Pertanika J. Sci. & Technol. 31 (4): 1881 - 1894 (2023)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Comparison of Count Data Generalised Linear Models: Application to Air-Pollution Related Disease in Johor Bahru, Malaysia

Zetty Izzati Zulki Alwani[1], Adriana Irawati Nur Ibrahim[1]*, Rossita Mohamad Yunus[1] and Fadhilah Yusof[2]

[1]*Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, 50603 UM, Kuala Lumpur, Malaysia*
[2]*Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia*

ABSTRACT

Poisson regression is a common approach for modelling discrete data. However, due to characteristics of Poisson distribution, Poisson regression might not be suitable since most data are over-dispersed or under-dispersed. This study compared four generalised linear models (GLMs): negative binomial, generalised Poisson, zero-truncated Poisson and zero-truncated negative binomial. An air-pollution-related disease, upper respiratory tract infection (URTI), and its relationship with various air pollution and climate factors were investigated. The data were obtained from Johor Bahru, Malaysia, from January 1, 2012, to December 31, 2013. Multicollinearity between the covariates and the independent variables was examined, and model selection was performed to find the significant variables for each model. This study showed that the negative binomial is the best model to determine the association between the number of URTI cases and air pollution and climate factors. Particulate Matter (PM_{10}), Sulphur Dioxide (SO_2) and Ground Level Ozone (GLO) are the air pollution factors that affect this disease significantly. However, climate factors do not significantly influence the number of URTI cases. The model constructed in this study can be utilised as an early warning system to prevent and mitigate URTI cases. The involved parties, such as the local authorities and hospitals, can also employ the model when facing the risk of URTI cases that may occur due to air pollution factors.

Keywords: Air pollution disease, count data, generalised linear model

ARTICLE INFO

Article history:
Received: 28 May 2022
Accepted: 02 November 2022
Published: 25 May 2023

DOI: https://doi.org/10.47836/pjst.31.4.16

E-mail addresses:
zettyziza@gmail.com (Zetty Izzati Zulki Alwani)
adrianaibrahim@um.edu.my (Adriana Irawati Nur Ibrahim)
rossita@um.edu.my (Rossita Mohamad Yunus)
fadhilahy@utm.my (Fadhilah Yusof)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

A generalised linear model (GLM) is commonly used to study the relationship between the response and the covariate(s). Generalised linear models are widely used in applying regression models for discrete data such as count data (Nelder & Wedderburn, 1972; Agresti, 2003; Cameron & Trivedi, 2013). Poisson regression is the most commonly used model to analyse count data. For example, Poisson regression was used to estimate the risk of hospitalisation for asthma in children after being exposed to air pollutants in a city in Southeast Brazil (Amâncio & Nascimento, 2012) and was used to study the relationship between air pollution concentration with upper respiratory tract infection (URTI) among children in Atlanta, Georgia (Darrow et al., 2014).

However, most real data are usually under-dispersed or over-dispersed. Therefore, real data may not fulfil the assumption of Poisson distribution which assumes equal dispersion. Negative binomial (NB) distribution is frequently used to handle over-dispersed data. Avcı (2018) used NB regression to study the effects of several people hospitalised with schizophrenia. Apart from that, several other count data models have been introduced to overcome both situations of under-dispersion and over-dispersion. The generalised Poisson (GP) distribution is one example of a distribution which can handle both cases. The GP regression model was applied in various areas to model over-dispersed data, such as household fertility data (Wang & Famoye, 1997) and accident data (Famoye et al., 2004).

Besides that, usually, the data is over-dispersed; thus, the response is usually truncated for some outliers or large values (Saffari et al., 2011). Using the usual regression for zero-truncated datasets may be difficult because it may try to predict zero counts despite the data, not including any zeros. Thus, zero-truncated models such as zero-truncated Poisson and NB may be more suitable. Zero-truncated Poisson was used to model the number of fishing trips taken during the fishing season in Alaska (Grogger & Carson, 1991), while zero-truncated NB was applied to estimate recreation demand in a national park in Canada (Martínez-Espiñeira & Amoako-Tuffour, 2008).

