



# A STUDY OF SEVERAL PROBLEMS OF HEART PATIENTS IN KOLHAPUR DISTRICT USING STATISTICAL TECHNIQUES

**Prakash S. Chougule<sup>1</sup>, Tejaswi S.Kurane<sup>2</sup>, Mr.Pavankumar T.Patil<sup>3</sup>.**

**Mr. Gawade Sandesh R.<sup>4</sup>, Mr. Suraj R. Patil<sup>5</sup>**

*Associate Professor<sup>1</sup>, Assistant Professor<sup>2</sup>, Research Student<sup>3,4,5</sup>*

*Department of Statistics, Rajarshi Chhatrapati Shahu College, Kolhapur (MS), India*

## ABSTRACT

Heart disease is the leading cause of death. The term "heart disease refers to several types of heart conditions that affect the heart, Coronary artery disease Heart Failure,. Arrhythmia, heart valve disease and Pericardial Disease, Cardiomyopathy (Heart Muscle Disease) , Congenital Heart Disease ( CHD), are epidemic in India. The Registrar General of India reported that CHD led to 17% of total deaths and 26% of adult deaths in 2001-2003, which increased to 23% of total and 32% of adult deaths in 2010-2013. The World Health Organization and Global Burden of Disease Study also have highlighted increasing trends in years of life lost and disability-adjusted life years from Congenital Heart Disease in India .The Cardiovascular diseases have now become the leading cause of mortality in India. In order to study several problems of Heart Patients we collect the information through systematic questionnaires containing number of attributes and the collected information are analyzed using several statistical tools and techniques. Our study shows, the prevalence of heart diseases and heart diseases related complications are more in male than in female.

**KEYWORDS:** CHD, Parametric Test, Nonparametric Test, Level of Significance.

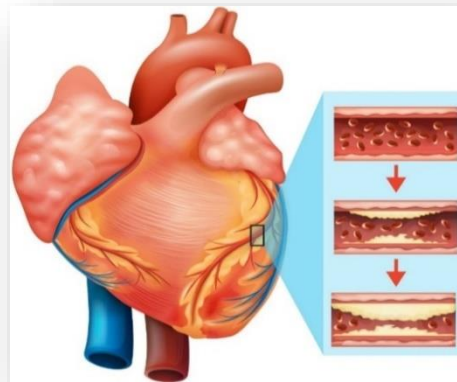
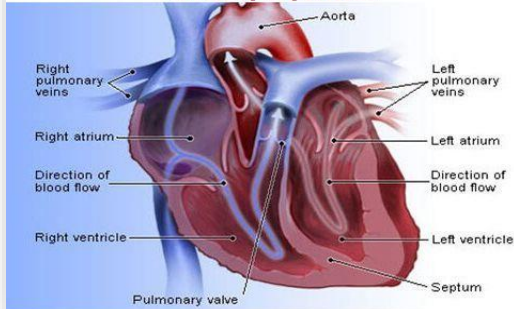
## INTRODUCTION

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications.

Heart is an important organ of the human body. It pumps blood to every part of our anatomy. If it fails to function correctly, then the brain and various other organs will stop working, and within few minutes, the person will die. Change in lifestyle, work related stress and bad food habits contribute to the increase in the rate of several heart-related diseases. Heart diseases have emerged as one of the most prominent causes of death all around the world. According to World Health Organisation, heart related diseases are responsible for taking 17.7 million lives every year, 31% of all global deaths. In India too, heart-related diseases have become the leading cause of mortality. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on 15<sup>th</sup> September 2017. Heart-related diseases increase the spending on health care and also reduce the productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India has lost up to \$237 billion, from 2005 to 2015, due to heart-related or Cardiovascular diseases. Thus, feasible and accurate prediction of heart-related diseases is very important. Medical organizations, all around the world, collect data on various health-related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend.

