from the map and provides a clearer picture of the spatial risk structure. The advantage of the mixture model approach to disease mapping is the graphical visual presentation of the prevalence of disease. The space-time mixture model more adequately investigates the dynamic nature of disease than does the mixture model.

## INTRODUCTION

Disease mapping, a method for displaying the geographical distribution of disease occurrence, has received attention for more than 2 decades. The traditional percentiles method frequently is applied to the Standardized Mortality Ratio (SMR) as the epidemiological measure under consideration. However, this approach has been criticized. One criticism is that classification based on percentiles is rather arbitrary, because there is no guarantee that such a classification can validly detect high or low risk areas (Schlattmann *et al*, 1993a). Another problem involves the instability of the crude SMR, especially when rare diseases are investigated in an area with a small population. In such a case, both the observed and the expected values are low. As a result, an area with a small population tends to present an extreme SMR, yielding a map which is dominated by the least reliable information (Bernardinelli *et al*, 1992; Heisterkamp *et al*, 1993).

Another traditional method, the significant method, is based on a classification using the p-value. However, a disease map which is based on this approach often faces the problem of mis-classification as well, because an area with a small population size has a greater chance of showing a significant result (Böhning, 1999). Additionally, the significant method approach faces the problem of multiple testing, and even adjusting for the number of comparisons does not lead to a consistent estimate of heterogeneity (Schlattmann *et al*, 1999).

Since both of these traditional approaches have some deficiencies and disadvantages in representing the geographical distribution of disease, many researchers have sought alternative solutions for

Correspondence: Sasivimol Rattanasiri, Clinical Epidemiology Unit, Office of the Dean, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, 10400 Bangkok, Thailand.
Tel: 66 (0) 2201-1269; Fax: 66 (0) 2201-1774
E-mail: tesvm@mahidol.ac.th

mapping disease. Empirical Bayes (EB) estimation provides a more stable relative risk estimate, and thereby overcomes some deficiencies of traditional maps which are based on the SMR. It was found that the EB approach removes the random variability which is present in data from small population counts (Böhning, 1999; Böhning and Schlattmann, 1999), leading to a smooth map with fewer extremes in the relative risk estimates (Clayton and Kaldor, 1987; Marshall, 1991; Mollie and Richardson, 1991; Devine and Louis, 1994). However, the EB approach lacks a post hoc classification of the posterior estimate of the epidemiological measure.

As a solution to the foregoing, mixture modelling has been proposed for disease mapping (Böhning and Schlattmann, 1999). The mixture model approach more appropriately reduces the random variation in the disease map than do the percentiles method, the significant method, or the EB estimation. A disease map based on the mixture model approach not only provides a shrinkage estimator in the form of the mean of the posterior distribution, but also provides an estimate of the underlying risk structure (Schlattmann *et al*, 1999). Another methodological advantage of using the mixture model for disease mapping is that an estimate of the number of components (each with its respective coloring pattern) is provided (Schlattmann *et al*, 1999; Böhning and Schlattmann, 1999). In a simulation study, the mixture model approach was compared with traditional approaches of map construction, such as using the percentiles method or the significant method (Böhning and Schlattmann, 1999). The results indicated that the mixture model approach provides a significantly higher percentage of correct classifications than do the traditional methods.

The investigation of the geographical distribution of malaria is essential for malaria control programs. The detection of geographical heterogeneity by disease mapping constitutes a simple screening procedure so that the managers of disease control programs are rationally able to use interventions which are most likely to succeed. Since malaria is an increasingly serious problem in some provinces of Thailand, the purpose of our study was twofold: (1) to investigate the geographical distribution of malaria in Thailand during 1995, 1996, and 1997 by applying the mixture model to disease mapping, and (2) to investigate the dynamic nature of malaria in Thailand during the 3-year time frame by applying the space-time mixture model.

## METHODS

### Data sources

The basic geographical aggregation unit for disease mapping in this study was the province, of which there are 76 in Thailand. The Standardized Incidence Ratio (SIR) was used to measure the occurrence of malaria in each province.

The total number of people by age group in each province was obtained from the database of the Statistical Data Bank and Information Dissemination Division, National Statistical Office, Thailand. The malaria morbidity data (consisting of the number of observed malaria cases by age group in each province) were taken from cases which had been reported to the Division of Epidemiology, Ministry of Public Health, Bangkok. As the age of the population affects the incidence of malaria, the age-standardized incidence ratio was considered to be the proper epidemiological measure for this study. We used the indirect standardized method for calculating the age-standardized incidence ratio to determine the rate of occurrence of malaria for each province (Böhning, 1998).