The study aims to investigate the link between air pollution and climate factors with the disease of upper respiratory tract infection (URTI). URTI is an upper respiratory tract infection that includes the throat, nose, pharynx, larynx, sinuses, and trachea (windpipe). URTI causes irritation and swelling of the upper airways (Thomas & Bomar, 2021). Several studies have shown that air pollution or/and climate can affect the number of URTI patients, using models such as Poisson GLM (Darrow et al., 2014; Çapraz & Deniz, 2021), Poisson generalised additive model (GAM) (Tam et al., 2014; Liu et al., 2015; Liu, Liu et al., 2015; Li et al., 2018; Zhang et al., 2019), GLM with quasi-Poisson (Liu et al., 2015) and case-crossover design with a fixed-effect model (Szyszkowicz et al., 2018). Since the Poisson distribution usually assumes equal dispersion, these models applied methods such as time series (Darrow et al., 2014) and GAM with quasi-likelihood (Tam et al., 2014; Liu, Guo et al., 2015) to allow for over-dispersion. However, these studies did not consider models with specific distributions catering to under-dispersion and over-dispersion.

Therefore, this research aims to compare several count data GLMs with different distributions when investigating the relationship between the number of URTI cases with air pollution and climate variables in Johor Bahru, Malaysia. Previously, Jamaludin et al. (2017) had applied the NB GLM to investigate the relationship between URTI and several air pollution and climate variables in Johor Bahru, Malaysia, while Alwani et al. (2021) used the zero-truncated Poisson and zero-truncated NB GLMs to examine the association between the number of dengue cases with similar air pollution and climate factors for similar data.

Hence, this study shall also consider generalised Poisson (GP), zero-truncated Poisson and zero-truncated NB GLMs to examine the relationship between URTI cases and their significant factors, apart from the NB GLM. GP is chosen as it can handle over and under-dispersed data, while the two zero-truncated models may be good alternatives since the number of URTI cases in our data is strictly nonzero. Since the data are nonzero, this study does not consider models catering for many zeros, such as zero-inflated Poisson or NB and negative binomial-Lindley. Note that Conway-Maxwell Poisson has also been considered for this study, but the model application showed that it is unsuitable for the data.

Apart from that, this study shall also investigate multicollinearity between the covariates, i.e., the independent variables and perform model selection to find the final best model with the significant covariates for each model, both of which were not previously considered in previous studies.

MATERIALS AND METHODS

Description of Data

The study investigates the link between upper respiratory tract infection (URTI) with air pollution and climate variables. The data is obtained from Johor Bahru, the southernmost city in Peninsular Malaysia; the data is obtained weekly, which covers 105 weeks starting from January 2012 up to December 2013. The numbers of URTI cases were obtained from the Department of Information and Informatics, Johor Bahru District Health Office. The environmental pollution variables are Ground Level Ozone (GLO), Nitrogen Dioxide (NO_2), Particulate Matter (PM_{10}) and Sulphur Dioxide (SO_2), while the climate variables are rainfall, temperature and relative humidity. Data on air pollution levels were obtained from the Johor Department of Environment, rainfall data were acquired from the Department of Irrigation and Drainage Malaysia, and data on temperature and relative humidity were acquired from the Malaysian Meteorological Department.

Methodology

The study aims to model the relationship between the number of URTI cases and the pollution and climate factors using the data aggregated weekly. The lag effects are needed to study this relationship. It is because the number of cases of URTI may be not only influenced by the values of the factors in the same week but also by the values in earlier weeks, as the influences of these factors may not be immediate (delayed effect). The delayed effect can be a few days for the short term or weeks for the long term.

Jamaluddin et al. (2017) applied a long-term association between air pollution and URTI where the lag used was the same week (lag 0) up to the previous 20 weeks (lag 20). However, in most of the previous studies, the usual lags used were single-day lag and cumulative lag, where the lags can go from the same day up to 3 days (Zhang et al., 2021), 6 days (Tao et al., 2014; Li et al., 2018) or 20 days (Saldiva et al., 1994). Thus, this paper uses the weekly lag, which is similar to Jamaluddin et al. (2017) but for a shorter period where the lag is used in the same week (lag 0) up to the previous 4 weeks (lag 4).

Pearson's correlation is then used to examine the association between URTI and each of the factors at the different lags, and the most significant lag is chosen for each factor which is the time lag with the highest correlation value (Montgomery et al., 2021). To avoid or at least reduce multicollinearity between the explanatory variables at their significant lags, partial (Pearson's) correlation is used to find the collinearity between these variables while excluding the influence of the other variables; the explanatory variables are then reduced for pair(s) of variables which have a high partial correlation.

