Spatio-temporal patterns of dengue in Malaysia: combining address and sub-district level

Cheong Y. Ling1,2, Oliver Gruebner3, Alexander Krämer4, Tobia Lakes1

1Geoinformation Science Lab, Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany; 2Medical Research Resource Centre, Institute for Medical Research, Jalan Pahang, Kuala Lumpur, Malaysia; 3Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA; 4Department of Public Health Medicine, University of Bielefeld, Bielefeld, Germany

Abstract. Spatio-temporal patterns of dengue risk in Malaysia were studied both at the address and the sub-district level in the province of Selangor and the Federal Territory of Kuala Lumpur. We geocoded laboratory-confirmed dengue cases from the years 2008 to 2010 at the address level and further aggregated the cases in proportion to the population at risk at the sub-district level. Kulldorff’s spatial scan statistic was applied for the investigation that identified changing spatial patterns of dengue cases at both levels. At the address level, spatio-temporal clusters of dengue cases were concentrated at the central and south-eastern part of the study area in the early part of the years studied. Analyses at the sub-district level revealed a consistent spatial clustering of a high number of cases proportional to the population at risk. Linking both levels assisted in the identification of differences and confirmed the presence of areas at high risk for dengue infection. Our results suggest that the observed dengue cases had both a spatial and a temporal epidemiological component, which needs to be acknowledged and addressed to develop efficient control measures, including spatially explicit vector control. Our findings highlight the importance of detailed geographical analysis of disease cases in heterogeneous environments with a focus on clustered populations at different spatial and temporal scales. We conclude that bringing together information on the spatio-temporal distribution of dengue cases with a deeper insight of linkages between dengue risk, climate factors and land use constitutes an important step towards the development of an effective risk management strategy.

Keywords: dengue, spatial and space-time scan statistics, relative risk, address and aggregation level, Malaysia.

Introduction

Dengue fever (DF) and dengue haemorrhagic fever (DHF) are arguably the most important infectious diseases in tropical areas (Gubler, 2006). Mainly distributed between latitudes 35° N and 35° S (WHO, 2009a), almost two fifths of the world’s population are at risk with about 50 million new cases emerging each year (Gubler, 1998a;b; WHO, 2009b). Dengue outbreaks commonly occurred in many parts of Africa, Australia, Latin America, Southeast Asia and throughout the Pacific Islands (Rogers et al., 2006). Vector control has proven difficult due to the biology of the mosquito and its ability to breed in domestic habitats (Eisen and Lozano-Fuentes, 2009). Moreover, little is known about the true determinants of high dengue rates and research gaps include the detailed influence of complex climatic issues that vary over space and time. Main determinants include rainfall and temperature (Chadee et al., 2007), presence of suitable larval development sites (Vanwambeke et al., 2011), behaviour and mobility of humans (Gubler, 1998b) and herd immunity (Halstead, 2008). The distribution patterns of dengue cases reflect the multifaceted interaction of all these risk factors. The risk of this disease is expected to significantly increase in the near future, and the absence of effective treatment makes the development of adequate vector control strategies more important than ever (Vanwambeke et al., 2007; Barrera et al., 2011).

In Malaysia, dengue disease has been endemic since the first case was found in Penang Island in 1901 (Ske, 1902). Since then, peaks have been reported many times; for example in 1974, 1978, 1982 and 1990, while the total number of infections has increased as well (Lam, 1993). Out of the four serotypes, DEN-3 was predominant from 1992 to 1995, while DEN-1, DEN-2 and DEN-3 have been alternating in recent years (WHO, 2008; Arima and Matsui, 2011). In 2008, a Malaysian study that included 1,000 individuals aged between 35 and 74...
years found a seropositive rate for dengue of 91.6% (Muhammad Azami et al., 2011). The incidence rate per 100,000 was the second highest in the Western Pacific region in 2010 (Arima and Matsui, 2011). For the year 2010, the Malaysia Ministry of Health (MoH) reported a DF incidence rate of 148.7 per 100,000 population (MoH, 2011), which far exceeds the national target of not more than 50 cases per 100,000 population (MoH, 2010). To the best of our knowledge, the great majority of dengue studies in Malaysia have focused solely on spatial clustering on a yearly basis rather than spatial and temporal distribution of the disease on a daily basis. The latest national dengue studies were purely spatial studies in Hulu Langat, a sub-district of Selangor in 2003 (Er et al., 2010), in Kuala Lumpur in 2009 (Aziz et al., 2012) and in both Selangor and Kuala Lumpur in 2010 (Hassan et al., 2012).

