A Study of the Dengue Epidemic and Meteorological Factors in Guangzhou, China, by Using a Zero-Inflated Poisson Regression Model
Chenggang Wang, Baofa Jiang, Jingchun Fan, Furong Wang and Qiyong Liu
Asia Pac J Public Health 2014 26: 48 originally published online 11 June 2013
DOI: 10.1177/1010539513490195

The online version of this article can be found at:
http://aph.sagepub.com/content/26/1/48

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Asia-Pacific Academic Consortium for Public Health

Additional services and information for Asia-Pacific Journal of Public Health can be found at:

Email Alerts: http://aph.sagepub.com/cgi/alerts
Subscriptions: http://aph.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Jan 12, 2014
OnlineFirst Version of Record - Jun 11, 2013
What is This?
A Study of the Dengue Epidemic and Meteorological Factors in Guangzhou, China, by Using a Zero-Inflated Poisson Regression Model

Chenggang Wang, MD1,2,3, Baofa Jiang, MD1,4, Jingchun Fan, MD2,4,5, Furong Wang, MD3, and Qiyong Liu, MPH2,4,5

Abstract
The aim of this study is to develop a model that correctly identifies and quantifies the relationship between dengue and meteorological factors in Guangzhou, China. By cross-correlation analysis, meteorological variables and their lag effects were determined. According to the epidemic characteristics of dengue in Guangzhou, those statistically significant variables were modeled by a zero-inflated Poisson regression model. The number of dengue cases and minimum temperature at 1-month lag, along with average relative humidity at 0- to 1-month lag were all positively correlated with the prevalence of dengue fever, whereas wind velocity and temperature in the same month along with rainfall at 2 months' lag showed negative association with dengue incidence. Minimum temperature at 1-month lag and wind velocity in the same month had a greater impact on the dengue epidemic than other variables in Guangzhou.

Keywords
public health, communicable diseases, epidemiology, climate change, occupational and environmental health

Background
Dengue fever is an acute, viral infection transmitted by the bite of infected Aedes mosquitoes throughout the tropics and subtropics.1 Guangdong province, located in southern mainland China, has a subtropical monsoon climate with long summers and abundant rainfall. Since the 1990s, frequent outbreaks of dengue have occurred in Guangdong province. From 2001 to 2006, a total of 897 dengue cases were reported in Guangdong, of which 73.89% were in Guangzhou, PR China.2

1Shandong University, Jinan, Shandong, PR China
2State Key Laboratory for Infectious Diseases Prevention and Control, National Institute for Communicable Disease Control and Prevention, China CDC, Beijing, China
3Shandong University of Traditional Chinese Medicine, Jinan, PR China
4Shandong University Climate Change and Health Center, Jinan, Shandong, PR China
5China CDC Key Laboratory of Surveillance and Early-Warning on Infectious Disease, Beijing, PR China

Corresponding Authors:
Qiyong Liu, State Key Laboratory for Infectious Diseases Prevention and Control, National Institute for Communicable Disease Control and Prevention, China CDC, Beijing 102206, PR China.
Email: liuqiyong@icdc.cn
Baofa Jiang, Shandong University, Jinan, 250012, Shandong, PR China.
Email: bjiang@sdu.edu.cn
the capital city of Guangdong Province. The primary vector for dengue in Guangzhou was female *A. albopictus*.

The transmission of dengue needs 3 essential constituents: dengue viruses, a sizeable number of vectors, and susceptible populations. It has been proposed that temperature and humidity play pivotal roles in dengue outbreaks. Environmental factors influence dengue transmission because vector biology and viral replication are temperature and moisture dependent. Meteorological factors affect the life cycle of *Aedes*, such as life span, survival rate, and biting behavior of the female adult mosquito. For example, appropriate temperature and rainfall can increase the availability or abundance of breeding places. Within the range of 20°C to 30°C, the extrinsic incubation period of the virus (a period of time required by the dengue virus for its development to an infective stage within the mosquito) and the life cycle of the mosquito will be shortened with increasing temperature.

