RESEARCH ARTICLE

Correlational study of air pollution-related diseases (asthma, conjunctivitis, URTI and dengue) in Johor Bahru, Malaysia

Aaishah Radziah Binti Jamaludin [a], Fadhilah Yusof [b,*], Rahmah Mohd Lokoman [c], Zainura Zainoon Noor [d], Noreliza Alias [e], Nurazimah Mohd Aris [f]

[a,b,c] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
[d] Centre for Environmental Sustainability and Water Security, Block C07, Level 2, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[e] Department of Hydraulics and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[f] Department of Health Johor Bahru, Jalan Abdul Samad,80100 Johor Bahru, Johor

* Corresponding author: fadhilahy@utm.my

Abstract

Four pollution related diseases, namely asthma, conjunctivitis, URTI and dengue will be studied in terms of their trend, behaviour and association with influential factors such as air pollution and climate variables. Two methods were chosen; Poisson Generalized Linear Model and Negative Binomial Model. These methods were used to determine the association between the diseases and their influential factors. This study shows that Sulphur Dioxide (SO2) is the most abundant source that contributes to the diseases. Therefore, the local authorities such as the Department of Environment need to reinforce the law in planning and monitoring the SO2 sources which are produced from fuel combustion in mobile sources and motor vehicles.

Keywords: Asthma, conjunctivitis, URTI, air pollution, negative binomial model, poisson generalized linear model

INTRODUCTION

Many of us experience some kind of air pollution-related symptoms such as watery eyes, coughing or wheezing. Even for healthy people, polluted air can cause respiratory irritation or breathing difficulties during exercise or outdoor activities. The actual risk depends on our current health status, the pollutant type and concentration, and the length of our exposure to the polluted air. According to Kjellstrom et al. [11], the particulate matter that may come from the combustion of biomass, fossil fuel, industry, motor vehicle engines and cigarette smoke can cause the Upper Respiratory Tract Irritation (URTI) and increased cardiorespiratory diseases. Among all the natural and anthropogenic sources of air pollution, the combustion of fossil fuels is a major contributor in urban and industrialized societies [2]. Tjønneland et al. [22] found that long-term exposure to traffic-related air pollution may contribute to the development of Chronic Obstructive Pulmonary Disease (COPD) with possibly enhanced susceptibility in people with diabetes and asthma. Yin et al. [27] estimated the associations between PM10 and mortality using daily time series data from 38 China cities and the analysis showed that PM10 was robustly associated with worse health outcomes in most cities. They found that PM10 had a larger impact on people with cardiorespiratory diseases, the old population, and the female population.

One of the common diseases related to air pollution is asthma disease. Asthma is a chronic, occasionally debilitating inflammatory disease of the airways that may be caused by air pollution from cars, factories or power plants. Asthma causes recurring periods of wheezing, chest tightness, shortness of breath, and coughing that are commonly associated with allergen exposures, air pollution, cigarette smoking and noxious particle insults [30]. The association between air

pollution and asthma disease has been well-established, particularly in the countries with rapid urbanization and industrialization [9, 17, 18, 23, 30] and Poisson regression generalized model was used broadly by many researchers [1, 3, 20, 26, 29].

The second type of disease that will be highlighted in this study is Conjunctivitis. Conjunctivitis is an inflammation of the conjunctiva secondary to a bacterial or viral infection [19]. In this paper, we assess the relationship between air pollutants and climates conditions with outpatient visits for allergic conjunctivitis in Johor Bahru's public clinics. Among the previous studies that applied regression analysis in conjunctivitis data are; Li et al. [12] who found that the air quality index at a two-day lag affected the probability of conjunctivitis diseases and it was greatly influenced by low temperature (cold season). Silvester et al. [19] used linear regression models to determine whether there was a change in the resistance pattern of bacteria isolated from cases of conjunctivitis. They found that resistance to chloramphenicol has remained stable since being made available over the counter. Hong et al. [7] applied univariate generalized least squares (GLS) model to evaluate the effect of the environmental factors on the outpatient attendance of allergic conjunctivitis. Most of the pollutants were found to show significant time-lag effects in the third week with conjunctivitis cases.

Another common problem is Upper respiratory tract infection (URTI) which may be exacerbated by air pollution. The previous studies that applied regression analysis are; Darrow al. [5] who used Poisson generalized linear models to investigate short-term changes in ambient air pollutant concentrations in relation with URTI problem among children aged 0 to 4 years old in Atlanta, Georgia. They found that three-day moving average concentrations of ozone, nitrogen dioxide, and the organic carbon fraction of particulate matter less than 2.5 µm in diameter (PM2.5) were associated with the number of URTI

patient. Ozone associations were strongest and were observed at low (cold-season) concentrations. Tam et al. [21] applied the generalized additive models to examine the association between air pollution and the daily number URTI cases general outpatient clinics (GOPCs) in Hong Kong. Significant associations were found between the daily number of consultations due to URTIs in GOPCs and the concentrations of air pollutants, implying that air pollution incurs a substantial morbidity and increases the burden of primary health care services. Liu et al. [13] analyzed the relationship between levels of air pollution and the number of children hospitalizations because of URTI in Shenmu, China. They built a generalized additive Poisson regression model and the linear fitting models of the air pollutant concentrations and meteorological factors. The results showed that the primary air pollutant in Shenmu is PM10 and its Pearson correlation coefficient with Air Pollution Index (API) is 0.917.

Several studies have been conducted to identify the relationship between the air pollution and Dengue cases. Wilder-Smith et al. [24] has conducted a study based on hypotheses proposed by Massad et al. [15] that the haze can reduce the spread of Dengue. Their studies have been conducted in Singapore and the haze is measured as an air pollution index (IPU). However, Wilder-Smith et al. [24] found that there was no significant relationship between the haze and Dengue cases. Meanwhile, Olívêr et al. [16] conducted a univariate correlation analysis between bush fires and Dengue cases in four states in Brazil for the period from 2000 to 2009. Their study found that bush fires and Dengue cases in Brazil had a negative correlation. Various studies have been conducted to determine the correlation between climate variables such as rain, temperature and humidity with Dengue cases. Cheong et al. [4] have reviewed climate impacts on Dengue diseases in Selangor, Kuala Lumpur and Putrajaya, Malaysia, from 2008 to 2010. They assessed the relationship between climate variables: temperature, humidity, precipitation and wind speed with daily dengue cases. Their study found that Dengue cases had a positive correlation with temperature and precipitation, and negative correlation with wind speed. However, there is no significant relationship between dengue and humidity cases. Horta et al. [8] studied climate impact on Dengue disease in Coronel Town Fabriciano, Brazil from 2004 to 2010. Their study found that Dengue cases had a positive correlation with temperature and precipitation. However, there is no significant relationship between dengue and humidity cases. Xu et al. [26] have reviewed the impact of climate on Dengue disease in Singapore from 2001 to 2009. Their study found that only temperature has a positive correlation with Dengue cases.

