Applied Mathematical Sciences, Vol. 8, 2014, no. 52, 2563 - 2572
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2014.42130

Association of Hypertension with Risk Factors Using Logistic Regression

Wan Muhamad Amir W Ahmad[1*], Mohamad Arif Bin Awang Nawi[1], Nor Azlida Aleng[1], Nurfadhlina Abdul Halim[1], Mustafa Mamat[2], Mohd Pouzi Hamzah[1] and Zalila Ali[3]

[1] School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, 21030 K Terengganu, Malaysia

[2] Director of International Student Centre, Universiti Sultan Zainal Abidin 21300 K Terengganu, Malaysia

[3] School of Mathematics Sciences, Universiti Sains Malaysia 11800 Minden, Pulau Pinang, Malaysia

*Coresponding author

Abstract

Hypertension is one of the important public health challenges worldwide because of its high frequency and concomitant risks of cardiovascular and kidney disease. A hypertension model is built to test the interaction and significance between the factors. In this present paper, we present the results that gained from multiple logistic regression method and used to model the relationship between the ordinal outcome variable. The significant variables is chosen based on the *p*-value associated to the significant level of model that lies on α = 0.05. Logit determination and the correlation between the variables are also discussed for further analysis. there are three factors that most significant of the six factors tested were identified as having influence significantly the performance of human blood pressure (hypertension). These factors are age (*p*-value <0.000), body mass index (*p*-value <0.001) and systolic (*p*-value <0.001). The use of mathematical software PASW version 18 is applied in this research as an alternatives calculation procedures derived from the methodology.

Keywords: Hypertension, Logistic Regression, multiple logistic regression, risk factors and blood pressure

INTRODUCTION

Hypertension is one of the important public health challenges worldwide because of its high frequency and concomitant risks of cardiovascular and kidney disease. It has been identified as a leading risk factor for mortality and ranked third as a cause of disability- adjusted life-years (Ezzati et al., 2002). The accelerating epidemic of hypertension in India was documented by studies done at various places across the country. The National Nutrition Monitoring Bureau (NNMB), which monitors the nutritional status of the population in nine States of India has estimated the prevalence of hypertension among the rural adult (aged 18 and above) population of India to be 25 per cent during 2004-2005 (NNMB, 2006).

Various risk factors have been associated with hypertension, including age, sex, race, physical activity, and socioeconomic class. Vast majority of cases of uncontrolled hypertension are amongst individuals more than 60 years of age (Thomas & Ramachandran, 2005). Population studies have also shown that blood pressure correlates with body mass index (BMI) and other anthropometric indices of obesity such as waist-hip ratio. In the Framingham Study, 70% of new cases of hypertension were related to excess body fat (Kotsis et al., 2005).

The reported prevalence of hypertension varies around the world, with rates as low as 5.2% in rural North India and as high as 70.7% in Poland. Blood pressure variations also exist from within communities in the same country depending upon the economic development and affluence. In economically developed countries, the prevalence of hypertension range between approximately 20 and 50% (Patricia et al., 2004). Prevalence of hypertension in the Asia-Pacific region ranges from 5 to 47% in men and from 7 to 38% in women (Lawes et al., 2004).

Hypertension is the most common cardiovascular disorder affecting approximately 1 billion people globally and accounts for approximately 7.1 million deaths annually. Some of the known risk factors for primary hypertension like age, heredity, and gender are non-modifiable. However, the majority of the other risk factors like tobacco use, alcohol use, unhealthy diet, physical inactivity, overweight and obesity can be effectively prevented (Brundtland, 2002).

The objectives of this study were to determine the prevalence of hypertension and its associated risk factors in Seremban, Negeri Sembilan Malaysia.

METHODOLOGY

A questionnaire specially prepared for this study was used for data collection. Data on age, sex, body mass index (BMI), systolic, diastolic and total cholesterol was collected. Positive diagnosis of hypertension was made when the systolic blood pressure was ≥ 140mmHg and/or diastolic blood pressure ≥ 90mmHg. BMI was calculated using a simple equation (body weight in kg divided by height in m2).

Logistic regression model was developed use by Truett et al. (1967), using multivariate analysis in the Framingham heart study. Since then, the logistic regression model has become a standard binary or dichotomous data analysis in various areas of medicine.