According to the World Health Organization (WHO), cardiovascular disease mortality will rise to almost 30 million by 2040 S. I. Ayon, M. M. et. al.(2020). Electrocardiogram (ECG), echocardiogram (heart ultrasound), cardiac magnetic resonance imaging (MRI), stress tests (exercise stress test, stress ECG, nuclear cardiac stress test), and angiography are commonly used tests by physicians to help identify cardiovascular problems. Poornima et. al. (2018) studied Effective heart disease prediction system using data mining techniques and conclude that This system performs realistically well even without retraining. Furthermore, the experimental results show that the system predicts heart disease with ~100% accuracy by using neural networks. In order to study the various pattern of heart patient in Kolhapur district. We prepare a systematic Questionnaire contains the number of attributes and collected information are analyzed using different statistical tools and techniques.

### List of Heart Diseases, causes and symptoms



## OBJECTIVES

- ❖ To identify the risk factors associated with heart diseases
- ❖ To study the association between diabetes & family history of diabetes
- ❖ To study association between diabetes & gender
- ❖ To study association between diabetes & exercise
- ❖ To study the association between blood pressure and diet
- ❖ To study the identify risk factors of heart disease
- ❖ To identify common types of heart disease
- ❖ To discuss the preventative measure
- ❖ To identify factors that contribute to heart disease
- ❖ To identify foods to avoid or limit in a cholesterol controlled diet

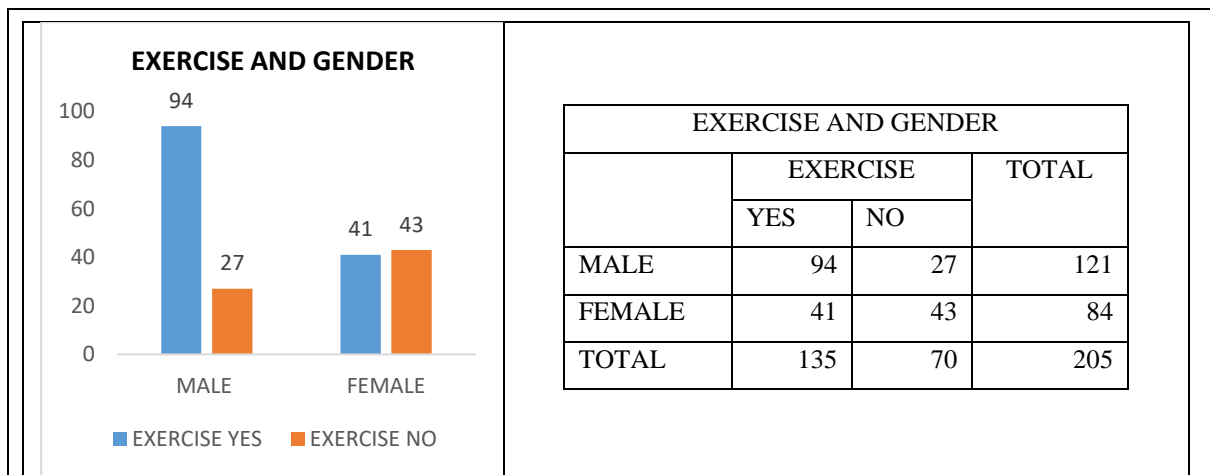
## METHODOLOGY

Primary data is collected through a structured questionnaire. The univariate, bivariate and multivariate analysis has been applied on the data. In univariate analysis the single variable has been uniquely examined by means of graph and percentages. In bivariate analysis, the relationships among variables were explored and in multivariate analysis, tried to identify risk factors of the diabetes along with their corresponding probabilities according to the survey conducted.

