Journal of the Korean Data & Information Science Society
2017, **28**(4), 899–910

http://dx.doi.org/10.7465/jkdi.2017.28.4.899
한국데이터정보과학회지

Nomogram plot for predicting chronic otitis media in Korean adults

Eun Jin Kang[1] · Jea Young Lee[2]

[12]Department of Statistics, Yeungnam University

Received 25 May 2017, revised 29 June 2017, accepted 5 July 2017

Abstract

Nomogram is useful for predicting the prevalence of each patient through the scoring system without a complex formula. Because there are few studies on chronic otitis media (COM) in adults, this study aims to identify the relevant risk factors for COM in Korean adults and to build a nomogram for the risk factors. The Health Interview Survey data subset, derived from the Sixth Korean National Health and Nutrition Examination Survey (KNHANES VI), was used to evaluate the participants. Of the participants, the weighted prevalence of COM was 5.3%. Residence, earphone use, atopic dermatitis, allergic rhinitis, chronic rhinosinusitis, and subjective hearing status were identified as risk factors for COM. Using 6 risk factors, we propose a nomogram for COM, and use AUC to verify the discrimination of the nomogram.

Keywords: Chronic otitis media, Korean adults, nomogram, risk factors.

1. Introduction

Many people suffer from illness due to lack of health care, nowadays. Especially, they tend to be unaware of the importance of small abnormalities in the ear, which can lead to chronic otitis media (COM). COM is an inflammatory disease in the middle ear. Otitis media has two main types. Acute otitis media (AOM) mainly affects children under two years old, and results from bacteria or viruses (Qureishi *et al.*, 2014). Otitis media with effusion (OME) is a chronic inflammatory condition. It typically affects children between three and seven years old. It is characterized by the effusion of a glue-like fluid from behind an intact tympanic membrane, without the signs and symptoms of acute inflammation (van Zon *et al.*, 2013). If an inflammation persists in the middle ear, it can result in long-term or permanent damage, e.g., perforation, atelectasis, retraction, tympanosclerosis, or cholesteatoma in the tympanic membrane. The prevalence of otitis media has declined. In 1990, 4,900 patients died of it; however, it was investigated only half that number, 2,400, died in 2013 (Collaborators, 2015). The prevalence differs by region. In Southeast Asia, Africa, and the Western Pacific

[1] Graduate student, Department of Statistics, Yeungnam University, Kyungsan 38541, Korea. E-mail: yu_ejk@naver.com
[2] Corresponding author: Professor, Department of Statistics, Yeungnam University, Kyungsan 38541, Korea.

countries, the prevalence of otitis media is 2-4%. In North America and Europe, it is less than 2% (Park *et al.*, 2015).

Many studies have been conducted on the prevalence and risk factors of COM. The prevalence of otitis media is significantly higher in children who have not fully developed internal organs of the ear compared to other ages. The prevalence of acute otitis media is 62% to 1-year old and 83% to 3 years old at least 1 time (Teele *et al.*, 1989). Consequently, most of Studies on COM targeting children have been conducted (Engel *et al.*, 2005; Gozal *et al.*, 2008), but studies of risk factors for otitis media in adults are lacking. In Park *et al.*, 2015, it was studied using data from the fifth Korean National Health and Nutrition Examination. For people who aged over 20 years old, the prevalence of COM was 3.8%, and it was analyzed whether general characteristics, medical conditions, otologic conditions, blood test, hearing threshold, and questionnaire of quality of life are relevant with COM. As a result of logistic regression analysis, people who have pulmonary tuberculosis, chronic rhinosinusitis, hearing impairment, or tinnitus, more suffered from COM. On the other hand, hepatitis B showed low odds ratio, and the patients who didn't suffer from hepatitis B, more suffered from COM.

In our study, we identified risk factors for adults older than 40 years with COM and established a nomogram for the risk factors. We used data from the Sixth Korean National Health and Nutrition Examination Survey (KNHANES VI; 2013015). Especially, we divided the data into training set and validation set. Training data was used to build a model, and a nomogram. Also, it was significantly determined using AUC (Area under the ROC Curve) of training set and validation set.

