Model to assess the factors of 10-year future risk of coronary heart disease among people of Framingham, Massachusetts

Azizur Rahman[1], Arifa Tabassum[2]
[1]Department of Community Health Sciences, University of Manitoba, Canada
[2]Department of Statistics, Jahangirnagar University, Bangladesh

Article Info

ABSTRACT

In earlier decade, heart disease was the most common cause of death in the US. Among the many important risk factors such as age, number of cigarettes smoke could help in determining the odds of having corona heart disease (CHD) when modeling with other important factors. We analyzed ongoing cardiovascular study on residents of the town of Framingham, Massachusetts, US to predict the 10-year risk of future CHD. We applied the Binary logistic regression model to assess the strength of the association of factors (such as gender, age, number of cigarettes smoke, total cholesterol level) in predicting the odds of having CHD in study population. Results showed that gender, age, number of cigarette smoke, systolic blood pressure were statistically significant and the increased age and cigarettes per day increase the odds of having 10-year risk of CHD. However, the noticeable finding was that patients with Diabetes at higher glucose level have the higher odds of having 10-year risk of CHD than with low level of glucose concentration among the residents of Framingham study.

Corresponding Author:

Azizur Rahman,
Department of Communication Health Sciences,
University of Manitoba,
Winnipeg, R3E 0M2, Canada.
Email: rahmana1@myumanitoba.ca

1. INTRODUCTION

Cardiovascular disease (CVDs) are a group of disorders of the heart and blood vessels and they include two major sub-type of disease: coronary heart disease (CHD) and cerebrovascular disease (stroke). Despite the advances in clinical care and medicine, cardiovascular diseases (CVDs) are continuing to be the principal cause of morbidity and mortality. Cardiovascular disease (CVDs) are a group of disorders of the heart and blood vessels and they include two major sub-type of disease: coronary heart disease (CHD) and cerebrovascular disease (stroke). The primary cause of CHD is atherosclerosis that reduces blood flow through the coronary arteries to the heart muscle and therefore malfunctioning starts in heart. CHD is now the leading cause of death worldwide. As estimated 3.8 million men and 3.4 million women die each year from CHD [1]. In developed countries heart disease is the leading cause of death in men and women [2, 3]. However, according to a report of world health organization (WHO) in 2005, the prevalence and mortality due to coronary heart disease (CHD) is declining in the developed nations. Although rates of CVD have long since peaked for many developed countries and mortality from the disease is declining, it still accounts for almost a third (32.8%) of all deaths in the U.S. and a large majority of cardiac deaths in the U.S. are due to coronary heart disease (CHD) [4].

In 2009, CHD accounted for 64% of all Cardiac deaths in the U.S [4]. Cognitive decline and CVDs share many vascular risk factors (VRFs) such as smoking, hypertension, and diabetes mellitus; furthermore,

CVDs can contribute to cognitive decline by causing cerebral hypoperfusion, hypoxia, emboli, or infarcts [5]. Also, hypertension is a leading modifiable risk factor for heart disease and stroke [6-8]. Again, for developing hypertension there are some factors [9], which may not directly related with CHD but secondarily associated with CHD. Therefore, it is important to identify the associated risk factors to predict the risk of coronary heart disease (CHD) among individuals. Several research [10-12] has been conducted so far to identify the association of risk factors to coronary heart diseases (CHD). The focus of this study was to develop a binary logistic regression model to explore the relationship between the associated risk factors such as gender, age, number of cigarettes consumptions, total cholesterol level, glucose level etc. with 10-year risk of coronary heart disease (CHD) for the residents in Framingham, Massachusetts, U.S, as our outcome variable is a dichotomous in nature.

In recent years, the rate of CHD is increasing rapidly and researchers of U.S are engaging themselves to identify the factors of future risk of coronary heart diseases (CHD) [13-18]. The associated Research Questions are: Are men more susceptible to heart disease than women? does increased age and number of cigarettes smoked, show increasing odds of having heart disease? does increase in total cholesterol level and glucose level, increases the likelihood of having 10-year risk of coronary heart disease CHD?. The organization of the paper was as follows: In Section 2, we described data collection procedure and description of variables used in the analysis. Section 3 discussed about the methodology and analysis procedure. In Section 4, we discussed results of the final fitted logistic regression model. This paper ends with conclusion and possible discussion about the findings in Section 5.

