www.cafetinnova.org

Indexed in
Scopus Compendex and Geobase Elsevier, Chemical Abstract Services-USA, Geo-Ref Information Services-USA

ISSN 0974-5904, Volume XX, No. XX

International Journal of Biological Sciences and Engineering

XXXXXX

A Multilevel Analysis to Determine the Impact of Demographic, Clinical and Climatological Factors on Type of Dengue

SHANIKA L WICKRAMASURIYA AND ROSHINI SOORIYARACHCHI
Department of Statistics, Faculty of Science, University of Colombo, Colombo 03, Sri Lanka
Email: *shaniw_21@yahoo.com, roshinis@hotmail.com*

Abstract: This study mainly aims to identify factors that affect different types of dengue infections. Information about 5059 dengue patients reported at 15 districts in Sri Lanka during the period of 2006-2008 was utilized for this purpose. Results indicate that there is an impact from the clustering variable, district and from demographic variables, Age, Sex; health and laboratory variables, Platelet count, Survival time and Packed Cell Volume; Climatological parameters, rainfall, temperature, humidity and their lags

Key words: *Dengue Fever, infectious disease, epidemiology, demographic, climatic, clinical, dengue hemorrhagic fever*

1. Introduction

A. Background

Dengue virus is most commonly transmitted by the female mosquitoes of the *Aedes aegypti* and more rarely *Aedes albopictus* genus. It is the most important mosquito-borne viral disease affecting humans. Mainly, it is caused by four closely related, however, antigenitcally distinct, virus serotypes (DEN-1, DEN-2, DEN-3 and DEN-4), of the genus *Flavivirus*. Infection with one of these serotypes does not provide cross-protective immunity; therefore, persons living in dengue endemic areas can have four dengue infections during their lifetime. As a result, people with a repeat (secondary) dengue virus infection have a greater risk of dengue than persons infected for the first time [1].

Infection with a dengue virus serotype can produce a spectrum of clinical illness, ranging from an influenza-like illness to a fatal shock syndrome [2].

This [2] indicates that dengue has many manifestations ranging from asymptomatic to flu-like state to a life-threatening condition. The symptomatic patients can be further divided into the main full blown dengue types of Dengue Fever (DF), Dengue Hemorrhagic Fever (DHF) – I and II and Dengue Shock Syndrome (DSS) in this order of severity.

If the patient's susceptibility to the deadly form can be diagnosed early then many lives could be saved. Protein biomarkers have led the way in developing the first accurate predictive model to differentiate between DF and DHF [3]. This will allow the doctors to provide better individual diagnostics and treatments for the different forms of dengue. There is no drug treatment for DHF and fatality rates can exceed 20%. However, early and intensive therapy can reduce the rates to less than 1% [3].

In addition to protein biomarkers there are various factors associated with the emergence of dengue infection types [4] and could be used for detecting the type of dengue to supplement the use of protein biomarkers. These factors are rather complex and could be categorized as viral, vector, host as well as environmental conditions. Figure 1 [4] depicts this graphically in a more meaningful way.

The four main factors making up figure 1 can be characterized as follows. Most effects relating to environment, vector and virus can be attributed to the weather measured in terms of rainfall, temperature and humidity, as weather patterns significantly affect the environment, mosquito vector and transmissibility of the virus [5].

A study conducted in Thailand showed that there is a strong influence of temperature on vector efficiency for dengue virus [6]. This study has concluded that when mosquitoes were incubated at 30˚C, the extrinsic incubation period for dengue-2 virus was 12 days, however, dropped to 7 days at 32˚C and 35˚C. Further, the host factor can be characterized by demographic and clinical factors.

B. Aims of the study

The main aim of this research is to study the impact of demographic, clinical and weather conditions on the type of dengue infection with the intention of determining the epidemiology of different types of dengue infections. As DSS is a rare condition and it is

difficult to find adequate data, DF and DHF1 and DHF2 were considered in this research. A comprehensive

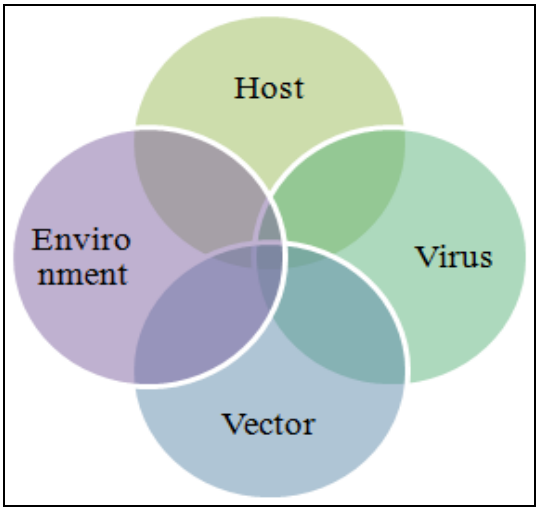

Fig1: *Factors associated with the dengue infections*

study such as this has not been done on different types of dengue infections. Thus the conclusions reached from this study will help the health sector to make policy decisions regarding the detection and control of dengue types.

C. Data for the study

The data for this study is from the Epidemiology Unit of Sri Lanka. Generally in Sri Lanka, cases of dengue typically vary throughout the year and assume a regular pattern, normally in association with changes of temperature, rainfall and humidity [7]. Since the collection of stagnant water favors the breeding of mosquitoes, Sri Lanka tends to experience dengue outbreaks twice a year due to the monsoon seasons.

D. Brief description of methods

Patients belonging to the same district tend to have similar behaviors and climatic conditions, which in turn results in an obvious hierarchical structure, where patients are level 1 unit and the districts are level 2 units. Therefore, it is not reasonable to perform a traditional statistical analysis by pooling all the records of dengue in island-wide fashion. If the districts are regarded as a random sample from a population of districts, and the primary interest is in making inferences about the variation between the districts in general, then the most appropriate structure for this dataset is multilevel [8]. This will lead to correct standard errors, confidence intervals and significance tests. In order to satisfy the objective of this research a hierarchical regression model in the form of a multilevel multinomial model was fitted to the data after suitable initial examination of the data using descriptive and univariate methods.

2. Methods

A. Study Population

In order to investigate the factors contributing to the spread of the disease, dengue records of Epidemiology Unit, Sri Lanka were mainly utilized. It consists of details about each and every dengue patient reported from island-wide government and private hospitals during 2006-2008.

Human Subjects Approval Statement

Ethical approval to use human subjects was obtained from the Epidemiology Unit of the Ministry of Health in Sri Lanka. Research carried out on these humans is in compliance with the Helsinki Declaration.

