Modeling and Forecasting Malaria Incidence in North-Western Thailand

Wattanavadee Sriwattanapongse

A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Research Methodology

Prince of Songkla University

2010

Thesis Title	Modeling and Forecasting Malaria Incidence in North-Western Thailand
Author	Mrs. Wattanavadee Sriwattanapongse
Major Program	Research Methodology

Major Advisor

.....

(Emeritus Prof. Dr. Don McNeil)

Co-advisor

.....

(Dr. Metta Kuning)

Examining Committee:

..... Chairperson

(Dr. Nittaya McNeil)

.....

(Emeritus Prof. Dr. Don McNeil)

.....

(Dr. Metta Kuning)

.....

(Dr. Suwich Thammapalo)

The Graduate School, Prince of Songkla University, has approved this thesis as fulfillment of the requirements for the Doctor of Philosophy in Research Methodology.

.....

(Assoc. Prof. Dr. Krerkchai Thongnoo)

Dean of Graduate School

ชื่อวิทยานิพนธ์	ตัวแบบและการพยากรณ์อุบัติการณ์ของโรคมาลาเรีย ในภาคตะวันตกเฉียงเหนือของประเทศไทย
ผู้เขียน	นางวัฒนาวดี ศรีวัฒนพงศ์
สาขาวิชา	วิธีวิทยาการวิจัย
ปีการศึกษา	2552

บทคัดย่อ

การศึกษานี้ประยุกต์วิธีการทางสถิติที่เหมาะสมสำหรับตัวแปรที่สนใจ และโครงสร้างของข้อมูลที่แตกต่างกัน (ข้อมูลรายบุคคลคล และข้อมูลที่เป็นจำนวนนับ) ซึ่งเป็นข้อมูลผู้ป่วยโรคมาลาเรียรายบุคคลที่ทางโรงพยาบาลได้จัดเก็บตั้งแต่ พ.ศ. 2542-2547 ในภาคตะวันตกเฉียงเหนือของประเทศไทย ทั้งพื้นที่ติดและไม่ติดชายแดน ด้วยวิธีการทางสถิติที่เหมาะสมสอดคล้องกับรูปแบบ ความสัมพันธ์ ระหว่างตัวแปรตามและตัวแปรอิสระ

วัตถุประสงค์ของการศึกษานี้ เพื่อหารูปแบบอุบัติการณ์การเกิดโรคมาลาเรียในภาคตะวันตกเฉียงเหนือของประเทศไทย และหาว่าพื้นที่ไหนมีอุบัติการณ์การเกิดโรคมาลาเรียสูงเพื่อดำเนินการรักษาและป้องกันในพื้นที่เป้าหมาย

การศึกษาครั้งนี้แบ่งออกเป็น 2 ส่วน คือ ส่วนที่หนึ่ง มีวัตถุประสงค์เพื่อหาตัวแบบที่เหมาะสมในการพยากรณ์อุบัติการณ์รายเดือนของการรายงานผู้ป่วยในอำเภอของสองจังหวัด คือ แม่ฮ่องสอน และตาก ซึ่งมีความเสี่ยงของการเป็นโรคสูงในภาคตะวันตกเฉียงเหนือของประเทศไทย การศึกษาในส่วนที่สอง เป็นการประยุกต์วิธีใหม่โดย มีวัตถุประสงค์เพื่อแยกแยะรูปแบบและแนวโน้มอุบัติการณ์การเกิดโรคมาลาเรียจากการรวมข้อมูลรายไตรมาสใน 65 อำเภอของ 6 จังหวัดภาคภาคตะวันตกเฉียงเหนือของประเทศไทย คือ ลำพูน แพร่ น่าน เชียงราย แม่ฮ่องสอน และ ตาก

การศึกษาในส่วนที่หนึ่ง ตัวแบบที่ใช้ คือ การถดถอยเชิงเส้น การถดถอยปัวส์ซอง และการถดถอยทวินามนิเสธ เพื่อพยากรณ์อุบัติการณ์โรคมาลาเรียตามอำเภอ กลุ่มอายุ (0-4, 5-14, 15-39, และ 40+) และช่วงเวลารายเดือนที่มีการระบาดเกิดขึ้นในเวลาอันใกล้เพื่อป้องกันการเกิดโรคโดยใช้มาตรการที่เหมาะสม การถดถอยเชิงเส้นได้ให้ค่าสัมประสิทธิ์การอธิบาย (r^2) ของจังหวัดแม่ฮ่องสอน และตาก เท่ากับร้อยละ 61.2 และ 81.0 ตามลำดับ การถดถอยปัวส์ซองได้ให้ค่า residual deviance ของจังหวัดแม่ฮ่องสอน และตาก เท่ากับ 8,135.9 และ 8,898.4 ตามลำดับ และการถดถอยทวินามนิเสธได้ให้ค่า residual deviance ของจังหวัดแม่ฮ่องสอน และตาก เท่ากับ 2,212.7 และ 2,834.2 ตามลำดับ ส่วนประกอบทั้งหมดในตัวแบบมีนัยสำคัญทางสถิติที่ p-value <

0.05 ระหว่างตัวแบบที่ใช้ เลือกตัวแบบที่ดีที่สุดโดยใช้ analysis of deviance ผลการศึกษาปรากฏอย่างชัดเจนว่า ตัวแบบทวินามนิเสธเป็นตัวแบบที่เหมาะสมที่สุด ตัวแบบที่ใช้ประกอบด้วยปัจจัยทางบวกสอดคล้องกับฤดูกาลของปี อำเภอ กลุ่มอายุ และอุบัติการณ์ในเดือนก่อนหน้า (lagged) และสามารถใช้ตัวแบบนี้ในการพยากรณ์ระยะสั้น ๆ ได้ ตัวแบบที่ได้ใช้พยากรณ์การเกิดขึ้นอย่างรุนแรงของโรคในทันที และเป็นประโยชน์ในการจัดสรรแหล่งทรัพยากรสำหรับการป้องกันโรค

การศึกษาในส่วนที่สอง ตัวแบบการถดถอยที่ใช้คือ ตัวแบบเชิงบวกปกติ-ล็อก และตัวแบบเชิงบวกผสมการคูณบนฐานการวิเคราะห์ตัวประกอบที่ใช้อธิบายรูปแบบเหล่านี้ ตัวแบบเชิงบวกปรกติ-ล็อกได้ค่าสัมประสิทธิ์การอธิบาย (r^2) เท่ากับร้อยละ 79.10 ตัวแบบเชิงบวกผสมการคูณได้ค่าสัมประสิทธิ์การอธิบาย (r^2) เท่ากับร้อยละ 81.7 ผลการศึกษาพบว่าตัวแบบเชิงบวกผสมการคูณเป็นตัวแบบที่เหมาะสมพอดีกับอุบัติการณ์การเกิดโรคมาลาเรียในพื้นที่ภาคตะวันตกเฉียงเหนือตามอำเภอ กลุ่มอายุและช่วงเวลาไตรมาส ตัวแบบแสดงแนวโน้มและความผันแปรเชิงพื้นที่ในอุบัติการณ์ของโรค จากการศึกษาในครั้งนี้อุบัติการณ์โรคมาลาเรียสูงในบุคคลที่มีอายุตั้งแต่ 5 ถึง 39 ปี ในอำเภอที่ติดเขตชายแดน ในทางตรงข้ามอุบัติการณ์โรคมาลาเรียสูงในเด็กอายุระหว่าง 0-4 ปี ในอำเภอที่ไม่ติดเขตชายแดน อำเภอส่วนใหญ่ใน 6 จังหวัดมีแนวโน้มอุบัติการณ์การเกิดโรคมาลาเรียลดลง ตามรูปแบบฤดูกาลโดยสูงสุดในไตรมาสที่ 2 คือ เดือนเมษายนถึงมิถุนายน แผนที่ได้แสดงความแตกต่างอุบัติการณ์มาลาเรียระหว่างอำเภอที่ติดชายแดนพม่าและอำเภออื่น ๆ ในขณะที่อำเภอส่วนใหญ่มีอุบัติการณ์โรคมาลาเรียลดลงตลอดช่วงเวลาหกปี มีสองอำเภอที่มีแนวโน้มการเกิดโรคมาลาเรียสูงมาก คือ อำเภอทุ่งหัวช้าง จังหวัดลำพูน กิ่งอำเภอวังเจ้า จังหวัดตาก มีสองอำเภอที่แนวโน้มการเกิดโรคมาลาเรียไม่มีเลย คือ อำเภอเมืองลำพูน และอำเภอบ้านโฮ่ง จังหวัดลำพูน มีหนึ่งอำเภอแนวโน้มการเกิดโรคมาลาเรียเพิ่มขึ้น คือ อำเภอลี้ จังหวัดลำพูน ผลการศึกษาแสดงให้เห็นว่า พื้นที่อำเภอส่วนใหญ่อุบัติการณ์การเกิดโรคมาลาเรียลดลงในช่วงเวลาที่ศึกษา แต่ยังคงสูงมากในอำเภอที่ติดเขตชายแดนประเทศพม่า

Thesis	Modeling and Forecasting Malaria Incidence in North-Western Thailand
Author	Mrs. Wattanavadee Sriwattanapongse
Major	Research Methodology
Academic Year	2009

ABSTRACT

In study suitable statistical methods have been applied for different data structures based on hospital case records of malaria routinely reported from 1999 to 2004 in the north-western area of Thailand including both districts bordering on Myanmar and non-border districts. Using the appropriate statistical methods, the associated patterns of outcomes and determinants can then be identified.

The objective of this study was to model the pattern of hospital-diagnosed malaria incidence in north-western Thailand, and to determine precisely where malaria incidence is high so that these areas can be targeted for treatment and prevention.

There are two parts in this study. The first study uses a conventional statistical model to find a suitable model for forecasting monthly incidence rates of reported hospital cases of malaria in districts of the two provinces, Mae Hong Son, and Tak, with high risk of disease in the north-western region of Thailand. The second study uses new methods 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 six north-western provinces of Thailand. The provinces in our study comprise Lamphun, Phrae, Nan, Chiang Rai, Mae Hong Son, and Tak.

In the first study, the model compared 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. Linear regression obtaining an r-squared of 61.2% for Mae Hong Son and 81.0% for Tak province. 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. All components in the model for each province are statistically significant at p-value < 0.05. 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.

In the second study, the model compared log-normal regression models (additive and additive plus multiplicative based on principal components) to describe these patterns. Additive log-normal gave an r-squared of 79.1%.The additive plus multiplicative model gave an r-squared of 81.7% .The additive plus multiplicative linear model provides an appropriate fit to the malaria incidence rates in the north-western region classified by districts, age-group and quarterly period. The models show trends and spatial variations in disease incidence. According to this study, the incidence of malaria was highest among persons aged from 5 to 39 years in border districts, but in contrast, children aged 0-4 had highest incidence rates among those in other 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. A thematic map 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 (Mueang Lamphun and Ban Hong in Lamphun) show no trend, and one district (Li in Lamphun) shows an increasing trend. 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.

Acknowledgements

This thesis would not have been possible without significant contributions and assistances from a number of people. First and foremost, I would like to thank my supervisor Emeritus Prof. Dr. Don McNeil, who came from Macquarie University, Sydney Australia, for his invaluable guidance, support and assistance throughout the completion of this thesis. Thanks also to my co-supervisors, Dr. Metta Kuning, for her encouragement and suggestions, and to Dr. Phattrawan Tongkumchum, Asst. Prof. Dr. Chamnien Choonpradub, Assoc. Prof. Supreeya Wongtra-ngan, MD, Sawitri Werasophon, MD, and Assoc. Prof. Dr. Jirawan Jitthavech for their helpful assistance, encouragement, and guidance.

I also wish to thank the reviewers from the *Songklanakarin Journal of Science and Technology* and the *Chiang Mai Journal of Science* for their review of papers incorporated into this thesis.

I am grateful to the Thailand Bureau of Epidemiology and the National Health Security Office, Ministry of Public Health, for providing the data. I also would like to acknowledge the Graduate School, Prince of Songkla University, for funding.

My sincere thanks also to Chiang Mai University for allow me to study PhD in Research Methodology at Prince of Songkla University, Pattani Campus.

Finally, I would like to thank my family and friends for their encouragement throughout my study, especially my Mother, Mrs. Somsri Jitpakdee.

Wattanavadee Sriwattanapongse

Contents

	Page
บทคัดย่อ	iii
Abstract	v
Acknowledgements	viii
Contents	ix
List of Tables	xii
List of Figures	xiii
List of Appendices	xv
Chapter	
1. Introduction	1
1.1 Rationale for study	1
1.2 Definition of incidence	4
1.3 Literature review	6
1.4 Outline of thesis	16
2. Methodology	17
2.1 Sources of data	17
2.2 Data management	19
2.3 Demographic factors	21
2.4 Statistical methods	22
2.4.1 Linear regression models	22
2.4.2 Poisson regression models	23

Contents (Cont.)

	Page
2.4.3 Negative binomial regression models	24
2.4.4 Additive log-normal models	26
2.4.5 Additive plus multiplicative models	27
2.4.6 Principal component analysis	29
3. Data Analysis and Models for Malaria	33
3.1 Malaria in North-Western Thailand	33
3.1.1 Characteristics of the data	33
3.1.2 Distributions of incidence rates	34
3.1.3 Fitting a model to the data	37
3.2 Modeling Malaria Incidence in North-Western Thailand	43
3.2.1 Characteristics of the data	43
3.2.2 Distributions of incidence rates	44
3.2.3 Age-specific malaria incidence rates	45
3.2.4 Fitting a model to the data	48
3.3 Summary for Malaria in North-Western Thailand	52
4. Conclusions and Limitations	56
4.1 Research methodology	56
4.2 Conclusions	58
4.2.1 Modeling malaria incidence	58
4.2.2 Malaria incidence rates	58

Contents (Cont.)

	Page
4.3 Discussion	60
4.4 Strengths of this study	63
4.5 Limitations and suggestions for further study	64
References	66
Vitae	73

List of Tables

Table		Page
1	Results from fitting linear model to log-transformed incidence rates	38
2	Results from fitting Poisson models to malaria disease counts	40
3	Results from fitting negative binomial models to malaria disease counts	41
4	Malaria incidence rates per 1,000 by province: 1999-2004	45
5	Malaria incidence rates per 1,000 by province and age group: 1999-2004	46

List of Figures

Figure		Page
1	Path diagram for studies	20
2	Time series of monthly disease rates for each age group in the two provinces	35
3	Average annual incidence rates for malaria in Mae Hong Son and Tak provinces, 1999-2004	36
4	Plots of Pearson residuals versus asymptotic scores after fitting Poisson and negative binomial models for malaria disease counts in the two provinces	42
5	Average annual incidence rates for malaria in North-Western of Thailand	44
6	Malaria incidence in provinces of North-Western Thailand by age group: 1999-2004	47
7	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 Thailand	49
8	Age group, trend and district components of malaria incidence in North-Western Thailand 1999-2004: additive model	50

List of Figures (Cont.)

