

Identifying Risk Factors for Female Cardiovascular Disease Patients in Malaysia: A Bayesian Approach

^{1,2}Nurliyana Juhan*, ³Yong Zulina Zubairi, ⁴Zarina Mohd Khalid
and ⁵Ahmad Syadi Mahmood Zuhdi

^{1,4}Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

²Preparatory Centre of Science and Technology,
Universiti Malaysia Sabah, 88400 Sabah, Malaysia

³Centre for Foundation Studies in Science,
University of Malaya, Kuala Lumpur, Malaysia

⁵Cardiology Unit,
University Malaya Medical Centre, 50603 Kuala Lumpur, Malaysia

*Corresponding author: yzulina@um.edu.my

Article history

Received: 16 October 2018

Received in revised form: 4 December 2018

Accepted: 17 December 2018

Published on line: 31 December 2018

Abstract Cardiovascular disease (CVD) includes coronary heart disease, cerebrovascular disease (stroke), peripheral artery disease, and atherosclerosis of the aorta. All females face the threat of CVD. But becoming aware of symptoms and signs is a great challenge since most adults at increased risk of cardiovascular disease (CVD) have no symptoms or obvious signs especially in females. The symptoms may be identified by the assessment of their risk factors. The Bayesian approach is a specific way in dealing with this kind of problem by formalizing a priori beliefs and of combining them with the available observations. This study aimed to identify associated risk factors in CVD among female patients presenting with ST Elevation Myocardial Infarction (STEMI) using Bayesian logistic regression and obtain a feasible model to describe the data. A total of 874 STEMI female patients in the National Cardiovascular Disease Database-Acute Coronary Syndrome (NCVD-ACS) registry year 2006-2013 were analysed. Bayesian Markov Chain Monte Carlo (MCMC) simulation approach was applied in the univariate and multivariate analysis. Model performance was assessed through the model calibration and discrimination. The final multivariate model of STEMI female patients consisted of six significant variables namely smoking, dyslipidaemia, myocardial infarction (MI), renal disease, Killip class and age group. Females aged 65 years and above have higher incidence of CVD and mortality is high among female patients with Killip class IV. Also, renal disease was a strong predictor of CVD mortality. Besides, performance measures for the model was considered good. Bayesian logistic regression model provided a better understanding on the associated risk factors of CVD for female patients which may help tailor prevention or treatment plans more effectively.

Keywords STEMI; cardiovascular disease; female; risk factor; Bayesian.

Mathematics Subject Classification 62P10, 62-07, 62H12.

1 Introduction

Cardiovascular disease (CVD) often be thought of as a health problem for males is the most common cause of death for both males and females in Malaysia and worldwide [1,2]. CVD includes coronary heart disease, cerebrovascular disease (stroke), peripheral artery disease, and atherosclerosis of the aorta [3]. Most females have misperception that they are most likely to die from cancer or breast cancer [3,4]. In reality, CVD poses a greater mortality threat than breast cancer [5] and kills about 10 times more women than all cancers combined [6,7].

All females face the threat of CVD. But becoming aware of symptoms and signs is a great challenge since most adults at increased risk of cardiovascular disease (CVD) have no symptoms or obvious signs especially in females [8]. Additionally, some of the CVD symptoms in females are different from those in males [9,10]. These may be identified by the assessment of the risk factors.

The Bayesian approach is a specific way in dealing with this kind of problem by formalizing a priori beliefs and of combining them with the available observations. Bayesian approach was introduced in medical research in 1982. However, its application in clinical research has only been approved by the US Food and Drug Administration (FDA) in 2010 [11]. Bayesian approach, which is learning from evidence as it accumulates, is now applied in all major areas of medical statistics, including clinical trials, epidemiology, and decision making for new treatments or technologies.

Thus, this study aimed to identify associated risk factors in CVD among female patients presenting with ST Elevation Myocardial Infarction (STEMI) in Malaysia using Bayesian logistic regression model and obtain the best model to describe the data. The organization of this study is as such; it starts with a brief information on CVD and Bayesian approach. Next is the results of proposed Bayesian logistic regression model and model overall fit. A discussion and the implication of the findings of the analysis is given in the proceeding section.

2 Materials

In Malaysia, a registry supported by the Ministry of Health (MOH) known as the National Cardiovascular Disease-Acute Coronary Syndrome (NCVD-ACS) registry, plays an important role in collecting and providing information about CVD across Malaysia [12]. In this study, a total of 874 female patients who were diagnosed with ST-elevation myocardial infarction (STEMI) were selected from the NCVD-ACS registry for the years 2006 to 2013. Data was collected from the time the female patient with STEMI was admitted to the hospital till 30 days post discharge.

