FASTING BLOOD SUGAR CONTROL AND TOTAL CHOLESTEROL ALGORITHM FOR DIABETES MELLITUS AND HYPERLIPIDAEMIA DISEASE PATIENTS: AN SAS METHODOLOGICAL PERFORMANCE

Wan Muhamad Amir W Ahmad1*, Nor Farid Mohd Noor2, Zainab Binti Mat Yudin3, Nor Azlida Aleng4, Ruhaya Hassan5

1,2,3,5 School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia
4 School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu (UMT), 21030 Kuala Nerus, Terengganu, Malaysia.

*Corresponding author: Wan Muhamad Amir W Ahmad, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia, wmmair@usm.my

ABSTRACT

Hyperlipidaemia disease is a common disease which is very high risk of developing heart disease while diabetes mellitus is the major risk factor of unstable coronary artery disease. Through this study, we identify the factors that related to the kind of disease by searching the factor which has a strong association among all studied variables. For such linear modeling, the use of regression are not new but in this paper an attempt is made to propose a covariate-dependent linear model to identify the factors that contribute to the estimation of Hyperlipidaemia disease model and diabetes mellitus model. This study used the data which are gained from Hospital Universiti Sains Malaysia (HUSM). The obtained model is employed to estimate the potential possibilities factors that contribute to the hyperlipidaemia disease model and diabetes disease model. This can provide a useful model for statistical inferences and other relevant characteristics.

Keywords: Linear Regression, Response Surface Methodology (RSM), Hyperlipidaemia disease, SAS.

1.0 Introduction

According to Kaul, Tarr et al, diabetes disease was first documented by the Egyptians and is characterised by weight loss and polyuria. (Kaul, Tarr et al., 2013). The chronic metabolic disorder, diabetes mellitus is a fast-growing disease global problem in all over the countries. It is estimated that in 2010 there were globally 285 million people (6.4%) suffering from this disease. This number is estimated to increase to 430 million in the absence of better control or cure (Kaul, Tarr et al. 2013). Usually patients with diabetes mellitus has a relationship with other diseases. Analyses from the Framingham Heart Study showed that men and women who
have high triglyceride levels (>1.7 mmol/liter) and a low high-density lipoprotein (HDL) level (<1.03 mmol/liter) have a higher chance of rate of coronary artery disease (Castelli, 1992). Many of studies also showed that the relationship of triglycerides to Coronary Artery Disease (CAD) is significant for the other major risk factors, such as total cholesterol, low-density lipoprotein (LDL) cholesterol, blood pressure, smoking, and glucose intolerance (Castelli, 1992).

Triglyceride level is one of the prognostic factors for Coronary Heart Disease (CHD) and it is assumed to have an effect on cholesterol metabolism. The triglyceride change had an inverse correlation with the pretreatment level of urea nitrogen \( r = -0.48 \), \( p = 0.009 \) but not with total cholesterol \(-0.20\), triglyceride itself \( r = 0.25 \). The change in triglyceride level did have a direct correlation with that of total cholesterol \( r = 0.56 \), \( p = 0.002 \), as has been mentioned. Triglyceride has no correlation with the change of creatinin reading but changes in triglyceride level tended to have a direct but not significant correlation with urea nitrogen \( r = 0.34 \), \( p = 0.07 \) (Segawa, Kataoka et al., 1998). According to Segawa, Kataoka et al., the change in total cholesterol level had direct correlations with changes in the triglyceride level \( r = 0.56 \), \( p = 0.002 \) and the urea nitrogen level \( r = 0.51 \), \( p = 0.006 \), but these changes in triglyceride and urea nitrogen level did not correlate significantly (Segawa, Kataoka et al. 1998). Triglyceride decreased in both groups during Ramadan, but cholesterol levels dropped considerably during and after Ramadan for those who concurrently engaged in physical activity and fasted \(-12.24 \) and \(-8.4\) mg/dL, respectively) Elevation of the low-density-lipoprotein (LDL) is considered a major cause of coronary heart disease (CHD). It is widely believed that the reduction of blood cholesterol levels by drug treatment can decreases the incidence of CHD among patients (Haghdoost and Poorranjbar, 2009). In Framingham Heart Study, HDL, VLDL, and LDL cholesterol as well as total cholesterol were measured in young adult and they found that parental total cholesterol levels measured to be significant predictors of present total cholesterol in the offspring. (Garrison, Castelli et al., 1992)