Figure		Page
9	District components of malaria incidence in North-Western Thailand 1999-2004 based on additive plus multiplicative model. The district trend amplitude is defined in the model as the extent to which a district follows the overall trend	51
10	Malaria incidence in North-Western border provinces of Thailand: 1999-2004	52

Chapter 1

Introduction

1.1 Rationale for study

Malaria is a serious health problem in many tropical countries of the world. Nearly one billion people are exposed to hyperendemic and mesoendemic malaria in Southeast Asia (Hay et al 2004). Malaria is a disease which is deeply rooted in poor communities and hence affects national development, particularly in Thailand.

Geographically, Thailand is located in the Southeast Asia tropical zone. The transmission of malaria is common, particularly in the upper northern region of the country. Malaria in Thailand is endemic in forest areas, and many cases occur along the national borders, particularly on the border with Myanmar to the west where there is evidence of much higher malaria incidence rates in some districts (Konchom et al 2003).

The border areas of Thailand are mostly forest-fringe foothills, especially along the north-western border with Myanmar. There is high malaria receptivity, and vulnerability that presents numerous problems in the control of malaria transmission. Malaria among foreign labourers is a major obstacle to an effective control program, although the number of reported cases has declined significantly since 2000. The disease is localized in four main areas: the Thai-Myanmar border to the north-west, the Thai-Lao to the north-east, the Thai-Cambodian border to the east, and the central part of the southern peninsula.

Since 1997, more than 60% of the total malaria cases in Thailand were detected in border provinces, and one-third of these were detected in Tak province on the Thai-Myanmar border. The total number of cases nationwide has been decreasing every year, while the number of cases detected in border provinces has been gradually increasing. This indicates that the disease has become more localized along the international borders (Konchom et al 2003).

By 1930, malaria was recognized as one of the most severe diseases in Thailand, with mortality rates exceeding 400 cases per 100,000 populations. In 1947, the mortality rate attributed to malaria was nearly 300 cases per 100,000 population, and represented a major cause of premature death in rural population. In the 1950s, intensified malaria control efforts using DDT significantly reduced malaria-related deaths over the ensuing years. By 1974, mortality was 16 per 100,000 population; and in 1994, deaths had decreased further to 1.6/100,000. In 1997, the mortality rate was 1.26/100,000, with 700 deaths. Mortality rates have been decreasing since 1987 in spite of a fluctuation of malaria incidence over the past 15 years. The continued decrease in mortality since 1974 has been attributed to the expansion of peripheral health care delivery in more remote areas of the country, and the establishment of malaria clinics providing rapid diagnosis and prompt treatment in areas of high transmission (Prasittisuk 1985). Despite these dramatic improvements, high rates of disease and death continue among people who reside in the forested hill areas, especially in economically depressed localities along the borders, which are most often associated with limited access to public health services (Chareonviriyaphap et al 2000).

Many research models have been used to study malaria in the border provinces of Thailand. Rattanasiri et al (2004) investigated the geographical distribution of malaria in Thailand during 1995, 1996, and 1997 by applying a mixture model to disease mapping, and examined the dynamic nature of malaria in Thailand during the three-year time frame by applying a space-time mixture model. Non-parametric maximum likelihood estimation was employed to calculate 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 based on the rate of malaria for each province. Malaria occurred extensively in four provinces on the Thai-Myanmar border, and in two provinces on the Thai-Cambodia border. Of the 76 provinces studied, 10 showed an increasing trend over the three-year period. A comparison of the map based on the mixture model with the map based on the traditional percentile 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 also more adequately investigates the dynamic nature of disease than does the percentile model. For these reasons, the investigation of modeling and forecasting malaria incidence in north-western Thailand were studied for this thesis.

The objective of this thesis is to model the pattern of hospital-diagnosed malaria incidence in North-Western Thailand, and to determine precisely where malaria incidence is high so that these areas can be targeted for treatment and prevention.

The results of this study will provide a useful short-term forecast, utilizing a model that produces such forecast of malaria incidence rates, as a basis for allocation of resources for disease prevention.

1.2 Definition of incidence

The incidence of disease is the rate at which the disease is spread among persons in the population. The prevalence of disease in a population is the number of people who are infected divided by the number of people in the population. The numerator is those who are ill, those who have specific symptoms of illness, or those who are have microbiology evidence of infection but do not exhibit symptoms. The denominator in the prevalence equation is also defined by the epidemiology researcher. It may be the number of persons in the population, regardless of known exposure status, or it may be persons who were exposed. In the former case, the measurement of prevalence defines the burden of disease in the population overall; in the latter case, the definition gives the prevalence of disease among those exposed. Where exposure is common, age-specific population prevalence is commonly measured. Where exposure is rare, prevalence rates by exposure group are used more frequently.
The other commonly used measure is the incidence of disease. The incidence is the rate at which persons acquire the disease or the rate at which the infections agent is being transmitted throughout the population. The incidence of disease always includes a unit of time, the number of case of malaria in a given year, month or quarter, for example.

The incidence and prevalence of disease are related to each other by the duration of disease. In case where the duration of disease is short, the prevalence of disease will

be approximately equaled to the incidence of disease because most infections will be relative recent. If, in contrast, the duration of the disease is long, the prevalence of disease will include both new and former case of disease and will be a more appropriate measure than the incidence of disease. This relationship can be described by the equation:

$$\text{Prevalence} = \text{Incidence} \times \text{Duration}$$

At any time, the incidence may be decreasing when the prevalence is rising. Such may be the case with HIV infection in the United States and Western Europe at present, because combined antiretroviral therapy has prolonged survival, but because of the effect of the drugs in reducing viral load, the transmission, or incidence of new cases, may be decreasing.

In other infectious disease that have short duration, and where infected persons remain susceptible to re-infection, the incidence may exceed the point prevalence. Persons may have several episodes of diarrhea or rhinovirus respiratory infection per year that last only a few days. In these diseases, the point prevalence may be low but the annual incidence may be quite high. It may be preferable to measure the impact of these diseases with annual incidence rates. In contract, in malaria hyperdemic areas, young children may receive hundreds of bites from infected mosquitoes every year. In this situation, the annual incidence of malaria is so high that is difficult to measure. However, a blood film will allow determination of the point prevalence of infection, because the parasites persist in the blood for some time. Malaria prevalence data are more useful to differentiate populations at very high risk or of hyperendemic foci in an endemic area. (Nelson and Williams 2007)

1.3 Literature review

Several publications have addressed the problem of modeling and forecasting malaria disease incidence.

Statistical modeling

McNeil and Tukey (1973) noted that two-way tables of all kinds often require diagnosis, usually of the residuals after some simple fit such as row-plus-column or row-times-column. The method is likely to be either one-degree-of-freedom for non-additivity, or a diagnostic plot (for which the former is the linear regression term). Not all diagnoses can be made at the first step. Some diagnostic plots appear as a diagonal cross; or, when such an appearance is unclear, become converted to oppositely tilted pictures when we look only at points with high fitted values and, separately, at those with low fitted values. Such behaviour signals a need for a more subtle re-expression than powers and logs, for example a re-expression such as $p^{\lambda} - (1-p)^{\lambda}$. The results showed that the appearance and treatment of such diagnoses lead to: the use of letter-value displays and the associated plots to study the character of non-normality; the instances of the effect of the method of fitting on the shape of the distribution of residuals; and convenient algorithms for the iterative, recursive fitting of a variety of additive, multiplicative, and mixed additive-multiplicative models to any kind of two-way table.

Booth et al (2002) investigated the Lee-Carter method of mortality forecasting, which assumes an invariant age component. Most applications have also adopted a linear time component. The use of the method with their Australian data is compromised by significant departures from linearity in the time component, and changes over time in

the age component. The researchers modified the method by adjusting the time component to reproduce the age distribution of deaths, rather than total deaths, and to determine the optimal fitting period in order to address non-linearity in the time component. In the Australian case the modification has the added advantage that the assumption of invariance is better met. The modifications result in higher forecast life expectancy than the original Lee-Carter method and official projections, and a 50% reduction in forecast error. The model is also expanded to take account of age-time interactions by incorporating additional terms, but these are not readily incorporated into forecasts.

Bi et al (2003) explored the impact of climate on the transmission of malaria in China. A time series analysis was conducted using data on monthly climatic variables and monthly incidence of malaria in Shuchen County for the period 1990-1991. The results, using Spearman's correlation analysis, showed that monthly mean maximum and minimum temperature, two measures of monthly mean relative humidity, and monthly amount of precipitation were positively correlated with the monthly incidence of malaria in the county. Regression analysis suggested that monthly mean minimum temperature and total monthly rainfall, with a one month lagged effect, were significant climate variables in the transmission of malaria in Shuchen County. Seasonality was also significant in the regression model, and there was a declining secular trend in the incidence of malaria. The results indicate that climatic variables should be considered as possible predictors for regions with similar geographic and socioeconomic conditions.

Hoshen and Morse (2004) described climate as a major driving force behind malaria transmission. Hence, climate data are often used to account for the spatial, seasonal

and inter-annual variation in malaria transmission. The researchers used a mathematical-biological model of the parasite dynamics in Africa, comprising the weather-dependent stages, both within vectors and within host stages. The results of numerical evaluations of the model in both time and space show that it qualitatively reconstructs the prevalence of infection. A process-based modeling structure has been developed that may be suitable for the simulation of malaria forecasts based on seasonal weather forecasts.

Jansakul and Hinde (2004) discussed the possible use of the Newton-Raphson algorithm to obtain maximum likelihood estimates of the linear mean-variance negative binomial regression model and of the overdispersion parameter. Constructing a half-normal plot with a simulated envelope for checking the adequacy of a selected linear mean-variance negative binomial model was also discussed. These procedures were applied to analyze data from a number of embryos in an orange tissue culture experiment in Ireland. The analysis showed that the NB1 regression model with a cubic response function over the dose levels was consistent with the data, including some areas where a few malaria cases were found as well as in malaria-free areas.

Kaewsompak et al (2005) studied the epidemic patterns of dengue hemorrhagic fever and other acute febrile illnesses in Yala province in Southern Thailand. They investigated the relationship between incidence rates in terms of their geographical distribution, and developed a methodology that could be applied routinely to geographical epidemiologic research for the spatio-temporal mapping of disease. Schematic range maps and statistical models were used to investigate disease distribution by year and location. The analytic methods used Poisson and negative binomial distribution models. The concept of a "risk alert" is suggested as a method

for highlighting subdistricts with unexpectedly high incidence rates in a given year. The authors found high correlations between the incidence rates for dengue fever: in 2002 there were two sub-districts with moderate risk alerts, namely KaYuBoKo (p-value = 0.030) and KuTaBaRu (p-value = 0.042); whereas in 2003 there was just one subdistrict with a high risk alert, KaYuBoKo (p-value = 0.003).

Gomez-Elipe et al (2007) developed a statistical 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 a time series of quarterly notifications of malaria cases from local health facilities, in addition to rain and temperature records, and the normalized difference vegetation index. An autoregressive integrated moving average methodology was employed to obtain a model showing the relationship between quarterly notification of malaria cases and environmental variables. The results showed that this model is a simple and useful tool for forecasting malaria incidence in Burundi.

Briët et al (2008) investigated developing a forecasting system which could assist in the efficient allocation of resources for malaria control in Sri Lanka. Exponentially weighted moving average models, autoregressive integrated moving average (ARIMA) models with seasonal components, and seasonal multiplicative autoregressive integrated moving average (SARIMA) models were compared on a monthly time series of district malaria cases for their ability to predict the number of malaria cases one to four months ahead. The addition of covariates, such as the number of malaria cases in neighboring districts, or rainfall amounts, were assessed for their ability to improve the predictive ability of selected (seasonal) ARIMA

models. The results showed that the addition of rainfall as a covariate improved prediction of selected (seasonal) ARIMA models moderately in some districts, but worsened prediction in others. Improvement by adding rainfall was more noticeable with longer forecasting horizons.

Correlation analysis

Gagnon et al (2002) 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 highest in the wet season. However, the authors also suggested that on an inter-annual scale, malaria was also associated with drought. They analyzed the relationship between El Niño Southern Oscillation events and malaria epidemics in a number of South American countries, including Colombia, Ecuador, French Guiana, Guyana, Peru, Suriname and Venezuela. A statistically significant relationship was found between El Niño and malaria epidemics in Colombia, Guyana, Peru and Venezuela. It was found that flooding engenders malaria epidemics in the dry coastal region of northern Peru, while droughts favor the development of epidemics in Colombia and Guyana, and epidemics lag a drought by one year in Venezuela. In Brazil, French Guiana and Ecuador, where malaria signal non-climatic factors such as insecticide spraying, variations in availability of anti-malaria drugs were not found. Population migration is likely to play a stronger role in malaria epidemics than ENSO-generated climatic anomalies. In some South American countries, El Niño forecasts show strong potential for improving public health efforts to control malaria.

Khasnis and Nettleman (2005) concluded that global warming has serious implications for all aspects of human life, including infectious diseases, and that the effect of global warming on disease depends on the complex interaction between the human host population and the causative infectious agent. From the human standpoint, changes in the environment may trigger human migration, causing disease patterns to shift. Crop failures and famine may reduce host resistance to infections. Disease transmission may be enhanced through the scarcity and contamination of potable water sources. Importantly, significant economic and political stresses may damage the existing public health infrastructure, leaving mankind poorly prepared for unexpected epidemics. Global warming will certainly affect the abundance and distribution of disease vectors. Altitudes that are currently too cool to sustain vectors will become more conducive to them. Some vector populations may expand into new geographic areas, whereas others may disappear. Malaria, dengue, plague, and viruses causing encephalitic syndromes are among the many vector-borne diseases likely to be affected. Some models suggest that vector-borne diseases will become more common as the earth warms, although caution is needed in interpreting these predictions. Clearly, global warming 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. A graphical display shows the improvement of risk prediction brought about by the model used in their study. Such a model, even if based purely on statistical data analysis, can provide a useful basis for allocation of resources for disease prevention.

