Analysis of Factors Affecting the Health Insurance Ownership with Binary Logistic Regression Model

D W Astari¹ and Kismiantini¹

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences, Yogyakarta State University, Jl Colombo No 1, Karangmalang, Depok, Sleman, Yogyakarta, Indonesia

E-mails: dhina.windy@student.uny.ac.id; kismi@uny.ac.id

Abstract. Having health insurance is an important decision for enjoying the security of a safe future. Health insurance can protect people from a large amount of medical costs. More Indonesian have health insurance now a days than ever before since the government is committed to supporting the universal health care. This study aims to determine factors affecting the health insurance ownership and to understand their relations in Indonesia using a binary logistic regression model. The data used in this study came from the fifth wave of the Indonesian Family Life Survey (IFLS). There were 29,508 respondents where 14,653 among them have health insurance. The logistic regression model suggested that job, education, chronic condition, marital status, and inpatient care were statistically significant to the health insurance ownership (yes/no), while not significant for gender and health condition. Fitting the logistic regression model with age as the only explanatory variable yielded that the probability of having health insurance increased in line with age.

1. Introduction
In today's life, a lot of people understand how importance of having insurance as an essential tool in protecting a family’s financial security. Various financial risks may happen in the future because of life, health, ageing, and education problems. When someone who as a breadwinner in the family dies or suffers from a permanent disability, it is more likely that the family will be in a difficult situation. A financial instability may also occur when the old age people are in the retirement. Hence, having life insurance may overcome the situation. Also, it is widely believed that the older people are most likely to have worst health outcomes so that having health insurance can help their large medical bills. A good education may also burden parent’s financial so that having education insurance can ensure the future educational needs for their children.

Several studies have investigated the importance of having life, property, and health insurances. Ward and Zurbruegg [5] investigated the determinants of life insurance in Asia. Hussels et al. [7] studied the determinants of life and property-casualty insurance demands. Kjosevski [16] identified the importance of a high level of education for life insurance in 14 countries in Central and South-Eastern Europe. Hsia et al. [4] examined the importance of health insurance status and insurance type relative to demographic, actual, and health variables as determinants of screening for several cancers. Blanton and Hoffman [8] reviewed the role of health insurance coverage in reducing racial/ethnic disparities in access to care in. Choi [12] studied the importance of health insurance among older
adults. Vidyattama et al. [20] studied the importance of providing subsidized health insurance in Indonesia to encourage people to seek health services when they feel sick or for preventive care.

In the study of the ownership of health insurance, Adams et al. [21] reported that the respondents who have health insurance were significantly associated with a willingness to participate in the plan than those without. Moreover, Amu and Dickson [22] found that the respondents with no education has the lowest insurance ownership. Their study confirmed that the wealth status, age, ecological zone, religion, ethnicity, education, marital status, parity were statistically significant to the ownership of health status in Ghana. Kimani et al. [18] found that the marital status, education, age, gender of household head, household wealth status were associated with having health insurance in Kenya. These findings showed the importance of the socio-demographic and economic variables in the study of the ownership of health status.

According to Swiss Reinsurance Company Ltd [25], Indonesia was ranked ninth out of twenty six countries in Asia on the insurance ownership by 2016, while in 2015 Indonesia was ranked in tenth. This indicates more Indonesians having insurances. Although many Indonesians appreciate the importance of having insurance, but not all of them know to use the insurance. According to Hoesen in [17], 89% of Indonesians know the importance of using insurance, but only 17% of them use insurance. One of the reason is that if they use the insurance then the insurance holders have to pay monthly premium payment.

When we want to analyze a binary response variable (i.e. having insurance or not) with several explanatory variables, a logistic regression model is a useful model for describing their relationships. It is commonly represented that the binary response variable with value of 1 states the existence of a characteristic and 0 expresses the non-existence [10]. Many researches on applying binary logistic regression models have been conducted. Kalhori et al. [14] used a logistic regression model for predicting the risk of patients failing in tuberculosis (TB) treatment due to completion. Their study found that higher probability of patients failing in the treatment were associated with old patients, male, low weight, foreign, prisoner, and history of relapse. Deb and Trivedi [2] studied factors which affect private insurance ownership (yes / no). They found that many chronic condition, many hospitalization, bad condition, good condition, length of study, and gender were the important factors influencing the private insurance ownership. A similar research had been also conducted by Antwi and Zhao [15] in Ghana. They found that gender, age, marital status, and length of stay were significantly affecting the ownership of health insurance.