In this study we apply Poisson Generalized Linear Model (PGLM) and Negative Binomial model (NBM) to determine the association between the diseases and their influential factors since these methods are appropriate to entertain count data and they are also the most cited methods in many studies. It is important to study the behaviour of pollution related diseases such as asthma, conjunctivitis, URTI and dengue fever in Johor Bahru to identify the pattern and trend of the diseases and the association with their influential factors such as air pollution and climates variables. The correlational study between the diseases and the climate variables such as temperature, rainfall, and relative humidity also will be determined.

STUDY AREA

Johor Bahru is the capital of the Malaysian state of Johor which sits at the southern part of the Peninsular Malaysia (Figure 1). Johor Bahru has a population of 497,097, and its metropolitan area is the third largest in Malaysia and also the southernmost city in the Peninsular Malaysia. Originally, the city area was only 12.12 km^2 (4.68 sq mi) in 1933 before being expanded to over 220 km^2 (85 sq mi) in 2000. The city has an equatorial climate with consistent temperatures, a considerable amount of rain, and high humidity throughout the course of the year. Temperatures range from 25.5 °C (77.9 °F) to 27.8 °C (82.0 °F) with an annual rainfall of around 2,000 mm (79 in), mostly from November until February. Although the climate is relatively uniform, it can change through the monsoon seasons with variation of wind speeds and direction, cloudiness, and

wet and dry seasons throughout the year. There are two monsoon periods every year, the first one happens between December and February, and is known as north-east Monsoon. It is characterised by heavy rains and winds from the north-east. The second one is the south-west Monsoon, characterised by relative dryness with winds driven from the south and south-west. It occurs between June and August. There are two inter-Monsoon periods from March to May and from September to November, which are relatively calm with less rain and weaker winds.

Fig 1 Location map of Johor Bahru, Malaysia (source: http://malaysiamap.facts.co)

As one of the three main urban centres in the Peninsular Malaysia (Kuala Lumpur and Penang being the other two), Johor Bahru is an important industrial, logistical and commercial centre. Its major industries include electronics, resource and petrochemical refinery and shipbuilding. The heavy industrial areas are Pasir Gudang and Tanjung Langsat, located east of the metropolitan area. They contain clusters of refineries, chemical processing plants, and shipbuilding factories. Light to medium industrial areas are mainly located north and north-west of the metropolitan area in Tebrau, Tampoi, Senai, Skudai and Kulai. Under the Iskandar Development Region masterplan, Johor Bahru is expected to grow to a large urban area of over 3 million residents by 2025, second in Malaysia only to the Klang Valley. The growth is also expected to spill over into the neighbouring districts of Pontian and Kota Tinggi, creating an urban area that would be even larger than that envisaged in the masterplan, possibly over five times the size of Singapore. Therefore the pollution episodes are commonly associated with rapid industrialization in this area. Thus, it is important to study the mechanism of air pollution in Johor Bahru and its association with asthma, conjunctivitis, URTI and dengue fever.

METHODOLOGY

This study is an ecological regression analysis study, which assessed the relationship between asthma, conjunctivitis, URTI and dengue in Johor Bahru, Malaysia from 1 January 2012 to 31 December 2013. Data were obtained from the Department of Information and Informatics, Johor Bahru District Health Office. Data on the level of environmental pollution such as Ground Level Ozone (GLO), Nitrogen Dioxide (NO2), Particulate Matter (PM10) and Sulfur Dioxide (SO2) were obtained from the Johor Department of Environment (workstation: CA0051). Three climate variables were used in this study, namely rainfall, temperature and humidity. Rainfall data were obtained from the Department of Irrigation and Drainage Malaysia, whereas temperature and humidity data were obtained from the Meteorological Department. Pearson correlation was used to determine the association between disease, pollutants and climatic variables at various lags. It is because there is a possibility of a gap for the effects of pollution on a disease, for example asthma cases occur not only on the same day (lag 0), but may also occur as a result of the

effects of pollution in earlier days (lag 1, lag 2 ...). Lag time was built for lag 0 (without lag) and Lag 1 (first week) to Lag 20 (20 weeks) for each of the pollutants and climatic variables in this study. To estimate the risk of asthma cases, conjunctivitis, URTI and Dengue, PGLM and NBM are used. The computer program used for this analysis is Excel and R software.

Pearson Correlation

The correlation between the variables measured by Pearson correlation, Egghe et al. [6] describes the definition of *r* as:

$$r = \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \sqrt{\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}} \quad (1)$$

where *x* and *y* are two different sets of data. It has a value between 1 and -1 inclusive, where 1 indicates a perfect positive linear correlation, 0 indicates an absence of a linear correlation and -1 indicates a perfect negative linear correlation.

Poisson Generalized Linear Model (PGLM)

In statistics, PGLM is a form of regression analysis used to model count data and contingency tables. This analysis assumes the response variable *Y* has a Poisson distribution and the logarithm of its expected value can be modeled by a linear combination of unknown parameters. This model can be expressed as follow Zeileis et al. [28]:

$$f(y; \mu) = \frac{e^{-\mu} \cdot \mu^y}{y!} \quad (2)$$

where $f(y; \mu)$ is probability density function of poisson distribution, *y* is observation data and μ is mean. The canonical link is $g(\mu)=log(\mu)$. The variance in the Poisson model is identical to the mean, thus the dispersion is fixed at $\varphi = 1$ and the variance function is $V(\mu) = \mu$.

Negative binomial model (NBM)

NBM is often used to model over-dispersed count data; where the observation data, y is assumed to follow a negative binomial distribution. The probability mass function is:

$$f(y; \mu, \theta) = \frac{\Gamma(y + \theta)}{\Gamma(\theta) . y!} \cdot \frac{\mu^y \theta^\theta}{(\mu + \theta)^{y+\theta}} \quad (3)$$

with mean, μ and shape parameter, θ and $\Gamma(.)$ is a gamma function .

RESULTS AND DISCUSSION

Descriptive data and statistical explanation

A set of weekly pollution data related to diseases such as asthma, conjunctivitis, URTI and dengue fever for the period of 2012 to 2013 were used for this study. The total number of cases related to the diseases in these two years can be illustrated in the Figure 2 below:
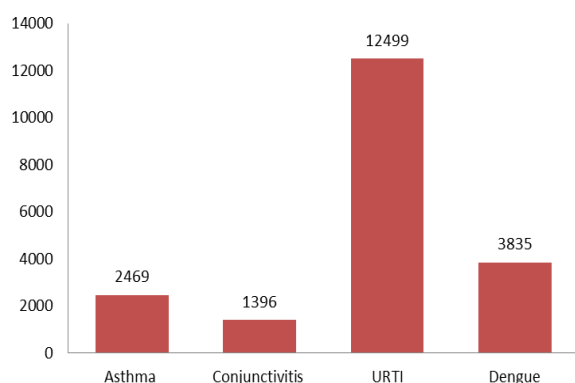

Fig 2 The total number of cases of Asthma, Conjunctivitis, URTI and Dengue were reported in 2012 to 2013 in Johor Bahru.