Devi and Jauhari (2006) investigated the effect of climatic factors on malaria incidence, with particular emphasis on capturing the essential events as a result of climatic variability in India. Mosquito sampling and identification was performed using WHO entomological methods, and follow-up of recognized keys and catalogues. Data on malaria incidence and meteorological information were gathered in a collaborative study with the District Malaria Office, and the Forest Research Institute in Dehradun, respectively. Pearson's correlation analysis was applied to establish the relationship between climate variables and malaria transmission. In this study, a higher positive correlation of association was found between monthly parasite incidence and climatic variables (temperature, rainfall and humidity). However, the highest significant correlation was found between rainfall and malaria incidence ($r = 0.718$, p -value< 0.0001) when the data were staggered to allow a lag of one month. Climatic variables that predict the presence or absence of malaria are likely to be the best pattern for forecasting the distribution of the disease.

Geographical Information System (GIS)

Kleinschmidt et al (2000) concluded that a good map of malaria risk is an important tool for malaria control in South Africa. The production of such maps relies on modeling to predict the risk for most of the map areas, with actual observation of malaria prevalence usually only known at a limited number of specific locations. Estimation is complicated by the fact that there are often variations of risk that cannot be accounted for by the known covariates, and because data points measuring malaria prevalence are not evenly or randomly spread across the area to be mapped. The researchers used a simple two-stage procedure to produce a map of predicted risk

areas, using logistic regression modeling to determine the appropriate risks on a larger scale, while employing geo-statistics to improve prediction at a local level. Malaria prevalence in children under 10 was modeled using climatic, population and topographic variables as potential predictors. After regression analysis, spatial dependence of the model residuals was investigated. Kriging on the residual was used to model local variation in malaria risk over and above that which is predicted by the regression model. Results from this study were illustrated by a map showing the improvement of risk prediction brought about by the second stage.

Srivastava et al (2001) investigated *Anopheles dirus*, which is found in deep-forested areas in India where manual surveys are very difficult because of inaccessibility. A Geographic Information System (GIS) and a Boolean operator were used to map areas where the species is likely to be found. Being a forest-based species, thematic maps of forest cover, altitude, rainfall and temperature were prepared. Overlaying and integration of thematic maps were done using Arc/Info NT and analysis by ArcView 3.1 (GIS ESRI) software. The results were validated through reported distribution and were found to be correct. This technique can cover vast and inaccessible areas, and is fast and easily duplicable in other parts of the world. Once the vector distribution is known, species-specific control measures can be formulated.

Statistical modeling for malaria epidemics in Thailand

Tiensuwan et al (2000) identified risk factors causing malaria in Tak province in the rainy season by using log-linear models. Tests of independence are used (chi-square and Cramer's V-value tests) to find out the relationships between any two variables. In addition, two- and three-dimensional log-linear models are used to obtain estimated

parameters and expected frequencies for these models. Among the models fitted, the best are chosen based on analysis of deviance. The results of this study show that most observed variables are significantly related, with p-value < 0.05. Causes of migration and reasons for staying overnight are highly related to personal variables. Thus, it can be concluded that two of the risk factors for malaria are causes of migration and reasons for staying overnight. Knowledge of prevention is also related to personal variables. Therefore, knowledge of prevention was concluded to be a risk factor affecting prevalence of malaria. For each set of three variables, the best model shows interaction terms of variables that have a relationship, but there are no interactions of three effects in these best models.

Malaria epidemics in Thailand

Konchom et al (2003) focused on the 30 provinces of Thailand situated next to neighboring countries, which can be divided into four groups: the Thai-Myanmar border (10 provinces), the Thai-Cambodia border (6 provinces), the Thai-Laos border (10 provinces) and the Thai-Malaysia border (4 provinces). The purpose of the study was to describe the pattern and trend of malaria incidence in the highly endemic provinces along the Thai borders for the 11 years from 1991 to 2001. Analysis of trends showed the distribution of malaria parasites to have shifted from a preponderance of *Plasmodium falciparum* to *Plasmodium vivax* along the western border with Myanmar, the northern border with Laos, and along the eastern border with Cambodia; whereas since 1997, along the southern border with Malaysia the pattern changed from a preponderance of *P. vivax* to *P. falciparum*. The results showed a significant difference in annual parasite incidence between border and non-

border districts, especially along the Thai-Myanmar and Thai-Cambodia borders. It is thus evident that all border districts should pay more attention to the control of malaria transmission and the activities of the malaria surveillance system, and that monitoring and evaluation of the Thai Malaria Control Program needs to be performed.

Chaveepojnkamjorn and Pichainarong (2005) used a cross-sectional study conducted from January 2001 to June 2002 among certain migrant populations living in malaria endemic areas in districts along the Thai-Myanmar border. They surveyed and examined residents in Mae Fah Luang and Mae Sai districts, Chiang Rai province, Northern Thailand. They focused on the knowledge and practice of primary malaria prevention, aimed at identifying the association between behavioral factors in migrant populations and malaria infection. *P. vivax* (51.8%) was detected more often than *P. falciparum* (47.7%). The proportion of malaria infections was 45.4% of the total of 421 blood examinations. The working age group (15-44 years) and males comprised the majority of the study subjects. Two age groups (0-14 and 15-34 years), and visiting or staying in the forest 14 days prior to the blood exam were significant risk factors. Thai-Yai and hill tribe ethnic groups demonstrated a significant protective factor (p-value < 0.05) compared to other Myanmar people. A poor knowledge of primary malaria prevention (63-68%), the presence of international migration, poverty, lack of malaria prevention resources (namely not using bed nets), and not using a smoky fire were factors which led to failure in primary prevention and control of malaria infections. Residence-workplace-living styles in the forest need a common effective method of primary prevention.

1.4 Outline of thesis

This thesis contains four chapters, including this chapter which, in addition to a rationale for study, gives a definition of incidence and includes a literature review. Chapter 2 is methodology. It provides a description of data including data sources, data management, demographic factors and an overview of the statistical methods used to forecast malaria incidence. Chapter 3 is data analysis and models for malaria. This chapter divides into two studies. First study is forecasting malaria in north-western Thailand. Second study is modeling malaria incidence in north-western Thailand. This involves characteristics of data, distribution of incidence rates and result from fitting statistical modeling of using the models described in Chapter 2. Chapter 4 is conclusion and limitations.

Chapter 2

Methodology

This chapter describes the statistical methodology used in this study, and statistical model fitting. These include linear regression models, Poisson regression models, negative binomial regression models, additive log-normal models, additive plus multiplicative models and principal components analysis.

Graphical and statistical analyses used programs written in R software (R Development Core Team 2008) for most of the analysis.

2.1 Sources of data

This study is a cross-sectional survey distributed by time and area. Data used in this study were obtained from a registry of hospital-diagnosed infectious disease cases collected routinely in each of Thailand's 76 provinces by the Ministry of Public Health. The records of every case have 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). This epidemiological data is called the 506 Surveillance Database. The study data consisted of malaria cases from the databases of provincial health offices in Lamphun, Phrae, Nan, Chiang Rai, Mae Hong Son, and Tak provinces from 1999-2004.

In 1963, the World Health Organization (WHO) joined with the Thailand Ministry of Public Health to set up an epidemiology project in order to conduct surveillance and

control of infectious diseases, and to develop an organization able to conduct epidemiological surveillance of infectious diseases over time. The organization responsible is the Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health.

The next step, the Bureau of Epidemiology was to establish epidemiology centers at the provincial, district and sub-district levels in order to collect patient data regarding the epidemiology of 78 infectious diseases. The epidemiology center at each level collected data on 506 Surveillance Database patient cards, creating a standard epidemiological infectious disease registry. Beginning in 1995, this data was processed into a computer instead of a register book. Surveillance of the epidemiology of infectious diseases comprised both government and private public health care. When a patient was diagnosed with one of the infectious diseases under surveillance, a 506 patient card was created and sent to the epidemiology center at the tambon or district in which the patient resided. The report was then entered into the computer program, and forwarded to the epidemiology center at a higher level. The epidemiology units of provincial public health offices then collected and sent the 506 reports to the Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health, for processing and analysis in order to plan prevention and disease control in particular areas.

The scope of this study is malaria cases from 1999-2004 in six provinces in north-western Thailand. The study models the pattern of hospital-diagnosed malaria incidence in order to determine where malaria incidence was high. The outcome variables are defined as the incidence rate in a cell indexed by district, age group and period, with district, age group and calendar period (allowing for a trend effect) as

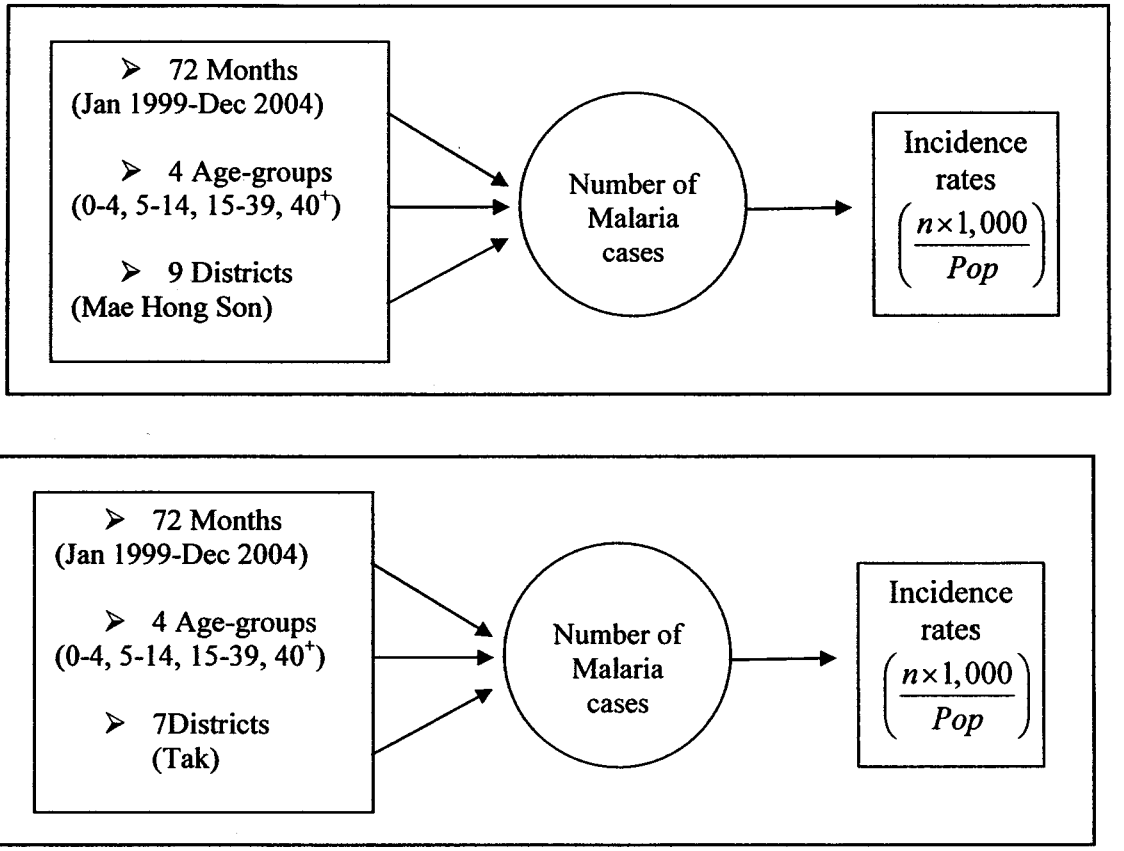
categorical determinant variables, and with reference to border and non-border districts as intervening variables. The target population was malaria cases in the provinces selected: Lamphun, Phrae, Nan, Chiang Rai, Mae Hong Son and Tak, all of which have high malaria incidence rates.

2.2 Data management

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 or quarter (72 months from January 1999 to December 2004 or 24 quarter 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 2000 Population and Housing Census of Thailand. Since there was little evidence of a sex effect the data for the two sexes were combined.

In this thesis a two-study path diagrams were followed: (1) Forecasting malaria in North-Western Thailand and (2) Modeling malaria incidence in North-Western Thailand. The first path diagram studies malaria incidence rate of an adverse event based on cells classified by demographic variables including age group, district, and period of time (monthly) in Mae Hong Son and Tak provinces. The second path diagram studies malaria incidence rate based on cells classified by demographic variables including age group, district, and period of time (quarterly) in six provinces (65 districts): Lamphun, Phrae, Nan, Chiang Rai, Mae Hong Son and Tak provinces.

Study 1: Forecasting malaria in North-Western Thailand: 1999-2004



Study 2: Modeling malaria incidence in North-Western Thailand: 1999-2004

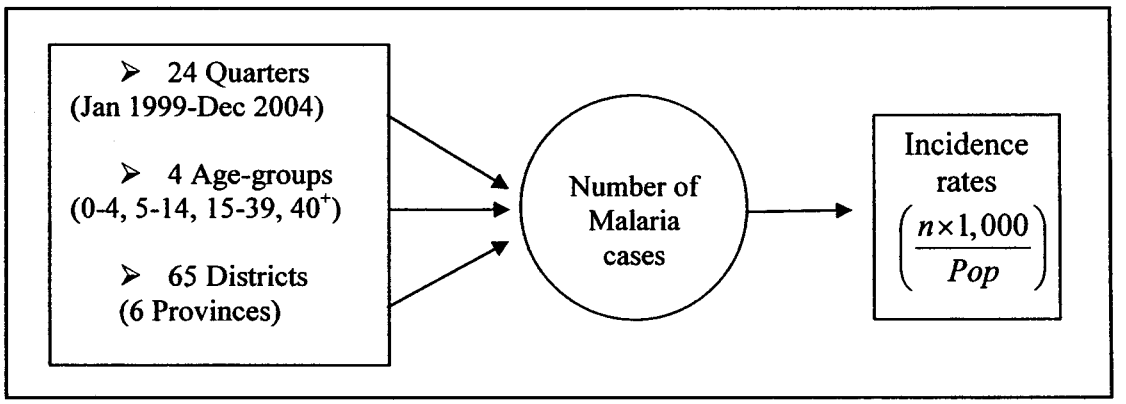


Figure 1: Path diagram for studies

2.3 Demographic factors

This section reviews the general demographic factors influencing malaria incidence rates. In the study of malaria infection in Mozambique by Abellana et al (2008), it was found that children under five years of age are at the highest risk, probably because they have not yet developed sufficient immunity, making these early years a particularly susceptible period.

Many studies of malaria have tended to focus on explaining patients' choice of "provider" or "outlet." For example, Asenso-Okyere et al (1997) found that the choice of provider of malaria care is influenced by facility price, travel time, waiting time for treatment, and a range of demographic factors (including education, age and sex). Another important factor is the quality of care measured in terms of drug availability (Goodman et al 2003). Charles et al (2005) investigated the changes in and factors associated with recent malaria notification trends in Western Australia (WA). The study design was a retrospective analysis of the WA Notifiable Infectious Diseases Database consisting of enhanced surveillance questionnaires completed by attending medical practitioners. The patients were those with cases of malaria reported between January 1990 and December 2001. Main outcome measurement was annual notification by demographic variables (including age, sex, occupation and place of residence), region/country of acquisition, chemoprophylaxis used, *Plasmodium* species, and outcome (There was little evidence of a sex effect on the data). From this study, demographic factors were defined as sex, age group, occupation, residence and migration.

