Penerbit Akademia Baru
Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 22, No. 1. Pages 13-22, 2016

Determinants of Comorbidity in Rheumatoid Arthritis: Influence of Demographic and Duration of Illness

S. N. Chin[*,a], *N. Abdullah*[b] *and L. P. Poh*[c]

Faculty of Science & Natural Resources,Universiti Malaysia Sabah, 88400, Kota Kinabalu, Sabah, Malaysia
[a,*]*snchin.ums@gmail.com*, [b]*norainiabdullah.ums @gmail.com*, [c]*lpohp@yahoo.com*

Abstract –*Rheumatoid arthritis (RA) is a chronic, disabling autoimmune disease which affects about 5 in 1000 people in Malaysia. Patients with RA are at increased risk of developing comorbid conditions. This research aims at determining these relationships between demographic, duration of illness and comorbidity in RA via a multiple binary logistic (MBL) regression analysis based on the 102 patients' information (23 males; 79 females) obtained from the rheumatoid clinic of the Queen Elizabeth Hospital in Kota Kinabalu. The relationship of the RA patients with comorbid conditions was studied with focus on the demographic and duration of illness. The variables obtained for analysis were the comorbid conditions namely, hypertension and hyperlipidemia, age, duration of illness, gender, ethnicity, household income and education level. From six independent variables, two were quantitative would be analyzed, while four were categorical, and would be transformed into dummy variables. Four phases in a model-building approach were executed where two models were formed where Model I predicted the probability of occurrence of hypertension with age of patients and first order interaction between duration of illness before diagnosis and household income of less than RM1000 had positive effects on the model, while Model II predicted the occurrence of hyperlipidemia among the RA patients with age of patients and first order interaction variable between gender(female) and age were the contributing factors.*

Keywords: Rheumatoid arthritis, Comorbid conditions, Multiple binary logistic regression, Model building.

1.0 INTRODUCTION

As stated in [1], an estimated 1.5 million people in the United States have RA, almost 1 percent of the nation's adult population. There are nearly three times women with this disease aged between 30 and 60 years compared to men. According to Arthritis Foundation Malaysia, this disease had affect about 5 in 1000 people in Malaysia [2]. Patients with rheumatoid arthritis (RA) are at increased risk of developing comorbid conditions, therefore many research had been carried out to gain a deeper understanding by analyzing this issue. [3] had carried their study in Saudi Arabia stated that most of the RA patients had suffered from comorbid disease. They found out that the most comorbidity among the rheumatoid arthritis was hypertension with 122 patients among 340, 105 diabetes (30.9%), 88 osteoporosis (25.8%) and 66 dyslipidemia (19.4%). In a retrospective cross-study carried by [4] had also reported that diabetes mellitus, hypertension and obesity were the most common comorbidities when the cause of death and associated comorbidities were registered and compared in relation to the mortality rate among rheumatoid arthritis patients.

Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 22, No. 1. Pages 13-22, 2016

[5] had carried a study in Finland to study the prevalence and importance of comorbidities in rheumatoid arthritis patients at the time of the diagnosis and after 15 years of follow up. In their study, they found that comorbidities had increased during the 15 years of RA, and the most common comorbidities were hypertension (30%), cardiovascular diseases (14%), and malignances (11%). [6] had also found that there was a high prevalence of hypertension in RA where a total of 400 RA patients were studied with 282 (70.5%) of them suffering from hypertension.

This research was then carried out to determine whether there are relationships between demographic, duration of illness and comorbidity in RA patients of Kota Kinabalu via a multiple binary logistic (MBL) regression analysis.

2.0 METHODOLOGY

2.1 3.1 Data Sample and Data Collection

A structured Case Report Form (CRF) from the National Inflammatory Arthritis Registry (NIAR) was distributed to the medical doctors for data collection from October 2013 to January 2014. 102 patients attending their regular clinic appointments were identified and registered. Consents were obtained from patients using the Patients Confidentiality Information form. The Demographic information obtained from the patient would include age, gender, marital status, social economic status, location of residence, ethic group/race, and highest level of education completed. On top of that, diagnosis criteria, comorbid conditions, extra-articular features were also obtained. Joint count assessments were then performed by the assessing doctor while other information necessary to fill into the CRF obtained from patients' medical records. The information was then entered into the online database. The next outcome date was then determined, and this coordinated with patients' scheduled clinic visit.

