Exploring Factors Associated with Heart Attack Risk Using Ordinal Logistic Regression

AKINBO R. Y. [1], OLATAYO T. O. [2], LASISI T. A. [3], ADEJUMO T. J. [3*], TITILOLA A. A. [2]

[1]Department of Statistics Federal Polytechnic Ilaro Ogun State, NIGERIA
[2]Department of Statistics Olabisi Onabanjo University, Ago-Iwoye Ogun State Nigeria, NIGERIA
[3]Department of Statistics Ladoke Akintola University of Technology Ogbomoso, NIGERIA

Abstract: Heart attack remains a major global public health concern, making the identification of key risk factors essential for effective prevention strategies. This study investigated factors associated with heart attack risk using ordinal logistic regression. Secondary data on 50,000 individuals were obtained from Kaggle.com and classified into three ordered risk levels: low, moderate, and high. Descriptive statistics, chi-square tests, and ordinal logistic regression were employed for analysis. The results showed that participants were predominantly middle-aged, with average body mass index, cholesterol level, and resting blood pressure falling within the overweight, borderline-high, and stage 1 hypertension ranges, respectively. Chi-square analyses indicated no significant association between heart attack risk and gender or chest pain type. Although the overall association between stress level and heart attack risk was not statistically significant, a significant linear trend suggested increasing risk with higher stress levels. Ordinal logistic regression further revealed that individuals with low stress levels were significantly more likely to belong to lower risk categories compared to those with high stress, while moderate stress showed no significant effect. Other conventional risk factors were not significant predictors. The model satisfied goodness-of-fit and proportional odds assumptions but exhibited low explanatory power. These findings highlight the importance of stress management in cardiovascular risk assessment and prevention.

Keywords: Heart attack risk, Ordinal logistic regression, Stress level, Cardiovascular risk factors. Chi-square analysis

Received: March 29, 2025. Revised: November 17, 2025. Accepted: December 14, 2025. Published: February 4, 2026.

1 Introduction

Heart attack remains a major global public health concern and continues to contribute significantly to morbidity and mortality worldwide. While numerous studies have investigated cardiovascular risk factors, inconsistencies remain regarding the relative importance of psychosocial variables such as stress when analyzed alongside traditional biomedical factors. This study extends existing literature by employing Ordinal Logistic Regression (OLR) to model ordered heart attack risk categories and by strengthening the statistical framework through additional diagnostics. In the same vein, according to World Health Organization (WHO) report, Cardiovascular diseases particularly heart attacks or myocardial infractions, remain one of the leading causes of mortality worldwide. Heart disease is one of the leading causes of death worldwide, with heart attacks (myocardial infarctions) being among the most severe manifestations. Heart disease remains the number one cause of death globally, responsible for approximately 17.9 million deaths each year, accounting for 32% of all global deaths [16]. In the United States alone, an estimated 805,000 heart attacks occur annually, with about 605,000 being first-time heart attacks and 200,000 being recurrent cases [1]. According to the Centers for Disease Control and Prevention (CDC), nearly one in five heart attacks is silent, meaning the person affected is unaware they had a cardiac event, increasing the likelihood of subsequent, more severe episodes (CDC, 2022). The economic burden of heart disease is also significant, with treatment costs and productivity losses estimated at $219 billion per year in the U.S. alone (AHA, 2021). [17] highlighted that limited access to preventive healthcare, poor dietary habits, increased tobacco use, and rising cases of obesity and diabetes contribute to the growing burden of heart attacks in these regions. The WHO (2021) reports that over 75% of cardiovascular disease-related deaths occur in LMICs, where healthcare systems often lack the resources for effective prevention and treatment. The integration of machine learning techniques, such as multinomial logistic regression, enables the identification of key risk factors and the prediction of heart attack risk levels.