2.4 Statistical methods

2.4.1 Linear regression models

At the beginning of 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 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

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

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:

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

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.

2.4.2 Poisson regression models

A Poisson time series model can be used as a starting point. Assume that N_t, $t = 1, \ldots, n$ is a time series of counts taking non-negative integer values. The observation driven Poisson model is defined in terms of its density function, that is,

$$\text{Prob}(N_t = n) = \frac{\lambda_t^n}{n!}\exp(-\lambda_t), \quad \text{where } N_t \text{ is } 0, 1, 2, 3, \ldots \qquad (4)$$

Variable N_t have a Poisson distribution with mean λ_t and variance λ_t.

Davis et al (2003) suggested observation-driven models for time series counts N_t based on the Poisson distribution with mean λ_t where $\ln(\lambda_t)$ is expressed as an additive function of determinants and lagged observations on N_t. 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 λ_{ijt} is the mean of N_{ijt}, a suitable generalized linear model based on the Poisson distribution could take the form:

$$\ln(\lambda_{ijt}) = \ln(p_{ij}) + \mu + \alpha_i + \beta_j + \eta_s + \sum_{k=1}^{m} \gamma_k y_{ij,t-k} + \delta_1 y_{ij,t-1}^{(\alpha)} + \delta_2 y_{ij,t-1}^{(\beta)} . \qquad (5)$$

2.4.3 Negative binomial regression models

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 γ and takes the form:

$$\text{Prob}(N_t = n) = \frac{\Gamma(n+\gamma)}{\Gamma(n+1)\Gamma(\gamma)} \left(\frac{\gamma}{\gamma+\lambda_t}\right)^{\gamma} \left(\frac{\lambda_t}{\gamma+\lambda_t}\right)^{n} . \qquad (6)$$

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

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 (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 (6) is (Wackerly et al 1996)

$$E[\exp(\theta N_t)] = \left(1+\frac{\lambda_t}{\gamma}-\frac{\lambda_t}{\gamma}e^{\theta}\right)^{-\gamma}, \qquad (7)$$

From equation (7), the moment generating function for the standardized residuals is

$$E(e^{\theta z}) = E\left[\exp\left(\frac{\theta(N_t-\lambda_t)}{\sqrt{\lambda_t+\lambda_t^2/\gamma}}\right)\right] = \left(1+\frac{\lambda_t}{\gamma}-\frac{\lambda_t}{\gamma}\exp\left(\theta/\sqrt{\lambda_t+\lambda_t^2/\gamma}\right)\right)^{-\gamma}\exp\left(-\theta\sqrt{\lambda_t}/\sqrt{1+\lambda_t/\gamma}\right). \quad (8)$$

The limiting distribution of $E(e^{\theta z})$ as $\lambda_t \to \infty$ is now obtained using a Taylor expansion, as follows:

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

$$=\left(1-\frac{\theta}{\gamma\sqrt{1/\lambda_t+1/\gamma}}-\frac{\theta^2}{2\gamma(1+\lambda_t/\gamma)}\right)^{-\gamma}\exp\left(-\theta\sqrt{\gamma}\right)+o(1/\lambda_t), \qquad (9)$$

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

$$(1-\theta/\sqrt{\gamma})^{-\gamma}\exp(-\theta\sqrt{\gamma}) \qquad (10)$$

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

2.4.4 Additive log-normal models

We consider alternative statistical models for describing the relation between malaria incidence and the age group, district and period factors, namely a 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

$$Y_{ijt} = \alpha_i + \beta_j + \gamma_t \qquad (11)$$

where the indexes i, j and t denote age-group, district, and period, respectively and 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

$$E\left[\log\left(y_{ijt}\right)\right] = \alpha_i + \beta_j + \gamma_t \qquad (12)$$

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 (Konchom et al 2003), 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 (12) to

$$E\left[\log\left(y_{ijt}\right)\right] = \alpha_{ik} + \beta_j + \gamma_t, \qquad (13)$$

where $k = 1$ for border districts, $k = 2$ otherwise. Since districts are nested within border location, one of the α_{ik} parameters in this model is redundant, requiring a constant to be specified, for definitions we assume $\alpha_{11} = \alpha_{42}$.

2.4.5 Additive plus multiplicative models

Since we also wish to investigate the possibility that the trend γ_t varies from district to district, we consider a further extension of model (Equation 13) to include such a multiplicative interaction by replacing γ_t by $\delta_j\gamma_t$, that is

$$E\left[\log\left(y_{ijt}\right)\right] = \alpha_{ik} + \beta_j + \delta_j\gamma_t \qquad (14)$$

This model, though non-linear, can be fitted by regression analysis after first computing γ_t 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 δ_j. This technique appears to have been first used by Fisher and McKenzie (1923). It was later used by McNeil and Tukey (1973) to model socio-demographic data in regions and has been applied to spectroscopy, chromatography and other fields to model additive mixtures of overlapping curves (Lawton and Sylvestre 1971, Sylvestre et al 1974, Theil 1983). The method has also been used more recently for modeling age-specific mortality rates (Good 1969, Wilmoth et al 1989, Booth et al 2002). Graphical plots against normal quantiles of deviance residuals (for the negative binomial model) and studentized residuals (for the additive 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 competed using sum contrasts.

Adjusted Incidence Rates

After fitting the model, adjusted incidence rates for each factor of interest are obtained by suppressing the subscripts in Equations 13-14 corresponding to the other factors and replacing these terms with a constant satisfying the condition that the sum of the counts based on the adjusted incidence rates matches the total (Liu et al 2005).

Sum Contrasts

Sum contrasts (Venables and Ripley 2002, Tongkumchum and McNeil 2009) are used to obtain confidence intervals for comparing adjusted incidence rates within each factor with the overall incidence rate. An advantage of these confidence intervals is that they provide a simple criterion for classifying levels of a factor into three groups according to whether each corresponding confidence interval exceeds, crosses, or is below the overall mean.

2.4.6 Principal component analysis

Malaria incidence rates were obtained for each combination of 24 quarters (24 periods from January-March 1999 to October-December 2004), 4 age groups (0-4, 5-14, 15-39 and 40+ years), and 65 districts. 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.

In order to characterize these district patterns, principal components analysis can be used. This technique appears to have first been used by Fisher and Mackenzie (1923). Later, McNeil and Tukey (1973) used it to model regional socio-demographic data. Principal components analysis has since been applied to spectroscopy, chromatography and other fields to model additive mixtures of overlapping curves (Lawton and Sylvestre 1971, Sylvestre et al 1974, Theil 1983).

The method of principal components is based on a data matrix Y with, in our case, dimension 65 × 24 whose columns correspond to the 24 quarter-year periods of interest and whose rows correspond to the 65 districts. A second-moment matrix V is computed as:

$$V_{p\times p} = Y'Y \tag{15}$$

where Y is a matrix with (i, j) element y_{ij} having n rows and p columns, and Y' is its transpose of.

Alternatively, the matrix V could be taken as the covariance matrix corresponding to the data matrixes Y. The choice between the second-moment matrix and the covariance matrix depends on whether it is preferable to correct the variables before

or after the principal component analysis. Similarly, the choice between the covariance matrix and the correlation matrix depends on whether it is preferable to scale the variables before or after the principal components are computed. In practice, these choices are influenced by the objective of the study and the application model. For example, when the objective is simply to find common factors among the variables of interest, a principal component analysis would preferably be based on the correlation matrix.

In our study we have chosen to use the second-moment matrix as the basis for modeling malaria incidence rates because the age-district-specific means contain important information, and fitting a model using a single procedure that incorporates these means is simpler and more direct than the alternative method of first subtracting these means and then adding them back after fitting a model on the principal components of the covariance matrix.

Denote by B the orthonormal $(p \times p)$ matrix with elements β_{kj} whose columns are the scaled eigenvectors of matrix V. These satisfy the constraints:

$$\sum_{j=1}^{p} (\beta_j^{(k)})^2 = 1 \quad for \quad j = 1, 2, ..., p \tag{16}$$

$$\sum_{j=1}^{p} \beta_j^{(k)} \beta_j^{(l)} = 1 \quad for \quad k \neq 1 \, . \tag{17}$$

The columns are ordered by descending values of the corresponding eigen values; then the values of the p principal components for the p variables can written in terms of an $(n \times p)$ matrix A as:

$$A = YB \tag{18}$$

where $$\sum_{t=1}^{n} a_t^{(k)} a_t^{(l)} = 0 \quad for\ k \neq 1. \tag{19}$$

Since $B^{-1} = B'$, post-multiplying both sides of equation $A = YB$ by B' gives:

$$Y = AB'. \tag{20}$$

Then, the least squares fitted model using the first m eigenvalues from $Y = AB'$ is:

$$\hat{Y}_{ij} = a_t^{(1)} \beta_j^{(1)} + a_t^{(2)} \beta_j^{(2)} + ... + a_t^{(m)} \beta_j^{(m)}. \tag{21}$$

In this model, $\hat{Y}_{ij}$ is the estimate of Y_{ij} on the first m principal components, $a_t^{(k)}$ is the i^{th} score from the k^{th} principal component, and $\beta_j^{(k)}$ is the j^{th} coefficient of the eigenvector from the k^{th} principal component.

Residuals

A measure of the discrepancy between Y_{ij} and $\hat{Y}_{ij}$ is the residuals $e_{ik} = Y_{jk} - \hat{Y}_{jk}$. These residuals contain information where the model might fail to fit the data, so it is worthwhile to check their behavior. Plotting the residual standard deviations for each variable indicates lack of homogeneity. The residual covariance matrix, having elements

$$c_{jk} = \frac{1}{n-1} \sum_{t=1}^{n} \left(e_{ij} - \hat{e}_j\right)\left(e_{ik} - \hat{e}_k\right), \tag{22}$$

also provides an indication of the adequacy of the model.

Covariance matrix

Alternatively, the principal components analysis can be based on the covariance matrix:

$$V_{p\times p} = (Y - \bar{y})'(Y - \bar{y}), \tag{23}$$

where $\bar{y}$ is the vector of p column means. In this case the fitted model based on the first m eigenvectors is:

$$\hat{Y}_{ij} = \mu + \gamma_j + a_t^{(1)}\beta_j^{(1)} + a_t^{(2)}\beta_j^{(2)} + ... + a_t^{(m)}\beta_j^{(m)} \tag{24}$$

where μ is the overall mean, γ_j is the j^{th} column effect (district-age effect, in our case) and $\sum_{j=1}^{p}\gamma_j = 0$. The constraints in equations 16 -17 also apply in this model, since $\sum_{j=1}^{p}\gamma_j = 0$. Thus the number of degrees of freedom for the denominator is now $\nu_1 = (n-m)(p-m) - p$. The number of degrees of freedom for the numerator is defined as $df_0 - df_1 = n + p - 2m + 1$ and $df_1 = (n-m)(p-m)$.

Chapter 3

Data Analysis and Models for Malaria

The analysis in this chapter is based on graphical and conventional statistical methods. Graphical and statistical analyses were carried out using programs written in R software (R Development Core Team 2008) were used for most of the preliminary analysis.

3.1 Malaria in North-Western Thailand

In this section we present a preliminary analysis and fitting a model of the time series of monthly malaria incidence over the period from January 1999 to December 2004. The analysis is based on the graphical method and statistical methods. First, the characteristics of the data are described. After that, distributions of incidence rates are explained. Finally, we present result from fitting a model (Equations 2-3 and 5-6, Chapter 2) to the data and finally its summary. The results presented in this section also appear in Sriwattanapongse et al (2008) and is reproduced in Appendix II, and the second of which has been submitted for publication.

3.1.1 Characteristics of the data

Data were obtained from a registry of hospital-diagnosed infectious disease cases collected routinely in each of Thailand's 76 provinces by the Ministry of Public Health. We created malaria disease counts for specific age groups and districts, either monthly (72 months from January 1999 to December 2004), Malaria incidence rates were calculated based on the number of malaria cases per 1,000 residents in the

district, according to the 2000 Population and Housing Census of Thailand, for each age group and period.

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 health offices. The provinces selected, Mae Hong Son and Tak, both border Myanmar and have high malaria incidence rates.

Based on an individual hospital case records routinely reported in each province from 1999-2004, linear regression models of log-transformed incidence rates (Equations 2-3, Chapter 2) 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 (Equation 4-6, Chapter 2)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.

3.1.2 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 1,000 in Tak. The time series of average monthly age-specific malaria rates per 1,000 populations at risk for all districts in each province are plotted in Figure 2.

