Predicting readmission risk after coronary artery bypass graft surgery using logistic regression model

V Febriani, D Lestari, S Mardiyati and S Devila
Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

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

Abstract. Coronary Artery Bypass Graft (CABG) post-operative readmission has a high incidence rate compared to other health cases. This particular case contributes to an increase in morbidity and hospital costs of patients. Therefore, an appropriate prediction model is needed while the model can be beneficial to the health financing institutions. There are many risk factors that will be used to predict CABG post-operative readmission. Of the many risk factors observed, some factors that have a significant influence on constructing the Logistic Regression model will be determined. This model is developed to generate probabilities which are then called Created Readmission Risk Scores (CRRS).

Keywords: Morbidity, risk factors, wald tests, weighted least square

1. Introduction
Coronary heart disease is a disorder of the heart caused by plaque that builds up in the arteries, thereby obstructing the oxygen supply to the heart. This disease is one of the most common cardiovascular diseases [1]. According to the World Health Organization (WHO), in 2015, 70% of all causes of death from heart disease are due to coronary heart disease (CHD) in 2015. Coronary heart disease has resulted in 17.5 million deaths, equivalent to 30.0% of total deaths worldwide [2]. Risk factors for CHD are determined through the interaction of two or more risk factors, namely factors that cannot be controlled (non-modifiable risk factors) such as heredity, age, and sex, and factors that can be controlled (modifiable risk factors) such as high blood pressure, Diabetes Mellitus, obesity, Dyslipidemia [3].

One procedure that can be used to cure coronary heart disease is Coronary Artery Bypass Graft (CABG). CABG is done by opening the chest wall through cutting the Sternum bone and then installing new blood vessels taken from the saphenous vein, mammaria interna or radialis artery. The process aims to overcome the oxygen supply bottlenecks that occur in the heart [4]. After undergoing CABG, patients will be faced with 2 conditions: recovery or complications that cause the patient to be hospitalized again (readmission). Readmission is a serious concern because it is directly related to the amount of costs that patients must incur in the hospital [5]. Research related to post-CABG readmission was conducted in California by Li et al. and concluded that the most significant factor was congestive heart failure [6]. In 2014 Lancey R et al. in New York obtained 6 significant factors: chronic obstructive pulmonary disease, Diabetes Mellitus, cerebrovascular disease, congestive heart failure, intra-aortic balloon pump, and blood products [7]. A similar research
conducted in Argentina by Espinoza et al. produced 5 significant factors such as Diabetes Mellitus, hematocrit before surgery, cardiopulmonary bypass, highest serum glycemic level and postoperative atrial fibrillation [8]. Therefore, this research aims to build a model that can predict how much risk a patient will return to be treated after undergoing CABG surgery. In this research, the prediction model that will be used is the logistic regression model.

2. Methodology

2.1. Logistic regression model

Logistic regression is a regression model used to find the best fitting model to describe the relationship between the binary characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables, where the response variable has only two possibilities: success or failure, expressed as 0 (fail) and 1 (succeed). The following are the general logistic regression equations [9]:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \]  
\[ Y_i = x_i^T \beta + \epsilon_i \]

where \( i = 1, 2, 3, ..., n \), and \( n \) is the number of observations, \( x_i^T = [1, x_{i1}, x_{i2}, ..., x_{ik}] \) and \( \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \).

\( Y_i \) is a response variable and \( Y_i \) is assumed to be Bernoulli distribution. \( Y_i \) has the following probability distribution:

Table 1 shows that the probability of success is stated as 1 and the probability of failure is stated as 0. The probability value is \( \pi_i \) for success and \( 1 - \pi_i \) for failure. Assuming \( E(\epsilon_i) = 0 \) the expected value for the response variable is obtained: \( E(Y_i) = (1 - \pi_i)0 + \pi_i \cdot 1 = \pi_i \). So

\[ E(Y_i) = E(x_i^T \beta + \epsilon_i) = E(x_i^T \beta) + E(\epsilon_i) = x_i^T \beta \]

\( Y \) is a binary response variable and the \( E(Y_i) \) has a form limit, then

\[ 0 \leq E(Y_i) = \pi_i \leq 1 \]

In logistic regression, logit transformation is defined as

\[ x_i^T \beta = \ln \frac{\pi_i}{1 - \pi_i} \]  
\[ \pi_i = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \]

Thus, the relationship between \( \pi_i \) and \( x_i^T \beta \) is defined as [10]:

| \( Y_i \) | Probability |
|---|---|
| 1 | \( P(Y_i = 1) = \pi_i \) |
| 0 | \( P(Y_i = 0) = 1 - \pi_i \) |
To estimate the parameters $\beta_0, \beta_1, \beta_2, ..., \beta_k$ in equation 4, the Weighted Least Square (WLS) method has the following form of equation [11]:

$$\hat{\beta} = (X^TWX)^{-1}X^TWy$$  \hspace{1cm} (5)

where $\hat{\beta}$ is a vector whose entries contain an estimated value of the parameter coefficient $\beta_i$, $U$ is a diagonal matrix with entries $u_{ii}$ on the main diagonal, where $u_{ii} = n_i \pi_i(x_i)(1-\pi_i(x_i))$ for $i = 1,2, ..., k$.

