

## Role of Metabolic Obesity and Body Mass Index in Patients with Coronary Artery Diseases

Nikhil Gupta, Nihar Maheshwari and C.V. Arjun  
Department of Information Technology, Manipal Institute of Technology,  
Manipal University, Manipal, India

**Abstract:** Obesity and its related consequences are some of the most alarming health problems the world is currently facing. Studies have shown that a little more than a fourth of the adults all over the world are overweight and a little more than a tenth are obese. Also, obesity is considered as an independent risk factor for determining the severity of a Coronary Artery Disease (CAD). The aim of this study is to determine the role of metabolic obesity and body mass index in patients with coronary artery diseases. We train a model capable of predicting the GENSINI score which determines the severity of CAD in four health groups, namely Metabolically Healthy and Normal Weight (MHNW), Metabolically Obese Normal Weight (MONW), Metabolically Healthy Obese (MHO) and Metabolically Abnormally Obese (MAO). Several factors like glycated Hemoglobin (HbA1C), triglyceride and insulin resistance are considered. Ridge regression, Lasso regression and neural networks are the mechanisms used to train a model for predicting the GENSINI score based upon health statistics of patients.

**Key words:** Body mass index, coronary artery diseases, metabolic obesity, phenotypic obesity, GENSINI score, health statistics

---

### INTRODUCTION

Although associations between the development of cardiovascular diseases and obesity are known, the relationship between obesity as reflected by body mass index and angiographically demonstrated coronary artery diseases is not fully understood. Previous studies suggest obesity promotes the severity of a coronary artery disease (Atique *et al.*, 2016) (also cardiovascular diseases in general (Burke *et al.*, 2008; Jahangir *et al.*, 2014; Kwon *et al.*, 2013; Lavie *et al.*, 2009) but its relationship with BMI is still unclear. A thorough understanding of such a relationship can lead to more patient-centric measures for the prevention of CAD. Our study aims to more fully characterize the association of metabolic obesity and body mass index with coronary artery disease. This study presents the results obtained by training a model capable of predicting the GENSINI score which determines the severity of CAD in the following groups of people:

- Metabolically Abnormal Obese (MAO)
- Metabolically Healthy Obese (MHO)
- Metabolically Obese Normal Weight (MONW)
- Metabolically Healthy Normal Weight (MHNW)

According to DeFronzo *et al.* (2015), the prevalence of MHO in our population is low and is more common in women and younger people.

Using these results, the group showing a good association to severity of CAD (that is the group that is more prone to CAD), the metabolically obese or phenotypically obese is found out. The prognostic markers for CAD among factors like HbA1C, FI, HOMA-IR, TC, TG, HDL, LDL and hsCRP are also found out. These parameters are described in the next study.

### MATERIALS AND METHODS

**Data collection and definition:** Our patient population consisted of people who required hospitalization and investigation with coronary angiography for the treatment of suspected CAD. Patients of all ages with stable or unstable presentation of CAD were included. Our study was conducted at Kasturba Medical College, Manipal. Data collected included demographics, angiographic findings, cardiovascular risk factors, procedural variables and the type of treatment provided. Height and weight of each patient were also measured and recorded. The data collected contained the following attributes-age, BMI (Body Mass Index), WC (Waist Circumference), HbA1c

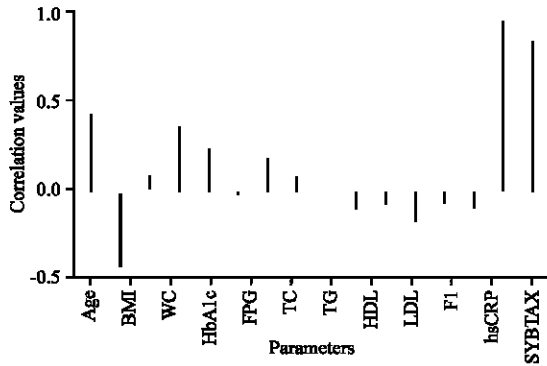


Fig. 1: Correlation with the GENSINI score in group 1

Table 1: Criteria for metabolic obesity

Criteria	Values
Waist Circumference (WC)	102 cm in men, 88 cm in women
Triglycerides (TG)	>150 mg/dL
HDL	<40 mg/dL in men, <50 mg/dL in women
Blood Pressure (BP)	130/85 mmHg
Glucose (FPG)	Fasting >110 mg/dL

