© Nofer Institute of Occupational Medicine, Łódź, Poland

International Journal of Occupational Medicine and Environmental Health 2006;19(4):224–7
DOI 10.2478/v10001-006-0032-7

USE OF GENERALIZED LINEAR MIXED MODELS TO EXAMINE THE ASSOCIATION BETWEEN AIR POLLUTION AND HEALTH OUTCOMES

MIECZYSŁAW SZYSZKOWICZ

Air Health Effects Division
Health Canada, Ottawa, Ontario; Canada

Abstract
Background: Time-series and case-crossover are two techniques that are widely used for assessing the short-term impact of ambient air pollution exposure on health. **Materials and Methods:** The generalized linear mixed model (GLMM) methodology is proposed here to study the association between ambient air pollution and health outcomes. Poisson random-effects models are applied to analyze the clustered counts, where the groups of days, determined by the triplet <day of week, month, year>, form the clusters. The proposed technique uses a nested structure for the clusters and allows random-effects for hierarchical factors. A random intercept in the models adjusts for different levels of counts among the clusters. A fixed slope represents a common response to the exposure. **Results and Conclusions:** The obtained results are consistent with those generated by a classical approach (for example the case-crossover technique). The GLMM technique is a valid alternative methodology for studying air health effects.

Key words:
Case-crossover, Cluster, Mixed effect, Poisson model, Random effect, Slope

INTRODUCTION

The time-series and case-crossover designs are frequently used for assessing the association between exposure to ambient air pollution and numbers of health events over time. In the time-series design, a sequence of numbers of health events, values of pollutant concentrations and confounders are considered in the time scenario. In this approach, time-dependent confounders are controlled by modeling. Sophisticated analytical tools have been developed to better control for potential confounders and to adjust for seasonal trends in the data [1]. Thus seasonal cycles, secular trends and days of the week are incorporated and controlled in the constructed model.

The case-crossover design is an alternative to time-series analysis [2]. This method is essentially an adaptation of the case-control study where cases serve as their own controls. The method is event driven. A subject's exposure to external factors at the time of the health event (case period) are compared with another time period when the subject was a non-case (control period). Matching of controls to case periods by day of the week adjusts for the influence of this important factor [3].

The purpose of this paper is to present an alternative to these methods. The presented technique is based on Poisson regression applied to clustered counts. In the time-series design days of the week are treated as factors and they are usually used in fitted models. In the case-crossover technique controls are matched to be on the same day of the week as the case. The method proposed here uses days of the week to construct clusters. A cluster contains four or five days of the same day of the week.

The clusters are formed naturally based on a calendrical structure; days are nested in the same days of the week,

Received: October 23, 2006. Accepted: November 24, 2006.
Address reprint requests to: Dr. M. Szyszkowicz, Air Health Effects Division, Health Canada, 269 Laurier Avenue, Room 3-030, Ottawa, On, K1A 0K9, Canada (e-mail: mietek_Szyszkowicz@hc-sc.gc.ca).

which are nested in the same month, and the months are nested in the same year. Each month forms seven clusters. The hierarchical construction of the clusters allows the model to incorporate level specific random effects. The behavior of an individual is associated with this hierarchical structure and by consequence many health outcomes are cluster-dependent. Also some pollutants are related to the calendrical structure of the cluster hierarchy, mainly those generated by traffic and industry.

The numbers of events, air pollution levels, temperature and other covariates have different values on the constructed clusters. All seasonal effects, trends and cycles observed along time are now broken and absorbed by the clusters. In this methodology, time is only applied to define clusters and to incorporate the nested structure. Each cluster has its own identity and exists independently from others. The idea behind this approach is to apply a generalized linear mixed-effects model to analyze clustered counts. The Poisson model is fitted with log as a link function. It is clear that some clusters have consistently higher outcomes than others. These differences are captured and controlled by including the random intercept option in the models. On the other hand it is assumed that the response (slope) to the pollutant is the same among the clusters. The constructed models use a fixed slope option to satisfy the assumption of a universal response.

MATERIALS AND METHODS

Data

Real data can be used to illustrate the behavior of the methods discussed here. In this example, the summarized numbers of daily emergency department (ED) visits, related to cardiac problems in a hospital in Vancouver, Canada, represent counts of health events. The period of study was from January 1999 to December 2002. During these 1461 days there were 5317 cases diagnosed and recorded as ED cardiac visits. The study population consists of people serviced by the emergency department. A 3-day lagged daily mean value of ground level ozone (O_3) is considered as the shared ambient air pollution exposure for this population. We have data for 1461 days and the following variables are used in the constructed models: daily counts of cardiac related problems visits, shared exposure, and calendar time (day, month, year).

Classical method

One of the classical methodologies is the case-crossover method. The main problem in the case-crossover design is to choose a schema to determine controls for a case. Recently, time-stratified referent selection strategy has been favored as a method to match controls to case. As already suggested, this version of the referent selection is localizable and ignorable, which are desirable properties of the case-crossover method [3]. The *phreg* procedure developed by the SAS Institute Inc. [4] was applied to fit the model. In this case only the pollutant is used on the right hand side of the model. Days of the week are applied as indicators to choose controls in the same month as cases.