Hyperlipidemia means there is an excessively cholesterol in the blood. Cholesterol is produced mainly by the liver and it is an essential for healthy cell membranes, brain functioning, hormone production and vitamin storage. An excessively cholesterol in the blood can cause progressive atherosclerosis. A chronic hyperlipidemia not only in the deposition of lipids in the atheromatous lesions but that it may produce the primary endothelial injury that initiates the process of atherosclerosis as well (Ross and Harker, 1976). To overcome the problem, different type of anti-hyperlipidemic agents have been developed depending on the symptoms and other conditions of the subjects. (Ross and Harker, 1976). Briefly, a 63-year-old patient with a high load of atherosclerotic risk factors (severe hyperlipidemia (triglycerides up to 500 mg/dl; low-density lipoprotein cholesterol up to 240 mg/dl)) suffered from end-stage renal disease owing to cholesterol emboli after coronary angiography because of symptomatic coronary artery disease in October 1997(Canaud, Bragg-Gresham et al., 2006).

Statistical analysis plays an important role in the most applied research. In 1973, Goldstein, Hazzard et al., comparing the cholesterol and triglyceride levels of controls and survivors of different years and sex, lipid levels were transformed by applying linear regression analysis. The formula used to be as follows: adjusted lipid value = (observed lipid value - control
means the lipid value of appropriate age and sex) + mean lipid value at age 45 year of the appropriate sex. Statistical analysis is very important in order to determine the factor as such control mean lipid values for men and women at different ages were derived from the regression equations for the appropriate sex: \( y = ax + b \), where \( y \) = mean cholesterol concentration or mean log triglyceride concentration; \( x \) = age in years; \( a \) = average annual change of cholesterol or logo triglyceride concentration; and \( b \) = cholesterol or log triglyceride concentration when \( x = 0 \). The constants for \( a \) (coefficient of linear regression) and \( b \) (y intercept) used in the calculation of the adjusted lipid values are summarized in Table 12. Under these conditions of the regression equations, mean lipid values for both men and women were nearly equivalent at age 45 yr. Since the logarithms of triglyceride levels in controls were found to be distributed more normally than the corresponding skewed, untransformed values, the log scale was used for age and sex adjustments of triglyceride with subsequent reconversion to the arithmetic scale. (Goldstein, Hazzard et al., 1973)

### 2.0 Materials and Methods

**Part I: Methodology of Linear Modeling and Response Surface Methodology RSM**

Below is the algorithm in the SAS language for the Linear Modeling (LM) and RSM for our first Case Study I. For the Case Study II till to the Case Study IV, we need to change to the related data by substituting data and their related variable. First step, we perform Linear regression modeling and the second step, is response surface counter plot using SAS software, through the options of the SAS RSREG and SAS PROC REG algorithm which is given as follows:

Data Diabetes;
Input FBS HbA1c Total_cholesterol Triglyceride LDL HDL; /*Variable input based on the research

| Cards; |
|---|---|---|---|---|---|
| 5.900 | 6.60 | 3.67 | 1.20 | 2.02 | 1.10 |
| 8.100 | 6.10 | 3.80 | 1.60 | 2.22 | 0.85 |
| 6.100 | 6.70 | 4.27 | 1.13 | 2.31 | 1.44 |
| 7.500 | 7.90 | 5.42 | 1.26 | 3.82 | 1.03 |
| 5.800 | 7.51 | 3.60 | 0.90 | 2.15 | 1.03 |
| 5.800 | 7.51 | 3.60 | 0.90 | 2.15 | 1.03 |
| 3.600 | 7.80 | 3.20 | 1.33 | 1.64 | 0.96 |
| 5.400 | 6.60 | 4.79 | 1.99 | 2.64 | 1.24 |
| 4.300 | 9.60 | 3.21 | 0.79 | 1.38 | 1.47 |
| 6.600 | 7.00 | 4.74 | 2.74 | 2.70 | 0.79 |
| ... | ... | ... | ... | ... | ... |
| 9.800 | 6.90 | 1.45 | 1.87 | 2.64 | 0.96 |
| 10.80 | 9.20 | 4.97 | 1.51 | 3.43 | 0.85 |
| 7.600 | 7.50 | 3.88 | 0.94 | 2.41 | 1.04 |
| 7.900 | 8.40 | 4.61 | 2.61 | 2.41 | 1.01 |
| 11.70 | 9.70 | 6.04 | 1.28 | 4.17 | 1.29 |

