Original article

FORECASTING THE MONTHLY INCIDENCE RATE OF PNEUMONIA IN MAE HONG SON PROVINCE, THAILAND

Wattanavadee Sriwattanapongse, M.S.,¹ Surin Khanabsakdi, M.S.,¹ Supreeya Wongtra-ngan, M.D.²

¹Department of Statistics, Faculty of Science, ²Department of Obstetrics and Gynecology, Faculty of Medicine, Chiang Mai University

Abstract

The objective of this study was to investigate the epidemic patterns of community-acquired pneumonia (CAP) incidence by month, district and age group for Mae Hong Son, the north-western border province of Thailand. The results may subsequently be applied to geographical epidemiologic research for the spatial-temporal mapping of diseases. This will facilitate data linkage to population demographic statistics as they become available. This study was a cross-sectional hospital-based survey of all cases of the disease, provided by the Provincial Health Office in Mae Hong Son province. The sample comprised all such data collected from 1999 to 2005. The model used was the linear regression time series to forecast the districts and age groups where epidemics are likely to occur in the near future, in order to aid in disease prevention. The model contains additive effects associated with the season of the year, district, age group, and the pneumonia incidence rates in previous months, and it 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.

We found that the largest residual obtained for Mae Hong Son was 2.00 – corresponding to 5 cases reported among the 5- to 14-year-old age group – in Pang Ma Pha district in April, 1999 (an incidence rate of 2.76 per 1,000). The pneumonia transmission rate in Mae Hong Son province during the study period was relatively low from January to June of each year.

Keywords: pneumonia incidence, Mae Hong Son province, linear regression model

According to the WHO, infectious diseases are estimated to account for about one-quarter of deaths worldwide.(¹) Pneumonia contributes to between 750,000 and 1.2 million neonatal deaths and an unknown number of stillbirths each year worldwide.(²)
The etiology depends on time of onset. Gram negative bacilli is predominate in the first week of life, and after that Gram positive bacteria, Streptococcus pneumonia, probably causes about 25% of neonatal pneumonia. Mortality from pneumonia generally decreases with age until late adulthood. Elderly individuals, however, are at particular risk of pneumonia and associated mortality.

The prevalence of pneumonia is high among children under 5 years old. Every year, 1.9 million children under 5 years of age die from pneumonia. There have been several recent studies of pneumonia morbidity and mortality in Thailand. Brady et al. evaluated pneumonia cases reported during 1999-2001 by the Ministry of Public Health surveillance system in Sa Kaeo province near the Cambodian border, and found that pneumonia deaths were underreported compared to data available from death certificates. Suwanjutha et al. focused on risk factors for children under 5 years of age. Reechaipichitkul and Tantiwong studied clinical features of community-acquired pneumonia among patients treated at Srinagarind Hospital in Khon Kaen province in the north-eastern region. People most at risk are those older than 65 or younger than 2 years of age, or those who already have a health problem.

Pneumonia is estimated to be the leading cause of mortality in the world among children less than 5 years of age, with more than 95% of all clinically diagnosed episodes occurring in developing countries. It is generally caused by viruses, bacteria or other infective agents getting into the respiratory tract. Acute pneumonia is one of ten diseases of which morbidity and mortality rates are under surveillance in Thailand. The number of acute pneumonia cases is staggering, with more than 1 million (1,055,393) in 2002. It is the leading cause of illness in Thailand.

In Thailand, all hospital-diagnosed infectious disease cases are routinely recorded by the Ministry of Public Health in each of its 12 administrative zones, and these records include pneumonia. Mae Hong Son is one of six provinces in the upper northern zone. Pneumonia accounted for 10.46% of all disease cases in the province over the seven-year period of 1999-2005. It was thus, the third most common disease reported, after diarrhea (33.96% of cases) and malaria (20.81%), followed by pyrexia of unknown origin (9.16%), and conjunctivitis (8.57%). Among these diseases, pneumonia was by far the most lethal, accounting for 0.38% of all deaths in the region from hospital-diagnosed cases of infectious diseases during the seven-year period. It should be noted, however, that 75.43% of these pneumonia deaths occurred among children less than 5 years of age.

