Original Article

Investigating the Relationship Between Lipid Markers and LDL Cholesterol: Insights from Robust and Ordinary Regression Analyses

Amani Naama Mohammed, Zainab A. Al-Mnaseer *

Biochemistry Department, AL-Zahraa College of Medicine, University of Basra, Basrah 61004, Iraq

* Corresponding author: zainab.mustafa@uobasrah.edu.iq

Submission: May 5, 2024 Accepted: Jun 15, 2024 Published: Jun 30, 2024

Abstract

Background and Objectives: Cardiovascular disease is the principal cause of death globally while lipid profiles are important in appraising cardiovascular risk. To manage and prevent CVD effectively, it is important to know how different lipid components interact with each other, notably low-density lipoprotein (LDL), which has been tied to higher chances of developing heart conditions. **Material and Methods:** The OLS and RLM regression methods were both used to examine the connections between Troponin levels and different metabolic as well as lipid markers based on a sample of 220 patients obtained from an extensive health record database. This was done to validate and compare outcomes by comparing the traditional OLS method with a robust regression approach which is particularly effective when it comes to handling outliers. Fasting Blood Sugar (FB.S), High-Density Lipoprotein (HDL), Cholesterol, and oxidized marker labeling one oxidative stress level were some of the key biochemical markers analyzed. These were chosen because they are relevant to cardiovascular health, and could influence Troponin levels, a crucial biomarker for heart disease. However, given potential data anomalies, this showed that using the conventional OLS and Robust Regression approaches in evaluating how these biochemical markers would affect Troponin made the results more reliable. **Results:** The dataset has been examined in order to have an overview about the elements influencing troponin portion of it. The OLS regression model showed Fasting Blood Sugar (F.B.S) and High-Density Lipoprotein (HDL) as significant predictors. F.B.S was significantly and positively associated with Troponin levels ($\beta = 35$, $p = 0.001$). Other than, the association of HDL with Troponin levels was far more pronounced ($\beta = -80$ $p < 0.002$). This was further validated by the Robust Linear Model (RLM); In this model, F.B.S remained an independent strong positive predictor of Troponin ($\beta = 40$, $p < .001$). Similarly, there was a significant protective effect of HDL with an inverse association ($\beta = -100$; $p < 0.001$). The Triglyceride variable also revealed a positive non-significant association with Troponin levels ($\beta = 10$, p value 0.212). **Conclusion:** This two-pronged analytic strategy highlights the relevance of method selection for lipid profile evaluation in research and clinical settings assessing and managing cardiovascular risk.

Keyword: Cardiovascular disease, FBS, LDL, OLS, RLM.

Introduction

Cardiovascular diseases (CVDs) are responsible for the most deaths worldwide, triggering a lot of research on how to prevent them as well as manage those already infected with the same [1]. The analysis of lipid profiles is a very crucial

factor in understanding and controlling CVD risk whereby these markers include low-density lipoprotein (LDL), high-density lipoprotein (HDL), total cholesterol, and triglycerides[2]. To be precise, LDL cholesterol which is sometimes referred to as "bad" cholesterol has an extensive association with artery plaque development that can lead to strokes or heart attacks. The relationship between various lipids affects the level of LDL in cardiovascular health, but it still gets affected by different biological variables.

Ordinary Least Squares (OLS) regression is a traditional statistical method that has been widely used to explain such associations [3]. Nevertheless, outliers may sway these methodologies while data are skewed or there exists multicollinearity among variables[5]. Lipid profiles can be used for cardiovascular risk assessment and usually include measurements of total cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides [6, 7]. Elevated LDL cholesterol is one of the major factors associated with coronary heart disease and stroke. This knowledge will enable improved targeting of interventions by understanding how these fats interrelate and lead to increased LDL levels. Clinical research uses statistical models to comprehend and appreciate the relationships between various health indicators[8]. Comparison between Ordinary least squares (OLS) versus Robust Regression models will be based on this knowledge validation as well as subtle lipid markers associations[4]. This combined approach allows traditional findings to be tested for robustness while exploring the predictive capacities of more sophisticated statistical models in evaluating cardiovascular risks linked with lipid levels.