Wan Muhamad Amir W Ahmad, Nor Farid Mohd Noor, Zainab Binti Mat Yudin, Nor Azlida Aleng and Ruhaya Hassan
5.400  7.90  4.03  2.69  2.13  0.68  
15.20  13.20  6.89  3.48  3.97  1.34  
8.600  10.50  4.77  1.47  2.48  1.57  

ods rtf file='abc.rtf' style=journal;   /*On Session of Creating Output in Microsoft Word*/

proc corr outp=R;                     /*Regression modeling based on Diabetes Mellitus Disease*/
var FBS HbA1c Total_cholesterol Triglyceride LDL HDL;
proc print data=R;
proc reg data= Diabetes;
model FBS= Triglyceride LDL HbA1c/ selection= rsquare VIF details;
run;

proc corr outp=G;                     /*Regression modeling based on Hyperlipidaemia Disease*/
var FBS HbA1c Total_cholesterol Triglyceride LDL HDL;
proc print data=G;
proc reg data= Diabetes;
model Total_cholesterol= LDL Triglyceride HDL/ selection= rsquare VIF details;
run;

ods graphics on;                     /*Plots=(surface) for all possible pairs*/
proc rsreg data=Diabetes plots=(surface); /*Pair I: Plots=(surface) FBS=Triglycerides LDL*/
model FBS=Tryglycerides LDL/lackfit;
run;
ods graphics off;

ods graphics on;                     /*Plots=(surface) for all possible pairs*/
proc rsreg data= Diabetes plots=(surface); /*Pair II: Plots=(surface) FBS=Tryglycerides HbA1c*/
model FBS= Triglyceride HbA1c/lackfit;
run;
ods graphics off;

ods graphics on;                     /*Plots=(surface) for all possible pairs*/
proc rsreg data= Diabetes plots=(surface); /*Pair III: Plots=(surface) FBS= LDL HbA1c*/
model FBS= LDL HbA1c/lackfit;
run;
ods graphics off;

ods graphics on;
proc rsreg data= Diabetes plots=(surface); /*Pair I:Plots=(surface) Total_cholesterol= LDL Tryglycerides*/
model Total_cholesterol= LDL Triglyceride /lackfit;
run;
ods graphics off;
ods graphics on;
proc rsreg data= Diabetes plots=(surface); /Pair II:Plots=(surface) Total_cholesterol= LDL HDL*/
model Total_cholesterol= LDL HDL /lackfit;
run;
ods graphics off;
ods graphics on;
proc rsreg data= Diabetes plots=(surface); /Pair III:Plots=(surface) Total_cholesterol= Tryglycerides*/
model Total_cholesterol= Triglyceride HDL /lackfit;
run;
ods graphics off;
ods rtf close; /*Off Session of Creating Output in Microsoft Word*/

Part II: Methodology of Item Analysis Through Correlation Analysis

Item analysis give a guidance to the researcher to determine the possible association between the studied variable, its also uses to assess how reliable multiple items in a test measure the same construct. The analysis also can performed by MINITAB through the options Statistics, Multivariate and Item Analysis. At the first stage of analysis we have to define the all related items which is contributing to dependent variable. After running the analysis, MINITAB will produce an important information that related to study model by providing matrix plot of related item

3.0 Result

Case I: Modeling on Diabetes Mellitus Disease Among Studied Patients

The variable selection was done based on the recommendations proposed by SPSS 24 to ensure that this model fits the data. To determine the suitability of the studied model, several measures of goodness of fit were evaluated as such the p-value and VIF criteria. Below is the three variables that the recommendations by SPSS through linear model.

<p>| Table 1: Description of the sample data |</p>
<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg</td>
<td>Triglyceride</td>
</tr>
<tr>
<td>LDL</td>
<td>Low Density Cholesterol</td>
</tr>
<tr>
<td>HbA1c</td>
<td>Haemoglobin A1c</td>
</tr>
<tr>
<td>FBS</td>
<td>Fasting Blood Sugar</td>
</tr>
</tbody>
</table>
Item analysis through the scatterplot shows that FBS has a linear relationship to triglycerides, HbA1c and Low Density Cholesterol. This suggests that the item measure has a linear relationship. In order to well-illustrated the case, modeling approach to all the potential items will be used. And the results are as follows:

**Table 1.1: Regression Modeling of Diabetes Mellitus Disease**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Unstd. Coefficient Beta (β)</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting Blood Sugar</td>
<td>HbA1c</td>
<td>0.966*</td>
<td>1.154</td>
</tr>
<tr>
<td></td>
<td>Triglycerides</td>
<td>1.053*</td>
<td>1.100</td>
</tr>
<tr>
<td></td>
<td>Low Density Cholesterol</td>
<td>0.501*</td>
<td>1.134</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td></td>
<td>-2.926</td>
</tr>
<tr>
<td>R²</td>
<td>.574</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjust R²</td>
<td>.560</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. F</td>
<td>39.538*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Significant levels: *p < 0.05*

Result provided in Table 1.1. This implies that 57.4% in fasting blood sugar can be explained by HbA1c (β = 0.966, p < 0.05), triglycerides (β = 1.053, p < 0.05) and low density cholesterol (β = 0.501, p < 0.05). Among the studied, factor triglycerides was found to has the highest positive effect on the blood sugar level. To see their association, contour plot analysis between three variables were plot separately and the further discussion will discussed in details.
Figure 1.2(a): Contour For Fasting Blood Glucose Vs triglycerides and HbA1c

The counter and surface plots indicate that the highest value of fasting blood sugar reading is obtained when HbA1c reading is high and triglycerides is high. This region appears at the upper right side of the plot.

Figure 1.2(b): Contour For Fasting Blood Glucose Vs Triglycerides and LDL

Figure 1.2(c): Contour For Fasting Blood Glucose Vs LDL and HbA1c

The counter and surface plots indicate that the highest value of fasting blood sugar reading is obtained when LDL reading is high and triglycerides is high. This region appears at the upper right side of the plot. Same interpretation goes to LDL vs HbA1c.
Figure 1.3: Fit Diagnostics for Fasting Blood Sugar

Figure 1.3 show the diagnostics plot for FBS model. The first plot (predicted value vs residual) seems to indicate that the residuals and the fitted values are uncorrelated (scattered randomly around zero), and they are homoscedastic linear model with normally distributed errors and constant variance. An overall relevance diagnostics indicator shows that model assumptions were fulfilled.

Case II: Modeling on Hyperlipidaemia Disease Among Studied Patients

Stepwise method was used for the variable selection of the hyperlipidaemia disease. As like to the case I, to determine the suitability of the studied model, several measures of goodness of fit were evaluated as such the p-value and VIF criteria. Below is the three variables that the recommendations by SPSS through linear model.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tc</td>
<td>Total Cholesterol</td>
</tr>
<tr>
<td>LDL</td>
<td>Low Density Cholesterol</td>
</tr>
<tr>
<td>HDL</td>
<td>High Density Cholesterol</td>
</tr>
<tr>
<td>Tg</td>
<td>Triglyceride</td>
</tr>
</tbody>
</table>
Item analysis through the scatterplot shows that total of cholesterol has a linear relationship to HDL and LDL. This suggests that the item measure has a linear relationship to each other’s. In order to well-illustrated the case, modeling approach to all the potential items will be used to model their relation. And the results are as follows:

Table 1.1: Regression Modeling of Total Cholesterol Disease

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Unstd. Coefficient Beta (β)</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Cholesterol</td>
<td>Low Density Cholesterol</td>
<td>1.092*</td>
<td>1.092</td>
</tr>
<tr>
<td></td>
<td>Triglycerides</td>
<td>0.417*</td>
<td>1.109</td>
</tr>
<tr>
<td></td>
<td>High Density Cholesterol</td>
<td>0.784*</td>
<td>1.073</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>-0.028</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td>.869</td>
<td></td>
</tr>
<tr>
<td>Adjust R²</td>
<td></td>
<td>.864</td>
<td></td>
</tr>
<tr>
<td>Sig. F</td>
<td></td>
<td>193.76*</td>
<td></td>
</tr>
</tbody>
</table>

Note: Significant levels: *p < 0.05

Result provided in Table 1.2 This implies that 86.9% in total cholesterol can be explained by low density cholesterol (β = 0.417, p < 0.05), triglycerides (β = 0.417, p < 0.05) and high density cholesterol (β = 0.784, p < 0.05). Among the studied, factor low density cholesterol was found to has the highest positive effect on the blood sugar level. To see their association, contour plot analysis between three variables were plot separately and the further discussion will discussed in details.
The counter and surface plots indicate that the highest value of total cholesterol reading is obtained when Triglycerides reading is high and LDL is high. This region appears at the upper right side of the plot.