Figure 2 shows the number of patients for asthma, conjunctivitis, URTI and dengue fever for Johor Bahru in 2012 to 2013. The overall number of asthma patients is 2469 patients, conjunctivitis with 1396 patients, URTI with 12,499 patients and dengue with 3835 patients. URTI recorded the highest number of patients followed by dengue fever, asthma and conjunctivitis.

Time series plot

Figure 3 shows a plot of the time series of Asthma disease, conjunctivitis, URTI and dengue fever for the period of 2012 to 2013. The data does not show any specific trend and it is fluctuating throughout the year. However to confirm the existence of the trends in data series, Mann Kendall trend test will be used to detect any existence of trend in the datasets and it will be elaborated in the next section.
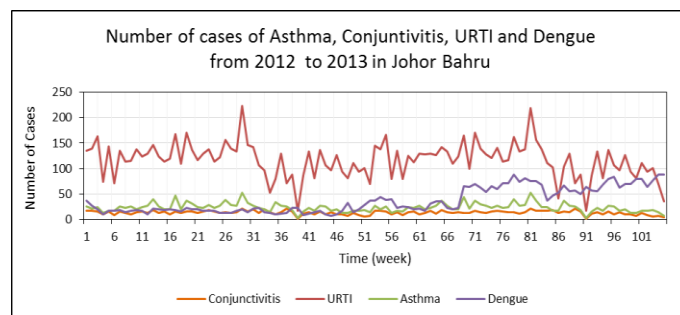

Fig 3 The time series plot of Asthma, conjunctivitis, URTI and Dengue in Johor Bahru

The maximum reading of asthma cases is 53 which were recorded in June 2012 and July 2013. Meanwhile, the minimum reading of asthma cases is 3 recorded in August 2013. The maximum reading of conjunctivitis is 22 which were recorded in June 2012 and June 2013, while the minimum reading of conjunctivitis is 4 which were recorded in December 2013. The maximum reading of URTI is 250 persons that were recorded in June 2012 and June 2013 and the minimum reading of URTI was recorded in August 2012 with a reading of 23. For dengue fever, the maximum reading was recorded in July 2013 with 393 cases and the minimum reading was recorded in November 2012 with 44 cases. The number of dengue cases exhibits an increasing trend starting from 71st week compared to the previous week. Overall, the number of patients of all diseases is higher in the middle of the year between June and July. Dengue and URTI cases show the highest number compared with other diseases. This may be due to the weather conditions and air pollution that occurred during the period.

Descriptive statistics for asthma, conjunctivitis, URTI and dengue fever

Before carrying out the analysis meticulously, it is important to look at the descriptive statistics as it can explain the initial information on each of the variables. Table 1 shows the descriptive statistics on the number of cases for all diseases from 2012 to 2013.

Table 1 Descriptive statistics on the number of cases for all diseases (2012-2013)

Diseases	Mean	Std. deviation	Skewness	Kurtosis	Variance
Asthma	22	9.2388	0.7848	4.4758	85.3547
Conjunctivitis	14	2.8617	0.0085	3.1496	8.1891
URTI	125	32.2328	0.1368	4.2927	1038.952
Dengue fever	37	24.7072	0.6555	1.9251	610.4441

Table 1 lists the basic characteristics of the number of cases for all diseases, including mean, standard deviation, skewness, kurtosis and

variance. The mean number of asthma cases is 22 patients with standard deviation of 9.2388, skewness 0.7847, kurtosis 4.4758 and variance 85.3547. The coefficient of variation for asthma disease is 0.42. The mean number of conjunctivitis cases is 14 patients with standard deviation of 2.8617, skewness 0.0085, kurtosis 3.1496 and variance 8.1891. The coefficient of variation of conjunctivitis disease is 0.204. The mean number of URTI cases is 125 patients with standard deviation of 32.2328, skewness 0.1368, kurtosis 4.2927 and variance 1038.952. The coefficient of variation of URTI disease is 0.252. The mean number of dengue cases is 37 with standard deviation of 24.7072, skewness 0.6555, kurtosis 1.9251dan variance 610.4441. The coefficient of variation of dengue fever is 0.668.

These data clearly show that asthma, URTI and dengue fever experienced over dispersion as the variance is greater than the mean. Conjunctivitis disease shows the lowest of near-zero skewness coefficient of variation compared with other diseases. This initial interpretation will be reconfirmed later with the over dispersion test that will be explained in the next section.

The effect of air pollutants on asthma, conjunctivitis, URTI and dengue fever

A set of weekly variables of air pollutants for the period of 2012 to 2013 was used to identify the association with the asthma, conjunctivitis, URTI and dengue fever. The air pollution parameters considered in this study consist of Ground Level Ozone (GLO), Nitrogen Dioxide (NO_2), Particulate Matter (PM10) and Sulfur Dioxide (SO_2).
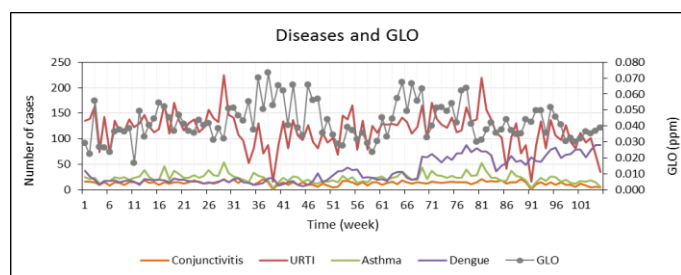

Fig 4 The weekly time series plot of asthma (green line), conjunctivitis (orange line), URTI (red line) and dengue (purple line) with an average weekly reading of GLO (the grey line)

Figure 4 shows a plot of the time series of all diseases and GLO. GLO is not emitted directly into the air, but it is created as a result of a chemical reaction between nitrogen dioxide (NO_2) and volatile organic compounds (VOC) in the presence of sunlight. Breathing the gas ozone can trigger a variety of health problems, especially for children, the elderly, and people of all ages who have lung diseases such as asthma. Ozone can also have harmful effects on sensitive vegetation and ecosystems. However, Figure 4 above showed no close link between ozone and asthma disease, but the plot series of URTI diseases are more likely to show affinity with ozone. To learn more about the relationship between these parameters, Pearson correlation will be used and will be explained in details in the next section.
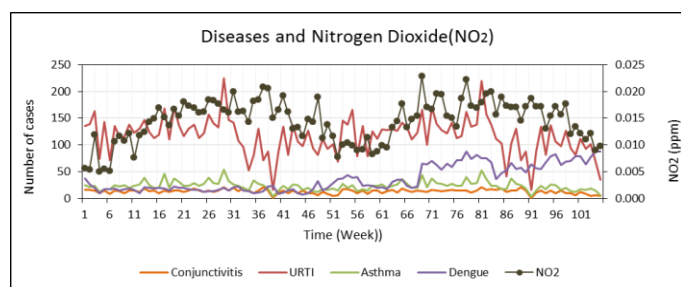

