Risk Prediction Model Occurrence of Obstructive Sleep Apnea (OSA) Risk in Chronic Obstructive Pulmonary Disease patient. Case Study: Cipto Mangunkusumo Hospital

I R Anisa¹, T Siswantining¹,*, T Kamelia² and D Sarwinda¹

¹ Departemen of Mathematics, FMIPA Universitas Indonesia, Kampus UI Depok, Depok 16242, Indonesia. ² Pulmonology Division, Internal Medicine Department, Universitas Indonesia/Dr Cipto Mangunkusumo Hospital, Jakarta, Indonesia

*Corresponding author: titin@sci.ui.ac.id.

Abstract. The complication occur in Chronic Obstructive Pulmonary Disease (COPD) patient is nocturnal hypoxemia. Obstructive Sleep Apnea (OSA) and COPD can occur at the same time which cause two times uncomfortable while breathing. The purpose of this research is to determine prediction model occurrence OSA risk in COPD patient based on factor affecting the risk of OSA occurring in COPD patients. Data used in this research is primary data from COPD patients who is diagnosed by doctor at RSCM by interviewing them using Berlin questionnaire and physical examination such as measuring the circumference of neck and waist. This study uses non-probability sampling i.e. purposive sampling method. Sample of this research is 111 patients with COPD. This research uses binary logistic regression to predict model occurrence of OSA risk in COPD patients. This study shows that waist circumference and COPD Assessment Test (CAT) 2 questionnaire (COPD patients with severe degree) are significant factor of OSA on COPD patient. In addition, COPD patients with severe degree are 4.39 times greater risk suffer from OSA than mild to moderate COPD patients and each centimetre increase of waist circumference has higher risk of OSA. Accuracy of our model is estimated using classification table with cut point at 0.5 and its accuracy is 73.9%.

1. Introduction
COPD and OSA are two common pulmonary diseases. Chronic Obstructive Pulmonary Disease (COPD) is a preventable and treatable pulmonary disease indicated by persistent respiratory system and airflow barriers caused by airway or alveolar abnormalities that are usually caused by hazardous particles and gases [1]. To determine the severity of COPD symptoms experienced by patients, it can be calculated based on the questionnaire that is COPD Assessment Test (CAT) and modified Medical Research Council (mMRC)[1].

COPD is the fifth leading cause of death in the world (in 2002) and cause 5% of death in the world (in 2005). It is also known that almost 90% of COPD patients’ death occur in low and middle-income countries [2]. Nocturnal hypoxemia is a complication suffered by COPD patients where oxygen intake decreases at night. This condition is exacerbated if the patient has Obstructive Sleep Apnea (OSA) [3].

On the other hand, OSA is a sleeping disorder caused by occluded airway. Normally, the soft tissue in the upper respiratory tract, including the tongue and the throat wall, will relax during sleep. In people with OSA, when the tissue relaxes, it blocks the airway, which causes temporary pauses in breathing. These pauses are brief and may only last a few seconds (at least 10 seconds) but it can
happen several times during an hour of sleeping [4]. In China, man workers aged 30-60 years old suffer from OSA is 41% [5]. However, the prevalence for OSA populations overall ranged from 9% to 38% and is higher in men. In some elderly groups, it reaches 90% in men and 78% in women [6].

Unfortunately, COPD and OSA can happen at the same time, known as Overlap Syndrome, causing two times uncomfortable condition while breathing [3]. Two recent researches observe the prevalence of COPD patients with positive OSA is 65.9% in lung rehabilitation [7] and 10.8% in Korea [8].

Polysomnography (PSG) is widely used to diagnose OSA. While the severity degree of OSA is calculated by Apnea Hypopnea Index (AHI). AHI is divided into 3 severity classification, which are 5-14 for low degree of OSA, 15-30 for middle degree, and more than 30 for high degree of OSA [9]. While, the Berlin questionnaire as a tool for capturing COPD patients at risk for OSA and Berlin questionnaire used is a validated questionnaire for the population of Indonesia [10].

Model to estimate the risk of OSA in COPD patients is not invented yet, both in Indonesia and other countries. Therefore, this study attempts to determine estimation model of OSA on COPD patients by observing COPD patients in Rumah Sakit Cipto Mangunkusumo (RSCM).

2. Method

Regression analysis has become an integral component of the analysis of data pertaining to describe the relationship between the response variable and one or more predictor variables [11].

