Use of an Artificial Neural Network to Predict Risk Factors of Nosocomial Infection in Lung Cancer Patients

Jie Chen², Qin-Shi Pan¹, Wan-Dong Hong³, Jingye Pan², Wen-Hui Zhang¹, Gang Xu¹, Yu-Min Wang¹*

Abstract

Statistical methods to analyze and predict the related risk factors of nosocomial infection in lung cancer patients are various, but the results are inconsistent. A total of 609 patients with lung cancer were enrolled to allow factor comparison using Student’s t-test or the Mann-Whitney test or the Chi-square test. Variables that were significantly related to the presence of nosocomial infection were selected as candidates for input into the final ANN model. The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the performance of the artificial neural network (ANN) model and logistic regression (LR) model. The prevalence of nosocomial infection from lung cancer in this entire study population was 20.1% (165/609), nosocomial infections occurring in sputum specimens (85.5%), followed by blood (6.73%), urine (6.0%) and pleural effusions (1.82%). It was shown that long term hospitalization (≥22days, \( P=0.000 \)), poor clinical stage (IIIb and IV stage, \( P=0.002 \)), older age (≥61year old, \( P=0.023 \)), and use the hormones were linked to nosocomial infection and the ANN model consisted of these four factors. The artificial neural network model with variables consisting of age, clinical stage, time of hospitalization, and use of hormones should be useful for predicting nosocomial infection in lung cancer cases.

Keywords: Artificial neural network (ANN) - predictors - lung cancer - nosocomial infection

Introduction

Lung cancer mortality is the highest in all tumors and its incidence is gradually increasing (Jemal et al., 2005; Malcolm et al., 2009; Elsayed et al., 2011). Despite surgical resection, chemotherapy and radiation therapy technology are continuously improving, patients with lung cancer remains extremely vulnerable to relapse and fatal (Gridelli et al., 2003). The cure rate of lung cancer is very low and the average 5-year survival of patients with lung cancer is below 15 % (Ogawa et al., 2008; Rachet et al., 2008; Chen et al., 2009; Stewart et al., 2010). The nosocomial infection rate of patients with lung cancer showed high trend (Kamboj et al., 2009). Nosocomial infection not only affected the treatment and rehabilitation, prolonged time of hospitalization, increased health care costs, but also significantly resulted in prognosis, even life-threatening (Bereket et al., 2012). To analyze the characteristics and risk factors of nosocomial infections in lung cancer patients, which will help to we adopt effective prevention and control of nosocomial infection for improving patient outcomes and prolonging survival in lung cancer patients.

As one of the clinical prediction rules (Simon et al., 2012), an artificial neural network (ANN) is composed of a series of interconnecting parallel nonlinear processing elements (nodes) with limited numbers of inputs and outputs (Hong et al., 2011). A systematic review suggested that ANN is potentially more successful than conventional statistical techniques at predicting clinical outcomes when the relationship between the variables that determine the prognosis is complex, multidimensional and non-linear (Bartosch-Harlid et al., 2008). The aim of this study was to develop an ANN to predict nosocomial infection in lung cancer.

Materials and Methods

Subjects

The 609 patients with lung cancer came from the First Affiliated Hospital of Wenzhou Medical University, China, from January April 2005 to January 2014. The above cases were confirmed by the histopathological results. The lung cancers consisted of 443 male and 166 women, aged between 32 to 88 years. The demographics characteristics of 609 cases lung cancer patients see Table 1. The criterion for the histopathologic diagnosis of lung cancer was the World Health Organization (WHO)/International Study of Lung Cancer (IASLC) lung cancer histological classification standards. The following
information was collected for each patient on admission: age, gender, clinical stage, histological classification, invasive procedures, mechanical ventilation, surgery, radiotherapy, chemotherapy, hemoglobin, serum albumin, white blood cell count, use of antibiotics, use of hormone, non-neoplastic lung disease, concurrent diabetes or renal insufficiency, smoking (smoking index= number of cigarettes smoked per day × smoking years), time of hospitalization. This study was approved by the Institutional Ethics Review Board of the First Affiliated Hospital of Wenzhou Medical University and all patients provided written informed consent to this study.

The culture and identification of bacterial and fungi
Specimens include 300 cases sputum, 121 cases blood, 121 urine sample and 67 cases pleural effusion. The bacterial are identified by BioMerieux automatic identification system (France, Meraux Corporation) and susceptibility testing is used by K-B method. The culture and identification of fungi were used with Kemaijia chromogenic medium (France, Meraux Corporation).