GLM using NB, COM-Poisson, GP, zero-truncated Poisson and zero-truncated NB are then fitted to the data using the reduced explanatory variables at their significant lags. In each of the models, the canonical link is the log link, such that $\log(\mu_i) = \boldsymbol{X}_i\boldsymbol{\beta}'$ where μ_i is the mean for observation y_i, $\boldsymbol{X}_i$ is the independent variables, and $\boldsymbol{\beta}$ is the regression parameters. The details for each model are given below:

1. Negative Binomial model

The NB model is usually used for over-dispersed count data, where the variance is assumed to be larger than the mean. The p.m.f. of NB distribution is expressed as Equation 1:

$$f(y_i;\mu_i,\theta) = \frac{\Gamma(y_i+\theta)}{\Gamma(\theta)\ y_i!}\cdot\frac{\mu_i^{y_i}\theta^{\theta}}{(\mu_i+\ \theta)^{y_i+\theta}}, \qquad y_i = 0,1,2,\ldots, \tag{1}$$

where θ is the shape parameter and gamma function $\Gamma(.)$.

2) Generalised Poisson model

P.m.f. of GP can be expressed as Equation 2 (Consul & Jain, 1973; Consul & Famoye, 1992):

$$f(y_i;\mu_i,\kappa|y_i>0)=\frac{\big((1-\kappa)\mu_i+\kappa y_i\big)^{y_i-1}(1-\kappa)\mu_i\exp(-(1-\kappa)\mu_i-\kappa y_i)}{y_i!}, \qquad [2]$$

where $0 < k < 1$.

3) Zero-truncated Poisson model

P.m.f. of zero-truncated Poisson is given as Equation 3 (Grogger & Carson, 1991; Zuur et al., 2009):

$$f(y_i;\mu_i|y_i>0)=\frac{\mu_i^{y_i}e^{-\mu_i}}{y_i!\,(1-e^{-\mu_i})}\ . \qquad [3]$$

4) Zero-truncated negative Binomial model

Zero-truncated NB (Grogger & Carson, 1991; Zuur et al., 2009) has the p.m.f. as in Equation 4:

$$f(y_i;\alpha,\mu_i|y_i>0)=\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)\,\Gamma(\,y_i+1)}\ \left(\frac{\alpha}{\mu_i+\ \alpha}\right)^{\alpha}\left(1-\frac{\alpha}{\mu_i+\ \alpha}\right)^{y_i}\ . \qquad [4]$$

During the model fitting of each model, stepwise selection based on the Akaike information criterion (AIC) (Montgomery et al., 2021) is used to find the final model, which contains only the significant variables. Normal quantile-quantile (Q-Q) plot and several tests such as the Shapiro-Wilk (SW), Jarque-Berra (JB) and D' Agostino (DA) tests are then used to check for the normality assumption of the residuals of the final model for each model considered in this study. The normal Q-Q plot compares the quantiles of the data distribution with the quantile of the standardised normal distribution. The SW test uses the order and normalised order statistics for normality (Shapiro et al., 1968), while the DA test assesses symmetry and evaluates the distribution shape using normal approximation (D'Agostino & Pearson, 1973). The JB test, on the other hand, combines skewness and kurtosis in a single statistic (Wuertz & Katzgraber, 2005). The scatter plot of the residuals of the final model is also used to check for the assumptions that the residuals are mostly random, centred around zero and have constant variance.

Finally, to select the best final model among all the proposed GLMs, AIC and Bayesian information criterion (BIC) are used. The AIC is a penalised log-likelihood measure. Let L be the likelihood function for a specific model. The AIC is given by Equation 5:

$$\mathrm{AIC}=-2\ln L+2p, \qquad [5]$$

where p is the number of parameters in the model. On the other hand, the Schwartz BIC is given by Equation 6:

$$\mathrm{BIC}=-2\ln L+p\ln n\,, \qquad [6]$$

where n is the size of the data set. Generally, the smaller the values of AIC and BIC are, the better the model. For more information on AIC and BIC, please refer to Montgomery et al. (2021).

To summarise the steps in modelling the relationship between the number of URTI cases and this study's pollution and climate factors, please refer to the flowchart in Figure 1. All analyses are applied using R and Excel. The R packages used for the GLMs are the *MASS* package for NB, and the *VGAM* package (Yee, 2021) for GP, zero-truncated Poisson and zero-truncated NB.

RESULTS

Descriptive Data

In this research, a weekly number of pollution-related disease cases, in particular URTI, and weekly level of environmental pollution (GLO, NO_2, SO_2 [all in ppm]; PM_{10} in µg/m^3 and climate [rainfall in mm), temperature in °C and relative humidity in %) throughout January 1, 2012, to December 31, 2013, were used. The total number of cases in these two years for URTI was 12244. The weekly time series plot of URTI is given in Figure 2, and the weekly number of cases for URTI fluctuated throughout the two years. Few cases occur between September to November of each year; September to November is one of the inter-monsoon periods, relatively calm with less rain and weaker winds (Jamaludin et al., 2017).