2. RESEARCH METHOD

The dataset is publicly available on the Kaggle website [19], and it is an ongoing cardiovascular study on residents of the town of Framingham, Massachusetts. The dataset provides the patients' information. To answer associated research questions and to obtain the best fit model, our dataset consists over 4,250 records and 15 attributes. Following data description provides the variable name and its description. Each attribute is a potential risk factor. There are both demographic, behavioral and medical risk factors: Sex: male or female of patient; Age: age of the patient, CurrentSmoker: whether the patient is a current smoker (yes, no), CigsPerDay: the number of cigarettes the person smoked on average in one day, BPMeds: whether the patient was on blood pressure medication (yes, no), prevalentStroke: whether the patient had previously had a stroke (yes, no), prevalentHyp: whether the patient was hypertensive (yes, no), diabetes: whether the patient had diabetes or not (yes,no); totChol: total cholesterol level of patient, sysBP: systolic blood pressure of patient; diaBP: diastolic blood pressure of patient; BMI: Body Mass Index of patient; glucose: glucose level of patient, heartRate: heart rate (continuous-In medical research, variables such as heart rate though in fact discrete, yet are considered continuous because of large number of possible values.) Predict variable (desired target): 10-year risk of coronary heart disease CHD (binary: "1", means "Yes", "0" means "No"). In the dataset, there were missing values for CigsPerDay (29), BPMeds (53), TotChol (50), BMI (19), heartRate (1), and glcLevel (388). We did not consider any missing value analysis technique in our data analysis section.

It is customary to use Binary Logistic Regression as our outcome variable is dichotomous or binary (1=the patient has 10-year risk of CHD, 0=the patient does not have 10-year risk of CHD). The Framingham dataset consists of binary (or nominal) and continuous independent variables [20-23]. To select the best model for our analysis we used purposeful selection of covariates to minimize the number of variables (parsimony) in the model such that the resultant model is more likely to be numerically stable and is more easily generalized. First, we run the univariate analysis for each of the covariates, to select out preliminary main effects model. We selected all the variables as they are all significant at 20% level of significance that can be seen in Table 1.

Then we run the multivariable analysis with all the selected variables and found CurrentSmoker (p=0.6326), BPMeds (p=0.5330), PrevalentStroke (p=0.1415), PrevalentHyp (p=0.1162), Diabetes (p=0.9944), DiaBP (p=0.5332), BMI (p=0.4085), and HeartRate (p=0.5833) were not statistically significant that can be seen in Table 2, we excluded these variables for our main effect model. However, BMI, Diabetes, CurrentSomker and PrevalentHyp are clinically more significant for developing coronary heart diseases (CHD) [24], therefore, our preliminary main effect model is:

$$g(x) = ln\left(\frac{\pi(x)}{1-\pi(x)}\right) =$$
$$\beta_0 + \beta_1 Sex + \beta_2 Age + \beta_3 BMI + \beta_4 CurrentSmoker + \beta_5 CigsPerDay +$$
$$\beta_6 PrevalentHyp + \beta_7 Diabetes + \beta_8 SysBP + \beta_9 TotChol + \beta_{10} glcLevel \quad\quad (1)$$

Table 1. Analysis of maximum likelihood estimates

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	Odds Ratio	95 % CI for Odds ratio	
Sex	0.49916	0.0859	32.72	<.0001	1.635	1.382	1.935
Age	0.0746	0.00527	201.016	<.0001	1.078	1.066	1.089
CurrentSmoker	0.1084	0.8560	1.6028	0.2055	1.114	0.942	1.368
CigsPerDay	0.0127	0.00340	13.96	0.0002	1.013	1.006	1.020
BPMeds	1.0635	0.1960	29.44	<.0001	2.896	1.973	4.253
PrevalentStroke	1.4921	0.4052	13.5601	0.0002	4.447	2.010	9.899
PrevalentHyp	0.9837	0.0872	127.26	<.0001	2.674	2.254	3.173
Diabetes	1.1219	0.2035	35.89	<.0001	3.385	2.272	5.045
TotChol	0.0048	0.00093	27.75	<.0001	1.005	1.003	1.007
SysBP	0.0241	0.00180	178.445	<.0001	1.024	1.021	1.028
DiaBP	0.0317	0.00342	86.06	<.0001	1.032	1.025	1.039
BMI	0.0486	0.0099	23.65	<.0001	1.050	1.029	1.071
HeartRate	0.0052	0.00351	2.229	0.1360	1.005	0.998	1.012
glcLevel	0.0105	0.00157	44.588	<.0001	1.011	1.007	1.014