(Glycated Hemoglobin), FPG (Fasting Plasma Glucose in mg/dL), TC (Total Cholesterol in mg/dL), TG (TriGlycerides in mg/dL), HDL (High Density Lipoprotein in mg/dL), LDL (Low-Density Lipoprotein in mg/dL), FI (Fasting Insulin), HOMA-IR (Insulin Resistance), hsCRP (High-Sensitive CRP), GENSINI and the SYNTAX score. Group values from 1-4 representing the four groups mentioned before were also recorded for each patient. For a person to be considered metabolically obese at least three of the following five criteria should be met (Table 1).

For a person to be considered phenotypically obese, their body mass index should be  $>25 \text{ kg/m}^2$ . The severity of coronary artery diseases is measured by the GENSINI and the SYNTAX scores. Greater scores represent higher severity.

**Statistics of the data:** Initially, the distribution of GENSINI scores with groups is considered. It is observed that in group 1: (Metabolically healthy and normal weight) and group 3: (Metabolically healthy and obese) a high majority of patients have low GENSINI scores which seems obvious considering their health status. On the other hand, patients in group 2: (Metabolically obese and normal weight) and group 4: (Metabolically abnormal and obese) tend to have high values for the GENSINI score.

In order to develop a regression model, it is necessary to understand the relationships between the input factors (health statistics of a patient in this case) and the prediction (GENSINI score) (Fig. 1-3). One strong correlation that is seen among all groups is that

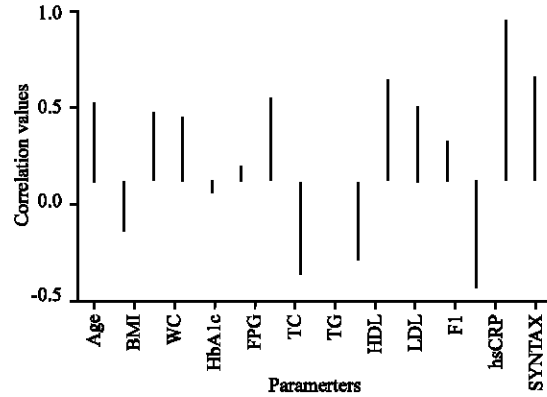


Fig. 2: Correlation with the GENSINI score in group 2

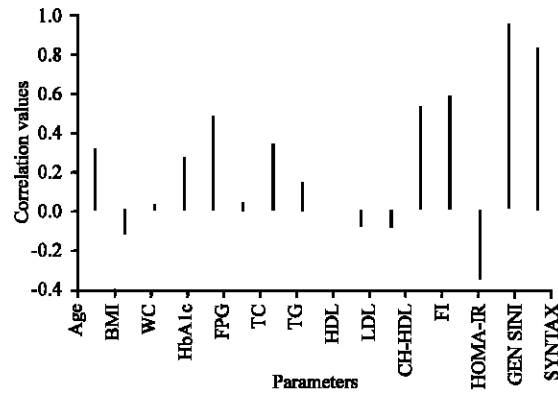


Fig. 3: Correlation with the GENSINI score in group 3

of GENSINI and SYNTAX. Considering that these are different metrics to measure the same factor, i.e., the severity of CAD, they are bound to be highly correlated ( $p\text{-value: } 2.22 \times 10^{-16}$  which is a strong indicator that the correlation holds good for the entire population). Therefore, variation in trends of the SYNTAX scores can be predicted from the trends of the GENSINI scores across various groups. Hence, SYNTAX scores have been eliminated from our models for training. Among other correlations which are significantly interesting, we see that among all groups, BMI is negatively correlated and age is positively correlated with the GENSINI score. Also, the total cholesterol is weakly correlated with GENSINI in the first and the third groups. This can be considered as an indication that the cholesterol values have little to do with the Coronary Artery Diseases in Metabolically Healthy patients. Also, Glycated Hemoglobin (HBA1C) is a parameter that is positively correlated amongst all the four groups and is an important factor in determining the GENSINI score (Fig. 4).

