Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 16s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Statistical Analysis on Myocardial Infarction

Vishwajit Vilas Khajekar[1], Vaishnavi Dnyaneshwar Kale[1], Amruta Arvind Sawant[1], Shradha Sanjay Bansode[1], Dipali Dnyaneshwar Kale[1], Aparnnaa Arun Kulkarni[2*]

[1]Department of Statistics, New Arts Commerce and Science College Ahmednagar

[*2]Aparnnaa Arun Kulkarni, Department of Statistics, New Arts Commerce and Science College Ahmednagar

Email ID: khajekarvishwajit@gmail.com

Email ID: vaishnavikale51@gmail.com

Email ID: amrutasawant1961@gmail.com

Email ID: shradhasbansode@gmail.com

Email ID: dipalikale8530@gmail.com

Email ID: aparnagurvale@gmail.com

Cite this paper as: Vishwajit Vilas Khajekar, Vaishnavi Dnyaneshwar Kale, Amruta Arvind Sawant, Shradha Sanjay Bansode, Dipali Dnyaneshwar Kale, Aparnnaa Arun Kulkarni, (2025) Statistical Analysis on Myocardial Infarction. *Journal of Neonatal Surgery*, 14 (16s), 769-774.

ABSTRACT

Myocardial Infarction (MI) is one of the most common causes of death globally, with numerous risk factors leading to its development. The purpose of this study is to investigate the relationship between demographic and lifestyle factors and the development of myocardial infarction. Variables examined were age, gender, blood group, body mass index, blood pressure levels, marital status, qualification, number of meals taken per day, meal type, consumption of meat, food eating habits, smoking habits, stress, acidity, and tiredness etc.

The chi-square test of independence was applied to show significant association between BMI, smoking, eating habits, alcohol drinking, blood group, and oily food eating with higher risk of heart disease. Moreover, risk ratios were obtained to measure the intensity of such relationship. Findings point toward main modifiable risk factors of inappropriate dietary intakes, tobacco use, and alcohol consumption significantly causing myocardial infarction. This study highlights the need to change such lifestyle dimensions to control heart disease.

We have used five different techniques to predict heart disease. The implemented algorithms are logistic regression, Random Forest, Decision tree, K-Nearest Neighbour, Support Vector Machine, Naïve bayes algorithm. The experiment shows that Support Vector Machine has the highest accuracy around (89.47%).

Keywords: *Myocardial Infarction, Dietary regime, Physical activity, Stress, Lifestyle.*

1. INTRODUCTION

Myocardial infarction is a severe cardiovascular condition resulting from the obstruction of blood supply to the heart muscle. It has been linked to several demographic and lifestyle variables such as age, gender, smoking status, dietary patterns, and prior health status. The growing incidence of Myocardial Infarction worldwide highlights the importance of knowledge and prevention of its risk factors, many of which are amenable to modification.

In Maharashtra COVID- 19 had a significant impact on cardiovascular disease. In 2019 the number of deaths in Mumbai due to cardiovascular disease before pandemic was 5849, in 2020 it decreases slightly to 5633. After pandemic the cases were 9490 i.e on an average 26 deaths per day, so to find the reason behind this we decide to do analysis on Myocardial Infarction. In Indore district, the number of heart attack cases nearly doubled from 1,716 in 2019-2020 to 3,235 in 2021-2022.

Our analysis aims to find what are the reasons behind increase in the cases of heart disease as well as increasing deaths due to heart disease. By using statistical methods to aim to identify key risk factors, patterns, trends associated with myocardial infarction.

A number of studies have established that people with relative (parents, siblings) who have MI are at much risk of developing the condition themselves. The studies also shown that people with diabetes are at much greater risk of developing heart disease, including myocardial infarction, than people without diabetes.

This research investigates the correlation between a variety of variables, such as BMI, blood pressure (BP), diet, lifestyle, and demographic variables with the incidence of MI. A thorough analysis was carried out using a sample population to investigate the correlation of these variables with heart disease. The chi-square test indicated that there were significant correlations between increased BMI, smoking, alcohol use, blood group, and the intake of oily foods with an elevated risk of MI. Additionally, risk ratios were determined to measure the strength of these correlations. The results indicate that lifestyles changes, including improved diet, less alcohol use, and avoidance of smoking, may contribute substantially to the prevention of MI. This research adds to the increasing evidence on modifiable risk factors and sheds important light for clinicians and policymakers seeking to enhance heart health outcomes.

2. LITERATURE REVIEW

Nayab Akhtar presented their paper on Heart Disease Prediction (February 2021). The purpose of this study was to find which classification technique is best for predicting Heart Disease. Among these techniques Naïve Bayes gives highest accuracy. And decision tree gives lowest accuracy.

