Analysis of the effect of smoking and exercise habits on coronary heart disease in Indonesia with logistic regression

R Amalia1 and Kismiantini2

1Department of Mathematics, Universitas Negeri Yogyakarta, Jl Colombo No 1, Karangmalang, Depok, Sleman, Yogyakarta, Indonesia

2Department of Statistics, Universitas Negeri Yogyakarta, Jl Colombo No 1, Karangmalang, Depok, Sleman, Yogyakarta, Indonesia

E-mail: [rahma5156fmipa2016@student.uny.ac.id](mailto:rahma5156fmipa2016@student.uny.ac.id); kismi@uny.ac.id

**Abstract**. The high number of people with coronary heart disease (CHD) is comparable to the high level of smoking in Indonesia. While Indonesia’s lack of physical activity pattern, based on Basic Health Research (Riskesdas) results in 2018, rose from 26.1% to 33.5%. The incidence of CHD in Indonesia is potentially increasing. The study aims to analyze the effects of smoking and exercise habits on the incidence of CHD using data from the fifth wave of Indonesian Family Life Survey (IFLS 5). The binary logistic regression model has been used to explain the factors that affect whether or not a respondent has CHD. The results showed that the smoking habit, age, hypertension, and high cholesterol were significant predictors for the incidence of CHD (yes or no) in Indonesia. The exercise habits, gender and education were not significant predictors for the incidence of CHD in Indonesia. The main result found that the CHD in Indonesia is 1.898 times more likely to occur among people who smoke than non-smokers.

1. Introduction

World mortality in 2016 indicated that 71% were caused by non-communicable diseases (NCDs), specifically 40.4 million deaths out of 56.9 million [18]. Of all the death caused by NCDs, 44% were caused by cardiovascular disease, that is 17.9 million deaths out of 40.4 million. The World Health Organization (WHO) predicts that the number of cardiovascular disease deaths (mainly heart attack and stroke) in 2030 will reach the number of 23.6 million [17].

Cardiovascular disease is a disease caused by heart and blood vessels disorders, one of which is Coronary Heart Disease (CHD). In Indonesia, CHD is the most common disease, with more than two million cases per years. The CHD can be prevented by overcoming the behavioral risk factors, such as the use of tobacco (smoking) and physical activity lack or irregularity.

In Indonesia, the smoking habit has become a comprehensible behavior. The society often serves the cigarettes as a food and drink, whether in daily life or in traditional events. The number of male smokers had reached 49.8 million people, whereas the number of female smokers is 3.9 million. In Indonesia, the number of cigarettes consumed is classified as very high to 1,675 cigarettes per person per year [15]. Smoking and heart disease are two harmful threats to health. Smokers are expected to have a 24 % higher risk of CHD suffered than non-smokers [1]. The content of nicotine and carbon monoxide in cigarettes can increase the risk of blood clots in the arteries [8]. Other chemicals in cigarettes can also damage the layers of coronary arteries with the result that the risk of CHD suffered is increased. A lifestyle is another trigger factor for CHD. Looking at the general population phenomena in productive age, many of them enjoyed the unhealthy lifestyle, as if they lacked physical activity.

The habit of regular exercise is capable of reducing systolic blood pressure, blood cholesterol levels, low density lipoprotein (LDL) levels, and increasing blood flow from the active organs to the less-active organs, as well as decreasing the risk factor of CHD. In Indonesia, the pattern of individuals with less physical activity (cumulative activities less than 150 minutes a week) rose from 26.1% in 2013 to 33.5% in 2018 [7]. This indicates that Indonesia is potentially encountering CHD improvement. The prevalence of CHD in Indonesia, based on doctor’s diagnosis, is 1.5% of the province of North Kalimantan with the highest prevalence rate, 2,2% of the special region of Yogyakarta, and 2% of Gorontalo [9]. In addition, there are 8 other provinces with a higher prevalence compared with the national prevalence, namely Aceh (1.6%), West Sumatera (1.6%), the special capital district of Jakarta (1.9%), West Java (1.6%), Central Java (1.6%), East Kalimantan (1.9%), North Sulawesi (1.8%), and Central Sulawesi (1.9%). The prevalence of female CHD (1.6%) is higher than male (1.3%).