To account for the impact of the environmental factors, the climatic data were obtained from the Meteorological Department, Sri Lanka. It contains monthly mean rainfall (mm), mean temperature (˚C) and mean humidity (%) records for all the districts from January 2006 to December 2008.

The response variable of interest is the type of the dengue infection diagnosed by the physician. In order to investigate the existence of any difference between the type of the dengue infection and demographic variables such as age of the patient and sex, health and lab variables such as, Survival time (time spent in the hospital), White Blood Cell count, Platelet and Packed Cell Volume of the patient, and climatic conditions such as the amount of rainfall, temperature and humidity of the respective district for the respective month were considered. Previous studies indicate that the peak incident seasons of dengue are identified to be around the monsoon seasons [9]. These facts prompted to consider the monthly rainfall, temperature and humidity as the explanatory variables of this study. Secondly, based on the findings of a previous study, the lag effects of these variables were taken into account [10]. Moreover, the choice of the lag effects was heavily governed by the fact that, places with pure stagnant water are more favorable for the breeding of *Aedes aegypti* mosquitoes after a rainfall season. The impact of time was incorporated by considering the month on which the patient had got the disease.

Finally, the dataset consists of 16 variables spread across two main levels. The 1st level unit comprises of individual patients, who are then nested within the 2nd level unit which is identified as the district to which her/his residence belongs. This type of data structures are known as “hierarchical data”.

Even though the climatic variables are measured at the district level, within a particular district, date of onset of the disease of the patient is different. Therefore, all the

variables including climatological variables are measured at the patient level.

B. Data Preparation

Except for the variables White Blood Cells (WBC) and Packed Cell Volume (PCV), all the other variables have approximately more than 90% of the data. However, the variables, WBC and PCV have less than 75% of data but, as these are important variables and have more than 60% of the observations; will not be eliminated from the study. Data with incomplete records were discarded from the study.

Missing data are frequently encountered in climate variables due to many reasons including failure in the observatory instruments, meteorological extremes and observation recording errors. One possible way of overcoming this difficulty is to impute missing values in a series by using complete data from correlated nearby climate stations' [11]. Therefore, inverse distance weighting was used in this study to estimate a point using sample points, which were weighted by a quantity proportional to the inverse of their distance from the estimated point.

The final dataset consists of 5059 observations (DF=3966; DHF1=398; DHF2=695), among 15 districts namely, Anuradhapura, Batticaloa, Colombo, Galle, Gampaha, Hambantota, Kalutara, Kandy, Kegalle, Kurunegala, Matale, Matara, Puttalam, Ratnapura, and Trincomalee.

Table 1 summarizes number of missing values for each climatic variable at each of the 15 stations.

Table1: *Number of missing observations for matara and trincomalee*

Climatic Variable	District	
	Matara	Trincomalee
Rainfall	4	5
Rainfall lag 1	1	6
Rainfall lag 2	5	7
Temperature	0	5
Temperature lag 1	0	5
Temperature lag 2	0	6
Humidity	0	5
Humidity lag 1	0	5
Humidity lag 2	0	6

As illustrated in the table 1, missing values were seen only in Matara, and Trincomalee. The number of missing observations is very low and thus imputed by using the inverse distance weighting procedure.

All the continuous variables, except time were categorized according to their percentiles so as to avoid the problem of nonlinearity between these and the log it in modeling. This will help to enhance the major findings of this study to a general audience. Moreover, based on the results of a simulation study it was recommended that dividing into three categories based on $(1/3)^{rd}$ split could be more informative than a binary split (median split), so as not to sacrifice the interpretability [12]. This is the motivation for categorization of the continuous variables into 3 levels.

C. Statistical Methods

1) Preliminary Analysis

The potential analytical factors/covariates, their notations, categories and coding mechanism are presented in Table 2.

Prior to performing the model fitting it is essential to conduct a comprehensive preliminary analysis for the dataset of interest to identify the special features of the dataset and the impact of various factors/covariates on the response. This was carried-out using descriptive and univariate analysis. In the univariate analysis the Generalized Cochran Mantel Haenszel test for correlated categorical data was used to adjust any correlations present as most of the traditional techniques for assessing the relationship between variables could not be used for stratified data structures [13].

Generalized Cochran Mantel Haenszel test provides three different test statistics (T_P, T_U and T_{EL}) to test the association between variables in the presence of a stratified data structure. However, during the univariate stage of this study one of the test statistics was used namely T_P as recommended [13].

The univariate analysis used in this study assesses the relationship between the response variable and the explanatory variables. However, it is essential to identify the influence of these explanatory variables jointly on the response. Thus, the univariate analysis was further followed by an advanced analysis technique known as multilevel modeling which accounts for the underling hierarchical nature of the dataset.

Table2: *Potential Analytical Factors/Covariates*

Variable	Notation	Category	Code
Response Variable			
Type of the dengue infection		DF DHF 1 DHF 2	1 2 3
Patient Level Explanatory Variables			
Demographic Variables			
Age	**Age**	< 20 years	1
		20 – 35 years	2
		> 35 years	3
Sex	**Sex**	Male	1
		Female	2
Health and Laboratory Variables			
Survival	**Survival**	< 8 days	1
		8 – 9 days	2
		> 9 days	3
White Blood Cells Count	**WBC**	< 3200	1
		3200 – 4800	2
		> 4800	3
Platelet Count	**Platelet**	< 40000	1
		40000 – 80000	2
		> 80000	3
Packed Cell Volume	**PCV**	< 40	1
		40 – 45	2
		> 45	3
Climatological Variables			
Rainfall	**RF**	< 121.2	1
		121.2 – 252.1	2
		> 252. 1	3
Rainfall lag 1	**RF_1**	< 128.8	1
		128.8 – 242.6	2
		> 242.6	3
Rainfall lag 2	**RF_2**	< 119.7	1
		119.7 – 250.5	2
		> 250.5	3
Temperature	**Temp**	< 27.1	1

		27.1 – 27.8	2
		> 27.8	3
Temperature lag 1	**Temp_1**	< 27.2	1
		27.2 – 27.9	2
		> 27.9	3
Temperature lag 2	**Temp_2**	< 27.3	1
		27.3 – 28.1	2
		> 28.1	3
Humidity	**Humid**	< 81.5	1
		81.5 – 83.5	2
		> 83.5	3
Humidity lag 1	**Humid_1**	< 82	1
		82 – 83.1	2
		> 83.1	3
Humidity lag 2	**Humid_2**	< 81.5	1
		81.5 – 83	2
		> 83	3
Time	**T**		

Note: *The survival time is calculated by subtracting the date of discharge from date of onset, thus all the observations are censored.*

2) Multilevel Multinomial Model [14]

Consider that y_{ij} is the categorical response for the i^{th} level 1 unit in the j^{th} level 2 unit, and denote the probability of being in category s by ${\pi_{ij}}^{(s)}$.

