Prediction model of exacerbations in patients with Chronic Obstructive Pulmonary Disease (COPD) at RSCM

V Amelia¹, T Siswantining¹ and T Kamelia²

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA) Universitas Indonesia, Depok 16424, Indonesia
²Division of Pulmonology, Department of Internal Medicine, Faculty of Medicine, Universitas Indonesia, Jakarta 10430, Indonesia

Corresponding author’s email: titin@sci.ui.ac.id

Abstract. Chronic Obstructive Pulmonary Disease (COPD) is a worldwide health problem. COPD has a tendency for exacerbations. Exacerbations are worsening of acute respiratory symptoms resulting in additional therapy. Exacerbations in COPD increase the risk of death. The objective of this study is to determine the prediction model of exacerbations in patients with COPD based on factors affecting exacerbations in patients with COPD at RSCM (Rumah Sakit Cipto Mangunkusumo). The data used in this study is secondary data from the medical records of patients with COPD in RSCM. The sample was chosen using purposive sampling technique. The samples in this study are 107 patients with COPD. The method used is binary logistic regression analysis. The results of this study indicate that the factors that significantly influence the exacerbations of COPD are breathlessness, history of ICS use, and history of antibiotics use. Appropriate logistic regression model has been obtained. The result indicates that patients with COPD who have breathlessness, have history of ICS use, and have history of antibiotics use are more at risk of exacerbations than those who don’t. Accuracy test has been conducted with classification table at cut point 0.5. The prediction model has an accuracy rate of 74.77 %.

Keywords: COPD, exacerbations, logistic regression

1. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a common, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases [1]. COPD is a worldwide health problem. According to World Health Organization (WHO), globally it is estimated that 3.17 million deaths were caused by COPD in 2015. In Indonesia, the prevalence of COPD is 3.7 % [2]. The state of COPD consists of stable state and exacerbations state. COPD exacerbations are defined as an acute worsening of respiratory symptoms that result in additional therapy [1]. The worsening of the previous condition is characterized by increase in dyspnea (shortness of breath), increase in the amount of sputum, and change in the color of sputum to yellow or green.

Previous studies of COPD exacerbations have been done. A study by Make et al in 2015 suggested that the factors associated to short-term (6 months) risk of exacerbations in COPD were COPD maintenance medications, daily reliever use, the number of exacerbations during previous year, the ratio of forced expiratory volume in 1 second (FEV1) to forced vital capacity (FVC), and gender [3].
These factors were used to develop a score called SCOPEX. Meanwhile, a study by Steer et al in 2012 suggested that the factors that most significantly affect the mortality in patients hospitalised with an exacerbation of COPD were dyspnea, eosinopenia, consolidation, acidaemia, and atrial fibrillation [4]. These factors were combined to form a score called DECAF score.

The characteristics of patients with COPD in foreign countries compared to Indonesia are suspected to be different. That’s why this research was done in Indonesia, especially at Cipto Mangunkusumo Hospital (RSCM). RSCM is chosen because it is a national referral hospital that not only function on health services but also function on research and education in health.

The problem in this study is how to determine the prediction model of exacerbations in patients with COPD based on factors affecting exacerbations in patients with COPD at RSCM. So, the objective of this study is to determine the prediction model of exacerbations in patients with COPD based on factors affecting exacerbations in patients with COPD at RSCM.

The scope of problem in this study:
1. This study uses medical records of RSCM from 2010 to 2017.
2. The patients studied were COPD patients who are inpatient or outpatient at pulmonology clinic of RSCM.

2. Variables in the study
The variables in this study consist of dependent variable and independent variables. The dependent variable is exacerbation, defined as worsening of acute respiratory symptoms resulting in additional therapy. This variable consists of two categories, exacerbation and not exacerbation.

