Estimating the average length of hospitalization due to pneumonia: a fuzzy approach

L.F.C. Nascimento[1], P.M.S.R. Rizol[2] and A.P. Peneluppi[1]

[1]Departamento de Medicina, Universidade de Taubaté, Taubaté, SP, Brasil
[2]Departamento de Engenharia Elétrica, Universidade Estadual Paulista, Guaratinguetá, SP, Brasil

Abstract

Exposure to air pollutants is associated with hospitalizations due to pneumonia in children. We hypothesized the length of hospitalization due to pneumonia may be dependent on air pollutant concentrations. Therefore, we built a computational model using fuzzy logic tools to predict the mean time of hospitalization due to pneumonia in children living in São José dos Campos, SP, Brazil. The model was built with four inputs related to pollutant concentrations and effective temperature, and the output was related to the mean length of hospitalization. Each input had two membership functions and the output had four membership functions, generating 16 rules. The model was validated against real data, and a receiver operating characteristic (ROC) curve was constructed to evaluate model performance. The values predicted by the model were significantly correlated with real data. Sulfur dioxide and particulate matter significantly predicted the mean length of hospitalization in lags 0, 1, and 2. This model can contribute to the care provided to children with pneumonia.

Key words: Air pollutants; Fuzzy logic; Pneumonia; Particulate matter; Sulfur dioxide

Introduction

Pneumonia is a multifactorial disease that is influenced by, among other factors, paternal education, maternal age, number of people in the home, birth weight, z score, and breastfeeding (1,2). In 2009, there were 60,000 hospitalizations due to pneumonia in children under 4 years of age in the State of São Paulo, generating an expense of R$47 million, and there were 533 admissions in São José dos Campos that cost over R$400,000 (3). In that same year, 2700 children died due to pneumonia, with a higher prevalence in the north (8.3%) of Brazil, followed by the northeast (5.7%) region of the country, and these two regions had the lowest declines in mortality due to pneumonia (4,5).

Exposure to air pollutants is associated with hospitalization for pneumonia in children, and particulate material is the most common pollutant associated with this outcome (6-8). In general, previous studies aimed to quantify the risk of hospitalization due to exposure to air pollutants with additive models using Poisson regression.

However, we must consider the length of hospitalization due to pneumonia in children, since, apart from the financial cost, hospitalization implies a social cost because it separates the child from his or her environment. In 2009,

hospitalizations due to pneumonia in the State of São Paulo accounted for more than 300,000 hospital days, distributed among more than 660 municipalities in the state, and for 4000 days in São José dos Campos.

Thus, a model to predict the length of stay based on pollutant concentrations can reinforce the importance of public policies oriented to more effective air quality control and can consequently decrease the length of stay or even prevent its occurrence.

The objectives of this study were to construct a model using fuzzy logic tools to predict the duration of hospitalization for pneumonia in children aged less than 1 year, and also to validate the model.

Material and Methods

A predictive model was constructed using fuzzy logic tools. The fuzzy logic approach differs from commonly used logic that is dichotomous, i.e., Yes or No, Present or Absent. In Boolean logic, one element belongs to one or another set, and its membership is self-excluding, whereas in fuzzy logic one element can belong to two sets simultaneously, but with different values of membership.

Particulate matter (PM$_{10}$), sulfur dioxide (SO$_2$), ozone (O$_3$), and effective temperature (ET) were the input variables, and the average time (days) of hospitalization due to pneumonia in children up to 1 year of age, was the output variable. Data on length of stay, referring to the city of São José dos Campos, SP, Brazil, were obtained from Datasus (Brazil) (9).

São José dos Campos is located in southeastern Brazil between the cities of São Paulo and Rio de Janeiro (23°11′ South and 45°53′ West) and has a population of approximately 700,000 inhabitants. The city is an industrial, commerce, and service center that serves towns in the Paraíba Valley and the State of Minas Gerais, which has a total population of approximately two million inhabitants. The Dutra highway, considered the most important highway in Brazil, runs through the city and is used by approximately 130,000 vehicles per day, including a large volume of heavy vehicles and buses.