**Solution:** The dataset possesses multicollinearity. This means that two or more predictor variables of the multiple

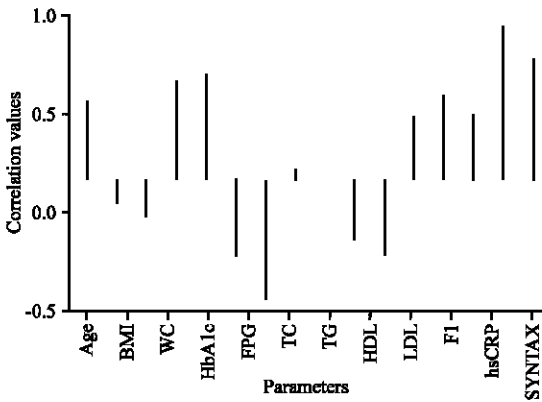


Fig. 4: Correlation of parameters with the GENSINI score in group 4

regression model are highly correlated. Hence, one variable can be linearly predicted from the others with a substantial degree of accuracy. This will lead to unfair weighing of the predictor variables and therefore must be removed.

It can be seen that for the entire dataset, TC and LDL as well as TC and CH-HDL are highly correlated. In groups 1, 3 and 4, FI shows high correlation with HOMA-IR and one of them has to be eliminated. This is one of the many highly correlated relations. A tabular description of all the variables showing high correlation is shown.

**Preprocessing of the data**

**Missing values:** Out of about 1500 values in the collected data, only 3 values all within FI were missing. Since, the numbers of missing values were proportionately small, a group wise average is taken to replace them. The average of FI for group 1 (35.9386) and group 4 (47.99) is used.

**Normalization:** The gradient descent algorithm is used as an optimization technique in order to estimate the coefficients of models. Such an optimization algorithm is shown to converge faster when the convex optimization function used is uniform. Here in our case, the gradient descent algorithm is used in order to develop regression models and the neural network. A uniform and symmetric convex optimization function can be obtained if all the entities under consideration are normalized. Normalization is done to permit this efficiency.

**Data splitting:** About 75% of the data was used for training the model and the remaining 25% to test it. Root mean square error is then found between the actual values and the predicted values of the 25% testing data.

The models considered for predicting the GENSINI scores from the given features are ridge regression, LASSO regression and neural networks. Since, the GENSINI score is a continuous value, the metric used for measuring the accuracy of predictions was the root mean square error. Ridge regression or weight decay method is a corrective measure in regression analysis taken to deal with the problem of multicollinearity amongst the several predictor variables under consideration. In a regression model involving many variables, if there exists a high correlation amongst predictor variables, then the contribution or the importance of a variable depends on which other variables are considered in the model. Hence, the input variable does not reflect its natural effect on the response as it is affected by other input variables. This results in poor predictions in cases with a large number of factors. Ridge Regression solves this problem to a fair extent by adding a bias to the variables. In our case, the initial idea to ensure that over-fitting does not occur is to give a penalty to the cost function for every unit increase in the parameter values. However, the problem with ridge regression is that it cannot zero out coefficients; thus, either all factors are included or none (Fig. 5).

LASSO (Least Absolute Shrinkage and Selection Operator) method allows selection of variables to some extent. LASSO is a regression method based on penalty, in a sense that it assigns a penalty to the size of the regression coefficients. Due to this, there may be situations where a few of the parameter estimates may be exactly zero, thus, allowing feature selection. As the penalty applied increases, size of the parameters on which the penalty is applied reduces causing them to zero out. This helps with the case of multiple highly correlated features where a regular regression will result in large predictor coefficients. A LASSO Model will enable an automatic selection for such predictors. Hence, it was thought of as a nice approach to our problem which consisted of a dataset possessing multicollinearity.

