COMPARING MODELS OF THE EFFECT OF AIR POLLUTANTS ON HOSPITAL ADMISSIONS AND SYMPTOMS FOR CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Mehmet Ali Cengiz, Yuksel Terzi
Department of Statistics, University of Ondokuz Mayıs, Samsun, Turkey

SUMMARY

There is an increasing interest in the use of hospital admission for Chronic obstructive pulmonary disease (COPD) in studies of short-term exposure effects attributed to air pollutants. However, little is known about the effect of air pollutants on COPD symptoms. This study was undertaken to determine whether there was an association between air pollutant levels and both hospital admissions and symptoms for COPD. For model comparison, we present Generalized Linear Model, Generalized Additive Model and a general approach for Bayesian inference via Markov chain Monte Carlo in generalized additive model. Furthermore, for comparing the predictive accuracy, Artificial Neural Networks (ANN) approach is given.

Key words: chronic obstructive pulmonary disease, Generalized Additive Model, Bayesian, WinBUGS, hospital admission, air pollution

Address for correspondence: M. A. Cengiz, Department of Statistics, University of Ondokuz Mayis, Samsun, Turkey. E-mail: macengiz@omu.edu.tr

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a group of diseases characterized by airflow obstruction that can be associated with breathing-related symptoms (e.g., cough, exertional dyspnea, expectoration, and wheeze). There is an increasing interest in the use of hospital admission data in studies of short-term exposure effects attributed to air pollutants. Considerable attention has been paid to vulnerable individuals such as subjects suffering from a chronic obstructive pulmonary disease (COPD). Numerous studies have investigated the relationship between air pollution and hospital admissions for COPD (1–18).

Generalised Additive Model (GAM) (19) has become the most widely used method for assessing the short-term health effects of air pollution. GAM models provide a flexible alternative to parametric regression models. GAM provides a powerful class of models for modelling nonlinear effects of continuous covariates in regression models with non-Gaussian responses. A huge variety of competing approaches are now available for modelling and estimating nonlinear functions of continuous covariates. Prominent examples are smoothing splines (20), local polynomials (21), regression splines with adaptive knot selection (22–24) and P-splines (25, 26). Currently, smoothing based on mixed model representations of GAMs and extensions is extremely popular (27–30). Schwartz and Marcus focuse on GAM for model selection in multiple Poisson regression for modelling associations between air pollution and increases in hospital admissions for respiratory disease (7).

The Bayesian inference for generalized additive model enjoys the flexibility of nonparametric models and the exact inference provided by the Bayesian inferential machinery. It is this combination that makes Bayesian nonparametric modelling so attractive (30, 31). Bayesian approaches are currently either based on regression splines with adaptive knot selection (32–37) or on smoothness priors (20, 38, 39). Crainiceanu et al. provide a simple set of programmes for the implementation of nonparametric Bayesian analysis in WinBUGS using Penalized Spline Regression (40). Brezger and Lang provide Bayesian semiparametric regression based on smoothness priors and Markov chain Monte Carlo (MCMC) simulation techniques (41).

While adverse effects of exposure to air pollutants and hospital admissions for COPD are well studied, little is known about the effect of air pollutants on COPD symptoms. This study focuses on modelling air pollution and both symptoms and hospital admissions (assuming that consecutive outcomes are independent) for COPD to the Afyon Respiratory Disease Hospital between 1 October 2007–30 September 2009.

The goal of this paper is not to discuss generalized additive model, Bayesian methodology, or provide novel modelling techniques. Firstly, we compare GLM (Multiple Poisson Regression), GAM and GAM with a Bayesian approach in WinBUGS (42), which has become the standard software for Bayesian analysis for modelling the association between air pollution, and both symptoms and hospital admissions for COPD using Akaiki Information Criteria (AIC), Bayesian Information Criteria (BIC), and Deviance Information Criteria (DIC).

Secondly, this study compares the predictive accuracy of four modelling techniques for modelling the association between air pollution, and both symptoms and hospital admissions for COPD. The methods used are GLM (Multiple Poisson Regression), GAM, Bayesian GAM and Artificial Neural Networks (ANN).