The independent variables in this study are selected based on previous studies. They are as follows:
(a). Gender. This variable is defined as the gender of COPD patients. It consists of two categories, female and male.
(b). Age. This variable is defined as the life duration of COPD patients counted from the patient’s birth date until the patient is admitted to the hospital due to COPD, expressed in years.
(c). Body Mass Index (BMI). This variable is defined as an assessment of adult nutritional status that is measured by weight in kilograms (kg) divided by the square of the height in metres (kg/m²). Low BMI is associated with high mortality in COPD patients [5].
(d). Smoking history. This variable is defined as a habit of smoking cigarettes in patients with COPD. It consists of two categories, smoking and not smoking. Around 20–25 % of smokers are at risk of developing COPD or lung cancer [6].
(e). Eosinophil level. This variable is defined as the percentage of eosinophils in the blood of patients with COPD based on the blood test results.
(f). Neutrophil level. This variable is defined as the percentage of neutrophils in the blood of patients with COPD based on the blood test results.
(g). pCO2. This variable is defined as the partial pressure of carbon dioxide in the blood of patients with COPD based on blood gas analysis. It is expressed in mmHg.
(h). pO2. This variable is defined as the partial pressure of oxygen in the blood of patients with COPD based on blood gas analysis. It is expressed in mmHg.
(i). Oxygen saturation. This variable is defined as the percentage of oxygen molecules that has bonded with haemoglobin in patients with COPD.
(j). Breathlessness. This variable is defined as the condition of shortness of breath or difficulty breathing. It consists of two categories, breathless and not breathless.
(k). Sputum production. This variable is defined as a condition when COPD patient’s airway is filled with phlegm or mucus. It consists of two categories, producing sputum and not producing sputum.
(l). History of bronchodilator use. Bronchodilator is medicine that can dilate the bronchi and bronchioles, decrease resistance in the respiratory airway, and increase airflow to the lungs. This variable consists of two categories, having history and not having history.
(m). History of antimuscarinic use. Antimuscarinic is medicine that can reduce mucus secretion. This variable consists of two categories, having history and not having history.

(n). History of ICS (Inhaled Corticosteroid) use. This variable is defined as the history of using corticosteroid in inhalation form. This variable consists of two categories, having history and not having history. A study suggests that COPD patients treated with ICS experience significantly fewer exacerbations than patients on placebo [7].

(o). History of antibiotics use. Antibiotics are medicine that can destroy or slow down the growth of bacteria. This variable consists of two categories, having history and not having history.

(p). Cardiovascular-related comorbidities. This variable is defined as comorbidities associated with the heart and blood vessels that are suffered by patients with COPD. It consists of two categories, having and not having.

(q). Respiratory-related comorbidities. This variable is defined as comorbidities associated with the respiratory that are suffered by patients with COPD. It consists of two categories, having and not having.

3. Experimental method
Population in this study is patients with COPD at RSCM who were listed on the medical records in 2010-2017. The sample is 107 COPD patients who are selected using purposive sampling. The data used is secondary data from the medical records of COPD patients at RSCM from 2010 to 2017. The method for data analysis is binary logistic regression.

3.1. Binary logistic regression model
Regression analysis is a statistical technique for investigating and modeling the relationship between variables [8]. One of the cases of regression is binary logistic regression. Binary logistic regression is an approach to analyze the relationship between binary (dichotomous) dependent variable with numerical or categorical independent variables. The dependent variable has only two possible outcomes, generically called success and failure and denoted by 0 and 1. Binary logistic regression model has the form:

\[
\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi})}
\]  \hspace{1cm} (1)

where \(\pi(x_i)\) is the probability of “success”, \(\beta_0, \beta_1, \beta_2, \ldots, \beta_p\) are unknown parameters, \(x_i = (x_{1i}, x_{2i}, \ldots, x_{pi})\) is the set of independent variables, \(i = 1, 2, \ldots, n\), and \(n\) is the number of observations.

If some of the independent variables are nominal, it is inappropriate to include them in the model as if they were interval scale variables. In this situation, the method of choice is to use design variables (dummy variables) [9]. In general, if a nominal scaled variable has \(k\) possible values, then \(k - 1\) dummy variables will be needed. Suppose that the \(j^{th}\) independent variable \(x_j\) has \(k_j\) levels. The \(k_j - 1\) dummy variables will be denoted as \(D_{jl}\) and the coefficients for these dummy variables will be denoted as \(\beta_{jl}\), \(l = 1, 2, \ldots, k_j - 1\). Thus, the model with \(p\) variables and the \(j^{th}\) variable being discrete would be:

\[
g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_p x_p
\]  \hspace{1cm} (2)