Effective vector control relies critically on the determination of the proper scale to address dengue virus transmission (Getis et al., 2003; Suaya et al., 2009). Different geographical scales have been suggested, ranging from finer scales based on the vector’s flight range (Getis et al., 2003; Yoon et al., 2012) to larger scales due to the relative importance the closeness of neighbouring communities and human movement (Shepard et al., 2012). Fine-scale studies, i.e. at the address level, allow for detailed spatio-temporal analysis in linking the incidence to the local environment as the complex process of viral transmission is spatially continuous (Morrison et al., 1998; Siqueira-Junior et al., 2008). However, they do not account for the background population at risk of contracting a disease and therefore standard risk ratios cannot be calculated. In contrast, aggregation-level studies include the baseline risk population (Jeffery et al., 2009) but encounter the “modifiable area unit problem” (MAUP) (Openshaw, 1983) that causes differences in the analytical results of the same dataset compiled under different, spatial aggregation levels. Both levels provide significant advantages and disadvantages. Furthermore, previous studies that explicitly deal with comparisons of spatial clustering at two scales have been conducted for various diseases, e.g. cancer at the level of town, zip code and census tract (Sheehan et al., 2000) or the level of exact coordinates, census block group, census tract and town (Gregorio et al., 2005). However, these studies did not combine results of the different levels, e.g., the address and aggregation level, where the latter can vary between administrative limitations such as province, district, subdistrict, etc.

We selected the state of Selangor and the Federal Territory of Kuala Lumpur to study patterns of dengue since a large number of dengue cases (69,702 cases) occurred there between 2008 and 2010; indeed accounting for 50.9% of all reported cases in Malaysia according to the Department of Statistics in Malaysia (DoSM, 2011a). In addition, the area is geographically heterogeneous and includes the strongly urbanised Kuala Lumpur with 100% urban population as well as surrounding suburban sub-districts and rural areas with great variations in population density. We aimed to investigate if a combination of two different geographical levels would provide new insights on dengue disease patterns. To our knowledge, this is the first spatio-temporal study combining address and sub-district level with respect to dengue. We mapped the spatio-temporal distribution of dengue risk in Selangor and Kuala Lumpur by quantifying high-risk clusters of serologically confirmed dengue cases. Specifically, our research goals were: (i) to assess the spatio-temporal patterns of dengue cases at the address level; (ii) to assess the spatio-temporal patterns of dengue cases at the sub-district level; and (iii) to combine insights from both spatial levels to provide better documentation assisting health interventions.

Materials and methods

Study area

As seen in Fig. 1, Selangor comprises nine districts: Klang, Petaling, Sepang, Gombak, Sabak Bernam, Hulu Langat, Hulu Selangor, Kuala Langat and Kuala Selangor, which are further divided into 54 “Mukims” or sub-districts, the smallest local governing units. Kuala Lumpur is enclosed by the state of Selangor and has eight sub-districts, namely Ampang, Bandar Kuala Lumpur, Batu, Cheras, Kuala Lumpur, Petaling, Setapak and Ulu Klang. Both states cover an area of 8,222 km² with a location between 2° 35’ N to 3° 60’ N latitude and 100° 43’ E to 102° 5’ E longitude. The study area is thus geographically heterogeneous varying from Kuala Lumpur to the surrounding suburban sub-districts and rural areas along the boundaries of Selangor. The population density varies considerably from Kuala Lumpur at 6,891 persons per km² to Selangor at 674 persons per km² in 2010 (DoSM, 2011a).