Table 2. Final model fit to the framingham CHD study data, n=3803

Parameter	Coeff.	Std. Err.	z	p
Intercept	-8.7215***	2.3584	13.675	0.0002
Sex	0.5276***	0.1066	24.5117	<0.0001
Age	0.1137***	0.0331	11.8026	0.0006
CurrentSmoker	0.0721	0.1534	0.2212	0.6381
CigsPerDay	0.0181**	0.00606	8.9462	0.0028
PrevalentHyp	0.2155	0.1330	2.6241	0.1052
Diabetes	-.7789	0.6041	1.3756	0.2408
TotChol	0.00242	0.00973	0.0616	0.8040
SysBP	0.0139***	0.00284	23.8504	<.0001
BMI	-0.0322	0.0641	0.2520	0.6157
glcLevel	-0.0132	0.0110	1.4463	0.2291
Age*TotChol	-0.00019	0.000136	2.0569	0.1515
glcLevel*TotChol	0.000070*	0.000042	2.7631	0.0965
BMI*TotChol	0.000169	0.00260	0.4210	0.5165
Diabetes*glcLevel	0.00668	0.00482	1.9215	0.1657

Almost all our independent variables (such as Age, BMI, CigsPerDay, SysBP, TotChol and glcLevel) are continuous, we checked the assumption of linearity in the logit for these variables. All other variables were linear in the logit except for glcLevel. We tried different transformations approache for this variable but none of the transformations support the assumption. We retain our original variable as it is for entering the final model. Then we included different pairs of interactions one by one into the main effects model which are clinically meaningful and keep only important pair of interactions such as age*totChol, bmi*totChol, diabetes*glcLevel and glcLevel*totChol based on the Wald statistic value that can be seen in Table 3, therefore our final preliminary model is:

$$g(x) = \beta_0 + \beta_1 Sex + \beta_2 Age + \beta_3 BMI + \beta_4 CurrentSmoker + \beta_5 CigsPerDay + \beta_6 PrevalentHyp + \beta_7 Diabetes + \beta_8 SysBP + \beta_9 TotChol + \beta_{10} glcLevel + \beta_{11} age * TotChol + \beta_{12} glcLevel * TotChol + \beta_{13} BMI * TotChol + \beta_{14} Diabetes * glcLevel \qquad (2)$$

In order to assess the fit of the model, from the Hosmer–Lemeshow goodness of fit statistic the corresponding p-value computed from the chi-square distribution with 8 degrees of freedom is 0.7591. Therefore, we do not reject the null hypothesis, and this indicates that the model seems to fit quite well. Deviance and Pearson Goodness-of-Fit Statistics also suggests that the model fits well as the p-values are 1.000 and 0.3396 respectively. The Percent Concordant has a high value of 74.2. This suggests that the model is good. We also check the Somers' D and Gamma statistic where both have moderate value 0.481. The area under the ROC curve of our selected model is 0.7416. This suggests that our selected model can describe discrimination appropriately and comparing with AUC of ROC for other models, our model is performing better in describing the discrimination. Therefore, we can say that our final model fits well. In order to find the subjects which seems to be poorly fit or over influential, we check the plots of Influence Diagnostic and predicted probability diagnostic as can be seen in Figure 1 and found that the subject ID 963 and ID 3488 are influential. The plot of standardized Pearson residuals against predicted values also suggests that the residuals are quite independent as can be seen in Figure 2.

Table 3. Interactions included one by one into the main effect model

Parameter	Coeff.	Std. Err.	Wald Chi-Sqr	P>Chi-Sqr
Age*Currentsmoker	-0.0133	0.0122	1.1879	0.2756
Age*BMI	-0.00066	0.00137	0.2336	0.6289
Age*CigsPerDay	0.000269	0.000479	0.3158	0.5741
Age*PrevalentHyp	-0.00142	0.0122	0.0135	0.9076
Age*Diabetes	-0.0284	0.0319	0.7909	0.3738
Age*TotChol	-0.00018	0.000134	1.8005	0.1796
Age*SysBP	-0.00015	0.000259	0.3247	0.5582
Age*glcLevel	0.000020	0.000231	0.0078	0.9298
Sex*totChol	0.00261	0.00215	1.4689	0.2255
Sex*SysBP	0.00593	0.00412	2.0692	0.1503
PrevalentHyp*glcLevel	0.00496	0.00342	2.1034	0.1470
BMI*glcLevel	0.000159	0.000296	0.2893	0.5907
TotChol*glcLevel	0.000067	0.000039	2.889	0.0892
BMI*TotChol	0.000277	0.000251	0.7499	0.3865
Diabetes*glcLevel	0.00612	0.00467	1.7157	0.1902