2. Data source

The Health Interview Survey data subset, derived from the publicly available KNHANES VI, was used for the study. KNHANES VI was conducted by the Korea Centers for Disease Control and Prevention from 2013-2015. Participants were selected from a representative sample of the general Korean population, and their information included their health and nutritional status. In addition, COM was diagnosed by professional research team.

In total, 22,948 people participated in KNHANES VI (8,018 in 2013, 7,550 in 2014, and 7,380 in 2015). People who did not have results for ear, nose, and throat (ENT) examinations, were under 40 years of age, suffered from gastric, liver, colon, breast, cervical, lung, or thyroid cancer at the time of the survey, or were not diagnosed with COM, were excluded. Also, people who did not have answer about smoking status or alcohol consumption were not included. Finally, 6,487 subjects were eligible for inclusion in this study. The data set was divided into the training data set (60%) and the validation set (40%), randomly.

3. Methods

3.1. Data collection and management

The patients filled out a questionnaire about their general characteristics. Our study is aimed at people over 40 years of age; the age was categorized by decade (40-49, 50-59, 60-69, or ≥ 70) and we analyzed the patient's gender (male or female) for its relation with COM. The following information on the patients' socioeconomic status was investigated: education level (less than middle school or beyond high school), income level ($< 25\%$, 25-50%, 50-75%, or $\geq 75\%$ according to the equivalized household income per month), occupation (white-collar: manager, professional, clerk, service/sales worker, unemployed, retired, student, or housewife; blue-collar: agriculture, forestry, fishery worker, craft or related trade worker, plant or machine operator or assembler, or simple laborer), residence (urban or rural area in accordance with the patient's official address). The body-mass index was categorized as either < 25 or ≥ 25 kg/m^2. Also, earphone use in noisy situations and temporary exposure to noise were considered. The smoking status and alcohol consumption were surveyed. If a patient has smoked more than five packs of cigarettes in their life and the patient is a smoker, they were categorized as "yes". If they are former smokers or a nonsmoker, they were categorized as "no". If a patient had consumed alcohol one or more times per month during the past year, their alcohol consumption was categorized as "yes", otherwise "no". The number of household members (1-2, 3-4, or 5-6) and the subjective health status (very good, good, average, poor, or very poor) were also surveyed. The patient's medical histories were obtained, including hypertension, diabetes mellitus, hepatitis B, atopic dermatitis, allergic rhinitis, and chronic rhinosinusitis. For the otologic investigation, participants were asked about their hearing status ("not discomfort", "a little discomfort", "a lot of discomfort", or "cannot hear") and in this study, "a lot of discomfort" and "cannot hear" were analyzed in one category as "a lot of discomfort". Tinnitus was also included.

To build a nomogram and validate the nomogram, raw data was separated into training set (n=3,959) and validation set (n=726). Using the training set, the analysis was adjusted for the complex sample design of the survey using sample weights. We used the KNHANES sampling weight variables along with a masked-variance primary sampling unit and stratum variables. This adjustment allowed for extrapolation from the samples to the non-institutionalized civilian population of Korea as a whole. The survey sample weights were used in the analyses. We identified candidate predictors of COM using the chi-squared test for categorical variables (p-value<0.05) and performed a logistic regression analysis to identify risk factors associated with COM using backward method (p-value<0.1). Using the risk factors, we built a nomogram for COM. AUC is used to verify the discrimination of the nomogram for training set and validation set.

All data were analyzed using statistical package for the social sciences (SPSS) Version 23.0 (SPSS In., Chicago, IL, USA).

3.2. Logistic regression model

Regression analysis, that a dependent variable is continuous, is a statistical process for estimating how more than one explanatory variable affects the continuous response variable. However, in the cases that the dependent variable has two outcome categories, such as pass and fail, or survival and death, logistic regression analysis is usually used. In this study, we consider the only binary dependent variable that are commonly used. The binary dependent variable follows Bernoulli distribution. When an explanatory variable $X = x$, the probability of response variable $Y = y$ is

$$P(Y = y \mid X = x) = p_x^y(1 - p_x)^{1-y}, \qquad (y = 0, 1).$$

In the above, it usually means that 1 is success and 0 is fail, and the probability of success is specified as p_x. The probability of success also indicated by linear probability model, p_x