MATERIALS AND METHODS

GAM is a statistical model for blending properties of generalized linear models with additive models. For a Generalized Linear Model (GLM) with a log link function, we specify the expectation of a random variable Yt as

$$E(Y_t|X_t) = \exp\left(\beta_0 + \sum_{i=1}^{r} \beta_i X_{t,i}\right) \quad (1)$$

Refer to (43) for a detailed discussion of GLMs. Here Yt denotes counts of the records of hospital admissions from respiratory disease and $x_t=(x_{t,1},..., x_{t,r})'$ denotes the explanatory variables at time t. We assume an overdispersed Poisson model estimated using a quasi-likelihood approach.

A nonparametric alternative to the parametric GLM is the Generalized Additive Model (GAM). GAM allows non-linear relations between the response variable and each explanatory variable (19). For GAM, we assume

$$E(Y_t|X_t) = \exp\left(\beta_0 + \sum_{i=1}^{r} g_i(X_{t,i})\right). \quad (2)$$

where each g_i is a smooth, possibly non-linear, univariate function. Any of the g_i can be made linear to obtain a semi-parametric model. As with GLM, we use quasi-likelihood estimation. Cubic smoothing spline's were used to estimate the non-parametric functions g_i.

This study presents the relations between the whole hospitalized patients, cases with respiratory disease in the Afyon Respiratory Disease Hospital and the measures of air pollution at the city centre. This study was performed by retrospective evaluation of the patient's records from 1 October 2006–30 September 2009. SO_2 (Sulfur dioxide)-PM10 (Particulate matter) values related to the same period were extracted from the archives of the Afyon Environmental Department Air Pollution Unit. Weekly records of hospital admissions, number of patients with caught, number of patients with exertional dyspnea, and number of patients with expectoration for COPD were obtained from the Afyon State Hospital for the period from October 2006–November 2009. Weekly average levels of SO_2 and PM10 were obtained from the environmental state agency. Weekly counts of hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration for COPD were considered as the dependent variable of pollutants in Poisson regression model.

The first goal of this analysis is to identify associations between air pollution and hospital admissions, number of patients with cough, number of patients with exertional dyspnea and number of patients with expectoration for COPD using multiple Poisson regression in GLM context. So we used PROC GENMOD in SAS ver. 9.3 Software to investigate the relationship among hospital admissions, number of patients with cough, number of patients with exertional dyspnea, number of patients with expectoration, and the predictors (SO_2 and PM10). We specified the lag between exposure (weekly admission counts and the number of patients) and response (weekly average levels of pollutants).

RESULTS

Results of the independent variable effects analysis are shown in Table 1.

The analysis of parameter estimates results show that the effect of PM10 on hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration is highly significant and the effect of SO_2 on hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration is insignificant at the 5% level.

Standard multiple Poisson regression assumes a strict linear relationship between the response and the predictors. We now investigate less restrictive models using GAM with moderately flexible spline terms for each of the predictors. We prefer additive models using a univariate smoothing spline for each term. Each term is fit using a univariate smoothing spline with three degrees of freedom.

We obtained the plots of predictions against predictor given in Figure 1 using PROC GAM in SAS. The plots in Figure 1 show that the partial predictions corresponding to SO_2 have a quadratic pattern while PM10 have relatively a linear pattern for all models.

The plots in Figure 1 show the fact that the GLM (multiple Poisson regression) only includes a linear effect in SO_2 for all

Table 1. *Results of the independent variable effects analysis using GLM*

Model	Parameter	Estimate	Standard error	P
Number of admissions	Intercept	0.1507	0.1805	0.4039
	SO_2	0.0015	0.0014	0.2809
	PM10	0.0045	0.0014	0.0018
Number of patients with cough	Intercept	−0.1112	0.2058	0.5889
	SO_2	0.0016	0.0016	0.3149
	PM10	0.0044	0.0015	0.0043
Number of patients with exertional dyspnea	Intercept	−0.1112	0.2058	0.5889
	SO2	0.0016	0.0016	0.3149
	PM10	0.0044	0.0015	0.0043
Number of patients with expectoration	Intercept	−0,2202	0.2066	0.2864
	SO_2	0.0013	0.0016	0.4243
	PM10	0.0055	0.0015	0.0003