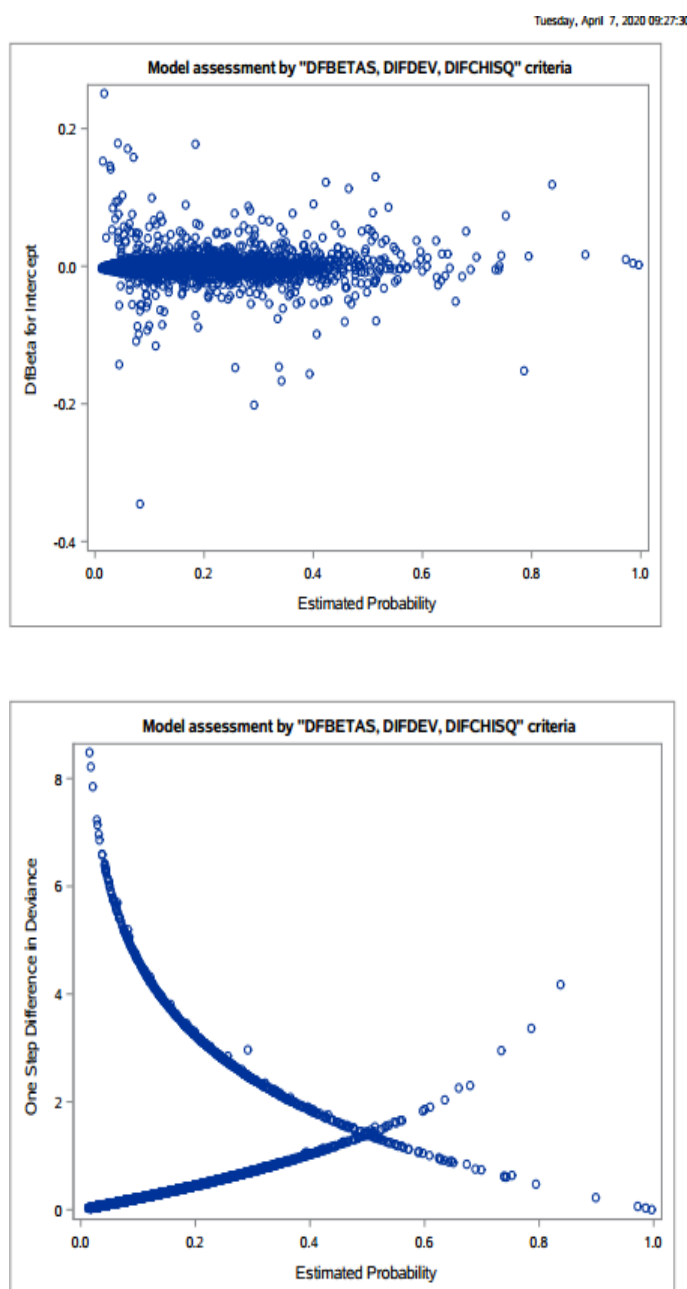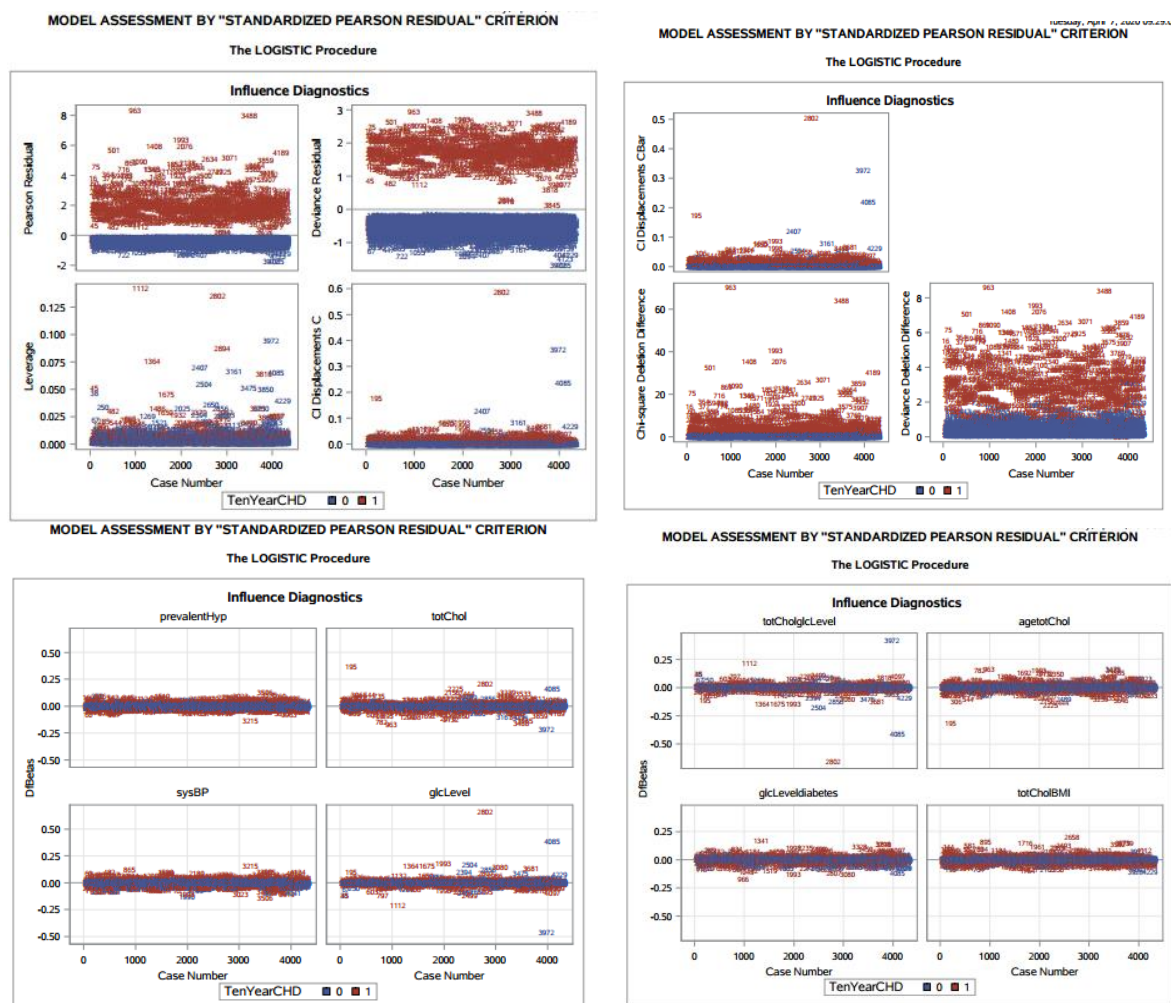Figure 1. Influence diagnostic and predicted probability