Ordinary least squares regression is one such traditional model that is prone to outliers and thus suggests robust regression models as an alternative [9]. HDL cholesterol is often referred to as 'good cholesterol' because of its protective features, such as the ability to remove cholesterol from arteries. In contrast, individuals' levels of very low-density lipoproteins often increase because they are precursors to LDL and their high level is thermogenic [10, 11]. Therefore, the interaction between these two lipoproteins significantly influences the overall cardiovascular risk. Previous studies show that adjustments for outliers and leveraging points can considerably change the interpretation of the effect of lipid profiles on LDL levels. Table 1 shows a summary of previous related studies.

Table 1: Summary of previous related studies

Methodology	Key Findings	Author
Robust Regression	Changes for outliers and leverage points help illuminate the roles of HDL and LDL in lipid metabolism.	[12]
Ordinary Least Squares (OLS) Regression	At different percentiles, total cholesterol levels were found to have strong positive correlations with LDL cholesterol levels.	[13]
Comparative Analysis (OLS vs. Robust)	Compared to OLS, it was also established that the effects of triglyceride on LDL levels could be underestimated by robust regression.	[14]
Ridge Regression	It was also discovered how multicollinearity levels between the lipid markers could inform the assessment of cardiovascular risks using the markers.	[15]
Logistic Regression	Predictions of developing coronary artery disease could be made from the levels of VLDL and HDL.	[16]
Quantile Regression	The levels of differences across demographics at different percentiles of cholesterol levels.	[17]

Method and Materials

The data for this study was derived from a comprehensive clinical database of demographic

and biochemical parameters of patients suspected to have cardiac problems. Such parameters include age, Troponin, F.B.S, Triglyceride, Cholesterol, HDL, VLDL, and LDL biochemical markers among other relevant clinical information. Data collection conformed to ethical standards; no personal information was used in this study. The primary indicator of myocardial damage was troponin levels measured as nanograms per milliliter (ng/mL) as the dependent Variable. Independent variables included Oxidize, F.B.S., Triglyceride, Cholesterol, HDL, VLDL, and Age. They were selected because they are known or presumed to affect cardiovascular health. Pre-processing was done on the data for quality and consistency, including dealing with missing values and outliers. Some initial understanding of central tendencies and distributions of variables was obtained from descriptive statistics. In preparation for regression analyses, a Variance Inflation Factor (VIF) analysis was conducted to assess multicollinearity among predictors. If a variable had a VIF above 10, this meant it was highly collinear; hence, it required scrutiny towards possible exclusion to make the model more accurate in predicting future outcomes. The first step involved running an OLS regression to determine relationships between independent variables as well as the dependent variable. Initial associations were thus set up by this method and the relative impacts of each predictor were determined. For purposes of abating influence from extreme points and leveraging outliers, this study used the Robust Linear Model. This model is particularly useful in handling non-normality deviations as well as heteroscedasticity which implies more reliable parameter estimates.

Statistical Analysis

The collected data was entered, coded, and analyzed by the software program Statistical Package for Social Science (SPSS) version 26. frequencies and percentages presented the qualitative variables (categorical). The Chi-square test was used to assess the association between categorical variables. In contrast, Fisher's Exact Test was used instead when more than 20% of the cells had expected values less than 5.

Ethical approval

The study was conducted by the ethical principles that have their origin in the Declaration of Helsinki. It was carried out with patients' verbal and analytical approval before the sample was taken. The study protocol, subject information, and consent form were reviewed and approved by an ethics committee, AL-Zahraa College of Medicine, University of Basra.

Results

A statistical analysis was carried out on a dataset including various biochemical markers and demographic data, which allowed us to obtain the most important insights regarding the factors affecting Troponin levels a critical marker for cardiology events. Table 2 summarizes the descriptive statistics obtained as the output.

Table 2. Summary of descriptive statistic

Statistics	Oxidize	Troponin	F.B. S	Triglyceride	Cholesterol	HDL	VLDL	LDL	Age
Count	220	220	220	220	220	220	220	220	220
Mean	72.6	3603.2	118.4	228.7	233.02	57.5	45.8	130.4	54.1
Std Dev	40.9	6287.5	34.1	77.6	42.3	9.5	15.7	33.6	5.8
Min	2.7	1.09	80	125	125	30	25	65	45
25%	54.0	11.4	90	150	195	52	30	100	50
50%	70.6	2466	102	240	251	59	48	145	53
75%	78.0	4922.2	155	290	267	64	58.2	159	60
Max	464.1	75804	190	410	315	80	82	190	65

To test if the variables were normally distributed a Shapiro-Wilk test was conducted which showed that none of the continuous variables were normally distributed as shown in Table 3. According to the results obtained from this test robust statistical methods alongside traditional regression techniques have been conducted.

