Geographical structure of dengue transmission and its determinants in Thailand

Y. NAGAO*, P. SVASTI, A. TAWATSIN AND U. THAVARA

1 Osaka University Graduate School of Medicine, Suita, Osaka, Japan
2 Department of Industrial and Operations Management, Thammasat University, Bangkok, Thailand
3 Department of Medical Sciences, National Institute of Health of Thailand, Nonthaburi, Thailand

(Accepted 18 May 2007)

SUMMARY

Expansion of dengue has been attributed to urbanization. To test this concept, we examined dengue transmission intensities in Thailand. We used the inverse of mean age of dengue haemorrhagic fever (DHF) cases as a surrogate of dengue transmission intensity (or force of infection). The transmission intensity in Bangkok decreased rapidly since the mid-1990s, to levels that are currently lower than in other regions. Regression analysis revealed that transmission intensity is highest in the Northeastern rural region, mainly due to scarcity of private water wells. Private wells reduce the need for household water containers, the major breeding sites for vectors. Cumulatively, these results show that urbanization is not necessarily associated with intense dengue transmission in Thailand. Paradoxically, the DHF incidence in Bangkok has surpassed other regions despite declines in transmission intensity. This finding implies the existence of endemic stability (i.e. low incidence of a clinical illness in spite of high transmission intensity).

INTRODUCTION

Dengue haemorrhagic fever (DHF), a life-threatening manifestation of dengue infection, emerged in the 1950s [1] and has rapidly increased [2]. In the years following its emergence, DHF was most often reported in urban areas [3]. For this reason, the expansion of dengue virus and DHF has been attributed to urbanization [4]. Bangkok, which is inhabited by 9% of the Thai population, was proposed to be the epicentre from which annual waves of DHF have been emitted to the other parts of Thailand [5].

The idea that urbanization constitutes a risk for dengue infection has not been thoroughly tested. To begin to test this idea, transmission intensities (or force of infection) of dengue virus among different regions of Thailand were compared. Our analysis centred on DHF, but not on the more benign dengue fever, since dengue fever is not regularly reported in endemic countries [6]. The incidence of DHF fluctuates dramatically with a cycle of 3–4 years [7], possibly due to the oscillation in population immunity [8]. Hence, the incidence may not accurately represent the transmission intensity. Instead, we assumed that the inverse of the mean age of DHF cases (IMA-DHF) could be used as a surrogate for the transmission intensity of dengue virus. The inverse of mean age of infected individuals is generally regarded as an approximation of the force of infection (i.e. ‘per capita’ rate at which susceptibles acquire infection”) for infectious diseases which confer life-long immunity [8]. This relationship has not been confirmed

* Author for correspondence: Dr Y. Nagao, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita, Osaka, 565-0871, Japan.
(Email: nagaoy@anes.med.osaka-u.ac.jp)
for DHF, since DHF occurs mainly in secondary infections [9, 10]. However, a secondary infection, as well as a primary infection, would be expected to occur at an earlier age under more intense transmission. Consistent with this, recent simulations have predicted that the mean age of DHF is negatively correlated with the force of infection of dengue virus (Nagao & Sakamoto, unpublished observations). In addition, an epidemiological study conducted in Thailand revealed a strong negative correlation between mean age of DHF cases and proportion of houses infested with larvae/pupae of vector mosquitoes (Thammapalo et al., unpublished observations). Therefore, we based our analysis upon the assumption that the IMA-DHF reflects the force of infection of dengue virus, albeit not necessarily linearly.

The present study illustrates the spatial patterns of transmission intensities of dengue across Thailand, as determined by IMA-DHF. In addition, the observed spatial patterns are explained by using climatic and socioeconomic conditions as independent variables. The results highlight the rapidly shifting epidemiological state of dengue transmission, which may also be present in other tropical countries.