### Statistical analysis

Non-parametric maximum likelihood estimation (NPMLE) was used to estimate the parameters of both the mixture model and the space-time mixture model (the number of components, and a mean of SIR and a weight for each component). After determining the number of components, a Maximum Likelihood Estimation (MLE) was used to estimate a mean of SIR and a weight for each component. Applying Bayes' theorem, we then assigned each province to a component risk level.

### Disease mapping with the mixture model approach

The simplest and most natural derivation of the mixture model arises when one sample comes from a population that consists of several homogeneous components, which then becomes a het-

erogeneous case (Lindsay, 1995). In observing only the sample $x_{1,...,}x_n$ from the marginal density of X (the mixture density), there is no consideration given to a specific component. The mixture model provides a solution to this problem (Böhning, 1999).

The mixture model approach to disease mapping assumes that the population under scrutiny consists of components with different risk levels of disease. Each component has a risk of disease ($\lambda_j$) and represents a certain proportion ($p_j$) of the total regional unit (Böhning and Schlattmann, 1999; Schlattmann *et al*, 1999). The first step of disease mapping with the mixture model approach is to estimate $\lambda_j$ and $p_j$ in each component. When we assume that the malaria cases ($o_i$) follow a Poisson distribution with mean $E_i\lambda$ given the area and $\lambda$), $o_i$ unconditionally follows a mixture of the Poisson distribution as:

$$o_i \sim \sum_{j=1}^{k} f(x_i;\lambda_j)\, p_j = \sum_{j=1}^{k} Po(o_i;\lambda_j E_i)\, p_j = f(o_i, E_i, P)$$

where

$x_i = o_i / E_i$ is the observed SIR in area *i*, $i$ = 1,...,n,

$o_i$ is the observed number of malaria cases in area *i*, $i$ = 1,...,n,

$E_i$ is the expected malaria cases in area *i*, $i$ = 1,...,n,

$\lambda_j$ is the level of disease risk in subpopulation j,

$p_j$ is the probability of belonging to the *j*th subpopulation, and

k is the number of components in the mixing distribution.

Estimation was done by maximum likelihood. $\hat{P}$, the non-parametric maximum likelihood estimator (NPMLE), was found by maximizing the (marginal) log-likelihood function which is defined as:

$$l(P) = \sum_{i=1}^{n} \log f(x_i P) = \sum_{i=1}^{n} \log \{\sum_{j=1}^{k} p_j f(x_i,\lambda_j)\}$$

The suitable algorithm for maximizing the log-likelihood function was taken from Böhning *et al* (1992). This algorithm is implemented in the computer packages DismapWin (Schlattmann, 1996) or C.A.MAN (Schlattmann *et al*, 1993b).

The second step was to determine the number of components. This was done by computing the Likelihood Ratio Statistic (LRS) for testing the hypothesis:

$H_0$: number of components = k

against

$H_a$: number of components = k+1.

The likelihood ratio test can be defined as:

$$LRS = 2\log\xi_n = 2 \times [l(\hat{P}_{k+1}) - l(\hat{P}_k)]$$

where

$\hat{P}_k$ is the maximum likelihood estimator under $H_0$, and

$\hat{P}_{k+1}$ is the maximum likelihood estimator under Ha.

The LRS test conventionally has been an asymptotic $\chi^2$ distribution with *d* degrees of freedom, where the degrees of freedom, *d*, are equal to the difference between the number of parameters both in the null and in the alternative hypotheses. However, because this theory is known to fail for the mixture problem, critical values in our study were obtained from a simulation technique which had been developed by Böhning (1999).

The final step in the mixture model approach for disease mapping was to classify the SIR value for each area into one of the components of the mixing distribution. This was accomplished by applying Bayes' theorem and by using the estimated mixing distribution as the prior distribution. Classification was done by computing the probability of each area belonging to a certain component. When $Z_{ij}$ is the unobserved variable which describes area *i* in subpopulation *j* ($Z_{ij}$), the posterior probability is defined as:

$$\Pr(Z_{ij} = 1 | o_i, \hat{P}, E_i) = \frac{\hat{p}_j f(o_i, \hat{\lambda}_j, E_i)}{\sum_{i=1}^{k} \hat{p}_j f(o_i, \hat{\lambda}_j, E_i)}, \text{ for } j = 1,...,k \text{ and } i = 1,...,n.$$

The *i*th area is then assigned to that subpopulation *j* for which it has the highest posterior probability of belonging (Böhning and Schlattman, 1999).