After obtaining the parameter values $\hat{\beta}_j$, $j = 0,1,2, ..., k$, Wald tests are then performed to determine the effect of each independent variable individually on the dependent variable. The results of this test indicate whether it is feasible for a free variable to enter the logistic regression model [10].

The hypothesis used in this test is as follows:

$H_0 = \text{There is no influence of risk factors on patient readmission.}$

$H_1 = \text{There is an influence of risk factors on patient readmission.}$

Test statistics used in this test are:

$$W = \frac{\beta_j}{SE(\beta_j)}$$  \hspace{1cm} (6)

where $W$ is Wald Statistics, $\beta_j$ is the parameter coefficient to $j$ and $SE(\beta_j)$ is the Standard Error from the parameter coefficient to $j$. With $\alpha = 0.05$, the decision rule used in this test is to reject $H_0$ if $W > Z_{\alpha/2}$ or p-value < $\alpha$.

3. Results and discussion

The data used in this study came from the European Society of Cardiology (ESC) guidelines recorded in 2010–2014. 388 patients were observed. Of these, 82 patients had complications after CABG; the remaining 306 patients had no complications after CABG. The following will show some data from these patients, where the number 0 in the complications column means that there are no postoperative complications while number 1 shows the opposite.

Table 2 shows that there were 18 risk factors observed, namely age, sex, weight, hypertension (HA), diabetes mellitus (DM), hyperlipidemia, chronic obstructive pulmonary disease (COPD), peripheral arterial disease (PAD), stroke, Canadian Cardiovascular Society (CCS), congestive heart failure (CHF) displayed as NYHA in data, myocardial infarction (MI), posterior cruciate ligament (PCL), chronic kidney disease (CKD), atrial fibrillation (AF), vessel diseases, left main coronary artery (LM), and ejection fraction (EF). Readmission served as the response variable.

In this research, the observed risk factors are used to build a logistic regression model. Using R software, the coefficient values (Estimate column) for each risk factor are showed in figure 1.

The results in figure 1 not only display the estimated coefficient $\hat{\beta}$ for the prediction model, but also the significance of the effect of each risk factor. From the value Pr ($>|z|>$) smaller than 5 %, we can see that four risk factors influence readmission after CABG, namely sex, weight, NYHA, and EF. The logistic regression model can therefore be rebuilt using only these four risk factors. In figure 2, we can see the results of logistic regression for four significant factors: gender, weight, the type of heart failure experienced by patients and ejection fraction.

Figure 2 shows the results of logistic regression with 4 selected significant risk factors, where $\hat{\gamma}$ is the probability of the occurrence of postoperative CABG readmission, $x_1$ is the sex of the patient, $x_2$ is the patient’s body weight, $x_3$ is the heart failure experienced by the patient and $x_4$ is the patient’s Ejection Fraction. The resulting logistic regression model is $\hat{\gamma} = 0.48916 + 0.90424x_1 - 0.02884x_2 + 0.28389x_3 - 0.0208x_4$. This model can determine the chances of readmission in patients, as shown in table 3.
### Table 2. Data of some patients after CABG surgery.

| No | Age | Sex | Weight | HA | DM | Hyperlipidemia | COPD | PAD | Stroke |
|----|-----|-----|--------|----|----|----------------|------|-----|--------|
| 1  | 86  | 1   | 61     | 1  | 1  | 0              | 0    | 0   | 0      |
| 2  | 85  | 1   | 84     | 1  | 0  | 0              | 0    | 1   | 1      |
| 3  | 87  | 1   | 75     | 1  | 1  | 0              | 0    | 0   | 0      |
| 4  | 87  | 1   | 95     | 1  | 0  | 1              | 0    | 0   | 1      |
| 5  | 84  | 1   | 52     | 0  | 0  | 0              | 1    | 0   | 0      |

| No | CCS | NYHA | MI | PCI | CKD | AF | DV | LM | EF | Complications |
|----|-----|------|----|-----|-----|----|-----|----|----|---------------|
| 1  | 3   | 3    | 1  | 1   | 0   | 1  | 3   | 0  | 36 | 0            |
| 2  | 2   | 2    | 1  | 0   | 1   | 0  | 3   | 1  | 48 | 1            |
| 3  | 3   | 3    | 1  | 1   | 1   | 1  | 3   | 0  | 40 | 0            |
| 4  | 4   | 3    | 1  | 0   | 0   | 0  | 3   | 0  | 50 | 0            |
| 5  | 3   | 3    | 2  | 0   | 0   | 0  | 3   | 1  | 18 | 1            |

```r
# call:
glm(formula = complications ~ Age + Sex + weight + MA + DM + 
    Hyperlipidemia + COPD + PAD + Stroke + CCS + NYHA + MI + 
    PCI + CKD + AF + DV + LM + EF, family = binomial(link = "logit"), 
    data = datapasten)
```