$= \alpha + \beta x$. However, it has some structural defects. First, the left term has a value between 0 and 1, but the right term has a whole real number. Also, when estimating α and β, least squares estimation no longer have minimum variance unbiased estimators for the parameters. Hence, the probability of success, p_x is indicated as

$$p_x = \frac{exp(\alpha + \beta x)}{1 + exp(\alpha + \beta x)}.$$

In the pathologic study or clinical research, logistic regression analysis identifies risk factors or important factors associated with the disease (Lee *et al.*, 2005). Because risk factor for a disease are rarely univariate, we can consider k explanatory variables $X_1, X_2, \cdots, X_k$.

$$p_x = \frac{exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}.$$

In the above equation, the logit model of success probability is

$$ln\left(\frac{p_x}{1 - p_x}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

In this paper, we apply to logistic regression analysis where the dependent variable is COM to identify risk factors, and propose a nomogram to predict the probability of COM using the results of logistic regression analysis.

3.3. Nomogram plot

In medical research, it is very important to find risk factors related to a disease, to predict the prevalence of the patient by using the risk factors, and to establish a treatment plan. Logistic regression model or Cox proportional hazard model is used to identify the risk factors. However, it is difficult for those who don't have enough statistics background. Recently, to predict the probability for a disease, a lot of nomograms have been developed (Ahn, 2013; Kim *et al.*, 2014).

Nomogram is useful for predicting the prevalence of each patient through the scoring system instead of using a complex formula. It graphically represents the numerical relationships between disease and its risk factors (Lee *et al.*, 2009). Without regard to statistical significance or sign of estimated regression coefficient, each predictor is assigned a score based on the estimated regression coefficients in a nomogram. Nomogram has points line, total points line, probability line, and predictor lines. First line is points line that assigns a score from each predictor. If we know predictor's value, we can draw a vertical line from predictor's value to point line. The most important factor is assigned 100 points, and sequentially other factors are assigned points in proportion to points of the most important factor. Because length of the lines is proportion to impact in the model, it is quickly recognized the importance of the risk factors influencing specific disease through nomogram. Then, we can calculate the total points by adding these scores to all predictors. Lastly, we can find the prediction probability by finding the corresponding points on the total points line and drawing a vertical line on the probability line. The composition of the nomogram can be obtained as follows (Iasonos *et al.*, 2008; Yang, 2014).

(1) Points line
Each risk factor is assigned a score between 0 and 100 points through the points line.

(2) Predictor line
Using the results from a logistic regression model, LP_{ij} (Linear predictor) value of each risk factor is calculated:

$$LP_{ij} = \beta_i \times X_{ij},$$

where $i = 1, 2, \cdots$, m: the number of risk factors, $j = 1, 2, \cdots, n_i$: The number of categories for each risk factor, β_i: the estimated regression coefficient in logistic regression, X_{ij} : the value of ith explanatory variables having jth category. A risk factor having the largest absolute value of the estimated regression coefficients is assigned 100 points and it is the most influential factor. The score of the remaining factors is calculated using calculated LP_{ij} value.

$$Point_{ij} = \frac{LP_{ij} - minLP_{ij}}{maxLP_{*j} - minLP_{*j}} \times 100,$$

where, LP_{*j}: the largest absolute value of the estimated regression coefficient. Each point of risk factors can be obtained by drawing a vertical line between points line and applicable category in each factor.

(3) Total points line and probability line
If the cumulative sum of the scores for each risk factor was calculated, the probability can be easily obtained through the nomogram. To calculate cumulative sum of the scores for each risk factor that matches the probability, we can use the following equation.

$$TotalPoints = ① \times (③ - ②),$$

① points per unit of linear predictor:

$$\frac{100}{maxLP_{*j} - minLP_{*j}}.$$

② $LP_{for\ TP=0}$ hold each factor at reference level (intercept).
③ $LP_{for\ TP>0} : ln\left(\frac{Risk\ of\ Y=1}{1-Risk\ of\ Y=1}\right)$.