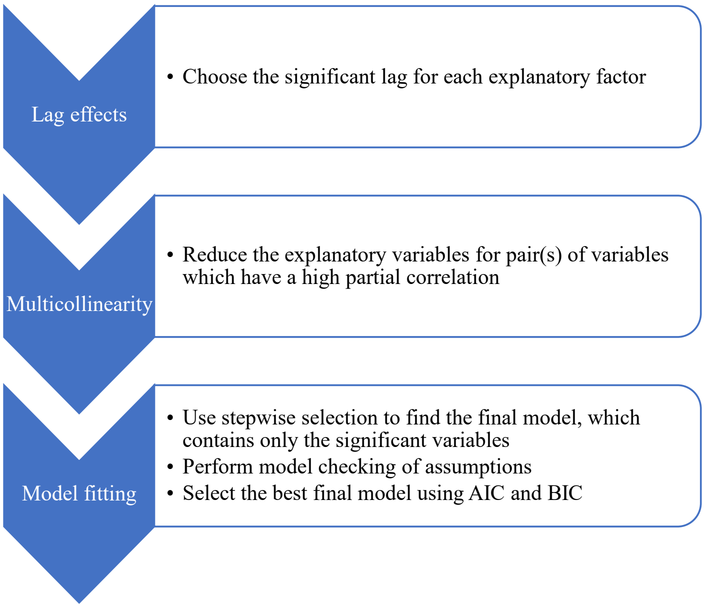

Figure 1. Flowchart of the methodology

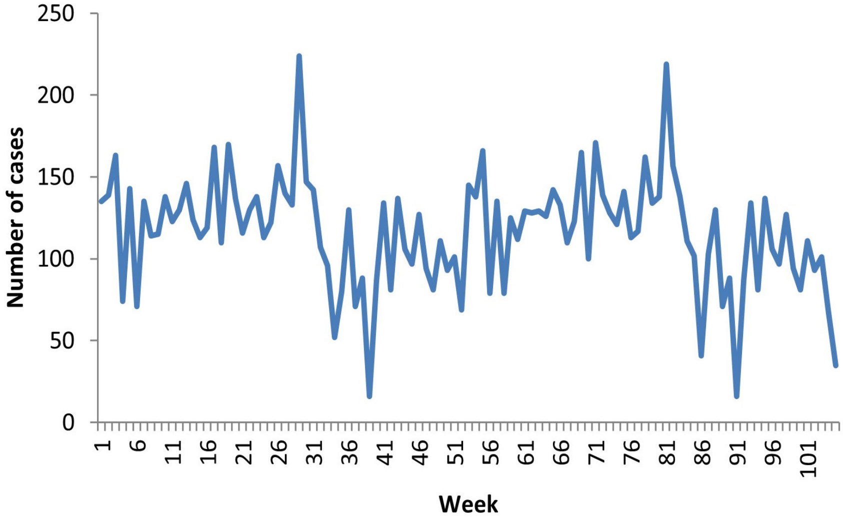

Figure 2. Weekly time series plot of URTI

Table 1 gives the descriptive statistics for several cases of URTI, environmental pollution and climate factors. From Table 1, the weekly numbers of cases recorded for URTI were all nonzero, and the variance for URTI is much larger than its mean. The table also shows that the levels of environmental pollution were very low, apart from PM_{10}, which had very high values. PM_{10} also had a large dispersion. For climate factors, rainfall had a very large dispersion due to Malaysia's wet and dry seasons.

Table 1

Descriptive statistics of the number of cases of URTI and level of environmental pollution and climate for 2012–2013

	Min	Max	Mean	Standard Deviation	Skewness	Variance
URTI	16	224	117	34.581	- 0.176	1195.86
GLO	0.017	0.073	0.043	0.012	0.557	0.0001
NO_2	0.005	0.023	0.014	0.004	- 0.491	0.00001
PM_{10}	22.30	225.49	42.70	20.728	6.880	429.67
SO_2	0.0002	0.0089	0.0031	0.0018	0.772	0.000003
Rainfall	0	194.50	53.43	51.006	0.989	2601.61
Temperature	25.24	29.00	26.78	0.704	0.354	0.496
Relative humidity	75.30	92.79	85.92	2.970	- 0.342	8.817

Relationship Between URTI And the Pollution and Climate Factors

Table 2 shows the result of Pearson's correlation between the URTI and the related factors; the values of each factor less influence the number of URTI cases in the same week but more influenced by values in the earlier weeks (lags). The factors affect the number of URTI cases at different lags. It may be due to different lagged/delayed effects of each pollution and climate factor on the patients, meaning that each factor's effects may differ based on the length of time (weeks) after exposure. Based on Table 2, NO_2 and SO_2 have a stronger relationship with the number of URTI cases in the same week (lag 0). However, some variables show stronger delayed effects on the number of URTI cases, such as GLO in the previous first week (lag 1), rainfall in the third week (lag 3) and PM_{10}, temperature and relative humidity in the fourth week (lag 4).