Fig 5 The weekly time series plot of asthma (green line), conjunctivitis (orange line), URTI (red line) and dengue (purple line) with an average weekly reading of Nitrogen Dioxide (black line)

Figure 5 shows a plot of the time series of all diseases and Nitrogen Dioxide (NO_2). Nitrogen Dioxide is known as the most atmospheric pollutants after SO_2. This source is believed to result from a chemical reaction from industry, vehicle exhaust, power plants, heat and furnace burners. Figure 5 shows URTI diseases is likely to be affected by nitrogen dioxide.
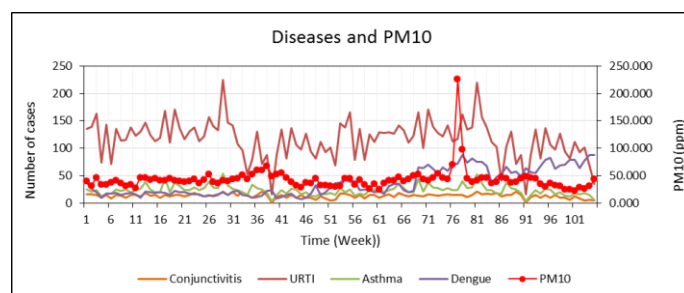

Fig 6 The weekly time series plot of asthma (green line), conjunctivitis (orange line), URTI (red line liver) and dengue (purple line) with an average weekly reading of fine particles, PM10 (red light)

Figure 6 shows a plot of the time series of all diseases and fine particles (PM10). PM10 refers to fine particles with a size of 2.5 to 10 micrometers in diameter, which are the source material from the crushing and grinding operations dust by vehicles on the road. From the plot of the time series above, a consistent pattern of PM10 can be seen and it only increased significantly between weeks 76 and 81. According to Figure 6, PM10 is likely not to have a significant impact on the change of all diseases.
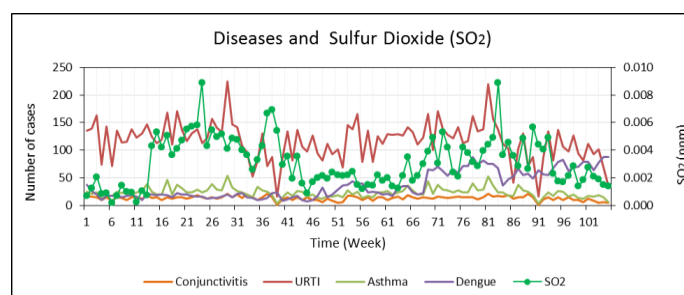

Fig 7 The weekly time series plot of asthma (light green line), conjunctivitis (orange line), URTI (red line) and dengue (purple line) with average weekly readings of sulfur dioxide (dark green line)

Figure 7 shows a plot of the time series for all diseases and sulfur dioxide (SO_2). SO_2 emissions result from fuel combustion sources, refining of crude petroleum, natural metal operations, sulfuric acid plants, paper manufacturing plants, open burning and incineration of municipal waste. It was found that the SO_2 was indirectly affecting the number of patients for all of the above diseases.

The influence of climate on asthma, conjunctivitis, URTI and dengue fever

Johor Bahru has an equatorial climate with consistent temperature, a large amount of rainfall and high humidity throughout the year. The ambient temperature ranges from 25.5°C to 27.8°C with an annual rainfall of about 2,000 mm (79 in), mostly from November to February. Although the climate is relatively uniform, it can be changed through the monsoon with the changes of wind speed, cloudiness, and wet and dry seasons throughout the year. There are two monsoon periods each year, the first takes place between December and February, and is known as the North-East Monsoon. It is characterized by heavy rain and wind from the north-east. The second is the southwest monsoon, characterized by relatively drought with windblown from the south and southwest. It takes place between June and August. There are two inter-monsoon period from March to May and from September to November, with moderate rain and weak

wind. Therefore, this study also focuses on relationship between asthma disease, conjunctivitis, URTI and dengue fever with climate variables such as rainfall, temperature and humidity.
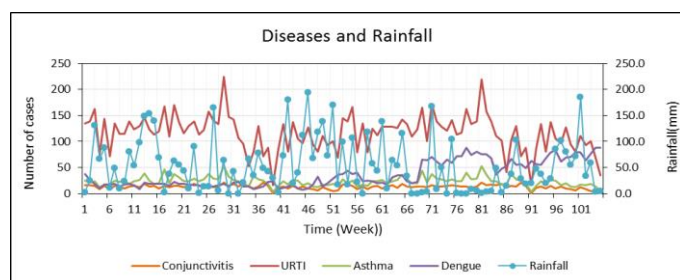

Fig 8 The weekly time series plot of asthma (green line), conjunctivitis (orange line), URTI (red line) and dengue (purple line) with an average weekly rainfall (blue line)

Figure 8 above shows the fluctuation pattern of rainfall throughout the year for 2012 to 2013. It was found that rainfall may have an impact on asthma, conjunctivitis, URTI and dengue diseases. However, a thorough checking will be done through Pearson correlation.
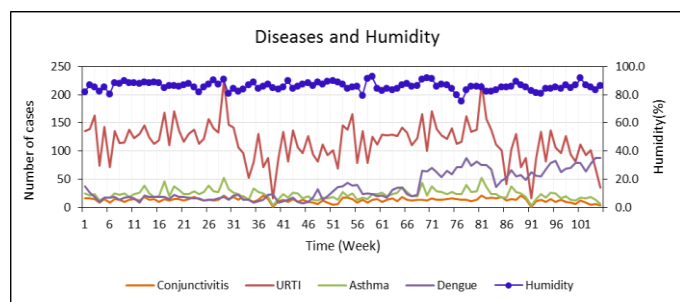

Fig 9 The weekly time series plot of asthma (green line), conjunctivitis (orange line), URTI (red line) and dengue (purple line) with an average weekly humidity (dark blue line)

From Figure 9 above, we can see the trend of humidity in Johor Bahru is consistent throughout the years. This may have little effect on the changes of all diseases.
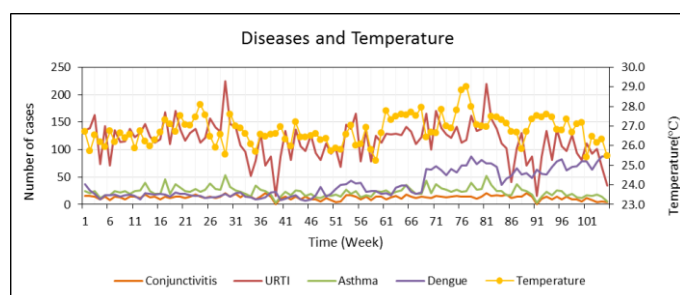

Fig 10 The weekly time series plot of asthma (green line), conjunctivitis (orange line), URTI (red line) and dengue (purple line) with an average weekly temperature (yellow line)

It can be seen in Figure 10 that the fluctuation pattern of the temperature is not significant as compared to the rainfall patterns shown in (Figure 8). Even though many previous studies proved that asthma disease is greatly influenced by the temperature as the asthma diseases is worse during cold weather, this did not happen in Johor Bahru, Malaysia.