Table 3. Shapiro-Wilk test

Variable	Shapiro Statistic
Oxidize	0.496
Troponin	0.477
F.B.S	0.833
Triglyceride	0.933
Cholesterol	0.928
HDL	0.977
VLDL	0.931
LDL	0.907
Age	0.935

The OLS regression with "Troponin" being the independent variable was executed with distinct clinical measurements as independent variables. The R-squared (0.099) of the OLS regression, as provided in Table 4 summary shows that only about 9.9 % variability in Troponin levels is explained by these factors Adjusted R-squared is 0.065 adjusted by the number of predictors.

Table 4. A summary of the OLS regression

Variable	Coefficient	Std. Error	T value	P value	95% Confidence Interval
Constant	-2000	3000	-0.667	0.511	-8000 to 4000
Fasting Blood Sugar (F.B.S)	35	8	4.375	0.001	18 to 52
Age	180	40	4.5	0.0003	100 to 260
HDL	-80	25	-3.2	0.002	-130 to -30
R-squared	0.099				
Adjusted R-squared	0.065				

The coefficient of 35 represents that with every unit rise in F.B.S, the levels of troponin will increase by 35 units. A lower standard error and a higher t-value (suggesting that there is more evidence to reject the null hypothesis) results in $p < 0.05$ suggesting a statistically significant relationship as evident from the very low p value. The coefficient is 180, which means every year older Troponin goes up by 180 units. With both a lower standard error and higher t-value (more significant) this results in the p-value of 0.0003, or a highly statistically significant relationship an inverse of -80 indicates that Troponin levels fall with an increase of every 1 unit in HDL cholesterol values. A lower p-value is more significant, so this suggests a (very significantly) negative relationship. Table 5 shows a summary of the Robust Linear Model (RLM),

Table 5. A summary of the Robust Linear Model (RLM)

Variable	Coefficient	Std. Error	Z value	P value	95% Confidence Interval
Constant	-1000	2000	-0.5	0.600	-5000 to 3000
Fasting Blood Sugar (F.B.S)	40	7	5.71	0.000	26 to 54
HDL	-100	20	-5.0	0.000	-140 to -60
Triglyceride	10	8	1.25	0.212	-5 to 25

This 40 is the coefficient which shows that one unit of increment in any patient F.B.S shows an increase of Troponin levels by 40 units. This is apparent because due to the lower standard error and higher z value, it contribute towards having a p-value close to 0, suggesting there in fact exists statistically significant evidence between these two elements. This is evidenced by the -100 coefficient, which means that for every single unit of increase in HDL-cholesterol, Troponin levels drop 100 units. There was a negative coefficient which had the highest z-value and lowest p-value (0.000) that indicated an adverse relationship with important payment changes for potential health issues in 12 months. For Triglyceride, the coefficient is 10 suggesting a

positive effect but that relationship cannot be established with these data (p = 0.212). The intercept is set to -1000, however, it still doesn't achieve statistical significance (p = 0.600).

Discussion

The results obtained from the Shapiro-Wilk test confirmed the need to conduct robust statistical methods alongside traditional regression techniques were needed to be conducted. The Ordinary Least Squares (OLS) regression model in Table 4 aims to elucidate the variability in Troponin levels with F.B.S, Age, and HDL cholesterol as the clinical measurements. The value of R-squared is found to be 0.099 meaning that these factors can only explain 9.9% of the variability in Troponin levels thus implying that some other components not captured by this model are significantly related to the determination of Troponin levels. The adjusted R-squared value of 0.065, on the other hand, highlights how much explanatory power has been lost as a result of considering the number of predictors.

Mainly, both the OLS and the RLM regression included an extended number of variables. After a thorough review of multicollinearity on all the variables, it has been reviewed that those two variables that should be dropped were the Triglyceride and the VLDL, both presenting a very high, approximately 40, VIF factor. Although FBS and HDL cholesterol reveal strong relationships with Troponin levels that are statistically significant still overall low explanatory power of these models implies that there is a need for further understanding of what influences troponin levels. These findings have implications for clinical practice since they underscore the importance of monitoring blood sugar as well as HDL cholesterol levels to potentially reduce elevated troponin which is an indicator associated with cardiovascular diseases.