**Disease mapping with space-time mixture modeling**

Because the space-time mixture model gives a valuable indication of an emerging pattern over time, it was introduced to investigate the fluctuation of the occurrence of malaria from 1995 through 1997. This model looks simultaneously

for all space-time components (clusters) (Böhning *et al*, 2000). Because the space-time mixture model gives fewer parameters for comparison, the fluctuation of a certain disease over time is easier to compare and interpret.

The basic idea of disease mapping with the space-time mixture model approach is to consider the space-time data as one data set which models a single mixture distribution. When $O_{it}$ and $E_{it}$ are respectively the observed and the expected cases for area $i$, $i = 1,...,n$, and time $t$, $t = 1,...,T$, the mixture density is defined as:

$$f(o_{it}, P, E_{it}) = \sum_{j=1}^{k} p_j f(o_{it}, \lambda_j, E_{it}) = \sum_{j=1}^{k} P_j f(x_{it}, \lambda_j)$$

with

$$\sum_{j=1}^{k} p_j = 1 \text{ and } p_j \geq 0, \text{ for } j = 1,...,k$$

and, in this case, the mixture log-likelihood is:

$$l(P) = \sum_{t=1}^{T} \sum_{j=1}^{n} \log\{\sum_{j=1}^{k} p_j f(x_{it}, \lambda_j)\}$$

where

$x_{it} = o_{it}/E_{it}$ is the observed SIR in area $i$ and time period $t$.

Estimation of the parameters of this model was done by repeating the foregoing process. Classification of the areas into the $T$ maps was done again with the posterior probability, which is defined as:

$$\Pr(Z_{itj} = 1 \mid o_{it}, \hat{P}, E_{it}) = \frac{\hat{p}_j f(o_{it}, \hat{\lambda}_j, E_{it})}{\sum_{l=1}^{k} \hat{p}_l f(o_{it}, \hat{\lambda}_l, E_{it})},$$

so that area $i$ in time period $t$ is classified in that component $j$ for which it has the highest posterior probability of belonging. (Note that each area is classified T times when using this classification rule).

## RESULTS

For 1995 we obtained 7 components from NPMLE. Based on the NPMLE, MLE for the lower numbered component, the corresponding log-likelihoods for all k components, and the associated LRS (Table 1), the smallest number of components compatible with malaria data in 1995 consisted of 5 components (Table 2).

Category $\lambda_1$ had the lowest risk with a mean SIR of 0.08 and a weight of 61%, and category $\lambda_5$ had the highest risk with a mean SIR of 39.55 and a weight of 1% (Table 2). The corresponding maps (with risk categories) for the three time periods (1995, 1996, and 1997) of malaria data are illustrated in Fig 1.

For application of the space-time mixture model to our data set for 1995 through 1997, we obtained 8 components from NPMLE. Based on the NPMLE, MLE for the lower numbered component, the corresponding log-likelihoods for all k components, and the associated LRS, we found that k=6 was the optimum component for this data set (Table 3). The resulting MLE which is compatible with malaria data for 1995 through 1997 is shown in Table 4.

Category $\lambda_1$ had the lowest risk with a mean SIR of 0.09 and a weight of 59%, and category $\lambda_6$ had the highest risk with a mean SIR of 29.16 and a weight of 2% (Table 4). Based on the classification rule for map construction (Böhning, 1999), we constructed the corresponding maps for the rate of malaria for each the three time periods (Fig 2). Sixty (78.95%) provinces did not change their allocation, and 16 (21.05%) provinces changed their allocation. Of the 10 (13.16%) provinces that changed from a lower to a higher risk component, 3 (3.94%) provinces increased significantly. Six (7.89%) provinces changed from a higher to a lower risk component.

