Inter-Relationships of Cholesterol With Cardiac Factors for Heart Patients

Jinseog Kim1, Rabindra Nath Das2,3, Youngjo Lee1, Sabyasachi Mukherjee4, Hyungmi An3, Sunit Kumar Medda3

1Department of Applied Statistics, Dongguk University, Gyeongju, Korea
2Department of Statistics, The University of Burdwan, Burdwan, West Bengal, India
3Department of Statistics, College of Natural Science, Seoul National University, Seoul, 151-747, Korea
4Department of Mathematics, NSHM Knowledge Campus, Durgapur, West Bengal, India

ABSTRACT

Objectives: The role of Cholesterol and its relationship with some cardiac risk factors for heart patients are examined in the current report using both Cholesterol level and two cardiac factors modeling.

Materials and methods: A real data set of 303 heart patients with 14 study characters are considered in the report. Statistical joint generalized linear models (JGLMs) are considered using both Gamma & Log-normal distributions.

Results: It is observed from Cholesterol level modeling that Cholesterol level is higher for female heart patients (P=0.0013) than male, or at older ages (P=0.0012) than younger. It is higher for the patients with high maximum heart rate (P=0.0877), or having resting electrocardiographic at normal level (P=0.0107), or with thalassemia at reversal defect (P=0.0486) and at fixed defect (P=0.0040) than at normal. It is also higher for the patients having heart disease diagnosis (angiographic disease status) value 0 (meaning less than 50% diameter narrowing) (P=0.0515) than others. Variance of Cholesterol level is higher for female patients (P=0.0265) than male, and it increases as ST depression induced by exercise relative to rest (Oldpeak) (P=0.0096) increases. From maximum heart rate modeling, it is noted that maximum heart rate increases as the Cholesterol level (P=0.0325) increases. In addition, variance of maximum heart rate decreases as the Cholesterol level (P=0.0058) increases. Also from resting blood pressure modeling, it is observed that mean resting blood pressure increases as the Cholesterol level increases, where it is a confounder in the model.

Conclusions: Cholesterol levels should be examined regularly at older ages, along with the maximum heart rate achieved, thalassemia status, and resting blood pressure for both male and female heart patients.

Key words: Blood pressure, Chest pain, Cholesterol level, Joint mean & dispersion modeling, Maximum heart rate, Non-constant variance.

INTRODUCTION

Generally, atherosclerosis is a preliminary form of heart disease, in which cholesterol plaques form in the artery walls, consequently blood flow is restricted. Practically, there is no definite cardiovascular disease (CVD) marker. It is known that total Cholesterol (TC) is considered as a CVD marker as it is a main cause of atherosclerosis and cardiac disease4,5. The relationship between TC and CVD has been focused in many research articles such as by Pasty et al4, Das4, Esrey et al6, Schupf et al7, Akerblom et al8, The Lipid Research Clinics Coronary Primary Prevention Trial results9. The relationship between TC and CVD in both sexes and in all ages has been established based on meta-analysis in the Prospective Studies Collaboration conducted by Lewington et al10. Even though the relationship between TC and CVD mortality is observed in both sexes and in all ages, but the risk decreases as the age rises, and it is minimal more than 80 years old, which was reported by Rijnsburger et al11. These are not reproduced in the report.

It is well-known that TC is the sum of high-density lipoprotein (HDL-C), low-density lipoprotein (LDL-C), and very low density lipoprotein (VLDL-C). Note that triglyceride is not included in TC. TC level below 200 (mg/dL) is considered as normal, and more or equal to 200 (mg/dL) is treated as high. The center for disease control and prevention has examined data from 2005-2008, and it has observed the prevalence, incidence, treatment, and control of high LDL-C levels12. It has been observed that nearly 71 million American adults (33.5%) had high LDL-C levels, just only 34 million (48.1%) received treatment, and 23 million (33.2%) had their LDL-C controlled12. High LDL-C level is correlated with a higher risk of CVD, but a high level of HDL-C is correlated with a lower risk of CVD, which was reported by Weverling-Rijnsburger et al12. TC levels indicate a patient’s cardiac risk suggested by Takata et al13, Akerblom et al14. For further guidance, medical practitioners divide the TC level by the HDL-C level, which should be less than 4 to 1. Heart disease risk can be reduced by increasing HDL-C level, and along with reducing TC level.