Figure 3.1 is a nomogram that predicts the probability of clear-cell (CC) histology in renal cell cancer (Iasonos *et al.*; 2008). Probability can be obtained as follows. When the nomogram is constructed according to the above compositions, including points line, 4 predictor lines, total points line and probability line, each patient can get the probability of getting cancer. In case of a patient who is flow (No), age (75), clinical size (14), and sex (M), total points are about 115 points. The probability can be obtained by drawing a vertical line between total points line and probability line. Finally, the probability that the patient has cancer is about 50%. Therefore, the probability can easily be obtained only by clinical features.

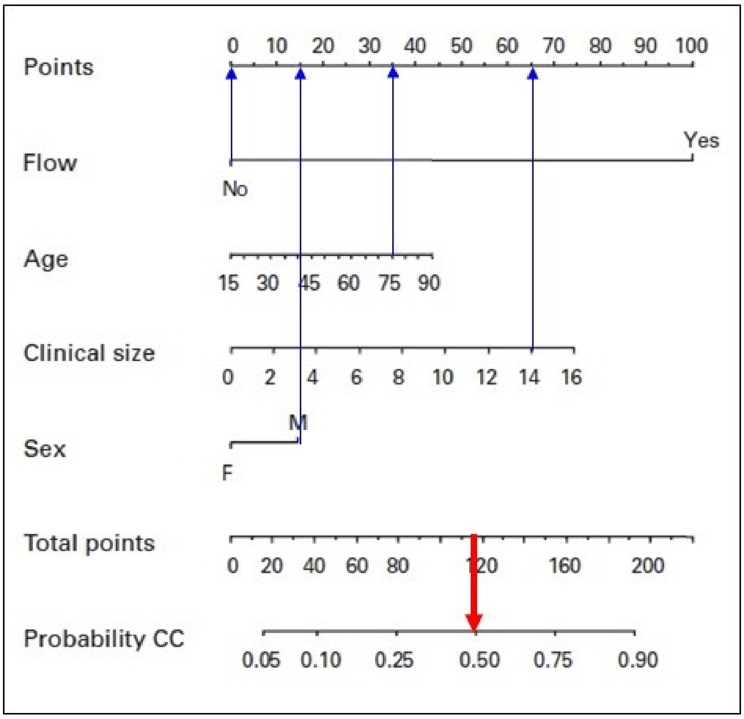

Figure 3.1 Example of nomogram (Iasonos *et al.*, 2008)

3.4. AUC (area under the ROC curve)

ROC (Receiver Operation Characteristic) curve is one of the methods that has traditionally been used in order to evaluate the performance of a predictive model using diagram in the field of discriminant analysis (Pearce and Ferrier, 2000). It is graphical plot drawing a connecting line with 'sensitivity' on vertical axis against '1-specificity' on horizontal axis. If AUC is close to 1, it indicates a good performance of the predictive model. On the other hand, If AUC is close to 0.5 indicate a poor performance of the predictive model.

4. Results

4.1. Chi-square test and multivariable logistic regression analysis for COM

Before establishing a nomogram for COM, Chi-square test and multivariable logistic regression analysis were used to identify the risk factors. In training set, the weighted prevalence of COM was 5.3%. In Table 4.1, Table 4.2, and Table 4.3, residence, earphone use, atopic dermatitis, allergic rhinitis, chronic rhinosinusitis, subjective hearing, and tinnitus are significant. People who live in urban, use an earphone in noisy situations, suffered from atopic dermatitis, allergic rhinitis, chronic rhinosinusitis, or tinnitus, or have very poor subjective hearing status have more COM.