2.1. Binary Logistic Regression

Binary logistic regression is a statistical technique used to analyse the relationship of response variables that have two categories i.e. success and failure. Therefore, in prediction variable data type can be a nominal, ordinal, interval nor ratio.

Independent variable shows by \( x_i \), \( i = 1, \ldots, p \) and dependent variable \( Y_i \) consists of two categories: success and failure, denoted \( y_i = 1 \) (success) and \( y_i = 0 \) (fail), let \( Y_i \) as Bernoulli distributed with these functions:

\[
f(y_i) = \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i}, y_i = 0,1
\]

where probability of success is \( \Pr(Y_i = 1|x_i) = \pi(x_i) \) and probability of failure is \( \Pr(Y_i = 0|x_i) = 1 - \pi(x_i) \). Regression model that formed \( \mathbb{E}(Y_i = 1|x_i) = \pi(x_i) \) is as follows.

\[
\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}}}
\]

In a logistic regression model required a connecting function in accordance with the logistic regression model is the logit function. Logit transformation as a function of \( \pi(x_i) \) stated as follows:

\[
\text{logit } [\pi(x_i)] = g(x_i) = \ln \left( \frac{\pi(x_i)}{1 - \pi(x_i)} \right)
\]

2.2. Parameter Estimator

Commonly used method for parameter logistic regression model is Maximum Likelihood Estimation (MLE). MLE estimated \( \beta \) parameter by maximizing likelihood function and requiring data must follow a distribution. Each observation is assumed mutually independent so that likelihood function is the multiplication of the probability function of each observation as follows.

\[
l(\beta) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i}
\]

For \( i = 1, 2, \ldots, n \), \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \) is an independent variable for the i-th observation and \( y_i \) is a dependent variable on binary i-th with values 0 and 1.

Mathematically, \( \beta \) parameter will be evaluated by maximizing the logarithm function probability as known as log likelihood defined as follows.
\[ L(\beta) = \sum_{i=1}^{n} \{ y_i \ln[\pi(x_{ji})] + (1 - y_i)\ln[1 - \pi(x_{ji})] \} \]  

(5)

The maximum value of \( \beta \) is obtained through the \( l(\beta) \) to \( \beta \) and the result is zero.

\[ \frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} x_{ji}[y_i - \pi(x_{ji})] = 0 \quad , j = 0,1,2, ..., p \]  

(6)

The value of \( \beta \) obtained from equation (6) called maximum estimated likelihood and denoted as \( \hat{\beta} \). Thus, prediction value for logistic regression model is \( \hat{\pi}(x_{ji}) \), where value from equation (2) is calculated using \( \hat{\beta} \) and \( x_{ji} \).

2.3. Parameter Significance Testing

Significance test consists of two stages: the significance test of model parameters simultaneously and test the significance of the model parameters partially.

1. **Simultaneously test**

Simultaneous test is used to determine the significance of \( \beta \) parameter to dependent variable together. The hypothesis is used as follows.

\[ H_0: \beta_1 = \beta_2 = \cdots = \beta_i = 0 \]
\[ H_1: \text{at least } \beta_1 \neq 0 ; i = 1,2, ..., p \]

Tests on these parameters is done through G test statistics are stated as follows.

\[ G = -2 \ln \left( \frac{L_0}{L_p} \right) \]  

(7)

\( L_0 = \) likelihood value without independent variable
\( L_p = \) likelihood value with \( p \) independent variable

The G test statistic is a Likelihood Ratio Test where the \( G \) value follows the Chi-Square distribution with the degrees of freedom as much as the independent variable \( p \). The decision is obtained from the comparison between the test statistic value of \( G \) and \( \chi^2 \) tables. \( H_0 \) is rejected if the test statistic value \( G > \chi^2_{(\alpha,p)} \).

2. **Partially test**

Partial test is a variable test performed one by one using Wald (W) test statistic. The partial test hypothesis is as follows.

\[ H_0: \beta_i = 0 \]
\[ H_1: \beta_i \neq 0 ; i = 1,2, ..., p \]

the test statistic used is as follows.

\[ W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \]  

(8)

with,

\( \hat{\beta}_i = \) estimator \( \beta_i \)

\( SE(\hat{\beta}_i) = \) standard error \( \beta_i \)
2.4. Coefficient Interpretation

Interpretation of the parameter coefficients is done to determine the trend / functional relationship between independent variables with dependent variables and show the effect of changes in value on the dependent variable and shows the effect of changes in the value of the variable concerned.