In this research, the data was collected from the rheumatoid clinic of the Queen Elizabeth Hospital (QEH), Kota Kinabalu, Sabah starting from October 2013, and the data collection period continued for 3 months. There would be at least 3 months of follow up period after the first date of notification. Patients were consulted with the doctors once in three months. After three months, the data collection was repeated again. As suggested by [7], patients needed to follow up their clinics every 3 to 6 months. Data were then obtained from the registry through QEH and interpreted according to the data definition provided by the registry.

2.2Multiple Binary Logistic Regression

Logistic regression is used to predict an outcome from a set of predictor variables but the dependent variable of a logistic regression is usually dichotomous, that is, the dependent variable takes the value of 1 with a probability of success , or 0 with probability of failure [8]. It is well suited for studying the relationships between categorical or qualitative outcome with one or more predictor variables [9]. The logistic regression model will be chosen from a set of data which requires the estimate of the values of the unknown parameters, and . Maximum likelihood method is used to estimate these values of the unknown parameter . Once the maximum likelihood estimates of and were found, we substituted these values into the response function to obtain the fitted logistic response function given as follows: . With multiple logistic regressions, estimates of vector will be obtained and the fitted multiple logistic regression shown as follows: . The Logit model is also based on the cumulative logistic distribution function which gives probability estimated bounded by 1 and 0 [10]. The Multiple Binary Logit

Journal of Advanced Research Design
ISSN (online): 2462-1943 I Vol. 22, No. 1. Pages 13-22, 2016

(MBL) Model can be represented as: ...(1), with …(2), and …(3) for which denotes the i-th probability for an event to occur where i = 1, 2,..., n and is the complement for the events above. In the model equation, , where is the binary dependant variable (viz. the occurrence of hypertension and hyperlipidemia) with '1' denoting the occurrence of hypertension and hyperlipidemia, and '0' denoting the non-occurrence of hypertension and hyperlipidemia; denotes the j-th independent variable which may exist as single independent variables, interaction variables (first order, second order, third order,...), generated variables (polynomial or dummy variable) and transformed variables (Ladder Transformation, Box-Cox Transformation); denotes the constant term for the model and denotes j-th coefficient of independent variable, and u denotes the error term or residuals of the model. The model has k number of independent variables and (k +1) denotes the number of parameters, which is the number of independent variables including the constant term.

The four phases in model building in the multiple regression analysis had been introduced by [11]. The model-building approach for the multiple binary logistic model based on the Deviance and Pearson Chi-square Tests has been introduced by [12]. However, in this work, Phase 3 and Phase 4 of the model-building procedures differ and can be briefly explained as follows:

Phase 1: All Possible Models

All the possible models consisting of the different combinations of variables and their interactions are listed out for consideration as in [10]. Table 1 summarizes the number of possible models for 2 single independent quantitative variables. It can be seen that only models up to first order interactions were formed since higher interaction orders of dummy variables are not relevant.

Table 1: Summary of all possible models

Number of Variables	Individual	Interactions Variables		
		First Order	Second Order	Total
1	1	1	-	2
2	1	1	-	2
Total	2	2	-	4
Models	M1-M2	M3-M4		

Phase 2: Selected Models

All the possible models listed out in Phase 1 will go through a variable filtering process so as to find these selected models[10]. The parameter tests would include the Global test, Multicollinearity test, Coefficient test and the Wald test. The models in each stage will be labeled as in Fig.1, where M stands for the multiple binary logit model, a is the model number, b is the number of multicollinearity source variables removed in the multicollinearity test, and c is the number of insignificant variables eliminated from the coefficient test. Before carrying out the logistic regression test, assumption of multicollinearity has to be tested in order to reduce the error of model. Presence of multicollinearity was detected using Variance Inflation Factor (VIF). In this test, the variables with VIF more than 10 are deleted as the value of VIF exceed 10 is frequently taken as indication that multicollinearity may be unduly influencing the least squares estimates [13].