## DISCUSSION

Disease mapping has been studied and developed by several researchers for a long time; however, the traditional methods of disease mapping still have some deficiencies. For instance, the percentiles method has the potential danger of misrepresenting the geographical distribution of the measure of interest (Cislaghi *et al*, 1995), that is, a considerable overestimation of high risk areas appears in a disease map based on this method. The mixture model application for disease mapping is an alternative approach which satisfactorily produces a smooth map in which random variability has been extracted from the

Table 1
NPMLE[a], MLE[b] for lower component models, corresponding log-likelihoods, and LRS[c] (1995).

| Component | Parameter[d] | Weight | Log-likelihood$_k$ | -2($\log_k$-$\log_{k+1}$) |
|---|---|---|---|---|
| k=7 | 0.0827 | 0.6073 | -193.3718 | |
| | 0.7383 | 0.2347 | | |
| | 3.1964 | 0.0926 | | |
| | 8.2728 | 0.0127 | | |
| | 18.6119 | 0.0251 | | |
| | 26.0796 | 0.0144 | | |
| | 39.5625 | 0.0131 | | |
| | | | | 3.6330 |
| k=6 | 0.0837 | 0.6119 | -195.1883 | |
| | 0.7702 | 0.2361 | | |
| | 3.6113 | 0.0991 | | |
| | 18.5737 | 0.0253 | | |
| | 26.0818 | 0.0144 | | |
| | 39.5925 | 0.0131 | | |
| | | | | 3.6012[e] |
| k=5 | 0.0837 | 0.6120 | -196.9889 | |
| | 0.7705 | 0.2361 | | |
| | 3.6168 | 0.0992 | | |
| | 21.0612 | 0.0395 | | |
| | 39.5474 | 0.0132 | | |
| | | | | 16.9312 |
| k=4 | 0.0837 | 0.6120 | -205.4545 | |
| | 0.7706 | 0.2361 | | |
| | 3.6173 | 0.0993 | | |
| | 23.4826 | 0.0526 | | |
| | | | | 12.5190 |
| k=3 | 0.1597 | 0.7922 | -211.7140 | |
| | 2.3510 | 0.1533 | | |
| | 20.7741 | 0.0545 | | |
| | | | | 30.2672 |
| k=2 | 0.2658 | 0.8684 | -226.8476 | |
| | 21.5824 | 0.1316 | | |
| | | | | 281.7522 |
| k=1 | 1.0025 | 1.0000 | -367.7237 | |

[a]Estimates of the number of components, parameter, and weight.
[b]Estimates of parameter and weight when we fixed the number of components.
[c]For testing hypothesis $H_0$: number of components=k; for $H_a$: number of components=k+1.
[d]Mean of SIR for each component.
[e]Critical value for rejecting the null hypothesis was 4.01; the smallest components were used.

Table 2
Results of fitting mixture model to disease mapping of malaria (1995).

| Parameter | Values | | | | |
|---|---|---|---|---|---|
| Means of SIR | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
| | 0.0837 | 0.7705 | 3.6168 | 21.0612 | 39.5474 |
| Weights | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ |
| | 0.6120 | 0.2361 | 0.0992 | 0.0395 | 0.0132 |

Table 3
NPMLE[a], MLE[b] for lower component models, corresponding log-likelihoods, and LRS[c] (1995 through 1997).

| Component | Parameter[d] | Weight | Log-likelihood | $-2(\log_k - \log_{k+1})$ |
|---|---|---|---|---|
| k=8 | 0.0898 | 0.5872 | -610.8472 | |
| | 0.7493 | 0.2539 | | |
| | 3.1631 | 0.0793 | | |
| | 8.8603 | 0.0272 | | |
| | 15.1165 | 0.0098 | | |
| | 16.9776 | 0.0163 | | |
| | 27.8392 | 0.0217 | | |
| | 38.8803 | 0.0043 | | |
| | | | | 0.0812 |
| k=7 | 0.0898 | 0.5875 | -610.8878 | |
| | 0.7493 | 0.2540 | | |
| | 3.1633 | 0.0793 | | |
| | 8.8712 | 0.0274 | | |
| | 16.3810 | 0.0257 | | |
| | 27.8237 | 0.0218 | | |
| | 38.8807 | 0.0043 | | |
| | | | | 2.0766[e] |
| k=6 | 0.0898 | 0.5875 | -611.9261 | |
| | 0.7493 | 0.2540 | | |
| | 3.1633 | 0.0793 | | |
| | 8.8711 | 0.0274 | | |
| | 16.4087 | 0.0260 | | |
| | 29.1590 | 0.0259 | | |
| | | | | 5.4226 |
| k=5 | 0.0898 | 0.5876 | -614.6374 | |
| | 0.7503 | 0.2540 | | |
| | 3.1778 | 0.0795 | | |
| | 11.3226 | 0.0467 | | |
| | 28.8048 | 0.0321 | | |
| | | | | 27.2600 |
| k=4 | 0.0950 | 0.6082 | -628.2674 | |
| | 0.8930 | 0.2571 | | |
| | 5.0975 | 0.0790 | | |
| | 20.3846 | 0.0557 | | |
| | | | | 96.8182 |
| k=3 | 0.1791 | 0.7827 | -676.6765 | |
| | 2.1912 | 0.1383 | | |
| | 13.3655 | 0.0789 | | |
| | | | | 22.0952 |
| k=2 | 0.2667 | 0.8576 | -687.7241 | |
| | 6.3014 | 0.1424 | | |
| | | | | 687.9818 |
| k=1 | 0.9984 | 1.0000 | -1,031.7150 | |