The result of chi-square analysis with the cross-sectional method by Saesarwati and Satyabakti [11] points out that there is no significant correlation between the habit of smoking and exercise with the incidence of CHD in the productive age, but there is a significant correlation between the status of passive smokers (non-smokers) with the incidence of CHD in the productive age. The result of Sudayasa et al [13] analysis using the Odds Ratio (OR) test states that a meaningful correlation exist between smoking and CHD incidence. The smoker respondent is 2.45 times more likely to suffer from CHD than the non-smoker. Another research carried out by Tsani [16] using the chi-square analysis and the case-controlled Odds Ratio (OR) test indicates that there is no correlation between the smoking habits with CHD; however there is a meaningful correlation between the exercise habits with the incidence of CHD. The respondent who did not possess the exercise habits has a 4.89 times risk of CHD being affected as compared to the respondent who has the exercise habits. The results differences might be caused by a selection bias and a research information bias. In addition, the distinction in the research method can also cause differences in the research results. The previous studies used the simple statistical analysis as if the chi-square test and the Odds Ratio (OR) test used to determine the effect of the independent variable towards the dependent variable. The more complex statistical analysis is necessary to answer the further problem of research. According this research method used the logistic regression model to analyze the effect of CHD risk factors in Indonesia.

Logistic regression used to model the correlation between one dependent variable with two or more categories and one or more the independent variable. The purpose of the logistic regression analysis is to discover the most appropriate and efficient model, and it also reasonable to describe the correlation between the response variable and a set of predictors [5]. The method of binary logistic regression in used in this study to analyze the effect of smoking and exercise habits along with other factors such as age, sex, education, hypertension, and high cholesterol towards the risk of coronary heart disease (yes or no) in Indonesia.

1. Method
   1. *Data*

The research data used secondary data from RAND Corporation and Indonesian Survey METER, namely the fifth wave of Indonesian Family Life Survey (IFLS) which was conducted in 2014/2015. The response variable (dependent) in the research is in the form of the binary variable revealed to have CHD or not (yes or no), whereas the independent variables used are age (X1), sex (X2), education (X3), smoking habit (X4), exercise habit (X5), hypertension (X6), dan high cholesterol (X7).

* 1. *Data Analysis*

In the study, observation aggregated 28,513 respondents aged of over 19 years provides complete information about the required variables. The binary logistic regression was applied to examine the relationship between the set of CHD factors and whether or not CHD has been present. The model of binary logistic regression is given as follows

|  |  |
| --- | --- |
|  | (1) |

where are the regression parameters, are the independent variables, and is the probability of having CH. The data analyses with the logistic regression model were conducted by following steps of (1) parameter estimation, (2) parameter testing, (3) goodness of fit test, and (4) parameter interpretation. All data analyses in this study were performed using the glm function in stats package from R program.

* + 1. Parameter Estimation

The parameters in the logistic regression model can be estimated using the maximum likelihood method. The function of log-likelihood of binary logistic regression is given below

|  |  |
| --- | --- |
|  | (2) |

where .

The parameter estimates of the logistic regression model are the ones obtained by establishing the first derivative of to and then equating to zero [5]. The estimate of maximizing is obtained from the first derivation of to and the second derivation should be shown as negative.

|  |  |
| --- | --- |
|  | (3) |

* + 1. Parameter Testing

The likelihood ratio test (LRT) is used to test the single parameter of the logistic regression model. This test is to examine the relationship between each predictor and a binary response. The null hypothesis is (the th predictor can be excluded from the model) and the alternative hypothesis is (the th predictor remains in the model). The test statistic based on the LRT is as follows:

|  |  |
| --- | --- |
|  | (4) |

where is the likelihood without independent variable and is the likelihood with independent variable. The test statistic Λ follows the chi-square distribution with degrees of freedom (number of independent variables). The rule of decision-making is that H0 is rejected if or the value of .

* + 1. Goodness of Fit Test

Goodness of fit test is used to evaluate whether the model is fit or not with the data. The obtained observation value is the same or closed to what is expected in the model. The null hypothesis is that the model is fit, and the alternative hypothesis is that the model is not fit. The statistic of Hosmer-Lemeshow test with is formulated as follows:

|  |  |
| --- | --- |
|  | (6) |

where is the total number of subject in group, symbolizes the number of subject with in group k, and is the probability of average success estimated that based on the suitable model with the subject in group -th with. The test statistic approaches the chi-square distribution with degrees of freedom [4]. The null hypothesis will be rejected if or if the -value .