**METHODS**

**Epidemiological data**

In Thailand, there are currently 926 districts in 76 provinces. Among these districts, 50 are considered to be part of the Bangkok metropolitan province. Detailed epidemiological data were obtained from the Bureau of Epidemiology, Ministry of Public Health of Thailand. Data regarding the annual number of DHF cases were stratified by age group (0, 1–4, 5–9, 10–14, 15–24, 25–34, 35–44, 45–54, 55–64, ≥ 65 years). The mean age of DHF cases was obtained by averaging the mid-point age of each group (i.e. 0.5, 3, 7.5, 12.5, 20, 30, 40, 50, 60, 75 years) weighted by the number of DHF cases in each group. Then, by adjusting the number of DHF cases in each group to the national demographic structure in 2000, the adjusted mean age of DHF was obtained. The DHF case data were also used to derive the annual DHF incidence, based upon population data obtained from the National Statistics Office of Thailand. Between 1981 and 1994, the spatial resolution of these data was province level. Between 1995 and 2004, it was district level. We used the district-level data for our analysis.

**Map data**

The geographic information system software employed was MapInfo 7.0 (MapInfo, Troy, NY, USA). A digital map that delineates the boundaries of the 926 districts was obtained from MapInfo Thailand.

**Detection of spatial clustering**

To examine the spatial clustering of transmission intensity, we employed a methodology proposed by Getis & Ord [11–13]. In this methodology, the $G^*_i$ spatial statistic describes clustering of the variable of interest, $Y$, around location $i$. Large $G^*_i$ values indicate local clustering of large $Y$ values; whereas, a small $G^*_i$ indicates clustering of low $Y$ values. We derived $G^*_i$ and associated one-tailed $P$ values for the variable of interest, IMA-DHF, at each district $i$. The binary distance matrix required by this methodology was generated based upon a cut-off distance of 200 km between paired district centres. Spatial clusterings (at the level of $P < 0.001$) were noted on the map.

**Climatic data**

Climatic data beginning from 1987 were obtained from the University Corporation of Atmospheric Research [14]. For each year between 1995 and 2004, we extracted the following data: the annual mean of daily mean temperatures ($T_{\text{mean}}$ in the official data definition, °C), the daily minimum temperature ($T_{\text{min}}$, °C), the daily maximum temperature ($T_{\text{max}}$, °C), the total reported precipitation per month (RPCP, mm/month), the average vapour pressure deficit (AVPD, hPa) which is the saturation vapour pressure minus the observed vapour pressure, and the average pan evapo-transpiration (APET, mm/day) which is the amount of water transformed into vapour from a pan of a standard size. This study enrolled 89 weather stations on the Indochina peninsula, which consistently reported all these variables between 1995 and 2004. These climatic variables were interpolated to the geographic centre (centroid) of each district, as reported previously [15].

**Socioeconomic data**

Over 250 socioeconomic variables have been surveyed from each of the 61 000 Thai villages on a bi-annual basis, as described previously [15]. Data were collected in even-numbered years between 1994 and 1998.
and in odd-numbered years since 2001. This was performed under the initiative of the National Rural Development Committee of Thailand. This rural database, officially coded ‘NRD2C database’, does not include Bangkok. We selected the variables that (i) were consistently recorded in all six surveys conducted between 1994 and 2005, (ii) covered more than 90% of all villages in any survey, and (iii) were not related to technical detail for a specific industry. Twenty-five variables were selected (Table 1). Among these, six variables were related to local water resources, four to education, two to public health, four to transportation, three to demographic characteristics, and six to other aspects. These variables were averaged for each district and linearly interpolated to the years intervening the surveyed years. This database can be accessed at: http://www.vector-borne-diseases.org/socioeconomics_thailand.

Regression analyses

To explain the spatial pattern of transmission intensity of dengue (represented by IMA-DHF) based upon climate and socioeconomics, regression analyses were performed. Two regression methodologies were used in tandem in order to select the most robust explanatory variables. The first methodology, the random-effect linear regression model, adjusted for the possible biases caused by repeated measurement or temporal autocorrelation. In the present study, this bias would have arisen from the fact that each district reported a maximum of ten annual values of IMA-DHF. The second methodology was employed to adjust for spatial autocorrelation [16, 17]. For this spatial regression analysis, both dependent and independent variables were averaged from the entire study period for each district.