Logistic Regression with Binary Independent Variable quantifying with 0 and 1. In general the value of y is continuous, but in this case the variable y becomes discrete, where the value is 0 and 1 so that the value of y is transformed like a probability value called the odds ratio. As for the logistic regression of continuous variables is to assume a linear logit function of the independent variable.

The interpretation is any increase in c unit independent variables, the risk of \( y = 1 \) is equal to \( e^{\beta_1} \) times bigger.

Odds ratio is defined.

\[
\text{OR} = e^{\beta_1}
\]

2.5. Classification Table

Table classification was conducted to determine the accuracy of the prediction model. Classification table is the result of cross-classification between the dependent variable (Y) with a new dichotomous variable, where value is derived from statistical probability estimates \( \hat{\pi}(\mathbf{x}_i) \).

**Table 1. Classification Table using Logistic Regression Model with Cut Point (c)**

| Prediction | Observation | Total |
|---|---|---|
| \( \hat{y} = 1 \) | \( y = 1 \) | \( n_{11} \) |
| \( \hat{y} = 1 \) | \( y = 0 \) | \( n_{10} \) |
| \( \hat{y} = 0 \) | \( y = 1 \) | \( n_{01} \) |
| \( \hat{y} = 0 \) | \( y = 0 \) | \( n_{00} \) |
| Total | \( n_1 \) | \( n \) |

Classification accuracy was measured with high sensitivity and specificity. Sensitivity is the ability to predict success events correctly. While specificity is the ability to predict the occurrence of failure and true. Sensitivity and specificity are calculated as follows.

\[
\text{Sensitivity} = \frac{n_{11}}{n_{11} + n_{01}} \times 100%
\]

\[
\text{Specificity} = \frac{n_{00}}{n_{00} + n_{10}} \times 100%
\]

3. Results and Discussion

3.1 Data

The population in this study were all patients of COPD at Cipto Mangunkusumo hospital. The sample selection for this research uses non-probability sampling with purposive sampling type, where the sample is not taken randomly but rather determined by the researcher based on certain criteria. The number of samples in this study were 111 patients who had been diagnosed with COPD by doctors at RSCM by meeting inclusion and exclusion criteria.

**Table 2. Inclusion and Exclusion Criteria in Patients with COPD**

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| COPD patients with age \( \geq 25 \) years | Patients refuse to be a research sample |
| Willing to take the study by interviewed by COPD Assessment Test (CAT) questionnaire and modified Medical Research Council | Pregnant mother |
3.2 Numerical Result

Characteristics of patients with COPD in this study are based on the risk of OSA, gender, age, body mass index (BMI), CAT questionnaire, mMRC questionnaire, smoking history, neck circumference and waist circumference. By gender showed that most respondents were male as many as 82 people (74%) while the female respondents 29 people (26%) with an average age of respondents was 59.77 years, the majority of respondents in this study experienced shortness of breath are 75 respondents (68%), a total of 39 respondents (35%) in this study has moderate COPD, 36 respondents (33%) with mild COPD and 36 respondents (32%) of COPD severity.

Based on questionnaires Berlin showed that the respondents were at risk of OSA totaling 47 respondents (47%) and respondents who are not at risk of OSA amounted to 64 respondents (58%) with a mean neck circumference and waist circumference was 37.5 cm and 91.47 cm.

Parameter significance testing consists of two stages: simultaneous test and partial test. In the simultaneous test p-value obtained by $0.0002449016 < \alpha = 0.05$, then $H_0$ rejected so that it can be concluded that the independent variables are used together influence on the risk of OSA. In the partial test in Table 3 that at $\alpha = 0.05$ seen independent variables have a significant effect, namely age, CAT questionnaire 2 and waist circumference.

After performing a significance test parameters, further suitability test will be conducted using a model of Hosmer & Lemeshow test to see whether model fit to the data. In Hosmer & Lemeshow test obtained chi-square value is 3.9829 with p-value of $0.8587 > 0.05$ then $H_0$ is not rejected so it can be concluded that the model used can explain the data.

Based on data analysis using the R software, found that a significantly influential factor is 2 or CAT questionnaire can be categorize as COPD severity and waist circumference bigger than it should. The results of parameter estimates are presented in Table 4.