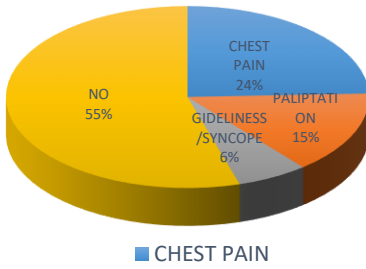
- **Statistical Tools Used**
  1. Graphical representation
  2. Descriptive Statistics
  3. Testing of hypothesis
- **Software Used:** MS-Excel MS-World



**GRAPHICAL REPRESENTATION**

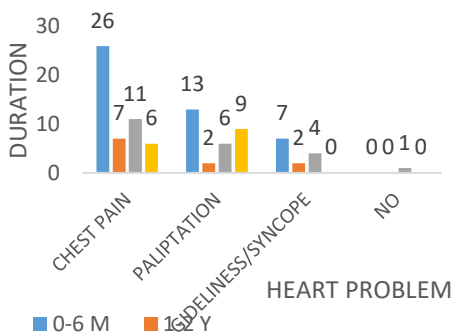


**SUFFERING OF HEART PROBLEM**



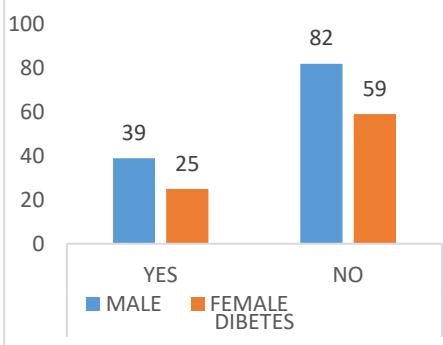
SUFFERING OF HEART PROBLEM	
CHEST PAIN	50
PALIPATATION	30
GIDELINESS/SYNCOPE	13
NO	112

**Duration and heart problem**



DURATION	HEART PROBLEM			
	CHEST PAIN	PALIPATATION	GIDELINESS/SYNCOPE	NO
0-6 M	26	13	7	0
1-2 Y	7	2	2	0
ABOVE 2 YEAR	11	6	4	1
6-12 M	6	9	0	0

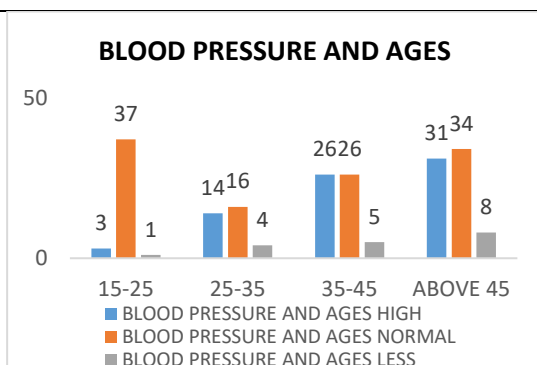
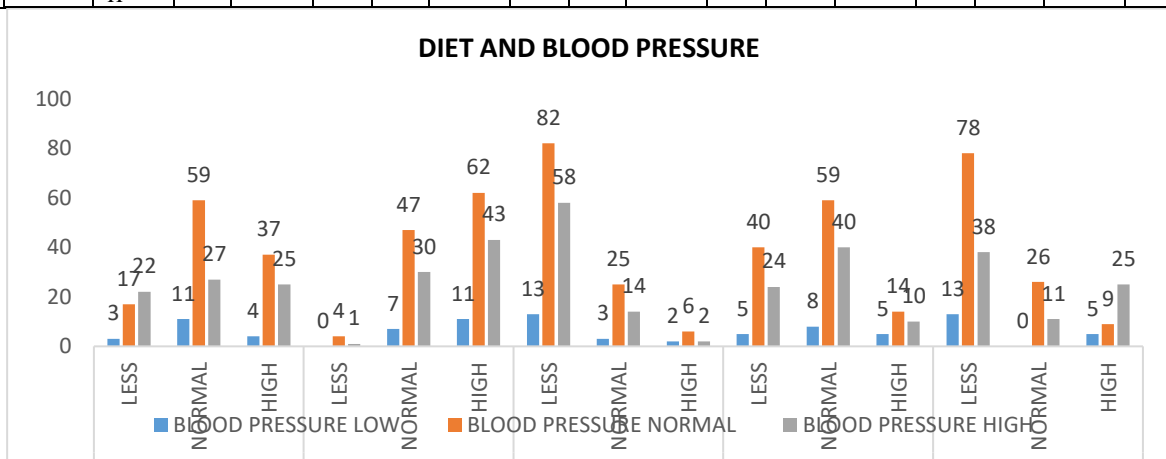
**DIABETES AND GENDER**