Data was tabulated, cross tabulated and analyzed statistically using PASW version 18. A probability value of $p < 0.05$ was considered to be significant. Binary logistic regression analysis was conducted with reporting of odds ratio to establish the risk for hypertension. To explore the underlying association between hypertension and the selected explanatory variables, a set of logistic regression models is fitted in this section. Let define the following dichotomous variables for the hypertension.

Y= 0 if there is no exist hypertension
Y= 1 if there is exist hypertension

Then, logit for multiple logistic regression given by the equation below:

$$g(x) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Sex} + \beta_3 \text{BMI} + \beta_4 \text{Systolic} + \beta_5 \text{Diastolic} + \beta_6 \text{CHOL}$$

The results shown by the conditional probability is

$$\pi(x) = P(Y_{ij} = 1 | x_{ij}) = \frac{e^{g(x_{ij})}}{1 + e^{g(x_{ij})}}$$

β_0 is the Y intercept, βs are regression coefficients, and Xs are a set of predictors (Age, Sex, BMI, Systolic, Diastolic and CHOL). β_0 and βs are typically estimated by the maximum likelihood (ML) method, which is preferred over the weighted least squares approach by several authors, such as Haberman (1978) and Schlesselman (1982). The ML method is designed to maximize the likelihood of reproducing the data given the parameter estimates. Data are entered into the analysis as 0 or 1 coding for the dichotomous outcome, continuous values for continuous predictors, and dummy codings (e.g., 0 or 1) for categorical predictors. The null hypothesis underlying the overall model states that all βs equal zero. A rejection of this null hypothesis implies that at least one β does not equal zero in the population, which means that the logistic regression equation predicts the probability of the outcome better than the mean of the dependent variable Y. The interpretation of results is rendered using the odds ratio for both categorical and continuous predictors.

2.1 Sample Size Calculation

In this study, we calculate the sample size with manual calculation, with the significance level (α) 0.05 and the power of study (1-β) of 80%. The detectable hazard ratio of the presence of prognostic factor relative to absence of prognostic factors was decided by the researcher and expert opinion by clinicians (Naing, 2003). Two Proportions formula is given as follows:

$$n = \frac{p_0(1-p_0)+p_1(1-p_1)}{(p_0-p_1)^2}\left(z_\alpha + z_\beta\right)^2$$

where:

P_0 = Based on literature review
P_1 = Based on expert opinion
$z_\alpha = z_{0.05} = 1.9600$ (one tailed)
$z_\beta = z_{0.20} = 0.8416$ (one tailed)

Table 1. Sample Size Calculation

No	Objective	*P_1	P_0	Type 1 Error	Power	Sample Size
1	Patients with hypertension and associated risk factors (Joshi & Pooja, 2010)	0.414	0.275	5%	80%	174 patients

Calculation

$$n = \frac{0.414(1-0.414)+0.275(1-0.275)}{(0.414-0.275)^2} \times (1.96+0.80)^2 = 174.2 \approx 174 \text{ patients}$$

After adding 10% estimated missing data, we get $n = 174 + (0.1 \times 174) = 191.4 \approx 191$ per group, which can be obtained as follows:

i. Patients with hypertension and associated risk factors = 191 patients
ii. Patients without hypertension and associated risk factors = 191 patients

Therefore, a total patient to be sampled is $(191 \times 2) = 382 \approx 383$ patients.

RESULTS AND DISCUSSION

3.1 Logistic Regression Models

Based on Table 1 below, there are three significant factors for the multiple logistic regression models. The first factor is age (p-value = 0.002), the second factor is the BMI (p-value = 0.002) and the third factor is Systolic (p-value = 0.028). Each entry of significant factors in the model gives the p-value varies depending on certain measures.