Further, artificial neural networks are computational models which are formed from a large collection of units (neural units). Each such unit is connected with many others, forming a network and links can enhance or inhibit the activation state of adjoining neural units just like their counterparts in nature. Each individual neural unit computes using the sigmoid function as an activation function. This activation function or limiting function on each connection ensures that the signal surpasses the limit before propagating to other artificial units. The system in this process learns primitive features of the problem in the initial layers of the network and starts learning deeper and more natural features through the

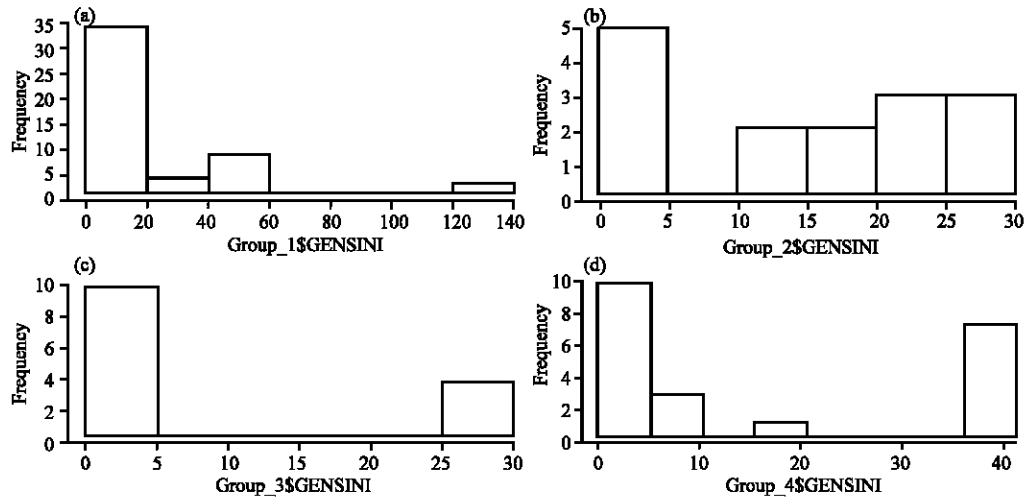


Fig. 5: Histograms showing frequency against GENSINI scores in each group: a) Histogram of group\_1\$GENSINI; b) Histogram of group\_2\$GENSINI; c) Histogram of group\_3\$GENSINI and d) Histogram of group\_4\$GENSINI

layers. These systems are not explicitly programmed but are self-learning machines trained based on the data available. Neural networks with varying number of inner layers were also used for predicting the values. The best results obtained for the test data were from using a network with only a single hidden layer. Including more layers caused the model to over-fit.

### RESULTS AND DISCUSSION

Using LASSO regression, ridge regression and neural networks individually on each group and collectively on the entire data, it is observed that neural networks give the best result that is the lowest RMSE (Root Mean Square Error) of 6.170 when it is treated on the entire dataset together. Considering each group individually, LASSO Model gives the best results for groups 2 and 3 and neural networks again, gives the best results for groups 1 and 4.

The ridge regression model is eliminated as a potential model because the results obtained using the test data are very inaccurate. Results obtained from the other selected models in each group with the test data are not very different from the actual recorded values. Hence, we can say that our model is not overfitting the training data. Results obtained after applying each model on the test data are given in Table 2.

**Relationship between obesity types and CAD:** A 95% confidence interval is found for the GENSINI scores within each group. For the metabolically obese normal weight group, the GENSINI scores lie within the range of

Table 2: Pearson's correlation coefficients

Variables	Values
Group 1	0.852783799856: TC and LDL 0.887927793492: FI and HOMA_IR
Group 2	0.87153020518: HbA1c and HOMA_IR -0.950003408973: HDL and CH_HDL
Group 3	0.827154839926: Age and HDL 0.810336764663: HbA1c and TC 0.820247281312: FPG and HDL 0.857193774502: FPG and HOMA_IR 0.984325207983: TC and LDL
Group 4	0.977675396139: FI and HOMA_IR 0.937306254471: TC and LDL 0.838328141656: TC and CH_HDL 0.865079811939: LDL and CH_HDL 0.959614719483: FI and HOMA_IR