Developing an early detection system of dengue risk based on meteorological and entomological indices is a priority for dengue research. In this study, we sought to identify the major meteorological influences on dengue fever in Guangzhou by using a zero-inflated Poisson regression model for the period 2000 to 2012.

**Methods**

**Study Setting**

Guangzhou is located at 112°57′E to 114°03′E and 22°26′N to 23°56′N (Figure 1), south adjacent to the South China Sea, which is the most important foreign trade window and the world-famous port city. The city covers an area of 7434.4 square kilometers, with more than 12 million residents in 2010. It has a humid subtropical climate, with high temperatures and humidity in the summer and relatively mild and dry in the winter. Annual average rainfall ranges from 1468 to 2530 mm, and annual average temperature is typically between 18°C and 25°C.
Materials

Monthly dengue cases in Guangzhou for the period of 2000-2012 were used in this study. As one of class B notifiable communicable diseases since 1989, dengue was first diagnosed in the local medical institutions according to the unified criteria issued by the Chinese Ministry of Health. The diagnosis criteria included epidemiological exposure history, clinical manifestations, and laboratory tests such as white blood cell count. For those who could not be clearly diagnosed, a specific IgG ELISA test was performed to confirm the diagnosis. All the diagnostic criteria were consistent from 2000 to 2012. To ensure scientific responses to the outbreaks of infectious diseases, the Health Ministry of China developed a series of strict report management systems for notifiable communicable diseases. Before 2004, information about notifiable communicable diseases was submitted in the form of cards by local medical institutions to the local CDC within 24 hours, then to China CDC level-by-level with examination and verification. Thus, data from 2000 to 2003 were obtained from the Chinese CDC database on the basis of notifiable communicable disease report cards. In 2004, a direct submission network of infectious diseases was established, from which we downloaded data from 2004 to 2012. Changes in the form of dengue data transfer had no obvious impact on the dengue epidemic. Monthly weather data, including average minimum temperature ($T_{min}$), average maximum temperature ($T_{max}$), average relative humidity (Hum), average wind velocity (Wind), and total rainfall (Rain) were obtained from the China meteorological data–sharing service system.

Model Structure

Poisson regression is widely used to model count data whose mean coincides with the variance. However, there are always too many zeros in count data, which leads to data dispersion, and it is beyond the predictive ability of the Poisson regression. Zero-inflated Poisson regression proposed by Lambert is a model for count data with excess zeros.

The zero-inflated Poisson regression model is given below. Here $P_i = \exp(\gamma_j x_{ij}) / \exp(1 + \gamma_j x_{ij})$ is the probability of zero in the first state, $\mu_i = \exp(\beta_j x_{ij})$ is the expectation of per-month dengue cases, $x_{ij} = x_{i1}, x_{i2}, \ldots, x_{ik}$ is the covariate matrix $X$, and $\beta_j = \beta_1, \beta_2, \ldots, \beta_k$ and $\gamma_j = \gamma_1, \gamma_2, \ldots, \gamma_m$ are parameters to be estimated.

$$\logit(p_i) = \log \left(\frac{p_i}{1 - p_i}\right) = \sum_{j=1}^{k} \gamma_j x_{ij},$$

(1)

$$\log(\mu_i) = \sum_{j=1}^{k} \beta_j x_{ij},$$

(2)

The zero-inflated Poisson distribution is a mixed linear model via logit (shown in the first equation) and log (in the second equation) link functions. In the zero-inflated Poisson distribution, the zeros are assumed to occur in 2 distinct states. In the first state, zeros are the only occurrences with probability $p_i$; the zeros are referred to as “structural” zeros. In the second state, it is a Poisson distribution; observations are zeros or the number of events with probability $(1 - p_i)$, and the zeros are called “sampling” zeros. The model parameters are estimated by the method of maximum likelihood.