Then, a two level random intercept model is as follows,

$$log\left(\frac{{\pi_{ij}}^{(s)}}{{\pi_{ij}}^{(t)}}\right) = {\beta_0}^{(s)} + {\beta_1}^{(s)} x_{ij} + {u_j}^{(s)},$$
$$s = 1, 2, \ldots, t-1 \quad \ldots\ldots\ldots\ldots\ldots (1)$$

Where ${u_j}^{(s)}$ is the 2nd level random effect, and assumed to be Normally distributed with mean zero and variance $\sigma_u^{2(s)}$. The random effects are contrast specific, as indicated by the s superscript, due to unobserved 2nd level factors that affect each contrast. However, the random effects may be correlated across contrasts:

$$Cov\left(u_j^{(s)}, u_j^{(r)}\right) = \sigma_u^{(s,r)} \; s \neq r$$

Furthermore, in order to calculate the predicted probabilities ${\pi_{ij}}^{(s)}$ $(s = 1, 2, \ldots, t)$ for different values of $x's$ the following formula could be used.

$${\pi_{ij}}^{(s)} = \frac{exp({\beta_0}^{(s)} + {\beta_1}^{(s)} x_{ij} + {u_j}^{(s)})}{1 + \sum_{k=1}^{t-1} exp({\beta_0}^{(k)} + {\beta_1}^{(k)} x_{ij} + {u_j}^{(k)})} \ldots\ldots (2)$$

The probability of being in the reference category t is obtained by subtraction:

$${\pi_{ij}}^{(t)} = 1 - \sum_{k=1}^{t-1} {\pi_{ij}}^{(k)} \quad \ldots\ldots\ldots (3)$$

This study uses MLwiN v2.19 to perform the multilevel modeling. The estimation procedure used was the 1st order predictive (Penalized) quasi-likelihood (PQL) followed by Markov Chain Monte Carlo (MCMC) methods, since it gives a DIC (Deviance Information Criteria) value which could be used as a measure to compare competing models [15]. The basic idea of MCMC is that prior distributions for each of the parameters are combined with the data to produce a posterior distribution for the parameter. Results are gained from a burn-in of 500 and a chain length of 15000 [15].

When performing the advanced analysis, all variables were taken into consideration irrespective of whether they were significant or not in the univariate analysis. Even though, some variables tend to be insignificant

when tested alone, these could be significant in the presence of other covariates/factors in the model.

Generally, it is recommended to build a complex multilevel model gradually by starting from the simplest model. Therefore, a forward selection procedure was adopted to determine the best model.

MLwiN v2.19 treats the level code with the lowest value as the base category. However, in order to get more practical results, default base categories were changed. Generally, many clinicians believe that females and children are more susceptible to suffer from dengue infections. This leads to consider males and patients with age group greater than 35 as the reference category for the variables Sex and Age respectively. Moreover, the severity of DHF1 and DHF2 is higher than DF, thus these patients have to stay quite a long time in the hospitals than that for the DF patients. Therefore, lowest category of the survival time is selected as the base category, as the primary objective of this study is to identify the factors that differentiate DHF1 and DHF2 with compared to DF. According to the dengue guidelines, it was observed that as the severity of dengue infection increases WBC and Platelet count would tend to decrease, whereas the PCV would tend to increase. Thus, high categories of WBC and Platelet count, and the lower category of PCV were chosen as the base category. As there is no clear idea about the climatological variables, the default base categories (lower category) were used as the base categories.

The DIC (Deviance Information Criteria), together with the Wald Statistic will be used to check the statistical significance of the added variables. Generally, it is recommended that a model with lower DIC value is preferred over a higher value [16].

Consider $\hat{\beta}$ to be an estimate for a parameter, whereas $SE(\hat{\beta})$ is the standard error of the estimate. Then for testing the following hypothesis, Wald statistic could be used.

$H_0: \beta = 0 \; Vs \; H_1: \beta \neq 0$

The Wald statistic is given by $W = \left(\frac{\hat{\beta}}{SE(\hat{\beta})}\right)^2$

Under the null hypothesis, this statistic is said to be distributed as Chi Squared with one degree of freedom.

If the calculated test statistic exceeds the critical value of the Chi-Squared distribution (i.e., $\chi^2_{1,\alpha}$), then it is possible to reject the null hypothesis at α level of significance.

Following the model building procedure, it is important to perform a residual analysis, to validate the model assumptions. However, in the multilevel logistic regression context, the only available methods of residual analysis are mostly graphical techniques. Among these, Caterpillar plot and the Normal plot are widely used. Moreover, this would be followed by the Anderson Darling test to assess the normality of the district level residuals in a statistically meaningful way.

3. Results

A. Descriptive Analysis

Table 3 shows the distribution of type of dengue infections by district.

Table3: *Distribution of type of dengue infection by district*

District	Average annual incidence rate for different types of dengue infections (per million)		
	DF	DHF1	DHF2
Anuradhapura	142.7	12.1	12.5
Batticaloa	35.7	8.9	9.6
Colombo	**450.1**	**34.1**	**72.3**
Galle	42.1	2.2	6.3
Gampaha	**242.1**	**30.4**	**65.9**
Hambantota	126.2	15.1	12.7
Kalutara	**276.9**	17.7	**30.3**
Kandy	65.5	8.2	4.8
Kegalle	**205.7**	**41.1**	**44.5**
Kurunegala	49.0	12.9	21.0
Matale	150.9	1.4	4.2
Matara	181.2	25.4	27.9
Puttalam	105.1	20.8	23.0
Ratnapura	24.6	3.1	7.1
Trincomalee	56.3	3.8	15.0

When annual incidence rates are considered highest rates, were observed for Colombo, Gampaha and Kegalle districts. Therefore, Western province (Colombo, Kalutara, and Gampaha) is a high risk zone for dengue and this may indicate that highly populated urban areas are more favorable for dengue transmission.

Table 4 represents the percentage incidence for different types of dengue infections for different levels of variables.

When comparing males with females, males have a higher percentage incidence of DF compared to females while females have a higher percentage of DHF2

compared to males. The percentage incidence of DHF1 is similar for both males and females. Patients with age greater than 35 years have a higher incidence of DF with compared to other age groups whereas patients with age less than 20 years have higher percentage of getting DHF1 and DHF2 with compared to other age groups.

