TEMORAL PATTERNS AND A DISEASE FORECASTING MODEL OF DENGUE HEMORRHAGIC FEVER IN JAKARTA BASED ON 10 YEARS OF SURVEILLANCE DATA

Monika S Sitepu1,2, Jaranit Kaewkungwal2,5, Nathanej Luplerdlop2, Ngamphol Soonthornworasiri2, Tassaneel Silawan3, Supawadee Poungsombat4 and Saranath Lawpoolsri2,5

1Directorate General of Health Care, Ministry of Health, Jakarta, Indonesia; 2Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok; 3Faculty of Public Health, Mahidol University, Bangkok; 4Bureau of Vector-Borne Disease, Ministry of Public Health, Nonthaburi; 5Center for Emerging and Neglected Infectious Diseases, Mahidol University, Thailand

Abstract. This study aimed to describe the temporal patterns of dengue transmission in Jakarta from 2001 to 2010, using data from the national surveillance system. The Box-Jenkins forecasting technique was used to develop a seasonal autoregressive integrated moving average (SARIMA) model for the study period and subsequently applied to forecast DHF incidence in 2011 in Jakarta Utara, Jakarta Pusat, Jakarta Barat, and the municipalities of Jakarta Province. Dengue incidence in 2011, based on the forecasting model was predicted to increase from the previous year.

Keywords: dengue hemorrhagic fever, time series analysis, seasonal autoregressive integrated moving average, forecasting

INTRODUCTION

Dengue infection is a major public health problem worldwide with a rapidly increased incidence and spread during the past five decades (WHO, 2009). Tropical countries are most afflicted by the transmission of this virus which threatens over 2.5 billion people who live in the tropics (Gubler, 1998). Fifty to 100 million dengue infections are reported annually from 100 countries worldwide (WHO, 2011). Two hundred thousand to five hundred thousand cases of dengue hemorrhagic fever (DHF) are reported annually worldwide with 24,000 deaths occurring (Monath, 1994; Gubler, 1998; Gibbons and Vaughn, 2002). DHF has been reported to be the leading cause of hospitalizations and death among children in several countries in Southeast Asia, including Indonesia (WHO SEARO, 1999).

The first reported outbreak of dengue fever in Indonesia occurred in Jakarta and Surabaya in 1968. Since 2005, Indonesia has reported the highest number
of dengue cases in Southeast Asia (WHO SEARO, 2006). In 2009, the Ministry of Health of Indonesia reported that the number of DHF cases reached 156,052 in 382 districts. A National Health Survey in 2007 found DHF was the fourth leading cause of death and second leading cause of hospitalization among toddlers in Indonesia (MOH, 2010).

Jakarta Province has the highest incidence of DHF among all the provinces in Indonesia (MOH, 2010). All municipalities in this province are endemic for dengue transmission with all serotypes circulating (Suwandono et al, 2006). Moreover, the Jakarta Provincial Health Office reported that DHF has the highest IR (200.8/100,000 population) of all communicable diseases in 2010.

Prevention and control measures for DF/DHF in Jakarta are based on national policies, which include improvement of the surveillance system, disease management and community participation (Kusristuti and Sutomo, 2005). Understanding patterns of disease transmission can significantly improve outbreak control (Allard, 1998).

Time series analysis has been used successfully in various studies of communicable diseases to determine temporal patterns and forecast disease occurrence (Helfenstein, 1986; Luz et al, 2008; Silawan et al, 2008; Wangdi et al, 2010). This technique is considered a useful tool for understanding the structure of data, forecasting, monitoring and providing feedback regarding control activity (Box et al, 1994). Time series analysis can be practically applied to routinely collected data (longitudinal data) in disease surveillance systems. The objective of this study was to demonstrate the use of time series analysis of routine dengue surveillance data to understand and forecast patterns of dengue occurrence in Jakarta, Indonesia.

**MATERIALS AND METHODS**

**Study area**

The Jakarta Province is located at 6°12’S and 106°48’E. It covers 662.33 km² of land area and 6,977.5 km² of sea (Governor Decree, 2007). This province is divided into five municipalities (Jakarta Barat, Jakarta Timur, Jakarta Pusat, Jakarta Selatan, and Jakarta Utara), occupying lowland areas, and a regency (Kepulauan Seribu) consists of 110 small islands in the Sea of Java (Governor Decree, 2007). In the 2010 census, the total population of Jakarta was 9,607,787 with a density of 14,476 people/km² (BPS-Statistics Indonesia, 2010). Being a tropical area, the climate in Jakarta tends to be hot and humid throughout the year. However, this province has distinct wet and dry seasons. Rainfall increases during the wet season from December to March and decreases significantly during the dry season from June to September.