Records of concentrations (in μg/m^3) of PM$_{10}$, SO$_2$, and O$_3$, as well as relative humidity (RH, %) and minimum temperature (°C) were obtained from the Environmental Agency of São Paulo State (CETESB). These records are related to the period from January 1, 2009, to December 31, 2009. We calculated the ET, which considers temperature and RH, according to the following equation

$$ET = T - 0.4 \times (T - 10) \times \left(1 - \frac{RH}{100}\right)$$

Each input variable was constructed with two membership functions (Figure 1), and the output variable was built with four membership functions (Figure 2).
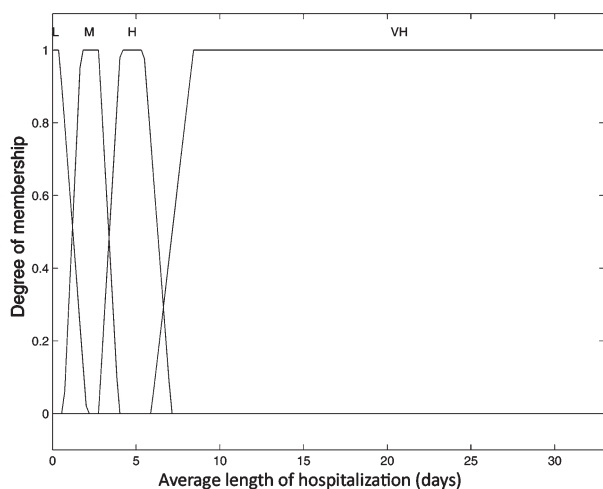
Membership functions for pollutants had a trapezoidal shape, and the straight lines representing low (L) and high (H) values intersected at the midpoint of the actual values of the variables. For each pollutant, the x-axis represents concentration (in μg/m^3) or ET (°C) and the y-axis represents membership functions between zero and one.

Membership functions (two) for each input variable (four) allowed the construction of 16 rules (2^4 rules); the length of stay was determined by the fuzzy inference system proposed by Mamdani (10), applying the set of the output model to the centroid defuzzification technique. This inference method has output built by the consequent overlap of individual rules. In this study, an example of a rule would be: if PM$_{10}$ is L and SO$_2$ is L and O$_3$ is L and ET is H, then average length of hospitalization is L. Another example would be: if PM$_{10}$ is H and SO$_2$ is H and O$_3$ is H and ET is L, then average length of hospitalization is H. If is considered to be related to the antecedent, and then is the consequent.

The fuzzy toolbox of Matlab software (MathWorks, USA) was used; this software allows importation of records from real input variables and provides output data given by the fuzzy model. The records provided by the model and real records were compared, yielding a Pearson correlation
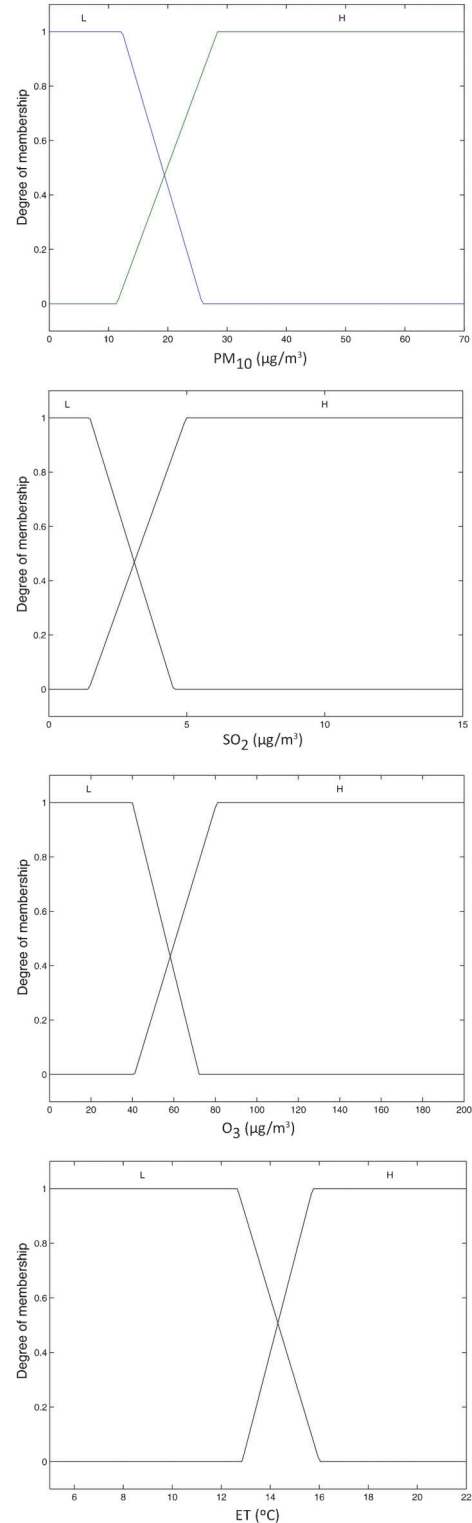
Figure 1. Membership function to particulate matter (PM$_{10}$), sulfur dioxide (SO$_2$), ozone (O$_3$), and effective temperature (ET), São José dos Campos, SP, Brazil, 2009.