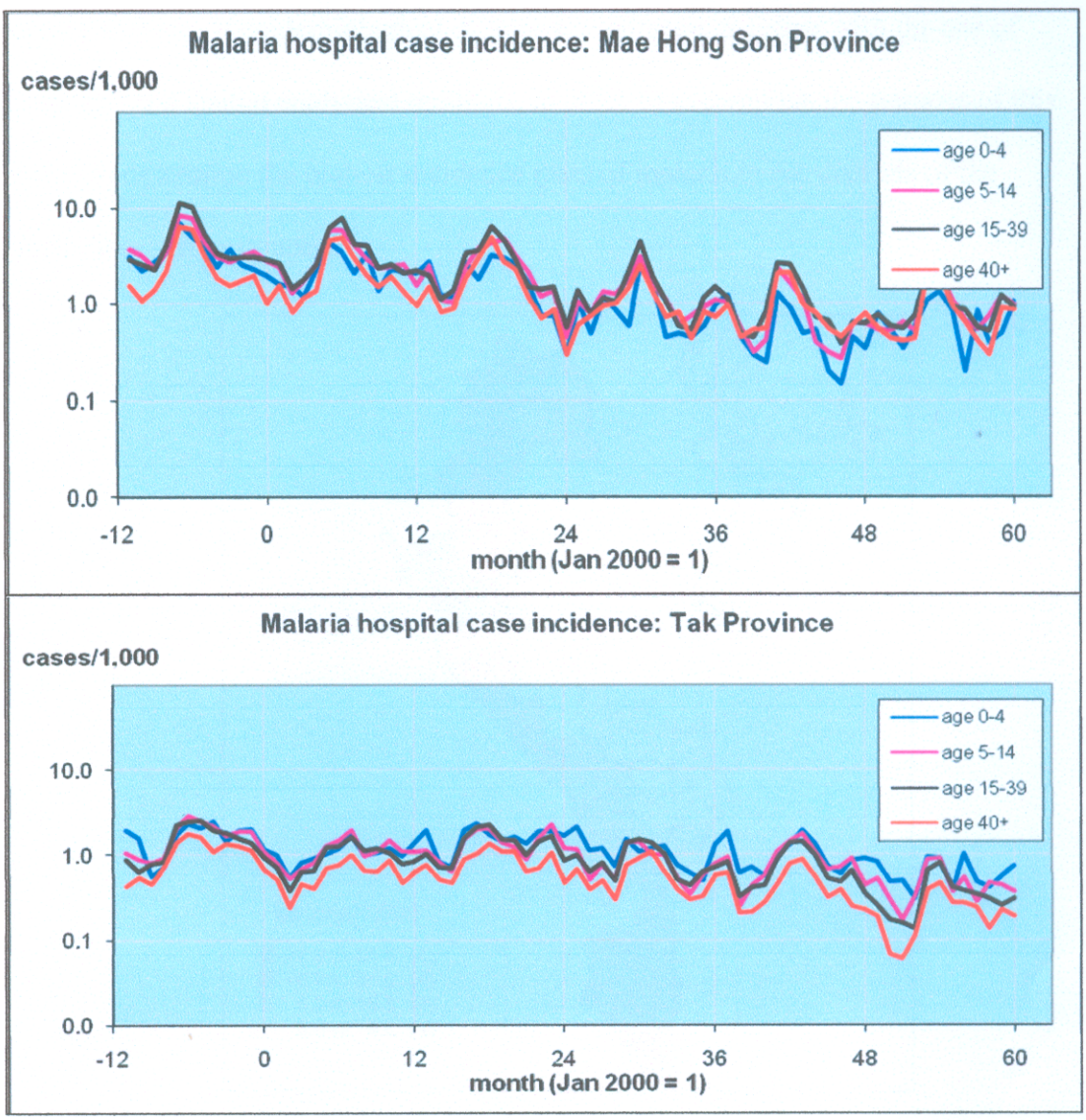


Figure 2: Time series of monthly disease rates for each age group in the two provinces

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.

Geographical Information System (GIS) has rich repertoire of powerful tools that can aid in decision making for public health issues. It has been applied in this work for spatial analysis of epidemiological disease surveillance, together with the use of statistical and spatial analytical methods, in order to accomplish the purpose of this study: to determine the highest incidence rates of malaria in the districts of the North-Western provinces.

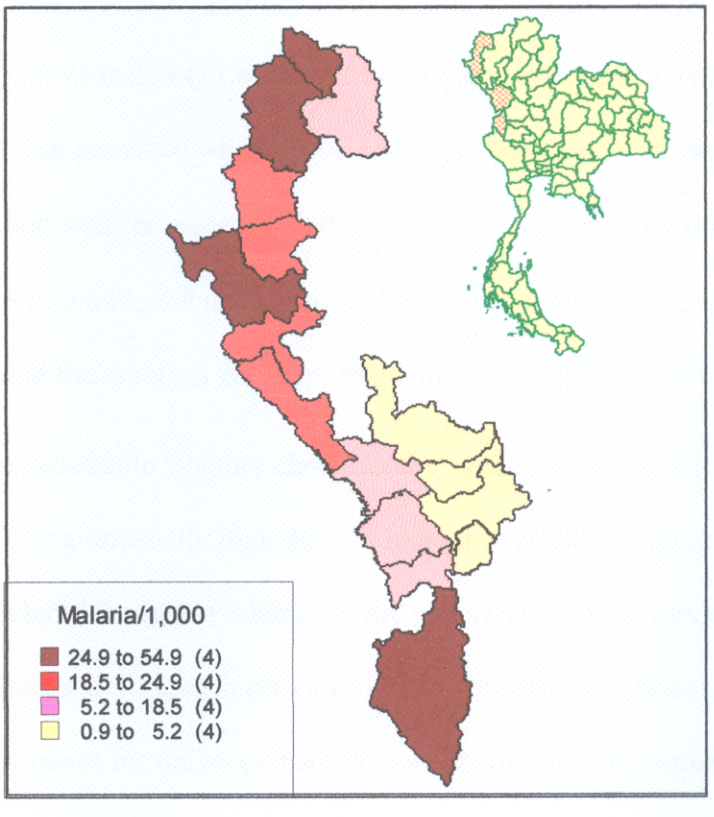


Figure 3: Average annual incidence rates for malaria in Mae Hong Son and Tak provinces, 1999-2004

Figure 3 thus 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 (Sam

Ngao, Ban Tak, Mueang Tak and King Amphoe Wang Chao), and the highest occurred in the southernmost district of Tak (Umphang: 49.5 cases/1,000/year) and Mueang Mae Hong Son in the north (31.3 cases/1,000/year).

3.1.3 Fitting a model to the data

Linear regression models

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 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).

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: δ_1	0.224	0.026	0.291	0.024
2: γ_2	0.091	0.024	0.146	0.021
3: γ_3	0.075	0.022	0.113	0.020
Other Age Groups: δ_1	0.341	0.034	0.172	0.028
Other Districts: δ_2	0.211	0.036	0.148	0.031
r-squared statistic	0.6121		0.8099	

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.

Poisson and negative binomial models

Turning to the Poisson and negative binomial regression models given by Equations (5) and (6), Figure 4 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 appropriate. 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 and Table 3 give the results obtained from fitting the Poisson and negative regression model given by Equations (5) and (6) to the malaria disease counts for each of the two provinces. For negative binomial model, the dispersion parameter estimates (γ) are 3.521 for Mae Hong Son and 3.725 for Tak.

Table 2: Results from fitting Poisson models to malaria disease counts

Determinant	*Mae Hong Son*		*Tak*	
	Coefficient	St. Error	Coefficient	St. Error
Constant	0.234	0.047	-0.977	0.057
Age Group: 0-4	0	0	0	0
5-14	0.060	0.025	0.038	0.020
15-39	0.124	0.024	0.015	0.019
40+	-0.015	0.026	-0.023	0.024
District: 1	0	0	0	0
2	-0.150	0.027	0.180	0.056
3	-0.402	0.036	0.130	0.065
4	-0.091	0.018	0.979	0.047
5	-0.134	0.021	1.096	0.050
6	-0.121	0.020	0.602	0.040
7	-0.126	0.026	0.807	0.045
8			1.214	0.057
9			0.619	0.051
Month: January	0	0	0	0
February	0.249	0.037	0.224	0.032
March	-0.375	0.043	-0.416	0.038
April	0.148	0.042	0.147	0.038
May	0.556	0.034	0.324	0.033
June	1.052	0.030	0.889	0.028
July	0.580	0.032	0.562	0.028
August	-0.174	0.034	0.294	0.029
September	-0.192	0.035	0.044	0.029
October	-0.197	0.037	-0.084	0.031
November	0.046	0.035	0.210	0.030
December	0.259	0.035	0.255	0.030
Autoregressive Lag: 1: γ_1	0.391	0.014	0.509	0.020
2: γ_2	0.101	0.012	0.142	0.019
3: γ_3	0.032	0.010	0.087	0.024
Other Age Groups: δ_1	0.201	0.016	0.050	0.017
Other Districts: δ_2	0.156	0.014	0.161	0.016
Residual Deviance	8,135.9		8,898.4	

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 (= 32+36+18+78+168+53) 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 (47, 18, 18, 38, 53, and 29, respectively) after that, this particular outlier heralded a large epidemic.

Table 3: 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: γ_1	0.292	0.027	0.412	0.032
2: γ_2	0.090	0.024	0.168	0.028
3: γ_3	0.069	0.022	0.145	0.026
Other Age Groups: δ_1	0.284	0.033	0.121	0.036
Other Districts: δ_2	0.161	0.032	0.247	0.038
Dispersion: γ	3.521	0.181	3.725	0.188
Residual Deviance	2,212.7		2,834.2	

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(6, 19, 25, 21, 12, and 24), and 26 cases followed by another 97 (28, 13, 5, 13, 23, and 15), 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 Mueang Tak in July 2003, with no further cases recorded in this district and age group in the following six months.

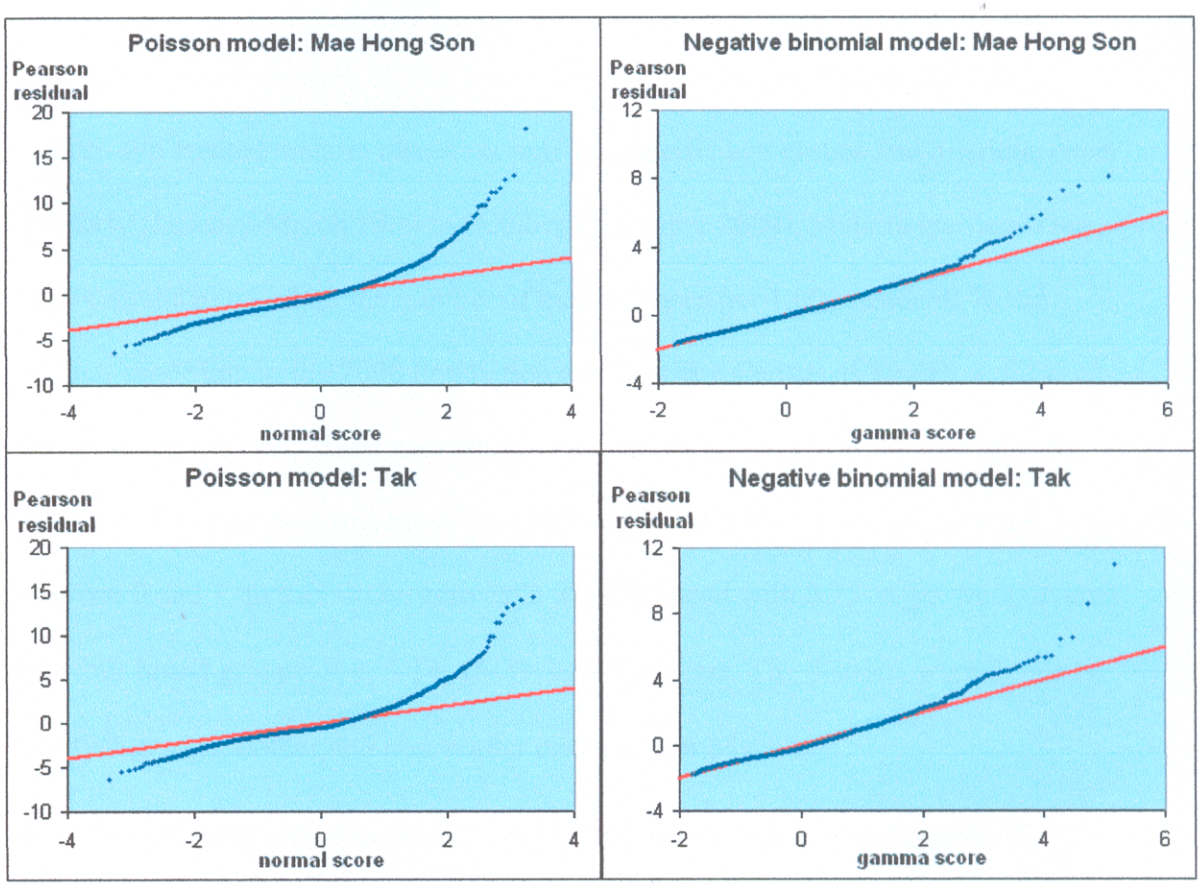


Figure 4: Plots of Pearson residuals versus asymptotic scores after fitting Poisson and negative binomial models for malaria disease counts in the two provinces

3.2 Modeling Malaria Incidence in North-Western Thailand

In this section we present some preliminary analysis and further analysis using statistical models in chapter 2 to the time series of quarterly malaria incident over the period from January-March 1999 to October-December 2004 with six provinces in north-western Thailand. The analysis also is based on the graphical and statistical methods. The results presented in this section also appear in Sriwattanapongse et al (2009) and is reproduced in Appendix III, and the second of which has been submitted for publication.

3.2.1 Characteristics of the data

Data were obtained from a registry of hospital-diagnosed infectious disease cases collected routinely in each of Thailand's 76 provinces by the Ministry of Public Health. We created malaria disease counts for specific age groups and districts, either quarterly (January-March 1999 to October-December 2004). Malaria incidence rates were calculated based on the number of malaria cases per 1,000 residents in the district, according to the 2000 Population and Housing Census of Thailand, for each age group and period. Many zero counts were recorded among the age groups. To allow for zero counts in this model, we replaced them with a specified constant between 0 and 1. In this study, zero cells were replaced with 0.25. Ages were divided into four broad groups: 0-4, 5-14, 15-39, and 40+ years. The districts were divided into two groups: border and non-border districts near Myanmar.

The objective of this 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.

3.2.2 Distributions of incidence rates

Between January 1999 and December 2004, 67,347 hospital cases of malaria 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.

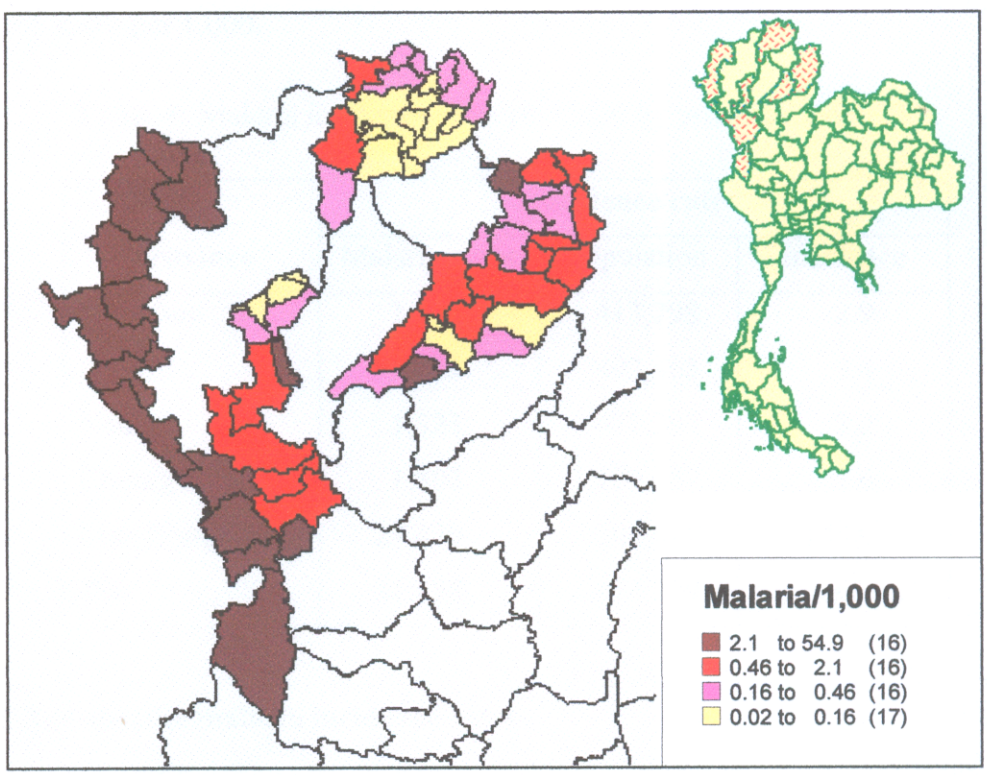


Figure 5: Average annual incidence rates for malaria in North-Western of Thailand

Figure 5 shows that how the average malaria incidence rate varied by district in the six provinces. The lowest rates occurred in the seventeen non-borer districts, and the

highest occurred in the 13 border district and 3 non-border districts. The highest malaria incidence in north-western is Umphang (54.89 cases/1,000/year).