3.2. Parameter estimation
The unknown parameters \(\beta' = (\beta_0, \beta_1, \ldots, \beta_p)\) should be estimated to form a logistic regression model. The method to estimate parameters in logistic regression is the Maximum Likelihood method.
This method is applied by looking for the parameters $\beta' = (\beta_0, \beta_1, ..., \beta_p)$ that maximize likelihood function. Each observation follows the Bernoulli distribution. Since each observation is assumed to be independent, the likelihood function is obtained as the product of the probability function of each observation as follows:

$$l(\beta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1 - y_i}$$  \hspace{1cm} (3)

It is easier mathematically to maximized the likelihood function in the form of logarithms. This expression, called log likelihood, is defined as follows:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} [y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]]$$  \hspace{1cm} (4)

The value of $\beta$ that maximizes $L(\beta)$ can be found by differentiating $L(\beta)$ with respect to parameters $\beta_0, \beta_1, ..., \beta_p$ and setting the resulting expressions equal to zero. These equations are:

$$\sum_{i=1}^{n} [y_i - \pi(x_i)] = 0$$  \hspace{1cm} (5)

and

$$\sum_{i=1}^{n} x_{ij} [y_i - \pi(x_i)] = 0$$  \hspace{1cm} (6)

for $j = 1, 2, ..., p$. The equation 5 and equation 6 are nonlinear in parameters $\beta_0, \beta_1, ..., \beta_p$, so it is necessary to solve it using iteration method.

### 3.3. Testing for significance of parameters

Testing for the significance of the parameters is performed to determine whether the independent variables in the model significantly affect the dependent variable. This test is performed simultaneously and partially. The simultaneous test is performed to know the significance of parameters $\beta$ to dependent variable simultaneously. The hypotheses are $H_0$: $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ and $H_1$: at least one $\beta_i \neq 0$, $i = 1, 2, ..., p$. The test statistic is $G$, formulated as $G = -2\ln[L_1/L_0]$, where $L_1$ is likelihood without independent variable and $L_0$ is likelihood with independent variables.

The partial test is performed to know the effect of each $\beta_i$ individually. The hypotheses are $H_0$: $\beta_i = 0$ and $H_1$: $\beta_i \neq 0$; $i = 1, 2, ..., p$. The test statistic is Wald ($W$), defined as $W = \hat{\beta}_i/S\dot{E}(\hat{\beta}_i)$, where $\hat{\beta}_i$ is estimator of $\beta_i$ and $S\dot{E}(\hat{\beta}_i)$ is standard error for $\beta_i$.

### 3.4. Goodness-of-fit test

Assessing the goodness-of-fit of the model is performed to know how effective the model describes the dependent variable. This test is performed using Hosmer and Lemeshow test. The hypotheses are $H_0$: Model fits the data and $H_1$: Model doesn’t fit the data. The test statistic is $\hat{C}$, formulated as $\hat{C} = \sum_{k=1}^{q}(o_k - n_k \bar{\pi}_k)^2/n_k \bar{\pi}_k(1 - \bar{\pi}_k)$.

### 3.5. Interpretation of the coefficients

The coefficients in logistic regression model are interpreted to find out the meaning of parameter estimation on independent variables. For categorical variables, the parameters are interpreted using odds ratio. Odds is the proportion of probability of a “success” with “failure” in a category. The odds ratio (OR) is defined as the ratio of the odds for $x = 1$ with the odds for $x = 0$. The odds ratio is expressed by $OR = e^{\hat{\beta}_1}$. The interpretation of this odds ratio is that the tendency for $y = 1$ at $x = 1$ is $e^{\hat{\beta}_1}$ times
compared to $x = 0$. For numerical variables, the interpretation of $\beta_i$ coefficients is that everytime the independent variable increase as much as $c$ unit, the risk of $y = 1$ is $e^{\beta_i}$ times greater.

3.6. Classification table
The classification table is used to determine the accuracy of prediction model. This table is the result of cross-classifying the dependent variable with a dichotomous variable whose values are derived from the estimated logistic probabilities $\pi(x_i)$ [9]. The derived dichotomous variable is obtained by defining a cutpoint $c$, $0 \leq c \leq 1$, and compare each estimated probability to $c$. If the estimated probability exceeds $c$, then let the derived variable be equal to 1; otherwise it is equal to 0. The most commonly used value for $c$ is 0.5. The classification table is presented in table 1.