Table4: *Percentage incidences for different types of dengue infections for different variables*

Variable	Category	Percentage incidences for different types of dengue infections		
		DF	DHF1	DHF2
Sex	Female	74.91	8.42	16.66
	Male	76.79	8.83	14.38
Age	<20	72.08	10.76	17.16
	20-35	76.68	8.38	14.94
	>35	80.40	6.31	13.30
Survival time	<8	78.57	7.75	13.68
	8-9	76.43	9.09	14.48
	>9	71.70	9.33	18.97
WBC	Low	76.27	8.60	15.14
	Moderate	74.90	8.52	16.58
	High	76.93	8.87	14.19
Platelets	Low	68.80	9.45	21.75
	Moderate	76.47	9.27	14.26
	High	83.23	7.20	9.57
PCV	Low	80.52	8.35	11.13
	Moderate	73.88	9.18	16.94
	High	73.77	8.67	15.29
RF	Low	77.18	8.05	14.78
	Moderate	78.05	8.87	13.08
	High	80.04	6.87	13.09
RF_1	Low	78.71	7.50	13.80
	Moderate	75.95	8.87	15.18
	High	80.73	7.25	12.02
RF_2	Low	79.43	8.02	12.55
	Moderate	74.65	7.98	17.37
	High	81.49	7.63	10.88
Temp	Low	81.31	8.24	10.45
	Moderate	77.42	7.45	15.13
	High	77.43	8.10	14.48
Temp_1	Low	81.04	8.04	10.92
	Moderate	78.93	7.91	13.17
	High	75.74	7.72	16.54
Temp_2	Low	81.73	8.01	10.25
	Moderate	78.02	8.25	13.74
	High	76.00	7.22	16.78
Humid	Low	82.60	7.68	9.72
	Moderate	75.39	8.33	16.28
	High	78.82	7.54	13.64
Humid_1	Low	80.80	8.49	10.71
	Moderate	75.38	6.93	17.69
	High	79.63	8.32	12.05
Humid_2	Low	79.33	9.61	11.05
	Moderate	75.75	7.20	17.05
	High	81.05	7.29	11.66

Patients with survival time less than 8 days have a higher incidence of DF with compared to other survival categories while, patients with survival time greater than 9 days have a higher percentage incidence of getting DHF2 with compared to other survival times. There is a minor difference between the patients with survival time 8-9 days and above 9 days for the DHF1 category. When considering the WBC count, low and high level of WBC counts have approximately the same percentage of incidence. However, moderate amount of WBC count have a higher incidence percentage of DHF2 when compared to other levels of WBC. Further, the percentage of incidence of DHF1 is similar in all WBC levels. Patients with high platelet count have a higher percentage incidence of getting DF with compared to other platelet categories, whereas patients with low platelet count have a higher percentage incidence of getting DHF2. However, patients with low and moderate platelet count have similar incidence for DHF1. Patients with low PCV have a higher percentage incidence of getting DF with compared to other categories while patients with moderate PCV have a higher percentage of DHF1 and DHF2.

A high amount of rainfall in the current month has a higher percentage incidence of DF when compared to other levels of rainfall in the current month, whereas low amount of rainfall in the current month has a higher incidence of DHF2. Moreover, for low and moderate levels of rainfall in the current month there is no significant difference in the percentage incidences for DHF1. When considering the impact of rainfall of the previous month, presence of a high amount of rainfall in the previous month has a higher percentage incidence of DF with compared to other levels of rainfall in the previous month, while a moderate amount of rainfall in the previous month has a higher percentage incidence for DHF1 as well as for DHF2. A high amount of rainfall in two months before has a higher percentage incidence of getting DF with compared to other levels

of rainfall in two months before, while a moderate rainfall in the two months before has a higher percentage incidence of DHF2 cases with compared to other levels of rainfall in two months before. However, the percentage incidence is similar for DHF1 among all levels of rainfall in two months before. When concerning the impact of temperature in the current month, presence of a low temperature in the current month has a higher percentage incidence of DF with compared to other categories of temperature in the current month, whereas a moderate and high amount of temperature in the current month have a similar percentage of incidence for DHF2 cases with compared to other categories of temperature in the current month. There is no significant difference in the percentage incidence for DHF1 among all categories of temperature in the current month. A low amount of temperature in the previous month has a higher percentage incidence for DF when compared to other categories of temperature in the previous month, while a high amount of temperature in the previous month has a higher percentage incidence for DHF2 when compared to other categories of temperature in the previous month. However, there is no significant variation in the percentage incidences for DHF1 among all categories of temperature in the previous month. When the impact of temperature in two months before is considered, a low amount of temperature in two months before has a higher percentage incidence of DF with compared to other categories of temperature in two months before, whereas high amount of temperature in two months before has a higher percentage incidence of DHF2 cases with compared to other categories of temperature in two months before.

For DHF1, there is no significant difference in the percentage incidence for low and moderate amount of temperature in two months before. When considering humidity of the current month, presence of a moderate amount of humidity in the current month has a higher percentage incidence of getting DF when compared to other categories of humidity in the current month, while moderate amount of humidity in the current month has a higher percentage incidence of DHF2 cases when compared to other categories. There is no significant variation in the percentage incidences for DHF1 among all the categories of humidity in the current month. There is similar percentage incidence for low and high amount of humidity in the previous month for DF and DHF1. However, presence of a moderate humidity in the previous month has a higher percentage incidence for DHF2 cases when compared to other categories of humidity in the previous month. When concerning the impact of humidity in two months before, presence of a low amount of humidity in two months before has a higher percentage incidence of DHF1 with compared to other categories of humidity in two months before, while moderate amount of humidity in two months before has a higher percentage incidence for DHF2 when compared to other categories of humidity in two months before. There is a similar percentage incidence for low and high amount of humidity in two months before for DF.

Generally, in each year it could be observed that two peak incidences have occurred within the periods of June-August and September-November for DF. For DHF 1 and 2, however, 2006 has been a high incidence year with one peak around June- August. The rest of the times show no specific pattern for DHF 1 and 2.

Figure 2 depicts the variation of number of dengue fever, dengue hemorrhagic fever 1 and 2 patients within the year for the period of 2006-2008.