Shazia Chowdhary, Vanita Gupta, Shagufta Chowdhary examined the study of association of blood group with ischemic heart disease. They discovered that the likelihood of ischemic heart disease was three times greater in blood group A and B compared to blood group O and AB. When genderwise distribution was done it was observed that males with blood group B and females with blood group A had highest prevalence of MI.

Santosh Kumar Sinha, Vinay Krishna, Ramesh Thakur, Ashutosh Kumar, Vikas Mishra, Mukesh Jitendra Jha, Karandeep Singh, Mohit Sachanm Rupesh Sinha, Mohammad Asif, Nasar Afdaali, Chandra Mohan Varma analyzed the MI in very young adults. The aim of the present study was to compare risk factors among young adults (<30 years). The most common symptoms was chest pain, smoking, family history of Heart Disease, Stress, Obesity, physical inactivity.

Marlou A. M. Limpens, Eralds Asllanaj, Lisanne J. Dommershuijsen, Eric Boersma, M. Arfan Ikram, Maryam Kavousi, Trudy Voortman investigated healthy lifestyles in older adults and life expectancy with and without heart failure. They found that overall healthy lifestyle containing a healthy diet, physically active, no smoking, a healthy weight and no to low alcohol use, can have a positive impact on prevention of Heart Failure and on total life expectancy and more years lived without heart failure.

3. MATERIAL AND METHODS

Chi-square test:

The chi-Square Test is a statistical method used to determine whether there is a significant association between two categorical variables. In this research, the test is applied to examine the relationship between dependent variable and independent variable. (e.g.: Alcohol intake with respect to heart disease.). The chi-square Test for independence tests whether these variables are associated or if the association is random.

Logistic Regression:

Logistic regression is a statistical method for binary classification, with the dependent variable having just two possible values (disease, no disease). Logistic regression is an extension of linear regression but is different in its formulation of the relationship between the independent variables and the dependent variable. Logistic regression does not predict the continuous values but rather estimates the probability that an instance belongs to a class. e.g.: disease or no disease. It's easy to interpret, efficient, and effective in most real-world classification problems.

Random Forest:

Random Forest is an advanced machine learning method applied to classification and regression tasks. It is constructed from several Decision trees, hence a stronger and more accurate model. It is commonly applied in real world scenarios because of its high accuracy, capacity to deal with missing values, and immunity to overfitting. Random Forest is a powerful and versatile machine learning algorithm that delivers high accuracy, robustness, and reliability across various domains.

Decision Tree Algorithm:

A Decision Tree is one of the most common supervised machine learning algorithms for both classification and regression problems. It is a tree-like model in which every internal node is a decision based on a feature, every branch is an outcome, and every leaf node is a final prediction. Decision Trees are very common because they are easy to understand, simple, and can handle both categorical and numeric data. Decision Trees are a straight forward but efficient machine learning algorithm employed in numerous real-world scenarios.

Support Vector Machine:

Support Vector Machine is a supervised learning algorithm that can be applied for classification or regression tasks. Though it's utilized for both classification and regression, it's mostly applied to classification problems owing to its capability to generate an optimal decision boundary between classes. SVM is well-versed in dealing with high-dimensional datasets and is also employed extensively in applications of text classification, image recognition, and medical diagnosis.

K-Nearest Neighbors:

K-Nearest Neighbors (KNN) is a supervised learning classifier and regression algorithm. It is a lazy ad non-parametric learning algorithm, as it does not make any assumption about the structure of the data and does not create a model at training time. KNN is a very strong and easy to use algorithm that performs well in classification and regression, particularly with small datasets. Effective feature selection, scaling and selecting an optimal K value are essential for its accuracy improvement.

Naïve Bayes Algorithm:

Naïve Bayes is a probabilistic machine learning classifier. It relies on Bayes theorem to determine the probability of a class given some features. The name "Naive" comes from the fact that it makes the assumption that all features are independent of one another, something that is not always the case but makes computation easier.

Odds Ratio:

Odds Ratio (OR) is a statistical measure to compare an event's odds of occurring in one group relative to another. OR is frequently applied in case control studies.