The coefficient test is performed on the reduced model after the multicollinearity source variables removals. It is an elimination procedure of insignificant variable by using the

Journal of Advanced Research Design
ISSN (online): 2462-1943 I Vol. 22, No. 1. Pages 13-22, 2016

backward elimination method [14]. After all the selected models are obtained, the Wald test is carried out to justify the removal of the insignificant variables [10].

M.a.b.c

Figure 1: Model labeling [11]

Phase 3: Regression Model Selection

In this work, the regression model is chosen by testing the significance of the coefficients in the model. This process is done by testing whether the independent variables are significantly contributing to the dependent variables. In this case, the observed values of the response variable are compared to predicted values obtained from models with and without the variable. If the predicted values with the variable in the model are better than when the variable is not in the model, then the variable in the model is "significant". As stated by [15], comparison of observed to predict value was based on the log likelihood function as:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1 - y_i)\ln[1 - \pi(x_i)]\} \ldots(4)$$

The comparison of observed to predict values using the likelihood function is based on the equation of $D = -2\ln\left[\frac{\text{likelihood of the fitted model}}{\text{likelihood of the saturated model}}\right] \ldots(5)$. The quantity inside the bracket in the expression (5) is called likelihood ratio. According to [15], using minus twice its log was necessary to obtain a quantity whose distribution was known and can therefore be used for the purpose of hypothesis testing. The test is known as the likelihood ratio test. Equating equation (4) into (5), giving: $D = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1 - y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \ldots(6)$ where $\hat{\pi}_i = \hat{\pi}_i(x_i)$. The statistic D in equation (6) is called deviance which plays the same role as that of sum-of-squares in linear regression. The calculation of log likelihood and the likelihood ratio test are the features in the logistic software and this had made our lives easier. In the simple case of a single independent variable, the model is fitted first with only constant term, and then fit the model with the independent along with the constant. This gives rise to the new log likelihood. The likelihood ratio test is then obtained by multiplying the difference of these two values by -2.

Phase 4: Model Goodness of Fit

Goodness of fit is the most important aspect of logistic regression analysis for testing whether the model fits well or not. [15] proposed various methods to test the fit of model and adequacy, which are (1) computation and evaluation of overall measure of fit, (2) examination of individual components of summary statistic, and (3) examination of other measures of the difference between observed and fitted values. In this research, the Hosmer-Lemeshow and classification tables are used to test for the goodness of fit of model.

Hosmer-Lemeshow Test

Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 22, No. 1. Pages 13-22, 2016

This is the one of the most common tool conveniently used in logistic regression analysis. It is first done by sorting the observations in increasing order of their estimated event probabilities and then divided into ten groups on the basis of estimated probabilities. According to [15], two grouping strategies were proposed:

i. Collapse the table based on percentiles of estimated probabilities.
ii. Collapse the table based on fixed values of the estimated probability.

With the first method as mentioned, used of g=10 groups resulted in the first group containing $n_1' = n/10$ subjects having the smallest estimated probabilities, and the last group containing $n_{10}' = n/10$ subjects having the largest probabilities. With second method, use of g =10 groups results in cutpoints defined at $k/10$, $k = 1,2,...,9$ and the groups contain all subjects with estimated probabilities between adjacent cutpoints. For both grouping strategies, the Homer-Lemeshow goodness-of-fit statistic, $\hat{C}$, is obtained by calculating the Pearson chi-square statistic from $(2 \times g)$ table of observed and estimated frequencies, where g is the number of groups. The statistic is given by $\hat{C} = \sum_{i-1}^{g} (O_i - n_i\bar{\pi}_i)^2 / n_i\bar{\pi}_i(1-\bar{\pi}_i)$ …(7) where n_i is the total number of subjects in the i^{th} group, O_i is the observed number of events in the i^{th} group, and $\bar{\pi}_i$ is the average estimated probability of an event in the i^{th} group.

