Poll Res. 42 (1) : 135-139 (2023)

ISSN 0257–8050
DOI No.: http://doi.org/10.53550/PR.2023.v42i01.021

IMPACT OF ENVIRONMENTAL FACTORS ON AIR POLLUTION RELATED TO A RESPIRATORY DISEASES USING STATISTICAL MODELS

S. RAGURAMAN[1] AND S. SUNDARABALAN[2]

Department of Statistics, Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai, T.N., India

(Received 8 August, 2022; Accepted 29 October, 2022)

ABSTRACT

Air pollution is a serious environmental problem with various health threatening outcomes. Hospital admission data were collected from 120 randomly selected respiratory diseases during May 2020 to April 2021 in Southern districts Tamil Nadu were taken into account of this research work. Statistical analyses were performed using generalized linear models, generalized additive models, parameter driven model and transitional regression model. This study investigates the hospitalization of respiratory diseases and air pollution levels in Southern district of Tamil Nadu.

KEY WORDS : Air pollutants, Respiratory diseases, Regression models.

INTRODUCTION

Human health is very closely linked to environmental quality, as etiologic of most of the human diseases being related to the status of living environment of man. According to statistics, 25% of all preventable illnesses are caused by detrimental environmental factors (UNEP, United Nation's Children Fund, WHO 2002). As far as air pollution is concerned, both developed and developing countries are at risk. The main sources of air pollution are the vehicular and industrial sector.

Urban air pollution problem is reaching crisis dimension in many cities of the developing countries. There are several endemic diseases resulting from air borne particulate materials which has drawn public attention. The latest epidemic disease outbreak occurred in April 2003, when Severe Acute Respiratory Syndrome (SARS) appeared suddenly, sending shockwaves throughout the world, distributing public health systems and economic security world-wide. Chemical agents, particularly air borne ones, are considered to be major factors in causing tuberculosis, bronchitis, heart diseases, cancers and asthma.

We explored the robustness of the pollution-respiratory disease relation using a variety of regression-based approaches, controlling for secular trends, seasonality and confounding effects of weather variables. These models include Generalized Linear Models (GLM), Generalized Additive Models (GAM), Parameter Driven Poisson Regression Models (PDM) and Transitional Regression Models (TRM). In each case, we considered models based on a Poisson distribution, incorporating over-dispersion and serial correlation where possible.

This study was undertaken to determine whether there is an association between air pollutants levels, climatic factors and number of hospital admissions for four respiratory diseases in Southern districts, Tamil Nadu.

Review of Literature

Koshal and Manjulika (1974) discussed a reduction of 50% in the air pollution would imply a social saving of the order of $1.9-$2.2 billion per year in terms of the respiratory diseases only and number of deaths, these saving would be even of the a higher order. Katsouyanni *et al.* (1990) studied air pollution and cause specific deaths in city of Athens in deaths were generally higher during the high pollution days in respiratory conditions, even though the number of deaths was smaller than another two categories in cardiac deaths and other deaths. Bente *et al.* (2003) studied traffic related air pollution including NO_2 and the Volatile Organic Compounds seemed to be of particular importance.

Chun-yuh yang *et al.* (2007) examined the association between levels of air pollutants and hospital admissions rates for asthma among the Taipei city in Taiwan. Ahmed *et al.* (2007) studied air pollution is not only the factor increasing the risk of Asthma in school children, some other factors such as respiratory diseases, infectious diseases, genetic and passive smoking also present a high-risk threat. Jennifer *et al.* (2009) contributed to adverse health for children living in areas with chronic exposure to higher levels of O_3 and $PM_{2.5}$ associated with children with lower exposures in a large nationwide sample of United States children.

Lokman (2009) determined hospital admissions due to respiratory diseases have been mostly affected adversely from pollutant concentrations compared to meteorological parameters other than temperature in the Balikesir city. Selma *et al.* (2010) evaluated the allergic diseases in adults are related to some geo-matric characteristic of the living area suggesting protective measures could come into priority in certain areas of the Turkey. Pieralhille *et al.* (2012) exposure to environmental pollution, even at typical low levels, can increase the risk of emergency room admission for acute respiratory diseases on exacerbations of obstructive lung diseases in the general population in Milan, the largest metropolis of Northern Italy. Hajat *et al.* (2015) investigated the relationship between air pollution and upper respiratory disease as reflected in number of consultations made at family practices on London.

MATERIALS AND METHODS

Hospital admission data in southern districts from May 2020 to April 2021 for respiratory diseases were obtained from the Nivya Respiratory Centre, Tirunelveli. Four diagnostic categories were Allergic Asthma, Allergic Bronchitis, Chronic Obstructive Pulmonary Disease (COPD) and Pulmonary Tuberculosis (PTB). In addition to hospital admission discharge data like hospital admission date, patient's age and gender are considered.