In this study, we want to examine factors which affect the ownership of health insurance in Indonesia using the national survey data of the fifth wave of Indonesian Family Life Survey (IFLS). Indonesia is one of the largest country in the world where in the rank of 51 out of 84 of the health care index [26]. Therefore, it is important to understand the factors which influence lower or higher probability of having health insurance.

2. Methods
2.1 Data
Data were taken from the fifth wave of the Indonesia Family Life Survey (IFLS) in 2014/2015. The IFLS 5 was conducted by RAND in collaboration with the Center for Population and Policy Studies (CPPS) of the University of Gadjah Mada and Survey METER. The sample was representative of about 83% of the Indonesian populations in 13 provinces in the country.

2.2 Measure
The outcome variable was whether a respondent had health insurance or not (yes or no). The explanatory variables examined in the study were selected from previous related studies which included respondents job, age, gender, highest education, chronic condition, health condition, marital status, and inpatient care. The job variable was found in Culwell and Feinglass [9] study; the age variable was found in Truett and Truett [1]; gender in Liao et al. [23] study; marital status and
in-patient care were found in Antwi and Zhao[15]; highest education and chronic conditions were found in Deb and Trivedi [2] study; health conditions in Gius [13] study.

2.3 Data Analysis
In this study, a total of 29,508 respondents were used in the analyses. Descriptive statistics and binary logistic regression were used to overview characteristics of the sample and to identify factors associated with health status ownership. The binary logistic regression analysis can be used to determine the relationship between a binary response and continuous or categorical explanatory variables. The binary logistic regression model is given below,

\[
\logit(\pi) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p
\]

where \(\beta_0, \beta_1, \ldots, \beta_p\) are regression parameters; \(X_1, X_2, \ldots, X_p\) are explanatory variables, and \(\pi\) is the probability of success. The steps in the analyses using the binary logistic regression are (1) calculating descriptive statistics, (2) estimating regression parameters using maximum likelihood estimation, (3) conducting the likelihood ratio test, (4) conducting the Wald test, (5) determining the final model, (6) interpreting the regression coefficients and the odds ratio, and (7) evaluating the goodness of fit of the model using accuracy of classifications.

A Chi-square test for independence was carried out to test the relationship between health insurance ownership and individual explanatory variables. The null hypothesis is that the two categorical variables are independent (or there is no relationship between the two categorical variables). The alternate hypothesis is that the two categorical variables are dependent (or there is a relationship between the two categorical variables). When the null hypothesis is rejected than there is an evidence that there is an association between the two variables. Thus, for health condition, the hypotheses are as follows: (i) \(H_0\): A respondent’s health insurance ownership status and their health condition are independent or unrelated; and (ii) \(H_1\): A respondent’s health insurance ownership status and their health condition are related (dependent on each other). All data analyses in this study were performed using R program.

3. Results
3.1 Descriptive statistics
Table 1 shows the frequency and percentage distributions of the outcome and explanatory variables in this study. Overall, 53.35\% of women in the sample had health insurance which is higher percentage than men of 46.65\%. Among women who had health insurance, 64.83\% were unemployed and 49.52\% were employed in the formal or informal sectors, respectively. Only 18.45\% of men who reported to have health insurance were unemployed and 81.55\% were employed in either formal or informal sectors, respectively. Surprisingly, about 30\% of the respondents who reported to have or no health insurance had the highest level of education in senior high school or elementary school. The majority of respondents were married, somewhat healthy conditions, not in chronic conditions, and inpatient cares.

Age is the only continuous variable in this study. The average and median age were 36.60 and 34 years, with a standard deviation of 14.30. The sample of respondents without health insurance had an average age of 36.23 (STD = 14.43) compared to 36.98 (STD = 14.16) in the group with insurance.