Pearson’s partial correlation values between the air pollution and climate factors at their significant lags are shown in Table 3. SO_2 and NO_2, relative humidity and temperature are highly correlated (these correlations are also significant at a 1% level). Therefore, this study includes NO_2 instead of SO_2, as there is evidence from previous studies that NO_2 is more associated with URTI (Wong et al., 2006; Li et al., 2018). The relative humidity in the model is included instead of temperature, as a previous study found that absolute humidity was significantly associated with URTI (Mäkinen et al., 2009). Thus, the variables included in the models in the next discussion are NO_2, GLO and PM_{10} for the air pollution factors and rainfall and temperature for the climate factors, each at their significant lags.

Table 2

Pearson’s correlation of URTI with air pollution and climate factors

Factor		Significant lag	Pearson correlation
Air pollution	GLO	1	- 0.169
	NO_2	0	0.195
	PM_{10}	4	0.325
	SO_2	0	0.142
Climate	Rainfall	3	- 0.219
	Relative humidity	4	- 0.100
	Temperature	4	0.4643

Table 3

Pearson’s partial correlation between air pollution and climate factors at their significant lags

	GLO	NO_2	PM_{10}	SO_2	Rainfall	Relative humidity	Temperature
GLO	1.00						

Table 3 *(Continue)*

	GLO	NO_2	PM_{10}	SO_2	Rainfall	Relative humidity	Temperature
NO_2	0.44*	1.00					
PM_{10}	-0.08	0.11	1.00				
SO_2	-0.18	0.65**	-0.01	1.00			
Rainfall	0.19	-0.23*	-0.01	0.01	1.00		
Relative humidity	0.09	0.06	-0.22*	0.10	-0.03	1.00	
Temperature	0.03	0.21*	0.10	0.03	-0.14	-0.66**	1.00

* significant at 5% level; ** significant at 1% level

The Selection of Best Model

The over-dispersion test was conducted on the URTI data. From this test, the *p*-values for the test on URTI are less than 0.05. Thus, the number of URTI cases is over-dispersed, and the NB model is preferred over the Poisson model. It is why the Poisson GLM is not considered in this study.

GLM using NB, GP, zero-truncated Poisson, and zero-truncated NB are each applied to the URTI data. Some final models have different significant variables after applying the stepwise selection procedure. NB, zero-truncated NB, and GP models have the same significant variables, PM_{10}, NO_2 and GLO, while the zero-truncated Poisson final model has similar significant variables but with additional variables of rainfall and relative humidity.

Looking at Table 4, the final NB GLM has lower AIC and BIC values compared to the GP and zero-truncated Poisson models and has the same values as the zero-truncated NB model. Note that the NB and zero-truncated NB models have the same significant covariates in the final model, but the zero-truncated NB model has an extra parameter to be estimated compared to the NM model. Thus, the NB GLM is chosen as the best model to examine the effect of air pollution and climate factors on URTI cases for this data.

Table 4

Result of normality tests, AIC and BIC for URTI data

	NB	Generalised Poisson	Zero-truncated Poisson	Zero-truncated NB
Normality tests (*p*-value)	SW: 0.010 DA: 0.009 JB: 0.001	SW: 0.033 DA: 0.019 JB: 0.004	SW: 0.007 DA: 0.008 JB: 0.001	SW: 0.010 DA: 0.009 JB: 0.001
AIC	1013.072	1021.855	1647.067	1013.072
BIC	1026.098	1034.881	1662.698	1026.098

For the residuals of the fitted NB model, the *p*-values for the three normality tests are smaller than 0.05, as shown in Table 4, indicating strong evidence to reject the null hypothesis that residuals are normally distributed. However, from the Q-Q plot for the residuals of the fitted NB model given in Figure 3, this model mostly falls on a straight line except for the few points on the right tail. Therefore, this study concluded that the normality assumption when fitting the data using the NB model is not severely violated. The residuals of the fitted NB model are also mostly random, centred around zero and have approximately constant variance, as seen in Figure 4.