[a]Estimates of the number of components, parameter, and weight.
[b]Estimates of parameter and weight when we fixed the number of components.
[c]For testing hypothesis $H_0$: number of components=k; for $H_a$: number of components=k+1.
[d]Mean of SIR for each component.
[e]Critical value for rejecting the null hypothesis was 4.01; the smallest components were used.

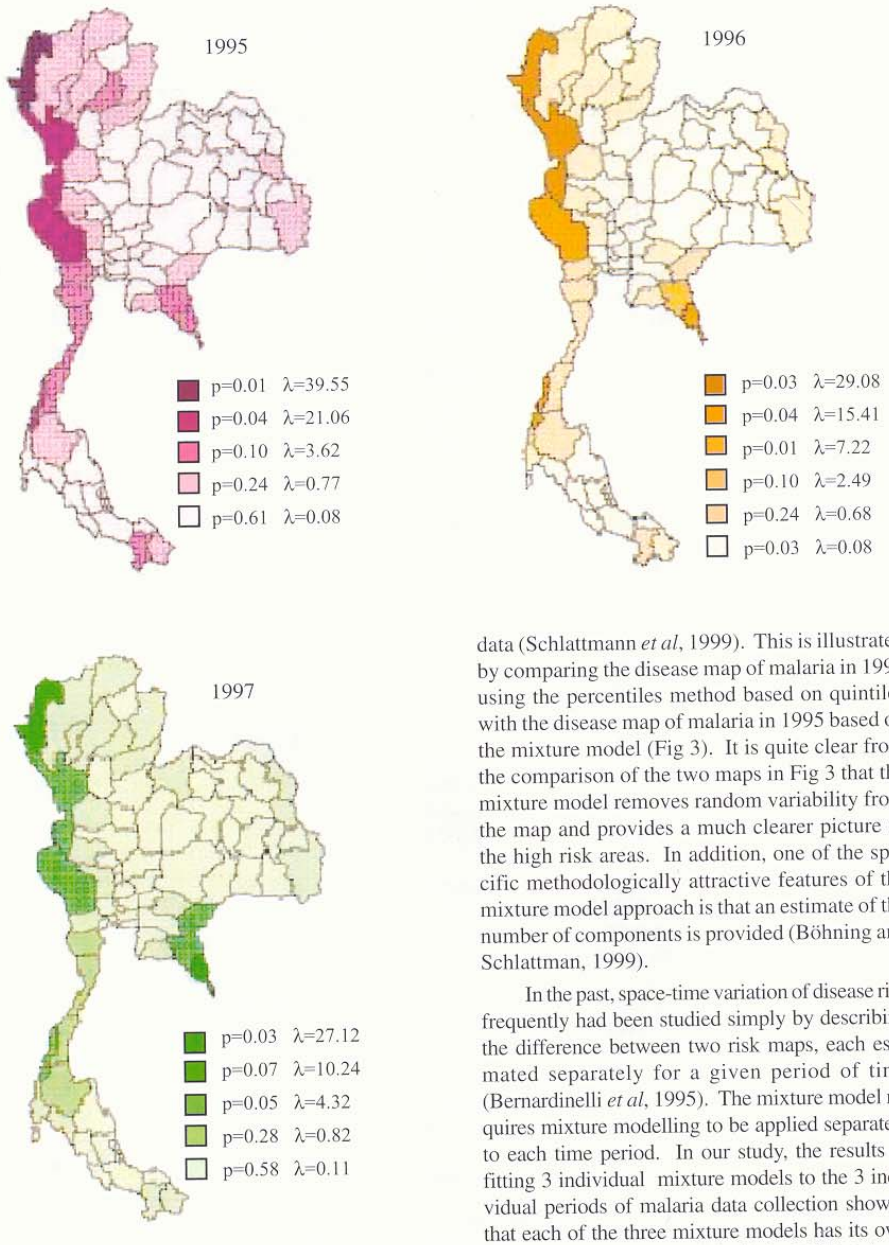


Fig 1–Disease maps of occurrence of malaria using a separate mixture model for each time period.

data (Schlattmann *et al*, 1999). This is illustrated by comparing the disease map of malaria in 1995 using the percentiles method based on quintiles with the disease map of malaria in 1995 based on the mixture model (Fig 3). It is quite clear from the comparison of the two maps in Fig 3 that the mixture model removes random variability from the map and provides a much clearer picture of the high risk areas. In addition, one of the specific methodologically attractive features of the mixture model approach is that an estimate of the number of components is provided (Böhning and Schlattman, 1999).

In the past, space-time variation of disease risk frequently had been studied simply by describing the difference between two risk maps, each estimated separately for a given period of time (Bernardinelli *et al*, 1995). The mixture model requires mixture modelling to be applied separately to each time period. In our study, the results of fitting 3 individual mixture models to the 3 individual periods of malaria data collection showed that each of the three mixture models has its own number of components-5 (k=5) for 1995, 6 (k=6) for 1996, and 5 (k=5) for 1997 (Fig 1). Although