Case data and pre-processing

We obtained dengue case data from the Vector Borne Disease Control Division (MoH) that were based on
passive surveillance, i.e. the routine notification of diseases by the state or local health departments to the health department using standardised reporting forms (WHO, 2009a). We used only dengue cases confirmed by single positive IgM antibodies using IgM capture enzyme-linked imunosorbent assay (ELISA). These represented about 46.2% of all reported dengue cases for the years 2008 to 2010. For the address level, we geocoded the geographical location according to the residency: 11,664 cases in 2008, 10,482 in 2009 and 8,300 in 2010. The geocoding accounted for 92.2% (2008), 93.4% (2009) and 94.9% (2010) of the addresses, respectively, while the remainder had to be excluded due to missing addresses. The addresses were geocoded using Google Maps Application Programming Interface (API) that has been shown to be a high-quality geocoding service (Roongpiboonsopit and Karimi, 2010) and successfully used for the geocoding of the location of health care facilities (Gu et al., 2010) and for mosquito surveying (Neteler et al., 2011). For the years 2008 to 2010, we aggregated the number of confirmed cases at the sub-district level (12,424, 10,982 and 8,581, respectively) based on the onset date of the dengue occurrence by daily information.

**Population and map data**

We used population data at the sub-district level from the latest census in 2010 (DoSM, 2011b). In addition, we applied the sub-district map of Selangor and Kuala Lumpur from the Department of Survey and Mapping in Malaysia (JUPEM), Putrajaya, the federal administrative capital located in the district Sepang, was included in Selangor (Fig. 1). ArcGIS, version 10.0 (ESRI; Redlands, USA) and R, version 2.15.0 with the “maptools” package (Lewin-Koh et al., 2012) and “ggplot2” package (Wickham, 2009) were used for data pre-processing, spatial analysis and visualization.

**Spatio-temporal pattern analysis**

At the address level, we delineated clusters of dengue cases on a daily data basis using the software “Spatial and Space-Time Scan Statistics” (SaTScan version 9.1.1) by Kulldorff (2010). The scan statistic technique has been successfully applied in a number of studies (Kulldorff and Nagarwalla, 1995; Song and Kulldorff, 2003; Kulldorff et al., 2004). Space-time...
permutation models were employed to check for outbreaks at the address level where there was no population data available to assess the at risk population structure (Kulldorff et al., 2005). With reference to the official MoH definition of an outbreak as the incidence of two or more dengue cases in a location where the onset date of the cases are less than 14 days apart and located within 200 m of each other (MoH, 1986, 2009), the maximum spatial cluster size was set at 200 m and the maximum temporal cluster size at 14 days. Only clusters with significant levels below 0.001 were reported after Monte Carlo simulation repeated 999 times. A window of a radius that can be varied up to 200 m is moved across the study area with the aim of identifying potential clusters within various window settings. The assessment of a cluster was achieved by comparing the number of cases within the circle with the number of expected cases (assuming random distribution). The cluster with the maximum log likelihood ratio was taken as the most likely cluster, i.e. the cluster least likely to be due to chance. The secondary clusters are those that are in rank order after the most likely cluster, by their likelihood ratio value.

At the sub-district level, the same software SaTScan version 9.1.1 was applied, but with slightly modified settings to better reflect the specific patterns encountered from the polygons of the study region. The analysis was done with the total number of cases observed and the estimated population numbers at each sub-district. A retrospective Poisson-based model was employed as the number of dengue cases in the study area was Poisson-distributed (Kulldorff et al., 2005). The circular scanning window was centred at the centroid of each sub-district polygon. To avoid unwanted effects linked to different sizes of the scanning window (Waller and Gotway, 2004; Chen et al., 2008), we performed sensitivity tests and considered percentages of the population at risk from 1% to 50 %. To make it possible to discover small, homogeneous clusters within larger, heterogeneous ones, while at the same time not failing to detect significant, regional-level clusters due to small window sizes (Chen et al., 2008), the maximum spatial cluster was finally set at 15% of the total population at risk. The maximum temporal cluster size was set at 14 days similar to the address level for comparison purpose. With the availability of the total population at risk, the relative risk can be estimated with the proportion of the estimation risk of dengue disease within the cluster to the estimated risk outside the cluster. To test the null hypothesis that the relative risk of dengue would be the same between any sub-districts and remaining sub-districts, the Monte Carlo approach with 999 repetitions was used. P-values <0.001 were considered statistically significant.