Since dengue transmission in the small islands (Kepulauan Seribu Regency) is reported to be low (Jakarta Provincial Health Office, 2010, Unpublished data), this study was conducted in the five inland municipalities: Jakarta Barat, Jakarta Timur, Jakarta Pusat, Jakarta Selatan, and Jakarta Utara (Fig 1). All municipalities are urban and endemic for dengue (Jakarta Provincial Health Office, 2010, Unpublished data).

**Data collection**

Dengue is a notifiable disease in Indonesia, hence, all dengue cases at health centers, hospitals, clinics, and private physician’s offices are required to report immediately to the Jakarta Province Health
Active surveillance is periodically performed as a part of the dengue surveillance system. However, due to limited resources, epidemiological investigations can not be done for all reported cases. Hence, data obtained from passive surveillance comprises the majority of data.

This study used DHF and DSS passive surveillance data collected between January 2001 and December 2010 from the Jakarta Province Health Office. The case definition for DHF/DSS was based on WHO criteria for DHF/DSS, which includes fever, hemorrhagic tendencies, thrombocytopenia, and evidence of plasma leakage due to increased vascular permeability (WHO, 1997). In this study, the term DHF refers to both DHF and DSS.

Monthly data sets from 2001 to 2010 for each municipality and all municipalities were used. The ten-year observations were divided into two segments: the first was the estimation segment, where the monthly incidence data from 2001 to 2008 were analyzed to determine the best time-series model and the second was the 2009 to 2010 data served as the validation segment to confirm the time series model.

Data analysis
The monthly incidence of DHF over the 10-year period was plotted to observe the temporal patterns of DHF transmission in Jakarta. A seasonal boxplot was created for each municipality to identify the seasonal pattern and its outliers for
DHF transmission during the 10-year period. Trend analysis using Poisson regression was performed to determine the annual trend for DHF incidence in five municipalities and for overall Jakarta, when seasonal effects were not taken into account.

Time series analysis was chosen based on the assumption the incidence of infectious diseases is related to the previous incidence and population at risk (Helfenstein, 1986). Using the data of monthly DHF incidence, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model was chosen to fit the data. Stationarity of the data were initially assessed. A Box-Cox transformation procedure was used to stabilize the variance, while differencing and seasonal differencing were used to stabilize the mean and remove seasonal autocorrelation, respectively. The SARIMA \((p,d,q,s)(P,D,Q)_s\) model was generated using the Box-Jenkins approach.

The autocorrelation function (ACF) and the partial autocorrelation function (PACF) were analyzed to identify the order of SARIMA \((p,d,q,s)(P,D,Q)_s\) model, where: (1) \(p\) and \(P\) are the order for autoregressive (AR) and seasonal autoregressive, respectively; (2) \(d\) and \(D\) are the order for differences and seasonal differences, respectively; (3) \(q\) and \(Q\) are the order for moving average (MA) and seasonal moving average, respectively, and (4) \(s\) is the length of the seasonal period.

Tentative models were then estimated using the maximum likelihood method. The significance of the estimated parameters was derived and considered if the \(p\)-value was less than 0.05. The Ljung-Box was performed to determine the adequacy of the tentative model. An adequate model was chosen if the residuals of ACF were statistically equal to zero or white noise. Akaike's Information Criterion (AIC) and Schwartz's Bayesian Information Criterion (BIC) were performed. The model with the lowest BIC and AIC was chosen as the best-fitting model.

An out-of-sample forecast was used to forecast the incidence from January 2009 to December 2010. The accuracy of forecasting was determined by comparing the actual with the forecasted DHF incidence using the Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE).

The data used in this study was obtained from the Jakarta Health Office. All data were collected anonymously. The study received ethical approval from the Committee on Health Research Ethics, the National Institute of Health Research and Development, the Ministry of Health, Republic of Indonesia and the Ethics Committee, Faculty of Tropical Medicine, Mahidol University, Thailand.