Deviance Residuals:
Min      IQ 1Q  Median    3Q    Max
-1.3525  -0.7086 -0.3499  -0.3413  2.4423

Coefficients:

|                | Estimate | Std. Error | z value | Pr(>|z|)   |
|----------------|----------|------------|---------|------------|
| (Intercept)    | -0.08689 | 0.08141    | -1.076  | 0.2837     |
| Age            | 0.01499  | 0.00990    | 1.499   | 0.1361     |
| Sex            | 0.88723  | 0.32211    | 2.770   | 0.0055 **  |
| Weight         | -0.03754 | 0.03137    | -1.197  | 0.2331     |
| HA             | 0.33314  | 0.38602    | 0.860   | 0.3891     |
| DM             | 0.13770  | 0.30832    | 0.447   | 0.6513     |
| Hyperlipidemia | -0.28834 | 0.41259    | -0.710  | 0.4756     |
| COPD           | 0.19000  | 0.30444    | 0.624   | 0.5343     |
| PCI            | 0.10807  | 0.38478    | 0.280   | 0.7770     |
| CKD            | 0.45938  | 0.32070    | 1.432   | 0.1520     |
| AF             | 0.59362  | 0.36499    | 1.626   | 0.1038     |
| DV             | 0.04833  | 0.16865    | 0.281   | 0.7744     |
| LM             | -0.06908 | 0.28257    | -0.244  | 0.8068     |
| EF             | -0.01583 | 0.01106    | -1.702  | 0.0884 .   |

***Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

**Figure 1.** Logistic regression results with 18 risk factors.

Table 3 shows, patient number 1 with male sex, body weight 61 kg, NYHA 3 and Ejection Fraction 36. This patient is at risk of being treated again with 43 % chance. Readmission risk is obtained by using equation 4.
Figure 2. Logistic Regression Results with 4 significant risk factors.

Table 3. Opportunities for patient readmission after CABG surgery.

| No | Sex | Weight | NYHA | EF | Readmission risk |
|----|-----|--------|------|----|------------------|
| 1  | 1   | 61     | 3    | 36 | 43 %             |
| 2  | 1   | 84     | 2    | 48 | 19 %             |
| 3  | 1   | 75     | 3    | 40 | 32 %             |
| 4  | 1   | 95     | 3    | 50 | 18 %             |
| 5  | 1   | 52     | 3    | 18 | 59 %             |

4. Conclusion
The risk of readmission in patients after undergoing a coronary artery bypass graft (CABG) can be predicted by analyzing patient health data. From these data, logistic regression can show the most significant factors, so that we can build a model that can ultimately identify high-risk patients.

Acknowledgments
This work was financially supported by Universitas Indonesia under research grant PITTA with grant contract number NKB-0669/UN2.R3.1/HKP.05.00/2019).

References
[1] National Institute of Health (NIH) 2018 About Health Topic available at https://www.nhlbi.nih.gov/health-topics
[2] World Health Organization 2017 World Health Organization Report: Cardiovascular Disease available at https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[3] World Heart Federation 2017 Cardiovascular Risk Factors available from https://www.world-heart-federation.org/resources/risk-factors
[4] Montaye M D, Bacquer D G and Amouye P 2000 Eur. Heart J. 21 808-13
[5] Jenkins S F, William M V and Coleman E A 2009 N. Engl. J. Med. 14 1418-28
[6] Li Z, Armstrong E J and Parker J P 2012 Circ. Cardiovasc. Qual. Outcomes 5 729-37
[7] Lancey R et al. 2015 J. Thorac Cardiovasc. Surg. 3 950-7
[8] Espinoza J, Camporrontondo M, Vrancic M, Piccinini F, Camou J, Benzadon M and Navia D 2016 *Clin. Trials Regul. Sci. Cardiol.* **20** 1-5

[9] Mendenhall W, Beaver R J and Beaver B M 2006 *Introduction to Probability and Statistics* 12th edition (Duxbury: Thomson Brooks/Cole)

[10] Hosmer D W and Lemeshow S 2000 *Applied Logistic Regression* (United States of America: John Wiley and Sons, Inc.)

[11] Grill J 2001 *Generalized Linear Model: A Unified Approach* (London: SAGE)