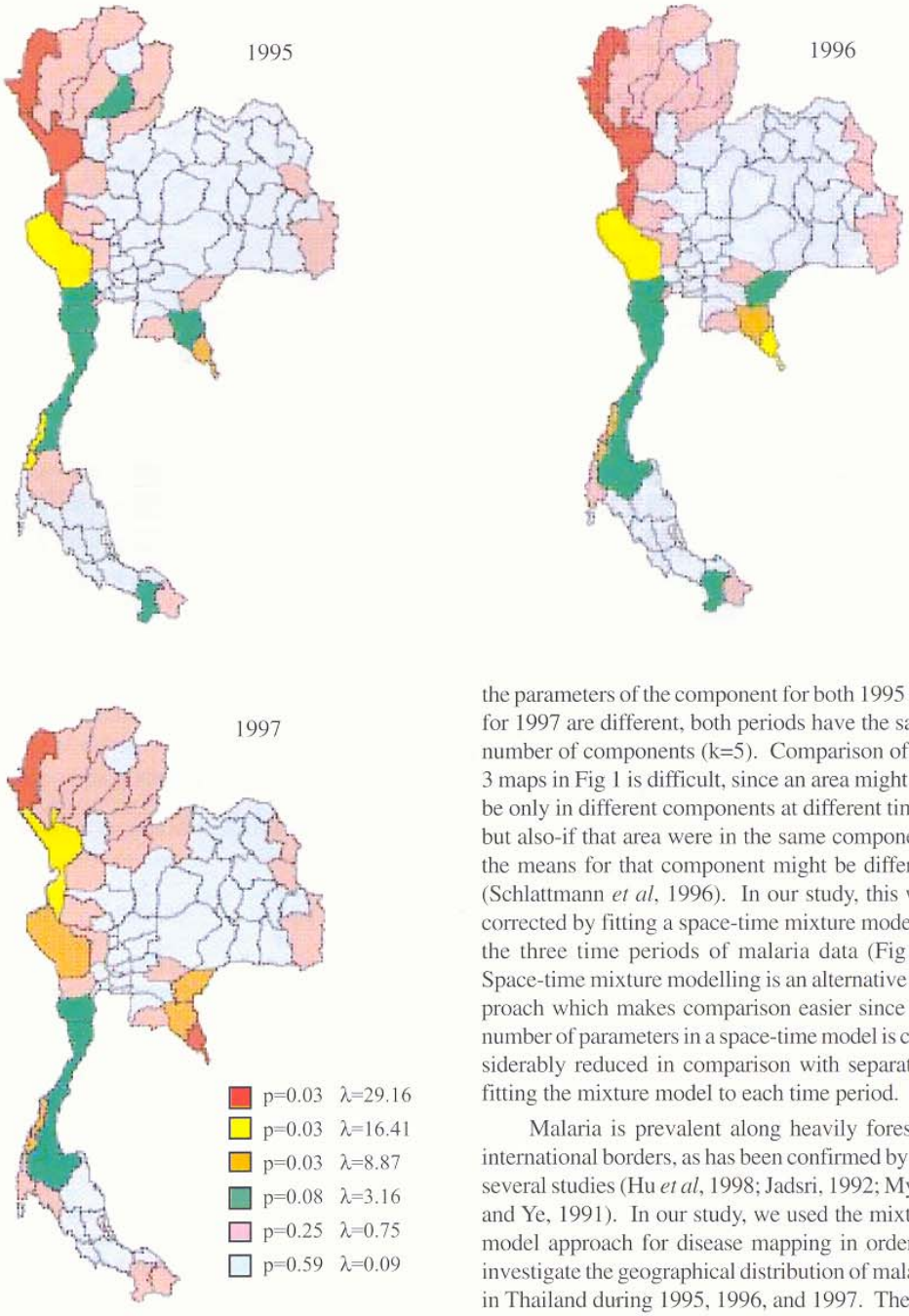


Fig 2–Disease maps of occurrence of malaria using space-time mixture models.

the parameters of the component for both 1995 and for 1997 are different, both periods have the same number of components (k=5). Comparison of the 3 maps in Fig 1 is difficult, since an area might not be only in different components at different times, but also-if that area were in the same component-the means for that component might be different (Schlattmann *et al*, 1996). In our study, this was corrected by fitting a space-time mixture model to the three time periods of malaria data (Fig 2). Space-time mixture modelling is an alternative approach which makes comparison easier since the number of parameters in a