Table 4 shows that during the study period from January 1999 to December 2004, 1,195 hospital cases of malaria were reported in Lamphun, 1,689 in Phrae, 1,523 in Nan, 1,803 in Chiang Rai, 29,498 in Mae Hong Son and 31,658 in Tak. Thus, a total of 67,347 hospital cases of malaria were reported in the 6 provinces (65 districts) considered in the study. The number of cases in any quarterly period for a particular age group and district varied from 0 to 1,023; the mean annual malaria incidence rate was 3.5 cases per 1,000.

Table 4: Malaria incidence rates per 1,000 by province:1999-2004

Province	Incidence rates		
	Malaria cases	Population	Incidence
Lamphun	1,195	413,298	2.9
Phrae	1,689	492,563	3.4
Nan	1,523	471,997	3.2
Chiang Rai	1,803	1,109,057	1.6
Mae Hong Son	29,479	204,842	144.0
Tak	31,658	474,709	66.7

3.2.3 Age-specific malaria incidence rates

The relationship between the pattern of age-specific malaria morbidity and malaria transmission intensity has been well documented (Molineaux 1988, Snow et al 1997, Snow and Marsh 1998). Kleinschmidt and Sharp (2001) calculated malaria incidence rates in South Africa in terms of five-year age groups based on a comprehensive small area malaria reporting system and from national census data for the period from mid-

1990 to mid-1999. Incidence rates were plotted against age groups for each of the nine malaria seasons, and by quintile of crude incidence rate. The results showed that age-specific incidence varied considerably in areas of high incidence and in years of high incidence. In these areas, malaria incidence rose with age until the late teens and either remained constant or decreased in young adults. Table 5 shows malaria incidence by age group in six of the North-Western provinces of Thailand.

The highest malaria incidence rate was age group below 5 year in Mae Hong Son province (118.65) because all of Mae Hong Son district were along the border Myanmar. It was workplace of Myanmar area and the mass migration of refugees in response to civil disturbances (Sriwattanapongse et al 2008).Malaria normally occurred in worker. It is commonly high in children below 5 years and namely, it is high in community also.

Table 5: Malaria incidence rates per 1,000 by province and age group: 1999-2004

Province	Malaria incidence by age group			
	0-4	5-14	15-39	40+
Lamphun	1.05	2.70	4.07	2.00
Phrae	0.99	1.12	5.23	2.86
Nan	1.57	3.40	4.34	2.40
Chiang Rai	2.52	2.21	2.34	0.24
Mae Hong Son	118.65	99.29	167.07	97.80
Tak	85.06	90.57	68.90	36.49

The maps in Figure 6 present malaria risk by using the incidence rates in four age groups to determine approximate risk on a larger scale. This map is compare malaria incidence rate in district to district by each age group. The incidence rate was suitable provide in to five classes in each age group. Malaria prevalence was presented as incidence rate by district. The method is illustrated by a map showing the

improvement of risk prediction brought about by the model. Kleinschmidt et al (2000) investigated malaria prevalence in children under 10 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 (Kleinschmidt et al 2000). Bohra and Adrianasolo (2001) 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.

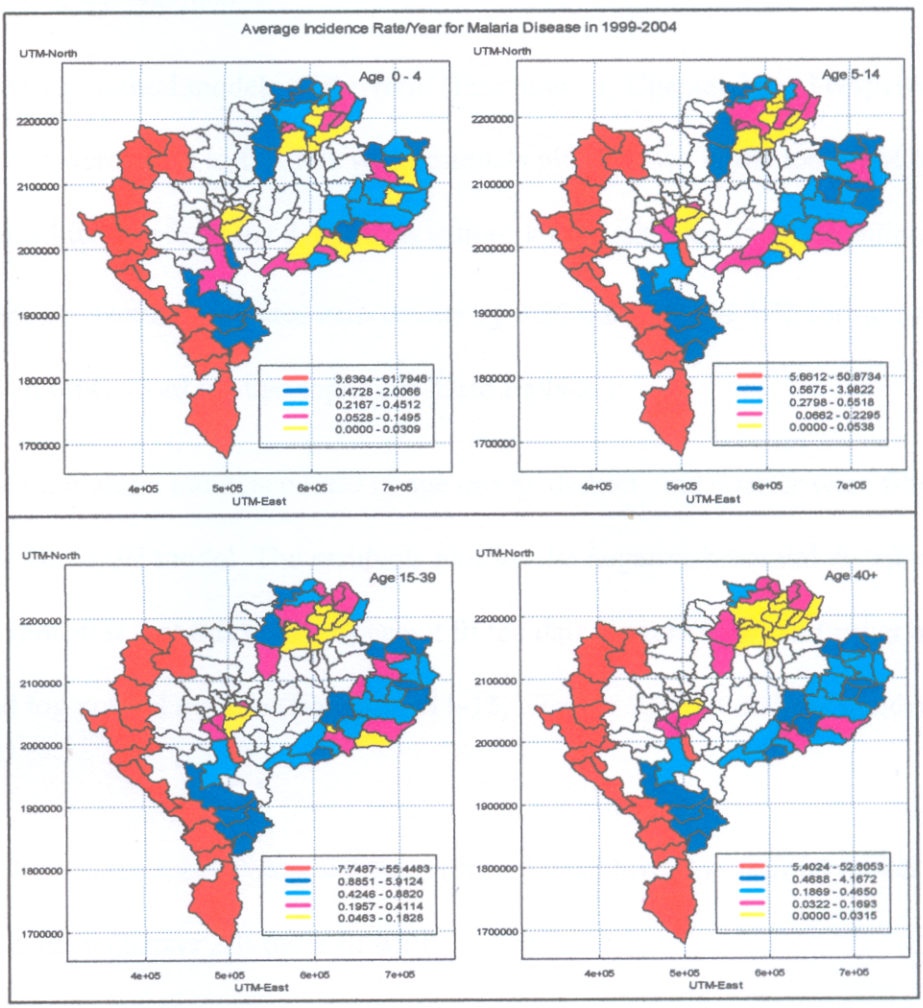


Figure 6: Malaria incidence in provinces of North-Western Thailand by age group: 1999-2004

The present study introduced a methodology to develop malaria social risk categories in the study area. The results provide valuable information for planning precautionary measures and in controlling the spread of malaria. The influence of these factors was obtained in quantitative terms using the Information Value method in the GIS environment. It was found that built-up areas have the highest incidence and hence constitute the highest risk zones.

3.2.4 Fitting a model to the data

Additive log-normal models

Additive log-normal models of the form (Equation 11, Chapter 2) and (Equation 12, Chapter 2) were fitted to the data with residuals plots shown in the top panel of Figure 6. 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.

First of all, malaria incidence rates in age group, district, and quarter were fitted by negative binomial model. The residuals plot for the negative binomial distribution clearly indicates that this model does not fit the data (Figure 6), so in further analysis we used log normal models (Equations 11-13, Chapter 2) to fit malaria incidence rates.

Figure 7 gives the results obtained from fitting the additive log-normal model give by equation 12 in chapter 2.It quite fit well.

Figure 7 shows confidence intervals for adjusted incidence rates by age group, trend and district after fitting model (Equation 13, Chapter 2). The sum contrasts (Venables and Ripley 2002, Tongkumchum and McNeil 2009) was used to obtain confidence

intervals for comparing adjusted incidence rates within each age group, district and trend with the overall incidence rate. The dotted line denotes the overall annual mean malaria incidence. In age group, Its constrains, $\alpha_{11} = \alpha_{42}$, first age group (0-4 years) effect in border district equal to last age group (equal or more than 40 years) effect in non-border district. Malaria incidence rates substantially decrease from very highest level in April- June 1999 quarter to October-December 2004 quarter.

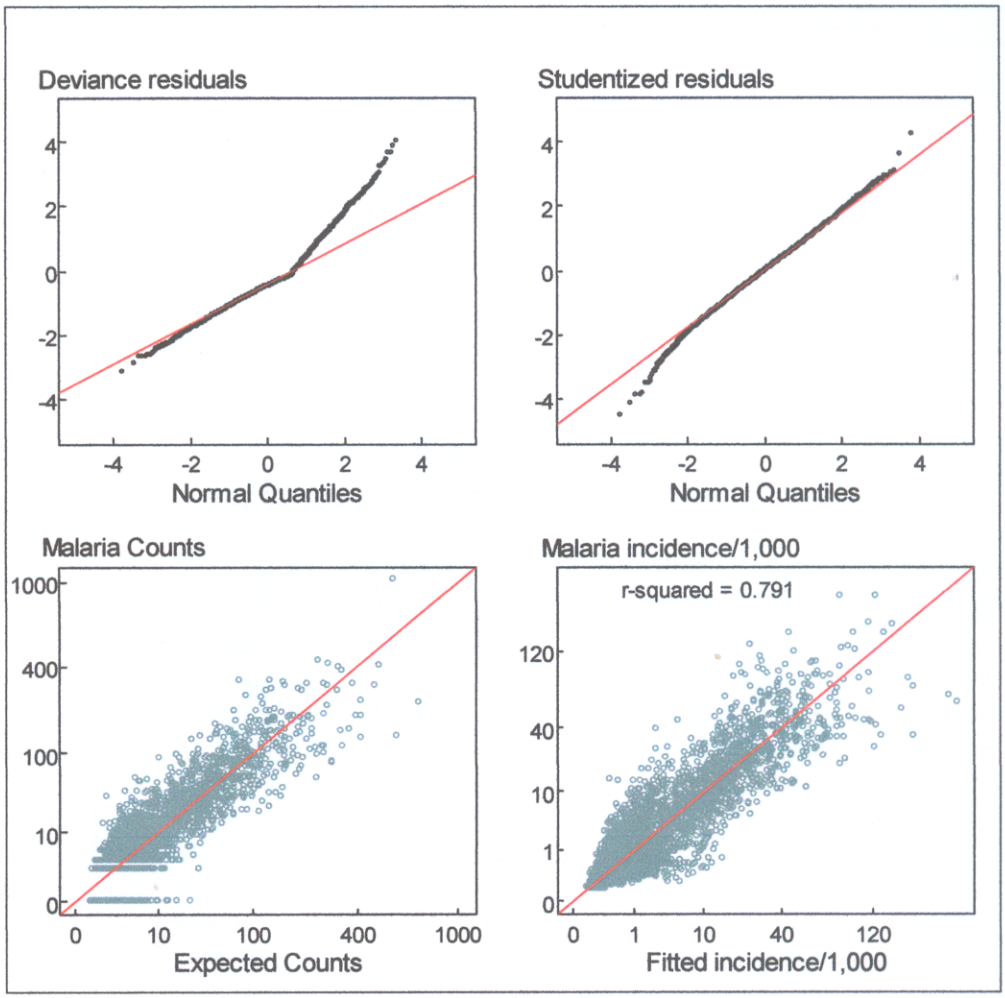
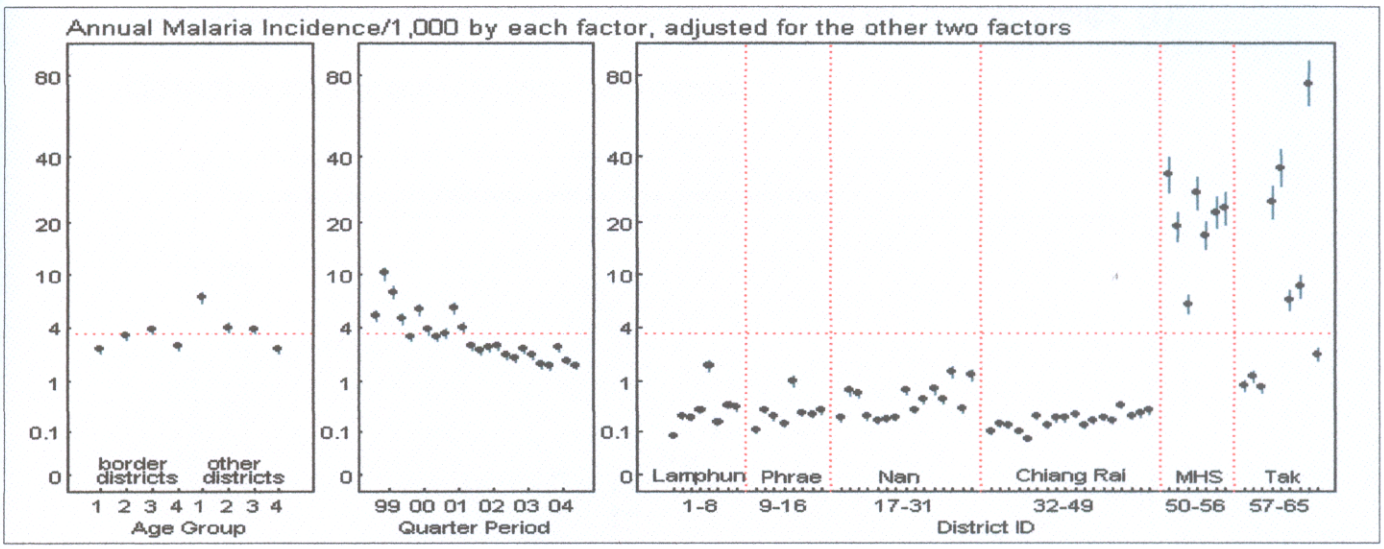


Figure 7: Diagnostic residual plots for negative binomial (top left) and log-linear (top right) models, and plots of counts and incidence rates for the log-normal model (lower panels) for malaria incidence rates in North-Western Thailand

Additive plus multiplicative models

Figure 8 shows malaria incidence rates for each district based on model (Equation 14, Chapter 2) after adjusting for age group and trend. The district specific incidence rates (denoted by the parameter β in the model) are plotted against the parameter δ, 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 81.70%.