**RESULTS**

**Temporal distribution**

The overall incidence of DHF suggests high endemicity in all municipalities of Jakarta Province. Table 1 summarizes the annual incidence for each endemic

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jakarta Utara</td>
<td>253</td>
<td>46</td>
<td>379</td>
</tr>
<tr>
<td>Jakarta Pusat</td>
<td>351</td>
<td>88</td>
<td>433</td>
</tr>
<tr>
<td>Jakarta Barat</td>
<td>171</td>
<td>58</td>
<td>224</td>
</tr>
<tr>
<td>Jakarta Selatan</td>
<td>293</td>
<td>70</td>
<td>450</td>
</tr>
<tr>
<td>Jakarta Timur</td>
<td>310</td>
<td>78</td>
<td>398</td>
</tr>
</tbody>
</table>
Table 2
Incidence rate ratio of annual DHF incidence and percent increase in the annual DHF incidence over time (2001 to 2010), by municipality and overall Jakarta Province.

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Incidence rate ratio</th>
<th>95% confidence interval</th>
<th>Percent increase in the incidence (per year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jakarta Utara</td>
<td>1.17</td>
<td>1.15 - 1.18</td>
<td>17</td>
</tr>
<tr>
<td>Jakarta Pusat</td>
<td>1.13</td>
<td>1.11 - 1.14</td>
<td>13</td>
</tr>
<tr>
<td>Jakarta Barat</td>
<td>1.10</td>
<td>1.08 - 1.12</td>
<td>10</td>
</tr>
<tr>
<td>Jakarta Selatan</td>
<td>1.13</td>
<td>1.12 - 1.15</td>
<td>13</td>
</tr>
<tr>
<td>Jakarta Timur</td>
<td>1.12</td>
<td>1.10 - 1.13</td>
<td>12</td>
</tr>
<tr>
<td>Jakarta Province</td>
<td>1.13</td>
<td>1.11 - 1.14</td>
<td>13</td>
</tr>
</tbody>
</table>

Jakarta Pusat and Jakarta Timur both had a median incidence of more than 300 per 100,000 during this period, while the median incidence was lowest in Jakarta Barat (171 per 100,000 population).

The trend of the annual DHF incidence was increased significantly during the study period. Trend analysis shows the overall incidence of DHF in Jakarta increased by approximately 13% yearly. The greatest increase was seen in Jakarta Utara with a 17% increase in incidence per year (IRR 1.17) (Table 2).

The monthly incidence of DHF had a similar seasonal pattern for each of the municipalities (Fig 2). DHF cases were found year long with varying intensity from month to month. The incidence of DHF usually increased in January with a peak around March or April, and then a gradual decrease until the end of the year. The epidemic cycle was unclear during the 10-year period. The highest peak was observed in 2004, suggesting an epidemic, after which the seasonal wave widened during the past five years with a slight increase in baseline incidence. In 2005 and 2006, a high DHF incidence was observed throughout the year, almost free of seasonality.

The presence of seasonal patterns is illustrated on Fig 3. During the past ten years, the transmission tended to start in December. A high monthly incidence was usually seen between January and June, when peak incidences were observed from March to April with a gradual decrease during the following months. The incidence during the high season months
Fig 3–Seasonal boxplot for monthly incidence in five endemic municipalities and Jakarta Province. Dot represents outliers of incidence for particular years.
was widely distributed compared to low season months, which indicates more inter-annual fluctuations in transmission during the high season months.

Outliers of DHF were seen in all study areas in February and March 2004 indicating the outbreaks occurred in 2004.

Jakarta Selatan had outliers in March 2007 without upper whiskers, suggesting the outbreak was characterized by a rapid increase in incidence in March compared to the same month in other years. Outliers were also seen during the low season in most study areas and in the overall Jakarta Province, except in Jakarta Utara. These outliers were observed between September and December in both 2005 and 2010, which suggests sustained high transmission during both those years.

**Time series analysis**

The best fitting model was obtained after estimation, diagnostic checking, and model selection for the individual and overall municipalities. The structure of the best model for each municipality and overall for Jakarta Province is shown in Table 3.

According to the model selection, most of the selected Box-Jenkins models are SARIMA $(1,0,1)(0,1,1)_{12}$ except for Jakarta Selatan and Jakarta Timur which are SARIMA $(0,0,2)(0,1,1)_{12}$ and SARIMA $(2,0,0)(0,1,1)_{12}$ respectively. The most commonly identified model,
Temporal Patterns of Dengue Transmission in Jakarta

Forecasting

The best models were fitted to a validation segment for the period of January 2009 to December 2010. Fig 4 shows the actual and predicted monthly incidences in the five municipalities and for Jakarta Province. The actual line was close to the predicted line and followed the actual pattern with a $R^2$ from 70.97% to 75.62%, which suggests the model may be used for disease forecasting.