Table 4.1 General characteristics of KNHANES participants

Characteristics	Unweighted total number	COM weighted, % (SE) Yes	No	p-value
Overall	3959	5.3 (0.4)	94.7 (0.4)	
Age (year)				0.5617
40-49	1303	5.1 (0.6)	94.9 (0.6)	
50-59	1182	5.0 (0.7)	95.0 (0.7)	
60-69	843	6.6 (1.0)	93.4 (1.0)	
$\geq$ 70	631	5.1 (1.1)	94.9 (1.1)	
Sex				0.6220
Male	1339	5.1 (0.7)	94.9 (0.7)	
Female	2620	5.5 (0.5)	94.5 (0.5)	
Education				0.4300
$\leq$ Middle school	2270	5.5 (0.5)	94.5 (0.5)	
$\geq$ High school	1689	4.9 (0.6)	95.1 (0.6)	
Income				0.8568
$<$ 25%	953	5.2 (1.0)	94.8 (1.0)	
25-50%	1006	5.6 (0.8)	94.4 (0.8)	
50-75%	1007	5.7 (0.9)	94.3 (0.9)	
$\geq$ 75%	993	4.8 (0.7)	95.2 (0.7)	
Occupation				0.3357
White collar	2905	5.6 (0.5)	94.4 (0.5)	
Blue collar	1054	4.7 (0.7)	95.3 (0.7)	
Residence				0.0044
Urban	3161	5.9 (0.5)	94.1 (0.5)	
Rural	798	2.9 (0.7)	97.1 (0.7)	
BMI (kg/m^2)				0.6265
$<$ 25	2432	5.2 (0.5)	94.8 (0.5)	
$\geq$ 25	1527	5.6 (0.7)	94.4 (0.7)	
Earphone use				0.0217
Yes	201	8.9 (2.1)	91.1 (2.1)	
No	3758	5.1 (0.4)	94.9 (0.3)	
Temporary exposure				0.7415
Yes	1369	5.1 (0.7)	94.9 (0.7)	
No	2590	5.4 (0.5)	94.6 (0.5)	
Smoking status				0.7721
Yes	1032	5.5 (0.8)	94.5 (0.8)	
No	2927	5.2 (0.5)	94.8 (0.5)	
Alcohol consumption				0.2866
Yes	2108	5.0 (0.5)	95.0 (0.5)	
No	1851	5.8 (0.6)	94.2 (0.6)	
Household members				0.7068
1-2	1648	5.5 (0.7)	94.5 (0.7)	
3-4	1908	5.4 (0.6)	94.6 (0.6)	
5-6	403	4.4 (1.1)	95.6 (1.1)	
Subjective health status				0.1203
Very good	169	2.2 (1.1)	97.8 (1.1)	
Good	873	4.7 (0.8)	95.3 (0.8)	
Average	2137	5.3 (0.5)	94.7 (0.5)	
Poor	596	6.8 (1.2)	93.2 (1.2)	
Very poor	184	7.4 (2.1)	92.6 (2.1)	

Logistic regression analysis was performed in Table 4.4. It is a result of the multivariable analyses using backward method to select variables. Independent predictors for COM were residence (adjusted OR: 2.13; 95% CI: 1.29-3.52), earphone use (adjusted OR: 1.59; 95%

Table 4.2 Medical conditions of KNHANES participants

Characteristics	Unweighted total number	COM weighted, % (SE) Yes	No	p-value
Hypertension				0.1052
Yes	2956	5.6 (0.5)	94.4 (0.5)	
No	1003	4.3 (0.7)	95.7 (0.7)	
Diabetes mellitus				0.9241
Yes	391	5.4 (1.5)	94.6 (1.5)	
No	3568	5.3 (0.4)	94.7 (0.4)	
Hepatitis B				0.2213
Yes	48	1.7 (1.7)	98.3 (1.7)	
No	3911	5.4 (0.4)	94.6 (0.4)	
Atopic dermatitis				0.0033
Yes	48	17.1 (6.7)	82.9 (6.7)	
No	3911	5.2 (0.4)	94.8 (0.4)	
Allergic rhinitis				0.0153
Yes	427	8.0 (1.4)	92.0 (1.4)	
No	3532	5.0 (0.5)	95.0 (0.5)	
Chronic rhinosinusitis				0.0003
Yes	298	9.8 (1.8)	90.2 (1.8)	
No	3661	4.9 (0.4)	95.1 (0.4)	

Table 4.3 Otologic conditions of KNHANES participants

Characteristics	Unweighted total number	COM weighted, % (SE) Yes	No	p-value
Subjective hearing				< 0.0001
Not discomfort	3190	3.7 (0.4)	96.3 (0.4)	
A little discomfort	638	12.7 (1.5)	87.3 (1.5)	
$\geq$ A lot of discomfort	131	16.4 (3.4)	83.6 (3.4)	
Tinnitus				0.0006
Yes	942	7.6 (1.0)	92.4 (1.0)	
No	3017	4.7 (0.4)	95.3 (0.4)	