Demographic variable such as patients' ethnicity was divided into four major ethnic groups namely Malay, Chinese, Indian and others. Age was categorized into two groups; age <65 years and age ≥ 65 years. The risk factors were diabetes mellitus, hypertension, smoking status, dyslipidaemia and family history of CVD. While myocardial infarction (MI) history, chronic lung disease, cerebrovascular disease, peripheral vascular disease and renal disease were classified as the comorbid variables. Killip class as the clinical presentation was divided into four classes. As for the treatment, the variables were percutaneous coronary intervention (PCI) and cardiac catheterization.

3 Methods

The Bayesian MCMC approach was applied to develop suitable models for CVD patients in Malaysia. This approach, in short, is a way to associate the past (prior) with present (current study) to make decisions about the future (posterior) [13]. Therefore, it was used to identify significant risk factors and to obtain the posterior estimates of the proposed model.

For the Bayesian model in this study, likelihood for the outcome variable was defined as a logistic regression model with "1" representing death and "0" alive or otherwise. Mathematically, logistic regression estimates a multiple linear regression function defined as:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n \quad (1)$$

where p is the probability of presence of the characteristic of interest, β_i is the logistic regression coefficients and x_i is the independent variables for $i = 1, \dots, n$.

For the type of prior, non-informative prior was applied due to lack of information on the regression parameters. Simulations were run for three multiple parallel chains with different initial point for each of it.

Univariate analysis was performed on all variables using Bayesian MCMC approach. Variables that satisfied the screening criteria were then fitted into the Bayesian multivariate logistic model. Results of the analysis were given in the form of odds ratio (OR) where it represents the odds that an outcome will occur given a specific contact, compared to the odds of the outcome occurring in the absence of that contact [14]. The standard error (SE) of the analysis were also obtained. All the analysis were performed using JAGS, a software that applies Gibbs sampling method in estimation of model parameters. The conditional distributions for β_i were calculate using the Gibbs sampling algorithm. The Gibbs sampling algorithm is used to generate an instance from the distribution of each variable in turn, conditional on the current values of the other variables [15]. Area under receiving operating characteristics curve (AUC) was used to measure model discrimination [16], while Hosmer-Lemeshow test was used to measure model calibration [17].

4 Results

Descriptive statistics was performed for female patients and the results are shown in Table 1. Majority of the STEMI female patients were from ethnic Malay with more than 50% followed by Chinese (20%) and Indian (19.9%). Predominantly, female patients fell into the less than 65-year-old age group. The most prevalence risk factor for STEMI female patients was hypertension (74.5%), followed by diabetes mellitus (55.6%) and dyslipidaemia (35.8%) respectively. Most of the STEMI patients were in Killip class I or II on presentation. As for the treatment, cardiac catheterization was the most undergone procedure.

Bayesian univariate analysis are performed for female patients and seven variables are considered significant namely age group, smoking status, family history of CVD, dyslipidaemia, MI history, renal disease and Killip class. The seven significant variables are then fitted into Bayesian multivariate model.

From the Bayesian multivariate analysis shown in Table 2, six variables are observed to be significantly associated with mortality of CVD female patients namely dyslipidaemia, MI

Table 1: Female Patients' Characteristics

Characteristic	<i>n</i> = 874 (%)	
Ethnicity	Malay	489 (55.9)
	Chinese	175 (20.0)
	Indian	174 (19.9)
	Others	36 (4.1)
Age group	<65	495 (56.6)
	≥ 65	379 (43.4)
Diabetes Mellitus	No	388 (44.4)
	Yes	486 (55.6)
Hypertension	No	223 (25.5)
	Yes	651 (74.5)
Smoking status	Never	786 (89.9)
	Active/former	88 (10.1)
Dyslipidaemia	No	561 (64.2)
	Yes	313 (35.8)
Family history of CVD	No	792 (90.6)
	Yes	82 (9.4)
MI History	No	784 (89.7)
	Yes	90 (10.3)
Chronic lung disease	No	859 (98.3)
	Yes	15 (1.7)
Cerebrovascular disease	No	842 (96.3)
	Yes	32 (3.7)
Peripheral vascular disease	No	870 (99.5)
	Yes	4 (0.5)
Renal disease	No	832 (95.2)
	Yes	42 (4.8)
Killip Class	Class I	521 (59.6)
	Class II	219 (25.1)
	Class III	44 (5.0)
	Class IV	90 (10.3)
PCI	No	629 (72.0)
	Yes	245 (28.0)
Cardiac catheterization	No	573 (65.6)
	Yes	301 (34.4)