The forecasting accuracy is given by the Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). The MAPE for each municipality and Jakarta Province ranged from 18.33% to 28.99% and the MAE ranged from 3.49 to 7.12. The MAPE and MAE suggest the fitted model is appropriate for forecasting.

DISCUSSION

In addition to data collection, data
analysis to describe and predict disease is an important component of the surveillance system. Longitudinal data is useful for understanding the temporal pattern of disease transmission. DHF in Jakarta fluctuated but remained at a high level over the past decade. The seasonal pattern of DHF incidence was observed over the past decade. Even though DHF cases were seen all year long, transmission increased significantly during the rainy season and reached a peak during March and April. An increase in rainfall led to an increase in DF/DHF incidence in one to two months later. This is probably due to the increasing in breeding sites during the rainy season (Arcari et al., 2007). A cyclical pattern was not clearly seen, particularly after 2006, which is consistent with the country-wide transmission pattern (Setiati et al., 2006). Climatic and socio-ecological conditions in Jakarta may have contributed to the persistently high transmission in this province. Trend analysis showed DHF incidence in Jakarta Province increased by about 13% per year. Without additional control efforts, DHF incidence may further increase at a similar rate. High temperature and humidity throughout the year makes virus transmission more efficient by increasing the lifespan of adult mosquitoes, increasing biting activity and shortening the extrinsic incubation and gonadotrophic periods (Halstead, 2007). Socio-economic changes, such as population growth, unplanned urbanization and

<table>
<thead>
<tr>
<th>Municipality</th>
<th>SARIMA model&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Parameter estimates&lt;sup&gt;b&lt;/sup&gt; ± standard errors</th>
<th>p-value&lt;sup&gt;c&lt;/sup&gt; (Ljung-Box test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jakarta Utara</td>
<td>(1,0,1)(0,1,1)&lt;sub&gt;12&lt;/sub&gt;</td>
<td>AR(1) : 0.703 ± 0.055</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA(1) : 0.418 ± 0.119</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMA(1) : -0.712 ± 0.172</td>
<td></td>
</tr>
<tr>
<td>Jakarta Pusat</td>
<td>(1,0,1)(0,1,1)&lt;sub&gt;12&lt;/sub&gt;</td>
<td>AR(1) : 0.593 ± 0.066</td>
<td>0.688</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA(1) : 0.470 ± 0.096</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMA(1) : -0.752 ± 0.197</td>
<td></td>
</tr>
<tr>
<td>Jakarta Barat</td>
<td>(1,0,1)(0,1,1)&lt;sub&gt;12&lt;/sub&gt;</td>
<td>AR(1) : 0.737 ± 0.068</td>
<td>0.819</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA(1) : 0.316 ± 0.117</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMA(1) : -0.775 ± 0.225</td>
<td></td>
</tr>
<tr>
<td>Jakarta Selatan</td>
<td>(0,0,2)(0,1,1)&lt;sub&gt;12&lt;/sub&gt;</td>
<td>MA(1) : 1.156 ± 0.110</td>
<td>0.451</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA(2) : 0.530 ± 0.126</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMA(1) : -0.576 ± 0.149</td>
<td></td>
</tr>
<tr>
<td>Jakarta Timur</td>
<td>(2,0,0)(0,1,1)&lt;sub&gt;12&lt;/sub&gt;</td>
<td>AR(1) : 1.074 ± 0.149</td>
<td>0.857</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AR(2) : -0.347 ± 0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMA(1) : -0.827 ± 0.226</td>
<td></td>
</tr>
<tr>
<td>Jakarta Province</td>
<td>(1,0,1)(0,1,1)&lt;sub&gt;12&lt;/sub&gt;</td>
<td>AR(1) : 0.664 ± 0.062</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA(1) : 0.534 ± 0.082</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMA(1) : -0.813 ± 0.257</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Seasonal ARIMA model fitted to the square root monthly incidence.

<sup>b</sup>Parameter estimates, p < 0.05.