CI: 0.93-2.70), atopic dermatitis (adjusted OR: 2.87; 95% CI: 1.18-7.01), allergic rhinitis (adjusted OR: 1.48; 95% CI: 0.95-2.30), chronic rhinosinusitis (adjusted OR: 1.70; 95% CI: 1.11-2.60), mild hearing impairment (adjusted OR: 3.86; 95% CI: 2.77-5.38), and critical hearing impairment (adjusted OR: 5.94; 95% CI: 3.53-9.99).

Table 4.4 Odds ratio for COM of KNHANES participants

	Adjusted OR, %(SE)	95 %CI	p-value
Residence (Urban)	2.13	1.29-3.52	0.0031
Earphone use	1.59	0.93-2.70	0.0891
Atopic dermatitis	2.87	1.18-7.01	0.0208
Allergic rhinitis	1.48	0.95-2.30	0.0852
Chronic rhinosinusitis	1.70	1.11-2.60	0.0152
Subjective hearing			
Not discomfort	1.00		
A little discomfort	3.86	2.77-5.38	< 0.0001
$\geq$ A lot of discomfort	5.94	3.53-9.99	< 0.0001

4.2. Nomogram construction for COM

The purpose of this paper is building a nomogram to predict the prevalence of COM through the result of logistic regression analysis (Figure 4.1). Nomogram graphically represents the numerical relationships between COM and 6 risk factors (residence, earphone use, atopic dermatitis, allergic rhinitis, chronic rhinosinusitis, and subjective hearing), instead of interpretation of complex models. In the nomogram, each patient receives a point for each factor. The composition of the nomogram is point line, 6 predictor lines, total point line, and probability line. For each risk factor, the assigned points are following: "residence in urban" is 42 points, "earphone use in noisy situations" is 26 points, "atopic dermatitis" is 59 points, "allergic rhinitis" is 22 points, "chronic rhinosinusitis" is 30 points, "a little discomfort hearing status" is 76 points, and "a lot of discomfort hearing status" is 100 points. The higher point in each factor, the more important factor of COM. As we can be seen from the length of the line, subjective hearing status is the most influential risk factor among the 6 factors. The total point line is the cumulative sum of the points assigned to each of the 6 risk factors. At probability line, prevalence of COM is obtained by drawing a vertical line between total points line and probability line.

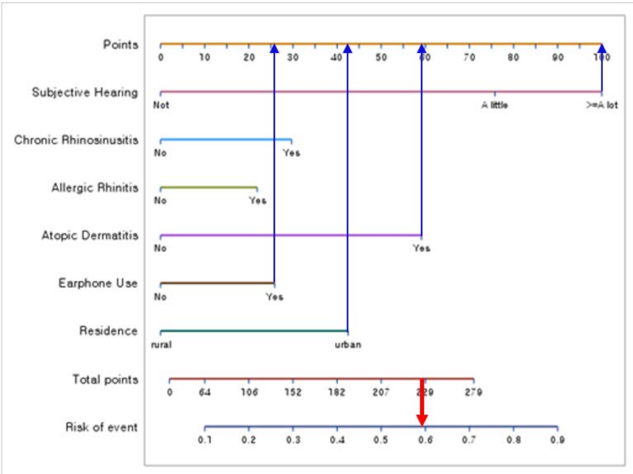

Figure 4.1 Proposed a nomogram for COM

For example, here is a patient who has a lot of discomfort hearing status, suffered from atopic dermatitis, use earphone in noisy situations, live in urban. Total point of the patient is about 227, and prevalence of COM is expected as about 60%. In this case, the risk of having COM is rather high, the patient is need to establish a treatment plan for COM.