		OUTCOME	
		Disease (Case)	No Disease (Controls)
EXPOSURE	Exposed	a	b
	Unexposed	c	d

$$Odds\ of\ Exposure\ in\ Cases = \frac{Number\ of\ Cases\ with\ Exposure}{Number\ of\ Cases\ without\ Exposure} = \frac{a}{c}$$

$$Odds\ of\ Exposure\ in\ Controls = \frac{Number\ of\ Controls\ with\ Exposure}{Number\ of\ Controls\ without\ Exposure} = \frac{b}{d}$$

$$Odds\ Ratio = \frac{Odds\ of\ Exposure\ in\ Cases}{Odds\ of\ Exposure\ in\ Controls} = \frac{a/c}{b/d} = \frac{a*d}{b*c}$$

Risk Ratio:

Risk Ratio (RR) compares the likelihood of an event happening in an exposed group relative to that of a non-exposed group. RR is frequently applied in cohort studies and clinical trials.

$$RR = \frac{\text{Risk in Exposed Group}}{\text{Risk in Unexposed Group}} = \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

Overview of the variables:

Blood Group, Body Mass Index, Blood Pressure, Marital Status, Qualification, Number of meal times per day, Types of meal, Meat consumption, Food habits, Types of Cooking Oil, Smoking Habits, Stress, Acidity, Tiredness.

Procedure:

The dataset is obtained by the online survey conducted through google forms. The survey consisting of questions on dietary regime, daily routine, health related questions, family history of MI disease.

4. DATA ANALYSIS

Chi-Square Test of Independence: This chi-square test is used to check is there any association between MI and variables viz blood group, BMI, Blood pressure, Marital status, Qualification, food habits, smoking habits, Stress, Acidity, Tiredness.

Table 1:

Attribute 1	Attribute 2	P-Value	Conclusion
Age BMI Family History	Myocardial Infarction	2.2e-16 4.651 e-8 1.788e-5	Age, BMI and family history are associated with Myocardial Infarction.
Oily Food Fast Food	Myocardial Infarction	0.0003876 0.009069	Oily food, fast food are associated with MI.
Smoking Alcohol Stress	Myocardial Infarction	0.0345 2.2e-16 0.02234	Smoking, alcohol intake, stress are associated with MI.
Physical Activity	Myocardial Infarction	0.03602	Physical activity and corona disease are associated with MI
Marital Status Cooking Oil Corona Disease	Myocardial Infarction	0.05624 0.5172 0.7651	Marital Status, cooking oil and corona disease are not associated with MI.

Modelling of the Data using Machine Learning Techniques:

Various supervised machine learning Techniques are used. Our aim is to find out accuracy of the data and which technique is best for predicting the model. The SVM algorithm gives more accuracy (89.47%) than other ML techniques. Reference [1]

Table 2:

Model	Logistic Regression	Random Forest	Decision Tree	SVM	KNN Algorithm	Naïve Bayes Algorithm
Train-Test Ratio	80:20	80:20	80:20	80:20	80:20	80:20
Accuracy (%)	87.72	82.46	73.68	89.47	84.21	57.89

Our aim is to find out among these variables which variable creates more risk to myocardial infarction. That's why we used risk ratio for the variables with respect to myocardial infarction.

Table 3:

Attribute 1	Attribute 2	Odds Ratio	Risk Ratio
Physical Activity	Myocardial Infarction	0.3	0.42
Alcohol Intake	Myocardial Infarction	17.22	0.615
Smoking	Myocardial Infarction	11.07	0.448

Fast Food	Myocardial Infarction	0.36	0.56
BMI	Myocardial Infarction	3.69	0.356
Acidity	Myocardial Infarction	1.7	0.155
Family History	Myocardial Infarction	3.13	0.296

5. CORRESPONDENCE ANALYSIS

Correspondence Analysis is a multivariate statistical technique used to explore relationships between categorical variables, visualizing associations and reducing data dimensions to a few key components.

This plot gives associations between health and lifestyle categorical variables. Important risk factors for heart disease, including Diabetes and BP (Blood Pressure), are located far from the central group, suggesting that they have a unique patterns and contribute significantly to cardiovascular conditions. The Age is located close to BP and Diabetes, hence we can conclude that older people are more likely to develop these conditions.

Lifestyle variables like Tobacco, Cigarette, and Paan are clustered, indicating a close relationship between the consumption of tobacco and health hazards. Stress, Tiredness, and Acidity are another cluster, and this could be an indication that chronic stress and digestive issues are interrelated. Poor eating habits like Cold Drinks, Oily Food, and Fast Food are clustered together, hence we can conclude that they are correlated to each other. Body Pain is far from the center hence we can interpret that the variable is independent on heart disease.

Overall, we can conclude that the variables high BP, Diabetes are likely to be more vulnerable to heart disease. Unhealthy living habits such as smoking, poor diet, and excessive stress can further add to health threats. Such understanding can be used in strategic health interventions based on dietary changes, stress regulation and preventive screening in high risk patients.