Diagnostic criteria of nosocomial infection
The nosocomial infection diagnostic criteria and incidence statistics with reference to the Chinese Ministry of Health nosocomial infection diagnostic criteria in 2001. Confirmed information according to clinical manifestations, laboratory tests and /or identification of bacteria or fungi, where the infection occurred at 48h after admission were classified as nosocomial infections; infections occur if the patient admission directly related last hospitalization, tied for nosocomial infection stained.

Statistical analysis
Continuous values were expressed as the mean±SD or medians and compared using Student’s t-test or the Mann-Whitney non-parametric test. Categorical values were described by counts and proportions and compared using the $X^2$ test. Variables that were significantly related to the presence of nosocomial infection were selected as candidates for input into the final ANN model. Sensitivity analysis (also known as independent variable importance analysis) was performed to determine the optimum variables for construction of the final ANN model (Hong et al., 2011). An exploratory three-layer multiplayer perceptron (MLP) ANN model with a back propagation algorithm was constructed for sensitivity analysis. The data were randomly divided into a training sample (487 cases, 80%) and a test sample (122 cases, 20%) in the exploratory ANN model. Sigmoid transfer functions were used in the hidden and output layers. Gradient descent was used to estimate the synaptic weights. The initial learning...
The 198 pathogens were isolated from 165 cases of patients with lung cancer. According to the results of the univariate and sensitivity analyses, a final three-layer feed-forward ANN model with a back propagation algorithm was constructed for all 609 patients. The ANN model was trained with a maximum of 500 iterations and 10 tours. The overfit penalty was assigned as 0.001, and the convergence criterion was 0.00001 (Hong et al., 2011). Fivefold cross-validation was used (Sall et al., 2007). The output of the ANN model was transformed to range from 0-1. Nosocomial infection was predicted if the output was greater than or equal to 0.5 (Hong et al., 2011). The sensitivity, specificity, negative predictive value, positive predictive value and diagnostic accuracy of the ANN model are reported herein. The clinical or patient-relevant utility of a diagnostic test is evaluated by a Fagan plot. The Fagan plot allows the reader to estimate the post-test probability of the target condition in an individual patient based on a selected pretest probability (Whiting et al., 2008). Forward conditional step-wise logistic regression analysis was performed to develop a logistic regression function (LR) for comparison. The conditional probabilities for stepwise entry and removal of a factor were 0.05 and 0.06, respectively. The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the performance of the ANN model and LR model. Differences were considered statistically significant if the two-tailed P value was less than 0.05.

SPSS 18.0 (SPSS Inc., Chicago, IL, USA) was used for ANN analysis.

Results

The distribution of nosocomial infection in patients with lung cancer

The prevalence of nosocomial infection from lung cancer in this entire study population was 20.09% (165/790), nosocomial infections occur in sputum specimens (85.45%), followed by blood (6.73%), urine (165/609), nosocomial infections in lung cancer as follows: -0.34+0.34 age (years) -1.01 clinical stage + 1.23 time of hospitalization (days).

The pathogen distribution of nosocomial infections in patients with lung cancer

The 198 pathogens were isolated from 165 cases of nosocomial infection in patients with lung cancer, including 41 cases Gram-negative bacteria, accounting for 20.71%; 71 cases Gram-positive bacteria, accounting for 35.86%; 86 cases fungi, accounting for 43.43%.

Chi-Square and T-test/Mann-Whitney test results for infection and non-infection from lung cancer patients

It was shown that clinical stage ($X^2=17.805$, $p=0.001$), surgery ($X^2=5.116$, $p=0.024$), mechanical ventilation ($X^2=11.766$, $p=0.001$), use of antibiotics ($X^2=6.241$, $p=0.012$), use of hormone ($X^2=12.256$, $p=0.000$), age ($t=-2.219$, $p=0.027$), hemoglobin ($t=2.007$, $p=0.045$), serum albumin ($t=2.994$, $p=0.003$), and time of hospitalization (Mann Whitney test =14465.000, $p=0.000$) were statistically different between infection and non-infection from lung cancer patients.