For the random-effect linear regression analysis, we first selected an ‘optimal time lag’ at which transmission intensity is optimally predicted by each climatic variable as follows. For each climatic variable, the mean value of the previous $j$ years ($j=0–8$) was incorporated as a single independent variable in the random-effect univariate linear regression. The time lag, $j$, which resulted in the maximum $R^2$, was selected as the optimal time lag for each climatic variable. Climatic variables averaged from their corresponding optimal time lags were incorporated into subsequent

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface water</td>
<td>Proportion of villages with surface water</td>
</tr>
<tr>
<td>Piped water</td>
<td>Proportion of villages with supply of piped water</td>
</tr>
<tr>
<td>Health station</td>
<td>Proportion of villages with health station</td>
</tr>
<tr>
<td>Kindergarten</td>
<td>Proportion of villages with kindergarten</td>
</tr>
<tr>
<td>Primary school</td>
<td>Proportion of villages with primary school</td>
</tr>
<tr>
<td>High school</td>
<td>Proportion of villages with high school</td>
</tr>
<tr>
<td>Electricity</td>
<td>Proportion of villages with electricity</td>
</tr>
<tr>
<td>Road to district centre</td>
<td>Proportion of villages connected to the district centre by roads</td>
</tr>
<tr>
<td>Public bus</td>
<td>Proportion of villages where public bus service is available</td>
</tr>
<tr>
<td>Firewood</td>
<td>Proportion of villages where firewood is used for cooking</td>
</tr>
<tr>
<td>Working remotely</td>
<td>Proportion of villages where residents work outside of the subdistrict</td>
</tr>
<tr>
<td>Public large wells</td>
<td>Per capita number of public large water wells</td>
</tr>
<tr>
<td>Private large wells</td>
<td>Per capita number of private large water wells</td>
</tr>
<tr>
<td>Public small wells</td>
<td>Per capita number of public small water wells</td>
</tr>
<tr>
<td>Private small wells</td>
<td>Per capita number of private small water wells</td>
</tr>
<tr>
<td>High-school graduates</td>
<td>Proportion of population who graduated from senior high school</td>
</tr>
<tr>
<td>Pickup trucks</td>
<td>Proportion of households which possess pickup trucks</td>
</tr>
<tr>
<td>Motorcycles</td>
<td>Proportion of households which possess motorcycles</td>
</tr>
<tr>
<td>Land ownership</td>
<td>Proportion of households which own land</td>
</tr>
<tr>
<td>Rent houses</td>
<td>Proportion of households which do not own any land</td>
</tr>
<tr>
<td>Area for agriculture</td>
<td>Proportion of area which is used for agriculture</td>
</tr>
<tr>
<td>Population density</td>
<td>Population per km$^2$</td>
</tr>
<tr>
<td>Household density</td>
<td>Number of households per km$^2$</td>
</tr>
<tr>
<td>Population per house</td>
<td>Average population per household</td>
</tr>
<tr>
<td>Birth rate</td>
<td>Number of births in the previous year per 1000 population</td>
</tr>
</tbody>
</table>
multivariate random-effect regression. Only the socio-economic variables that were recorded in the same year as the dependent variable were used as independent variables. To consider the possible confounding by ecological differences among the regions, dummy variables were included to stratify the analysis by region. The year value of each record was also incorporated as an independent variable to represent the temporal trend. The independent variables (i.e., climate, socioeconomics, region-specific dummy variables and year) that significantly contributed to the multivariate random-effect model were selected following a stepwise process of elimination.

Finally, multivariate spatial regression analysis, which has been frequently used in socioeconomic and health studies [16, 17], was performed using the selected variables. As a result, the variables that remained after stepwise elimination in both random-effect and spatial regressions should be the most robust.