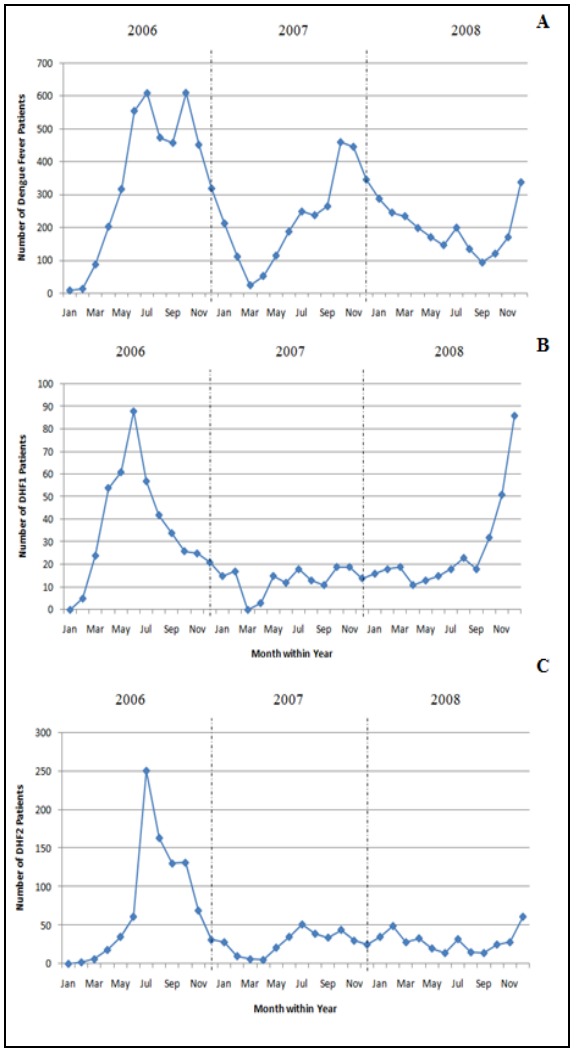

Fig2: *Variation of number of dengue fever (A), dengue hemorrhagic fever 1 (B) and 2 (C) patients within a year*

B. Univariate Analysis

The major objective of the univariate analysis is to evaluate the associations of each variable one at a time with the response variable. This section will describe the results derived in the univariate phase of this study.

When closely looking at the dataset, it can be seen that patients are nested within the district to which her/his residence belongs. Therefore, it is possible to consider the district as a stratification factor for use in the Generalized Cochran Mantel Haenszel test. Table 5 gives the results obtained by applying the correlation unadjusted Cochran Mantel Haenszel (CMH) test and the correlation adjusted Cochran-Mantel-Haenszel-test.

Table5: *Test Results of the Generalised Cochran Mantel Haenszel Test*

		Unadjusted CMH			**Correlation adjusted CMH**		
Variable	**Category**	**CMH**	**d.f**	**p-value**	**Tp**	**d.f**	**p-value**
Sex	Male	1.1792	2	0.5546	1.1565	2	0.5609
	Female						
Age	<20	24.3076	4	6.93e-05	24.2560	4	7.097112e-05*
	20-35						
	>35						
Survival	<8	20.4489	4	0.0004071	19.12230	4	0.0007436303*
	8-9						
	>9						
WBC	Low	4.5849	4	0.3326	4.354847	4	0.3601075
	Moderate						
	High						
Platelets	Low	146.5077	4	<2.2e-16	143.3866	4	5.314709e-30*
	Moderate						
	High						
PCV	Low	37.2051	4	1.634e-07	41.03896	4	2.638362e-08*
	Moderate						
	High						
RF	Low	8.1675	4	0.08563	7.923496	4	0.09442118
	Moderate						
	High						
RF_1	Low	10.5019	4	0.03277	10.30531	4	0.03558715*
	Moderate						
	High						
RF_2	Low	18.9255	4	0.0008129	20.07799	4	0.0004820034*
	Moderate						
	High						
Temp	Low	38.0133	4	1.113e-07	39.81199	4	4.733773e-08*

	Moderate						
	High						
Temp_1	Low	46.9556	4	1.558e-09	47.41399	4	1.250267e-09*
	Moderate						
	High						
Temp_2	Low	33.0587	4	1.162e-06	35.23104	4	4.164387e-07*
	Moderate						
	High						
Humid	Low	53.3621	4	7.157e-11	56.17209	4	1.845312e-11*
	Moderate						
	High						
Humid_1	Low	40.3115	4	3.732e-08	36.13466	4	2.714859e-07*
	Moderate						
	High						
Humid_2	Low	76.6399	4	8.964e-16	70.05392	4	2.21112e-14*
	Moderate						
	High						

** Coefficient is significant at 5% level.*

According to the results obtained it can be noted that there is a significant difference between the correlated unadjusted Cochran Mantel Haenszel test and the correlation adjusted Cochran Mantel Haenszel test statistic. Therefore, this indicates the existence of some correlation between the patients within a district.

Moreover, except 'Sex', 'White Blood Cells' and 'Rainfall' all the other variables are highly significant. This implies that there is a strong association between the type of dengue infection and all the other variables [17].

C. Advanced Analysis

1) Multilevel Multinomial Model

One of the major objectives of this study is to fit a multilevel multinomial logistic regression model to the dataset of interest to identify the factors that differentiate DHF1 and DHF2 from DF. The reason behind the above objective is, to test the claim of many clinicians that there is no major difference between DHF1 and DHF2 categories. It is vital to conduct a study such as this to assess this claim in a statistically meaningful way [2].

The implication of having a between district variance of zero is that a multilevel model is not required as there is no level 2 variation [14]. It would be equivalent to fitting a single level model. Therefore, in order to check the suitability of the multilevel concept it is essential to test the significance of the district level variance.

The DIC value of the random intercept model given in (1) (6242.72) is less than that of fixed effects model (6377.98). This suggests that there is an improvement of 135.26 due to the random intercept model. Thus, it is possible to conclude that, it is appropriate to fit a multilevel multinomial model to analyze the dataset of interest.

The random effect covariance is positive (0.612), hence it indicates that districts with high (low) incidence of DHF1 tends to have high (low) incidence of DHF2 as well [18]. Moreover, this relationship can be seen more clearly when the correlation coefficient is taken into account. It can be clearly seen that there is a high positive correlation between DHF1 and DHF2 (0.750).

The table 6 shows the results from the multinomial multilevel models. It includes the parameter estimate $(\hat{\beta})$ and its standard error $(SE(\hat{\beta}))$ and associated odds ratio (OR) pertaining to the two logits for every variable in the model. The logits, $\log(DHFi/DF)$ where $i = 1, 2$ indicate the log odds of belonging to $DHFi$ category when compared to the DF category.

The odds of getting $DHF1/DF$ for a patient with age less than 20 years is twice as much higher than that for a patient with age limit in excess of 35, whereas the odds of getting $DHF2/DF$ for a patient with age limit less than 20 is approximately 1.7 times higher than that for a patient with age greater than 35. Moreover, considering the impact of variable Sex, it could be observed that the odds ratio was significant only in $DHF2/DF$ category and the odds of getting $DHF2/DF$ for a female is nearly one and half times higher than that for a male patient.