The results of the chi-square test of the association between health insurance ownership and categorical explanatory variables are shown in the last column of Table 1. All the computed Chi-Squared for gender, job, highest education, chronic condition, health condition, marital status, and inpatient care are greater than their respective critical Chi-Squared values (at 5\% significance level), so that all the null hypotheses are rejected and concluded that they are dependent. Therefore in the case of health condition, the alternate hypothesis (\(H_1\)) is accepted so that a respondent’s health insurance ownership status and their health condition are related (dependent on each other).
Table 1. Frequency and percentage for the categorical explanatory variables

| Variables               | Have health insurance: frequency (%) [N=14653] | Do not have health insurance: frequency (%) [N=14855] | Chi-Square (p-value) |
|-------------------------|-----------------------------------------------|------------------------------------------------------|---------------------|
| Gender                  |                                               |                                                      |                     |
| Male                    | 6836 (46.65)                                  | 6988 (47.04)                                         | 0.430 (p = 0.5108)  |
| Female                  | 7817 (53.35)                                  | 7867 (52.96)                                         |                     |
| Job                     |                                               |                                                      |                     |
| Yes                     | 5693 (38.85)                                  | 6057 (40.77)                                         | 11.292 (p = 0.0008) |
| No                      | 8960 (61.15)                                  | 8798 (59.23)                                         |                     |
| Education               |                                               |                                                      |                     |
| Elementary School       | 4000 (27.30)                                  | 5109 (34.39)                                         | 560.700 (p < 0.0001) |
| Junior High School      | 2655 (18.12)                                  | 3315 (22.32)                                         |                     |
| Senior High School      | 5264 (35.92)                                  | 4903 (33.01)                                         |                     |
| University              | 2734 (18.66)                                  | 1528 (10.29)                                         |                     |
| Chronic condition       |                                               |                                                      |                     |
| Yes                     | 116 (0.79)                                    | 69 (0.46)                                            | 12.153 (p = 0.0005) |
| No                      | 14537 (99.21)                                 | 14786 (99.54)                                        |                     |
| Health condition        |                                               |                                                      |                     |
| Very unhealthy          | 146 (1.00)                                    | 167 (1.12)                                           | 16.754 (p = 0.0008) |
| Somewhat unhealthy      | 2982 (20.35)                                  | 2813 (18.94)                                         |                     |
| Somewhat healthy        | 8850 (60.40)                                  | 8949 (60.24)                                         |                     |
| Very healthy            | 2675 (18.26)                                  | 2926 (19.70)                                         |                     |
| Marital status          |                                               |                                                      |                     |
| Unmarried               | 2771 (18.91)                                  | 3053 (20.55)                                         | 19.553 (p < 0.0001) |
| Separated               | 935 (6.38)                                    | 1041 (7.01)                                          |                     |
| Married                 | 10947 (74.71)                                 | 10761 (72.44)                                        |                     |
| Inpatient care          |                                               |                                                      |                     |
| Yes                     | 982 (6.70)                                    | 429 (2.89)                                           | 234.810 (p < 0.0001) |
| No                      | 13671 (93.30)                                 | 14426 (97.11)                                        |                     |