Figure 2. Membership function to output (average length of hospitalization), São José dos Campos, SP, Brazil, 2009.

coefficient and a corresponding significance level.

Because the effects of exposure to air pollutants can manifest later, a lag of 0-2 days before hospitalization was considered for analysis. A receiver operating characteristic (ROC) curve was used as the cutoff, by taking the length of stay of 4 days to evaluate the performance of the model.

Results

The mean values, minimum, maximum, and standard deviations of the study variables are shown in Table 1.

The values obtained from correlation between the variables are shown in Table 2. It is possible to identify significant correlations between length of stay in days and the concentrations of PM_{10} and SO_2 and also with the output provided by the fuzzy model, (i.e., duration of hospitalization). The significant correlations between the fuzzy output and atmosphere and climate can be explained by the fact that the model contains these variables as inputs.

The values of the ROC curves and their confidence intervals are shown in Table 3. Particulate matter was the

Table 1. Mean $\pm$ SD values, minimum, and maximum (Min-Max) of the study variables, São José dos Campos, SP, Brazil, 2009.

	Mean $\pm$ SD	Min-Max
Particulate matter ($\mu g/m^3$)	21.6 $\pm$ 8.8	7-57
Sulfur dioxide ($\mu g/m^3$)	2.8 $\pm$ 1.6	1-9
Ozone ($\mu g/m^3$)	82.3 $\pm$ 31.1	16-196
Temperature (°C)	15.9 $\pm$ 2.8	7.6-21.0
Real time	4.2 $\pm$ 4.1	0-30
Fuzzy time	8.3 $\pm$ 6.1	0-22

Table 2. Pearson's correlation coefficient values for the atmospheric variables, average length of hospitalization (real time) and time provided by the model (fuzzy time), São José dos Campos, SP, Brazil, 2009.

	Real time	Fuzzy time	PM_{10}	SO_2	O_3	ET
Real time	1					
Fuzzy time	0.12*	1				
PM_{10}	0.17**	0.58**	1			
SO_2	0.15*	0.63**	0.50**	1		
O_3	−0.03	0.22**	0.34**	0.26**	1	
ET	−0.17**	−0.65**	-0.32**	−0.42**	0.14*	1

*$P<0.05$; **$P<0.01$.

best performing pollutant (0.60 area under the curve when the real records were analyzed); the results for ozone were not significant. Particulate material was able to predict the duration of hospitalization for pneumonia of 4 days or more, with an area under the curve of 0.80 with respect to fuzzy modeling. The inflection of the curve corresponds to a sensitivity of 87% and specificity of 53%. These values correspond to the ROC curve concentration of 15 $\mu g/m^3$ of PM_{10}, but sulfur dioxide exhibited the best results with an area under the curve of 0.86, and the inflection of the curve corresponds to a sensitivity of 94% and specificity of 60% corresponding to concentrations on the order of 2 $\mu g/m^3$.