6. RESULTS

Chi-Square test gives significant association between myocardial infarction and variables like age, BMI, family history, oily food eating, fast food eating, smoking, alcohol drinking and exercise were found.

Marital status, cooking oil type and history of COVID-19 disease were independent variables that did not have a statistically significant effect on myocardial infarction.

The odds of MI is 17 times more likely in the patients exposed to alcohol intake than patients not exposed to alcohol intake. Similarly, 11 times more likely in the patients exposed to smoking than the patients not exposed to smoking. And 4 times more likely in the patients exposed to BMI than the patients not exposed to BMI.

The person who drinks and don't have MI would be 39% less likely to have risk of MI than the person who are exposed to MI. Similarly person who smokes and don't have MI would be 55% less likely to have risk of MI than the person who exposed to MI. Person who eats fast food and don't have MI would be 44% less likely to have risk of MI than the person who expose to MI.

7. DISCUSSION

The results shows that SVM and Logistic Regression models are the best in predicting myocardial infarction risk. The models proved to be highly accurate, implying that they can be utilized for early diagnosis and risk prediction.

Lifestyle variables including smoking, alcohol consumption, unhealthy eating habits, and physical inactivity were strongly linked with myocardial infarction. The results are consistent with current medical literature.

The machine learning method, specifically SVM and Logistic Regression, can prove to be very useful in decision- making as well as for assessing patient risk.

From odds ratio we got odds of different variables with respect to myocardial infarction. And from risk ratio we got risk factors among all these variables.

8. CONCLUSION

In the present research, the Chi-Square test was applied to examine the association between different risk factors and myocardial infarction. Dependent factors were age, BMI, food habits (oily food, fast food), smoking, alcohol intake, exercise, hereditary which means that as age increases the chances of occurrence of myocardial infarction also increase. Person whose BMI is greater than 30 i.e. overweight have more chances of getting myocardial infarction. Similarly excessive consumption

of oily food and fast food also increases the risk of myocardial infarction. Frequently consumption of alcohol and smoking also increases the risk of MI. Physical activity such as walking, running, yoga, gym, rope skipping, swimming helps us to decrease risk of MI. Family history of patient of heart disease were not affected as compared other variables. and independent factors were marital status, type of cooking oil and other lifestyle factors. Reference[2] which means that risk of MI is not depend on type of cooking oil such as sunflower oil, peanut oil, soyabean oil, rice bran oil etc.

Also, machine learning models were used in forecasting the probability of myocardial infarction.

In general, the results indicate that SVM and Logistic Regression are the most accurate models for predicting myocardial infarction. Reference [1]. The outcomes highlight the role of lifestyle parameters like diet, smoking, alcohol consumption, and exercise in myocardial infarction risk.

The analysis proves great value to healthcare professionals for the identification of high-risk subjects and the initiation of preventive measures.

The odds of alcohol intake, smoking, BMI are large as compared to other variables and factors like alcohol intake, smoking, BMI, physical activity creates major risk to myocardial infarction.

REFERENCES

[1] Akhtar, N. and Taqdees, S., 2021. Heart disease prediction. *ResearchGate*, p.11.

[2] Sinha, S.K., Krishna, V., Thakur, R., Kumar, A., Mishra, V., Jha, M.J., Singh, K., Sachan, M., Sinha, R., Asif, M. and Afdaali, N., 2017. Acute myocardial infarction in very young adults: A clinical presentation, risk factors, hospital outcome index, and their angiographic characteristics in North India-AMIYA Study. *ARYA atherosclerosis*, *13*(2), p.79.

[3] Limpens, M.A., Asllanaj, E., Dommershuijsen, L.J., Boersma, E., Ikram, M.A., Kavousi, M. and Voortman, T., 2022. Healthy lifestyle in older adults and life expectancy with and without heart failure. *European Journal of Epidemiology*, *37*(2), pp.205-214.

[4] Gupta, R. and Gupta, V.P., 1996. Meta-analysis of coronary heart disease prevalence in India. *Indian heart journal*, *48*(3), pp.241-245.

[5] Chandola, T., Britton, A., Brunner, E., Hemingway, H., Malik, M., Kumari, M., Badrick, E., Kivimaki, M. and Marmot, M., 2008. Work stress and coronary heart disease: what are the mechanisms?. *European heart journal*, *29*(5), pp.640-648.

[6] Chowdhary, S., Sharma, V. and Chowdhary, S., 2016. Study of association of ABO blood group with ischemic heart disease. *International Journal of Medical Science and Public Health*, *5*(03), p.468.