Original Article Pol J Public Health 2019;129(3): 72-74

MIECZYSŁAW SZYSZKOWICZ

Application of quantile regression in environmental epidemiology

Abstract

Introduction. Among many problems present in studies evaluating associations between health conditions and exposure to ambient air pollution, there is the correlation between environmental factors. These issues are usually resolved by providing a correlation matrix for the parameters of interest.

Aim. To explore correlations between environmental factors.

Material and methods. As sample data we use environmental factors presented in Milan mortality data (Italy, 1980-1989) and emergency department visits for asthma in Windsor (Canada, 2004-2010). Here, we propose to use a series of quantile regression evaluations to emphasize and identify dependency among environmental factors.

Results. This presentation outlines an important role to investigate the potential correlations among ambient air pollutants, weather factors, and the values of the Canadian Air Quality Health Index (AQHI). In environmental epidemiology studies, these components are usually used in a common statistical model. Their correlations affect the values of the estimated relative risks, odds ratios or other estimated health effects. The presented approach examines associations among the factors as well as changes in correlations along quantiles. The examples used in this study explain various environmental phenomena; for example, the negative relationship between ambient ozone and nitrogen dioxide.

Conclusions. By a consequence, this work can aid in further developing policies aimed at reducing the health impacts of air pollution as it allows to identify highly correlated factors in the constructed models.

Keywords: air pollution, ambient, exposure, correlation, model.

DOI: 10.2478/pjph-2019-0017

INTRODUCTION

In many environmental epidemiological studies, ambient air pollutant concentration levels are used to represent exposure [1]. Usually, a fundamental goal of such studies is to estimate potential associations between health outcomes and the defined exposure [2]. The main results of these studies are presented as the assessed relative risk, proportional hazard or odds ratio values. Traditionally, the reported results are accompanied by descriptive statistics on air pollutants and weather factors including their mean, median, some specific percentiles, minimum, and maximum value. In the case of examining more than one ambient air pollutant, a correlation matrix among the air pollutants is usually also provided [3]. Correlation between two variables provides information on statistical relationships including dependence. It is a measure of the linear connection between two variables. These correlations may have strong effects on the estimated associations of the exposure with health endpoints. Very often in multi-pollutant models a signal for associations of an individual air pollutant may be absent when examining that air pollutant with others in the same statistical model.

AIM

The purpose of the article is to explore correlations between environmental factors.

MATERIAL AND METHODS

Here we propose to use a quantile regression technique to more thoroughly investigate the relationships among analyzed environmental variables. This proposed approach allows one to identify relationships between variables more accurately than simple linear correlation. Rather than having just one number to characterize the estimated correlation between two variables, it is possible to see the association along a series of the specified quantiles.

To simplify our material, we are only presenting references on the methodology and its statistical background [4,5]. Simplified quantile regression produces coefficient estimations when regressed on the quantile of a response. Consequently, it provides information regarding which quantiles are primarily affected in relation to the independent variable.

In this work, for an illustrative purpose, we consider two sets of databases. One database contains the daily mortality counts in Milan, Italy. Other data were organized to investigate

Environmental Health Science and Research Bureau, Health Canada, Canada

the associations between emergency department (ED) visits for asthma in Windsor with ambient air pollution exposure. For the exposure data, the air quality health index (AQHI) was applied. The AQHI value is reported hourly and it encapsulates three ambient air pollutant concentrations (fine particulate matter, nitrogen dioxide, and ozone) weighted by the estimated (fixed) risk coefficients [6]. Here we are using an average of 24 measurements of the AQHI to represent daily levels.

Milan mortality data (Milan, Italy)

The Milan mortality data contain daily mortality counts for 3652 consecutive days for the period from January 1, 1980 to December 30, 1989. The data are freely accessible from the package SemiPar in the R software [7]. In our analysis we do not use the health responses included in these data (number of deaths), instead, we analyzed the air pollutants and weather variables which accompanied the health data. There are a few publications using these data which reports the results on the associations between the number of deaths and exposure to total suspended particulates (TSP). These publications also list descriptive statistics for the environmental variables included in the Milan mortality database [8,9]. In this paper, we consider TSP values as the dependent variable and other environmental parameters present in the database as independent variables. We are using the following convention: Y symbolizes dependent variable, and X represents one or more independent variables. Using the Milan mortality data, we use regression of the form Y=X, where Y=TSP and X= ambient sulphur dioxide, temperature, and relative humidity.