* + 1. The Odds Ratio

Odds ratio is one of the results from the logistic regression which is useful to interpret the findings. The odds ratio can be calculated from each of independent variable, for example of the -th independent variable is [2]

|  |  |
| --- | --- |
|  | (7) |

1. Result
   1. *Binary Logistic Regression Model*

The parameter estimates through maximum likelihood method of the binary logistic regression model with seven independent variables is displayed in Table 1.

Table 1. Binary logistic regression results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Category | Estimate | S.E. | Wald | -value |
| (Intercept) | | 3.832 | 0.249 | 14.066 | **< 0.001** |
| Age |  | -0.029 | 0.003 | -9.206 | **< 0.001** |
| Sex | Male (reference) |  |  |  |  |
| Female | 0.126 | 0.114 | -1.115 | 0.104 |
| Education | Unknown (reference) |  |  |  |  |
| SD (Elementary school) | 0.506 | 0.363 | 1.394 | 0.223 |
| SMP(Junior High School) | 0.325 | 0.345 | 0.941 | 0.422 |
| SMA(Senior High School) | 0.600 | 0.291 | 2.061 | 0.051 |
| College | 0.071 | 0.292 | 0.242 | 0.818 |
| The smoking status | Non-smoking (reference) |  |  |  |  |
| Smoking | 0.712 | 0.142 | 4.999 | **0.042** |
| Exercise intensity |  | -0.001 | 0.015 | -0.046 | 0.828 |
| Hhypertension | No (reference) |  |  |  |  |
| Yes | 0.871 | 0.104 | 8.347 | **< 0.001** |
| High cholesterol | No (reference) |  |  |  |  |
| Yes | 1.070 | 0.125 | 7.966 | **< 0.001** |

The estimated binary logistic regression equation for the CHD incidence in Indonesia as follows:

|  |  |
| --- | --- |
|  | (8) |

where is the probability of having CHD in Indonesia. In this study, the Wald test is used to test the set of hypotheses vs for individual regression slope coefficient. The Wald values are obtained by dividing the slope coefficients by their standard error. If the null hypothesis is true, the Wald value has an approximate standard normal distribution for a large sample and the null hypothesis is rejected if the Wald value is greater than the critical standard normal value or the -value is less than the significance level. For example, the coefficient for age is -0.032 and the corresponding standard error is 0.003, and the Wald value is indeed -9.206 as specified in Table 1. Since the corresponding -value for this test is < 0.001, hence there is strong evidence that age is important to include in the model holding all other explanatory variables constant. The Wald tests suggested that the age, smoking status, hypertension, and high cholesterol were statistically significant at the significance level 0.05 on each variable.

*3.2. LR Test*

Table 2 shows the results of LRTs in a sequential manner based on the order of the model fit explanatory variables.

Table 2. Likelihood Ratio Testing Result

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | L.R. | Degree of  Freedom | -value |
| Age | 82.056 | 1 | < 0.001 |
| Sex | 1.226 | 1 | 0.268 |
| Education | 7.297 | 4 | 0.121 |
| Smoking status | 25.859 | 1 | < 0.001 |
| Exercise intensity | 0.002 | 1 | 0.963 |
| Hypertension | 65.227 | 1 | < 0.001 |
| High cholesterol | 61.923 | 1 | < 0.001 |

It can be concluded at the significant level of 0.05 that age, smoking status, hypertension and high cholesterol affect the probability having CHD, but not sex and exercise intensity variables.

* 1. *Goodness of Fit Test*

Using a hoslem.test() function from ResourceSelection package in R , the Hosmer-Lemeshow chi-squared statistic of (df = 8, p-value =0.731) yields that the model fits to data.

* 1. *Odds Ratio*

The odds ratio is performed to determine the influence of each independent variable on dependent variable. The odds ratio of the model is shown in Table 3.