The odds ratios of factor 'Survival time (time spent in the hospital)' indicates that the odds of getting $DHF2/DF$ for a patient with survival time greater than 9 days is nearly 1.3 times higher than that of a patient with survival time less than 8 days. Moreover, it is interesting to note that this variable only affects $DHF2/DF$ category.The calculations of the odds ratios, indicates that, the odds of getting $DHF1/DF$ for a patient with low platelet count is two times higher than that for a patient with high platelet count, whereas the odds of getting $DHF2/DF$ for a patient with low platelet count is three times higher than that for a patient with high platelet count. On the other hand, the odds of getting $DHF1/DF$ for a patient with moderate platelet count is one and half times as much higher than that for a patient with high platelet count, whereas the odds of getting $DHF2/DF$ for a patient with moderate platelet count is nearly twice than that of a patient with high platelet count. One interesting point that could be observed with respect to Packed Cell Volume (PCV) is that, it only affects $DHF2/DF$ category. The odds of getting $DHF2/DF$ for a moderate level of PCV is nearly 1.8 times higher than that of a low PCV level. Moreover, the odds of getting $DHF2/DF$ for a high level of PCV is approximately 1.9 times higher than that of a low PCV level. Therefore, it could be seen that, the odds ratios are approximately the same for moderate and high level of PCV.

Table6: *Parameter Estimates of Multilevel Multinomial Model*

Factor/Covariate	Category	DHF1		DHF2	
		$\hat{\beta}(ste(\hat{\beta}))$	OR	$\hat{\beta}(ste(\hat{\beta}))$	OR
Demographic Variables					
Age	<20	0.806(0.143)	2.2389*	0.570(0.110)	1.7863*
	20-35	0.284(0.151)	1.3284	0.117(0.114)	1.1241
Sex	Female	0.057(0.117)	1.9587	0.382(0.095)	1.4652*
Health and Laboratory Variables					
Survival	8-9	0.145(0.125)	1.1560	-0.006(0.101)	0.9940
	>9	0.228(0.142)	1.2561	0.256(0.112)	1.2918*
Platelet	Low	0.740(0.143)	2.0960*	1.302(0.119)	3.6766*
	Moderate	0.466(0.140)	1.5936*	0.668(0.122)	1.9503*
PCV	Moderate	0.148(0.132)	1.1595	0.569(0.112)	1.7665*
	High	0.104(0.149)	1.1096	0.624(0.123)	1.8664*
Climatological Variables					
RF_1	Moderate	0.424(0.144)	1.5281*	-0.279(0.118)	0.7565*
	High	0.023(0.151)	1.0233	-0.168(0.122)	0.8453
Temp	Moderate	0.281(0.180)	1.3244	0.222(0.139)	1.2486
	High	0.483(0.234)	1.6209*	-0.325(0.195)	0.7225
Temp_1	Moderate	0.276(0.195)	1.3178	0.027(0.162)	1.0274
	High	0.431(0.260)	1.5388	0.208(0.216)	1.2312

Temp_2	Moderate	0.100(0.164)	1.1052	0.173(0.147)	1.1889
	High	-0.611(0.210)	0.5428*	0.186(0.183)	1.2044
Humid	Moderate	0.290(0.149)	1.3364	0.544(0.129)	1.7229*
	High	-0.056(0.171)	0.9455	-0.120(0.147)	0.8869
Humid_2	Moderate	-0.530(0.144)	0.5886*	0.663(0.122)	1.9406*
	High	-0.666(0.156)	0.5138*	0.195(0.138)	1.2153
Time Cons		0.000(0.006)	1.0000	-0.020(0.005)	0.9802*
		-3.438(0.354)		-3.519(0.300)	
Deviance Information Criteria (DIC) : 6242.72					

When the impact of rainfall in the previous month is considered, the odds of getting $DHF1/DF$ for moderate amount of rainfall in the previous months is nearly one and half times higher than that of a low rainfall level, whereas the odds of getting $DHF2/DF$ for moderate amount of rainfall in the previous month is approximately 0.76 times than that of a low rainfall level. The odds of getting $DHF1/DF$ for a high level of temperature in the current month is approximately one and half times higher that of a low temperature. It was not possible to observe such relationship for the other category. It is interesting to note that the temperature of two months before only affects the$DHF1/DF$. Moreover, the odds of getting $DHF1/DF$ when there is a high amount of temperature two months before is nearly 0.5 times than that of a low temperature level. The humidity of the current month only affects the $DHF2/DF$ category. Furthermore, the odds of getting $DHF2/DF$ for a moderate level of humidity in the current month is nearly 1.7 times higher than that of a low humidity level. Considering the impact of humidity of the two months before, it could be seen that, the odds of getting $DHF1/DF$ for moderate amount of humidity in the two months before is approximately 0.6 times than that of a low humidity, whereas the odds of getting $DHF2/DF$ for a moderate amount of humidity in the two months before is approximately two times higher than that of a low humidity. On the other hand, the odds of getting $DHF1/DF$ for high amount of humidity in the two months before is approximately 0.5 times than that of a low humidity. However, this odds ratio was not significant for $DHF2/DF$.

The covariate 'time' only affects $DHF2/DF$ category. The odds of getting $DHF2/DF$ decreases at a rate of 0.98 for one unit increment in the variable.

2) Residual Analysis

Figure 3 gives the Caterpillar Plot, Normal Plot and Anderson Darling Test results for the two logits of the multilevel multinomial model.

According to the Caterpillar plot of the 1st logit (i.e. $logit\left(\frac{DHF1}{DF}\right)$), only 4 districts exhibit 95% confidence intervals that does not include zero. Namely, they are districts 7 (Kalutara), 3 (Colombo), 13 (Puttalam) and 9 (Kegalle) respectively. This indicates that these districts are different in behavior from the other districts. The Normal Plot depicts that most of the points are approximately distributed around the $45°$ axis. Therefore, it is reasonable to assume that the normality assumption of the residuals is satisfied

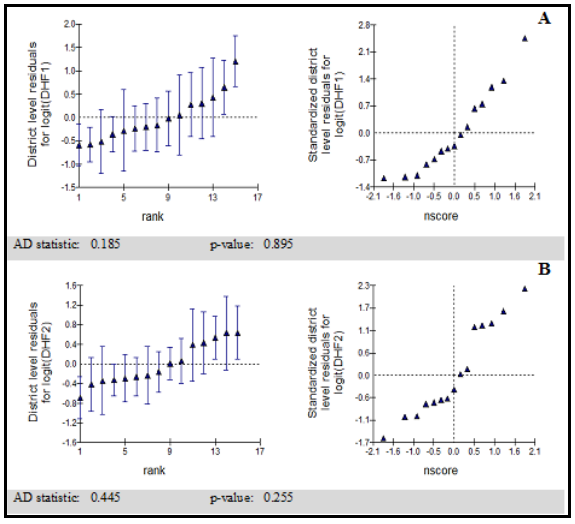

Fig3: *Caterpillar Plot, Normal Plot and Anderson Darling Test results of the (A), Caterpillar Plot, Normal Plot and Anderson Darling Test results of the (B)*

The p-value of the Anderson Darling Test statistic exceeds 0.05, since $p-value(0.895) > 0.05$, it is not

feasible to reject the null hypothesis of residuals are normally distributed at 5% level of significance. Thereby, it concludes that the district level residuals of the 1st logit adhere to a normal distribution.

