The Risk Factors of Laryngeal Pathology in Korean Adults Using a Decision Tree Model

*,†Haewon Byeon, *†Gwangju, Republic of Korea

Summary: Objective. The purpose of this study was to identify risk factors affecting laryngeal pathology in the Korean population and to evaluate the derived prediction model.
Study Design. Cross-sectional study.
Methods. Data were drawn from the 2008 Korea National Health and Nutritional Examination Survey. The subjects were 3135 persons (1508 male and 2114 female) aged 19 years and older living in the community. The independent variables were age, sex, occupation, smoking, alcohol drinking, and self-reported voice problems. A decision tree analysis was done to identify risk factors for predicting a model of laryngeal pathology.
Results. The significant risk factors of laryngeal pathology were age, gender, occupation, smoking, and self-reported voice problem in decision tree model. Four significant paths were identified in the decision tree model for the prediction of laryngeal pathology. Those identified as high risk groups for laryngeal pathology included those who self-reported a voice problem, those who were males in their 50s who did not recognize a voice problem, those who were not economically active males in their 40s, and male workers aged 19 and over and under 50 or 60 and over who currently smoked.
Conclusions. The results of this study suggest that individual risk factors, such as age, sex, occupation, health behavior, and self-reported voice problem, affect the onset of laryngeal pathology in a complex manner. Based on the results of this study, early management of the high-risk groups is needed for the prevention of laryngeal pathology.
Key Words: Decision tree–Voice disorder–Risk factor–Laryngeal pathology–Dysphonia.

INTRODUCTION

Laryngeal pathology is very frequent in the general population. In Korea, the prevalence of laryngeal pathology has been reported to be 7%.[1] Also, considering that the lifetime prevalence of laryngeal pathology has been reported to be approximately 29% in a recent epidemiologic study,[2] approximately 12 million Koreans experience a laryngeal pathology at least once in their lifetime. Laryngeal pathology had a higher prevalence than stroke in the same period in 2008.[3] However, the importance of laryngeal pathology has been overlooked in terms of public health because it does not cause death directly. However, laryngeal pathology limits daily life including labor activities due to functional problems in communication, resulting in economic loss. As an example, social costs in teachers caused by dysphonia are estimated at $2.5 billion/y in the United States of America.[4] Accordingly, it is important to clarify risk factors clearly in laryngeal pathology as taking precautions is more efficient than giving treatment in reducing economic loss due to laryngeal pathology and in decreasing the prevalence rate of laryngeal pathology.

According to the recent epidemiological studies, age, gender, self-reported voice problem, occupation, alcohol drinking, smoking, high education level, xerostomia, and asthma were reported as independent risk factors for dysphonia.[2,5–8] However, the preceding studies have differences in the applied confounding variables for each study and they applied regression models as a method to estimate risk factors. Hence, they are effective in exploring independent risk factors but have difficulties in showing the priority of risk factors. In addition, it is likely that the assumptions that are made in a regression model may be violated in the case when the data of diseases or disorders are used in the model, because linear regression models need assumptions to be made, including assumptions about the linearity, normality, and homoscedasticity of the data among others.

Recently, data mining, which is used in a decision tree and neural network analysis among others, has been applied to methods in classifying or estimating the objects of interest in various areas including health studies, medicine, and so on. In particular, the decision tree analysis not only can analyze nonparametric data but also is expressed by a tree structure for its analysis process, and the basis for prediction can be easily understood compared with the neural network or cluster analysis.[9] In addition, decision tree model is effective to use in clinical settings because the primary related factors, among other several related factors, can be identified.

The aim of this study was to investigate risk factors of laryngeal pathology and develop a prediction model for laryngeal pathology.

METHODS

Subjects

This research targeted the general population aged 19 years or older who had completed the 2008 Korea National Health and Nutrition Examination Survey (KNHANES), a nationwide survey of noninstitutionalized residents in South Korea, and who then participated in an otolaryngology examination. The multi-stage cluster sampling design and administration of KNHANES are described in detail elsewhere.[3] Briefly, from

Accepted for publication April 4, 2014.
From the *Department of Speech Language Pathology & Audiology, Nambu University, Gwangju, Republic of Korea; and the †Speech-Language Pathology Center, Nambu University, Gwangju, Republic of Korea.
Address correspondence and reprint requests to Haewon Byeon, Department of Speech Language Pathology & Audiology, Nambu University, Cheomdan Jungangro 23, Gwangsan-gu, Gwangju 506-706, Republic of Korea. E-mail: byeon@nambu.ac.kr
Journal of Voice, Vol. 29, No. 1, pp. 59-64
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2014.04.004

each of the 200 sampling units, 20–23 households were selected by systematic sampling, yielding 12 528 persons in 4600 households. The field survey was conducted by specially trained interviewers at mobile centers and in the participants' households. Among questionnaire surveys, survey on level of education and economic activity was conducted by individual interviews, and survey on the self-reported voice problem and behaviors regarding health such as smoking was executed with self-administered questionnaires from January to December in 2008. These surveys were completed by 9308 participants. This research targeted 3626 men and women who completed all of the health surveys, otolaryngology surveys and examinations, and laryngoscope examinations. Among them, 15 nonrespondents and 476 persons whose laryngoscopic findings could not be determined were excluded from the research, and 3135 persons (1310 men and 1825 women) were analyzed.