**Objectives**

This study aimed to investigate the epidemic patterns of community-acquired pneumonia (CAP) incidence by month, district and age group in Mae Hong Son province, based on routinely collected data available from provincial health offices. More specifically, the study also aimed to find out whether any district had experienced an unusually high incidence of pneumonia, so that preventive disease-control measures can be put into place.

Mae Hong Son is the north-western border province of Thailand with Myanmar, and has a high rate of pneumonia incidence. This province occupies an area of 12,700 square kilometers, and borders the Myanmar states of Shan, Kayah and Kayin. Mae
Hong Son province consists of seven districts: Mueang Mae Hong Son, Khun Yuam, Pai, Mae Sariang, Mae La Noi, Sop Moei and Pang Mapha. Mae Hong Son is mountainous and thus enjoys a cooler climate than other areas of Thailand. The Salween River forms part of the boundary with Myanmar. The rainy season extends from May to October. The average monthly rainfall during these months, over the 25-year period from 1961 to 1990, varied from 115 mm in October to 252 mm in August.\(^{(10)}\)

Based on individual hospital case records, routinely reported in the province from 1999 to 2005, linear regression models of log-transformed incidence rates were used to assess the effects of age, location and season of the year. Autoregressive terms were included to account for time series and spatial correlations. Given that the monthly disease counts in individual cells, defined by age group and district, are often small numbers with many zero occurrences, linear regression models were used. These statistical models can identify cells with an unexpectedly high occurrence of disease. Where substantial autocorrelation exists in the time series, such episodes might thus enable public health authorities to establish strategies for preventing outbreaks before they occur.

**Research methodology**

**Study design**

Data used in this study were taken from a registry of hospital-diagnosed infectious disease cases collected routinely in Mae Hong Son province by the Ministry of Public Health. For each year after 1998, these data were available in computer files. There was a record for each case, with fields comprising characteristics of the subject and the disease, including dates of sickness and disease diagnosis, the subject’s age, gender, and address, and severity of the illness (including date of death for mortality cases).

After cleaning to correct or impute data-entry errors, the records for the province were stored in a Structure Query Language (SQL) database. SQL programs were used to create pneumonia disease counts by month (84 months from January, 1999, to December, 2005), age group (0-4, 5-14, 15-39 and 40+ years), and district. Incidence rates were computed as the number of cases per 1,000 residents in the district, according to the 2000 Population and Housing Census of Thailand. Since there was little evidence of a gender effect, data for the two sexes were combined.

**The linear regression model**

This method described data structures with multi-categorical determinants (i.e., a multi-categorical determinant regression model). If variables were denoted by \(\chi_1, \chi_2, \ldots, \chi_c\), the extended model for Mae Hong Son used the form:

\[
y = a + \sum_{j=1}^{c} b_j \chi_j
\]

(1)

The constants \(b_1, b_2, \ldots, b_c\) were called regression coefficients. These measured the associations between the determinants and the outcome variable.\(^{(11)}\) For this paper, the outcome was the incidence rate, as the determinants were district, age group, season and lagged time.

The simplest model was based on linear regression, with the outcome variable defined as the incidence rate in a cell indexed by district, age group, and calendar month (allowing for a seasonal effect) as categorical determinants. Such incidence rates generally
had positively skewed distributions, so it was conventional to transform them by taking logarithms. Since monthly disease counts based on small regions were often zero, it was necessary to make some adjustments to avoid taking logarithms of zero.

The method used to define the outcome was:

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

(2)

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 were included as additional determinants. Such an observation-driven model with \( m \) lags could take the form:\(^{(12-13)}\)

\[ Y_{ijt} = \mu + \alpha_i + \beta_j + \eta_s + \sum_{k=1}^{m} Y_{ij,t-k} + \epsilon_{ijt} \]  

(3)

where \( Y_{ijt} \) is a random variable denoting the incidence of pneumonia disease in age group \( i \), district \( j \) and month \( t \) for the region of interest, and \( \eta_s \) is the corresponding number observed. \( Y_{ijt} \) is the outcome variable specified in Equation (2), and \( \epsilon_{ijt} \) comprises a set of independent, normally distributed random variables with a mean of 0, and \( s = \text{mod} (t, 12) \). In this model, we constrained the parameters so that \( \alpha_i = 0, \beta_j = 0 \) and \( \eta_s = 0 \). While linear time trends could be included in the model, they were less useful for short-term forecasting purposes in the presence of high serial correlations, and were not considered in this study.