First of all, descriptive analysis was performed for all variables. An overdispersion test proposed by Böhning was performed for dependent variables, which compares the sample mean with the sample variance. The Vuong test was also performed to determine whether the apparent overdispersion is induced by the extra number of zeros and which Poisson regression model is more suitable for this study.

Because of the autocorrelation of the monthly dengue cases, it was introduced into the model as an independent variable at 1-month lag. To avoid multicollinearity between the variables of
average minimum and maximum temperature, Pearson’s correlation was performed with Pearson’s correlation coefficient being 0.980 and $P$ value less than .001; the average minimum temperature adopted for it was often the limiting factor for the development of *Aedes* eggs. To control the seasonal variation of dengue in Guangzhou, the seasonal factor was extracted by using an additive model, which was also introduced into the model. No obvious long-term trend of dengue incidence in Guangzhou was found during this study period (2000-2012).19

To initially identify meteorological variables and their lag effects for introducing into the multivariate regression model, the cross-correlation analysis was adopted with data of 2000-2011, which provided Pearson’s product moment correlation coefficients between 2 time series of monthly dengue cases and meteorological variables at various lags and leads. To select the most significant meteorological variables, the backward elimination method was adopted in the multivariable zero-inflated Poisson regression analysis. In this study, the response variable was monthly dengue cases; independent variables of the initial full model were determined by results of cross-correlation analysis.

To verify the internal validity (with data of 2000-2011) and external validity (with data of 2012), the predicted values $[\mu_i = \exp (\beta_j x_{ij})]$ were obtained by using data of 2000-2012 with the final zero-inflated Poisson regression model. Intraclass correlation analysis was also performed to verify the consistency between the actual and predicted data.20 All analyses were carried out by Stata/SE 10.0 for Windows (StataCorp, College Station, TX), with a significance level of $P < .05$.

**Ethical Review**

The present study was reviewed by the research institutional review board of Shandong University and the China CDC. What we used in this study was disease surveillance data from which identifiers had been permanently removed, and hence, no specific individual could be discerned. So this study did not require ethics clearance.

**Results**

Descriptive statistical results for all variables are shown in Table 1 and Figure 2. Meteorological variables showed seasonal variation, peaking from June to August. The dengue epidemic in Guangzhou also showed an apparent seasonal pattern, with no cases in January to March and the peak (accounting for 88.00% of all cases) in August to October.

Table 2 shows the results of cross-correlation analysis of the relationship between monthly dengue cases (2000-2011) and meteorological variables with a lag of 0 to 3 months. The significant variables included average minimum temperature and average relative humidity at lags of 0 to 3 months, total rainfall at lags of 1 to 3 months, and average wind velocity at the same month, which were completely introduced into the initial multivariable regression model.

A total of 2050 dengue fever cases were reported from 2000 to 2011, with a mean of 14.24 cases per month, significantly less than the variance (3547.41). Results of the overdispersion test for the dependent variable are shown in Table 1. The statistic of $O$ was 2098.01, with $P$ value less than

**Table 1.** Descriptive Statistics of Monthly Meteorological Factors and Dengue Cases.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Variance</th>
<th>Minimum</th>
<th>Maximum</th>
<th>The Statistic of $O$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>14.24</td>
<td>3547.41</td>
<td>0.0</td>
<td>454.0</td>
<td>2098.01</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Wind (m/s)</td>
<td>1.61</td>
<td>0.17</td>
<td>1.1</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rain (mm)</td>
<td>156.91</td>
<td>24152.71</td>
<td>0.0</td>
<td>834.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum Temperature (°C)</td>
<td>19.50</td>
<td>31.81</td>
<td>6.5</td>
<td>27.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humidity (%)</td>
<td>72.72</td>
<td>54.76</td>
<td>48.0</td>
<td>87.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aThe Statistic of $O$ is an index of dispersion, $O = \sqrt{(n-1)/2(S^2-X)/X}$.*

---

Wang et al

51

UNIV AUTO DE YUCATAN on October 27, 2014

Downloaded from aph.sagepub.com at UNIV AUTO DE YUCATAN on October 27, 2014
Table 2. Cross-correlation Coefficients Between Monthly Dengue Cases (2000-2011) and Meteorological Variables in Guangzhou, China.