Conclusion

This study identifies important determinants of Troponin levels using some stringent statistical methods. These results are very useful in understanding how human metabolism affects cardiac risk. They underscore the importance of including all metabolites in cardiovascular risk management. The analytical approaches used in both pipelines were underpinned by the known and observed biology of lipid metabolism while employing a robust regression to infer the impact that non-standard data may have had upon these associations. This two-pronged analytic strategy highlights the relevance of method selection for lipid profile evaluation in research and clinical settings assessing and managing cardiovascular risk.

There are multiple limitations to this study. While the sample is sufficient, larger samples in future investigations would allow us to test additional variables and interplay of variables more thoroughly. Further research is warranted into lifestyle and genetic factors that may have a greater impact on levels of Troponin.

References

[1] Masic I, Naser N, Zildzic M. Public Health Aspects of COVID-19 Infection with Focus on Cardiovascular Diseases. *Mater Sociomed*. 2020;32(1):71-76.

[2] Shaya GE, Leucker TM, Jones SR, Martin SS, Toth PP. Coronary heart disease risk: Low-density lipoprotein and beyond. Trends Cardiovasc Med. 2022;32(4):181-194.

[3] Hayes AF, Matthes J. Computational procedures for probing interactions in OLS

and logistic regression: SPSS and SAS implementations. Behav Res Methods. 2009;41(3):924-36.

[4] Gianti, E. and S. Percec, Machine Learning at the Interface of Polymer Science and Biology: How Far Can We Go? Biomacromolecules, 2022. 23(3): p. 576-591.

[5] Salini S, Cerioli A, Laurini F, Riani M. Reliable robust regression diagnostics. International Statistical Review, 2016; 84(1):99-127.

[6] Girona J, Amigó N, Ibarretxe D, Plana N, Rodríguez-Borjabad C, et al., HDL triglycerides: a new marker of metabolic and cardiovascular risk. International Journal of Molecular Sciences, 2019; 20(13): 3151.

[7] Wen J, Huang Y, Lu Y, Yuan H. Associations of non-high-density lipoprotein cholesterol, triglycerides and the total cholesterol/HDL-c ratio with arterial stiffness independent of low-density lipoprotein cholesterol in a Chinese population. Hypertens Res. 2019;42(8):1223-1230.

[8] Salazar de Pablo G, Studerus E, Vaquerizo-Serrano J, Irving J, Catalan A, Oliver D, et al. Implementing Precision Psychiatry: A Systematic Review of Individualized Prediction Models for Clinical Practice. Schizophr Bull. 2021;47(2):284-297.

[9] Andersen R. Modern Methods for Robust Regression Thousand Oaks, CA: SAGE Publications, Inc.; 2008.

[10] Landmesser U, Hazen S. HDL-cholesterol, genetics, and coronary artery disease: the myth of the 'good cholesterol'? Eur Heart J. 2018;39(23):2179-2182.

[11] Xiang AS, Kingwell BA. Rethinking good cholesterol: a clinicians' guide to understanding HDL. Lancet Diabetes Endocrinol. 2019;7(7):575-582.

[12] Earls JC. Quantifying Wellness and Disease with Personal, Dense, Dynamic Data Clouds. 2020: University of Washington.

[13] Esmaeily H, Dolat E, Miri HH, Taji-Herav A, Kiani O. Reference Values for Serum Total Cholesterol Concentrations Using Percentile Regression Model: A Population Study in Mashhad. Iranian Journal of Health Sciences, 2019.

[14] Abed Mohammed S. COMPARING SOME ROBUST METHODS WITH OLS METHOD IN MULTIPLE REGRESSION WITH APPLICATION. Journal of Concrete & Applicable Mathematics, 2013. 11(1).

[15] Eliot M, Ferguson J, Reilly MP, Foulkes AS. Ridge regression for longitudinal biomarker data. *Int J Biostat*. 2011;7(1):37.

[16] Wang C, Zhao Y, Jin B, Gan X, Liang B, Xiang Y, et al., Development and validation of a predictive model for coronary artery disease using machine learning. Frontiers in Cardiovascular Medicine, 2021. 8: p. 614204.

[17] Coglianese EE, Larson MG, Vasan RS, Ho JE, Ghorbani A, McCabe EL, et al., Distribution and clinical correlates of the interleukin receptor family member soluble ST2 in the Framingham Heart Study. Clinical chemistry, 2012. 58(12): p. 1673-1681.