3.2 Binary logistic regression analysis

The results of the binary logistic regression analysis for determinants of health insurance ownership are shown in Table 2. Table 2 provides coefficients and their standard errors (S.E.), Wald test values, p values, odd ratios, and 95% confidence intervals for odds ratios. When all explanatory variables treat as continuous variables, we found that each explanatory variable was statistically significant at the significance level 0.05 (results not shown). However, in this study we prefer treating all explanatory variables as categorical with the exception for age. The Wald test is used to test the set of hypotheses \( H_0: \beta_r = 0 \) vs \( H_1: \beta_r \neq 0 \) for individual regression slope coefficients. 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 p-value is less than the significance level. For example, the coefficient for age is 0.008 and the corresponding standard error is 0.001, and the Wald value is indeed 7.578 as specified in Table 2. Since the relevant p-value for this test is < 0.0001, hence there is strong evidence that age is important to include in the model given the other explanatory variables in the model. The Wald tests suggested that the age, job, education, chronic condition, married, and inpatient care were statistically significant at the significance level 0.05 on each variable. Thus, we interest for testing the null hypothesis of the logistic regression model without gender and health condition, and the alternate hypothesis of the logistic regression model with all explanatory variables using the Likelihood Ratio (LR) test. The LR statistic (A) is defined by
dividing the maximum of likelihood function under $H_0$ by the maximum of likelihood function under $H_0$ or $H_1$. When $q$ regression parameters are set to 0 in the null hypothesis and if the null hypothesis is true, the $-2 \log(\Lambda)$ statistic has an approximate $\chi^2_q$ distribution for a large sample, and the null hypothesis is rejected if $-2 \log(\Lambda) > \chi^2_{q,1-\alpha}$ or p-value < $\alpha$. In this study, the LR statistic value was obtained 940.563 and p < 0.0001 which indicated that there is strong evidence of the importance of gender and health condition given that age, job, education, chronic condition, marital status, and inpatient care are in the model. Hence, all explanatory variables were kept in the model.

Table 2. Binary logistic regression results

| Variable                  | $\hat{\beta}$ | S.E. | Wald | p-value | Odds Ratio | 95% CI for Odds Ratio |
|---------------------------|----------------|------|------|---------|------------|-----------------------|
| Age                       | 0.008          | 0.001| 7.578| < 0.0001| 1.008      | (1.006, 1.011)        |
| Gender (female)           | 0.038          | 0.027| 1.415| 0.157   | 1.039      | (0.985, 1.095)        |
| Job (yes)                 | 0.069          | 0.027| 2.564| 0.111   | 1.072      | (1.016, 1.131)        |
| Junior High School        | 0.126          | 0.035| 3.551| < 0.0001| 1.134      | (1.058, 1.216)        |
| Senior High School        | 0.452          | 0.032| 14.032| 0.011 | 1.571      | (1.475, 1.673)        |
| University                | 0.926          | 0.040| 22.921| < 0.0001| 2.523      | (2.331, 2.731)        |
| Chronic condition (yes)   | 0.335          | 0.156| 2.144| 0.032   | 1.398      | (1.029, 1.899)        |
| Somewhat unhealthy        | 0.186          | 0.118| 1.571| 0.116   | 1.204      | (0.955, 1.519)        |
| Somewhat healthy          | 0.093          | 0.117| 0.796| 0.426   | 1.097      | (0.873, 1.380)        |
| Very healthy              | 0.057          | 0.119| 0.476| 0.634   | 1.058      | (0.838, 1.336)        |
| Separated                 | -0.040         | 0.064| -0.625| 0.532 | 0.961      | (0.848, 1.089)        |
| Married                   | 0.093          | 0.036| 2.544| 0.011   | 1.097      | (1.021, 1.178)        |
| Inpatient care (yes)      | 0.834          | 0.060| 13.843| < 0.0001| 2.302      | (2.046, 2.591)        |

In general, the probability of having health insurance increased significantly as the age increased. Women were reported to have similar the probability of having health insurance with men (OR = 1.039, 95% CI: 0.985-1.095). Being employed in the formal or informal sectors was significantly associated with a higher probability of having health insurance compared to being unemployed (OR = 1.072, 95% CI: 1.016-1.131). Education was a significant explanatory variable of having health insurance. Respondents who had attained junior high school (OR = 1.134, 95% CI: 1.058-1.216), senior high school (OR = 1.571, 95% CI: 1.475-1.673), and university (OR = 2.523, 95% CI: 2.331-2.731) were associated with a higher probability of having health insurance compared to those with elementary school. Respondents who had chronic condition were associated with a higher likelihood of having health insurance compared to those non chronic education (OR = 1.398, 95% CI: 1.029-1.899). Health condition was not significant associated with the probability of having health insurance. Respondents who had in somewhat unhealthy (OR = 1.204, 95% CI: 0.955-1.519), somewhat healthy (OR = 1.097, 95% CI: 0.873-1.380), and very healthy (OR = 1.058, 95% CI: 0.838-1.336) were reported to have similar the probability of having health insurance with very unhealthy respondents. Being separated had similar probability of having health insurance with being unseparated (OR = 0.961, 95% CI: 0.848-1.089). Other significant determinants of having health insurance were the married and inpatient care. Being married was significantly associated with a higher probability of having health insurance compared to being unmarried (OR = 1.097, 95% CI: 1.021-1.178). About 5% of respondents reported being inpatient care was positively associated with higher probability of having health insurance compared to outpatient (OR = 2.302, 95% CI: 2.046-2.591).