Two useful summaries of predictive power are

sensitivity = $P(\hat{y} = 1 | y = 1)$, specificity = $P(\hat{y} = 0 | y = 0)$ (7)

Another summary of predictive power from the classification table is the overall proportion of correct classifications. This estimates

\[
P(\text{correct classification}) = P(y = 1 \text{ and } \hat{y} = 1) + P(y = 0 \text{ and } \hat{y} = 0) = P(\hat{y} = 1 | y = 1)P(y = 1) + P(\hat{y} = 0 | y = 0)P(y = 0)
\]

which is a weighted average of sensitivity and specificity [10].

4. Results and discussion
The descriptive analysis show that majority of patients with COPD at RSCM experience exacerbations, are male, have smoking history, experience breathlessness, produce sputum, have history of bronchodilators use, have history of antimuscarinic use, have history of ICS use, have history of antibiotics use, have cardiovascular-related comorbidities, and have respiratory-related commorbidities. In addition, on average, they are 64 years old, have normal weight, have normal eosinophil level, have high neutrophil level, have low pCO$_2$ value, have high pO$_2$ value, and have low oxygen saturation.

The data obtained is analyzed using binary logistic regression analysis with R software version 3.3.3. In this study, there are 10 independent variables used as dummy variables because they are categorical. Those are gender, smoking history, breathlessness, sputum production, history of bronchodilator use, history of antimuscarinic use, history of ICS use, history of antibiotics use, cardiovascular-related comorbidities, and respiratory-related comorbidities. Each of these variables has 2 categories, so the dummy variable required is $2 - 1 = 1$ each. Furthermore, the parameters in the logistic regression are estimated using the maximum likelihood method. The result is represent in table 2.

| Table 1. Classification table based on the logistic regression model using a cutpoint $c$. |
|-----------------------------------------------|
| Predicted | Observed | Total |
| $\hat{y} = 1$ | $y = 1$ | $n_{11}$ | $n_{10}$ | $n_1$ |
| $\hat{y} = 0$ | $y = 0$ | $n_{01}$ | $n_{00}$ | $n_0$ |
| Total | | $n_{1}$ | $n_{0}$ | $n$ |

$n_{11}$ : the number of observations from $y = 1$ that is precisely classified as $\hat{y} = 1$

$n_{10}$ : the number of observations from $y = 0$ that is misclassified as $\hat{y} = 1$

$n_{01}$ : the number of observations from $y = 1$ that is misclassified as $\hat{y} = 0$

$n_{00}$ : the number of observations from $y = 0$ that is precisely classified as $\hat{y} = 0$

$n$ : total overall observations
Table 2. Parameter estimation of logistic regression model using R software.

| Variables                  | Coefficients | p-value |
|----------------------------|--------------|---------|
| (Constant)                 | -15.85841    | 0.00823 |
| Gender                     | -0.23067     | 0.80766 |
| Age                        | 0.01063      | 0.66459 |
| Body Mass Index (BMI)      | 0.04071      | 0.49002 |
| Smoking history            | 1.41382      | 0.17073 |
| Eosinophil level           | 0.01246      | 0.90593 |
| Neutrophil level           | 0.04016      | 0.17129 |
| pCO2                       | 0.02633      | 0.17277 |
| pO2                        | -0.01364     | 0.06633 |
| Oxygen saturation          | 0.05797      | 0.21561 |
| Breathlessness             | 2.51255      | 0.03795 |
| Sputum production          | 0.93698      | 0.12242 |
| History of bronchodilator use | 0.05451   | 0.95904 |
| History of antimuscarinic use | 0.95014   | 0.17412 |
| History of ICS use         | 1.28260      | 0.04410 |
| History of antibiotics use | 1.56011      | 0.02098 |
| Cardiovascular-related commorbidities | 0.01774 | 0.97672 |
| Respiratory-related commorbidities | -0.78112  | 0.20155 |

Table 3. Parameter estimation of logistic regression model with significant independent variables using R software.

| Variables                  | Coefficients | p-value |
|----------------------------|--------------|---------|
| (Constant)                 | -4.6430      | 0.000279|
| Breathlessness             | 2.7922       | 0.012007|
| History of ICS use         | 1.3349       | 0.004470|
| History of antibiotics use | 1.8007       | 0.001515|

Testing for the significance of parameters simultaneously is performed to know the significance of parameters $\beta$ to dependent variable simultaneously. At $\alpha = 0.05$, the p-value is $5.359945 \times 10^{-5}$. The p-value < $\alpha$, then $H_0$ is rejected so the independent variables simultaneously affect the exacerbations.