4.3. Validation of nomogram for COM: AUC

To verify the discrimination of the nomogram, AUC was calculated in Figure 4.2. ROC curves were illustrated using the training data and the validation data. The AUC of training data is 0.70 ($p < 0.001$), and the AUC of validation data is 0.63 ($p = 0.009$). The AUC for the training data, the same data using for selecting the risk factors and building the nomogram, was statistically significant, and validation data, the new data, was also statistically significant. Therefore, we show a statistically significant determination.

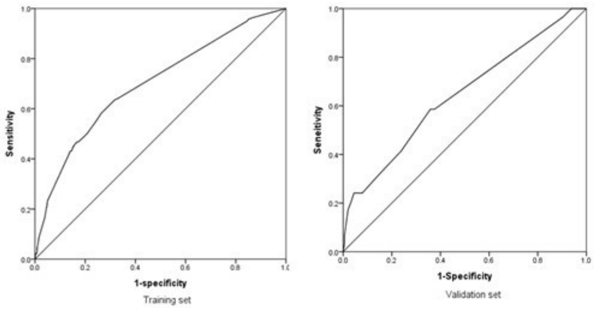

Figure 4.2 ROC (receiver operation characteristic) curve for COM

5. Discussion

In medical study, it is important to identify risk factors related to diseases. Logistic regression model or Cox proportional hazard model is usually used to find them. However, these methods are difficult for researchers who don't have statistics background to use them. Nomogram is useful for predicting the prevalence or the survival rate of each patient through the scoring system without a complex formula. Regardless of statistical significance or sign of estimated regression coefficient, each predictor is assigned a score based on the estimated regression coefficients in nomogram. The most important factor is assigned 100 points, and sequentially other factors is assigned points in proportion to points of the most important factor. As a result, length of the lines is proportion to impact in the model. If a nomogram for a disease is established, anyone can simply calculate their prevalence. Many studies have been conducted on the prevalence and risk factors of COM. The prevalence of otitis media is significantly higher in children who have not fully developed internal organs of the ear compared to other ages. There are many Studies on COM targeting children, but studies of risk factors for otitis media in adults are lacking. The purpose of our study is to construct a nomogram for COM using risk factors of COM, and to validate the nomogram. The data were collected from people who participated in the Korean National Health and Nutrition Examination Survey for three years from 2013 to 2015. There is few community-based study examining the prevalence and risk factors of COM among the Korean population, and data from KNHANES are comprehensive and nationally representative (Park and Moon, 2014). The weighted prevalence of COM in Korean adults aged ¿ 40 years was 5.3%; we considered 21 factors associated with COM. First, we performed the chi-square test to identify the significance of COM. 6 factors associated with COM were tested in the multiple logistic regression model. Finally, these factors were selected as risk factors for COM: residence, earphone use in noisy situations, atopic dermatitis, allergic rhinitis, chronic rhinosinusitis, subjective hearing status. Using these factors, we built a nomogram for COM. In the nomogram, the higher point in each factor, the more important factor of COM. As we can be seen from the length of the line, subjective hearing status is the most influential risk factor among the 6 factors, and atopic dermatitis, residence, chronic rhinosinusitis, earphone use in noisy situations, and allergic rhinitis have influenced on COM, orderly. Also, we validated the model with AUC. AUC is 0.70 and 0.63 in training set and validation set, respectively. It shows our performance was good for predicting the probability.

The association between chronic rhinosinusitis and COM is shown in (Park *et al.*, 2015). Due to the deterioration of the living environment, studies on allergic rhinitis and atopic dermatitis have been steadily increasing in Korea (Han and Park, 2016; Lee and Kim, 2014). The association between allergic rhinitis and COM is well documented (Gultekin *et al.*, 2010; Yeo *et al.*, 2007). Furthermore, in our study, COM is significantly affected form residence in urban and earphone use in noisy situations. The association found in our study seems to reflect the special situations of Korean adults aged over 40 years. In our study, subjective hearing discomfort are associated with COM. This results are related with Park *et al.*, 2015. Meanwhile, we were able to see the degree of COM's prevalence by the risk factors. Therefore, health care workers can predict the risk of COM for an individual patient early on, which will be useful for making a treatment plan.

In conclusion, our study built a nomogram based on the results of logistic regression analysis. Only with the nomogram, each patient receives a point for each risk factor, and the higher the total points of these points, the higher the prevalence of COM. It is easy to compare the relative magnitudes of each risk factor to COM, and to establish a treatment plan through the prevalence for COM easily obtained through the nomogram.