The Caterpillar plot of the 2nd logit (i.e. $logit\left(\frac{DHF2}{DF}\right)$), only 3 districts exhibit 95% confidence intervals that does not include zero. Namely, they are districts 1 (Anuradhapura), 9 (Kegalle), 10 (Kurunegala), respectively. The Normal Plot shows that the points are slightly deviated from the $45°$ axis.

The p-value of the Anderson Darling Test statistic exceeds 0.05, since $p-value(0.255) > 0.05$. Thus it is not feasible to reject the null hypothesis that the residuals are normally distributed at 5% level of significance. Thereby, it concludes that the district level residuals of the 2nd logitfollows a normal distribution.

3) Predictive Accuracy of the Multilevel Multinomial Model

Table 7 gives the classification table corresponds to the fitted multilevel multinomial model. In order to perform this task, $\pi_{ij}^{(1)}, \pi_{ij}^{(2)}$ and $\pi_{ij}^{(3)}$ are calculated for all i, j as given in (2). Then the patient was allocated to the category having highest probability. According to the calculation it could be concluded that the predictive accuracy of the main effects model is 86.02%

Table7: *Classification Accuracy of the Multinomial Model*

		True Category			Total
		DF	**DHF1**	**DHF2**	
Predicted Category	DF	3944	199	308	4451
	DHF1	14	148	127	289
	DHF2	8	51	260	319
Total		3966	398	695	5059

$$Accuracy = \frac{3944 + 148 + 260}{5059} = \frac{4352}{5059} = 0.8602$$

4. Discussion

The central aim of this study is to investigate the association between different types of dengue infections in Sri Lanka and demographic, health and laboratory and climatological variables using the multilevel multinomial model.

According to the multilevel multinomial model, females are more prone to suffer from Dengue Hemorrhagic Fever 2 than Dengue Fever when compared to males and there is no significant difference between Dengue Hemorrhagic Grade 1 and Dengue Fever with respect to gender. However, when DHF1 and DHF2 categories are combined together as DHF, females have more chance of getting Dengue Hemorrhagic Fever than Dengue Fever with compared to males [19]. Furthermore, patients under age of 20 years have more chance of getting Dengue Hemorrhagic Grade 1 or Grade 2 when compared to patients with Dengue Fever. A previous study also found that the younger children (less than 15 years) are more prone to DHF than older children and adults who are more likely to get DF [20].

White Blood Cells count is not a good indicator to differentiate Dengue Hemorrhagic Fever 1 and 2. As in the Guidelines on Management of Dengue Fever & Dengue Haemorrhagic Fever in Children and Adolescents by Health Ministry of Sri Lanka, a lower WBC level was seen in both DF and DHF patients [2]. Patients with Dengue Hemorrhagic Fever Grade 1 and Grade 2 have high chance of reduced platelet counts than patients with Dengue Fever. Literature shows that there is a significant negative correlation between the disease severity (grade of DHF) and platelet counts [21]. The Packed Cell Volume of the Dengue Hemorrhagic Fever 2 patients increases more than that of Dengue patients. However, there is no significant difference in the Packed Cell Volume of patients with Dengue Fever and Dengue Hemorrhagic Fever 1. This result could be further verified as PCV > 50 was significantly associated with the presence of bleeding manifestation [21]. Even though, literature shows that, on average, patients with DHF have a more severe illness and may require hospitalization for a more extended period of time in comparison to DF [22], when it is sub-divided as DHF1 and DHF2, the above relationship was not seen in both the categories separately. The survival time of patients with Dengue Hemorrhagic Fever 2 is higher than that of patients with Dengue Fever. This would imply that DHF2 patients have to stay quite a longer time in the hospital than patients with DF. However, there is no significant difference in the survival time of patients with Dengue Fever and Dengue Hemorrhagic Fever 1.

In addition to the health and laboratory variables, presence of moderate rainfall in the previous month favors having Dengue Hemorrhagic Fever 1 than Dengue Fever, whereas, it reduces the presence of Dengue Hemorrhagic Fever 2 than Dengue Fever. Moreover, presence of a moderate temperature in the current month favors having Dengue Hemorrhagic Fever 1 than Dengue Fever. On the other hand, presence of a high temperature in two months before would reduce the chance of having Dengue Hemorrhagic Fever 1 than Dengue Fever, whereas it increases the presence of Dengue Hemorrhagic Fever 2 than Dengue Fever. Existence of a moderate humidity in two months before would reduces the chance of having Dengue

Hemorrhagic Fever 1 than Dengue Fever, whereas, it increases the presence of Dengue Hemorrhagic Fever 2 than Dengue Fever. Moreover, high level of humidity in two months before would reduce the chance of having Dengue Hemorrhagic Fever 1 than Dengue Fever.

No studies could be found for climatic factors affecting the type of dengue namely, DHF1 and DHF2. However, the effect of climatic parameters, rainfall, temperature and humidity and their lag-terms have been evaluated on the overall dengue incidence [23]. Moreover, according to the Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control by WHO [24], several factors can influence the virus transmission including environmental and climatic factors, and climate directly influences the biology of the vector. This implies that the behavior of climate is different for DHF1 and DHF2. However, no direct studies have been done to quantify the impact of the climatic variables on the type of dengue, DHF1 and DHF2.

There is a decreasing trend of patients with Dengue Hemorrhagic Fever 2 than Dengue Fever over time. However, such relationship is not evident for Dengue Hemorrhagic Fever 1 with reference to Dengue Fever.

Some problems arose in this study. The first significant problem was encountered during the data preparation stage. The original dataset consisted of missing observations. Even though techniques such as missing value imputation are available, percentage of missing values relative to the available data is small. Therefore, it was decided to list wise delete records with incomplete observations for all the variables except the climatic variables. Not all 25 districts were used due to small number of observations in some of the districts. Furthermore, the entire Northern Province was unrepresented, therefore could not generalized the findings to the entire country.