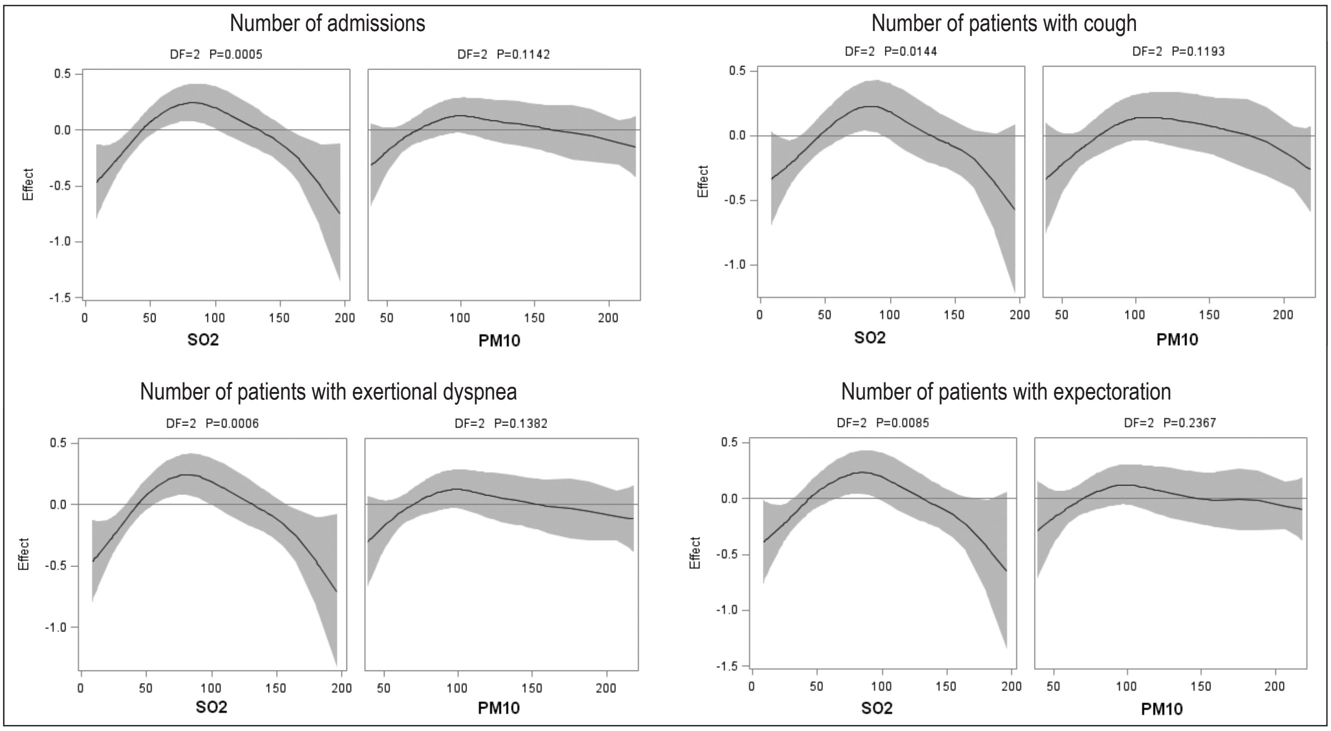

Fig. 1. *The plots of predictions against predictor for all models.*

models whereas the GAM model allows a more complex relationship, which the plots indicate, is quadratic. Having used the GAM procedure to discover an appropriate form of the dependence of hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration on each of two independent variables, we use semiparametric model including linear SO_2, PM10 and quadratic $g(SO_2) = SO_2*SO_2$ and PM10 terms. Results of the analysis of parameter estimates were given in Table 2.

For Bayesian approach, we investigated model efficacy in Bayesian semiparametric regression using MCMC and WINBUGS to generate chains of length 5000 after a burn-in of 5000, resulting in posterior samples of size 1000. Also we used diffuse priors β0, β1, β2 independent N (0,108). For model comparison, the Akaike information criterion (AIC), the Bayesian information criterion (BIC) for GLM and GAM and Deviance Information Criteria (DIC) for GAM with Bayesian approach were calculated for all models. The results are shown in Table 3.