Research has consistently identified a wide range of factors associated with heart attack risk. These include both controllable and non-controllable risk factors. Controllable factors such as high blood pressure, smoking, high cholesterol, diabetes, poor diet, physical inactivity, and obesity have been extensively studied [8]. Non - controllable factors include age, sex, ethnicity, and family history. The incidence of heart attack increases with age, particularly among men over 45 and women post-menopause [11]. Logistic regression has been widely used due to its interpretability and effectiveness in binary, ordinal and multinomial classification problems. Logistic regression and survival analysis have been extensively used for cardiovascular risk prediction. The Framingham Heart Study, [3] developed a risk prediction score based on age, cholesterol, smoking, and blood pressure levels. Several studies have established a strong correlation between controllable risk factors and heart attack incidence. High blood pressure, also known as hypertension is a well-established risk factor for heart attacks. Hypertension is a primary contributor to cardiovascular diseases as it places excess strain on the heart and arteries, leading to long-term damage [15]. High cholesterol levels, particularly low-density lipoprotein (LDL) cholesterol, have been identified as a major cause of atherosclerosis, where plaque buildup restricts blood flow to the heart. [5]. Smoking is another significant risk factor, as the harmful chemicals in tobacco smoke damage blood vessels and reduce oxygen supply to the heart, thereby increasing the likelihood of a heart attack [2]. Other important controllable risk factors include diabetes, obesity, lack of physical activity, and an unhealthy diet. Studies indicate that individuals with diabetes have an increased risk of heart disease due to high blood sugar levels damaging blood vessels [4]. Similarly, obesity is linked to high cholesterol, high blood pressure, and insulin resistance, all of which contribute to cardiovascular complications [7]. Beyond controllable factors, several non- controllable risk factors also play a role in heart attack susceptibility. Age is a well-established determinant, with research showing that men over the age of 45 and women over 55 have a significantly higher risk of heart attacks [11]. Gender differences further influence risk levels, as men tend to develop cardiovascular diseases earlier than women, though postmenopausal women experience an increased risk due to hormonal changes. [2]. Family history and genetic predisposition are also significant contributors, as individuals with a family history of heart disease have a higher likelihood of developing cardiovascular conditions [6]. With advancements in artificial intelligence and machine learning, researchers have explored more sophisticated methods for heart attack prediction. Studies comparing logistic regression with machine learning techniques such as decision trees, Support Vector Machines (SVM), and neural networks suggest that while deep learning models offer high predictive accuracy, logistic regression remains a preferred choice in medical settings due to its interpretability [10]. Additionally, research has demonstrated that integrating logistic regression with machine learning techniques can improve prediction accuracy while maintaining explainability [14]. Electronic Health Records (EHR) and AI-driven models have also contributed to the automation of heart attack risk prediction, allowing for real-time analysis of patient data [12]. However, despite these technological advancements, multinomial logistic regression continues to be a widely used and trusted method for classifying patients into different heart attack risk categories based on clinical and lifestyle factors. Logistic regression has been widely used due to its interpretability and effectiveness in binary, ordinal and multinomial classification problems. Studies such as [9] have demonstrated that logistic regression models using patient demographics, cholesterol levels, and lifestyle factors can achieve significant accuracy in heart disease prediction. Similarly, research by [13] highlighted that factors like hypertension, smoking, and diabetes significantly contribute to cardiovascular risk and that machine learning models can improve predictive accuracy compared to traditional statistical approaches. However, challenges such as data imbalance and feature selection impact the robustness of these models. This study builds upon existing research by applying ordinal logistic regression to a large dataset to classify individuals into low, moderate, and high-risk categories.

2 Materials and Methods

2.1 Research Design and Data Collection

Secondary data comprising 50,000 observations were obtained from Kaggle.com. The dataset was selected due to its large sample size, accessibility, and prior usage in peer-reviewed health analytics studies. Also, the study adopted a quantitative research design employing an ordinal logistic regression model to explore the factors responsible for heart attack risk. The analysis seeks to identify demographic, lifestyle, and medical indicators that significantly influence an individual's likelihood of experiencing varying levels of heart attack risk.