Classification Table

A classification table shows the actual measurement and the classification for the data. If the model fits well, most individuals from the sample fall into the correctly predicted categories. To measure the predictive power of the model finds sensitivity, specificity, and the overall proportion of correct classifications. In order to obtain derived dichotomous variable, we must first define a cut point, c and compare each estimated probability to c, value commonly used for c is 0.5 [15]. If the probability for success, $\hat{\pi}(x)$ exceeds c, then we let the derived variable be equal to 1, then it is predicted that a success for the individuals. Likewise, if $\hat{\pi}(x)$ is smaller than c, then one would predict a failure for this individual. Therefore each individual is classified as success if $\hat{\pi}(x) \geq 0.5$ and as a failure if $\hat{\pi}(x) < 0.5$.

3.0 RESULTS AND DISCUSSION

3.1 Data Preparation, Transformation and Description

Data were obtained from 102 RA patients of the Queen Elizabeth Hospital (QEH) in Kota Kinabalu in October 2013 to January 2014, with 79 (77.45%) females and 23(22.55%) male patients. There were 29 Chinese, 44 Kadazan Dusun patients and 29 other ethnic patients regardless of gender. There were six different types of information contained in the data set consisting of personal details such as age, race (ethnicity) and gender, and socio-economic information, such as household income and education level, were encoded into a categorical form. However, in this study, these information had been selected with a dependent variable (type of comorbid condition i.e. hypertension or hyperlipidaemia), two quantitative (age in years and duration of illness) and four qualitative independent variables (viz. gender, ethnicity, household income and education level). In this research, ethnicity of other ethnic, no income and no formal education was defined as the reference group. The four qualitative independent variables were transformed into dummy variables of values ‘0’ and ‘1’. For practicality,

Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 22, No. 1. Pages 13-22, 2016

dummy variables with zero occurrences, and low frequency of cases or of small variances were removed so as to reduce the chances of multicollinearity. Table 2 showed the data variables used and their descriptions in this study.

3.2 Demographic Profiles

There were 79 (77.45%) female patients and 23 (22.55%) male patients with RA with comorbid conditions. 102 patients that were registered. From the total of 102 patients, there were 79 female patients and 23 male patients. It consisted of 29 Chinese patients, 44 Kadazan Dusun patients, and 29 other ethnic patients regardless of gender. It was more than 3 times compared to the male patients as similar to the study of [1]. Another study by [16] found that ratio of women to men of RA population in Spanish was slightly higher with the ratio of 4:1.

Table 2. Data Variables and Their Descriptions

Variables	Description	Variable Type
Age	Age in years	Continuous
Duration of illness	Duration of illness before diagnosis in months	Continuous
Gender	1. Male 2. Female	Categorical
Ethnicity	1. Chinese 2. Kadazan Dusun 3. Others	Categorical
Household income	1. No income 2. < RM1000 3. RM1001-RM3000 4. >RM3001	Categorical
Education level	1. No formal education 2. Primary school 3. Secondary school 4. Tertiary	Categorical

Mean age of the patients was 54.2years. The youngest was 20 and the oldest was 85 years old. This was similar with the study by [3] where the average patient age was 53.5 ± 11.3 years. The age range with the most patients was in the 51 to 60 years category, which is 37.25%. Kadazan Dusun was the biggest ethnic group with 43.14% in the registry. Chinese and other Malaysians were equally distributed with 28.43% respectively. Other Malaysians included Bajau, Malay, Iban, and India. Although ‘Malay’ and ‘India’ were the biggest ethnicity in Malaysia, however they were included in the category of ‘Other Malaysian’ because the population of ‘Malay’ and ‘India’ were very small in Sabah compared to Chinese and Kadazan Dusun. The finding was slightly different from the previous study carried by [17] who found Chinese had highest prevalence of rheumatoid arthritis with 33.6% among 128 patients.

Household income of the 102 rheumatoid arthritis patients was categorized into 4 groups, which were no income, less than RM1000, RM1001 to RM3000, RM3001 and above. However, there was one unknown income in this sample and was dropped since it did not contribute to the analysis. In this study, most of the patients had lower social economic group. Almost 70% of them had an income less than RM3000; patients with less than RM 1000 were 44.55%. This finding was similar with the study of [18] who stated that RA was more prevalence in the lower social economic group. Among the RA patients, more than half (53.54%) were having only up to secondary school education and below. However, among these 102 patients, there were 3 unknown education level and these had been excluded in the analysis. The results were consistent with the study by [19] who found inverse relationship between level of education and risk of developing RA.