ID	Districts	ID	Districts	ID	Districts	ID	Districts	ID	Districts
1	Mueang Lamphun	14	Song	27	Santi Suk	40	Mae Sai	53	Mae Sariang
2	Mae Tha	15	Wang Chin	28	Bo Kluea	41	Mae Suai	54	Mae La Noi
3	Ban Hong	16	Nong Muang Khan	29	Song Khwae	42	Wiang Pa Pao	55	Sop Moei
4	Li	17	Mueang Nan	30	Phu Phiang	43	Phaya Mengrai	56	Pang Mapha
5	Thung Hua Chang	18	Mae Charim	31	Chaloem Phra Kiat	44	Wiang Kaen	57	Mueang Tak
6	Pa Sang	19	Ban Luang	32	Mueang Chiang Rai	45	Khun Tan	58	Ban Tak
7	Ban Thi	20	Na Noi	33	Wiang Chai	46	Mae Fah Luang	59	Sam Ngao
8	Wiang Nong Long	21	Pua	34	Chiang Khong	47	Mae Lao	60	Mae Ramat
9	Mueang Phrae	22	Tha Wang Pha	35	Thoeng	48	Wiang Chiang Rung	61	Tha Song Yang
10	Rong Kwang	23	Wiang Sa	36	Phan	49	Doi Luang	62	Mae Sot
11	Long	24	Thung Chang	37	Pa Daet	50	Mueang Mae Hong Son	63	Phop Phra
12	Sung Men	25	Chiang Klang	38	Mae Chan	51	Khun Yuam	64	Umphang
13	Den Chai	26	Na Muen	39	Chiang Saen	52	Pai	65	Wang Chao

Figure 8: Age group, trend and 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 8 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 9 shows a thematic map based on this classification, with the district specific trend effects also shown on the map. Note that (Figure 8) 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.

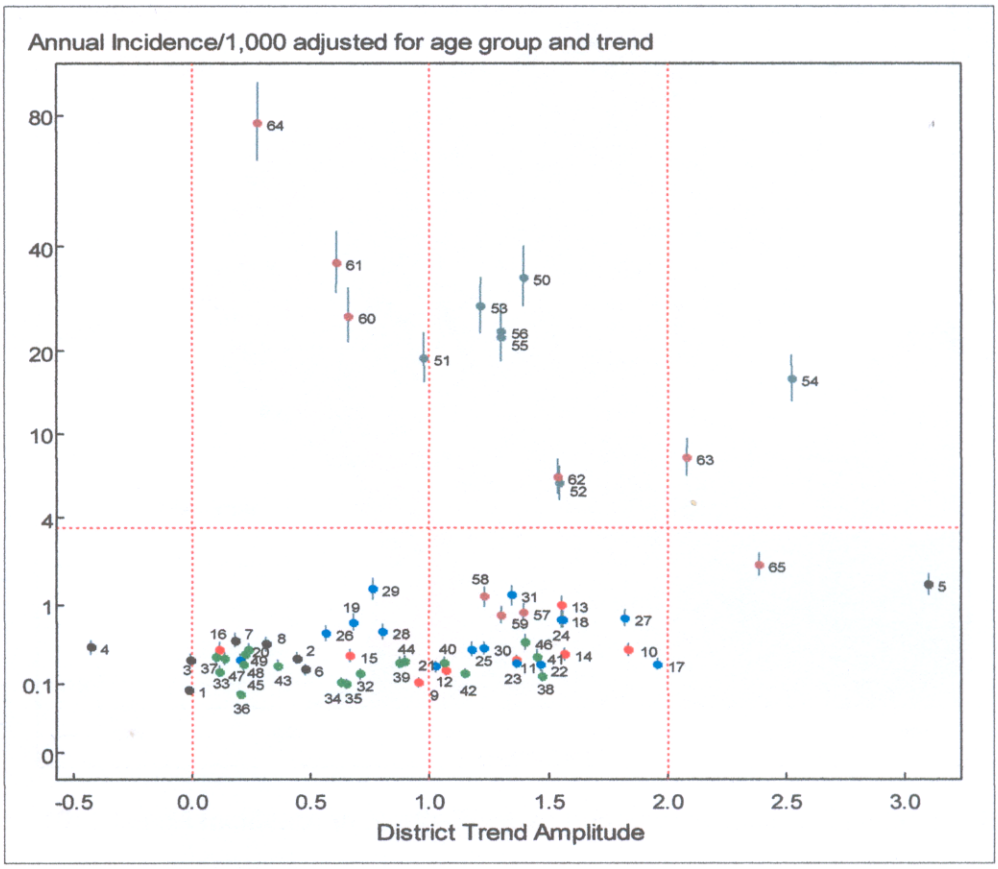


Figure 9: District components of malaria incidence in North-Western Thailand 1999-2004 based on additive plus multiplicative model. The district trend amplitude is defined in the model as the extent to which a district follows the overall trend

Figure 10 shows a substantial difference in annual malaria incidence between districts bordering Myanmar and other districts. Malaria incidence trend in each district were district parameter estimate (β) from model in equation (14). 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 (Mueang Lamphun and Ban Hong in Lamphun) show no trend, and one district (Li in Lamphun) shows an increasing trend.

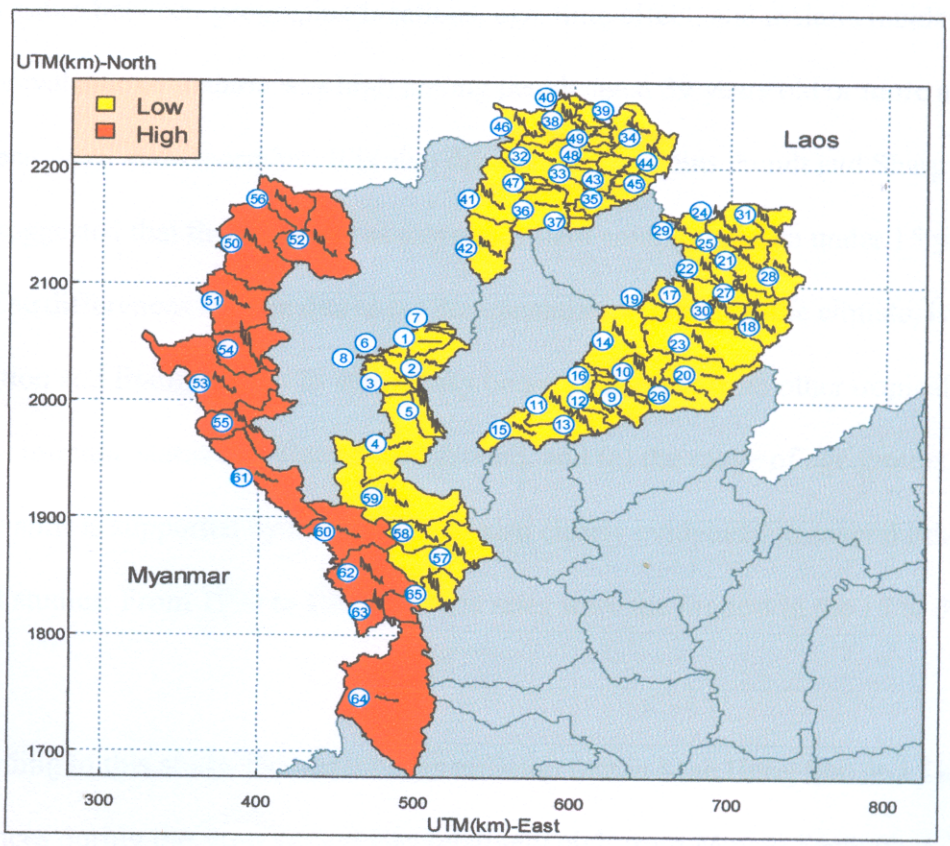


Figure 10: Malaria incidence in North-Western border provinces of Thailand: 1999-2004

3.3 Summary for Malaria in North-Western Thailand

We have shown that malaria is a serious health problem in Mae Hong Son and Tak. The negative binomial generalized linear model provides a appropriate fit to age-

group, districts, and month. The probability plot (Figure 4) 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. It also provides information for malaria prevention. We found that the prevalence of malaria was high among aged 15-39 years. In contrast, Kaewsompak et al (2005) used the negative binomial distribution to model the incidence rate of commonly 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 and 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 years old.

According to this study, the malaria transmission rate in Mae Hong Son and Tak provinces 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 (2000) 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 Hong Son is in Mueang Mae Hong 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 Umphang and Tha Song Yang due to their location. Umphang 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 et al 2004).

In addition, additive log-normal and additive plus multiplicative were fitted incidence rates by age group, district and quarter period in north-western Thailand. Tiensuwan et al (2000) use log-normal liner model to identified risk factors causing malaria in Tak province in the raining season, two-and- three-dimension log-linear model are used to obtain estimated parameter and expected frequencies. Among the models fitted, the best models chosen base on analysis of deviance. Furthermore, the additive plus multiplicative linear model provides an appropriate fit to the malaria incidence rates in the north-western region classified by districts, age-group and quarterly 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 12-14 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 (Faraway 2002), but this is beyond the scope of the present study.

According to north-western Thailand 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. 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. As remarked by Khasnis and Nettleman (2005), 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.

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.

Chapter 4

Conclusions and Limitations

All of the methods in this thesis have involved modeling malaria incidence. Populations at risk and malaria incidence rates are linked in geographical entities that become the units of analysis. These studies have produced a more detailed map of the distribution of malaria in parts of north-western Thailand. They have also given us a clear idea of the factors associated with malaria incidence in these areas, and how malaria incidence rates in different areas have an area has changed over time.

4.1 Research methodology

The objective of this thesis is to investigate the epidemic patterns of hospital-diagnosed malaria incidence by month, district and age-group in north-western Thailand, and to determine precisely where malaria incidence is high. The scope of this study is malaria cases from 1999-2004 in six provinces in north-western Thailand. The study models the pattern of hospital-diagnosed malaria incidence in order to determine where malaria incidence was high. The outcome variables are defined as the incidence rate in a cell indexed by district, age group and period (monthly or quarterly). The determinant variables are defined as category variable as demographic factors: age group, district and period factor (monthly or quarterly). The district factor was referenced to border and non-border districts as intervening variables. The target population was malaria cases in the provinces selected: Lamphun, Phrae, Nan, Chiang Rai, Mae Hong Son, and Tak, all of which have high malaria incidence rates.

This study design is a cross- sectional survey distributed by time and area. 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 or quarter (allowing for a seasonal effect) as categorical determinants. Data collect used in the current study are taken from a registry of hospital-diagnosed infectious disease cases collected routinely 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, 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. Next step, we create malaria disease counts by age group, district and month or quarter,. 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 is little evidence of a gender effect the data for the two sexes were not combined.

First of all, we will try to fit linear regression, Poisson and negative binomial regression models for malaria disease counts by age group, district and month in Mae Hong Son and Tak provinces. Later, we will try to fit additive log-normal models and additive plus multiplicative models for malaria incidence rate by age group, district and quarter in six provinces. We hope to be able to use these models to forecast the districts and age groups where epidemics are likely to occur in the near future. Among the models fitted, the best are chosen based on residual deviance.

4.2 Conclusions

4.2.1 Modeling malaria incidence

In brief, the negative binomial generalized linear model provides a appropriate fit to age group, districts, and month in Tak and Mae Hong son provinces; whereas, the additive plus multiplicative linear model provides an appropriate fit to age-group, districts and quarterly period in six provinces.

Second, the probability plot in study of malaria counts in Mae Hong Son and Tak provinces shown that the distributional assumption of a model verified and an effective way in capturing any unexpected increase in malaria counts. Also it provides information for malaria prevention. In addition, the probability plots of residuals against normal quantiles in six provinces indicate that the negative binomial model does not provide a satisfactory fit but a linear model for log-transformed incidence rates does fit malaria incidence rates in six provinces.

Finally, the estimated standard errors for the parameters within each factor in models are the same; therefore, its do not depend on the number of cases or the population at risk in each cell. The data in six provinces which the models are fitted only comprise incidence rates within each cell, in contrast the negative binomial model, which takes into account both the number of cases and the population at risk in each cell.

4.2.2 Malaria incidence rates

First of all, we found that the prevalence of malaria was high among aged 15 to 39 years old in Mae Hong Son and Tak provinces. According to six provinces 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. This is because in border districts are work place for worker from Myanmar migrant.

Second, the malaria transmission rate in Mae Hong Son and Tak during the study period was relatively high from May to September. Although, malaria trends for most districts in the six provinces showed a consistent decreasing trend, with a seasonal pattern peaking in the April-June quarter.

Finally, the highest malaria incidence rates in Lamphun province is Thung Hua Chang district, Phrae province is Den Chai district, Nan province is Song Khwae district, Chiang Rai province is Mae Fah Luang district, Mae Hong Son province is Mueng Mae Hong Son district and Tak province is Umphang district.

All in all, we present maps of malaria prediction risk by use of additive plus multiplicative regression models to determine approximate risk on a larger scale and we employ geo-statistical approaches to improve prediction at a local level. Malaria prevalence was predicted by means of a model which used district, age group and quarterly period variables as potential predictors. After the analysis, spatial dependence of model local variation in malaria 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. This thesis introduced a new methodology to develop malaria risk categories in the area. A thematic map showed both the districts with high incidence rates and the trend in each district.

4.3 Discussion

According to this study shown that malaria is a serious health problem in Mae Hong Son and Tak provinces which along border Myanmar. The negative binomial generalized linear model provides a suitable fit to age-group, districts, and month. Kaewsompak (2005) used the negative binomial distribution also to model the incidence rate of commonly occurring acute febrile illnesses in subdistricts of Yala province between 2002 and 2003 and found that the relationship between geographic locations, age, period effect and malaria incidence. Furthermore, Ruru and Barrios (2003) used Poisson and Classical Regression to fit malaria incidence. Poisson regression and classical regression are comparable based on the mean absolute percentage error. In short, additive plus multiplicative linear model provides an appropriate fit to the malaria incidence rates in north-western Thailand (Sriwattanapongse and Kuning 2009). Gomez-Elipe et al (2007) showed that a time series of quarterly notifications of malaria cases from local health facilities (rain, temperature records, and the normalized difference vegetation index) model is a simple and useful tool for forecasting malaria incidence in Burundi.

The prevalence of malaria in Mae Hong Son and Tak provinces 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 10 years old in Mali, West Africa. 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.

According to this study, the malaria transmission rate in Mae Hong Son and Tak provinces 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 (2000) 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 Hong Son province is in Mueang Mae Hong Son and Mae La Noi which is explained by these geographical location and significant social economic status. For Tak province is the high malaria incidences are in Umphang and Tha Song Yang due to their location. Umphang is extremely vulnerable to natural deep forest area such as Thung Yai Naresuan.