Figure 1 shows the estimated probability of having health insurance (solid line) and the 95% individual confidence level (dash line) for an age as the only continuous explanatory variable in the binary logistic regression model. It can be seen that there is a slightly increasing curve between the
estimated probability and age. The 95% individual confidence interval was quite wider for the old age in comparison with adults. On average, the estimated probability of having health insurance is 0.495. This indicates that the importance of having health insurance is being perceived for only a half of respondents. The old age corresponds to higher and lower probability of having health insurances than other respondents.

![Figure 1. Estimated probability of having health insurance for an age](image)

A good model is a model which has a minimal chance of misclassification [3]. The method used in the binary classification is a Fisher linear discriminant analysis. Table 3 provides the results of the classification accuracy using this method. The total percentage of accuracy is 56.5% which indicates a moderate model.

| Observed          | Predicted          | Total   | % accuracy |
|-------------------|--------------------|---------|------------|
|                   | Do not have health insurance (no) | Have health insurance (yes) |         |           |
| Do not have health insurance (no) | 9291               | 5564    | 14855      | 63        |
| Have health insurance (yes)       | 7276               | 7377    | 14653      | 50        |
| Total              | 16567              | 12941   | 29508      | 56.5      |

4. Discussion
This study describes the factors which influence the ownership of health insurance using the IFLS 5. The model obtained indicates that eight factors (age, job, junior high school, senior high school, university, having chronic condition, being married, having inpatient care) were significantly associated with the health insurance ownership (all p-values < 0.05). The gender, somewhat unhealthy, somewhat healthy, very healthy, and separated were not associated with the probability of having health insurance. The results showed that the highest education (university) has more contributions than other factors (OR = 2.523) and followed by the respondent is being inpatient care (OR = 2.302).
The probability of having insurance was reported gradually increasing as age increased. The respondents who had job were seen to have high probability of having insurance. These similar results can be found in Kirigia et al. [9]. This study showed that respondents who had the highest education in the university were more likely to have health insurance than those who had the highest education in the elementary, middle, and high school. This indicates that high education affects someone to have health insurance, this is in line with the results of Amu and Dickson [22]. The respondents who had history of hospitalization were more likely to have health insurance compared to the respondents who had no history of hospitalization (p-value < 0.0001, OR = 2.302). The respondents who had chronic condition were more likely to have insurance compared to the respondents who had no chronic condition. These similar results can be found in Kushel et al. [6]. The respondents who were being married were more likely to have health insurance compared to those who were not married. This finding corresponds to result of Kirigia et al. [9] and Kimani et al. [8]. To improve the percentage of accuracy model may be incorporated an income variable in the future study.

5. Conclusion
The binary logistic regression model suggested that eight variables (job, age, gender, highest education, chronic condition, health condition, marital and inpatient care) were significantly associated with the ownership of health status. The chance of female respondents, had the highest education of junior high school or high school or university, had chronic condition, had health condition some what unhealthy or somewhat healthy or very healthy, being married, and had in patient care were higher to have health insurance than those who did not have health insurance. The probability of having insurance increased with age.

6. Acknowledgment
The authors thank to the Institute of Research and Community Service, Yogyakarta State University, due to the funding of DIPA LPPM UNY-2018 for this work.