Table 2: Variables in the Multivariate Model for Female Patients

Variable	Posterior Mean	SE	OR	(95% Credible Interval)
Dyslipidaemia	−0.666	0.048	0.514	(0.310, 0.866)
MI History	0.626	0.062	1.870	(0.980, 3.490)
Smoking	−0.984	0.089	0.374	(0.141, 0.979)
Renal disease	0.827	0.077	2.286	(1.026, 3.938)
Killip class II	0.851	0.054	2.342	(1.362, 3.088)
Killip class III	1.120	0.083	3.065	(1.274, 7.308)
Killip class IV	2.768	0.057	15.927	(9.291, 20.995)
Age (≥ 65)	1.079	0.045	2.942	(1.819, 4.731)

history, smoking, renal disease, Killip class and age group of the patients. The OR indicated that female patients with dyslipidaemia were less likely to die (OR = 0.51). Besides, mortality for female patients with MI history were 1.87 times higher than those without it. Surprisingly, smoking patients were less likely to die than those non-smokers. The risk of death of those with renal disease were 2.29 times more than those without it. As expected, patients in Killip class IV were 15.9 times more likely to die than those from Killip class I. Also, risk of mortality was 2.94 times higher for female patients from the age group ≥ 65 than those from the age group <65 .

Results validation and model performance were performed using another 374 female patients' datasets. Discrimination in the Bayesian model was considered good with an AUC of 0.8012. Also, the Hosmer-Lemeshow test showed a good fit as the p -value is greater than 0.05

5 Discussion

This study has shown that Bayesian approach was capable of developing a good prognostic model and identifying the risk factors of CVD. Malaysia has an increasing prevalence of CVD risk factors in the population. Six significant variables from the Bayesian model were associated with the mortality of CVD female patients namely dyslipidaemia, MI, smoking, renal disease, Killip class and age group of the patients.

Dyslipidemia is mostly prevalent in women, especially at post-menopausal age and it has been closely related with CVD risk [18]. In line with the previous study, this study found that dyslipidaemia is one of the significant risk factors of CVD among Malaysian women. Besides, dyslipidaemia is always associated with obesity. As the degrees of obesity increase, there is an increase likelihood of developing dyslipidaemia so thus CVD in women [19].

The next risk factor was MI history, the OR suggested that females with MI history were more likely to die than those without it. This is similar with other studies, regardless of age, within a year of the first MI, more women than men died, have heart failure, or suffered from a stroke [20,21].

Smoking is always associated with a higher mortality risk [20-22]. However, in Malaysia this is not always the case as the OR in this study suggested that those who were smoker were less likely to die than non-smoker. This could be explained by a phenomenon called the smoker's

paradox where smokers had lower observed in-hospital mortality compared with nonsmokers [23-25].

In this study, higher mortality risk was observed for patients with renal disease. This is supported by other studies where renal disease or decreased kidney function is known as an independent risk factor for the development of new CVD or worsening the existing CVD [3,26,27]. Besides, patients in Killip class IV have higher risk of death as compared to patients in Killip class I. This result was similar to previous studies worldwide where patients in Killip class IV have higher number of admission and significantly higher in-hospital mortality rate than the Killip I patients [28-31].

Other CVD risk factor is age where in this study, female patients with increasing age have higher mortality. This is consistent with the findings of other studies whereby at older age or post-menopausal age, women tend to have higher risk of dying as there is less protection from the estrogen hormone [32,33]. In fact, each year CVD causes a larger number of deaths in older women than in older men [32].

Adopting Bayesian approach allowed us to incorporate prior knowledge into the current study. As in this study, non-informative prior was able to produced better prediction of risk factors of CVD, similar results were found in previous studies [35-36]. Furthermore, Bayesian approach provided more natural results and interpretations, especially in terms of being able to make probability statements of the variables of interest [37].

Also, the credible interval produced by Bayesian analysis offered a more direct expression of uncertainty [38]. However, there is a practical disadvantage of Bayesian approach whereby the long computational time since iterative sampling techniques are used [39].

6 Conclusion

In conclusion, Bayesian logistic regression model provided a better understanding on the associated risk factors of CVD for female patients which may help tailor prevention or treatment plans more effectively.