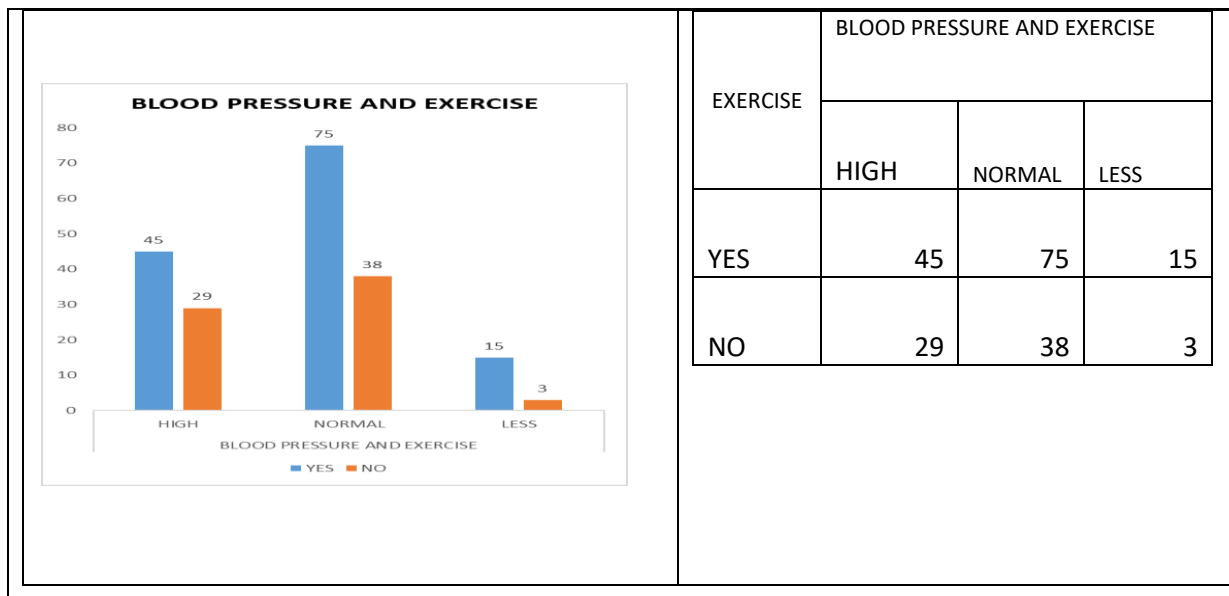
GENDER	DIBETES		TOTAL
	YES	NO	
MALE	39	82	121
FEMALE	25	59	84
TOTAL	64	141	



		DIET														
		NON-VAG			VEG			BAKERY PRODUCT			DAIRY PRODUCT			OILY PRODUCT		
		LESS	NORMAL	HIGH	LESS	NORMAL	HIGH	LESS	NORMAL	HIGH	LESS	NORMAL	HIGH	LESS	NORMAL	HIGH
BLOOD PRESSURE	LOW	3	11	4	0	7	11	13	3	2	5	8	5	13	0	5
	NORMAL	17	59	37	4	47	62	82	25	6	40	59	14	78	26	9
	HIGH	22	27	25	1	30	43	58	14	2	24	40	10	38	11	25



	BLOOD PRESSURE AND AGES		
	HIGH	NORMAL	LESS
15-25	3	37	1
25-35	14	16	4
35-45	26	26	5
ABOVE 45	31	34	8



**Testing of Hypothesis**

**a) Chi-square test for testing Independence between area and gender of heart Patients :-**

H<sub>0</sub>: Area and gender of heart disease are not independent V/S

H<sub>1</sub>: Area and gender of heart disease are independent.