3.3 Clinical Profiles

More than half of patients were diagnosed late, which was more than a year after the symptoms had onset. Among the patients, the maximum duration of patients who visited to the clinic after symptom onset was 337 months, which was about 28 years after the symptom had onset. This finding was close to the report by [20] who stated that 40% among 25 patients were diagnosed more than 6 months after symptom onset.

Most of the patients were also diagnosed late in almost every income groups. 44.12% of the patients were in category of income less than RM1000, two third of them were diagnosed late, which was more than a year after the symptom onset. 58.49% of patients who were in category of secondary had diagnosis of more than 12 months after the symptom onset. The duration of illness before diagnosis was shorter in patients who had higher education level where 46.67 % of patients who had tertiary education level were diagnosed less than 6 months after the symptom onset. They were more alert on the prevalence of the disease. Various education level categories had shown other duration of illness before diagnosis. However, in previous study by [21], the mean duration of symptoms was the same in all education levels. But they concluded that difference in severity of RA between patients with different education levels were present in the early stage of the RA disease.

3.4 Multiple Binary Logistic (MBL) Regression Models

The model-building procedures of Phase 1 to Phase 4 were carried out on the all possible models [12]. After the multicollinearity test, the backward elimination procedures of the coefficient test were then carried out to eliminate any insignificant variables in the models. Two models were obtained each representing hypertension (Model 1) and hyperlipidemia (Model 2) respectively. After the multicollinearity test in Model 1 and Model 2, with 17 variables remaining from the original 29 variables, the models were rerun to test the significance of remaining variables. Any insignificant variable (variable with $p > 0.05$) was eliminated from the model [14]. Table 3 indicated the goodness-of-fit tests using the Hosmer-Lemeshow Test and the classification table for Model 1 and Model 2 respectively.

The Hosmer-Lemeshow test of Table 3 showed that the chi-square test yielded a p-value of 0.236 for Model 1, and a p-value of 0.724 for Model 2, thus suggesting that these models are with good predictive value.The classification table in Table 3 for Model 1 also showed that in this study, 56.5% were correctly classified for having hypertension as comorbid of RA, and 76.8% for not having hypertension as comorbid of RA. Overall 67.6% were correctly classified. On the other hand, the classification table in Table 3 for Model 2 showed that 33.3% were correctly classified for having hyperlipidemia as comorbid of RA, and 93.1% for not having hyperlipidemia as comorbid of RA. Overall 75.5% were correctly classified as having hyperlipidemia.

Table 3: Goodness-of-Fit Tests of Multiple Binary Logistic (MBL) Regression Models

Journal of Advanced Research Design
ISSN (online): 2462-1943 I Vol. 22, No. 1. Pages 13-22, 2016

Model 1	Model 2

Model 1 – Hosmer-Lemeshow Test

Step	Chi-square	Df	*P*
1	7.265	8	.508
13	10.432	8	.236

Model 2 – Hosmer-Lemeshow Test

Step	Chi-square	Df	*P*
1	5.091	8	.748
16	5.313	8	.724

Model 1 – Classification Table

<table>
<tr><th colspan="3" rowspan="3">Observed</th><th colspan="3">Predicted</th></tr>
<tr><th colspan="2">Hypertension</th><th rowspan="2">Percentage Correct</th></tr>
<tr><th>no</th><th>yes</th></tr>
<tr><td rowspan="3">Step 1</td><td rowspan="2">Hypertension</td><td>no</td><td>43</td><td>13</td><td>76.8</td></tr>
<tr><td>yes</td><td>16</td><td>30</td><td>65.2</td></tr>
<tr><td colspan="2">Overall Percentage</td><td></td><td></td><td>71.6</td></tr>
<tr><td rowspan="3">Step 13</td><td rowspan="2">Hypertension</td><td>no</td><td>43</td><td>13</td><td>76.8</td></tr>
<tr><td>yes</td><td>20</td><td>26</td><td>56.5</td></tr>
<tr><td colspan="2">Overall Percentage</td><td></td><td></td><td>67.6</td></tr>
<tr><td colspan="6">a. The cut value is .500</td></tr>
</table>