To allow for possible spatial correlations between observations in different districts at the same time, and also correlations between different age groups, additional terms allowing for these effects might be included as determinants in the model. A simple extension \( Y_{ijt} = \mu + \alpha_i + \beta_j + \eta_s + \sum_{k=1}^{m} Y_{ij,t-k} + \delta_{i}Y_{ijt-1} + \delta_{j}Y_{ijt-1} + \epsilon_{ijt} \) takes the form:

\[ y_{ij}^{(a)} \quad y_{ij}^{(b)} \]  

(4)

where \( y_{ij}^{(a)} \) and \( y_{ij}^{(b)} \) denote the observed (transformed) incidence rates in all age groups other than \( i \), and in all districts other than \( j \), respectively.

Results

The total number of pneumonia cases reported by hospitals in Mae Hong Son province during the seven-year period of 1999-2005 was 16,185 cases. The number of outpatient cases, which did not admit to hospital (walking pneumonia), was 9,048 (55.91%). The pneumonia transmission rate in this group was relatively high from August to December of each year. The number of cases per month for a particular age group and district varied from 0 to 129 cases, and the corresponding maximum disease rate was 5.89 cases per 1,000 per month. [Fig. 1. a,b]

In the age group below 5 years old (Fig. 1a), the incidence rates show a marked seasonal periodicity, which decreased from very high levels (12,208 cases) in 1999 to around 1 case per 1,000 per month in recent years (Jan 2000=1 ). However, in other age groups (Fig. 1b) the reported pneumonia rates showed a less pronounced seasonal pattern, and decreased relatively slightly over
the same period. Incidence rate patterns were similar for all of these (5 years and older) age groups.

Figure 2 shows how the average pneumonia incidence rates varied by district in Mae Hong Son province. The lowest rates occurred in Sop Moei district (6.02 cases/1,000/year), while the highest was in Pang Mapha district (15.04 cases/1,000/year).

The results in Table 1 were obtained by fitting the linear regression model given by Equation (4) to the log-transformed incidence rates for Mae Hong Son province. In each case we took $K = 10,000$, giving reason-
ably linear, normal score plots. The number of zero counts was 497 (21.13%) for Mueang district, Mae Hong Son, where the disease incidence rates were much lower. While all components in the model for Mae Hong Son province were statistically significant, the lagged incidence rates accounted for the largest single contribution to the $R^2$ statistic (80.82% for Mae Hong Son Province); and the coefficients incorporating further correlations between age groups and districts were also quite substantial.

The largest residual obtained for Mae Hong Son province was 2.00, corresponding to 5 cases reported among the 5- to 14-year-old age group in Pang Ma Pha district in April, 1999 (an incidence rate of 2.76 per 1,000). However, this residual does not appear as an outlier on the normal score plot, and the number of cases and age groups reported in the same district in the following five months (May, Jun., Jul., Aug. and Sep., 1999) was small (1, 2, 0, 0 and 0, respectively).

**Discussion**

According to this study, the pneumonia transmission rate in Mae Hong Son province was relatively high during the study period. The majority of deaths were among residents in Mueang district of Mae Hong Son province (54.8%). The death rate was highest during the months of July through December. Much of this period encompasses the height of the “rainy season”. Although rainfall does not appear to have a direct effect on the relative incidence of pneumonia, the data show a consistent reduction in cases beginning in January and continuing through the
Table 1. The results of fitting a linear model for Mae Hong Son province