Daily observations of air pollution data such as sulphur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter (PM_{10}) were obtained from **cpcb.gov.in** and weather variables data such as temperature (degree Celsius), humidity (cubic meter) and wind speed (meters per second) for months May 2020 through to April 2021 were obtained from weatherbase.com and tnenvis.nic.in.Air pollution data is measured in $\mu g/m^3$.

Regression is a method for estimating the functional relationship between a dependent variable and one or more independent variables. The only difference among the models considered is in the outcome variable, related to the explanatory variable or variables in a linear function. Such models are collectively referred to as generalized linear model, generalized additive model, parameter driven model and transitional regression model.

The Poisson regression analysis was used the distribution of respiratory cases data is in a form of Poisson distribution. Poisson distribution appears to be appropriate when the response variable consists of nonnegative integers and is not normally distributed (9) and (10). Furthermore, the occurrence must be random and independent of each other. Poisson regression describes count or discrete data of the occurrences of some event over a specified interval (11).

For each of the different models we began with a base model comprising a temporal time trend, day-of-week effects, and sine and cosine pairs for the seasonal variation in the outcome variable. Each of the pollutants, with confounding weather variables, were entered based on its significance and AIC criterion.

We used a number of regression type approaches to describe the associations between ambient levels of air pollution, environmental factors and respiratory diseases. We began with a Generalized Linear Models (GLM) analysis. Let be the linear predictor defined by

$$\eta_i = g[E(y_i)] = g(\mu_i) = X'_i \beta$$

Note that the expected response is just

$E(y_i) = g^{-1}(\eta i) = g(X)$

We call the function g the link function.

A parameter driven model (PDM) by Zeger and Liang (1986) was used to incorporate the presence of serial correlation, where the serial correlation is set up through an unobservablelatent process.

The Poisson regression model can be written as

$y_i = E(y_i) + \varepsilon_i, \quad = 1, 2, \ldots, n$

We assume that the expected value of the observed response can be written as

$E(y_i) = \mu_i$

and that there is a function g that relates the mean of the response to a linear predictor, say

$E(y_i) = \mu_i$

$= \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$

$= x_i \beta$

The function g is usually called the link function. The relationship between the mean and the linear predictor is

$\mu_i = g^{-1}(\eta_i) = g^{-1}(x_i' \beta)$

There are several link functions that are commonly used with the Poisson distribution. One of these is the identity link

$g(\mu_i) = \mu_i = x_i' \beta$

When this link is used, since another popular link function for the Poisson distribution is the log link

$g(\mu_i) = In\ (\mu_i) = x_i' \beta$

For the above log link equation, the relationship between the mean of the response variable and the linear predictor is

$\mu_i = g^{-1}(x_i' \beta)$

$= e^{xi'\beta}$

The log link is particularly attractive for Poisson regression because it ensures that all of the predicted values of the response variable will be nonnegative. The method of maximum likelihood is used to estimate the parameters in Poisson regression.

In the development to generalized linear models, we use the link function g to relate the conditional mean μ(x) to the linear predictor η(x). But really nothing in what we were doing required η to be linear in x. In particular, it all works perfectly well if η is an additive function of x. We form the effective responses z_i as before and the weights w_i, but now instead of doing a linear regression on x_i we do an additive regression, using back tting (or whatever). This gives us a generalized additive model (GAM).

In addition, we utilized Transitional Regression Models (TRM) introduced by Brumback *et al.* (2000). Here, we present a special case of a TRM, dened as GLM with time series errors. For a Poisson with AR(1) errors the conditional mean is dened as

$$\mu_t = exp(X_i'\beta) + \psi_1 e_{t-1} \sqrt{\nu_t}$$

where $e_t = (Y_t - Y_t) / \sqrt{\nu_t}$ and $\nu_t = \exp\ (X_i'\beta)$.

Here e_t is scaled to give constant variance. Note that et = $\psi_1 e_{t-1} + \delta_t$ where $\{\delta_t\}$ is an independent series with zero mean.

Both the PDM and the TRM are adjuncts to the core regression models. However, the PDM and TRM differ with respect to modelling the serial correlation. The GLM with AR(1) errors, a special case of the TRM, is an observation driven model where the serial correlation is handled by including past values of the response variable as a covariate, whereas the serial correlation in a PDM is set up through an unobservable latent process. GLM, GAM, PDM and TRM all assume a Poisson distribution for respiratory diseases of hospital admissions data.

To allow comparison across models we used the following goodness-of-fit indices:

- Mean square error (MSE) = mean $\{(Y_t - Y_t)^2\}$, where Y_t are the (inverse link transformed) fitted values.
- Mean square proportional error (MSPE) = mean $\{(Y_t - Y_t)^2 / Y_t\}$.
- AIC = n log(σ^2)+2p, where σ^2 is the variance of the raw residuals (response minus fitted values), and p is the number of degrees of freedom in each model.