To derive new insights on spatio-temporal patterns of dengue we focused on the two separate spatial scales, i.e., the address and the sub-district, and then visually compared the results by overlays.

Results

The address level

Using the MoH definition of an outbreak, our spatio-temporal analysis at the address level identified a total number of 107 outbreaks (P <0.001) for the whole study period from 2008-2010 (Fig. 2). Spatially, the outbreaks were concentrated in the central and south-eastern region of the study area. Temporally, they occurred most frequently in the period December to March over the 3-year study period (see supplement). The highest monthly number of outbreaks for the 3-year period was observed in December 2009 with a total of nine outbreaks and 54 cases. From September to November 2009 and from May to September 2010, we only observed occasional outbreaks (see supplement).

With reference to the population density (Fig. 1), 67% of the detected address-level outbreaks occurred at sub-districts with more than 2,000 people per km², while 21% occurred at sub-districts with population densities from 1,001 to 2,000 people per km². The top-20 of the outbreaks were found in high-density sub-districts, except the seventh largest cluster that occurred in a low population density sub-district, namely Bestari Jaya (Fig. 2).

The sub-district level

The spatio-temporal analysis discovered several significant high-risk clusters, again concentrated in the central and south-eastern region of the study area (Fig. 2). Using the maximum temporal window of 14 days, we identified seven outbreak clusters (P <0.001) in total (Fig. 2). The most likely cluster, with a relative risk (RR) of 3.94 (P <0.001, 14.0% of the total population), covered three out of Petaling’s four sub-districts in the central region between 7 and 20 January 2010 (Fig. 1). One month later, from 26 February to 11 March 2010, a 2nd secondary cluster (Table 1) was detected in the neighbouring Kuala Lumpur region (RR = 3.69, P <0.001, 14.4% of the total population). Both clusters occurred in the high-population density
Fig. 2. Spatio-temporal clusters of dengue cases at the address level (circular shape) and sub-district level (shaded area) 2008-2010 in the 62 sub-districts of Selangor and Kuala Lumpur, Malaysia.

Supplement. Monthly temporal clusters of dengue cases at the address level in the 3-year period from 2008-2010.
region with more than 2,000 people per km$^2$ (Fig. 1) and were close in space and time. The 1st secondary cluster in the districts of Sepang and Hulu Langat from 30 December 2008 to 12 January 2009 (RR = 5.64, P <0.001, 8.1% of the total population) and the 6th secondary cluster in Petaling and Kuala Lumpur from 1 to 14 January 2009 (RR = 2.44, P <0.001, 12.5% of the total population) were similarly close in space and time (Table 1).

The cluster with the highest number of sub-districts was the 3rd secondary cluster in Hulu Selangor and Gombak with seven sub-districts and RR of 4.99 (P <0.001, 7.2% of the total population). The extraordinary high RR of 16.32 compared to the average of 2.44 to 5.64 was observed in Bestari Jaya of Kuala Selangor (5th secondary cluster) from 27 April to 10 May 2008 (P <0.001, 0.4% of the total population).

All clusters occurred in the first three months of the year except one in April and May 2008 (5th secondary cluster). Of all the clusters, 71.4% were seen in the same month (January) of the three years (Table 1). Three out of seven clusters (1st secondary cluster, 3rd secondary cluster and 6th secondary cluster) occurred in the end of 2008 stretching into the early part of 2009.