Acknowledgements

The authors would like to thank all medical and non-medical staffs who participated in the Malaysian National Cardiovascular Disease Database-Acute Coronary Syndrome registry. The authors also wish to thank the Ministry of Education Malaysia with vote no. 07397.

References

- [1] Department of Statistics Malaysia. *Statistics on Causes of Death, Malaysia, 2017 Deaths*. 2017.
- [2] WHO. WHO | Cardiovascular diseases (CVDs) [Internet]. WHO. World Health Organization; 2017. <http://www.who.int/mediacentre/factsheets/fs317/en/>
- [3] Miller P. Women and cardiovascular disease: what can health care providers do to reduce the risks? *N. C. Med. J.* 2016. 77(6):406–9.

- [4] Scheideler JK, Taber JM, Ferrer RA, Grenen EG, Klein WMP. Heart disease versus cancer: understanding perceptions of population prevalence and personal risk. *J. Behav. Med.* 2017. 40(5): 839–45.
- [5] Mehta, L. S., Watson, K. E., Barac, A. Beckie, T. M., Bittner, V. and Cruz-Flores, S. Cardiovascular disease and breast cancer: where these entities intersect: a scientific statement from the American Heart Association. *Circulation.* 2018. 137(8): 30–66.
- [6] Bairey Merz, C. N., Andersen, H., Sprague, E., Burns, A., Keida, M. and Walsh, M. N. Knowledge, attitudes, and beliefs regarding cardiovascular disease in women: The Women’s Heart Alliance. *J. Am. Coll. Cardiol.* 2017. 70(2): 123–32.
- [7] Center for Disease Control and Prevention (CDC). Leading Causes of Death in Females 2013 - Women’s Health - CDC. 2013. <http://www.cdc.gov/women/lcod/2013/index.htm>
- [8] Gonsalves, C. A., McGannon, K. R., Schinke, R. J. and Pegoraro, A. Mass media narratives of women’s cardiovascular disease: a qualitative meta-synthesis. *Health Psychol. Rev.* 2017. 11(2): 164–78.
- [9] Legato, M. J., Johnson, P. A. and Manson, J. E. Consideration of sex differences in medicine to improve health care and patient outcomes. *JAMA.* 2016. 316(18): 18–65.
- [10] Mosca L, Barrett-Connor E, Kass Wenger N. Sex/gender differences in cardiovascular disease prevention: What a difference a decade makes. *Circulation.* 2011. 124(19): 2145–2154.
- [11] FDA. *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. Guidance for Industry and FDA Staff.* 2010.
- [12] Wan Ahmad, W. A. and Sim, K. H. *National Cardiovascular Disease Database (NCVD) Report of the Acute Coronary Syndrome (ACS) Registry 2006.* Kuala Lumpur: National Cardiovascular Disease Database. 2006.
- [13] Wen, Y. H., Wu, S. S., Lin C.H. R., Tsai, J. H., Yang P. and Chang, Y. P. A Bayesian approach to identifying new risk factors for dementia. *Medicine (Baltimore).* 2016. 95(21): e3658.
- [14] Szumilas, M. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry.* 2010. 19(3): 227–229.
- [15] Gearhart, C. and Kasturiratna, D. Implementation of Gibbs sampling within Bayesian inference and its applications in actuarial science. *SIAM Journal.* 2018. 219–231
- [16] Hanley A. J. and McNeil, J. B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982. 143:29–36.
- [17] Demler, O. V., Paynter, N. P. and Cook, N. R. Tests of calibration and goodness-of-fit in the survival setting. *Stat Med.* 2015. 34(10):1659–80.
- [18] Cífková, R. and Krajoviechová. A. Dyslipidemia and cardiovascular disease in women. *Curr. Cardiol. Rep.* 2015. 17(7).
- [19] Chomistek, A. K., Chiuvé, S. E., Eliassen, A. H., Mukamal, K. J., Willett, W. C. and Rimm, E. B. Healthy lifestyle in the primordial prevention of cardiovascular disease among young women. *J. Am. Coll. Cardiol.* 2015. 65(1): 43–51.