<table>
<thead>
<tr>
<th>Determinant</th>
<th>Coefficient</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.1315</td>
<td>0.2404</td>
</tr>
<tr>
<td>Age group: 0-4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5-14</td>
<td>-1.7626</td>
<td>0.1705</td>
</tr>
<tr>
<td>15-39</td>
<td>-2.2008</td>
<td>0.2128</td>
</tr>
<tr>
<td>40+</td>
<td>-1.6915</td>
<td>0.1621</td>
</tr>
<tr>
<td>District: 1 Mueang Mae Hong Son</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2 Khun Yuam</td>
<td>-0.2228</td>
<td>0.0572</td>
</tr>
<tr>
<td>3 Pai</td>
<td>0.1917</td>
<td>0.0571</td>
</tr>
<tr>
<td>4 Mae Sariang</td>
<td>-0.1927</td>
<td>0.0571</td>
</tr>
<tr>
<td>5 Mae La Noi</td>
<td>-0.1843</td>
<td>0.0561</td>
</tr>
<tr>
<td>6 Sop Moei</td>
<td>-0.4108</td>
<td>0.0636</td>
</tr>
<tr>
<td>7 Pang Mapha</td>
<td>-0.1387</td>
<td>0.0558</td>
</tr>
<tr>
<td>Month: January</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>February</td>
<td>0.7698</td>
<td>0.0749</td>
</tr>
<tr>
<td>March</td>
<td>0.3582</td>
<td>0.0762</td>
</tr>
<tr>
<td>April</td>
<td>0.6409</td>
<td>0.0747</td>
</tr>
<tr>
<td>May</td>
<td>0.3774</td>
<td>0.0721</td>
</tr>
<tr>
<td>June</td>
<td>0.2069</td>
<td>0.0709</td>
</tr>
<tr>
<td>July</td>
<td>0.3401</td>
<td>0.0706</td>
</tr>
<tr>
<td>August</td>
<td>0.4103</td>
<td>0.0707</td>
</tr>
<tr>
<td>September</td>
<td>0.5519</td>
<td>0.0711</td>
</tr>
<tr>
<td>October</td>
<td>0.4843</td>
<td>0.0722</td>
</tr>
<tr>
<td>November</td>
<td>0.3838</td>
<td>0.0719</td>
</tr>
<tr>
<td>December</td>
<td>0.2244</td>
<td>0.0712</td>
</tr>
<tr>
<td>Autoregressive lag: 1</td>
<td>0.2636</td>
<td>0.022</td>
</tr>
<tr>
<td>2</td>
<td>0.1037</td>
<td>0.0218</td>
</tr>
<tr>
<td>3</td>
<td>0.0752</td>
<td>0.021</td>
</tr>
<tr>
<td>Other age groups</td>
<td>-0.0367</td>
<td>0.0491</td>
</tr>
<tr>
<td>Other distric</td>
<td>0.0787</td>
<td>0.0242</td>
</tr>
</tbody>
</table>

R-squared statistic 0.8082

months of May and June, when the rainy season generally begins. Seasonal variation was evident for various pneumonia categories.\(^{(14)}\)

To gain a better understanding of the risk factors associated with pneumonia, further studies should be undertaken to examine the relationships between climate or seasonal variations in temperature and the occurrence of pneumonia. Bacteria thrive in higher temperatures, while viruses spread more easily during winter. There were positive
associations between reported cases of pneumonia and high temperatures or extremes of rainfall. These results are consistent with previous research, and suggest that global climate change is likely to exacerbate pneumonia illnesses in many Pacific island countries.\(^{(15)}\) This may be associated with the high risk of pneumonia in a study by Simonsen,\(^{(16)}\) who reported that pneumonia continues to be a major cause of morbidity and mortality among children in developing countries.

We have shown that pneumonia is a serious problem in Mae Hong Son province. Consequently, we recommend that a study be done to determine if the data fit a Poisson distribution,\(^{(17)}\) as well as a negative binomial generalized linear model, which would provide an appropriate fit for age group, district and month. Barros and Hirakata\(^{(18)}\) report that Cox or Poisson regression, with robust variance and log-binomial regression, provide correct estimates, and are a better alternative for the analysis of cross-sectional studies with binary outcomes rather than logistic regression. This is because a prevalence ratio is more interpretable and easier to communicate to non-specialists than an odds ratio. Kaewsompak et al.\(^{(19)}\) used negative binomial distribution to model the incidence rate of commonly occurring acute febrile illnesses in subdistricts of Yala province in 2002 and 2003, and found a relationship among geographic location, age, time effect, and pneumonia incidence. Keola et al.\(^{(20)}\) showed that the correlation coefficient of pneumonia was found to be the highest, 0.94, and considered a widespread disease in terms of spatial diffusion, especially in urban areas with a high tendency for population growth.