Table 2. *Results of the analysis of parameter estimates using GAM*

Model	Parameter	Estimate	Standard error	P
Number of admissions	Intercept	−0.3711	0.2498	0.1374
	SO_2	0.0199	0.0058	0.0006
	$g(SO_2)$	−0.0001	0.0000	0.0012
	PM10	0.0038	0.0013	0.0047
Number of patients with cough	Intercept	−0.4932	0.2773	0.0753
	SO_2	0.0152	0.0065	0.0184
	$g(SO_2)$	−0.0001	0.0000	0.0300
	PM10	0.0039	0.0015	0.0114
Number of patients with exertional dyspnea	Intercept	−0.3598	0.2509	0.1515
	SO_2	0.0195	0.0058	0.0008
	$g(SO_2)$	−0.0001	0.0000	0.0016
	PM10	0.0037	0.0013	0.0063
Number of patients with expectoration	Intercept	−0.6722	0.2817	0.0170
	SO_2	0.0175	0.0066	0.0081
	$g(SO_2)$	−0.0001	0.0000	0.0119
	PM10	0.0049	0.0015	0.0014

AIC and BIC are criterions for model selection among a finite set of models. DIC is a hierarchical modelling generalization of AIC and BIC. It is particularly useful in Bayesian model selection problems where the posterior distributions of models have been obtained by MCMC simulation. The idea in comparison is that models with smaller AIC, BIC and DIC should be preferred to models with larger ones.

In order to compare the predictive accuracy for all models, an Artificial Neural Network approach was also considered. The main concept is to use pollution data as input parameters to predict the weekly hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration (output parameter). In our study a feed forward multi layer network architecture (ANN1) and a radial basis function approach (ANN2), which are widely used in ANN applications, were employed. For training and testing the proposed ANN model of the overall data set was randomly divided into two separate sets. The first set, namely the training set, consisted of 75% data records while the remaining 25% of the data records formed the test set. For comparison, the overall data set was divided the same way with ANN approach for GLM (Multiple Poisson Regression), GAM and Bayesian GAM. Table 4 provides RMSE and MAPE values for all models with five different approaches.

CONCLUSIONS

In this study, we used a standard Poisson regression and found association between SO_2 and hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration and no association between PM10 and hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration. Then we used the generalized additive model (GAM) of Poisson regression with a cubic spline and realised that the partial predictions corresponding to PM10 have a linear pattern and to SO_2 have not a quadratic pattern at 0.05 significance level for all models used. Quadratic terms for SO_2 for all models have negative effects on the number of admissions and the other dependent variables. We can interpret that as the level of SO_2 increases until a certain level, the number of admissions and the other dependent variables (number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration) will increase. After that certain level, the number of admissions will be stable.

We selected the new independent variable structure using GAM with a cubic spline. An important difference between the first analysis of this data with a standard Poisson regression (GLM) and the subsequent analysis with GAM is that GAM indicates that SO_2 is a significant predictor of the weekly hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration. The difference is due to the fact that the standard Poisson regression model only includes a linear effect in SO_2 whereas the GAM model allows a more complex relationship, which the plots indicate, is nearly quadratic. Having used the GAM procedure to discover an appropriate form of the dependence of hospital admissions, number of patients with cough, number of patients with exertional dyspnea, and number of patients with expectoration on each of two independent variables, you can use the standard Poisson regression to fit and assess the corresponding parametric model. So we see that GAM is very useful in visualizing the data and detecting the nonlinearity among the variables.

We also used a Bayesian approach for all models used. AIC, BIC for GLM and GAM and Deviance Information Criteria (DIC) for GAM with Bayesian approach were calculated for all models. The results in Table 3 show that Bayesian GAM approach gives smaller AIC and BIC values than GLM and GAM.