Results are presented as relative risks for an increase from the 10th to 90th percentile of the respective pollutant.

The associations between hospital admissions for respiratory diseases and air pollutants, weather variables were estimated using the odds ratio (OR) and their 95% confidence intervals (CI) were using statistical models. All statistical analysis were performed using R package.

RESULTS AND DISCUSSIONS

There were 350 hospital admissions with principal diagnoses of respiratory diseases during the month of May 2020 to April 2021.

Population and regional characteristics

Table 1 presents an overview of descriptive statistics of patient'sage, air pollutants and weather data. Air pollutants and weather variables vary considerably by region. The proportion of the population that is 208 male and 142 female are hospital admissions of respiratory diseases in southern districts.

Association of Air Pollutants with Allergic Asthma using different Statistical Models

Table 2 shows the characteristics of study subjects by air pollutants with allergic asthma and RRs for risk factors and percentage change in 95% CI associated with each air pollutants measure is reported together with the p value by using different statistical models.

Table 1. Descriptive analyses of the variables in the study

Variables		Mean	Std. Dev.	Minimum	Maximum
Age	Child	12.6	2.95	7	14
	Youth	22	2.31	20	24
	Adults	50.4	10.52	25	64
	Seniors	71.08	5.81	65	93
Pollutants:	SO_2	26.35	13.71	12	64
	NO_2	38.17	14.92	16	76
	PM_{10}	80.31	20.67	67	147
Weather Variables:	Temperature	76.19	2.61	74	84
	Humidity	79.85	5.46	55	89
	Wind Speed	6.07	0.84	4.34	7.33

Table 2. Association of Air Pollutants with Allergic Asthma using different Statistical Models

Pollutants	GLM		GAM		PDM		TRM	
	RR	UCL - LCL	RR	UCL - LCL	RR	UCL - LCL	RR	UCL – LCL
SO_2	0.0129	0.0118-0.0141	0.0253	0.0240-0.0259	0.0150	0.0139-0.0162	0.0289	0.0272-0.0297
NO_2	0.0937	0.0924-0.0949	0.0167	0.0152-0.0171	0.0115	0.0102-0.0124	0.0224	0.0211-0.0236
PM_{10}	0.0799	0.0781-0.0811	0.0113	0.0100-0.0122	0.0548	0.0533-0.0561	0.0262	0.029-0.0275
MSE	6.6138	4.8868	4.4140	4.3161				
MSPE	0.0066	0.0298	0.0270	0.0198				
AIC	481	372	236	738				

Table 3. Association of Air Pollutants with Allergic Bronchitis using different Statistical Models

Pollutants	· GLM		GAM		PDM		TRM	
	RR	UCL - LCL	RR	UCL - LCL	RR	UCL - LCL	RR	UCL – LCL
SO_2	0.0103	0.0091-0.0112	0.0253	0.0240-0.0263	0.0693	0.0679-0.0709	0.0200	0.0191-0.0212
NO_2	0.0937	0.0922-0.0949	0.0167	0.0152-0.0178	0.0516	0.0502-0.0524	0.0136	0.0126-0.0143
PM_{10}	0.0799	0.0781-0.0813	0.0113	0.0101-0.0124	0.0554	0.0543-0.0562	0.0136	0.0122-0.0147
MSE	6.6138	4.8868	5.9147	1.3371				
MSPE	0.02467	0.9298	0.0771	0.1332				
AIC	189	172	147	352				

Table 4. Association of Air Pollutants with COPD using different Statistical Models

Pollutants	GLM		GAM		PDM		TRM	
	RR	UCL - LCL	RR	UCL - LCL	RR	UCL – LCL	RR	UCL – LCL
SO_2	0.0102	0.0091-0.0112	0.0252	0.0240-0.0259	0.0186	0.0172-0.0196	0.0429	0.0420-0.0437
NO_2	0.0937	0.0924-0.0946	0.0167	0.0159-0.0174	0.0997	0.0982-0.0109	0.0167	0.0159-0.0173
PM_{10}	0.0799	0.0783-0.0809	0.0113	0.0109-0.0119	0.0164	0.0156-0.0173	0.0291	0.0280-0.0308
MSE	6.6138	4.8868	1.5987	1.7371				
MSPE	0.9131	0.9298	0.1490	0.3997				
AIC	282	364	307	632				

Table 3 shows the characteristics of study subjects by air pollutants with allergic bronchitis and RRs for risk factors and percentage change in 95% CI associated with each air pollutants measure is reported together with the p value by using different statistical models.

Table 4 shows the characteristics of study subjects by air pollutants with COPD and RRs for risk factors and percentage change in 95% CI associated with each air pollutants measure is reported together with the p value by using different statistical models.