The plot of the time series of the cumulative number of weekly cases of asthma, conjunctivitis, URTI and dengue fever with pollutants and climates variables have been designated in Figures 5 to 10. However, the relationship between disease and climate variables and pollutants cannot be clearly identified through a plot of time series. Therefore, Pearson correlation, PGLM and NBM will be carried out to identify a statistically significant relationship between the diseases and their influential factors.

Mann Kendall (MK) trend test

The basic operation of MK test is to use a probability test that will confirm the presence of increasing or decreasing trends in data series. The null hypothesis for this test is that there is no monotonic trend in the series whereas the alternate hypothesis is that a trend exists. This trend can be positive, negative, or non-null. The downward trend can be represented by a score of negative S while the upward trend can be represented by the positive S [10]. Based on the test trend MK shown in Table 2, two variables (Asthma and URTI) shows a decreasing trend with a negative S score value. However, conjunctivitis and dengue fever showed an increasing trend with positive S score. Conjunctivitis and URTI demonstrate a statistically significant trend with a *p*-value greater than 0.05. However, the trend of the disease and Dengue Fever Asthma is not significant because the *p*-value was less than 0.05.

Table 2 Result of MK test

Diseases	MK trend test	
	S-score	*p-value*
Asthma	-1504	0.00033
Conjunctivitis	281	0.49640
URTI	-247	0.55697
Dengue fever	3091	< 2.22x10⁻¹⁶

RESULTS AND DISCUSSION

The study of the relationship between Asthma, URTI, Conjunctivitis, Dengue fever and the air pollution index (API)

Pearson correlation was used to find the relationship between asthma disease and pollutants such as ozone (GLO), nitrogen dioxide (NO2), fine particles (PM10) and sulfur dioxide (SO2). Duration gap used is of the same week (Lag 0) to 20 weeks (Lag 20). The result of Pearson correlation can be summarized in Table 3 below:

Table 3 The Pearson correlation of the diseases with their influential factors at the significant lags

Disease	Influential factor		Significant lag	Pearson correlation
Asthma	Air pollution	GLO	Lag 18	**-0.198**
		NO2	Lag 19	**-0.479**
		PM10	Lag 4	**0.381**
		SO2	Lag 19	**-0.507**
	Climates	Rainfall	Lag 16	**0.285**
		Temperature	Lag 5	**0.396**
		Humidity	Lag 14	**0.209**
Conjunctivitis	Air pollution	GLO	Lag 9	**-0.149**
		NO2	Lag 19	**-0.345**
		PM10	Lag 14	**-0.307**
		SO2	Lag 20	**-0.390**
	Climates	Rainfall	Lag 18	**0.289**
		Temperature	Lag 0	**0.196**
		Humidity	Lag 15	**0.284**
URTI	Air pollution	GLO	Lag 18	**-0.249**
		NO2	Lag 19	**-0.440**
		PM10	Lag 14	**-0.307**
		SO2	Lag 19	**-0.441**
	Climates	Rainfall	Lag 17	**0.231**
		Temperature	Lag 14	**-0.334**
		Humidity	Lag 9	**0.281**
Dengue	Air pollution	GLO	Lag 13	**0.215**
		NO2	Lag 8	**0.344**
		PM10	Lag 18	**0.191**
		SO2	Lag 11	**0.137**
	Climates	Rainfall	Lag 13	**-0.288**
		Temperature	Lag 15	**-0.245**
		Humidity	Lag 5	**0.456**

Table 3 above shows the correlation between the diseases and the influential factors at significant lags. The significant correlation was selected based on Pearson correlation from lag 1 (week 1) until lag 20 (week 20). The air pollution has the effect on the asthma cases in which we can see that the PM10 reading will influence the cases after 4 weeks (lag 4) which is the fastest effect among the other parameter follow by GLO (lag 18), NO2 (lag 19) and SO2 (lag 19). In terms of climate variables, it was found that asthma cases were influenced by the previous rainfall reading at 16 weeks before (lag 16). Meanwhile, the previous temperature reading at 5 weeks before (lag 5) will effected the asthma cases. It took 14 weeks (lag 14) for humidity to trigger the cases. Among the three climate variables, temperature gave the fastest effect to the asthma cases. Besides the asthma cases, from the table we can see other diseases like conjunctivitis, URTI and dengue and their correlation with the pollutants and the climate variables. Their correlation was found at various lags. These results will be used in the modeling approach that will be explained in the next section.

The selection of the model

Based on the previous study, the appropriate model for count data that is widely used by many researchers is Poisson model. However to reconfirm whether Poisson model is suitable for all diseases, an over-dispersion test was done. By looking at descriptive statistics, asthma and URTI data shows over dispersion as the variance of the data is greater than the mean whereas conjunctivitis exhibits under dispersion with the variance being less than the mean (refer Table 1). Therefore, further checking is required to reconfirm the selection of the model. Hence, an over-dispersion test has been done on PGLM of each disease to check whether the model is adequate or not and the result is as illustrated in the table below:

Table 4 Over-dispersion result for all diseases

Diseases	Estimate dispersion (φ)	*p*-value
Asthma	1.7206	0.002522
Conjunctivitis	0.9446	0.6551
URTI	7.1596	1.458e-07
Dengue	10.8173	5.027e-10

Asthma, URTI and dengue have shown the evidences of over dispersion (φ is estimated to be 1.72, 7.16 and 10.82 respectively) which express strongly against the assumption of equi-dispersion (i.e. $\varphi = 0$). Meanwhile conjunctivitis show the data is under dispersion with φ equal to 0.94 and the *p*-value is greater than 0.05. Therefore, conjunctivitis disease is appropriate with PGLM while the rest of the data is not appropriate. The best approach to treat the over dispersed data is through NBM. It can be considered as a PGLM since it has the same mean structure as PGLM but it has an extra parameter to model the over-dispersion. If the conditional distribution of the outcome variable is over-dispersed, the confidence intervals for the NBM are likely to be narrower as compared to those from a PGLM.