- [20] Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J. and Cushman, M. Executive summary: heart disease and stroke statistics-2015 update. *Circulation*. 2015. 131(4): 434–41.
- [21] Mehta, L. S., Beckie, T. M., DeVon, H. A., Grines, C. L., Krumholz, H. M. and Johnson, M. N. Acute myocardial infarction in women: a scientific statement from the American Heart Association. *Circulation*. 2016. 133: 916-947.
- [22] American Diabetes Association AD. Standards of medical care in diabetes-2015 abridged for primary care providers. *Clin Diabetes*. 2015. 33(2): 97–111.
- [23] Gupta, T., Kolte, D. , Khera, S., Harikrishnan, P., Mujib, M. and Aronow, W. S. Smoker’s paradox in patients with ST-segment elevation myocardial infarction undergoing primary percutaneous coronary intervention. *J. Am. Heart Assoc*. 2016. 5(4): e003370.
- [24] Symons, R., Masci, P. G., Francone, M., Claus, P., Barison A. and Carbone, I. Impact of active smoking on myocardial infarction severity in reperfused ST-segment elevation myocardial infarction patients: the smoker’s paradox revisited. *Eur. Heart J*. 2016. 37(36): 2756–2764.
- [25] Ali, S. F., Smith, E. E., Reeves, M. J., Zhao, X., Xian, Y. and Hernandez, A. F. Smoking Paradox in Patients Hospitalized With Coronary Artery Disease or Acute ischemic stroke. *Circ. Cardiovasc Qual. Outcomes*. 2015. 8(6 suppl 3): 73–80.
- [26] Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O’Callaghan, C. A. and Lasserson, D. S. Global prevalence of chronic kidney disease - a systematic review and meta-analysis. In Remuzzi, G. (Ed.). PLoS One. 2016. 11(7): e0158765.
- [27] Mills, K. T., Xu, Y., Zhang, W., Bundy, J. D., Chen, C. S. and Kelly, T. N. A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. *Kidney Int*. 2015. 88(5): 950–7.
- [28] Milwidsky, A., Greidinger, D. and Topilsky, Y. *Echocardiographic Killip Classification*. <https://events.eventact.com/dan/29113/114953.pdf>
- [29] Taguchi, E., Konami, Y., Inoue, M., Suzuyama, H., Kodama, K. and Yoshida, M. Impact of Killip classification on acute myocardial infarction: data from the SAIKUMA registry. *Heart Vessels*. 2017. 32(12): 1439–1447.
- [30] Fach, A., Bünger, S., Zabrocki, R., Schmucker, J., Conradi, P. and Garstka, D. Comparison of outcomes of patients with ST-Segment elevation myocardial infarction treated by primary percutaneous coronary intervention analyzed by age groups (85 years); (results from the Bremen STEMI Registry). *Am. J. Cardiol*. 2015. 116(12): 1802–9.
- [31] Shiraishi, J., Kohno, Y., Nakamura, T., Yanagiuchi, T., Hashimoto, S. and Ito, D. Predictors of in-hospital outcomes after primary percutaneous coronary intervention for acute myocardial infarction in patients with a high Killip class. *Intern Med*. 2014. 53(9): 933–939.
- [32] Ley, S. H., Li, Y., Tobias, D. K., Manson, J. E., Rosner, B. and Hu, F. B. Duration of reproductive life span, age at Menarche, and age at menopause are associated with risk of cardiovascular disease in women. *J. Am. Heart Assoc*. 2017 6(11): e006713.
- [33] Gerbarg, P. L. and Brown, R. P. Pause menopause with *Rhodiola rosea*, a natural selective estrogen receptor modulator. *Phytomedicine*. 2016. 23(7): 763–769.
- [34] WHO. *10 Leading Causes of Death in Females*. WHO. 2018.

- [35] Wong, R. S. Y. and Ismail, N. A. An application of Bayesian approach in modeling risk of death in an intensive care unit. *PLoS One*. 2016. 11(3): 1–17.
- [36] Pascale, A., Nicoli, M. and Spagnolini, U. Cooperative Bayesian estimation of vehicular traffic in large-scale networks. *IEEE Trans Intell Transp Syst*. 2014 . 15(5): 2074–88.
- [37] Walley, R. J., Smith, C. L., Gale, J. D. and Woodward, P. Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: a case study. *Pharm. Stat*. 2015. 14(3): 205–215.
- [38] Feng, D., Cortese, G. and Baumgartner, R. A comparison of confidence/credible interval methods for the area under the ROC curve for continuous diagnostic tests with small sample size. *Statistical Methods in Medical Research*. 2017. 26(6): 2603-2621.
- [39] Torman, V. B. L. and Camey, S. A. Bayesian models as a unified approach to estimate relative risk (or prevalence ratio) in binary and polytomous outcomes. *Emerg Themes Epidemiol*. 2015. 12(1).