Table 3. Odds Ratio

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Category | OR | 95% C.I. | |
| Upper | Lower |
| (Intercept) | | 3.339 | 5.152 | 1.864 |
| Age |  | 0.969 | 1,613 | -1.676 |
| Sex | Male (reference) |  |  |  |
| Female | 1.135 | 1.770 | -1.518 |
| Education | Unknown (reference) |  |  |  |
| SD (Elementary school) | 1.658 | 2.150 | -1.138 |
| SMP(Junior High School) | 1.384 | 1.969 | -1.319 |
| SMA(Senior High School) | 1.822 | 2.244 | -1.044 |
| College | 1.073 | 1.715 | -1.573 |
| The smoking status | Non-smoking (reference) |  |  |  |
| Smoking | 2.038 | 2.356 | -0.932 |
| Exercise intensity |  | 0.999 | 1.643 | -1.645 |
| Hhypertension | No (reference) |  |  |  |
| Yes | 2.390 | 2.516 | -0.773 |
| High cholesterol | No (reference) |  |  |  |
| Yes | 2.614 | 2.714 | -0.575 |

The probability of having CHD increased significantly as the age increased. Female had higher probability of having CHD than male (OR = 1.135, 95%CI: -1.518 – 1.770). Respondents who had attained elementary school (OR = 1.658, 95%CI: -1.518 – 1.770), junior high school (OR = 1.384, 95%CI: -1.319 – 1.969), senior high school (OR = 1.822, 95%CI: -1.044 – 2.244) had higher probability of having CHD than those whose education was unknown, while respondents who had attained college (OR = 1.073, 95%CI: -1.573 – 1.715) had equally probability of having CHD with respondents whose education was unknown. Being smokers had higher probability of having CHD than being non-smokers (OR = 2.038, 95%CI: -0.932 – 2.356). The probability of having CHD was equally likely as the intensity of exercises increased since the odd ratio value was close to one (OR = 0.999). Respondents with hypertension had higher probability of having CHD than respondent who lacked hypertension (OR = 2.390, 95%CI: -0.773 – 2.516). High-cholesterol respondents had higher probability of having CHD than respondents who lacked high cholesterol (OR = 2.614, 95%CI: -0.575 – 2.714).

* 1. *Discussion*

This study describes the factors which influence CHD in Indonesia using the fifth wave of Indonesian Family Live Survey (IFLS). There were 28,513 respondents aged of over 19 years. The explanatory variable included age, sex, education, smoking status, exercise intensity, hypertension and high cholesterol, and the response variable is whether having CHD or not. The parameters testing of the logistic regression model using LR test shows that the explanatory variables that have a significant influence with CHD in Indonesia are age, smoking status, hypertension, and high cholesterol, while sex, education, and exercise intensity do not influence.

The model fit test using GOF statistical test from Hosmer and Lemeshow test show that the model (equation 8) is fit. There is no difference between the observed results and the possible prediction results.

The probability of having CHD increased significantly as the age increased. Davidson [3] states that increasing age will increase the risk of having CHD. Females are 1.135 more likely to have CHD than males. These similar results can be found at Rahajoe [10] is that the risk and mortality rate of CHD in female is higher than male. However, another result obtained by Salim and Nurrohmah [12]. Their study showed that male more likely to have CHD than female. Respondents who had attained elementary school, junior high school, and senior high school were more likely to have CHD than respondents whose education were unknown, while respondents who had attained college had similar probability of having CHD with respondents whose education were unknown. Smokers have a greater probability of having CHD than non-smokers.

Theoretically, based on Tandra [14], the risk of CHD for smokersincreases by 2 to 4 times compared to non-smokers. The risk of CHD increases as age and number of cigarettes increase. The risk of death by CHD is diminished into 50% in the first year after stopping smoke. Smoking habit is one of the CHD factors which can be controlled by changing smoking action. Besides, sport is one of the most effective ways to decrease the CHD risk. By doing sport activities such as aerobics (walking, running, swimming, and cycling) , heart and lung capacity can be increased. These activities make someone not easily feel tired and widen the heart blood vessels so that the blood rush becomes more switched, decreased blood cholesterol, decreased blood presure, the main factors of CHD [19].

The probability of having CHD is the same for the increased exercise intensity. Exercise will increase the physical activity, which will strenghten the heart muscle, accelerate blood rush, increase blood vessel glucose supply to the muscle so that diabetes is controlled. Wijayanto [19] study states that good exercise recommended by the association of cardiologists is a moderate intensity exercise which is done 3 times a week for at least 30 minutes. In this study the analysis shows different results from the general theory. There is no difference in the effect of having CHD in Indonesia among people who exercise different intensities.