For the entire data together; 0.70889724141: GROUP and BMI; 0.897603908389: TC and LDL; 0.717531489708: TC and CH-HDL

Table 3: Obtained root mean square errors

Models/Groups~	Group 1	Group 2	Group 3	Group 4
Ridge regression	21.18	24.14	510.90	20.47
LASSO regression	14.33	8.82*10E-15	4.79	10.55
Neural networks	4.11	0.14	8.81	0.55

Table 4: Obtained root mean square errors

Factors	Correlation	p-values	Statistical significance ( $\alpha = 0.05$ )
hsCRP	0.10900	0.279100	No
FI	-0.01330	0.895000	No
BMI	-0.22400	0.024900	Yes
WC	0.02280	0.821400	No
AGE	0.41650	0.000016	Yes
HBA1C	0.34300	0.000475	Yes
HOMA IR	0.09821	0.330900	No
TC	-0.07320	0.468800	No
TG	0.05696	0.555600	No
HDL	0.07350	0.465800	No
LDL	-0.14321	0.155100	No

9.5-19.7. For the metabolically healthy but obese group, this range is found to be 3.49-14.64 (Table 3-5). It is between 9.53 and 22.3 for the metabolically

Table 5: Obtained root mean square errors

Groups	Feature importance
Group 1 (MHNW)	WC>TG>HOMA_IR>AGE
Group 2 (MONW)	WC>LDL>F>HOMA-IR
Group 3 (MHO)	BMI>hsCRP>WC>FI
Group 4 (MAO)	hsCRP>LDL>CH_LDL>HbA1c
Entire dataset	hsCRP>FI>BMI>WC

abnormal obese group. It can be seen that the severity of CAD follows the trend: metabolic and phenotypic obese>metabolic obese>phenotypic obese.

### CONCLUSION

Overall if a single model is to be used (the group number is used as a parameter here) then neural networks give the best results. A shift from phenotypic to metabolic obesity indicates an increase in the severity of CAD. Also, it can be seen that BMI, AGE and HBA1C can be used as prognostic markers to determine severity because they are statistically significant. HBA1C and Age have positive correlations with GENSINI. BMI has a negative correlation (similar to result by Gomez-Huelgas *et al.*, 2013) with it. These features affect GENSINI scores directly while the other important features obtained previously from neural networks given affect GENSINI as a combination with one another and not directly. They play an important role in severity prediction but we cannot say how they affect GENSINI directly without further analysis.

### REFERENCES

Atique, S.M., B. Shadbolt, P. Marley and A. Farshid, 2016. Association between body mass index and age of presentation with symptomatic coronary artery disease. *Clin. Cardiol.*, 39: 653-657.

Burke, G.L., A.G. Bertoni, S. Shea, R. Tracy and K.E. Watson *et al.*, 2008. The impact of obesity on cardiovascular disease risk factors and subclinical vascular disease: The multi-ethnic study of atherosclerosis. *Arch. Internal Med.*, 168: 928-935.

DeFronzo, R.A., E. Ferrannini, P. Zimmet and George Alberti, 2015. *International Textbook of Diabetes Mellitus*. John Wiley & Sons, Hoboken, New Jersey, Pages: 1240.

Gomez-Huelgas, R., D. Narankiewicz, A. Villalobos, J. Warnberg and J. Mancera-Romero *et al.*, 2013. Prevalence of metabolically discordant phenotypes in a mediterranean population the IMAP study. *Endocr. Pract.*, 19: 758-768.

Jahangir, E., D.A. Schutter and C.J. Lavie, 2014. The relationship between obesity and coronary artery disease. *Transl. Res.*, 164: 336-344.

Kwon, B.J., D.W. Kim, S.H. Her, D.B. Kim and S.W. Jang *et al.*, 2013. Metabolically obese status with normal weight is associated with both the prevalence and severity of angiographic coronary artery disease. *Metab.*, 62: 952-960.

Lavie, C.J., R.V. Milani and H.O. Ventura, 2009. Obesity and cardiovascular disease. *J. Am. Coll. Cardiol.*, 53: 1925-1932.