Malaria in Thailand is forest-related and most prevalent along the international borders, especially on the Thai-Myanmar border, especially on the Thai-Myanmar border where young men working in or near forests are at high risk (Sriwattanapongse et al 2007). This is 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. Taking the Tak Province for example, it is estimated that nearly a third of its population are refugees, migrants, or displaced populations. Because of the large border-crossing population movement, it may not be surprising that Tak is one of the most malaria endemic provinces in Thailand. 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 et al 2004). In my opinion, malaria transmissions may depend on meteorological and environmental factors.

According to study malaria in six provinces, the incidence of malaria was highest among persons aged between 15 and 39 years in border districts; but in contrast, children aged 0-4 had the highest incidence rates among those in non-border districts. This is because in border districts, malaria trends for most districts in the six provinces showed a consistent decreasing trend, with a seasonal pattern peaking in the April-June quarter. As remarked upon by Khasnis and Nettleman (2005), the ability of mankind to adapt is dependent upon the magnitude and speed of change; it also depends on recognizing epidemics early, containing them effectively, providing appropriate treatment, and committing resources toward prevention and research.

Chaveepojnkamjorn and Pichainarong (2004) study malaria infection among the migrant population along the Thai-Myanmar border area. The results showed that the study subjects were predominantly Thai-Yai and Myanmar, and revealed that residence located in the forest increased the risk of malaria infection.

In my view, malaria epidemic in border district due to foreign labor from Myanmar influx to work place in border district. Along the Thai–Myanmar border, it is estimated by the World Health Organization (WHO) Border Health Program (World Health Organization Thailand 2005) that at the end of 2004, there were approximately 120,000 registered refugees living in camps, 400,000 registered migrant workers, and another 500,000 undocumented workers.

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. In my view, geo-statistical are important tools malaria control in order to treatment and prevention.

4.4 Strengths of this study

The results of this study provide a model that can produce useful short-term forecasts of malaria incidence rate in north-western Thailand. There is evidence of much higher malaria incidence rates in districts bordering Myanmar; the incidence rates between border and non-border districts differ substantially. One main advantage of this study is that the model is able to separately consider data from border and non-border districts.

The important result is provided valuable information for planning precautionary measures and in controlling the spread of malaria. It was found that built-up areas have the highest incidence and constitute the highest risk zones. This study found that border district areas have affect on malaria epidemics. The map of malaria prevalence predicted risk in north-western border provinces of Thailand can be used by the Public Health Department as a base map for applying preventive measures to control the malaria outbreak. This will help in focussing the preventive measures being according to priority in high, average and low zones and help in saving time and money.

4.5 Limitations and suggestions for further study

The methods used to estimate malaria incidence from malaria transmission risk provide a limitation guide in forecasting malaria cases in any given area because:

The delimitation of only four categories of age group (0-4, 5-14, 15-39, and 40+) provides a fine model; a particular change in age group category may reveal a different range of malaria incidence.

The model to determine malaria transmission was based on a six provinces. Other provinces such as Chiang Mai, Lampang and Phayao were not included in this study because their data is not sufficient and more missing. Particularly, Chiang Mai is a big city in north-western and many border districts. Ministry Public Health report that it is high risk of malaria such as Chiang Dao district.

The studies used to derive basic incidence rates within categories of endemicity, district and age group were primarily conducted during 1999-2004. If a study were to be conducted for a period longer than six years, the resulting model would be less prone to forecasting error.

Further study, climatic variable was included in model. Bi et al (2003) indicate that climatic variables should be considered as possible predictors for regions with similar geographic and socioeconomic conditions. Climatic variables that predict the presence or absence of malaria are likely to be the best pattern for forecasting the distribution of the disease (Devi and Jauhari 2006). Briët et al (2008) showed that the addition of rainfall as a covariate improved prediction of selected (seasonal) autoregressive integrated moving average (ARIMA) models moderately in some districts, but

worsened prediction in others. Improvement by adding rainfall was more noticeable with longer forecasting horizons.

Further study is also needed in order to fully understand the behavior of the model. Applying it on simulated data derived from known tessellation designs would reveal whether the model is able to correctly capture the different spatial processes. Model comparison between a stationary model and a model which is based on random tessellations will show if parsimony is preferred over complexity. The tessellation-based model was applied in mapping malaria prevalence data in other parts of Thailand (Beil et al 2006). The non-stationary features present in malaria data observed over large areas has never been addressed previously.

Local characteristics such as human activities, meteorological patterns, and land use or malaria interventions can alter spatial correlation in different parts of the region. It is more likely to think of it as a mixture of spatial processes affecting large areas, rather than as a single process. Ignoring non-stationary features may partly explain differences between the various malaria maps produced so far. Proper modeling of the spatial process will lead to more accurate parameter estimation and prediction. A random tessellation approach allows the data to decide on the number of tiles, and thus determines the spatial processes. As long as we have a large amount of data, inferences are driven by the data rather than by prior specifications. Maps of the spatial covariance parameters can be useful for controlled interventions. Regions where the spatial correlation reduces rapidly over short distances may indicate local unmeasured factors which influence malaria risk. Without additional information the model will not be able to find causal explanations. It will only identify the areas which have geographical dependencies over larger or shorter distances.

References

Abellana, R., Ascaso, C., Aponte, J., Saute, F., Nhalungo, D., Nhacolo, A. and Alonso, P. 2008. Spatio-seasonal modeling of the incidence rate of malaria in Mozambique. Malaria Journal, 7: 228.

Asenso-Okyere, W.K., Dzator, J.A and Osei-Akoto, I. 1997 The behaviour towards malaria care – a multinomial logit approach. Social Indicators Research, 39: 167-186.

Beil, M., Eckel, S., Fleischer, F., Schmidt, H., Schmidt, V., Walther, P. 2006 Fitting of random tessellation models to keratin filament networks. Journal of Theoretical Biology, Volume 241(1): 62-72.

Bi, P., Tong, S., Donald, K., Parton, K.A. and Ni, J. 2003. Climatic variables and transmission of malaria: a 12-year data analysis in Shuchen County, China. Public Health Reports, 118(1): 65-71.

Bohra, A. and Andrianasolo, H. 2001. Application of GIS in modeling of dengue risk based on sociocultural date: Case of Jalore, Rajasthan, India. Dengue Bulletin, 25: 92-102.

Booth, H., Maindonald, J. and Smith, L. 2002. Applying Lee-Carter under conditions of variable mortality decline. Population Studies, 56: 325-336.

Briët, O.J.T., Vounatsou, P., Gunawardena, D.M., Galappaththy, G.N.L. and Amerasinghe, P.H. 2008 Models for short term malaria prediction in Sri Lanka Malaria Journal, 7: 76.

Chareonviriyaphap, T., Bangs, M.J. and Ratanatham, S. 2000. Status of malaria in Thailand. Southeast Asian Journal of Tropical Medicine and Public Health, 31: 225-237.

Charles, D.M., Hart, J., Davis, W.A., Sullivan, E., Dowse, G.K. and Davis, T.M.E. 2005. Notifications of imported malaria in Western Australia, 1990-2001: incidence, associated factors and chemoprophylaxis. Medical Journal of Australia, 182: 164-167.

Chaveepojnkamjorn, W. and Pichainarong, N. 2004. Malaria infection among the migrant population along the Thai-Myanmar border area. Southeast Asian Journal of Tropical Medicine and Public Health, 35(1): 48-52.

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): 1293-1301.

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

Devi, N.P. and Jauhari, R.K. 2006. Climatic variables and malaria incidence in Dehradun, Uttaranchal, India, Journal of Vector Borne Diseases, 43: 21-28.

Faraway, J.J. 2002. Practical regression and Anova using R. Boca Raton, FL: Chapman & Hall/CRC.

Fisher, R.A. and Mackenzie, W.A. 1923. The manurial response of different potato varieties. Journal of Agricultural Science, 13: 311-320.

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

Gomez-Elipe, A., Otero, A., van Herp, M. and Aguirre-Jaime, A. 2007. Forecasting malaria incidence based on quarterly case reports and environmental factors in Karuzi, Burundi, 1997-2003. Malaria Journal, 6: 129.

Good, I.J. 1969. Some applications of the singular decomposition of a matrix, Technometrics, 11: 823-831.

Goodman, C., Hanson, K., Mills, A., Wiseman, V. and Worrall E. 2003. The economics of malaria and its control. Paper for the WHO/TDR Scientific Working Group on Malaria; Geneva, Switzerland, 24-27 March, 2003.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M. and Snow, R.W. 2004. The global distribution and population at risk of malaria: past, present, and future. Lancet Infectious Diseases, 4(6): 327-36.

Hoshen, M.B. and Morse, A.P. 2004. A weather-driven model of malaria transmission. Malaria Journal, 3: 32.

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. Hoboken, NJ: John Wiley & Sons.

Khasnis, A.A. and Nettleman, D.M. 2005. Global warming and infectious disease. Archives of Medical Research, 36: 689-696.

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 and International Health, 6(12): 986.

Kleinschmidt, M., Bagayoko, G.P.Y., Clarke, M., Craig, D. and Sueui, L. 2000. A spatial statistical approach to malaria mapping. International Journal Epidemiology, 29: 355-361.

Konchom, S., Singhasivanon, P., Kaewkungwal, J., Chupraphawan, S., Thimasarn, K., Kidson, C., Rojanawatsirivet, C., Yimsamran, S. and Looareesuwan, S. 2003. Trend of malaria incidence in highly endemic provinces along the Thai borders 1991-2001. Southeast Asian Journal of Tropical Medicine and Public Health, 34(3): 486-494.

Lawton, W.H. and Sylvestre, E.A. 1971. Self modeling curve resolution. Technometrics, 13: 617-632.

Liu, G.F., Wang, J., Liu, K. and Snavely, D.B. 2005. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Statistics in Medicine, 25 (8):1275-1286.

McNeil, D.R. and Tukey, T.W. 1973. Higher-order diagnosis of two-way tables, illustrated on two sets of demographic empirical distributions. Biometrics, 31: 487-510.

Molineaux, L. 1988. The epidemiology of human malaria as an explanation of its distribution, including some implications for its control. In: Wernsdorfer, W.H. and McGregor, I. (Eds) Malaria: Principles and practice of malariology (Vol. 2). London: Churchill Livingstone.

Nelson, K.E., Williams, C.F. 2007. Infectious disease epidemiology: theory and practice. (2nd ed). Boston, MA: Jones and Bartlett Publishers.

Prasittisuk, C. 1985. Present status of malaria in Thailand. Southeast Asian Journal of Tropical Medicine and Public Health, 16: 141-5.

R Development Core Team. 2008. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Rattanasiri, S., Böhning, D., Rojanavipart, P. and Athipanyakom, S. 2004. A mixture model application in disease mapping of malaria. Southeast Asian Journal of Tropical Medicine and Public Health, 35: 38-47.

Ruru, Y. and Barrios, E.B. 2003. Poisson Regression Models of Malaria Incidence in Jayapura: Indonesia, The Philippine Statistician, 52(1-4): 27-38.

Snow, R.W. and Marsh, K. 1998. New insights into the epidemiology of malaria relevant for disease control. British Medical Bulletin, 54(2): 293-309.

Snow, R.W., Omumbo, J.A., Lowe, B., Molyneux, C.S., Obiero, J.O., Palmer, A., Weber, M.W., Pinder, M., Nahlen, B., Obonyo, C., Newbold, C., Gupta, S. and Marsh, K. 1997. Relation between severe malaria morbidity in children and level of Plasmodium falciparum transmission in Africa. Lancet, 349: 1650-1654.

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.

Sriwattanapongse, W.,Kuning, M.,Jansakul, N. 2007, Forecasting monthly malaria cases in districts of border provinces in Thailand, 7th National Grad Research Conference: GRAD-RESEARCH , April 4-5,Prince of Songkla University, Surat Thani Campus.

Sriwattanapongse, W., Kuning, M. and Jansakul, N. 2008. Malaria in north-western Thailand. Songklanakarin Journal of Science and Technology, 30(2): 207-214.

Sriwattanapongse W.,Kuning M.,2009, Modeling Malaria Incidence in North-Western Thailand, Chiang Mai Journal of Science, 36(3): 403-410.

Sylvestre, E.A., Lawton, W.H. and Maggio, M.S. 1974. Curve resolution using a postulated chemical reaction. Technometrics, 16: 363-368.

Theil, H. 1983. Linear algebra and matrix methods in econometrics. In: Griliches, Z. and Intriligator, M.D. (Eds). Handbook of Econometrics. Amsterdam: North-Holland Publishing.

Tiensuwan, M. 2000. Modelling recurrent malaria in Thailand. Statistical Methods in Medical Research, 9(3): 249-258.

Tongkumchum, P. and McNeil, D. 2009. Confidence intervals using contrasts for regression model. Songklanakarin Journal of Science and Technology, 31 (2): 151-156.

Venables, W.N. and Ripley, B.D. 2002. Modern applied statistics with S (4th ed). New York: Springer-Verlag.

Wackerly, D.D., Mendenhall, W. and Schaeffer, R.L. 1996. Mathematical statistics with applications (5th ed). New York: Duxbury Press.

Wilmoth, J. R., Vallin, J. and Caselli, G. 1989. Quand certaines generations ont une mortalite differente de celle que I'on Pourrait attendre, Population; 44: 335-376.

World Health Organization Thailand, 2005. Overview of Thai-Myanmar border health situation 2005, 210.

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 States of America, 101(37): 2375-2380.

Vitae

Name: Mrs. Wattanavadee Sriwattanapongse

Student ID: 4848006

Educational Attainment:

Degree	Name of institution	Year of Graduation
B. Sc. (Education)	Prince of Songkla University	1979
M. Sc. (Applied Statistics)	National Institute of Development Administration	1983

Work-Position and Address:

Position: Associate Professor

Address: Department of Statistics, Faculty of Science,
Chiang Mai University, 50200

List of Publications and Proceedings:

Publications: (1) Sriwattanapongse, W., Kuning, M. and Jansakul, N. 2008. Malaria in North-Western Thailand, Songklanakarin Journal of Science and Technology, 30 (2): 207-214.

(2) Sriwattanapongse, W. and Kuning M. 2009. Modeling Malaria Incidence in North-Western Thailand. Chiang Mai Journal of Science, 36(3): 403-410.

Proceedings: Sriwattanapongse, W., Kuning, M. and Jansakul, N. Forecasting monthly malaria cases in districts of border provinces in Thailand. 7th National Grad Research Conference: GRAD-RESEARCH 2007, April 4-5 Prince of Songkla University, Surat Thani Campus.