7. References

[1] Truett D B and Truett L J 1990 The demand for life insurance in Mexico and the United States: a comparative study J Risk Insur. 57321-328
[2] Deb P and Trivedi P 1997 Demand for medical care by the elderly: a finite mixture approach. J Appl Econ. 12 313-336.
[3] Hosmer D and Lemeshow S 2000 Applied logistic regression (New York: John Wiley & Sons)
[4] Hsia J, Kemper E, Kiefe C and et al 2000 The importance of health insurance as a determinant of cancer screening: Evidence from the women’s health initiative Prev med. 31 261–27
[5] Ward D and Zurbruegg R 2002 Law, politics and Life insurance consumption in asia GENEVA PAP R I-ISS P. 27 395–412.
[6] Kushel M B, Vittinghoff E, and Haas J S 2001 Factors as sociated with the health care utilization of homeless persons JAMA. 285 200-206.
[7] Hussels S, Ward D and Zurbruegg R 2005 Stimulating the demand for insurance Risk Management And Insurance Review. 8 257–78
[8] Blanton M L and Hoffman C 2005 The role of health insurance coverage in reducing racial/ethnic disparities in health care Health Aff (Millwood). 24 398-408.
[9] Kirigia J M, Sambo L G, Nganda B, Mwabu G M, Chatora R, and Mwase T 2005 Determinants of health insurance ownership among South African women BMC Health Serv Res. 51 7.
[10] Agresti A 2007 Categorical data analysis second edition (New York: John Willey & Sons)
[11] Culwell K R and Feinglass J 2007 The association of health insurance with use of prescription Contraceptives Open Med. 39 226-230.
[12] Choi B P 2010 The U.S. property and liability insurance industry: firm growth, size and age
Risk Management and Insurance Review. 13 207-224.

[13] Gius M P 2010 An analysis of the health insurance coverage of young adults Int J Appl Econ. 7
[14] Kalhori S R, Nahesi M and Zeng X-J 2010 A logistic regression model to predict high
risk patients to fail in tuberculosis treatment course completion IAENG Int J Appl Math. 40
[15] Antwi S and Zhao X 2012 A logistic regression model for Ghana national health insurance
claims International Journal of Business and Social Reasearch (IJBSR). 2 139-147
[16] Kjosevske J 2012 The determinants of life insurance demand in central and south eastern europe
International Journal of Economics and Finance. 4 237-247
[17] BPJS Humas. (2014, May 21). BPJS kesehatan. Downloaded on March 23, 2018, from BPJS
kesehatan website: https://bpjs-kesehatan.go.id/bpjs/index.php/post/read/2014/88/BPJS-
Kurangi-Kesenjangan-Kebutuhan-Asuransi-Kesehatan.
[18] Kimani J K., Ettarh R, Warren C, and Bellows B 2014 Determinants of health insurance
ownership among women in Kenya: evidence from the 2008–09 Kenya demographic and
health survey Int J Equity Health. 13 27.
[19] Owusu-Sekyere E and Chiaraah A 2014 Demand for health insurance in Ghana: what factors
influence enrollment? AmJ Public Health Res. 2 27-35. doi: 10.12691/aiphr-2-1-6.
[20] Vidyattama Y, Miranti R and Resosudarmo B 2014 The role of health insurance membership
in health service utilisation in Indonesia Bull Indones Econ Stud. 50 393-413.
[21] Adams R, Chou Y J, and Pu C 2015 Willingness to participate and pay for a proposed national
health insurance in St. Vincent and the grenadines: a cross-sectional contingent valuation
approach. BMC Health Serv Res 15 148.
[22] Amu H and Dickson K S 2016 Health insurance subscription among women in reproductive age
in Ghana: dosocio-demographics matter? Health Econ Rev. 6 24.
[23] Liao Y, Gilmour S and Shibuya K 2016 Health insurance coverage and hypertension control in
china: results from the china health and nutrition survey PLoS One. 11
[24] Strauss J F W and B Sikoki 2016 The fifth wave of the indonesia family life survey (IFLS5):
Overview and Field Report WR-1143/1-NIA/NICHD
[25] Swiss Reinsurance Company Ltd 2017 World insurance 2016 (Zurich: Swiss Re)
[26] Health Care Index for Country 2019. https://www.numbeo.com/health-
care/rankings_by_country.jsp