The data used for this research is secondary data obtained from the publicly available dataset hosted on Kaggle.com. The dataset was preprocessed and analyzed using statistical techniques, with the ordinal logistic regression model serving as the primary inferential tool. The dataset consists of 50,000 records and 20 attributes representing demographic, lifestyle, and medical factors.

Demographic Factors: Age, Gender
Lifestyle Factors: Smoking, Alcohol Consumption, Physical Activity
Medical Indicators: Blood Pressure, Cholesterol Levels, Heart Rate, Diabetes, Stress Levels.

2.2 Variables Description

Table 1: Description of dependent Variable (Y) (Heart Attack Risk Level)

Category	Description	Code
Low	Individuals with minimal risk factors, generally healthy	0
Moderate	Individuals with some risk factors but no immediate threat	1
High	Individuals at severe risk requiring urgent medical intervention	2

Table 2: Description of Independent Variables (X: Predictors of Heart Attack Risk)

Variable	Description	Value Labels
X_1	Age (Continuous)	18–35: Younger adults, 36–55: Middle-aged, 56+: Older adults
X_2	Gender	0 = Female, 1 = Male
X_3	Smoking Status	0 = Non-smoker, 1 = Smoker
X_4	Alcohol Consumption	0 = Non-drinker, 1 = Occasional drinker, 2 = Frequent drinker
X_5	Physical Activity	0 = Low, 1 = Moderate, 2 = High
X_6	Diabetes	0 = No diabetes, 1 = Has diabetes
X_7	Hypertension	0 = No hypertension, 1 = Has hypertension
X_8	Cholesterol Level	Normal (<200 mg/dL), Borderline (200–239 mg/dL), High (≥240 mg/dL)
X_9	Heart Rate	Normal (60–100 bpm), Elevated (101–120 bpm), High (>120 bpm)
X_{10}	Stress Level	0 = Low, 1 = Moderate, 2 = High
X_{11}	Blood Pressure	Normal (<120/80), Elevated (120–129/80), High (≥130/80)

[[[

Categorical variables such as gender status were encoded into numeric values to enable model computation (Male = 1, Female = 0). To eliminate the effect of differing units among continuous variables (e.g., age, cholesterol), z-score normalization was applied:

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

where X′ = standardized value, X = original value, μ = mean, σ = standard deviation.

2.3 Descriptive Statistics

Descriptive analysis was conducted using tables, frequency distributions, and charts to summarize variable characteristics and patterns. Inferential analysis was performed to examine relationships between variables using Chi-square tests and ordinal logistic regression.

The Chi-square test determined the association between categorical predictors and heart attack risk levels.

Chi-square statistic:

$$X^2 = \sum\sum\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \quad (2)$$

Assumptions: Random sampling, normal distribution, independent observations, and cell frequency ≥ 5.

2.4 Ordinal Logistic Regression Model

The ordinal logistic regression model predicts the cumulative probability of a respondent belonging to a given heart attack risk category or below.

$$P(Y \leq j \mid X) = \frac{e^{\theta_j - (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}{1 + e^{\theta_j - (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}} \quad (3)$$

Where:
j represent the intercept for the ordered categories (Low, Moderate, High),
$P(Y \leq j)$ is the cumulative probability of a patient being in categories j or lower,
e is the exponential function,
θ_j is the intercept for category j,
$X_1, X_2, \cdots, X_n$ are the independent variables (predictors),
$\beta_1, \beta_2, \cdots, \beta_n$ are the coefficients that determine the influence of each predictor.

2.5 Log-Odds Formulation

The log-odds of belonging to category j relative to the reference (Low Risk) is expressed as:

$$log\left(\frac{P(Y = j)}{P(Y = Low)}\right) = \beta_{0j} + \beta_{1j}X_1 + \beta_{2j}X_2 + \cdots + \beta_{nj}X_n \quad (4)$$

Where:
β_{0j} is the intercept for category j,
$\beta_1, \beta_2, \cdots, \beta_n$ are regression coefficients,
$X_1, X_2, \cdots, X_n$ are predictor variables.