space-time model is considerably reduced in comparison with separately fitting the mixture model to each time period.

Malaria is prevalent along heavily forested international borders, as has been confirmed by the several studies (Hu *et al*, 1998; Jadsri, 1992; Myint and Ye, 1991). In our study, we used the mixture model approach for disease mapping in order to investigate the geographical distribution of malaria in Thailand during 1995, 1996, and 1997. The results are illustrated by maps (Fig 1) showing that malaria is most prevalent in 4 provinces on the Thai-Myanmar border and in 2 provinces on the Thai-

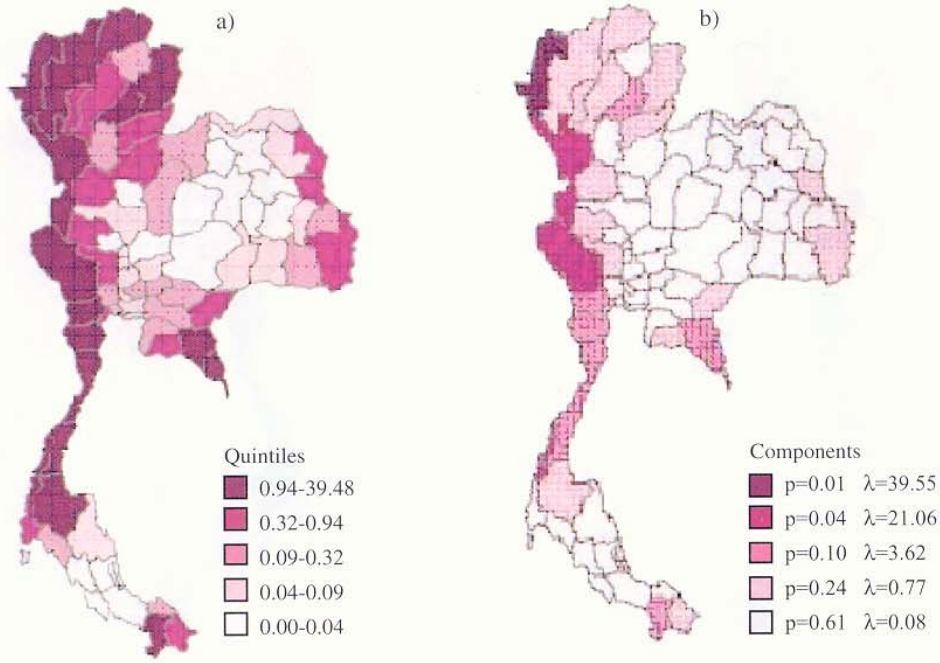


Fig 3–Disease maps of occurrence of malaria using a) Percentiles method, b) mixture model.