Discussion

This study shows the effect of exposure to air pollutants on hospitalization duration using fuzzy logic tools. The model performs well for predicting time of admission of 4 or more days, a situation that suggests a more severe pneumonia.

Studies on the effects of exposure to air pollutants on human health have used the additive model of Poisson regression because events count as variables. This

Table 3. Areas under the receiver operating characteristic (ROC) curves for air pollutants and real and fuzzy values, São José dos Campos, SP, Brazil, 2009.

	PM_{10}	SO_2	O_3
Fuzzy			
Lag 0	0.80 (0.74-0.86)	0.86 (0.81-0.91)	0.80 (0.73-0.87)
Lag 1	0.65 (0.58-0.72)	0.65 (0.58-0.72)	0.57 (0.49-0.66)*
Lag 2	0.54 (0.47-0.61)*	0.64 (0.57-0.72)	0.51 (0.42-0.59)*
Real			
Lag 0	0.60 (0.54-0.67)	0.60 (0.54-0.67)	0.45 (0.38-0.52)*
Lag 1	0.58 (0.51-0.64)	0.63 (0.56-0.69)	0.47 (0.40-0.54)*
Lag 2	0.57 (0.50-0.63)	0.64 (0.58-0.71)	0.50 (0.43-0.57)*

*$P>0.05$.

approach allows estimation of the risk of hospitalization due to exposure. Moreover, hospitalization duration is usually not considered in the models, but this variable is very important because it represents a significant social cost to the child and his or her family.

The fuzzy model presented here showed good sensitivity, demonstrated by the ROC curve for PM_{10} and SO_2. A similar fuzzy approach was adopted in studies on diabetic neuropathy (11), neonatal mortality (12), and evaluation of myocardial perfusion (13). Fuzzy logic tools have been applied to classify the severity of pulmonary fibrosis (14) and pneumonia in children (15) and in pediatric emergency diagnoses (16). No Brazilian studies on the effects of exposure to air pollutants and hospital respiratory disease were found; therefore, a comparison of the performance of the model presented in this study is not possible. Chaves (17) recently developed a model to estimate the role of exposure to pollutants in hospitalization for respiratory diseases, using fuzzy and neurofuzzy modeling, and reported better results with fuzzy modeling.

The average concentrations of PM_{10}, SO_2, and O_3 found in São José dos Campos are below those found in large cities and also have similar demographic and socioeconomic characteristics (18). The distribution of length of stay according to the concentrations of PM_{10}

and SO_2 exhibited a dose response in this population; so if higher concentrations of these pollutants occur, it is expected that hospitalizations may be longer in duration.

ROC curves showed good results for both PM_{10} and SO_2, considering that lags 0, 1, and 2 were better for SO_2 at lag 0; when compared to the actual data, the ROC curves showed slightly lower results.

The robustness of the model constructed in this study was validated with real data because fuzzy logic does not require a training phase as for neural networks. Other formats for the membership functions can be tried, such as triangular or bell curves, in an attempt to improve the model. Other advantages of this model are its application in any location where there are available data on the same air pollutants and the ease of obtaining the predicted results. Creation of software, which does not require a large investment, together with the use of a portable instrument may allow estimation of hospitalization duration and provide support to local health managers.

Our findings demonstrate that it is possible to construct and validate a predictive model using fuzzy logic with input values corresponding to pollutants.

Acknowledgments

References

1. Nascimento LFC, Marcitelli R, Agostinho FS, Gimenes CS. Hierarchical approach to determining risk factors for pneumonia in children. *J Bras Pneumol* 2013; 30: 445-451, doi: 10.1590/S1806-37132004000500008.