Based on the results of parameter estimation in Table 2, the logistic regression model equation can be written as follows.

$$\hat{\pi}(\mathbf{x}) = \frac{\exp(-6.59615 + 1.47908x_1 + 0.06409x_2)}{1 + \exp(-6.59615 + 1.47908x_1 + 0.06409x_2)}$$

Based of the equation above, COPD patients with severe degree are 4.39 times greater risk suffer from OSA than mild to moderate COPD patients and each centimetre increase of waist circumference has higher risk of OSA. Accuracy of our model is estimated using classification table with cut point at 0.5 and its accuracy is 73.9%.

### Table 3. Results of Parameter Estimation of Logistic Regression Model

| Variable | Coefficient | p-value |
|----------|-------------|---------|
| (Constant) | -7.77206 | 0.02444 |
| Gender | -1.11772 | 0.09797 |
Body Mass Index (BMI) & 0.02482 & 0.73195 \\
Age & 0.03828 & 0.04052 \\
Smoking History & 0.23752 & 0.70713 \\
mMRC questionnaire & -0.21011 & 0.71167 \\
CAT 1 questionnaire & 0.49683 & 0.40077 \\
CAT 2 questionnaire & 2.02619 & 0.00478 \\
Neck circumference & -0.04305 & 0.70683 \\
Waist circumference & 0.05791 & 0.04443 \\

| Table 4. Results of Logistic Regression Model Parameter Estimation |
|-------------------|-----------------|-----------------|
| Variable & Estimate & p-value |
| (Constant) & -6.59615 & 0.000474 |
| CAT 2 questionnaire & 1.47908 & 0.002308 |
| Waist circumference & 0.06409 & 0.001523 |

4. Conclusion

Based on data analysis, there are 111 patients who have been diagnosed with COPD by a doctor at the RSCM. From the risk factors that has been studied by previous researchers such as age, gender, body mass index (BMI), smoking history, mMRC questionnaires, CAT questionnaires, neck circumference and waist circumference the factors significantly affecting the risk of OSA in COPD patients are waist circumference and CAT 2 questionnaires or patients with severe COPD. From the analysis of the data obtained and the logistic regression model coefficients interpretation that the variables that affect significantly the risk of OSA in patients COPD with weight and waist circumference degree. In this study it was found that a more severe degree of COPD patients at risk of OSA 4.39 times greater than COPD patients with mild to moderate degree, and every increase of 1 cm in waist circumference at risk of OSA.

5. References

[1] Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2017 *Pocket Guide to COPD Diagnosis, Management, and Prevention* (Global Initiative for Chronic Obstructive Lung Disease, Inc.)

[2] World Health Organization 2016 *Chronic Obstructive Pulmonary Diseases* (World Health Organization)

[3] Owens R L and Malhotra A 2010 Sleep-Disordered Breathing and COPD: The Overlap Syndrome *Respiratory Care* 55(10) pp. 1333–1346

[4] American Sleep Association 2017 *Obstructive Sleep Apnea-Symptoms, Causes & Treatments* (America: American Sleep Association)

[5] Marry S M, Lamm B, Lauder I J, Tsang K W T, Chung K, Mok Y, and Lam W 2001 A community study of sleep-disordered breathing in middle-aged Chinese men in Hongkong *Chest Journal* 119(1) pp. 62-69

[6] CV Senaratna 2017 Prevalence of Obstructive Sleep Apnea in the general population: A systematic review *Sleep Medicine Review* 34 pp. 70-81

[7] Soler X, Gao E, Powell F L, Ramsdell J W, Loredo J S, Malhotra A, and Ries A L 2015 High prevalence of Obstructive Sleep Apnea in Patients with Moderate to Severe Chronic
Obstructive Pulmonary Disease *Annals of the American Thoracic Society* 12(8) pp. 1219-1225  
[8] Choi K, Thomas R J, Kim J, Lee S K, Yoon D W, and Shin C 2017 Overlap Syndrome of COPD and OSA in Koreans *Medicine* 96(27) p. 7241  
[9] American Academy of Sleep Medicine 2008 *Obstructive Sleep Apnea* (New York: American Academy of Sleep Medicine)  
[10] Kamelia T 2017 Validation and Realiability Study to Screening Obstructive Sleep Apnea for Indonesian Population. *Sleep Medicine* 40(1) p. 152  
[11] Hosmer D W and Lemeshow S 2013 *Applied Logistic Regression (Third Edition)* (United States of Amerika: John Wiley & Sons, Inc)

**Acknowledgments**

This research is fully supported by PITTA Universitas Indonesia Grant 2018