Model 2 – Classification Table

<table>
<tr><th colspan="3" rowspan="3">Observed</th><th colspan="3">Predicted</th></tr>
<tr><th colspan="2">Hyperlipidaemia</th><th rowspan="2">Percentage Correct</th></tr>
<tr><th>no</th><th>yes</th></tr>
<tr><td rowspan="3">Step 1</td><td rowspan="2">Hyperlipidemia</td><td>no</td><td>66</td><td>6</td><td>91.7</td></tr>
<tr><td>yes</td><td>16</td><td>14</td><td>46.7</td></tr>
<tr><td colspan="2">Overall Percentage</td><td></td><td></td><td>78.4</td></tr>
<tr><td rowspan="3">Step 16</td><td rowspan="2">Hyperlipidemia</td><td>no</td><td>67</td><td>5</td><td>93.1</td></tr>
<tr><td>yes</td><td>20</td><td>10</td><td>33.3</td></tr>
<tr><td colspan="2">Overall Percentage</td><td></td><td></td><td>75.5</td></tr>
</table>

The MBL regression model equations for Model 1 and Model 2 are given as follows:

Model 1 (Hypertension):

$$\ln\left(p_i/1-p_i\right)=-2.943+0.074x_1-1.934d_4-1.418d_8-0.043x_1d_9+0.016x_2d_4\ldots(8)$$

where $\ln\left(p_i/1-p_i\right)$ is the estimated probability of getting hypertension, x_1 is the age of patients in years, d_4 is the household income less than RM 1000, d_8 is the education level of secondary, x_1d_9 is the interaction of age of patients and education level of tertiary, x_2d_4 is the interaction of duration of illness before diagnosis and household income of less than RM1000. Previous study [6] revealed age, BMI and prednisolone use to be independently associated with occurrence of hypertension.

Model 2 (Hyperlipidemia):

$$\ln\left(h_i/1-h_i\right)=-6.0103+0.071x_1+0.027x_1d_1\ldots(9)$$

where, $\ln\left(h_i/1-h_i\right)$ is the estimated probability of getting hyperlipidemia, x_1 is the age of patients in years, x_1d_1 is the interaction between age and female gender.

Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 22, No. 1. Pages 13-22, 2016

4.0 CONCLUSION

Using multiple binary logistic regressions, the model on hypertension comprising of one continuous variable, two categorical variables, and two interactions of variables contributed significantly contributed to the dependent variable of hypertension. However, when using the same independent variables to test for the probability of occurrence of hyperlipidemia in rheumatoid arthritis patients, there were only two variables contributed significantly to the model, which were age and interaction of female gender with age. Both MBL models were found to have age of diagnosis was significantly contributing to the predictor of comorbid, however, the range was found to be between 20 and 85 years of age.

REFERENCES

[1] Arthritis Foundation. Rheumatoid arthritis. Available from: www.arthritis.org/conditions-treatments/disease-center/rheumatoid-arthritis/, 2014.

[2] Arthritis Foundation Malaysia. Rheumatoid arthritis. Available from: http://www.afm.org.my/info/ra.htm, 2010.

[3] Al-Bishri, J., S. M. Attar, Nawal Bassuni, Yasser Al-Nofaiey, Hamed Qutbuddeen, Salma Al-Harthi, and Sarah Subahi. "Comorbidity profile among patients with rheumatoid arthritis and the impact on prescriptions trend."Clinical medicine insights. Arthritis and musculoskeletal disorders 6 (2013): 11.

[4] Al-Ghamdi, Aisha A. "The co-morbidities and mortality rate among rheumatoid arthritis patients at the western region of Saudi Arabia." Medical Science 16, no. 3 (2009).

[5] Tiippana-Kinnunen, T., H. Kautiainen, L. Paimela, and M. Leirisalo-Repo. "Co-morbidities in Finnish patients with rheumatoid arthritis: 15-year follow-up." Scandinavian journal of rheumatology 42, no. 6 (2013): 451-456.