2.6 Variable Selection Procedure

A stepwise variable selection technique was employed. Variables were retained based on: - Statistical significance at the 5% level, and - Improvement in model fit measured using the Akaike Information Criterion (AIC). This approach enhances reproducibility and model parsimony

2.7 Model Evaluation

Model performance was evaluated using Accuracy, Precision, Recall, and F1-score metrics. Confusion matrices were used to assess misclassification rates, while odds ratios and coefficients interpreted the magnitude and direction of predictors' effects.

3 Analysis of the Data

3.1 Descriptive Analysis of data

Table 3: Descriptive Analysis

	N	Minimum	Maximum	Mean	Std. Deviation
Age	50000	18	89	53.40	20.799
BMI	50000	15.00	40.00	27.5194	7.22518
Cholesterol Level	50000	150.0	300.0	225.036	43.3174
Resting BP	50000	90	179	134.43	25.908
Heart Rate	50000	60	129	94.53	20.187
Max. Heart Rate Achieved	50000	100	199	149.31	28.824
Valid N (listwise)	50000				

Table 3 illustrates the ages of the participants in the study ranged from 18 to 89 years, with a mean age of 53.4 years. This indicates that the sample population is largely composed of middle-aged and older adults, a demographic group typically associated with an increased susceptibility to heart-related diseases. The Body Mass Index (BMI) of participants ranged from 15.00 to 40.00, with an average value of 27.52, which falls within the overweight category. This finding suggests a general tendency toward higher body weight among the participants, highlighting a potential risk factor for cardiovascular complications. The cholesterol levels of participants varied between 150.0 and 300.0 mg/dL, with a mean value of 225.04 mg/dL, placing the average within the borderline high range. This result implies that a substantial portion of the population studied may be at increased risk of heart disease due to elevated cholesterol levels. The resting blood pressure of respondents ranged from 90 to 179 mmHg, with a mean of 134.43 mmHg, which falls within the Stage 1 hypertension range. This finding supports the likelihood of hypertension prevalence within the study population, which is a major contributing factor to cardiovascular risk. The heart rate among participants ranged between 60 and 129 beats per minute (bpm), with a mean of 94.53 bpm. This value is slightly above the normal resting heart rate range of 60–100 bpm, possibly indicating moderate stress levels or cardiovascular strain among some individuals. Finally, the maximum heart rate achieved during exertion varied from 100 to 199 bpm, with an average of 149.31 bpm. This result aligns with the expected physiological response during physical activity, depending on factors such as age and fitness level. Overall, these descriptive findings suggest that several biological and lifestyle indicators within the population point toward a moderate to high risk of cardiovascular disease.

3.2 Prevalence of Heart Attack Risk on Gender

Table 4: The Cross tabulation between Gender and Heart attack risk

			Heart Attack Risk			**Total**
			Low	**Moderate**	**High**	
Gender	Female	Count	12515	7510	5061	25086
		% within Gender	49.90%	29.90%	20.20%	100.00%
		% within Heart Attack Risk	50.00%	50.40%	50.20%	50.20%
	Male	Count	12509	7394	5011	24914
		% within Gender	50.20%	29.70%	20.10%	100.00%
		% within Heart Attack Risk	50.00%	49.60%	49.80%	49.80%
Total		Count	25024	14904	10072	50000
		% within Gender	50.00%	29.80%	20.10%	100.00%
		% within Heart Attack Risk	100.00%	100.00%	100.00%	100.00%

3.3 Prevalence of Heart Attack Risk on Stress Level

Table 5: The Cross tabulation between Stress Level and Heart attack risk

			Stress Level			**Total**
			Low	**Moderate**	**High**	
Heart attack risk	Low	Count	7452	12582	4990	25024
		% within Heart Attack Risk	29.80%	50.30%	19.90%	100.00%
		% within Stress Level	49.30%	50.40%	50.40%	50.00%
	Moderate	Count	4534	7413	2957	14904
		% within Heart attack risk	30.40%	49.70%	19.80%	100.00%
		% within Stress Level	30.00%	29.70%	29.90%	29.80%
	High	Count	3142	4976	1954	10072
		% within Heart Attack Risk Output	31.20%	49.40%	19.40%	100.00%
		% within Stress Level	20.80%	19.90%	19.70%	20.10%
Total		Count	15128	24971	9901	50000
		% within Heart attack risk	30.30%	49.90%	19.80%	100.00%
		% within Stress Level	100.00%	100.00%	100.00%	100.00%