Univariate and multivariate analysis

Eighteen variables considered relevant to the presence of nosocomial infections were tested using univariate and multivariate analyses. Multivariate analysis by logistic regression identified the following three independent variables as predictive of persistent nosocomial infections in lung cancer: age ($p=0.028$), clinical stage ($p=0.006$), and time of hospitalization ($p=0.000$). A logistic regression function (LR model) was developed to predict nosocomial infections in lung cancer as follows: -0.34+0.34 age (years) -1.01 clinical stage + 1.23 time of hospitalization (days).

Use of an Artificial Neural Network to Predict Risk Factors of Nosocomial Infection in Lung Cancer Patients

As shown in Figure 1, time of hospitalization, clinical stage, age and use of hormone were the most important predictors of nosocomial infections by sensitivity analysis (the exploratory ANN model constructed for the sensitivity analysis is not shown). The final three-layer 5-5-1-feed-forward back propagation ANN model with variables consisting of time of hospitalization, clinical stage, age and use of hormone was developed and trained in 609 patients (Figure 2). The sensitivity, specificity, positive likelihood ratio, negative likelihood ratio of the ANN was 56.0%, 85.0%, 3.73, and 0.52, respectively. The ROC curves for the ANN model and LR model for predicting nosocomial infection from lung cancer patients.

Table 3. T-test/Mann-Whitney Test Results for Infection and Non-infection from Lung Cancer Patients

| Variable name         | Cases           | T-test/Mann-Whitney test | P   |
|-----------------------|-----------------|-------------------------|-----|
| age (years)           | 66.75±11.18     | 63.96±11.06             | -2.219 | 0.027 |
| hemoglobin (g/L)      | 106.96±21.69    | 111.62±20.05            | 2.007  | 0.045 |
| serum albumin (g/L)   | 33.07±5.31      | 34.84±5.17              | 2.994  | 0.003 |
| white blood count (+10%) | 10.19±6.44 | 9.47±5.54               | -1.107 | 0.269 |
| time of hospitalization (days) | 27.00 (15.25-43.00) | 19.00 (12.00-32.00)   | 14465.000* | 0.000 |
| Smoking index         | 600.00 (15.00-1000.00) | 400.00 (0.00-815.00) | 16643.500* | 0.602 |

*Mann-Whitney test

Table 4. The Sensitivity, Specificity, and AUC of ANN Model and LR Model

|                | Sensitivity (%) | Specificity (%) | AUC     | LR+    | LR-    | PPV    | NPV    |
|----------------|-----------------|-----------------|---------|--------|--------|--------|--------|
| ANN model      | 56.0            | 85.0            | 0.860±0.034 | 3.73   | 0.52   | 75.7   | 69.9   |
| LR model       | 38.0            | 86.7            | 0.759±0.042 | 2.85   | 0.72   | 70.4   | 62.7   |

LR+, positive likelihood ratio, LR-, negative likelihood ratio, PPV, positive predictive value, NPV, negative predictive value

DOI:http://dx.doi.org/10.7314/APJCP.2014.15.13.5349

Use of an Artificial Neural Network to Predict Risk Factors of Nosocomial Infection in Lung Cancer Patients
infections in lung cancer patients are shown in Figure 3. The AUC of the ANN model (AUC=0.860±0.034) was statistically higher than the AUC of the LR model (AUC=0.759±0.042). We further analyze the time of hospitalization, clinical stage, age and found long time of hospitalization (≥22days, p= 0.000), poor clinical stage (IIIb and IV stage, p=0.002), older (≥61year old , p=0.023) were apt to nosocomial infection.

**Discussion**

The results of this study demonstrate that time of hospitalization, clinical stage, age and use of hormone were the most important predictors of nosocomial infections. Based on ROC analysis, the diagnostic performance of the ANN model was superior to both the LR model.

Of the standard ANNs, the MLP is perhaps the most popular network architecture currently in use (Saftoiu et al., 2012). An MLP model consists of an input layer, a hidden layer and an output layer. All of the artificial neurons are arranged in a layered feed–forward topology. Our ANN model was developed using the SPSS neural networks program and JMP software, which can both run the MLP model (Sall et al., 2007; Hong et al., 2011). ANNs are nonlinear statistical data modeling tools. They can take into account outliers and nonlinear interactions among variables and can reveal previously unrecognized and/or weak relationships between given input variables and an outcome (Sall et al., 2007). Therefore, ANNs often include parameters that may not reach significance using conventional statistics, as evidenced by the fact that use of hormone included in our ANN model was not significant in logistic regression analysis.