<table>
<thead>
<tr>
<th>Lag (months)</th>
<th>Wind</th>
<th>Rain</th>
<th>Minimum Temperature</th>
<th>Humidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.115a</td>
<td>0.096</td>
<td>0.177a</td>
<td>0.047a</td>
</tr>
<tr>
<td>1</td>
<td>-0.072</td>
<td>0.166a</td>
<td>0.242a</td>
<td>0.120a</td>
</tr>
<tr>
<td>2</td>
<td>-0.014</td>
<td>0.172a</td>
<td>0.256a</td>
<td>0.159a</td>
</tr>
<tr>
<td>3</td>
<td>0.033</td>
<td>0.228a</td>
<td>0.219a</td>
<td>0.157a</td>
</tr>
</tbody>
</table>

*p < .05.

Figure 2. Monthly weather measurements for 2000-2011 and actual and predictive monthly dengue cases for 2000-2012 in Guangzhou city.

In the last figure, the area to the left of the red line represents internal validity (with data of 2000-2011), with the other side representing external validity (with data of 2012).
which meant that the data were overdispersed. Results of the Vuong test also showed that the null hypothesis (no difference on the zero predicted probability between the Poisson regression model and zero-inflated Poisson regression model) was rejected, with the Z value being 2.43 and P value .008. This demonstrated the apparent overdispersion induced by the extra number of zeros (there being no dengue case within 93 months). Therefore, the zero-inflated Poisson regression model was more appropriate for this study than the standard Poisson regression model.

To validate the effectiveness of the zero-inflated Poisson regression model, predictive analysis was carried out by using data from 2000-2012. The sequences were drawn to compare the actual and predictive values (the last plot in Figure 2). Forecast data presented the obvious seasonal characteristics with a single peak. The predicted and actual values were reasonably consistent both in high- and low-risk years. Results of the intraclass correlation analysis demonstrated that the coefficient was 0.922 (95% confidence interval [CI] = 0.893-0.943), with a P value less than .001, which indicated almost perfect agreement between the predicted and actual values.

Results of the final zero-inflated Poisson regression model are presented in Table 3. In the first equation, dengue cases at 1-month lag were positively correlated with the probability of dengue occurrence. In the meantime, we tried to introduce other variables, such as minimum temperature, relative humidity, and rainfall, at 1 to 3 months’ lag, but they were not statistically significant. In the second equation, the number of dengue cases and minimum temperature at 1-month lag together with relative humidity at 0- to 1-month lag were all positively correlated with the prevalence of dengue fever. Wind velocity and temperature at the same month, along with rainfall at 2 months’ lag were negatively associated with dengue incidence. But seasonal factors were not significant at the .05 level. Incidence rate ratios (ie, exp β) of dengue were calculated, with 95% CIs, at 1 month. As we can see from Table 3, the minimum temperature at 1-month lag and wind velocity at the same month had greater impact on the dengue epidemic than other variables.

Discussion

In this study, we correctly identified and quantified the relationship between dengue and meteorological factors in Guangzhou by using the zero-inflated Poisson regression model, which may be useful for developing predictive tools for dengue control.
The epidemic of dengue fever in Guangzhou showed a seasonal pattern, with the number of monthly dengue cases being overdispersed and no cases in some months. The zero-inflated Poisson regression model is suitable for modeling the dengue epidemic in Guangzhou and has the advantage of being easier to interpret than the negative binomial model. A mixed probability distribution is developed to model the zero counts and nonzero counts, which could explain too many zero values in the data. The estimated results are efficient and unbiased.