Figure 2. Standardize Pearson residuals

3. RESULTS AND DISCUSSION

Each of our fourteen covariates was statistically significant in the univariate analysis with 20% level of significance. The variable Prevalent Stroke has high odds ratio (4.447 odds of having 10-year risk of CHD for those who have previous history of stroke compared to those who do not have) compared to other covariates in the univariable analysis of Logistic regression that can be seen in Table 1. But in final Logistic Regression model, after excluding BPMeds, Prevalent Stroke, DiaBP and Heartrate, and adding four interactions that are clinically significant, only one (totChol*glcLevel) among them are clinically and marginally statistically significant, we found that CurrentSmoker (z=0.2212, p=0.6381).

PrevalentHyp(z=2.6241, p=0.1052), BMI, Diabetes, totChol, glcLevel, Diabetes*glcLevel, BMI*totChol, and age*totChol were not statistically significant, but we keep them in our model as they have clinical significance. Age is statistically significant (z=11.8026, p=0.0006) and the effect of age is modified by the effect of total cholesterol levels of the patients. When all the modification effects are at level of zero, the increase in age by 5 years increases the odds of having 10-year risk of CHD by 76% [exp(5*0.1137)=1.76]. Sex is highly statistically significant (z=24.5117, p<0.0001) and no other covariates modifies Sex effects. Also from Table 4 we observed the odds ratio for Sex (OR=1.695, CI=(1.375, 2.089)) with confidence interval, indicates that odds of having 10-year risk of CHD among males are 69.5% more than females by controlling all other independent variables. More specifically, males are more susceptible of having 10-year risk of CHD than females.

CigsPerDay is statistically significant (z=8.9462, p=0.0028) and no other covariates modifies CigsPerDay effects on outcome. Therefore, the patients who has taken CigsPerDay by 5(say) units higher than other patients, the odds of having 10-year risk of CHD for the patients with higher frequency of CigsPerDay is 9.4% [exp(5*0.0181)=1.094] higher than the patients with lower consumption of CigsPerDay. To the end, higher consumption of CigsPerDay increases the likelihood of having 10-year risk of CHD for our study population. SysBP is statistically significant as can be seen in Table 2, z=23.8504,

p<0.0001) and no other covariates modifies SysBP effects on outcome variable. Therefore, the patients who has SysBP by 5(say) units higher than other patients, the odds of having 10-year risk of CHD for the patients with higher SysBP is 7.2% [exp(5*0.0139)=1.072] higher than the patients with lower SysBP.