The counter and surface plots indicate that the highest value of total cholesterol reading is obtained when LDL reading is high and HDL is high. This region appears at the upper right side of the plot. Same interpretation goes to HDL vs Triglycerides.
Figure 1.5 show the diagnostics plot for total cholesterol model. The predicted value vs residual seems to indicate that the residuals and the fitted values are uncorrelated, and they are homoscedastic linear model with normally distributed errors and constant variance. An overall relevance diagnostics indicator shows that model assumptions were fulfilled.

4.0 Summary and Discussion

This paper provides only a preliminary overview of the influencing factors that contribute to the estimation of hyperlipidemia and diabetes mellitus through the linear modeling techniques. The rapid growth of health education may depend on the related factors that are associated directly or indirectly for a certain disease. These related factors should be measured carefully and have to consider as factors related to the health status of an individual. In this paper, three different methods have been used for the both case study which is include: (i) scatterplot analysis, (ii) multiple linear model analysis, and (iii) response surface methodology (RSM).

Case Study I

Diabetes mellitus grows very fast in global scenario. It has been estimated that in 2010 there was approximately 6.4% of the adult population suffering from this disease (Kaul, Tarr et al., 2013). According to Ghazanfari and Haghdoost, the association of HbA1c with FBS was relatively strong particularly in diabetic patients. Generally, FBS was a more accurate predictor for HbA1c compared with HbA1c as a predictor of FBS (Ghazanfari, Haghdoost et
In our study, scatterplot methodology was applied in determining relationship among fasting blood sugar with HbA1c, triglycerides and low density cholesterol. All independent variables have positive association and it appears significantly to the fasting blood sugar in regression modeling. Results provided in Table 1.1. This implies that 57.4% in fasting blood sugar can be explained by HbA1c ($\beta = 0.966$, $p < 0.05$), triglycerides ($\beta = 1.053$, $p < 0.05$) and low density cholesterol ($\beta = 0.501$, $p < 0.05$). Gabbay et al. had pointed out that there is a significant correlation between HbA1c levels and glucose, as well as between HbA1c and cholesterol concentrations. This indicates that HbA1c levels have association with sugar level. The observation has been confirmed to include an even stronger correlation between HbA1c and triglyceride levels in patients with uncontrolled diabetes. Besides that, partial correlation coefficients revealed that the HbA1c level was mainly dependent upon the glucose response rather than the fasting concentration (Gonen, Rochman et al., 1977).

**Case Study II**

Hyperlipidaemia model shows that three factors (low density cholesterol, triglycerides and high density cholesterol) had a strong association with a total cholesterol. Hyperlipidemia is caused by fatty substances (or overabundance of lipids) in the blood and it is an important risk factor in development of atherosclerosis and heart disease, it may be caused metabolic disorders like diabetes mellitus, excessive alcohol intake, and many more. Alteration in cholesterol, triglycerides and LDL can cause possible complications in human body as such acute pancreatitis, occlusion of blood vessels and reduced elasticity of the lumen of the artery. Moreover risk increases with diabetes mellitus, hypothyroidism, nephrosis, alcoholism, use of oral contraceptives, family history of hyperlipidemia and improper diet that is high in fat and cholesterol. (Phogat, Deep et al., 2010). In our study, especially for the second case study which focusing on hyperlipidaemia disease. The scatterplot methodology was applied in determining relationship among fasting blood sugar with HbA1c, triglycerides and low density cholesterol. Results from multiple linear regression shows that total cholesterol has a significant association with low density cholesterol ($\beta = 0.417$, $p < 0.05$), triglycerides ($\beta = 0.417$, $p < 0.05$) and high density cholesterol ($\beta = 0.784$, $p < 0.05$). Results in Table 4.13 indicated that 53.7% variances in total cholesterol can be explained by Low Density, Cholesterol, Triglycerides and High Density Cholesterol.

This study tried to find the associated factors and improved the understanding of hyperlipidaemia disease and diabetes mellitus among public. The findings, which get from the linear regression modeling, RSM methodology can helps researcher to study more about the variable behaviour (factor that related to hyperlipidaemia disease and diabetes mellitus) and the changes according to their association.

**Declaration**

Authors declare that the article has not been published or submitted for publication in any other journal.
References