Poisson Generalized Linear Model (PGLM) and Negative Binomial model (NBM) for diseases with their influential parameter

A more formal way to accommodate over-dispersion in a count data regression model is to use a negative binomial model. Therefore, the data that possess over dispersion was modeled with negative binomial model. Two models have been selected in this study namely; Poisson Generalized Linear Model (PGLM) and Negative Binomial model (NBM). PGLM was used to model conjunctivitis disease due to its appropriateness whereas NBM was used to model Asthma, URTI and dengue fever since these data experienced over-dispersion. The results of the model estimated parameters for each disease is displayed in Table 5.

Table 5 Summary of PGLM and NBM for Asthma, Conjunctivitis and URTI

Model		Coeff.	Std. error	*p*-value
PGLM-Conjunctivitis	Intercept	0.4535	1.584	0.7747
	GLO (Lag$_9$)	2.602	2.327	0.2636
	NOX (Lag$_{19}$)	-0.2254	8.567	0.9790
	PM10 (Lag$_{14}$)	-4.260×10^{-3}	2.011×10^{-3}	**0.0341**
	SO2 (Lag$_{20}$)	-47.50	20.24	**0.0189**
	Rainfall (Lag$_{18}$)	6.331×10^{-4}	5.196×10^{-4}	0.2230
	Humidity (Lag$_{15}$)	1.940×10^{-2}	1.020×10^{-2}	0.0572
	Temperature (Lag$_0$)	3.100×10^{-2}	4.267×10^{-2}	0.4675
NBM-Asthma	Intercept	-1.369	1.360	0.3139
	GLO (Lag$_{18}$)	-2.604	2.452	0.2882
	NOX (Lag$_{19}$)	-14.44	10.36e	0.1633
	PM10 (Lag$_4$)	2.451×10^{-3}	1.282×10^{-3}	0.0559
	SO2 (Lag$_{19}$)	-52.10	22.00	**0.0179**
	Rainfall (Lag$_{16}$)	9.332×10^{-4}	5.004×10^{-4}	0.0622
	Humidity (Lag$_{14}$)	2.104×10^{-2}	9.526×10^{-3}	**0.0272**
	Temperature (Lag$_5$)	1.114×10^{-1}	4.487×10^{-2}	**0.0131**
NBM-URTI	Intercept	6.479	1.276	3.79×10^{-7}
	GLO (Lag$_{18}$)	-1.591	2.350	0.49837
	NOX (Lag$_{19}$)	0.4570	10.00	0.96355
	PM10 (Lag$_{14}$)	-3.680×10^{-3}	1.621×10^{-3}	**0.02319**
	SO2 (Lag$_{19}$)	-58.63	21.34	**0.00600**
	Rainfall (Lag$_{17}$)	6.984×10^{-5}	5.217×10^{-4}	0.89350
	Humidity (Lag$_9$)	2.896×10^{-2}	9.180×10^{-3}	**0.00161**
	Temperature (Lag$_{14}$)	-0.1435	4.396×10^{-2}	**0.00110**
NBM-Dengue	Intercept	-2.5799	2.663295	0.3327
	GLO (Lag$_{13}$)	5.200250	4.993259	0.2977
	NOX (Lag$_8$)	29.109237	18.484418	0.1153
	PM10 (Lag$_{18}$)	0.003768	0.002620	0.1504
	SO2 (Lag$_{11}$)	-32.776176	35.713633	0.3587
	Rainfall (Lag$_{13}$)	-0.001098	0.001133	0.3327
	Humidity (Lag$_{15}$)	-0.0413	0.019263	**0.0319**
	Temperature (Lag$_5$)	0.3372	0.086335	**9.38×10^{-5}**

For PGLM-Conjunctivitis model, only PM10 and S02 are significant with respect to $\alpha = 0.05$. So the rest of the parameters will not be included in the model. For NBM-Asthma model, SO$_2$, humidity and temperature are found to be significant since the *p*-value is less than 0.05. Thus the rest of the variables will not be included in the model. For NBM-URTI model, PM$_{10}$, SO$_2$, humidity and temperature are significant since the *p*-value is less than 0.05. The rest of the parameters will not be included in the model. For NBM-Dengue, all the values are not significant except humidity and temperature since the *p*-value is less than 0.05. Therefore, only humidity and temperature will be considered in the model. The appropriateness of the model will be evaluated in the next section. One advantage of the negative binomial model is that it is associated with a formal likelihood so that information criteria are readily available. Furthermore, the expected number of zeros can be computed from the

fitted densities [23]. The models constructed can be summarized as follow:

Table 6 Model's equation

Model	Model's equation
PGLM-Conjunctivitis	$Log(Conjunctivitis) = 0.4535 - 4.260 \times 10^{-3}(PM10\ Lag_{14}) - 47.50\ (SO_2Lag_{20})$
NBM-Asthma	$Log(Asthma) = -1.369 - 52.10\ (SO_2\ Lag_{19}) + 2.104 \times 10^{-2}(Humidity\ Lag_{14}) + 1.114 \times 10^{-1}(Temperature\ Lag_5)$
NBM-URTI	$Log(URTI) = 6.479 - 3.680 \times 10^{-3}(PM10Lag_{14}) - 58.63\ (SO2Lag_{19}) + 2.896 \times 10^{-2}(HumidityLag_9) - 0.1435\ (Temperature\ Lag_{14})$
NBM-Dengue	$Log(Dengue) = -2.5799 - 0.0413\ (Humidity\ Lag_{19}) + 0.3372(Temperature\ Lag_5)$

The number of conjunctivitis cases was greatly influenced by air pollution parameter such as PM10 at lag 14 and S02 at lag 20 whereas the climates variables does not influence this disease at all. Meanwhile, the number of asthma cases was greatly influenced by S02, humidity and temperature at particular lag. The number of URTI cases was greatly influenced by PM10, S02, humidity and temperature at certain lag. The number of dengue cases was greatly influenced by climates variable only; humidity and temperature at particular lag. It is meaning that air pollution does not affect the number of dengue cases at all.

Adequacy checking

The adequacy checking is based on Q-Q plot, residuals vs fitted plot, Jacque-Berra (JB) test, Shapiro-Wilk test and *t*-test. Q-Q plot is a graphical technique for determining if two data sets come from populations with a common distribution. A Q-Q plot is a plot of the quantiles of the first data set against the quantiles of the second data set. Meanwhile, residuals versus fitted plot are a scatter plot of residuals on the *y* axis and fitted values (estimated responses) on the *x* axis. The plot is used to detect non-linearity, unequal error variances, and outliers.