Previous research articles have studied only the relationship of cardiac factors such as systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate with Cholesterol based on simple and multiple regression, logistic regression, meta-analysis etc., which are not adequate statistical methods for modeling heterogeneous positive physiological data sets. Very few articles have studied the inter-relationships of Cholesterol and cardiac risk factors in both ways such as based on the modeling of a cardiac risk factor on Cholesterol along with other factors, and the modeling of Cholesterol on many cardiac factors. The current report examines the inter-relationships of Cholesterol in both ways using a real data set. In addition, the effects of Cholesterol on cardiac factors have been derived.

MATERIALS AND METHODS

Materials

The inter-relationship of Cholesterol on cardiac factors is examined herein using a real data set of 303 heart patients containing 14 study attributes, which is displayed in UCI machine learning repository. Donor of the data set is David W. Aha [(aha '@' ics.uci.edu) (714) 856-8779]. Readers can find the data set in UCI machine learning repository, and its collection method and patients population are clearly described in the study by Detrano et al15. These are not reproduced in the report.
The 14 study attributes are as follows: Sex (1=male; 0=female), Age (in years), Chest pain (CP) (1= typical angina; 2= atypical angina, 3= non-anginal pain or asymptomatic), Serum Cholesterol (Chol) (in mg/dl), Resting blood pressure (Trestbps) (in mm Hg on admission to the hospital), Fasting blood sugar (Fbs) (if Fbs> 120 mg/dl) (1 = true; 0 = false), Maximum heart rate achieved (Thalach), Exercise induced angina (Exang) (1= yes; 0= no), Resting electrocardiographic (Restecg) (resting electrocardiographic results -- value 0= normal; 1= having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of >0.05 mV)), ST depression induced by exercise relative to rest (Oldpeak), Ca (number of major vessels (0-3) colored by flouroscopy), Slope (the slope of the peak exercise ST segment -- value 1= upsloping, 2= flat, 3= downsloping), Target (num: diagnosis of heart disease (angiographic disease status) value 0 < 50% diameter narrowing; 1> 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)), Thalassemia (Thal) (3 = normal; 6 = fixed defect; 7= reversible defect).

Statistical methods

The considered data set herein is a multivariate form, and the related three responses are Cholesterol level, maximum heart rate achieved, and resting blood pressure. These three responses are continuous, heterogeneous, positive and non-normally distributed. Heterogeneous data can be modeled by using suitable transformation if the variance is stabilized by that transformation. But it is not always true. Under that cases, the response should be modeled by joint generalized linear models (JGLMs). Here we have examined that these three responses are not stabilized with any suitable transformation. So, we have considered to model these responses using JGLMs under both the Log-normal and Gamma distributions. JGLMs are clearly described in the book by Hastie et al25. All the included effects in both the models are accepted considering the lowest Akaike information criterion (AIC) value (within each class), which minimizes both the squared error loss and predicted additive errors, and the AIC rule is clearly given in the book by Hastie et al25. All the included effects in both the models are significant. The Cholesterol level analysis results are presented in Table 1. Based on AIC value, both the model fits are similar, and the Gamma fit (AIC=3166.688) is slightly better than Log-normal fit (AIC= 3167).

Data developed Cholesterol level probabilistic model should be tested by model checking tools before receiving it as the valid final model, which concludes all valid decisions. The derived Gamma fitted Cholesterol level model (Table 1) has been tested by model diagnostic plots in Figure 1. In Figure 1(a), the Cholesterol level Gamma fitted (Table 1) absolute residuals are plotted against the fitted values, which is quite a flat line, concluding that variance is constant with the running means. Figure 1(b) displays the mean Cholesterol level Gamma fitted normal probability plot (Table 1), which does not present any lack of fit. Therefore, both the Figures 1(a) and 1(b) confirm that that Gamma model fits the data well (Table 1).