Table 4. Odds ratio estimates and profile-likelihood confidence intervals

Effect	Unit	Estimate	95% Confidence limits	
Sex	1.0	1.695	1.375	2.089
currentSmoker	1.0	1.075	0.796	1.452
cigsPerDay	1.0	1.018	1.006	1.030
prevalentHyp	1.0	1.240	0.956	1.610
sysBP	1.0	1.014	1.008	1.020

TotChol and BMI are clinically significant factor for cardiovascular diseases, however, they are not statistically significant in our case. Similarly, diabetes and glcLevel are potential risk factor for developing the cardiovascular disease, but these are not statistically significant. Only interaction that are clinically and marginally statistically significant for our data set is totChol*glcLevel (z=2.7631, p=0.09). It means that the effect of totChol for patients with 10-year risk of CHD varies according to different glucose level, keeping BMI and age values at average level that can be seen in Figure 3. To the end, the increase in total cholesterol level and glucose level, increases the likelihood of having 10-year risk of coronary heart disease CHD for our study population. Also, the effect of Diabetes was modified by the effect of glucose level. For subjects who have glucose concentration level above 100 and have Diabetes mellitus, the odds of 10-year risk of having CHD is much higher for them as can be seen in Figure 4.
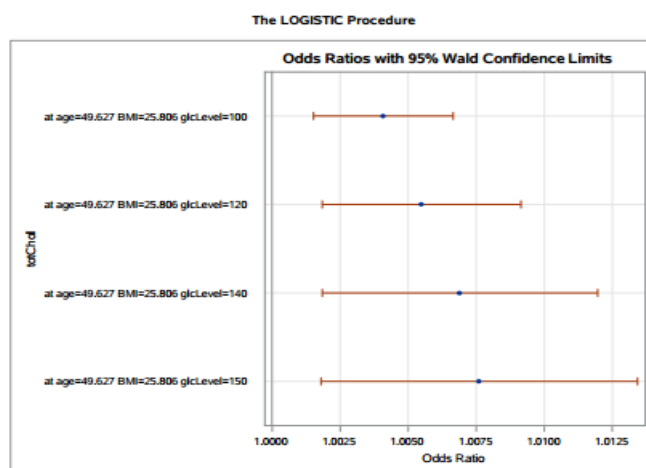

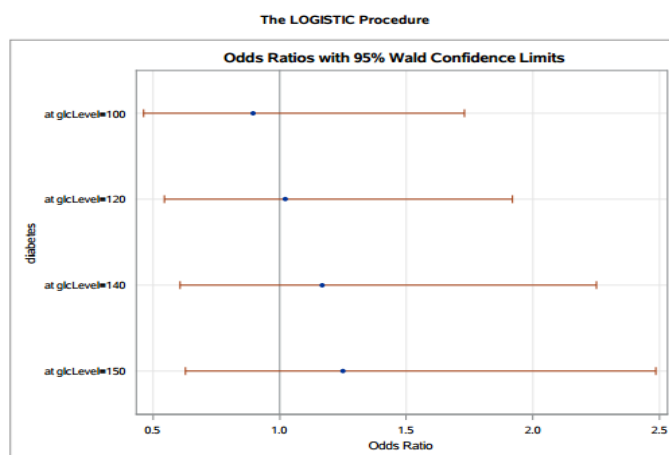Figure 3. Odds ratio of interaction effect of *totchol*glclevel* in final model

Figure 4. Odds ratio of interaction effect diabetes*glclevel in the final model

Our final model is both statistically and clinically significant and the goodness of model fit also support the model ability to perform well as the Area Under the ROC curve is moderately high to discriminate between those subjects who experience the outcome of interest and versus those who do not. In diagnostic checking using the standardized Pearson residuals and other techniques like DFBETAS, DIFDEV and DIFCHISQ, the model exhibits a good fitting considering no influential and outliers were present in our dataset. All the predictors of the model are clinically meaningful and then we see that except CurrentSomoker, PrevalentHyp, BMI, Diabetes, totChol, glcLevel, Diabetes*glcLevel, BMI*totChol, and age*totChol all other covariates are statistically significant in our final model. We found Sex, CigsPerDay, SysBP and Age which are risk factors for having cardiovascular disease, and are highly influential in predicting CHD that supports the literature [25-27]. We also found that the increased effect of age increased the likelihood of having CHD which supports the recent findings [28]. In addition, the effect of totChol is modified by the effect of glucose concentration level of the patients that also supports literature [29]. Moreover, the effect of Diabetes is modified by the effect of glucose concentration level for patients [30]. We found that the odds ratio of having CHD is highest at 1.249 (CI: 0.628, 2.485) when the glucose concentration level is high as of 150 units with diabetes cases which also supports the findings [31].