Mixed-random effects model

The generalized linear mixed models (GLMMs) extend GLMs by allowing for random, or cluster-specific effects in the linear predictor. The inclusion of random effects in the linear predictor reflects the idea that there is natural heterogeneity across clusters in their regression coefficients. The likelihood related to a GLMM involves an integral, which cannot in general be calculated explicitly. Three different approaches are used to calculate this integral, which can be briefly identified as: Laplace approximation, penalized quasi likelihood (PQL) calculation, and Gauss-Hermite numerical quadrature. Diggle et al. [5], and Molenberghs and Verbeke [6] are useful references for the full details on GLMMs. This new approach uses an implementation of this methodology. We used the R statistical package and its two functions; *glmmPQL* and *lmer* [7]. To enable the Gauss numerical quadrature approach to calculate the likelihood integral, the *gllamm* function [8] was also used. The *gllamm* function applies Gauss-Hermite quadrature with different options to control the number of points used in numerical integration. Two other functions, *glmmPQL* and *lmer,* were used with the PQL approach to calculate the integral. The *lmer* function has two other options, Laplace and adaptive Gauss quadrature (AGQ), but the recent version does not provide this for the AGQ method. To simplify the explanation, one of the possible models is presented below.

glmmPQL(Cardiac ~ *O3L3*, random = list(year = ~1,
month = ~1, day of week = ~1),
family = poisson()).

The R syntax is case sensitive thus spelling of Poisson family is in small letters. The model has two parts: fixed and random. In the random part the nested structure of the clusters is expressed. A fixed slope and a random intercept option (the notation ~1) are declared in the model. We can also study a random slope with respect to hierarchical levels.

RESULTS

The data and results are presented in generic form; thus a risk estimation, pollutant details, and confounders are not provided. The models have been simplified to focus on the proposed methodology. By consequence, the pollutant is the only variable used as an independent variable. It should be clear that other covariates may be easily incorporated into the constructed models. In other statistical software, if necessary, natural cubic splines can be added as the smoother for chosen covariates. Results are reported in the form of the estimated slopes and their standard errors.

Table 1 contains a summary of the results for 3-level nested clusters. The results from the case-crossover method and GLMMs are close, but not identical. The GLMM functions also provide estimation of variances of random effects. For example, the *gllamm* function estimated the following variances for random effects: 0.0145, 0.0045, and 0.0233 for days of the week, months and years, respectively.

The regular Poisson model (calculated in STATA as: *Poisson Cardiac O3L3*) estimated the slope as 0.0061 (standard error = 0.0018). This model provided the value of log likelihood = -3078.6, and the *gllamm* function on 3-level clusters provided log likelihood = -3062.8. We have a statistically significant improvement in the fit by using the GLMM technique.

Table 1. The results from different methodologies (software). The generalized linear mixed models (GLMM) technique was used on the nested structures; <day of week, month, and year>

Software	Slope	SE
phreg	0.0052	0.0029
gllamm	0.0061	0.0021
glmm PQL	0.0063	0.0022
lmer {PQL}	0.0063	0.0021
lmer {Laplace}	0.0062	0.0021

SE – standard error.

For illustrative purposes we used the same responses (all cardiac cases) and exposure to nitro dioxide (NO_2) and temperature in the current day. Table 2 shows the results. The obtained slope and its standard error may be used to estimate the relative risk for ED visits for cardiac problems related to nitrogen dioxide exposure.

Table 2. The results from two approaches: the classical Poisson model (log likelihood = -3080.8, from the *poisson* function) and the Poisson model (log likelihood = -3076.2, from the *gllamm* function) on the clusters

Approches	Slope	SE	P	95%CI	
Poisson:					
NO_2	0.0082	0.0032	0.012	0.0018,	0.0147
Temperature	0.0035	0.0028	0.230	-0.0022,	0.0091
gllamm:					
NO_2	0.0068	0.0009	<0.0005	0.0051,	0.0085
Temperature	0.0026	0.0015	0.083	-0.0003,	0.0055

SE – standard error; P – level of significance; CI – confidence interval.

DISCUSSION AND CONCLUSIONS

The time-series and case-crossover designs are well established as methods to examine the association between exposure to ambient air pollution and health outcomes. Both methods are supported by a broad methodological literature. Many scientific papers have been published with results based on time-series and case-crossover techniques.

The GLMM approach proposed here is a new technique in the sense of using it to examine the association between air pollution and health. The GLMM methodology is well documented and has been implemented for various purposes. The results presented here show that this technique is also valid for use in the air pollution health effects domain.

REFERENCES

1. Hastie TJ, Tibshirani RJ. *Generalized additive models*. London: Chapman and Hull; 1990.
2. Maclure M. *The case-crossover design: A method for studying transient effects on the risk of acute events.* Am J Epidemiol 1991;133:144–53.
3. Jane H, Sheppard L, Lumley T. *Case-crossover analyses of air pollution exposure data. Referent selection strategies and their implications for bias.* Epidemiology 2005;16:717–26.
4. *SAS Version 8.* Cary (NC): SAS Institute Inc.; 2004.
5. Diggle P, Heagerty P, Liang KY, Zeger SL. *Analysis of longitudinal data.* Oxford: Oxford University Press; 2002.
6. Molenberghs G, Verbeke G. *Models for discrete longitudinal data.* New York: Springer; 2005.
7. R Development Core Team. R: *A language and environment for statistical computing* [cited 2006 Oct 23]. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org.
8. Rabe-Hesketh S, Skrondel A, Pickles A. *The gllamm module (STATA), 2005* [cited 2006 Oct 23]. Available from:. http://www.gllamm.org.