RESULTS

Summarized JGLMs results for Cholesterol level analysis are shown in Table 1. Gamma fitted Cholesterol level model (Table 1) shows that mean Cholesterol level is higher for female heart patients (P=0.0013).

Figure 1: For the joint Gamma fitted cholesterol level model (Table 1), the (a) absolute residuals plot with respect to the fitted values, and (b) the normal probability plot for the mean model.

Statistical and Graphical Analysis

The response Cholesterol level is treated as the dependent variable and the rest other variables and factors are considered as the explanatory variables. Here Cholesterol level has been modeled by JGLMs using both the Log-normal and Gamma distributions. The final model has been accepted considering the lowest Akaike information criterion (AIC) value (within each class), which minimizes both the squared error loss and predicted additive errors, and the AIC rule is clearly given in the book by Hastie et al25. All the included effects in both the models are significant. The Cholesterol level analysis results are presented in Table 1. Based on AIC value, both the model fits are similar, and the Gamma fit (AIC=3166.688) is slightly better than Log-normal fit (AIC= 3167).

For log-normal distribution, JGLM of the mean and dispersion model are presented by

\[ \log(\mu) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Sex} + \beta_3 \text{Thalach} + \beta_4 \text{Oldpeak}, \]

where, \( \beta_0 \) and \( \beta_1 \) are the vectors of explanatory variables related with the regression coefficients \( \beta \) and \( \gamma \) respectively.

**Gamma JGLMs:** Instead of modeling \( \log(\mu) \), we can model \( \mu \) directly. Here \( V(\gamma) \) denotes the dispersion function with two components such as \( \gamma_1^2 \) (independent of mean parameters) and \( V(\mu) \) (depends on mean parameters). Generally, GLM family distribution is identified by \( V(\mu) \). For example, if \( y(\mu) = \mu \) it is Poisson, Gamma if \( y(\mu) = \mu^2 \), and Normal if \( V(\mu) = 1 \), etc. Gamma JGLMs mean and dispersion models are given by

\[ \eta_0 = g(\mu) = \gamma_1^2 \beta \quad \text{and} \quad \eta = h(\gamma) = w^T \gamma, \]

where, \( g(\cdot) \) and \( h(\cdot) \) are the GLM link functions corresponding to the mean and dispersion linear predictors respectively, and \( w^T \) are the explanatory variables vectors associated with the mean and dispersion parameters respectively. Maximum likelihood (ML) method is applied to estimate mean parameters, while the restricted ML (REML) method is used to estimate dispersion parameters, which are clearly described in the book by Lee et al25.
The mean and variance of Cholesterol level models are displayed above by two equations. It is found that mean Cholesterol level is expressed by Age, Sex, Thalach, Restingecg, Thal, Target, while its variance is presented by Sex and Oldpeak only.

**DISCUSSION**

Inter-relationships between TC level and cardiac factors are reported herein from the models of TC level, and cardiac risk factors such as maximum heart rate achieved and resting blood pressure. Final Gamma fitted mean TC level model (Table 1) shows that it is positively associated with age (P=0.0012), indicating that TC level is higher for older heart patients than younger. It is negatively associated with sex (1=male; 0=female) (P=0.0013), concluding that TC level is higher for female patients than male. It is positively associated with Thalach (P=0.0877), interpreting that TC level increases as Thalach increases. It is negatively associated with Restecg (resting electrocardiographic results -- value 0 = normal; 1 = having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)) (P=0.0107), denoting that TC level is higher for the patients with Restecg at normal level than others. Mean TC level is positively associated with Thal (3 = normal; 6 = fixed defect; 7 = reversible defect) at fixed defect (P=0.0940) and at reversal defect (P=0.0466), concluding that TC level is higher for the patients with Thal at reversal and at fixed defect levels than at normal. Mean TC level is negatively associated with Target (num: diagnosis of heart disease (angiographic disease status) value 0: < 50% diameter narrowing; 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)) (P=0.0515), interpreting that it is higher for the patients having heart disease diagnosis (angiographic disease status) value 0 (indicating less than 50% diameter narrowing) than others.