2. Boccolini CS, Carvalho ML, Oliveira MI, Boccolini PM. Breastfeeding can prevent hospitalization for pneumonia among children under 1 year old. *J Pediatr* 2011; 87: 399-404.

3. Brasil. Ministério da Saúde. DATASUS - Departamento de Informática do SUS. http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/nrsp.def. Accessed August 22, 2013.

4. Brasil. Ministério da Saúde. DATASUS - Departamento de Informática do SUS. http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10uf.def. Accessed August 22, 2013.

5. Rodrigues FE, Tatto RB, Vauchinski L, Leaes LM, Rodrigues MM, Rodrigues VB, et al. Pneumonia mortality in Brazilian children aged 4 years and younger. *J Pediatr* 2011; 87: 111-114.

6. Nascimento LF, Pereira LA, Braga AL, Modolo MC, Carvalho JA, Jr. [Effects of air pollution on children's health in a city in Southeastern Brazil]. *Rev Saúde Pública* 2006; 40: 77-82, doi: 10.1590/S0034-89102006000100013.

7. Cesar AC, Nascimento LF, de Carvalho JA Jr. [Association between exposure to particulate matter and hospital admissions for respiratory disease in children]. *Rev Saúde Pública* 2013; 47: 1209-1212, doi: 10.1590/S0034-8910.2013047004713.

8. Rosa AM, Ignotti E, Hacon SS, Castro HA. Analysis of hospitalizations for respiratory diseases in Tangara da Serra,

Brazil. *J Bras Pneumol* 2008; 34: 575-582, doi: 10.1590/S1806-37132008000800006.

9. Brasil. Ministério da Saúde. DATASUS. Informações de Saúde. Epidemiológicas e morbidade. http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/nrsp.def. Accessed April 24, 2014.

10. Tanaka K. *An Introduction to Fuzzy Logic for Practical Application*. Springer-Verlag; 1996.

11. Picon AP, Ortega NR, Watari R, Sartor C, Sacco IC. Classification of the severity of diabetic neuropathy: a new approach taking uncertainties into account using fuzzy logic. *Clinics* 2012; 67: 151-156, doi: 10.6061/clinics/2012(02)10.

12. Nascimento LF, Rocha Rizol PM, Abiuzi LB. Establishing the risk of neonatal mortality using a fuzzy predictive model. *Cad Saúde Pública* 2009; 25: 2043-2052, doi: 10.1590/S0102-311X2009000900018.

13. Duarte PS, Mastrocolla LE, Farsky PS, Sampaio CR, Tonelli PA, Barros LC, et al. Selection of patients for myocardial perfusion scintigraphy based on fuzzy sets theory applied to clinical-epidemiological data and treadmill test results. *Braz J Med Biol Res* 2006; 39: 9-18, doi: 10.1590/S0100-879X2006000100002.

14. Lopes AJ, Capone D, Mogami R, Lanzillotti RS, Melo PL, Jansen JM. Severity classification for idiopathic pulmonary fibrosis by using fuzzy logic. *Clinics* 2011; 66: 1015-1019, doi: 10.1590/S1807-59322011000600016.

15. Pereira JC, Tonelli PA, Barros LC, Ortega NR. Clinical signs of pneumonia in children: association with and prediction of

diagnosis by fuzzy sets theory. *Braz J Med Biol Res* 2004; 37: 701-709, doi: 10.1590/S0100-879X2004000500012.

16. Grigull L, Lechner WM. Supporting diagnostic decisions using hybrid and complementary data mining applications: a pilot study in the pediatric emergency department. *Pediatr Res* 2012; 71: 725-731, doi: 10.1038/pr.2012.34.

17. Chaves LE. Fuzzy and neuro-fuzzy computational models to evaluate air pollutants effects. [PhD thesis]: Universidade do Estado de São Paulo; 2013.

18. São Paulo. Environmental Agency of São Paulo State (CETESB). Report 2009. http://www.cetesb.sp.gov.br/ar/qualidade-do-ar/31-publicacoes-e-relatorios.