Windsor asthma data (Windsor, Canada)

These data were organized to investigate ED visits for asthma in Windsor, Canada, for the period from April 2004 to December 2010 [10]. Again, it did not concern the health data that were studied but rather the properties of the ambient factors that were of interest. The environmental data used here can be easily reconstructed by retrieving the recorded concentration values from the National Air Pollution Surveillance network (NAPS, See NAPS Web site: https://www.ec.gc.ca/rnspa-naps/). In this example we consider two kinds of regression; one is of the form Y=AQHI and X=sulphur dioxide, temperature, and relative humidity; and the other is Y=ozone and X= temperature, nitrogen dioxide, and fine particulate matter (PM2.5). Two regressions for fine particulate matter are illustrated and were considered in warm (April - September) and cold (October - March) periods.

We used SAS software to calculate the quantile regressions. Analyses were performed in SAS 9.3 (SAS Institute, Cary NC) using the QUANTREG procedure for quantile regression models. We defined a series of the quantiles (from 0.02 to 0.98 by 0.02) for the considered dependent variable. The quantile regressions were performed and the estimated regression coefficients and their 95% confidence interval (CI) limits were plotted along quantile values. Such a plot allows us to visualize and identify quantiles of the dependent variable which were affected by the independent variables.

RESULTS

The numerical results are series regression coefficients with 95% CI values. These values are plotted for the corresponding quantile levels. The main results of this presentation are shown in the form of two figures with four panels each.

Figure 1 represents the estimated parameters by quantile levels for TSP in Milan, showing intercept and coefficients for sulphur dioxide, temperature and relative humidity. The results indicate that TSP quantiles have positive and statistically significant regression coefficients for ambient sulphur dioxide. Relative humidity also has positive coefficients which are statistically significant above the value of 0.3 quantile level, and the coefficients are increasing with increasing quantile levels. The estimated intercept shows levels of TSP for the specified quantiles.

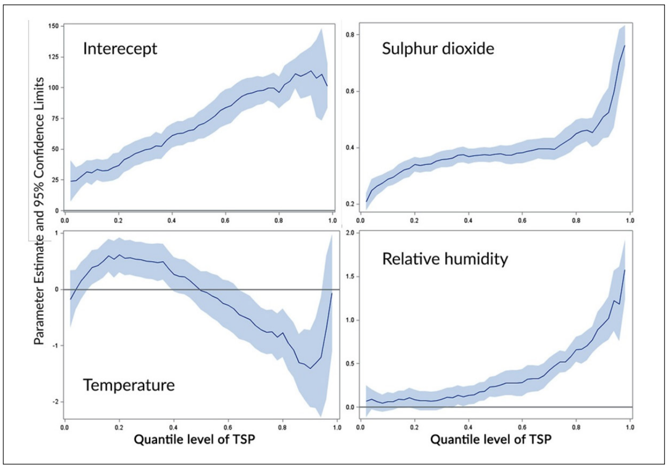

FIGURE 1. Quantile regression coefficients and their 95% CIs for variables from the Milan mortality data, Italy, 1980-1989.

Figure 2 shows the results related to Windsor. It is interesting to observe, very clearly marked on the figure, the associations between quantiles and the values of the regression coefficients for sulphur dioxide. The coefficients are positive and statistically significant. The relationship indicates that the estimated AQHI also maps the concentration levels of ambient sulphur dioxide; even though sulphur dioxide is not used in the formula to estimate the AQHI. The association for ambient carbon monoxide (data not shown) was represented by an almost flat horizontal line with statistically non-significant coefficients.

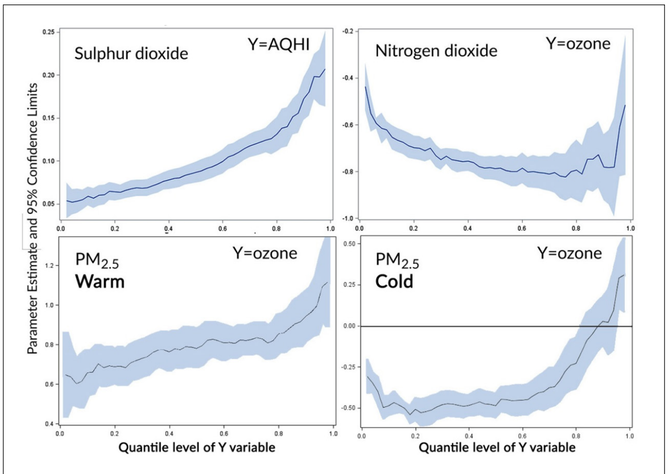

FIGURE 2. Quantile regression coefficients and their 95% CIs for variables from the data related to the study in Windsor (warm: April - September, cold: October - March), Canada, 2004-2010.