### Discussion

Our studies of dengue outbreaks in the province of Selangor and the Federal Territory of Kuala Lumpur at the address level found that the outbreaks occurred mainly in the central and south-eastern regions and that they peaked from December to March. This result supports the findings of the spatial mapping study in Hulu Langat district by Er et al. (2010), who reported the most important outbreaks in the south-eastern region. At the address level, the dengue outbreaks were predominately located in the high-population density region (>1,000 people per km$^2$) implying that the urbanised area played a major role in the dengue epidemic. This is in line with earlier findings such as those reported by Patz et al. (2005) and by Cheong et al. (2014). Probable reasons for the varying areas of identified clusters in the three-year period reported here may be the complex interaction of vector abundance and herd immunity (Salje et al., 2012). Furthermore, *Aedes albopictus* was found to be dominant compared to *Aedes aegypti* in the study area by Saleeza et al. (2011) and a previous ovitrap surveillance conducted by Chen et al. (2006) in the same area indicates that mixed breedings of the two main dengue mosquitoes, *Ae. aegypti* and *Ae. albopictus*, in a single ovitrap are known to lead to higher breeding rate and higher risk of dengue transmission. Recent studies in Japan (Nihei et. al., 2014) and western United States (Kesavaraju et. al., 2014) have stressed that *Ae. albopictus* represents an invasive vector that dominates other *Aedes* species, something that has been shown to lead to serious outbreaks of various arbovirus diseases including dengue (Gratz, 2004).

Spatio-temporal analysis at the address level allows for the detection of potential outbreaks but lacks information about the background population at risk.
(Kulldorff et al., 2005). However, by including the sub-district level, our analysis could also provide this information. At this level, we found that the dengue outbreaks covered a wide area from the central to the north-eastern and south-eastern regions and this identified spatial pattern not only supports, but also augment our findings at the address level. The most likely outbreak cluster was found to be in the Petaling district. The identified clusters at the sub-district level were close in space and time and may have been driven by human mobility rather than spatial action on the vector part (Stoddard et al., 2009; Teurlai et al., 2012) as there is a relatively high daily movement of commuters across the interconnecting sub-districts (Barter, 2002; Bunnell et al., 2002). The 2010 outbreak clusters in this area (Table 1) confirm the findings by Hassan et al. (2012), who associated these clusters with rainfall. Special focus should be given to the dengue cluster with extremely high RR, which occurred in a low-population density sub-district, i.e. Bestari Jaya in 2008 (Fig. 2). A larvae surveillance study by Saleeza et al. (2011) in this area showed that the problem was artificial containers discarded outside houses thereby providing breeding places for mosquito larvae. Proper waste management from the authorities, as suggested by Saleeza et al. (2011) and more community participation (Gubler and Clark, 1996) would improve the situation.

In our study, the sub-district level analysis overestimated of the cluster area. For example, the 3rd secondary cluster had a radius of 16.7 km and included seven sub-districts (Fig. 2), while clusters were only found in two sub-districts at the address level. Overestimation was also present with regard to the 1st secondary cluster where the Ulu Semenyih sub-district, characterised by low numbers of dengue cases, was included only due to its adjacency to Kajang sub-district that was found to have an excess number of outbreaks at the address level (Fig. 2). Jones and Kulldorff (2012) report that the aggregation (sub-district in our case) level has a greater likelihood to include areas that do not belong to the actual cluster in question. They state that as the scan statistics circle grows, it logically encloses more case observations in order to reach the relatively more spatial disparate aggregation level centroids. Hence, careful selection of scanning parameters is needed to better represent true outbreaks (Chen et al., 2008). Furthermore, the MAUP cause arbitrarily results at the aggregation level as the boundaries are not necessarily related to the spatial spread of dengue disease and population density, land use and breeding site availability (Jeffery et al., 2009; Tipayamongkholgul and Lisakulruk, 2011). Additionally, meteorological parameters may vary distinctly within a sub-district. However, across the sub-district level, we were able to estimate the background population at risk for contracting dengue. Thus, this approach provided more valuable information than just looking at the address level, where the likelihood of the outbreak cluster stayed the same across the study area. Therefore, in agreement with findings from Schwartz (1994), who states that information at the address level is not superior to the aggregation level and, conversely, the aggregation level is no substitute for the address level. In our study, overlying the address level results provided the exact location of the outbreaks. In one area we could even show that excess risk was concentrated at cross bordering regions, i.e. as depicted in the 4th secondary cluster, the border crossing regions of the two sub-districts Kapar and Klang (Fig. 2) showed excess numbers of address level outbreaks. This situation provides valuable input for the local authorities of both sub-districts and it would be useful for them to intensify collaboration in vector control.