We further analyze the time of hospitalization, clinical stage, age and found long time of hospitalization (≥22days, p= 0.000), poor clinical stage (IIIb and IV stage, p=0.002), older (≥61year old , p=0.023) were apt to nosocomial infection. Jiang Y et al. (2004) found that pulmonary fungal infection rate was 6.35% (78/1229). The major fungus was Candida albicans (68.18%). The main risk factors were age of > or =50 years (p<0.005), primary site (lung cancer, p<0.001), cancer stage (stage IV, p<0.005), pulmonary radiotherapy (p<0.001), chemotherapy (p<0.001), and long-term hospitalization (>2 weeks, p<0.005). Our results were mostly consistent to Jiang Y. Prolonged hospitalization increased the chances of opportunistic infections. Older were prone to infection for decreased immunity. Also The IIIb and IV stage of lung cancer had occurred distant metastasis or lymph node metastasis and their immune function is poor. In clinical practice, people often seek hormones as “antipyretics”, resulting in further development of disease or concurrent bacterial and fungal infection on the basis of viral infection.

In conclusion, an artificial neural network model with variables consisting of time of hospitalization, clinical stage, age and use of hormone may be useful for predicting the nosocomial infection in patients with lung cancer.

**Acknowledgements**

This study was financially supported by the Zhejiang Provincial Health Department (No. 2014KYA133).

**References**

Bartosch-Harlid A, Andersson B, Aho U, Nilsson J, Andersson R (2008). Artificial neural networks in pancreatic disease. *Br J Surg*, 95, 817-26.

Bereket W, Hemalatha K, Getenet B, et al (2012). Update on bacterial nosocomial infections. *Eur Rev Med Pharmacol Sci*, 16, 1039-44.

Chen CH, Lai JM, Chou TY , et al (2009). VEGFA upregulates FLJ10540 and modulates migration and invasion of lung cancer via PI3K/AKT pathway. *PloS One*, 4, 5052.

Gridelli C, Rossi A, Maione P (2003). Treatment of non-small-cell lung cancer: state of the art and development of new biologic agents. *Oncogene*, 22, 6629-38.

Hong WD, Ji YF, Wang D, Chen TZ, Zhu QH (2011). Use of artificial neural network to predict esophageal varices in
patients with HBV related cirrhosis. Hepat Mon, 1, 544-7.
Jemal A, Murray T, Ward E, et al (2005). Cancer statistics, 2005.
CA Cancer J Clin, 55, 10-30.
Jiang Y, Li JY, Li M, et al (2004). Clinical analysis of nosocomial pulmonary fungal infection in patients with cancer. Ai Zheng, 23, 1707-9 (in Chinese).
Kamboj M, Sepkowitz K A (2009). Nosocomial infections in patients with cancer. Lancet Oncol, 10, 589-97.
Moore MA, Ariyaratne Y, Badar F, et al (2009). Cancer epidemiology in South Asia - past, present and future. Asian Pac J Cancer Prev, 10, 49-67.
Ogawa E, Takenaka K, Katakura H, et al (2008). Perimembrane Aurora-A expression is a significant prognostic factor in correlation with proliferative activity in non-small-cell lung cancer (NSCLC). Ann Surg Oncol, 15, 547-54.
Rachet B, Woods LM, Mitry E, et al (2008). Cancer survival in England and Wales at the end of the 20th century. Br J Cancer, 99, 2-10.
Saftoiu A, Vilmann P, Gorunescu F, Janssen J, Hocke M, et al (2012). Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. Clin Gastroenterol Hepatol, 10, 84-90.
Salim E, Jazieh AR, Moore MA (2011). Lung cancer incidence in the arab league countries: risk factors and control. Asian Pac J Cancer Prev, 12, 17-34.
Sall J, Creighton L, Lehman A, Safari Tech Books Online (2007). JMP start statistics a guide to statistics and data analysis using JMP. SAS Press series, 4th ed. Cary, N.C., SAS Pub.
Simon TA, Stephen HL (2012). Clinical prediction rules. BMJ, 344, 8312.
Stewart DJ (2010). Tumor and host factors that may limit efficacy of chemotherapy in non-small cell and small cell lung cancer. Crit Rev Oncol Hematol, 75, 173-234.
Whiting PF, Sterne JA, Westwood ME, et al (2008). Graphical presentation of diagnostic information. BMC Med Res Methodol, 8, 20.