Measurement

Laryngeal pathology. Laryngeal pathology in this study were included as benign vocal fold lesions (eg, vocal nodules, vocal polyp, and vocal fold cyst), Reinke edema, laryngeal granuloma, laryngeal keratosis, laryngeal papilloma, sulcus vocalis, laryngitis, and suspected malignant neoplasm of the larynx. Although a variety of definitions of the term vocal fold lesions have been suggested, this study used the definition suggested by Rosen et al.[10] Laryngoscopic examination of adults above the age of 19 was carried out by otolaryngologists using 70° endoscopes, and abnormalities of the larynx were studied. The index of coincidence evaluation was executed twice, and the quality improvement committee reevaluated the pictures and videos examined by the otolaryngologists and computed the results. The index of coincidence for the laryngoscopic examination was 75%. Voice ailment data that were classified from the laryngoscopic examination were reclassified as "laryngeal pathology" and "no laryngeal pathology" by the researcher.

Demographic factors. The age, gender, education level, and occupation were examined. Education levels were classified into below elementary school graduation, below , below high school graduation, and above college graduation. The occupation was classified as economically inactive (unemployed persons, housewives, and students), nonmanual (managers, clerical workers, and service and sales workers), and manual (skilled agricultural, forestry and fishery workers, craft and related trades workers, and elementary occupations).[11]

Healthy behavior factor. Health behavioral factors included smoking and drinking status. Current smokers were defined as those who had smoked 100 cigarettes or more during their lifetime and were currently smoking. Nonsmokers were those who had never smoked or had smoked less than 100 cigarettes in their lifetime but currently did not smoke. Alcohol consumption was defined as less than once a week, twice or three times a week, or above four times a week during the last 1 year.

Self-reported voice problem. Self-reported voice problems were surveyed based on the question "Do you think that you have any voice problem now?" in the otolaryngology survey. Those who currently reported having problems with their voice were classified as having self-reported voice problems.

Data analysis

The weighted mean, standard error, and percentile were presented using a descriptive analysis. The weights for the KNHANES were placed such that individuals participating in the survey represented the entire population of Korea.[3] Weighted one-way analysis of variance (ANOVA) and the Rao-Scott chi-square test were used to compare the age, sex, education level, occupation, smoking, alcohol consumption, and self-reported voice problem in the laryngeal pathology group and the no laryngeal pathology group. The explanatory variables with a significance level below 0.1 were defined as the related factors of laryngeal pathology and were included in the decision tree model.

When the related factors of laryngeal pathology were identified in the Rao-Scott chi-square test and weighted one-way ANOVA, the related factors of laryngeal pathology were statistically classified and a prediction model was established, using a decision tree, which is a data mining technique. The classification and regression tree (CART) algorithm was used to predict the related factors in the decision tree model. Measuring impurity with the Gini index, CART is an algorithm that performs a binary split, where only two child nodes are formed from the parent node.[12] In the CART algorithm, the alpha value for the criteria of splitting and merging was set at 0.05. The number of parent nodes was 200 and that of child nodes was 80, and the number of branches was limited to 5. To make up for the imbalance in data distribution, the weights for misclassification costs were set asymmetrically, considering the prevalence of laryngeal pathology of Koreans.[13] The validity of the model was first tested using a misclassification table, and then the risks of the model were compared using the 10-fold cross validation. *Decision Tree version 18.0* (SPSS Inc., Chicago, Illinois) was used for all the analyses. Significance level was set at 0.01 in the paired test only for the chi-square test to investigate the related factors of laryngeal pathology and at 0.05 in the paired test for all other analyses.

RESULTS

Characteristics of participants and prevalence of laryngeal pathology

The general characteristics of the subjects by the presence of laryngeal pathology are presented in Table 1. Of the entire 3135 subjects, 2887 (92.1%) were in the normal group and 248 (7.9%) were in the laryngeal pathology group. The mean ages of the normal group and laryngeal pathology group were 43.4 and 47.9 years, respectively. The weighted one-way ANOVA test result showed that the age of the laryngeal pathology group was statistically significantly higher ($P < 0.001$). The result of the Rao-Scott chi-square test indicated that the normal group and laryngeal pathology group