References

Ahn, J. H. (2013). *Nomogram for prediction of prostate cancer in Korean men with serum prostate-specific antigen less than 10ng/mL*, Master's Degree Thesis, Pusan National University, Busan.

Collaborators (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. *The Lancet*, **385**, 117-171.

Engel, J. A., Straetemans, M. and Zielhuis, G. A. (2005). Birth characteristics and recurrent otitis media with effusion in young children. *International Journal of Pediatric Otorhinolaryngology*, **69**, 533-540.

Gozal, D., Kheirandish-Gozal, L., Capdevila, O. S., Dayyat, E. and Kheirandish, E. (2008). Prevalence of recurrent otitis media in habitually snoring school-aged children. *Sleep Medicine*, **9**, 549-554.

Gultekin, E., Develioğlu, O. N., Yener, M., Ozdemir, I and Külekçi, M. (2010). Prevalence and risk factors for persistent otitis media with effusion in primary school children in Istanbul, Turkey. *Auris Nasus Larynx*, **37**, 145-149.

Han J. Y. and Park H. S. (2016). Prevalence of allergic diseases and its related factors in Korean adolescents-Using data from the 2013 Korea youth risk behavior web-based survey. *Journal of the Korean Data & Information Science Society*, **27**, 155-168.

Iasonos, A., Schrag, D., Raj, G. V. and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis. *Journal of Clinical Oncology*, **26**, 1364-1370.

Kim, S. H., Shin, K., Kim, H. Y., Cho, Y. J., Noh, J. K., Suh, J. S. and Yang, W. I. (2014). Postoperative nomogram to predict the probability of metastasis in Enneking stage IIB extremity osteosarcoma. *BioMed Cental Cancer*, **12**, 666.

Lee, J. W., Park, M. R. and Yu, H. N. (2005). *Statistical method for bioscience research*, freedom academy, Seoul.

Lee, J. Y. and Kim, H. J. (2014). Identification of major risk factors association with respiratory diseases by data mining. *Journal of the Korean Data & Information Science Society*, **25**, 373-384.

Lee, K. M., Kim, W. J. and Yun, S. J. (2009). A clinical nomogram construction method using genetic algorithm and naïve Bayesian technique. *Journal of Korean Institute of Intelligent Systems*, **19**, 769-801.

Park, M., Lee, J. S., Lee, J. H., Oh, S. H. and Park, M. K. (2015). Prevalence and risk factors of chronic otitis media: the Korean national health and nutrition examination survey 2010-2012. *PLoS ONE*, **10**, 1-13.

Park, R. J. and Moon, J. D. (2014). Prevalence and risk factors of tinnitus: The Korean national health and nutrition examination survey 2010011, a cross-sectional study. *Clinical Otolaryngology*, **39**, 89-94.

Pearce, J. and Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic. *Ecological Modelling*, **133**, 225-245.

Qureishi, A., Lee Y., Belfield, K., Birchall, J. P. and Daniel, M. (2014). Update on otitis media - prevention and treatment. *Infection and Drug Resistance*, **7**, 15-24.

Teele D. W., Klein J. O. and Rosner B. (1989) Epidemiology of otitis media during the first seven years of life in children in greater Boston: a prospective, cohort study. *The Journal of Infectious Diseases*, **160**, 83-94.

Van Zon, A., Van der Heijden, G. J., Van Dongen, T. M. A, Burton, M. J. and Schilder, A. G M. (2013). Antibiotics for otitis media with effusion in children. *Clinical Otolaryngology*, **38**, 54-55.

Yang, D. (2014). Build prognostic nomograms for risk assessment using SAS. *In Proceedings of SAS Global Forum 2013. Paper 264013*. http://support.sas.com/resources/papers/proceedings13/264-2013.pdf.

Yeo, S. G., Park, D. C., Eun, Y. G. and Cha, C. I. (2007). The role of allergic rhinitis in the development of otitis media with effusion: effect on eustachian tube function. *American Journal of Otolaryngology*, **28**, 148-152.