We used two Artificial Neural Network approaches (ANN1 and ANN2) for comparing the predictive accuracy for all models. Although multi-layer network architecture (ANN1) and a radial basis function approach (ANN2) gives similar RMSE and MAPE values, Bayesian approach gives much smaller RMSE and MAPE values than ANN approaches, GLM and GAM.

Table 3. *Results for model comparison*

Model	GLM		GAM		Bayesian GAM		
	AIC	BIC	AIC	BIC	AIC	BIC	DIC
Number of admissions	395.34	397.59	381.38	391.84	351.23	353.21	329.87
Number of patients with cough	374.88	377.16	341.32	351.78	331.12	333.18	327.36
Number of patients with exertional dyspnea	391.43	395.93	380.84	390.88	350.13	351.18	328.36
Number of patients with expectoration	364.14	367.12	334.62	345.08	298.12	299.18	278.36

Table 4. *Results for the predictive performance for all models*

Model	GLM		GAM		Bayesian GAM		ANN1		ANN2	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Number of admissions	2.67	4.98	2.42	4.68	2.12	3.99	2.43	4.71	2.38	4.54
Number of patients with cough	1.68	3.85	1.61	3.72	1.44	3.35	1.58	3.68	1.58	3.59
Number of patients with exertional dyspnea	2.54	4.83	2.33	4.71	2.02	3.55	2.35	4.64	2.25	4.41
Number of patients with expectoration	1.50	3.34	1.38	3.21	1.31	2.88	1.41	3.19	1.36	3.01