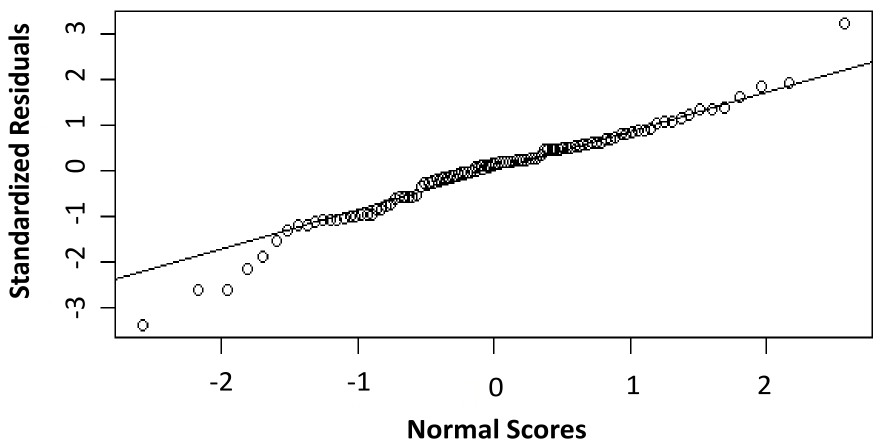

Figure 3. Q-Q plot for residuals of the fitted negative Binomial model for URTI data

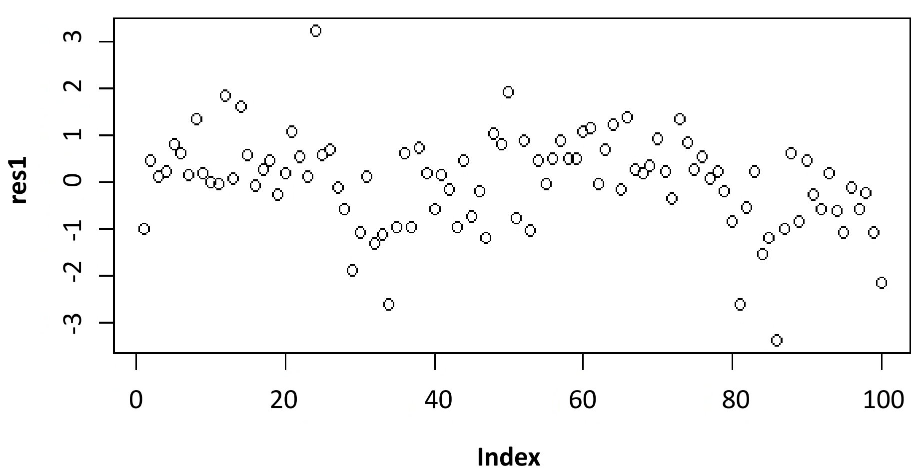

Figure 4. Scatter plot for residuals of the fitted negative Binomial model for URTI data

The parameter estimates for the fitted NB model are given in Table 5; the fitted model then gives the best final model as in Equation 7:

$$\hat{y} = \exp\left(4.621 + 2.346 \times 10^{-3}(\mathrm{PM}_{10}) + 18.236(\mathrm{NO}_2) - 5.700(\mathrm{GLO})\right) \quad [7]$$

where the fitted data follow the NB distribution (Equation 1) with an estimated shape parameter equal to 10.54. Note that although the parameter estimate for PM_{10} is small, the effect of this factor is still significant as the range of values for PM_{10} is quite large, as can be seen from Table 1; the minimum value for PM_{10} is around 22 μg/m^3, and the maximum value is around 225 μg/m^3.

Table 5

Summary of estimates for fitted negative Binomial model for URTI data

Parameter	Estimate	Std. error	p-value
Intercept	4.621	0.160	<0.001
PM_{10} (Lag 4)	2.346×10^{-3}	1.596×10^{-3}	0.142
NO_2 (Lag 0)	18.236	10.405	0.080
GLO (Lag 1)	-5.700	3.114	0.067

DISCUSSION

In this study, relative risk (RR) is used to estimate the impact of the significant factors on the number of URTI cases. The estimation of RR is $RR = \exp(\beta \times x_i)$, where β is the regression coefficient associated with a unit increment of the significant factors x_i (Çapraz & Deniz, 2021). A unit increment for each significant factor may differ according to the range of values for each factor. This study will present the result of excess relative risk (ERR), which is the percentage increase (%) in the number of URTI cases with a unit increment. ERR is derived from RR, calculated as $ERR = (RR\text{-}1) \times 100$. The values of RR and ERR for the final NB model are shown in Table 6. Note that the unit increment for each significant factor is chosen based on the range of values for each factor (Table 1). Hence, for the final NB model, an increment of 10 μg/m^3 in the concentration of PM_{10}, 0.005 ppm in the concentration of NO_2, and 0.01 ppm in the concentration of GLO correspond to an increase of URTI patients by 2.37% and 9.55% and a decrease of 5.54%, respectively.