<sup>c</sup>Seasonal ARIMA model fitted to natural log monthly incidence.
modern transportation in Jakarta may play a role in the persistence of dengue transmission (Gubler, 2002). Circulation of multiple dengue virus serotypes in Jakarta may have contributed to the increase pre-existing immunity in the susceptible host, which is a risk factor for increasing the incidence of DHF (Gubler, 1997; Guzman and Kouri, 2002).

There is a growing interest in integrating the disease-forecasting method into the surveillance system. Studies have shown the SARIMA model can produce a reliable model to forecast DHF incidence (Choudhury et al., 2008; Luz et al., 2008; Silawan et al., 2008; Gharbi et al., 2011; Martínez and Silva, 2011). In our study, the best fitting SARIMA models for each endemic municipality and Jakarta Province as a whole were parsimonious. The model for a non-seasonal structure (p,d,q) varied by municipality suggesting specific patterns exist in each municipality, defining which components influence the occurrence of dengue. The same order for seasonal components (P,D,Q) was seen in all municipalities and in overall Jakarta Province in the form of SARIMA (0,1,1)_{12}. This Seasonal Random Trend (SRT) indicates the dengue cases may be determined by the trend the previous year. The adjacent municipalities of Jakarta Utara, Jakarta Pusat, and Jakarta Barat had similar statistical structures for both seasonal and non-seasonal components, which may be in part due to geographically linked similarities, such as environmental and socio-economic factors.

However, some limitations should be considered when using data from routine passive surveillance for disease forecasting. Even though DHF cases in Indonesia are diagnosed using WHO guidelines, the data may include DF cases due to various diagnostic criteria applied by physicians at different health centers and hospitals. This may lead to misclassification bias, where non-DHF cases are included in the DHF database. However, the number of DF cases is expected to be relatively small compared to the number of DHF cases, and should not significantly skew the data or change disease patterns.

In this study, we were only interested in predicting DHF incidence, not DF incidence, because DHF is a major cause of morbidity and mortality in this country. Forecasting DHF cases may help allocate appropriate control activities in a timely manner. However, to understand overall disease transmission, both DHF and DF cases should be considered, because DF cases can play a role in dengue transmission in the community (Endy et al., 2002). Besides clinical diagnosis, laboratory confirmation and serotype identification are necessary for inclusion in the surveillance system to enhance the use of surveillance data, to understand the patterns of disease transmission and to develop an early warning system.

Finally, this forecasting model was based on DHF incidence over the years with the assumption that all other conditions, such as meteorological factors, DHF prevention and control programs, and socio-ecological factors remain constant. Hence, results from the forecasting should be carefully considered, especially when these conditions vary. Since a forecast in this study was based on univariate analysis, forecasting is less accurate during epidemic years. Several studies have been conducted highlighting the various climatic factors associated with dengue transmission during epidemic years in Indonesia (Corwin et al., 2001; Bangs et al., 2006; Arcari et al., 2007). Incorporating a climate variable into the SARIMA model might improve the accuracy of prediction.
as shown in previous studies (Wu et al, 2007; Gharbi et al, 2011). However, this study revealed the Box-Jenkins ARIMA model may be useful for public health authorities to understand trends and forecast incidence in dengue endemic areas. The ARIMA model is a practical tool for developing a forecasting system based on routine data collection within the existing surveillance system, which can strengthen an early warning system and can be used to initiate rapid response activities to anticipate future dengue epidemics.

ACKNOWLEDGEMENTS

This study was supported by the Thailand International Development Cooperation Agency (TICA). We would like to thank the Head of the Jakarta Province Health Office, Dien Emmawati, for allowing us to use the surveillance data. We are also grateful to all the staff at the Health Control Unit, Jakarta Provincial Health Office, and our colleagues from the Sub-directorate of Arbovirosis, MOH, for providing necessary information. Thanks to Mr Irwin F Chavez for editing the final manuscript. SL and JK were supported by the Office of Higher Education Commission and Mahidol University under the National Research Universities Initiative.

REFERENCES


Governor Decree of Jakarta in 2007, No.171 [Indonesian Language].


Kusriastuti R, Sutomo S. Evolution of dengue prevention and control programme in
Temporal Patterns of Dengue Transmission in Jakarta


Martinez EZ, Silva EA. Predicting the number of cases of dengue infection in Ribeirao Preto, Sao Paulo State, Brazil, using a SARIMA model. *Cad Saude Publica* 2011; 27: 1809-18.


Monath TP. Dengue: the risk to developed and developing countries. *Proc Natl Acad Sci USA* 1994; 29; 91: 2395-400.