Meteorological factors may directly or indirectly affect vector survival, life span, development and reproductive rates, and viral replication. This study demonstrated that dengue incidences correlated positively with some meteorological variables, such as relative humidity at 0- to 1-month lag and average minimum temperature at 1-month lag, and negatively with wind velocity, temperature of the same period, and rainfall at 2-month lag. A study in Guadeloupe, French West Indies, documented a positive correlation between dengue incidence with variables of relative humidity at 7 weeks’ lag and minimum temperature at 5 weeks’ lag. A longitudinal survey in Brasilia found that relative humidity and temperature were the only meteorological factors that correlated with these entomological indices. A study in Guangzhou by using time series Poisson regression analysis with data of 2001-2006 also indicated similar results, but total rainfall at 2-month lag, minimum temperature, and relative humidity at the same month failed to enter the best-fitting predictive model.

Temperature affects each stage in the life cycle of the mosquito. Within 20°C to 30°C, rising temperatures can increase the survival rate of adult and immature mosquitoes, accelerate their larval growth, enhance dengue virus replication, and shorten the extrinsic incubation period. \( A \) albopictus begins biting when the temperature is more than 12°C. The bite rate will increase as the temperature rises, with the highest bite rates in the 25°C to 30°C range. But when the temperature is too low (below 20°C) or too high (more than 35°C), the mortality of larvae of \( A \) albopictus will rise. When the temperature is above 32°C, the mosquito bite rates will drop. These may explain why the temperature of the same period has a negative effect on the prevalence of dengue fever in this study.

Relative humidity has an influence on the survival of mosquito eggs and adults, the biting behavior of female adult mosquitoes, and laying of eggs. At the same temperature, egg hatch-ability of \( A \) albopictus increases as the relative humidity rises. The optimum relative humidity is 75% for saving eggs. When the humidity is too low, laying eggs will be affected, and adult mosquito mortality will increase. Increasing humidity will also facilitate feeding for the adult mosquito, enhancing its survival.

Extreme wind velocity tends to suppress mosquito flight, thus affecting the biting rates of adult mosquitoes. A negative correlation was also found between dengue epidemic and wind velocity at 3 weeks’ lag in Barbados. Past rainfall will influence vector abundance in subsequent weeks or months by creating more breeding habitats for mosquitoes. Meanwhile, heavy rain can destroy existing mosquito breeding sites and affect maturation of mosquito eggs or larvae. In this study, similar results were obtained.

The dengue epidemic in Guangzhou was characterized by low-level epidemic from imported cases, followed by sudden occurrence and rapid transmission, resulting in an unpredictable large-scale outbreak without cyclic change. Being an infectious disease, dengue cases of the past month had a greater impact on the epidemic of the current month. This may affect the probability of dengue occurrence and the number of cases. In 2002 and 2006, the local outbreaks of dengue in Guangzhou should be mainly attributed to lack of early control of the epidemic, except for the relatively high minimum temperature and low wind velocity compared with other years.

According to the epidemic characteristics of dengue in Guangzhou (a low-level epidemic from imported cases, an apparent seasonal pattern, and no cases in some months, etc), this study analyzed the impact of meteorological factors on dengue fever by using the zero-inflated Poisson regression model. With the final model used in this study, we can predict the probability of
dengue occurrence and the number of cases in Guangzhou 1 month ahead. This may be useful for developing a dengue early warning system and taking measures for dengue control in advance. We should note that the present study also had some limitations. For example, we did not take socioeconomic effects into consideration because information about those factors was not available and not easy to be obtained.

**Conclusion**

Taken together, our findings shed new light on the relationship of meteorological factors and dengue fever in Guangzhou city of China by using the zero-inflated Poisson regression model. This may contribute to forecasting dengue epidemics in the future.

**Acknowledgments**

We are very grateful to the 2 anonymous reviewers who put forward invaluable comments, which were very helpful in improving our article.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by National Basic Research Program of China (973 Program) (Grant No. 2012CB955504), the National Natural Science Foundation of China (NSFC; Grant No. 30972563), and Health Scientific Research Special Funds Programs, China (Grant No. 201202006).

**References**