Table 4
Results of fitting space-time mixture model to disease mapping of malaria (1995 through 1997).

| Parameter | Value | | | | | |
|---|---|---|---|---|---|---|
| Means of SIR | λ1 | λ2 | λ3 | λ4 | λ5 | λ6 |
| | 0.0898 | 0.7493 | 3.1633 | 8.8711 | 16.4087 | 29.1590 |
| Weight | p1 | p2 | p3 | p4 | p5 | p6 |
| | 0.5875 | 0.2540 | 0.0793 | 0.0274 | 0.0260 | 0.0259 |

Cambodian border; in each of the 6 provinces the forest covers more than 50% of the landscape.

The advantage of the mixture model approach to disease mapping is the graphical visual presentation of the prevalence of disease. The space-time mixture model more adequately investigates the dynamic nature of disease than does the mixture model.

## REFERENCES

Bernardinelli L, Montomoli C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Stat Med* 1992; 11: 983-1007.

Böhning D. General epidemiology and its methodological foundations. Munich (Germany): Oldenburg, 1998 (in German).

Böhning D. Computer-assisted analysis of mixtures and applications. Meta-analyses, disease mapping and others. Boca Raton (Florida): Chapman & Hall, 1999.

Böhning D, Schlattmann P. Disease mapping with hidden structure using mixture models. In: Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J, Bertollini R, eds. Disease mapping and risk assess-

ment for public health. New York: John Wiley & Sons, 1999: 49-60.

Böhning D, Dietz E, Schlattmann P. Space-time mixture modelling of public health data. *Stat Med* 2000; 19: 2333-44.

Böhning D, Schlattmann P, Lindsay B. Computer-assisted analysis of mixtures (C.A.MAM): statistical algorithms. *Biometrics* 1992; 48: 283-303.

Cislaghi C, Biggeri A, Braga M, Lagazio C, Marchi M. Exploratory tools for disease mapping in geographical epidemiology. *Stat Med* 1995; 14: 2363-81.

Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; 43: 671-81.

Devine OJ, Louis TA. A constrained empirical Bayes estimator for incidence rates in areas with small populations. *Stat Med* 1994; 13: 1119-33.

Heisterkamp SH, Doornbos G, Gankema M. Disease mapping using empirical Bayes and Bayes methods on mortality statistics in The Netherlands. *Stat Med* 1993; 12: 1895-913.

Hu H, Singhasivanon P, Salazar NP, *et al.* Factors influencing malaria endemicity in Yunnan Province, PR China (analysis of spatial pattern by GIS). Geographical Information System. *Southeast Asian J Trop Med Public Health* 1998; 29: 191-200.

Jadsri S. A study of the risk factors in forest aquired malaria at Pong Nam Ron, Chanthaburi. Bangkok: Faculty of Tropical Medicine, Mahidol University; 1992. MS thesis.

Lindsay BG. Mixture models: theory, geometry, and applications. Hayward (California): Institute of Statistical Mathematics, 1995.

Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. *J R Stat Soc Ser C Appl Stat* 1991; 40: 283-94.

Mollie A, Richardson S. Empirical Bayes estimates of cancer mortality rates using spatial models. *Stat Med* 1991; 10: 95-112.

Myint L, Ye H. Study of the malaria situation in forested foothills and nearby plain areas of Myanmar. *Southeast Asian J Trop Med Public Health* 1991; 22: 509-14.

Schlattmann P. The computer package DismapWin. *Stat Med* 1996; 15: 931.

Schlattmann P, Böhning D. Mixture models and disease mapping. *Stat Med* 1993a; 12: 1943-50.

Schlattmann P, Böhning D. Computer packages C.A.MAN (computer assisted mixture analysis) and Dismap. *Stat Med* 1993b; 12: 1965.

Schlattmann P, Dietz E, Böhning D. Covariate adjusted mixture models and disease mapping with the program DismapWin. *Stat Med* 1996; 15: 919-29.

Schlattmann P, Böhning D, Clark A, Lawson A. Lung cancer mortality in woman in Germany 1995: A case study in disease mapping. In: Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J, Bertollini R, eds. Disease mapping and risk assessment for public health. New York: John Wiley & Son; 1999: 411-21.