4. CONCLUSION

All the analysis has been done using SAS 9.2 software. From our analysis, we found that the coefficient of the coefficient of BMI, Diabetes, Glucose level, Age*glcLevel are negative which suggests that the odds of having 10-year risk of CHD will decrease with the increased value of those terms. On the other hand, the coefficient of SEX, Age, CurrentSmoker, CigsPerDay, PervalentHyp, totChol, SysBP, totChol*glcLevel, diabetes*glcLevel, BMI*totChol have positive signs, therefore the odds of having 10-year risk of CHD will increase with the increased value of those terms. The higher the magnitude of these coefficients the odds of having CHD will increase or decrease based on the positive or negative sign, respectively, of the respective coefficients. In conclusion, we found that Sex, Age, CigsPerDay and SysBP are important risk factors in predicting 10-year risk of CHD for Framingham study population and the effects of Age was modified with the totalChol level. However, the effect of totChol also is modified with glucose concentration level and BMI factors. The increased Age and CigsPerDay increases the odds of having 10-year risk of CHD, but the noticeable finding is that patients with Diabetes at higher level of glcLevel have the higher odds of having 10-year risk of CHD than with low level of glucose concentration. The limitation of this study was that the variable *glc*Level had highest number of missing values. With improving this limitation and including other important covariates a better prediction model could be developed to identify risk of having CHD for Framingham study population.

REFERENCES

[1] J. Mackay and G. A. Mensah, "The Atlas of Heart Disease and Stroke," *World Health Organization*, Geneva, 2004.

[2] Kirwan, "Coronary heart disease," *Health Knowledge*, 2006. [Online]. Available: https://www.healthknowledge.org.uk/public-health-textbook/disease-causation-diagnostic/2b-epidemiology-diseases-phs/chronic-diseases/coronary-heart-disease (accessed Jan. 07, 2020).

[3] R. S. Kovats, et al., Spring 2006 Reports : Annual Updates, vol. 29. 2006.

[4] A. S. Go et al., "Heart Disease and Stroke Statistics-2013 Update A Report From the American Heart Association," *Circulation*, vol. 127, no. 1, pp. e6-245, 2012.

[5] Qiu, C., Fratiglioni, L., "A major role for cardiovascular burden in age-related cognitive decline," *Nat Rev Cardiol*, vol. 12, no. 5, pp. 267-277, 2015.

[6] Wall HK, et al., "Vital signs: prevalence of key cardiovascular disease risk factors for Million Hearts 2022-United States, 2011-2016," *Morbidity and Mortality Weekly Report*, vol. 67, no. 35, pp. 983-91, 2018.

[7] Ritchey MD, et al., "US trends in premature heart disease mortality over the past 50 years: Where do we go from here?," *Trends Cardiovasc Med.*, vol. 19, no. S1050-1738, pp. 30134-3, 2019.

[8] Go AS, Bauman MA, et al., "An effective approach to high blood pressure control: a science advisory from the American Heart Association, the American College of Cardiology, and the Centers for Disease Control and Prevention," *Journal of the American College of Cardiology*, vol. 63, no, 12, pp. 1230-8, 2014.

[9] A. Talukder and M. Ali, "Assessing the true association between hypertension status and stature of individuals in Bangladesh: propensity score analysis," *Journal of Human Hypertension*, 2020. doi: 10.1038/s41371-020-0328-2.

[10] D. Mozaffarian, et al., "Heart disease and stroke statistics-2015 update : A report from the American Heart Association," *Circulation*, vol. 131, no. 4, pp. e29-322, 2015.

[11] A. Barham, R. Ibraheem, and S. H. Zyoud, "Cardiac self-efficacy and quality of life in patients with coronary heart disease : a cross-sectional study from Palestine," *BMC Cardiovascular Disorders*, vol. 19, no. 290, pp. 1-12, 2019.

[12] H. Das Lohano, "A Logistic Regression Analysis of the Ischemic Heart Disease Risk," *Journal of Applied Science*, vol. 6, no. 4, pp. 785-788, 2006.