Testing for the significance of parameters partially is performed to find out which independent variables are worth entering into the model. The results of this test are showed in table 2. At $\alpha = 0.05$, the independent variables that have significant effect, those with p-value < 0.05, are breathlessness, history of ICS (Inhaled Corticosteroid) use, and history of antibiotics use.

The logistic regression model will be formed using these three significant variables. The parameter estimation is presented in table 3.

Based on table 3, the logistic regression model equation can be written as follows:

$$
\hat{p}(x) = \frac{\exp(-4.6430 + 2.7922x_1 + 1.3349x_2 + 1.8007x_3)}{1 + \exp(-4.6430 + 2.7922x_1 + 1.3349x_2 + 1.8007x_3)}
$$

(9)
where \( x_1 \) represents breathlessness variable, \( x_2 \) represents history of ICS use variable, and \( x_3 \) represents history of antibiotics use variable.

The goodness-of-fit test of the model is performed to determine whether there is a difference between the observations with the predictions. Based on the result, the chi-square value is 6.144 with p-value is 0.6311. Because p-value \( > \alpha = 0.05 \), \( H_0 \) is not rejected. So, it can be concluded that the model fits the data.

Since all significant independent variables are dichotomous, the interpretation uses odds ratio as shown in table 4.

Based on the odds ratio in table 4, it can be interpreted as follows:

(a). Breathlessness

The odds ratio for breathlessness is 16.32. It means that the risk of exacerbations of the patients with COPD who have breathlessness is 16.32 times greater than the risk of exacerbations of patients with COPD who don’t have breathlessness.

(b). History of ICS use

The odds ratio for history of ICS use is 3.8. It means that the risk of exacerbations of the patients with COPD who have history of ICS use is 3.8 times greater than the risk of exacerbations of patients with COPD who don’t have history of ICS use.

(c). History of antibiotics use

The odds ratio for history of antibiotics use is 6.05. It means that the risk of exacerbations of the patients with COPD who have history of antibiotics use is 6.05 times greater than the risk of exacerbations of patients with COPD who don’t have history of antibiotics use.

The logistic regression model is used to classify COPD patients based on the prediction of exacerbations. The accuracy of logistic regression model in classification is calculated using classification table with cutpoint \( c = 0.5 \). The classification table is shown in table 5.

Sensitivity and specificity are calculated as follows:

\[
\text{Sensitivity} = \frac{40}{58} = 68.97\%
\]

\[
\text{Specificity} = \frac{40}{49} = 81.63\%
\]

| Table 4. Odds ratios for significant variables. |
|-----------------------------------------------|
| Variables                                 | Odds ratio       |
|-----------------------------------------------|
| Breathlessness                                  | 16.316542177 |
| History of ICS use                             | 3.799601883    |
| History of antibiotics use                     | 6.054075137    |

| Table 5. Classification table based on logistic regression model with cutpoint 0.5. |
|-----------------------------------------------|
| Predicted | Observed | Total |
|-----------------------------------------------|
| \( \hat{y} = 1 \) (Exacerbation) | \( y = 1 \) (Exacerbation) | 40 | 9 | 49 |
| \( \hat{y} = 0 \) (Not exacerbation) | \( y = 0 \) (Not exacerbation) | 18 | 40 | 58 |
| Total |                                               | 58 | 49 | 107 |
The percentage of classification accuracy is calculated as the ratio between the number of observations classified precisely by the total number of observations. The percentage of classification accuracy is calculated as follows:

\[
\text{Percentage of classification accuracy} = \frac{40 + 40}{107} \times 100\% = 74.77\%
\]

Based on the logistic regression model and interpretation of odds ratio, it is found that patients with COPD who have breathlessness, history of ICS use, and history of antibiotics use are more at risk of exacerbations than those who don’t. Most breathlessness indicates that the condition of lung is getting worse. The worsening condition of the lungs indicates the occurrence of exacerbations. A study suggests that the rate of breathlessness affects mortality in COPD exacerbations [4]. Patients with COPD exacerbations who have higher degrees of breathlessness have higher risk of mortality as well.