REFERENCES

1. Anderson HR, Spix C, Medina S, Schouten JP, Castellsague J, Rossi G, et al. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project. Eur Respir J. 1997 May;10(5):1064-71.
2. Burnett RT, Dales R, Krewski D, Vincent R, Dann T, Brook JR. Associations between ambient particulate sulfate and admissions to Ontario hospitals for cardiac and respiratory diseases. Am J Epidemiol. 1995 Jul 1;142(1):15-22.
3. Chen L, Yang W, Jennison BL, Omaye ST. Air particulate pollution and hospital admissions for chronic obstructive pulmonary disease in Reno, Nevada. Inhal Toxicol. 2000 Apr;12(4):281-98.
4. Chen Y, Yang Q, Krewski D, Shi Y, Burnett RT, McGrail K. Influence of relatively low level of particulate ar pollution on hospitalization for COPD in elderly people. Inhal Toxicol. 2004 Jan;16(1):21-5.
5. Moolgavkar SH, Luebeck EG, Anderson EL. Air pollution and hospital admissions for respiratory causes in Minneapolis - St. Paul and Birmingham. Epidemiology. 1997 Jul;8(4):364-70.
6. Morgan G, Corbett S, Wlodarczyk J. Air pollution and hospital admissions in Sydney, Australia, 1990 to 1994. Am J Public Health. 1998 Dec;88(12):1761-6.
7. Schwartz J, Marcus A. Mortality and air pollution in London: a time series analysis. Am J Epidemiol. 1990 Jan;131(1):185-94.
8. Schwartz J. Air pollution and daily mortality: a review and meta analysis. Environ Res. 1994 Jan;64(1):36-52.
9. Schwartz J. What are people dying of on high air pollution days? Environ Res. 1994 Jan;64(1):26-35.
10. Schwartz J. Air pollution and hospital admissions for the elderly in Birmingham, Alabama. Am J Epidemiol. 1994 Mar 15;139(6):589-98.
11. Schwartz J. Air pollution and hospital admissions for the elderly in Detroit, Michigan. Am J Respir Crit Care Med. 1994 Sep;150(3):648-55.
12. Schwartz J. PM10, ozone, and hospital admissions for the elderly in Minneapolis - St. Paul, Minnesota. Arch Environ Health. 1994 Sep-Oct;49(5):366-74.
13. Schwartz J. Air pollution and hospital admissions for respiratory disease. Epidemiology. 1996 Jan;7(1):20-8.
14. Wordley J, Walters S, Ayres JG. Short term variations in hospital admissions and mortality and particulate air pollution. Occup Environ Med. 1997 Feb;54(2):108-16.
15. Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA. 2006 Mar 8;295(10):1127-34.
16. Yang CY, Chen CJ. Air pollution and hospital admissions for chronic obstructive pulmonary disease in a subtropical city: Taipei, Taiwan. J Toxicol Environ Health A. 2007 Jul;70(14):1214-9.
17. Terzi Y, Cengiz MA. Using of generalized additive model for model selection in multiple poisson regression for air pollution data. Sci Res Essays. 2009;4(9):867-71.
18. Milutinović S, Nikić D, Stosić L, Stanković A, Bogdanović D. Short-term association between air pollution and emergency room admissions for chronic obstructive pulmonary disease in Nis, Serbia. Cent Eur J Public Health. 2009 Mar;17(1):8-13.
19. Hastie TJ, Tibshirani RJ. Generalized additive models. New York: Chapman and Hall; 1990.
20. Hastie T, Tibshirani R. Bayesian back fitting. Stat Sci. 2000;15(3):196-223.
21. Fan J, Gijbels I. Local polynomial modelling and its applications. London: Chapman and Hall; 1996.
22. Friedman JH, Silverman BW. Flexible parsimonious smoothing and additive modeling. Technometrics. 1989;31(1):3-21.
23. Friedman JH. Multivariate adaptive regression splines. Ann Stat. 1991;19(1):1-67.
24. Stone CJ, Hansen MH, Kooperberg C, Truong YK. Polynomial splines and their tensor products in extended linear modeling: 1994 Wald memorial lecture. Ann Stat. 1997;25(4):1371-470.
25. Eilers PHC, Marx BD. Flexible smoothing B-splines and penalties. Stat Sci. 1996;11(2):89-121.
26. Marx BD, Eilers PHC. Direct generalized additive modeling with penalized likelihood. Comput Stat Data Anal. 1998;28(2):193-209.
27. Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. J R Stat Soc Series B Stat Methodol. 1999;61(2):381-400.
28. Currie ID, Durban M. Flexible smoothing with P-splines: a unified approach. Stat Modelling. 2002;2(4):333-49.
29. Wand MP. Smoothing and mixed models. Comput Stat. 2003;18(2):223-49.
30. Ruppert D, Wand MP, Carroll RJ. Semiparametric regression. Cambridge: Cambridge University Press; 2003.
31. Berry SM, Carroll RJ, Ruppert D. Bayesian smoothing and regression splines for measurement error problems. J Am Stat Assoc. 2002;97(457):160-9.
32. Smith M, Kohn R. Nonparametric regression using Bayesian variable selection. J Econom. 1996;75(2):317-43.
33. Denison DGT, Mallick BK, Smith AFM. Automatic Bayesian curve fitting. J R Stat Soc Series B Stat Methodol. 1998;60(2):333-50.
34. Biller C. Adaptive Bayesian regression splines in semiparametric generalized linear models. J Comput Graph Stat. 2000;9(1):122-40.
35. Dimatteo I, Genovese CR, Kass RE. Bayesian curve-fitting with free-knot splines. Biometrika. 2001;88(4):1055-71.
36. Biller C, Fahrmeir L. Bayesian varying-coefficient models using adaptive regression splines. Stat Modelling. 2001;1(3):195-211.
37. Hansen MH, Kooperberg C. Spline adaptation in extended linear models. Stat Sci. 2002;17(1):2-51.
38. Fahrmeir L, Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. J R Stat Soc Ser C Appl Stat. 2001;50(2):201-20.
39. Fahrmeir L, Lang S. Bayesian semiparametric regression analysis of multicategorical time-space data. Ann Inst Stat Math. 2001;53(1):11-30.
40. Crainiceanu CM, Ruppert D, Wand MP. Bayesian analysis for penalized spline regression using WinBUGS. J Stat Softw. 2005;14(14):1-24.
41. Brezger A, Lang S. Generalized structured additive regression based on Bayesian P-splines. Comput Stat Data Anal. 2006;50(4):967-91.
42. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS version 1.4 user manual. Cambridge: Medical Research Council Biostatistics Unit; 2003.
43. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. London: Chapman and Hall; 1989.

Received January 2, 2012
Accepted in revised form August 28, 2012