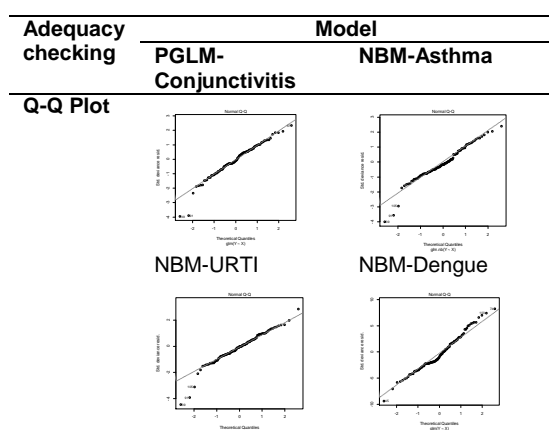| Adequacy checking	Model	
	PGLM-Conjunctivitis	NBM-Asthma
Q-Q Plot		
	NBM-URTI	NBM-Dengue

Fig 11 Q-Q plot of all models

From Figure 11 above and Figure 12 below, the Q-Q plot of all models appears to have come from populations with a common distribution. The residuals versus fitted show the pattern of red line deviating from the estimated regression line (the dotted line). All models show the residuals bouncing randomly around the 0 line. This suggests that the assumption that the relationship is linear is reasonable. However, only NBM-Asthma shows the regular pattern of the residuals along the horizontal band of the dotted line. This

suggests that the variances of the error terms are equal. The rest of the models comprise unequal variance of error. All the models display the residuals which stand out from the basic random pattern of residuals. Thus, this suggests that the outliers are present in the data.
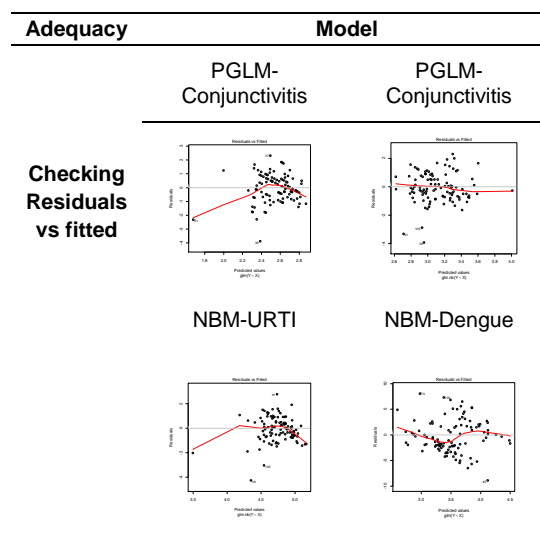
Adequacy	Model	
	PGLM-Conjunctivitis	PGLM-Conjunctivitis
Checking Residuals vs fitted		
	NBM-URTI	NBM-Dengue

Fig 12 Residuals versus fitted of all models

Table 7 Result of Jarque-Bera (JB) test, Shapiro-Wilk (SW) test and *t*-test for all models

Model	Adequacy test	Estimates	*p*-value
PGLM-Conjunctivitis	JB test	20.158	4.195×10^{-5}
	SW test	0.965	0.0069
	t-test	-0.070	**0.5101**
NBM-Asthma	JB test	14.734	0.0006
	SW test	0.961	0.0033
	t-test	-0.077	**0.467**
NBM-URTI	JB test	39.387	2.8×10^{-9}
	SW test	0.952	0.0008
	t-test	-0.108	**0.3206**
NBM-Dengue	JB test	0.872	**0.6468**
	SW test	0.992	**0.8083**
	t-test	-0.189	**0.06635**

However, from the normality test above, only NBM-dengue model exhibits statistically significance since the *p*-value of all tests; JB test, SW test and *t*-test are greater than 0.05 (the null hypothesis is failed to be rejected). The rest of the models are failed to fulfill the assumptions of JB test and SW test. The null hypothesis of residuals were normally distributed is strongly rejected (*p*-value is less than 0.05). Thus, further analysis is required to treat the problem.

CONCLUSION

In this study, the behaviors of the diseases namely asthma, conjunctivitis, URTI and dengue fever in Johor Bahru Malaysia have been studied from 2012 to 2013. The relationship between the diseases and their influential factors such as air pollution parameters and climate variables has been identified through the Pearson correlation. The lag time has also been applied to independent variables. The Poisson Generalized Linear Model (PGLM) and Negative Binomial Model (NBM) were developed to find the association between the diseases and the influential factors. Pollutants

variables that significantly affect the number of asthma, conjunctivitis and URTI are SO2 which are believed to come from the emitted of fossil fuel combustion at power plants and other industrial facilities, as well as fuel combustion in mobile sources such as locomotives, ships, and other equipment. The climate variables that significantly impact the number of asthma, URTI and dengue cases are humidity and temperature. Rainfall does not significantly impact the number of diseases at all.

Regression analysis model constructed in this study can be applied in the early warning system to assist in the prevention and mitigation of disease cases. However, there are also some weaknesses in this study. Among the shortcomings are, first, this study only examines the effects of air pollutant variables on disease cases based on recorded data for a particular period of time. Therefore, the effects of air pollutants on patients beyond the study period are still remain unknown. Secondly, most of the models do not statistically fulfill the assumption of normality. Therefore, further studies should find an alternative to improve the shortcomings.

Although there are some disadvantages, this study can be one of the models for an early warning system to assist the prevention and reduction of diseases in the district of Johor Bahru. The involved parties, such as hospitals can do the preparation early in facing the risk of diseases that may occur because the model in this study applied the time lag. In addition, the community also needs to take early care and work together in maintaining the surroundings so that the diseases can be prevented and mitigated.

ACKNOWLEDGEMENT

REFERENCES

[1] Amancio, C. T., & Luiz, F. C. N. (2012). Asthma and air pollutants : a time series study. *Revista da Associação Médica Brasileira (English Edition)*, 58: 302-307.

[2] Brook, R. D., Franklin, B., Cascio, W., Hong, Y., Howard, G., Lipsett, M.,Tager, I. (2004). A Statement for Healthcare Professionals From the Expert Panel on Population and Prevention Science of the American Heart Association. *American Health Association*, 109:2655-2671.

[3] Chen, K., Glonek, G., Hansen, A., Williams, S., Tuke, J., A. S. and P.B. (2016). The effects of air pollution on asthma hospital admissions in Adelaide , South Australia , 2003 – 2013 : time-series and case – crossover analyses Experimental Allergy. *Clinical & Experimental Allergy*, 46: 1416–1430.

[4] Cheong, Y.L., Burkart, K., Leitão, P.J., & Lakes, T. (2013). Assessing Weather Effects on Dengue Disease in Malaysia. *International Journal of Environmental Research and Public Health*, 10: 6319-6334.