The observed spatio-temporal patterns of dengue cases may be linked to complex interacting factors, e.g. abundance of Ae. aegypti females, transportation networks, land use change, human behaviour and climate variation (Rogers et al., 2006; Halstead, 2008; Eisen and Lozano-Fuentes, 2009; Lambin et al., 2010). Many have tried to combine such influencing factors in the search for a way of predicting areas of dengue risk (Nakhapakorn and Tripathi, 2005; Kolivras, 2006; Porcasi et al., 2012). We have also done so and investigated the influence of weather and land use in association with dengue cases and reported the results in previous work in the study area (Cheong et al., 2013, 2014). The former paper shows that the relative risk of dengue cases is positively associated with increased minimum temperature and increased rainfall with a lag time of 51 days and 26-28 days, respectively, but negatively associated with wind speed, while the latter points out that the most important land use factors associated with dengue cases are human settlement, followed by presence of water bodies, mixed horticulture, open land and neglected grassland. Due to the complexity of this disease, scrutinizing when and where the dengue outbreaks occurred in the past can be a useful guide for future outbreak prediction (WHO, 2009b). In addition, identification of high-risk dengue areas should focus public health activities such as immunization campaigns, once a vaccine against the dengue virus becomes reality (Lam et al., 2011).

This study has some limitations. Firstly, the dengue
data used did probably not include all those infected in the area studied as reporting is commonly influenced by limitation in finding 100% of the cases (Shepard et al., 2012). However, the study is based on large number of laboratory-confirmed dengue cases, and even if there was some degree of under-reporting, our analyses still provide useful insights with regard to areas at risk for infection. Secondly, we used circular scan statistic in this study, which implies that all detected outbreaks are approximately circular (Kulldorff et al., 2005), which is not always the case. Although the elliptic scan statistic performs well with non-circular true outbreak (Christiansen et al., 2006), longer computing times and the arbitrary setting of the eccentricity penalty must then be taken into account (Kulldorff et al., 2006). Thirdly, the home addresses used to allocate the dengue cases may not always be the correct ones, reflecting the mobility of patients with respect to recreation and work activities. Nevertheless, the higher probability of multiple infections in the household strengthens the home address as the main infection site (Halstead, 2008; de Melo et al., 2012). Additional illustrations of the monthly spatio-temporal clusters of dengue cases at the address level from year 2008 to 2010 can be found in the Supplement.

Conclusion

Combining address and sub-district level information assists the identification of differences and coincidences of high-risk clusters of dengue. Although additional resources may be needed in assuring patient residencies, this strategy should improve routine epidemiological surveillance and vector control and initial high-risk cluster areas can be targeted before the disease spreads into a larger area. Further, studies on the geographical relationships between the various factors influencing the development of high-risk disease clusters would shed light regarding the aetiology of dengue.

The three main findings were: (i) that dengue outbreaks show a significant spatio-temporal pattern at the address level; (ii) that the dengue risk show similar spatio-temporal patterns at the sub-district level; and (iii) that a combination of the outbreak information from address and sub-district level facilitates the planning of effective health interventions.

Acknowledgements

The authors would like to thank the Director General of Health, Malaysia for his permission to publish this paper. We gratefully acknowledge research support by the Ministry of Health Malaysia doctoral scholarship, DFG priority programme 1233 “Megacities-Megachallenge - Informal Dynamics of Global Change” (HO 2568/5-3), and DFG postdoctoral research fellowship (GR 4302/1-1). Appreciation is also extended to the Department of Statistic Malaysia, Disease Control Division of Ministry of Health Malaysia and Department of Survey and Mapping Malaysia for providing the data for this research. We also thank Maduni Madanayake, Rainer Sauerborn, Gerd Schilling and Rohani Ahmad for their advice.

References


de Melo DPO, Scherrer LR, Eiras ÁE, 2012. Dengue fever occurrence and vector detection by larval survey, ovitrap and
northern limit area in Japan. Geospat Health 8, 417-427.