Table 6

Relative risk (RR) and excess relative risk (ERR) for the final Negative Binomial model

Significant factor	Unit increment	RR	ERR (%)
PM_{10}	10	1.0237	2.37
NO_2	0.005	1.0955	9.55
GLO	0.01	0.9446	-5.54

CONCLUSION

In this paper, the weekly number of URTI cases in the city of Johor Bahru, Malaysia, from January 2012 to December 2013 have been studied. The explanatory factors have been reduced after checking for multicollinearity between the factors at their significant lags. The NB GLM is then found to be the best model to model the association between URTI and the air pollution and climate factors at the corresponding significant lags. Air pollution factors that have significant effect on URTI are PM_{10}, NO_2 and GLO. However, climate factors do not significantly affect URTI. The model constructed in this study can be utilized as an early warning system to assist in the prevention and mitigation of URTI cases. The involved parties, such as the local authorities and hospitals can also employ the model as a precaution when facing the risk of URTI cases that may occur due to the air pollution factors. Note however that the results from this study are limited to the observed data for a particular period. Therefore, the effects of these air pollution and climate factors on URTI cases beyond the study period still remain unknown.

ACKNOWLEDGEMENTS

This work was financially supported by the Ministry of Higher Education, Malaysia, under the Fundamental Research Grant Scheme No: FRGS/1/2019/STG06/UM/02/2 (FP110-2019A).

REFERENCES

Agresti, A. (2003). *Categorical Data Analysis.* John Wiley & Sons.

Alwani, Z. Z., Ibrahim, A. I. N., Yunus, R. M., & Yusof, F. (2021). Application of zero-truncated count data regression models to air-pollution disease. *Journal of Physics: Conference Series*, *1988*(1), Article 012096. https://doi.org/10.1088/1742-6596/1988/1/012096

Amâncio, C. T., & Nascimento, L. F. (2012). Asthma and ambient pollutants: A time series study. *Revista da Associacao Medica Brasileira*, *58*(3), 302-307. https://doi.org/10.1590/S0104-42302012000300009

Avcı, E. (2018). Using count regression models to determine the factors which affects the hospitalization number of people with schizophrenia. *Journal of Data Science*, *16*(3), 511-530. https://doi.org/10.6339/JDS.201807_16(3).0004

Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data* (Vol. 53). Cambridge University Press.

Çapraz, Ö., & Deniz, A. (2021). Assessment of hospitalizations from asthma, chronic obstructive pulmonary disease and acute bronchitis in relation to air pollution in İstanbul, Turkey. *Sustainable Cities and Society*, *72*, Article 103040. https://doi.org/10.1016/j.scs.2021.103040

Consul, P. C., & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, *15*(4), 791-799. https://doi.org/10.1080/00401706.1973.10489112

Consul, P., & Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, *21*(1), 89-109. https://doi.org/10.1080/03610929208830766

D'Agostino, R. B., & Pearson, E. S. (1973). Tests for Departure from Normality. *Biometrika*, *60*, 613-622. https://doi.org/10.2307/2335012

Darrow, L. A., Klein, M., Flanders, W. D., Mulholland, J. A., Tolbert, P. E., & Strickland, M. J. (2014). 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. https://doi.org/10.1093/aje/kwu234

Famoye, F., Wulu, J. T., & Singh, K. P. (2004). On the generalized Poisson regression model with an application to accident data. *Journal of Data Science*, *2*(3), 287-295. https://doi.org/10.6339/JDS.2004.02(3).167

Grogger, J. T., & Carson, R. T. (1991). Models for truncated counts. *Journal of Applied Econometrics*, *6*(3), 225-238. https://doi.org/10.1002/jae.3950060302

Jamaludin, A. R. B., Yusof, F., Lokoman, R. M., Noor, Z. Z., Alias, N., & Aris, N. M. (2017). Correlational study of air pollution-related diseases (asthma, conjunctivitis, URTI and dengue) in Johor Bahru, Malaysia. *Malaysian Journal of Fundamental and Applied Sciences*, *13*, 354-361. https://doi.org/10.11113/mjfas.v13n4-1.897

Li, Y. R., Xiao, C. C., Li, J., Tang, J., Geng, X. Y., Cui, L. J., & Zhai, J. X. (2018). Association between air pollution and upper respiratory tract infection in hospital outpatients aged 0-14 years in Hefei, China: A time series study. *Public Health*, *156*, 92-100. https://doi.org/10.1016/j.puhe.2017.12.006