[5] Darrow, L. A., Klein, M., Flanders, W. D., Mulholland, J. A., Tolbert, P. E., & Strickland, M. J. (2014). Original Contribution Air Pollution and Acute Respiratory Infections Among Children 0 – 4 Years of Age : An 18-Year Time-Series Study. *American Journal of Epidemiology* ,180(10): 968–977.

[6] Egghe, L., Leydesdorff, L., & Diepenbeek, C. (2000). The relation between Pearson ' s correlation coefficient r and Salton ' s cosine measure. *Unpublished note*, 1–24.

[7] Hong, J., Zhong, T., Li, H., Xu, J., Ye, X., Mu, Z., & Lu, Y. (2016). Ambient air pollution , weather changes , and outpatient visits for allergic conjunctivitis : A retrospective registry study. *US National Library of Medicine*, 1–8.

[8] Horta, M.A., Bruniera, R., Ker, F., & Catita, C. (2014). Temporal relationship between environmental factors and the occurrence of dengue fever. *International Journal of Environmental Health Research*, 24(5): 471-481.

[9] Ji, H., Myers, J. M. B., Brandt, E. B., Brokamp, C., Ryan, P. H., & Hershey, G. K. K. (2016). Air pollution , epigenetics , and asthma. *Allergy, Asthma & Clinical Immunology*, 12(51): 1–14.

[10] Karmeshu, N. (2012). Trend Detection in Annual Temperature & Precipitation using the Mann Kendall Test – A Case Study to Assess Climate Change on Select States in the Northeastern United States Trend Detection in Annual Temperature & Precipitation using the Mann. *Master Thesis, University of Pennsylvania*.

[11] Kjellstrom, T. E., Neller, A., & Simpson, R. W. (2002). Air pollution and its health impacts : the changing panorama. *US National Library of Medicine*, 177: 604–608.

[12] Li, Z., Bian, X., Yin, J., Zhang, X., & Mu, G. (2016). The Effect of Air Pollution on the Occurrence of Nonspecific Conjunctivitis. *Journal of Opthalmology*, 2016: 10–13.

[13] Liu, Y., Guo, Y., Wang, C., Li, W., Lu, J., & Shen, S. (2015). Association between Temperature Change and Outpatient Visits for Respiratory Tract Infections among Children in Guangzhou , China. *International Journal of Environmental Research and Public Health*, 12: 439–454.

[14] Liu, Y., Liu, J., Chen, F., Haider, S. B., Wang, Q., Jiao, F., & Shi, Y. (2015). Coal Mine Air Pollution and Number of Children Hospitalizations because of Respiratory Tract Infection : A Time Series Analysis. *Journal of Environmental and Public Health*, 2015(2):1-7.

[15] Massad, E., Coutinho, F. aB., Ma, S., & Burattini, M. N. (2010). A hypothesis for the 2007 dengue outbreak in Singapore. Epidemiology and Infection. *US National Library of Medicine*,138(7): 951-957.

[16] Olívêr, L., Burattini, M.N., Coutinho, F.A., Coelho, G.E., Struchiner, C.J., & MD, M. E. (2014). A negative correlation between dengue and bushfires in Brazil. *Journal of Environmental Health*, 76(6): 66–67.

[17] Orellano, P., Quaranta, N., Reynoso, J., Balbi, B., & Vasquez, J. (2017). Effect of outdoor air pollution on asthma exacerbations in children and adults : Systematic review and multilevel meta- analysis. *PlosOne*,(2): 1–15.

[18] Ren, M., Li, N., Wang, Z., Liu, Y., Chen, X., Chu, Y., & Li, X. (2017). The short-term effects of air pollutants on respiratory disease mortality in Wuhan , China : comparison of time-series and case-crossover analyses. *US National Library of Medicine*, 2016: 1–9.

[19] Silvester, A., Neal, T., Czanner, G., Briggs, M., Harding, S., & Kaye, S. (2016). Adult bacterial conjunctivitis : resistance patterns over 12 years in patients attending a large primary eye care centre in the UK. *BMJ Open Opthalmology*, 1–5.

[20] Souza, A. De, Kofanovski, A. Z., Sabbah, I., Débora, A., & Santos, S. (2016). Allergy & Therapy Asthma and Environmental Indicators : A Time-series Study. *Allergy & Theraphy*, 7(1): 1–5.

[21] Tam, W. W. S., Wong, T. W., Ng, L., Wong, S. Y. S., Kung, K. K. L., & Wong, A. H. S. (2014). Association between Air Pollution and General Outpatient Clinic Consultations for Upper Respiratory Tract Infections in Hong Kong. *PlosOne*, 9(1): 1–6.

[22] Tjønneland, A., Overvad, K., & Raaschou-nielsen, O. (2006). Chronic Obstructive Pulmonary Disease and Long-Term Exposure to Traffic-related Air Pollution A Cohort Study. *American Journal of Respiratory and Critical Care Medicine*, 183: 455-461.

[23] Tuluri, F., & Ak, G. (2017). Asthma Trends in Mississippi Coastal Region with Air Pollutants and Meteorological Factors. *Austin Journal of Allergy*, 4(2): 1–6.

[24] Wilder-Smith, A., Earnest, A., Tan, S. B., Ooi, E. E., & Gubler, D. J. (2010). Lack of association of dengue activity with haze. *Epidemiology and Infection*, 138(7): 962-967.

[25] Xu, H., Fu, X., Kim, L., Lee, H., Ma, S., Goh, K.T., Wong, J., Habibullah, S., Kee, G., Lee, K., Lim, T.K., Tambyah, P.A., Lim, L., & Ng, L.C. (2014). Statistical Modeling Reveals the Effect of Absolute Humidity on Dengue in Singapore. *PLOS Neglected Tropical Diseases*, 8(5): 2805-2813.

[26] Xu, Z., Huang, C., Su, H., Turner, L. R., Qiao, Z., & Tong, S. (2013). Diurnal temperature range and childhood asthma : a time-series study. *Environmental Health*,12: 1–5.

[27] Yin, P., He, G., Fan, M., Chiu, K. Y., Fan, M., Liu, C., & Zhou, M. (2017). Particulate air pollution and mortality in 38 of China's largest cities : time series analysis. *The BMJ*, 667: 1–12.

[28] Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. *Unpublished note*.

[29] Zhang, Y., Peng, L., Kan, H., Xu, J., Chen, R., Liu, Y., & Wang, W. (2014). Effects of Meteorological Factors on Daily Hospital Admissions for Asthma in Adults : A Time-Series Analysis. *Plos One*, 9(7): 1-7.

[30] Zheng, X., Ding, H., Jiang, L., Chen, S., Zheng, J., & Qiu, M. (2015). Association between Air Pollutants and Asthma Emergency Room Visits and Hospital Admissions in Time Series Studies : A Systematic Review and Meta-Analysis. *Plos One*, 18: 1-24.