Level of significance =  $\alpha = 5\% = 0.05$

**Observation Table**

AERA AND GENDER			
	AERA		TOTAL
	RUAL	URBAN	
MALE	67	54	121
FEMALE	45	39	84
TOTAL	112	93	205

$\therefore \chi^2_{Cal} = 0.0648$  ,  $\chi^2_{tab} = 3.8414$  and  $\therefore \chi^2_{Cal} < \chi^2_{tab}$

**a) Chi-square test for testing Independence between gender and blood pressure of heart Patients:-**

H<sub>0</sub>: Gender and blood pressure of heart disease are independent.

H<sub>1</sub> : Gender and blood pressure of heart disease are not independent.

Level of significance =  $\alpha = 5\% = 0.05$

**Observation Table**

GENDER AND BLOOD PRESSURE				
	BLOOD PRESSURE			TOTAL
	LESS	NORMAL	HIGH	
MALE	9	68	44	121
FEMALE	9	45	30	84
TOTAL	18	113	74	205

$$\therefore \chi_{Cal}^2 = 0.6739, \chi_{tab}^2 = 5.9914 \text{ and } \therefore \chi_{Cal}^2 < \chi_{tab}^2$$

**b) Chi-square test for testing association between Gender and Exercise of heart Patients:-**

H<sub>0</sub>: Gender and Exercise of heart disease are dependent.

H<sub>1</sub> Gender and Exercise of heart disease are not dependent.

Level of significance =  $\alpha = 5\% = 0.05$

**Observation Table**

EXERCISE AND GENDER			
	EXERCISE		TOTAL
	YES	NO	
MALE	94	27	121
FEMALE	41	43	84
TOTAL	135	70	205

$$\chi_{Cal}^2 = 18.3854, \chi_{tab}^2 = 3.8414 \text{ and } \therefore \chi_{Cal}^2 > \chi_{tab}^2$$

**c) Chi-square test for testing association between Gender and Diabetes of heart Patients:-**

H<sub>0</sub>: Gender and diabetes of heart disease are independent.

H<sub>1</sub> Gender and diabetes of heart disease are not independent.

Level of significance =  $\alpha = 5\% = 0.05$

**Observation Table –**

GENDER AND DIABETES			
	DIABETES		TOTAL
	YES	NO	
MALE	39	82	121
FEMALE	25	59	84
TOTAL	64	141	205



$$\therefore \chi_{\text{Cal}}^2 = 0.1408, \chi_{\text{tab}}^2 = 3.8414 \quad \text{and} \quad \therefore \chi_{\text{Cal}}^2 < \chi_{\text{tab}}^2$$

**d) Chi-square test for testing association between Surgery and Gender of heart Patients:-**

H<sub>0</sub>: Surgery and gender of heart disease are independent.

H<sub>1</sub>: Surgery and gender of heart disease are not independent

Level of significance =  $\alpha = 5\% = 0.05$

**Observation Table**

SURGERY AND GENDER			
	SUEGERY		TOTAL
	YES	NO	
MALE	38	83	121
FEMALE	16	68	84
TOTAL	54	151	205

$$\therefore \chi_{\text{Cal}}^2 = 3.9020, \chi_{\text{tab}}^2 = 3.8414 \quad \text{and} \quad \therefore \chi_{\text{Cal}}^2 > \chi_{\text{tab}}^2$$

**e) Chi-square test for testing association between Diabetes and Exercise of heart Patients:-**

H<sub>0</sub>: Diabetes and Exercise of heart disease are independent.

H<sub>1</sub>: Diabetes and Exercise of heart disease are not independent

Level of significance =  $\alpha = 5\% = 0.05$

**Observation Table**

DIABETES AND EXERCISE			
DIABETES	EXERCISE		TOTAL
	YES	NO	
YES	41	23	64
NO	94	47	141
TOTAL	135	70	205

$$\therefore \chi_{\text{Cal}}^2 = 0.1327, \chi_{\text{tab}}^2 = 3.8414 \quad \text{and} \quad \therefore \chi_{\text{Cal}}^2 < \chi_{\text{tab}}^2$$

**f) Chi-square test for testing association between Blood Pressure and Salt of heart Patients:-**

H<sub>0</sub>: Blood pressure and Salt of heart disease are dependent.