Mortality from pneumonia is most likely underreported in the passive surveillance system, since some patients may have died at home. For those who died in the hospital, the health personnel might not have sent a follow-up report to change the patient’s status. In other studies, the frequency of death among pneumonia patients ranged between 5% among adults in Spain,\(^{(21)}\) to 9% in hospitalized adults in Ohio,\(^{(22)}\) to 21% in hospitalized adults in Khon Kaen, Thailand.\(^{(23)}\) It is important to accurately measure and report deaths, since they reflect the severity of the disease and can greatly influence health policy decisions.\(^{(24)}\)

The opportunities for treating and preventing pneumonia have multiplied in recent years, outpacing the capacity of many existing surveillance systems to provide the information needed for evidenced-based public health decisions. National public health decisions will require accurate information on the disease burden and costs associated with pneumonia. It is not clear whether current pneumonia surveillance in Thailand is suitable for providing adequate information in order to make such decisions. Pneumonia incidence rates were different in each area because of (a) natural deep forest area and (b) climate variables including rainfall, humidity and temperature. It would be advantageous to gain more factual understanding associated with the risk of pneumonia.

Acknowledgements

We would like to express our gratitude to the Mae Hong Son Provincial Public Health Office for allowing us to use the disease data from their office.
References


การสร้างตัวแบบและการพยากรณ์อุบัติการณ์การเกิดโรคปอดบวมในจังหวัดแม่ฮ่องสอนโดยใช้ตัวแบบอนุกรมเวลาเชิงเส้น และการประยุกต์ใช้กับระบบสารสนเทศภูมิศาสตร์ (GIS)

วัฒนาวดี ศรีวัฒนพงศ์, พ.บ.ม.,¹ สุวัฒน์ ขนานพิกุล, พ.บ.ม.,¹ สุปรียา วงษ์ตะกร้อ, พ.บ.²

¹ภาควิชาสถิติ, คณะวิทยาศาสตร์, มหาวิทยาลัยเชียงใหม่
²ภาควิชาสุตติศาสตร์และนรีเวชวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่

บทคัดย่อ
วัตถุประสงค์ เพื่อศึกษารูปแบบการแพร่กระจายของอุบัติการณ์การเกิดโรคปอดบวมตามเดือน อายุ และกลุ่มอายุของผู้ป่วยที่เกิดขึ้นในจังหวัดแม่ฮ่องสอน ผลการศึกษานำไปประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ ตรวจสอบอัตราการระบาดของโรค การศึกษานี้เป็นการศึกษาแบบตัวแปรพักพักของข้อมูลผู้ป่วยที่ได้จากสำนักงานสาธารณสุขจังหวัดแม่ฮ่องสอน ระหว่างปี พ.ศ. 2542 ถึง พ.ศ. 2548 โดยการใช้ตัวแบบอนุกรมเวลาเชิงเส้น เพื่อพยากรณ์การระบาดของการเกิดโรคปอดบวมในอนาคตตามเดือน อายุ และกลุ่มอายุ และเพื่อป้องกันการเกิดโรค ตัวแบบเป็นปีจิจฉาๆที่เกี่ยวข้องกับฤดูกาลของปี อายุกลุ่มอายุ และอัตราการเกิดอุบัติการณ์โรคปอดบวมในเดือนต่ออายุที่นั้น ตัวแบบที่ได้นำไปใช้ในการพยากรณ์อุบัติการณ์การเกิดโรคปอดบวมในแต่ละเดือน จากการศึกษาพบว่า residual ที่ค่ามากที่สุดคือ 2.00 สอดคล้องกับผู้ป่วยอายุ 5-14 ปี จำนวน 5 รายในเดือนเมษายน ปี พ.ศ. 2542 (อุบัติการณ์โรคปอดบวมที่กับ 2.76 ต่อประชากร 1,000 คน) อีกทั้งการคัดค้านโรคปอดบวมในเดือนต่ออายุที่มีอุปทัศน์จากเดือนมกราคมถึงมิถุนายนของแต่ละปี เช่นไทยเพาะราษฎร 2552;48(3):85-94.

คำสำคัญ: อุบัติการณ์การเกิดโรคปอดบวม จังหวัดแม่ฮ่องสอน ตัวแบบการถดถอยเชิงเส้น