[6] Panoulas, Vasileios F., Karen MJ Douglas, Haralampos J. Milionis, A. Stavropoulos-Kalinglou, Peter Nightingale, Marina D. Kita, Andreas L. Tselios, Giorgos S. Metsios, Moses S. Elisaf, and George D. Kitas. "Prevalence and associations of hypertension and its control in patients with rheumatoid arthritis." Rheumatology 46, no. 9 (2007): 1477-1482.

[7] Bykerk, Vivian P., Pooneh Akhavan, Glen S. Hazlewood, Orit Schieir, Anne Dooley, Boulos Haraoui, Majed Khraishi et al. "Canadian Rheumatology Association recommendations for pharmacological management of rheumatoid arthritis with traditional and biologic disease-modifying antirheumatic drugs." The Journal of rheumatology 39, no. 8 (2012): 1559-1582.

[8] Yahaya, Aminatul Hawa, Noraini Abdullah, and Zainodin Jubok. "Logistic Regression Model of Hazardous Event on Floating Offshore Units."MATEMATIKA 29 (2013): 11-16.

[9] Peng, Chao-Ying Joanne, and Tak-Shing Harry So. "Logistic regression analysis and reporting: A primer." Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences 1, no. 1 (2002): 31-70.

Journal of Advanced Research Design
ISSN (online): 2462-1943 I Vol. 22, No. 1. Pages 13-22, 2016

[10] Jubok, Zainodin, and G. Khuneswari. "Model-building approach in multiple binary logit model for coronary heart disease." Malaysian Journal of Mathematical Sciences 4, no. 1 (2010): 107-133.

[11] Zainodin, H. J., A. Noraini, and S. J. Yap. "An alternative multicollinearity approach in solving multiple regression problem." Trends in Applied Sciences Research 6, no. 11 (2011): 1241.

[12] Noraini, A., H. J. Zainodin, and L. B. Rick. "Risk Factors Determination on UGIB Patients in Kota Kinabalu, Sabah, Malaysia." Journal of Medical Sciences 13, no. 7 (2013): 526.

[13] Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. Applied linear statistical models. Vol. 4. Chicago: Irwin, 1996.

[14] Abdullah, Noraini, Zainodin HJ Jubok, and JB Nigel Jonney. "Multiple regression models of the volumetric stem biomass." WSEAS Trans. Math 7 (2008): 492-502.

[15] Hosmer Jr, David W., and Stanley Lemeshow. Applied logistic regression. John Wiley & Sons, 2004.

[16] Carmona, Loreto, Virginia Villaverde, César Hernández-García, Javier Ballina, Rafael Gabriel, Armando Laffon, and EPISER Study Group. "The prevalence of rheumatoid arthritis in the general population of Spain."Rheumatology 41, no. 1 (2002): 88-95.

[17] Sulaiman, Wahinuddin, Ashraful Toib, Giriyappanavar Chandrashekhar, and Anwar Arshad. "The trends of DMARDS prescribed in rheumatoid arthritis patients in Malaysia." Oman medical journal 24, no. 4 (2009): 260.

[18] Jacobi, Catharina E., Geert D. Mol, Hendriek C. Boshuizen, Ines Rupp, Huibert J. Dinant, and Geertrudis AM Van den Bos. "Impact of socioeconomic status on the course of rheumatoid arthritis and on related use of health care services." Arthritis care & research 49, no. 4 (2003): 567-573.

[19] Pedersen, Merete, Søren Jacobsen, Mette Klarlund, and Morten Frisch. "Socioeconomic status and risk of rheumatoid arthritis: a Danish case-control study." The Journal of rheumatology 33, no. 6 (2006): 1069-1074.

[20] Chan, KIN-Wei A., David T. Felson, Robert A. Yood, and Alexander M. Walker. "The lag time between onset of symptoms and diagnosis of rheumatoid arthritis." Arthritis & Rheumatism 37, no. 6 (1994): 814-820.

[21] Vlieland, TP Vliet, Neeltje A. Buitenhuis, Derkjen van Zeben, Jan P. Vandenbroucke, Ferdinand C. Breedveld, and J. M. Hazes. "Sociodemographic factors and the outcome of rheumatoid arthritis in young women." Annals of the rheumatic diseases 53, no. 12 (1994): 803-806.