Corticosteroids are medicine that are commonly used to relieve symptoms of swelling, itching, and allergic reactions. ICS (Inhaled Corticosteroid) is a corticosteroid used in inhalation form. It is possible that use of ICS comes with an increased risk for exacerbations of COPD [11]. Patients with COPD use ICS because their COPD symptoms worsen. The worsening of COPD symptoms leads to exacerbations.

Antibiotics are recommended for patients whose exacerbations symptoms indicate bacterial infections, that is increased sputum volume, sputum purulence, and increased breathlessness [12]. Unnecessary use of antibiotics, such as in patients who don’t have increased purulent sputum, should be avoided because overuse of antibiotics may cause resistance, so the antibiotics don’t work to overcome the infection resulting in exacerbation.

5. Conclusion
This section contains conclusion based on the research and suggestions for further research. The limitations of this study are that the amount of sample is small, that is only 107 patients and the sample is selected using non-probability sampling technique. In addition, this study don’t involve the results of spirometry test as an independent variable due to inadequate data. Thus, based on the results of data analysis using binary logistic regression method, the factors that significantly influence the exacerbations of COPD are breathlessness, history of ICS use, and history of antibiotics use. The prediction model of exacerbations in patients with COPD based on factors affecting exacerbations in patients with COPD at RSCM has been obtained. The result indicates that patients with COPD who have breathlessness, have history of ICS use, and have history of antibiotics use are more at risk of exacerbations than those who don’t. The accuracy of the model is calculated using classification table at cutpoint 0.5. The result shows that the accuracy rate of the prediction model is 74.77 %, the specificity is 81.63 %, and the sensitivity is 68.97 %.

Based on the limitations of this research, the suggestions for the development of this research are:
1. The next similar research is expected to use a larger number of sample so the data can describe the real situation better.
2. The next similar research is expected to consider other factors that may affect exacerbations in patients with COPD, such as the spirometry test.

References
[1] Global Initiative for Chronic Obstructive Lung Disease 2017 Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease 2018 Report. Global Initiative for Chronic Obstructive Lung Disease Inc. available at https://goldcopd.org/wp-content/uploads/2017/11/GOLD-2018-v6.0-FINAL-revised-20-Nov_WMS.pdf
[2] Kementrian Kesehatan RI 2013 Riset Kesehatan Dasar 2013 (Jakarta: Badan Penelitian dan Pengembangan Kesehatan, Kementrian Kesehatan RI).
[3] Make B J, Eriksson G, Calverley P M, Jenkins C R, Postma D S, Peterson S, Oslund O and Anzueto A 2015 *Int. J. Chron. Obstruct. Pulmon. Dis.* 10 201-9.

[4] Steer J, Gibson J and Bourke S C 2012 *Thorax* 67 970-6.

[5] Yamauchi Y, Hasegawa W, Yasunaga H, Sunhara M, Jo T, Takami K, Matsui H, Fushimi K and Nagase T 2014 *Int. J. Chron. Obstruct. Pulmon. Dis.* 9 1337-46.

[6] Perhimpunan Dokter Paru Indonesia 2003 *Penyakit Paru Obstruktif Kronik: Diagnosis dan Penatalaksanaan* (Jakarta: Perhimpunan Dokter Paru Indonesia).

[7] Gartlehner G, Hansen R A, Carson S S and Lohr K N 2006 *Annals of Family Medicine* 4 253-62.

[8] Montgomery D C, Peck E A and Vining G G 2012 *Introduction to Linear Regression Analysis* 5th edition (USA: John Wiley & Sons, Inc.).

[9] Hosmer D W and Lemeshow S 2000 *Applied Logistic Regression 2nd Edition* (USA: John Wiley & Sons, Inc.).

[10] Alan A 2007 *An Introduction to Categorical Data Analysis 2nd Edition* (USA: John Wiley & Sons, Inc.) pp 142-143.

[11] Husebo G R, Bakke P S, Aanerund M, Hardie J A, Ueland T, Grenseth R, Persson L J P, Aukrust P and Eagan T M 2014 *PLoS ONE* 9 e109721.

[12] Halpin D M G, Miravitlles M, Metzdorf N and Celli B 2017 *Int. J. Chron. Obstruct. Pulmon. Dis.* 12 2891-908.