Table 1: Estimates of parameters of logistic regression model for hypertension

Factors	B	S.E.	Wald	Sig.	Exp(B)
Age	0.054	0.017	9.997	0.002	1.055
Sex	-0.168	0.293	0.328	0.567	0.846
BMI	0.100	0.032	9.618	0.002	1.105
Systolic	0.023	0.011	4.802	0.028	1.024
Diastolic	0.005	0.017	0.101	0.751	1.005
CHOL	-0.129	0.141	0.842	0.359	0.879
Constant	-8.693	2.028	18.376	0.000	0.000

Based on the significant factors above, there are significant factor values for the coefficients for age, BMI and systolic. Based on Table 1, multiple logistic regression models were as follows:

$$\pi(\mathrm{x}) = \frac{\ell^{-8.693+0.054\times \mathrm{Age}+0.100\times \mathrm{BMI}+0.023\times \mathrm{Systolic}}}{1+\ell^{-8.693+0.054\times \mathrm{Age}+0.100\times \mathrm{BMI}+0.023\times \mathrm{Systolic}}}$$

Predicted logit of Hypertension= $-8.693 + 0.054 \times \mathrm{Age} + 0.100 \times \mathrm{BMI} + 0.023 \times \mathrm{Systolic}$

According to the model a three-predictor logistic model was fitted to the data. From the model, Age ($p<0.05$), BMI ($p<0.05$) and Systolic ($p<0.05$) were positively related to Hypertension. For each point increase on the Age, the odds of being Hypertension also increase from 1.0 to $e^{0.054}$ (1.055). For variable BMI, increasing of BMI from 1.0 will increase the odds of Hypertension to $e^{0.100}$ (1.105) and for the Systolic increase from 1 to $e^{0.023}$ (1.024).

3.2 Assess the Goodness-of-Fit

3.2.1 Hosmer and Lemeshow

Hosmer and Lemeshow Test is based on grouping cases into deciles of risk. It compares the observed probability with the expected probability within each deciles. The P-value is checked. If it is >0.05, there is no significant difference between the observed probability and the expected probability. Based on the Table 2 below, *p*-value = 0.126 obtained is greater than 0.05 (p-value> 0.05) and suggesting that the model was fit to the data well. In other words, the null hypothesis of a good model fit to data was tenable (Hosmer and Lemeshow, 2000).

Table 2:Hosmer and Lemeshow Test

Chi-square	df	Sig.
12.603	8	0.126

3.2.2 Receiver Operating Characteristic Curve (ROC)

The ROC curve is a fundamental tool for diagnostic test evaluation. It's a graphical plot of the sensitivity which measure of the overall performance of a diagnostic test. ROC Curve take on any value between 0 and 1, since both the x and y axes have values ranging from 0 to 1. The closer the area is to 1.0, the better the test is, and the closer the area is to 0.5, the worse the test is. The larger the area, the better the diagnostic test achieved. If the area is 1.0, we have an ideal test, because it achieves both 100% sensitivity and 100% specificity. If the area is 0.5, then we have a test which has effectively 50% sensitivity and 50% specificity (Ahmad et al. 2010 & 2011). In practice, a diagnostic test is going to have an area somewhere between these two extremes.

Area under the ROC curve is 0.718 (95% CI: 0.661, 0.776). It is significantly different from 0.05 (0.000). The model can accurately discriminate 71.8% of the cases.

Table 3: Area Under the Curve

Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval Lower Bound	Upper Bound
0.718	0.029	0.000	0.661	0.776