Variance model of TC level (Table 1) shows that TC level variance is negatively associated with sex (1=male; 0=female) (P=0.0265), indicating that it is higher for female heart patients than male. It is positively associated with ST depression induced by exercise relative to rest (Oldpeak) (P=0.0095), concluding that it increases as Oldpeak increase.

Thalach analysis has been reported by Das et al21. For ready reference, JGLMs for both mean and dispersion models for Thalach are reported herein as follows. The Gamma fitted Thalach mean ($\mu$) model (from Das et al21; Table 1) is $\mu = \exp \left( 5.0286 - 0.0060 \text{Age} + 0.0490 \text{Chest pain}^2 + 0.0330 \text{Chest pain} + 0.0010 \text{Resting BP} + 0.0002 \text{Cholesterol} -0.0585 \text{Exercise induced angina} + 0.0919 \text{Thal2} + 0.0852 \text{Thal3} + 0.0628 \text{Target} \right)$, and the Gamma fitted Thalach variance ($\sigma^2$) model is $\sigma^2 = \exp \left( -4.0324 + 0.0330 \text{Age} -0.6336 \text{Thal3} - 0.0054 \text{Cholesterol} - 0.6336 \text{Target} \right)$.

From mean Thalach model reported by Das et al21, it is identified that Thalach is positively associated with TC (P=0.0325), indicating that Thalach increases as TC level increases. Identical result is shown herein using TC level modeling (Table 1). Also, variance of Thalach is negatively associated with TC (P=0.0058), concluding that it increases as TC level decreases. Also from Trestbps modeling, it can be shown that Trestbps is positively partially associated with TC level (P=0.2547), indicating that it increases as TC level increases. It is not shown herein. It will be reported in our future research article. Note that in epidemiology, partially significant effect is considered as confounder.

A similar study conducted by Das22 has shown that variance of TC level is negatively associated with DBP (P=0.001), while it is positively associated with the interaction effect SBP*DBP (P=0.084). Mean HDL-C is positively associated with DBP (P=0.039), while its variance is also positively associated with DBP (P=0.058). Also variance of lipid ratio is negatively associated with DBP (P=0.002). Mean DBP is positively associated HDL-C (P=0.002). There are very few studies regarding the inter-relationship between Cholesterol level and cardiac factors based on appropriate modeling, so the present results cannot be compared with the earlier results.

**CONCLUSIONS**

Both the Log-normal and Gamma JGLM fits for TC level have similar interpretations (Table 1). The standard errors of the estimators are very small, indicating that estimates are stable. AIC value and graphical diagnostic tools are used to select the appropriate model. Some of the above conclusions are reported in earlier research articles, which are observed in practice. Research should have greater faith in these results than those emanating from multiple linear regression, Logistic regression, Log-Gaussian (with constant variance), as the current derived modes have been examined with graphical diagnostic tools, AIC value and comparing between two distributions such as Log-normal and Gamma.

The present study only considers two cardiac factors such as Thalach and Trestbps. It cannot include other cardiac factors such as SBP, DBP, mean arterial pressure, men central venous pressure, cardiac index, ejection fraction etc., as these are not included in the considered data set. In addition, present study cannot examine the associations of LDL-C, or HDL-C, or VLDL-C, or triglyceride with many cardiac factors, as these factors are not included in the data set. Future researchers may consider more cardiac factors, along with the components of TC, which may give many interesting results.

The current report focuses many new results to the cardiology literature. Cardiac patients and medical experts will be benefitted from the current results. An appropriate statistical model can give correct...
interpretations. Cholesterol level should be examined regularly at older ages, along with maximum heart rate achieved, thalassemia status and resting blood pressure for both male and female heart patients.

**ACKNOWLEDGEMENT**

The authors are very much indebted to the referees and the editor who have provided valuable comments to improve this paper. This research was supported by the Dongguk University Research Fund of 2019 and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2C1002408). The authors are very grateful to the principal data investigators Andras Janosi, William Steinbrunn, Matthias Pfisterer and Robert Detrano who provided the data freely for using scientific study.

**CONFLICT OF INTEREST**

The authors confirm that this article content has no conflict of interest.

**REFERENCES**