Ghani [4] states that hypertension is a dominant risk factors for coronary heart disease. In this study, respondents with hypertension are more likely to have CHD (2.390 times more likely) than respondents without hypertension. Based on Iskandar [6], one of the most influential factors for CHD is cholesterol. High-cholesterol respondents are more likely (2.914 times more likely) to have CHD than respondents who lack high cholesterol.

* 1. *Conclusion*

The binary logistics regression model is fitted well with the data used in this study. The age, smoking status, hypertension and high cholesterol were reported to have significant factors for the logit of probability having CHD, while not for sex, education and exercise intensity. The probability of smokers to have CHD increases by 2.038 compared to non-smokers. The risk of having CHD increased significantly as the age increased. People with hypertension are reported to have 2.390 times compared to those without hypertension. While, people with high cholesterol have 2.914 times of CHD compared to those without high cholesterol.

References

1. Alodokter 2018 *Lebih Jauh tentang Penyakit Jantung Koroner* (accessed December 5, 2019) <https://www.alodokter.com/lebih-jauh-tentang-penyakit-jantung-koroner>
2. Bilder C R and Loughin T M 2015 *Analysis of Categorical Data with R* (New York: CRC Press)
3. Davidson C 2003 *Seri Kesehatan Bimbingan Dokter pada Penyakit Jantung Koroner* (Jakarta: Dian Rakyat)
4. Ghani L, Susilawati M D, and Novriani H 2016 Faktor Risiko Dominan Penyakit Jantung Koroner di Indonesia *Buletin Penelitian Kesehatan* **44**(3) pp 153 – 164
5. Hosmer D W, Lemeshow S, and Sturdivant R X 2013 *Applied Logistic Regression,* Third Edition (Hoboken: John Wiley & Sons)
6. Iskandar A H and Alfridsyah 2017 Faktor Risiko Terjadinya Penyakit Jantung Koroner pada Pasien Rumah Sakit Umum Meuraxa Banda Aceh *J. AcTion (Aceh Nutrition)* **2**(1) pp 32 – 42
7. Kemenkes Republik Indonesia 2014 *Info Datin: Situasi Kesehatan Jantung* (Jakarta Selatan: Germas)
8. Kemenkes Republik Indonesia 2017 *Hidup Sehat Tanpa Rokok* (Jakarta Selatan: Kemenkes RI)
9. Kemenkes Republik Indonesia 2018 *Hasil Utama Riskesdas 2018* (Jakarta Selatan: Kemenkes RI)
10. Rahajoe A U 2007 Penyakit Jantung pada Perempuan *J. Kardiologi Indonesia* **28**(3) pp 169 – 170
11. Saesarwati D and Satyabakti P 2016 Analisis Faktor Risiko yang Dapat Dikendalikan pada Kejadian PJK Usia Produktif *J. Promkes* **4**(1) pp 22 – 33
12. Salim A Y and Nurrohmah A 2013 Hubungan Olahraga dengan Kejadian Penyakit Jantung Koroner di RSUD dr. Moewardi J. GASTER **10**(1) pp 48 – 56
13. Sudayasa I P, Subijakto S, and Sahrul W O A 2014 Analisis Faktor Risiko Merokok, Stres, dan Riwayat Keluarga yang Berhubungan dengan Kejadian Penyakit Jantung Koroner *J. Sci. Halu Oleo University Faculty of Medicine* **1**(2) pp 48 – 56
14. Tandra H 2003 *Merokok dan Kesehatan* (accessed February 10, 2020) http://www.academia.edu/download/35644938/Merokok-dan-Kesehatan1.doc
15. The Tobacco Atlas 2016 *Fact Sheet: Indonesian Country* (accessed December 5, 2019) <https://tobaccoatlas.org/country/indonesia/>
16. Tsani F R 2013 Hubungan Antara Fakor Lingkungan dan Perilaku dengan Kejadian Penyakit Jantuk Koroner (Studi Kasus di Rumah Sakit X Kota Semarang) *J. Public Health UNNES*. **2**(3) pp 1 – 9
17. World Health Organization 2017, May 17 *Cardiovascular Diseases (CVDs)* (accessed December 5, 2019) <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)>.
18. World Health Organization 2018, May 24 *The Top 10 Causes of Death* (accesssed December 5, 2019) <https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death>
19. Wijayanto M A 2019 Olahraga Sebagai Pencegahan Penyakit Jantung Koroner *INA-Rxiv* *papers* doi:10.31227/osf.io/det2v