As we had sufficient amount of data for this study, missing value imputation was not performed. However, for small datasets with missing values, missing value imputation plays a very important role and this can be studied as further work. Furthermore, as the civil war is over data is available for Northern Province, thus this study could be extended to represent the whole country.

5. Conclusion

Finally, it would be possible to conclude that different factors affect $DHF1/DF$ and $DHF2/DF$ categories in a different way.This may be due to the fact that dengue is caused by four different virus serotypes and each may need different conditions and have different incubation periods. This could be further verified as DENV-2 was significantly associated with increased odds of DHF 1 compared to DF, when using DENV-1 as the reference [25].

6. Acknowledgements

Special thanks must be made to the Epidemiology Unit, Sri Lanka for providing the necessary information required for this study. Moreover, we would like to acknowledge Dr. S. Pieris in the Epidemiology Unit for providing necessary background knowledge required to gain a better understanding regarding the variables used in the study. In addition, we would like to acknowledge Prof. H. Goldstein, and Prof. B. Browne at the Centre for Multilevel Modeling, University of Bristol for their assistance regarding the multilevel models.

7. References

[1] Kanakaratne N, Wahala WMPB, Messer WB, Tissera HA, Shahani A, Abeysinghe N, de-Silva AM, Gunasekera M: Severe Dengue Epidemics in Sri Lanka 2003-2006. *Emerg Infect Dis* 2009, 15:192-199.

[2] Guidelines on Management of Dengue Fever & Dengue Haemorrhagic Fever In Children and Adolescents [http://www.epid.gov.lk/web/attachments/article/141/Guidelines%20on%20Management%20of%20DF%20and%20DHF%20in%20Children%20and%20Adolescents.pdf].

[3] First accurate predictive model to differentiate between dengue fever and dengue hemorrhagic fever[http://www.news-medical.net/news/20120217/First-accurate-predictive-model-to-differentiate-between-dengue-fever-and-dengue-hemorrhagic-fever.aspx].

[4] Current situation & epidemiology of dengue in Sri Lanka [http://www.healthedu.gov.lk/web/images/pdf/msp/current_situation_epidemiology_of_dengue.pdf]

[5] Wilson ME: Microbes, Vectors of Infection, Environment and Climate, Malaria, Cholera, HIV/AIDS, Tuberculosis, Dengue, Coccidiodomycosis. [http://www.libraryindex.com/pages/3241/Disease.html].

[6] Watt DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A: Effect of Temperature on the Vector Efficiency of AdesAegypti for Dengue 2 virus. *Am J Trop Med Hyg* 1987, 36:143-152.

[7] Pathirana S, Kawabata M, Goonatilake R: Study of potential risk of dengue disease outbreak in Sri Lanka using GIS and Statistical Modelling.*Journal of Rural and Tropical Public Health* 2009, 8:8-17.

[8] Goldstein, H:*Multilevel Statistical Models.*London: Institue of Education; 1999.

[9] Hongisto K: Dengue: Stooping a Potentially Deadly Threat in Sri Lankan. [http://www.childfund.org/media/article.aspx?id=1107]

[10] Perera HL, Sooriyarachchi MR: Fitting correlation adjusted generalized linear models to clustered dengue data measured over time.*Sri Lankan Journal of Applied Statistics* 2008,9:159-175.

[11] Yozgatlıgil C, Aslan S, İyigün C, Batmaz I: Comparison of missing value imputation methods in time series: The case of Turkish meteorological data. [http://rd.springer.com/article/10.1007/s00704-012-0723-x]

[12] Gelman A, Park DK: Splitting a predictor at the upper quarter or third and the lower quarter or third.*The American Statistician* 2008, 63:1-8.

[13] Zhang J, Boos DD: Generalized Cochran-Mantel-Haenszel Test Statistics for Correlated Categorical Data. *Communications in Statistics* 1997, 26:1813-1837.

[14] Rashbash J, Steele F, Browne WJ, Goldstein H: A User's Guide to MLwiN Version 2.10. Centre for Multilevel Modelling, University of Bristol, Bristol; 2008.

[15] Browne WJ: MCMC Estimation in MLwiN. London: Institute of Education; 2009.

[16] Stow CA, Lamon EC, Qian SS, Soranno PA, Reckhow KH:Bayesian hierarchical/multilevel models for inference and prediction using cross-system lake data. InReal World Ecology: Large-Scale and Long-Term Case Studies and Methods. Edited by Miao S, Carstenn S, Nungesser M. New York: Springer; 2009:101-136.

[17] De Silva DBUS, Sooriyarachchi MR: Generalized Cochran Mantel Haenszel test for multilevel correlated categorical data: an algorithm and R function.*J Natn Sci Foundation Sri Lanka* 2012, 40:137-148.

[18] Durrant GB, Steele F: Multilevel modeling of refusal and non-contact in household surveys: evidence from six UK Government surveys.*JR Statist Soc* 2009, 172:361-381.

[19] Potts JA, Thomas SJ, Srikiatkhachorn A, Supradish P, Li W, Nisalak A, Nimmannitya S, Endy TP, Libraty DH, Gibbons RV, Green S, Rothman AL, Kalayanarooj S: Classification of Dengue Illness Based on Readily Available laboratory Data.*Am J Trop Med Hyg* 2010, 83:781-788.

[20] Gubler DJ: Dengue and dengue hemorrhagic fever.*Clinic Microbial Rev* 1998, 11:480-496.

[21] Malavige GN, Velathanthiri VGNS, Wijewickrama ES, Fernando S, Jayaratne SD, Aaskov J, Seneviratne SL: Patterns of disease among adults hospitalized with dengue infections.*Q J Med* 2006, 99:299-305.

[22] Potts JA, Rothman AL: Clinical and laboratory features that distinguish dengue from other febrile illnesses in endemic populations.*Trop Med Int Health* 2008, 13:1328-1340.

[23] Nakhapakorn K, Tripathi NK: An information value based analysis of physical and climate factors affecting dengue fever and dengue haemorrhagic fever incidence.*Int J Health Geogr* 2005, 4:1-13.

[24] Dengue: guidelines for diagnosis, treatment, prevention and control [http://whqlibdoc.who.int/publications/2009/9789241547871_eng.pdf]

[25] Fried JR, Gibbons RV, Kalayanarooj S, Thomas SJ, Srikiatkhachorn A, Yoon IK, Jarman RG, Green S, Rothman AL, Cummings DA: Serotype-Specific Differences in the Risk of Dengue Hemorrhagic Fever: An Analysis of Data Collected in Bangkok, Thailand from 1994 to 2006.*PLosNeql Trop Dis* 2010, 4: e617.