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. https://doi.org/10.3390/ijerph120100439

Mäkinen, T. M., Juvonen, R., Jokelainen, J., Harju, T. H., Peitso, A., Bloigu, A., Silvennoinen-Kassinen, S., Leinonen, M., & Hassi, J. (2009). Cold temperature and low humidity are associated with increased occurrence of respiratory tract infections. *Respiratory Medicine*, *103*(3), 456-462. https://doi.org/10.1016/j.rmed.2008.09.011

Martínez-Espiñeira, R., & Amoako-Tuffour, J. (2008). Recreation demand analysis under truncation, overdispersion, and endogenous stratification: An application to Gros Morne National Park. *Journal of Environmental Management*, *88*(4), 1320-1332. https://doi.org/10.1016/j.jenvman.2007.07.006.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370-384. https://doi.org/10.2307/2344614

Saffari, S. E., Adnan, R., & Greene, W. (2011). Handling of over-dispersion of count data via truncation using poisson regression model. *Journal of Computer Science and Computational Mathematics*, *1*(1), 1-4. https://doi.org/10.20967/jcscm.2011.01.001

Saldiva, P. H., Lichtenfels, A. J. F. C., Paiva, P. S. O., Barone, I. A., Martins, M. A., Massad, E., Pereira, J. C. R., Xavier, V. P., Singer, J. M., & Bohm, G. M. (1994). Association between air pollution and mortality due to respiratory diseases in children in São Paulo, Brazil: A preliminary report. *Environmental Research*, *65*(2), 218-225. https://doi.org/10.1006/enrs.1994.1033

Shapiro S. S., Wilk M. B., & Chen V. (1968). A comparative study of various tests for normality. *Journal of American Statistical Association*, *63*, 1343-1372. https://doi.org/10.2307/2285889

Szyszkowicz, M., Kousha, T., Castner, J., & Dales, R. (2018). Air pollution and emergency department visits for respiratory diseases: A multi-city case crossover study. *Environmental Research*, *163*, 263-269. https://doi.org/10.1016/j.envres.2018.01.043

Tam, W. W., Wong, T. W., Ng, L., Wong, S. Y., Kung, K. K., & Wong, A. H. (2014). Association between air pollution and general outpatient clinic consultations for upper respiratory tract infections in Hong Kong. *PLoS One*, 9(1), Article e86913. https://doi.org/10.1371/journal.pone.0086913

Tao, Y., Mi, S., Zhou, S., Wang, S., & Xie, X. (2014). Air pollution and hospital admissions for respiratory diseases in Lanzhou, China. *Environmental Pollution*, *185*, 196-201. https://doi.org/10.1016/j.envpol.2013.10.035

Thomas, M., & Bomar, P. A. (2021). *Upper Respiratory Tract Infection.* StatPearls Publishing.

Wang, W., & Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics*, *10*(3), 273-283. https://doi.org/10.1007/s001480050043

Wong, T. W., Tam, W., Yu, I. T. S., Wun, Y. T., Wong, A. H., & Wong, C. M. (2006). Association between air pollution and general practitioner visits for respiratory diseases in Hong Kong. *Thorax*, *61*(7), 585-591. http://dx.doi.org/10.1136/thx.2005.051730

Wuertz, D., & Katzgraber, H. G. (2005). *Precise Finite-Sample Quantiles of the Jarque-Bera Adjusted Lagrange Multiplier Test.* ETHZ Preprint.

Yee T. W. (2021). *VGAM: Vector generalized linear and additive models R package version 1.1-5*. https://CRAN.R-project.org/package=VGAM

Zhang, D., Tian, Y., Zhang, Y., Cao, Y., Wang, Q., & Hu, Y. (2019). Fine particulate air pollution and hospital utilization for upper respiratory tract infections in Beijing, China. *International Journal of Environmental Research and Public Health*, *16*(4), Article 533. https://doi.org/10.3390/ijerph16040533

Zhang, F., Zhang, H., Wu, C., Zhang, M., Feng, H., Li, D., & Zhu, W. (2021). Acute effects of ambient air pollution on clinic visits of college students for upper respiratory tract infection in Wuhan, China. *Environmental Science and Pollution Research*, *28*(23), 29820-29830. https://doi.org/10.1007/s11356-021-12828-7

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). GLM and GAM for count data. In *Mixed Effects Models and Extensions in Ecology with R* (pp. 209-243). Springer. https://doi.org/10.1007/978-0-387-87458-6_9