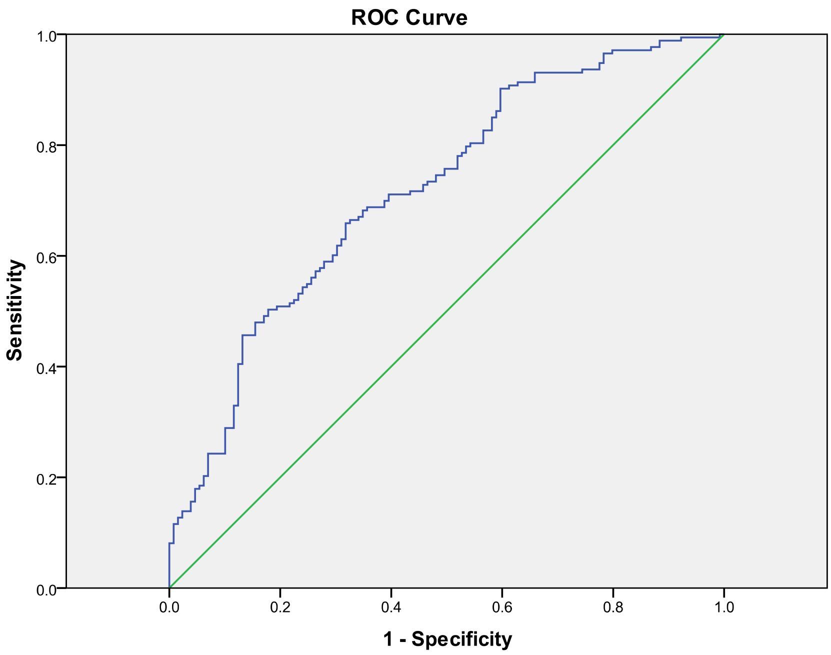

Fig.1. ROC Curve

3.3 Final Model of Logistic Regression

The final interpretation of the data is on the analysis of models produced.

Table 4: Final Model of Logistic Regression

Factors	B	S.E.	Wald	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Age	0.048	0.014	12.661	0.000	1.050	1.022	1.078
BMI	0.098	0.028	11.978	0.001	1.103	1.043	1.166
Systolic	0.028	0.008	11.624	0.001	1.028	1.012	1.044
Constant	-9.088	1.542	34.715	0.000	0.000		

A person with an increase in age has a 1.050 times the odds to have Hypertension (95% CI: 1.022 to 1.078, p-value <0.000) when when adjusted for BMI and systolic.

A person with an increase in BMI= mass (kg) / (height (m))^2 has a 1.103 times the odds to have Hypertension (95% CI: 1.043 to 1.166, p-value <0.001) when when adjusted for age and systolic.
A person with an increase in 1 mmHg of systolic blood pressure has a 1.028 times the odds to have Hypertension (95% CI: 1.012 to 1.044, p-value <0.001) when when adjusted for age and BMI.

DISCUSSION

In our study an attempt has been made to find out the association between different risk factors with hypertension by binary logistic regression analysis. Prevalence of hypertension increases with increasing age. Similarly a health survey in England 2003 (Primatesta and Poulter, 2006) and in the United States reported strong correlation between age and blood pressure. In central Malaysia the prevalence of hypertension among those aged 55 years and above living in a community was shown to be 25.6% and 51.1% among those living in old folks home. The prevalence of hypertension among the elderly in this study is comparable with the finding of another study which was conducted in northern Malaysia (Rashid et al., 2008).

In this study, variable systolic blood pressure was significant than others variables. This finding is consistent with the results from a study done by Jagmeet et al. 1999, where the strong relation observed between exaggerated systolic response and incident hypertension in univariate analysis was attenuated in multivariate analysis, suggesting that exercise systolic response was a weaker predictor of hypertension than the diastolic response. This finding is at odds with several other studies that have reported exercise systolic BP as a strong predictor of hypertension (Tanji et al., 1989).

Obesity is a well-established risk factor for hypertension (Rankinen et al., 2007). In this study the prevalence of hypertension increased with BMI. In the Ansan Study conducted in Korea, BMI and abdominal circumference was found to be a risk factor for hypertension (Jo et al., 2001). Elsewhere in Asia, the prevalence of overweight and hypertension was most common in Japan, followed by Iran, urban India, Singapore, urban Sri Lanka, and urban Philippines (Singh et al., 2010). Waist-hip ratio, which is another anthropometric measurement of obesity, has shown strong correlation with hypertension.

Risk factors identified were not the same in all the studies conducted in different places and it emphasized the need for identification of risk factors in the specific area for better prevention and control of hypertension and its consequences.

CONCLUSION

Over the last decade the logistic regression model has become a standard method of analysis in many situations. Logistic regressions are used extensively in the medical research and also help to make medical decisions. The present investigation were associated with performance the human blood pressure reacts to the influence of certain factors to determine the probability of a person to eventual incidence of hypertension. Performances of direct human blood pressure have a significant impact on heart disease.

Through multiple logistic regression analysis, there are three factors that most significant of the six factors tested were identified as having influence significantly the performance of human blood pressure (hypertension). These factors are age, body mass index (BMI) and systolic. These factors are selected based on the criteria to achieve the level of model significance (p-value < 0.05). As a conclusion, all three of these factors can affect the performance of blood pressure at risk for hypertension.