In both random-effect and spatial regressions, the final models were retested after replacing the crude IMA-DHF with the IMA-DHF calculated from the adjusted mean age, to exclude confounding by heterogeneity in the demographic structure. Statistical significance was based upon a two-sided \( \alpha \)-level of 0.05. In all regression analyses, the records with no DHF cases were omitted from the analysis. In preliminary regression analyses, we normalized IMA-DHF [18], and obtained largely similar results to the results from non-normalized IMA-DHF. Considering the difficulty in the interpretation of results, we presented the results from non-normalized IMA-DHF. Stata 9.0 (Stata Corp., College Station, TX, USA) was used for statistical analysis. Software for the spatial analysis was provided by Maurizio Pisati (University of Milano, Bicocca).

RESULTS

dHF epidemiological data

We divided Thailand into four regions, with Bangkok located in the central region (Fig. 1). Changes in the mean age of DHF cases were gradual, while DHF incidence fluctuated dramatically across all regions (Fig. 2a, b). The mean age of DHF cases was consistently the lowest in Northeastern region (Fig. 3). The mean age of DHF cases in the entire country increased gradually beginning in the 1980s, and this increase accelerated beginning in the mid-1990s (Fig. 2a). The increase in mean age of DHF, which
is equivalent to a decrease in transmission intensity, was most prominent in Bangkok. In Bangkok, the mean age of DHF increased from 11 to 22 years between 1995 and 2004, while the average age in the population increased only from 31 to 34 years. DHF incidence in Bangkok was generally lower than other regions before the mid-1990s, as has been reported [19]; however, Bangkok frequently surpassed other regions afterwards (Fig. 2b).

Cluster of transmission intensities

A clustering of high transmission intensities (represented by high IMA-DHF values) was consistently present in the Northeastern region (Fig. 4). In contrast, a clustering of low transmission intensities initially existed in the Northern region. In 2000, the mean age in this region was comparable to the Central region, hence this cluster of low transmission intensities disappeared. Another clustering of low transmission intensities subsequently emerged in the late 1990s in the area surrounding Bangkok and expanded thereafter.

Regression analyses

The optimal time lag was determined by univariate random-effect regression analysis for each climatic variable as follows: 5 years for $T_{\text{mean}}$, $T_{\text{min}}$, $T_{\text{max}}$ and RPCP; 8 years for APET and AVPD. These climatic variables with corresponding optimal time lags, along with socioeconomic variables, were incorporated into multivariate random-effect linear regression to select variables with significant contributions to IMA-DHF [Table 2, column (a)]. The mean age was then predicted based upon this multivariate random-effect model. The trend towards a low mean age in the Northeastern region was well reproduced (Fig. 5a vs. Fig. 3, 2004). This was mainly because public ‘large water wells’ (bobadan in Thai), a positive risk factor, were most prevalent in this region (Fig. 5b). Conversely, private ‘small water wells’ (bonamten), a negative risk factor, were scarce (Fig. 5c).

The variables selected by random-effect linear regression were further tested by spatial regression analysis [Table 2, column (c)]. The results suggested that private small water wells are potential risk reducers of the transmission intensity. APET, public large/small water wells, high school, land ownership, and birth rate remained positive risk factors. The spatial patterns of some of these variables are
presented in Fig. 5. In both random-effect and spatial regressions, the adjustment of mean age of DHF affected the results only slightly [Table 2, columns (b), (d)], ensuring that the interference by the demographic structure was minimal.

**DISCUSSION**

It has often been held that dengue transmission is more intense in urban areas, based on the observation that cities are often associated with a high incidence of DHF. However, DHF incidence was highly volatile, suggesting that incidence may not accurately reflect the transmission intensity. In addition, incidence is subject to biases caused by over-/under-reporting and local modification of diagnostic criteria for DHF [6, 10, 20, 21]. In contrast, mean age of DHF showed higher continuity and is less likely to be affected by over-/under-reporting or modified diagnostic criteria. Based on this rationale, the inverse of mean age of DHF was employed as a surrogate of...
transmission intensity. Use of this methodology revealed clustering of high transmission intensity in the Northeastern rural region and clustering of low transmission intensity in Bangkok and the Northern region.