Figure 2 also illustrates the associations for ambient ozone (Y=ozone) and two ambient air pollutants, nitrogen dioxide and fine particulate matter, regressed separately in addition to ambient temperature. Nitrogen dioxide has a negative statistically significant relationship with the quantile values of ambient ozone. In contrast, ambient fine particulate matter has a positive significant association and the values of the coefficients are increasing in the warm (April – September) period. This may suggest that the higher quantiles of ozone are more related to particulate matter in the air and volatile organic components or other organic and nonorganic particulate matters. In the cold (October – March) period we have negative statistically significant associations. Only for a high quantile (above 0.9) the coefficients are positive.

DISCUSSION

The proposed approach is a simple technique but provides additional useful information on the correlations between considered environmental variables. Very often in environmental epidemiology we are interested in various concentration levels. The presented method shows correlations along the considered concentration values. Traditional approach gives just one value of correlation coefficient for two data sets. For example, using the Excel function CORREL (it estimates correlation coefficients), for the Milan data we obtained the following values of the coefficient correlations: -0.44, 0.17, and 0.63, between TSP and temperature, relative humidity, and sulphur dioxide, respectively. As we see (Figure 2.; ozone and fine particulate matters) the correlation may change its forms and intensities by the climate seasons. Knowledge of the form and intensity of the correlation allows constructing the statistical models with adequate adjustments for the parameters in the used statistical model; include/exclude, amount of smoothing or other adjustments.

In our opinion, such relatively fast and elegant analyses should always be done when studying more than one air pollutant in a single model. The shapes formed by the estimated coefficients allow for more thorough examination of the correlations between two variables. The analysis also can be done for lagged exposures or, as in the case of the AQHI, multiple components, such as Y=AQHI, and X=PM2.5, NO2 and O3. Quantile regression allows the impact of individual air pollutants to be assessed against multiple values, such as air pollutants in the AQHI index.

CONCLUSIONS

Air pollutants may negatively affect various aspects of human health. Air pollutants of interest are frequently correlated. Such an association for two parameters is usually reported as one value, between -1 and 1. This value characterizes the connection between the two factors. In this study, we propose to determine and visually represent the relationship between parameters at various levels of their quantiles. This technique allows a better understanding of the interconnection between two air pollutants. The data used in this study explain various environmental phenomena; for example, the negative relationship between ambient ozone and nitrogen dioxide. This type of information is useful for refining statistical models that estimate health risk of ambient air pollutants. Consequently, this work can aid in further developing policies aimed at reducing the health impacts of air pollution as it allows identification of highly correlated factors in the constructed models.

Funding: This research received no external funding
Conflicts of Interest: The author declares no conflict of interest.

REFERENCES

1. Szyszkowicz M. Remarks on ambient air pollution and health outcomes. ISRN Public Health. 2013. http://dx.doi.org/10.1155/2013/846297.
2. Szyszkowicz M, Kousha T, Castner J, Dales R. Air pollution and emergency department visits for respiratory diseases: A multi-city case crossover study. Environ Res. 2018;163:263-9.
3. Szyszkowicz M, Thomson EM, Colamn I, Rowe BH. Ambient air pollution exposure and Emergency Department visits for substance abuse. PLoS One. 2018;13(6):e0199826.
4. Koenker R, Bassett G. Regression Quantiles. Econometrica. 1978;46:33-50.
5. Koenker R, Hallock KF. Quantile regression. J Econ Perspect. 2001;15(4):143-56.
6. Stieb DM, Burnett RT, Smith-Doiron M, Brion O, Shin HH, Economou V. A new multipollutant, no-threshold air quality health index based on short-term associations observed in daily time-series analyses. J Air Waste Manag Assoc. 2013;58:435-50.
7. Wand M. The package "SemiPar" in the R library. In the package milan mortaity data frame. Functions for semiparametric regression analysis, to complement the book: D. Ruppert, M. P. Wand, R. J. Carroll. Semiparametric Regression. Cambridge University Press; 2015. http://www.use-r.com/SPmanu.pdf.
8. Zanobetti A, Wand MP, Schwartz J, Ryan LM. Generalized additive distributed lag models: quantifying mortality displacement. Biostatistics. 2000;1(3):279-92.
9. Szyszkowicz M, Burr WS. Distributed Lag Models: an analysis of milan mortality aata. J Pollut Eff Cont. 2014;2:109.
10. Szyszkowicz M, Kousha T. Emergency department visits for asthma in relation to the Air Quality Health Index: a case-crossover study in Windsor, Canada. Can J Public Health. 2014;105(5):336-341.

Corresponding author
Dr Mieczysław Szyszkowicz
Environmental Health Science and Research Bureau, Health Canada
101 Tunney's Pasture, Ottawa, ON, K1A 0K9, Canada
E-mail: mietek.szyszkowicz@canada.ca