H<sub>1</sub>: Blood pressure of heart disease are not dependent

Level of significance =  $\alpha = 5\% = 0.05$



**Observation Table**

BLOOD PRESSURE AND SALT				
SALT	BLOOD PRESSURE			TOTAL
	HIGH	NORMAL	LESS	
HIGH	16	10	2	28
NORMAL	52	92	13	157
LESS	6	11	3	20
TOTAL	74	113	18	205

$$\chi_{\text{Cal}}^2 = 7.2377 \chi_{\text{tab}}^2 = 7.8147 \quad \text{and} \quad \therefore \chi_{\text{Cal}}^2 < \chi_{\text{tab}}^2$$

**g) Chi-square test for testing association between Age and Cholesterol level of heart Patients:-**

$H_0$ : Age and Cholesterol level of heart disease Patients are independent.

$H_1$ : Age of Cholesterol level of heart disease Patients are not independent

Level of significance =  $\alpha = 5\% = 0.05$

**Observation Table**

AGES	CHOLESTROL			TOTAL
	HIGH	NORMAL	LESS	
15-25	4	37	0	41
25-35	8	26	0	34
35-45	19	38	0	57
ABOVE45	18	55	0	73
TOTAL	49	156	0	205

$$\therefore \chi_{\text{Cal}}^2 = 7.2377 \chi_{\text{tab}}^2 = 7.8147 \quad \therefore \chi_{\text{Cal}}^2 < \chi_{\text{tab}}^2$$

**h) Kruskal-Wallis Test**

$H_0$ : There is significance difference between different periods of heart problem in male and female is same

$H_1$ : There is significance difference between different period of heart problem in male and female is not same

$$\text{Test statistic} - H = \frac{12}{N(N+1)} * \sum \frac{R_i^2}{n_i} - 3(N+1)$$

**Observation Table**

PERIOD OF HEART PROBLEM		
Duration	MALE	FEMALE
0-6 M	25	11
1-2 Y	5	3
ABOVE 2 Y	11	9
6-12 M	2	3
NO	35	32



$$H = \frac{12}{N(N+1)} * \sum \frac{T_i^2}{n_i} - 3(N + 1)$$

$$H = \frac{12}{10(10+1)} \left[ \frac{(14.5)^2}{2} + \frac{(6.5)^2}{2} + \frac{(11.5)^2}{2} + \frac{(3.5)^2}{2} + \frac{(19)^2}{2} - 3(10 + 1) \right]$$

$$H = 8.345455$$

$$\chi_{tab}^2 = 3.8141$$

Calculated value H < Tabulated value

### OVERALL CONCLUSION

- Despite of having lower BMI, people are getting caught with diabetes. Hence, we can conclude that high BMI or overweight is not only the risk factor of diabetes.
- The rate of heart patient is equally and rapidly increasing in rural areas as well as urban area.
- The Prevalence of heart diseases and heart diseases related complications are more in male than in female.
- When the sample data is classified according to the different age groups, we observed that in the younger age group (i.e. 25-40) more number of females are getting caught with diabetes. While, there is high probability of males being suffering from the disease in the middle (40-55) and older (55 & above) age groups.
- Study shows that there is significant association between diabetes and severe health complications such as cardiovascular (High B.P., Heart disease), ocular (impaired vision problems) and renal (kidney related) complications.

### REFERENCES

1. S. I. Ayon, M. M. Islam, and M. R. Hossain, "Coronary artery heart disease prediction: a comparative study of computational intelligence techniques," *IETE Journal of Research*, 2020.
2. R. Aggrawal and S. Pal, "Sequential feature selection and machine learning algorithm-based patient's death events prediction and diagnosis in heart disease," *SN Computer Science*, vol. 1, no. 6, 2020.
3. Gao, A. A. Ali, H. S. Hassan, and E. M. Anwar, "Improving the accuracy for analyzing heart diseases prediction based on the ensemble method," *Complexity*, vol. 2021, Article ID 6663455, 10 pages, 2021