REFERENCES

A Rashid K, KA Narayan, Azizah Hj Ab Manan. The Prevalence of Hypertension among the Elderly in Fourteen Villages in Kedah, Malaysia. Malaysian Journal of Medicine and Health Sciences Vol. 4(2) June 2008:33-39.

Brundtland GH. 2002. From the World Health Organization. Reducing risks to health, promoting healthy life. Jama. 288:1974.

D. W. Hosmer and S. Lemeshow, Applied logistic regression, second edition, John Wiley and Sons, 2000.

Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ; Comparative Risk Assessment Collaborating Group. Selected major risk factors and global and regional burden of disease. Lancet 2002; 360: 1347-60.

Haberman, S. (1978). Analysis of qualitative data (Vol. 1). New York: Academic Press.

Jo I, Ahn Y, Lee J, Shin KR, Lee HK, Shin C. Prevalence, Awareness, Treatment, Control and Risk Factors of Hypertension in Korea: the Ansan Study. Journal of Hypertension 2001; 9:1523-1532.

Joshi and Pooja V. 2010. Prevalence of hypertension in the U.S. adult population and its association with various behavioral risk factors**.** San Diego State University Library & Information Access.

Kotsis V, Stabouli S, Bouldin M, Low A, Toumanidis S, Zakopoulos N. Impact of Obesity on 24-Hour Ambulatory Blood Pressure and Hypertension. Hypertension 2005; 45:602-607.

Lawes CM, Rodgers A, Bennett DA, Parag V, Suh I, Ueshima H. Blood Pressure and Cardiovascular Disease in the Asia Pacific Region. Journal of Hypertension 2004; 21(4):707-16.

N. N. Naing. 2003. Determination of sample size. Malaysian Journal of Medical.

National Nutrition Monitoring Bureau, National Institute of Nutrition. Diet & nutritional status of population and prevalence of hypertension among adults in rural areas. NNMB Technical Report No. 24. Hyderabad: National Institute of Nutrition; 2006.

Patricia K, Megan W, Kristi R, Paul K, Whelton P, Jiang H. Worldwide Prevalence of Hypertension: A Systematic Review. Journal of Hypertension 2004; 22(1):11-19.

Primatesta P, Poulter NR. Improvement in Hypertension Management in England: Results from the Health Survey for England 2003. Journal of Hypertension 2006; 24(6):1187-92.

Rankinen T, Church TS, Rice T, Bouchard C, Blair SN. Cardiorespiratory Fitness, BMI, and Risk of Hypertension: The HYPGENE Study. Medicine & Science in Sports and Exercise 2007; 39(10):1687-92.

Schlesselman, J. J. (1982). Case control studies: Design, control, analysis. New York: Oxford University Press.

Singh RB, Suh IL, Singh VP, Chaithiraphan S, Laothavorn P,Sy RG. Hypertension and Stroke in Asia: Prevalence, Control and Strategies in Developing Countries for Prevention. Journal of Human Hypertension 2010; 14(10/11), 749-763.

S. K. Lwanga and S. Lemeshow, sample size determination in health studies: A practical manual. World Health Organization, 1991, 1-3.

Tanji JL, Champlin JJ, Wong GY, Lew EY, Brown TC, Amsterdam EA. Blood pressure recovery curves after submaximal exercise: a predictor of hypertension at ten-year follow-up. Am J Hypertens. 1989; 2: 135–138.

Thomas JW, Ramachandran SV. Epidemiology of Uncontrolled Hypertension in the United States. Circulation 2005; 112: 1651- 1662.

Truett J, Cornfield J, Kannel W. 1967. A multivariate analysis of the risk of coronary heart disease in Framingham. J Chron Dis 20: 511–524.

W. M. A. W. Ahmad, N. A. Aleng and Zalila Ali, binary logistic regression analysis technique used in analyzing the categorical data in education science: a case study of Terengganu state, Malaysia. World Appl. Sci. J.,9(9), 2010, 1062-1066.

W. M. A. W. Ahmad, N. A. Aleng, Zalila Ali and Arif Bin Awang Nawi, 2011. modelling associated factors of hiv-infected tuberculosis (tb) patients using path model analysis. World Appl. Sci. J., 12(9), 2011, 1580-1584.

Received: February 21, 2014