[13] Ritchey MD, et al., "Notes from the Field: Characteristics of Million Hearts Hypertension Control Champions, 2012-2019," *Morbidity and Mortality Weekly Report (MMWR)*, vol. 69, no. 7, pp. 196-197, 2020.

[14] R. Moonesinghe, et al., "Prevalence and cardiovascular health impact of family history of premature heart disease in the United States: Analysis of the National Health and Nutrition Examination Survey, 2007-2014," *Journal of the American Heart Association*, vol. 8, no. 14, cir. e012364, 2019.

[15] J. Fang, et al., "Awareness of heart attack symptoms and response among adults-United States, 2008, 2014, and 2017," *Morbidity and Mortality Weekly Report*, vol. 68, no. 5, pp. 101-6, 2019.

[16] A.S. Vaughan, et al., "Widespread recent increases in county-level heart disease mortality across age groups," *Annals of Epidemiology*, vol. 27, no. 12, pp. 796-800, 2017.

[17] S. Greer, L. et al., "County health factors associated with avoidable deaths from cardiovascular disease in the United States 2006–2010," *Pub Health Rep*. 2016, vol. 131, no. 3, pp. 438-48, 2016.

[18] M. Casper, et al., "Changes in the geographic patterns of heart disease mortality in the United States, 1973 to 2010," *Circulation*, vol. 133, no. 12, pp. 1171-80, 2016.

[19] Kaggle, "Kaggle.com, Datasets. Framingham CHD database." 2020. [Online]. Available: https://www.kaggle.com/uciml/framingham-chd-database (accessed Feb. 10, 2020).

[20] B. T. Zewude and K. M. Ashine, "Binary Logistic Regression Analysis in Assessment and Identifying Factors That Influence Students ' Academic Achievement : The Case of College of Natural and Computational Science, Wolaita Sodo University, Ethiopia," *Journal of Education and Practice*, vol. 7, no. 25, pp. 3-7, 2016.

[21] J. Road, "Binary logistic regression analysis on ADDMITING students using jamb score," *International Journal of Current Research*, vol. 8, no. 1, pp. 25235-25239, 2016.

[22] N. Shrestha, "Application of Binary Logistic Regression Model to Assess the Likelihood of Overweight," *American Journal of Theoretical and Applied Statistics*, vol. 8, no. 1, pp. 18-25, 2019.

[23] H. E. Akyuz and H. Gamgam, "Comparison of binary logistic regression models based on bootstrap method: An application on coronary artery disease data," *Gazi University Journal of Science*, vol. 32, no. 1, pp. 318-331, 2019.

[24] Yoshihiro Kokubo, et al., "Hypertension is a risk factor for several type of Heart Diseases: Review of Prospective Studies," *Advances in Experimental Medicine and Biology*, vol. 956, pp. 419-426, 2017.

[25] Khan, SS, et al., "Association of Body Mass Index with Lifetime Risk of Cardiovascular Disease and Compression of Morbidity," *JAMA Cardiol.*, vol. 3, no. 4, pp. 280-287, 2018.

[26] Huo X, et al., "Risk of non-fatal cardiovascular disease in early-onset versus late-onset type2 diabetes in China: a cross-sectional study," *The Lancet Diabetes & Endocrinology*, vol. 4, no. 2, pp. 115-124, 2016.

[27] William B. Kannel, "Is the Relation of Systolic Blood Pressure to Risk of Cardiovascular Disease Continuous and Graded, or Are There Critical Values?," *Hypertension*, vol. 42, no. 4, pp. 453-456, 2003.

[28] Brain J. N., and David, A. S., "The Interaction between Aging and Cardiovascular Disease," *Circ Res.*, vol. 110, pp. 1097-1108, 2012.

[29] Maarit, H., et. al., "Interaction between cholesterol and glucose metabolism during dietary carbohydrate modification in subjects with the metabolic syndrome," *The American Journal of Clinical Nutrition*, vol. 84, no. 6, pp. 1385-1392, 2006.

[30] Huo X, et al., "Risk of non-fatal cardiovascular disease in early-onset versus late-onset type2 diabetes in China: a cross-sectional study," *The Lancet Diabetes & Endocrinology*, vol. 4, no. 2, pp. 115-124, 2016.

[31] Peterson, S., et al., "Coronary Heart Disease Statistics-2005 edition," *British Heart Foundation Health Promotion Research Group, Department of Public Health, University of Oxford*, 2006.