Table 5 shows the characteristics of study subjects by air pollutants with PTB and RRs for risk factors and percentage change in 95% CI associated with each air pollutants measure is reported together

Table 5. Association of Air Pollutants with PTB using different Statistical Models

Pollutants	GLM		GAM		PDM		TRM	
	RR	UCL - LCL	RR	UCL - LCL	RR	UCL - LCL	RR	UCL – LCL
SO_2	0.0103	0.0092-0.0113	0.0252	0.0243-0.0261	0.0302	0.0294-0.0314	0.0223	0.0214-0.0231
NO_2	0.0937	0.0926-0.0942	0.0167	0.0156-0.0172	0.0242	0.0235-0.0255	0.0192	0.0186-0.0199
PM_{10}	0.0799	0.0783-0.0806	0.0113	0.0104-0.0122	0.0528	0.0516-0.0536	0.0305	0.0297-0.0310
MSE	6.6138	4.8868	4.5753	3.0765				
MSPE	0.2853	0.9298	0.2983	0.1776				
AIC	238	247	275	531				

with the p value by using different statistical models.

CONCLUSION

In this study we are not arguing for any particular method of analysis, the conclusions drawn are highly dependent on the choice of model used. This study is common in practice and the length of series we used here is typical of what is used in practice. Consequently, our study has important implications for the type of analysis that is commonly carried out in airpollution and four respiratory diseases. The findings from this study show that the relationship between ambient concentrations of nitrogen dioxide and four respiratory diseases of hospital admissions isrobust to different statistical methodology.

REFERENCES

Ahmed SqualliHoussaini, HafidaMassaouri, IsaadNasri, Marie Roth, ChakibNejjari and Mohammed Benchekroun, 2007. Air Pollution as a Determinant of Asthma among Schoolchildren in Mohammedia, Morocco. *International Journal of Environmental Health Research*. 17 (4): 243-257.

Bente Oftedal, Per Nafstad, Per magnus, Sonja Bjorkly and Anders Skrondal, 2003. Traffic Related Air Pollution and Acute Hospital Admission for Respiratory Diseases in Drammen, Norway 1995-2000. *European Journal of Epidemiology.* 18(7): 671-675.

Chun-Yuh Yang, Chih-Cheng Chen, Chi-Yu Chen and Hsin-We Kub, 2007. Air Pollution and Hospital Admissions for Asthma in a Subtropical city: Taipei, Taiwan. *Journal of Toxicology and Environmental Health*, *Part A.* 70 :111-117.

Hajat, S., Anderson, H.R., Atkinson, R.W. and Haines, A. 2015. Effects of Air Pollution on General Practitioner Consultations for Upper Respiratory Diseases in London. *Occupational and Environmental Medicine*. 59 : 294-299.

Jennifer, D. Parker, Lara, J. Akinbami and Tracey J. Woodruff, 2009. Air Pollution and Childhood Respiratory Allergies in the United States. *Environmental Health Perspectives.* 117 (1) : 140-147.

Katsouyanni, K., Karakatsani, A., Messari, I., Touloumi, G., Hatzakis, A., Kalandidi, A., Trichopoulos, D. 1990. Air Pollution and Cause Specific Mortality in Athens. *Journal of Epidemiology and Community Health*, Vol. 44, pp.321-324.

Lokman Hakan Tecer, 2009. A Factor Analysis Study: Air Pollution, Meteorology and Hospital Admissions for Respiratory Diseases. *Toxicological and Environmental Chemistry*. 91 (7) :1399-1411.

Pierachille Santus, Antonio Russo, Enzo Madonini, Luigi Allegra, Francesco Blasi, Stefano Centanni, Antonia Miadonna, Gianfranco Schiraldi and Sandro Amaducci, 2012. How Air Pollution Influences Clinical Management of Respiratory Diseases. A Case Crossover Study in Milan. *Respiratory Research*. 13 : 95-106.

Koshal, R.K. and Manjulika, K. 1974. Air Pollution and the Respiratory Disease Mortality in the United States - A Quantitative Study. *Social Indicators Research*. 1(3) : 263-278.

Ronald N. Forthofer, Eun Sul Lee and Michael Hernandez, 2007. *Biostatistics: A Guide to Design, Analysis and Discovery,* A Division of Reed Elsevier India private Limited, New Delhi - 110001.

Saxena, H.M. 2006. *Environmental Studies*, Rawat Publications, Jaipur - 302004.

Selma Metintas, Emel Kurt and PARFAIT Study Group, 2010. Geo-Climate Effects on Asthma and Allergic Diseases in Adults in Turkey: Results of PARFAIT Study. *International Journal of Environmental Health Research.* 20 (3) : 189-199.

Zeger, S.L. and Liang, K.Y. 1986. Longi-tudinal data analysis for discrete and continuous outcomes. *Biometrics.* 42(1) : 121-130.