Regression analysis was employed to explain the transmission intensity in light of climatic and socio-economic conditions. As an initial step, optimal time lag was selected for each climatic variable. For all the climatic variables, the optimal time lags were longer than 4 years. This may suggest that the transmission intensity at a time-point is determined by the accumulated effect of climate during the preceding years.

The highest transmission intensity in the Northeastern region could be explained by a combination of the highest prevalence of public large water wells and the lowest availability of private small water wells, as shown by multivariate regression analysis. The residents who obtain water from public wells store the water in household containers, which provide breeding sites of vector mosquitoes, *Aedes*. On the other hand private wells reduce the necessity of storing water in houses. These results are consistent with our previous report that public wells and private wells were positive and negative risk factors, respectively, for infestation by *Aedes* [15].

The birth rate remained a positive risk factor perhaps because it represents the supply of susceptible population. The finding that ‘land ownership’ was a positive risk factor may be explained by the fact that the large water containers frequently observed in rural Thailand require a yard. In addition, garbage, used tyres, and tree hollows that may be found in a yard often provide breeding sites [22, 23]. The proportion of villages in which high schools are present remained a positive risk. This might suggest that high schools represent a centre of contact between hosts and vectors, persisting after decades-long vector control efforts targeted at primary schools. An alternative explanation might be that pupils in districts with fewer high schools may lodge near their school, in another district, affecting the age profile of DHF in these districts. APET remained a positive risk factor possibly because people in the arid area have to store water in household containers. The high APET in the Northeast (Fig. 5d) explains the high transmission intensity observed in this region. Taken together, biological causality can be explained by most of the independent variables selected in the final model.

Moreover, these findings provide reciprocal support...
for the use of IMA-DHF as a surrogate for dengue transmission intensity.

The variable ‘year’ showed strong negative contribution in the random-effect regression suggesting that there has been a downward trend in dengue transmission intensity that cannot be explained solely by the climatic and socioeconomic factors examined. This trend may reflect, to some extent, vector control efforts. Regional vector control offices have been distributing larvicide freely and conducting educational campaigns within communities and schools. It is probable that such vector control efforts were more intense in urban areas than in remote areas due to the fact that regional offices are located in cities. If so, the rapid increase of mean age of DHF cases around Bangkok might reflect these intense vector control activities. Air-conditioned or window-screened houses, which have been increasingly prevalent in Bangkok, may also have contributed to the rapid decrease of transmission intensity there.

Dengue transmission intensity, represented by IMA-DHF, has been lower in Bangkok than in other regions since the mid-1990s. In contrast, DHF incidence in Bangkok surpassed other regions during the same period. These paradoxical findings are consistent with a state of ‘endemic stability’ [24, 25]. Endemic stability is defined as a state in which the incidence of a clinical illness is low, while the transmission intensity is high. In support of the existence of endemic stability, we recently found that the correlation between DHF incidence and vector abundance were partially ‘negative’ (Thammapalo et al., unpublished observations; Nagao et al., unpublished observations). We hypothesize that the unique aetiology of DHF (i.e. enhancement in secondary infections) may give rise to this endemic stability (Nagao & Sakamoto, unpublished observations). Regardless of its underlying mechanism, the possibility of such a partially negative correlation between DHF incidence and transmission intensity has an important implication. It implies that insufficient vector reduction might transiently increase DHF incidence in areas at high transmission intensity; Bangkok may provide such an example. Hence, a greater awareness of the relationship between DHF incidence and transmission intensity is warranted.

ACKNOWLEDGEMENTS
We are grateful to Professor Arthur Getis for advice on spatial statistics. We thank the Ministry of Public Health of Thailand and the National Statistics Office of Thailand for the preparation of epidemiological and demographic data, respectively.

DECLARATION OF INTEREST
None.

REFERENCES
day/month observations (http://dss.ucar.edu/datasets/ds512.0/). Accessed 14 February 2006.