3.4 Prevalence of Heart Attack Risk on Chest pain type

Table 6: The Cross tabulation between Chest pain type and Heart Attack Risk

			Chest Pain type				Total
			Non-angi	Asymptomatic	Typical	Atypical	
Heart attack	Low	Count	6223	6173	6320	6308	25024
		% within Heart attack risk	24.90%	24.70%	25.30%	25.20%	100.00%
		% within Chest Pain type	49.70%	49.90%	50.00%	50.60%	50.00%
	Moderate	Count	3748	3660	3777	3719	14904
		% within Heart attack risk	25.10%	24.60%	25.30%	25.00%	100.00%
		% within Chest Pain type	29.90%	29.60%	29.90%	29.80%	29.80%
	High	Count	2544	2550	2545	2433	10072
		% within Heart attack risk	25.30%	25.30%	25.30%	24.20%	100.00%
		% within Chest Pain type	20.30%	20.60%	20.10%	19.50%	20.10%
Total		Count	12515	12383	12642	12460	50000
		% within Heart Attack Risk	25.00%	24.80%	25.30%	24.90%	100.00%
		% within Chest Pain type	100.00%	100.00%	100.00%	100.00%	100.00%

The cross-tabulation analyses presented in Tables 4, 5, and 6 collectively reveal patterns that suggest limited differentiation in heart attack risk levels across gender, stress level, and chest pain type. In Table 4, the distribution between male and female participants is notably balanced, with approximately 50% of each gender falling within the low-risk category, followed by similar proportions in the moderate- and high-risk groups. This indicates that gender does not serve as a meaningful predictor of heart attack risk in this dataset, a finding supported by the chi-square test result ($p = 0.755$), which confirms the absence of a statistically significant association. Similarly, the cross-tabulation between stress level and heart attack risk (Table 5) shows a consistent trend across all stress categories, where most individuals fall into the low- or moderate-risk groups, and moderate stress emerges as the most common level among participants. Contrary to expectations that higher stress would correlate with greater heart attack risk, the results reveal no strong or direct relationship, suggesting that stress level alone may not significantly influence heart attack risk without considering other interacting factors. Furthermore, Table 6 shows that chest pain type also does not significantly affect heart attack risk levels, as individuals with non-anginal, asymptomatic, typical, and atypical chest pain display similar proportions across the risk categories. This uniform distribution is further supported by the chi-square test result ($p = 0.497$), indicating no statistically significant association between chest pain type and heart attack risk. Overall, these findings suggest that gender, stress level, and chest pain type do not individually have a substantial effect on determining heart attack risk within this dataset.

3.5 Test of Association between Heart Attack Risk and some explanatory variables

3.5.1 Chi-square test of association between Heart Attack Risk and (Gender, Stress Level and Chest Pain type)

Table 7: Chi-square test of association between Heart Attack Risk and (Gender, Stress Level, and Chest Pain type)

Heart Attack Risk and Gender	**Value**	**Df**	**Asymp. Sig.(2-sided)**
Pearson Chi-Square	0.561	2	0.755
Likelihood Ratio	0.561	2	0.755
Linear-by-Linear Association	0.297	1	0.586
Heart Attack Risk and Stress Level			
Pearson Chi-Square	7.305	4	0.121
Likelihood Ratio	7.291	4	0.121
Linear-by-Linear Association	5.573	1	0.018
Heart Attack Risk and Chest Pain type			
Pearson Chi-Square	5.376	6	0.497
Likelihood Ratio	5.389	6	0.495
Linear-by-Linear Association	3.312	1	0.069
N of Valid Cases	50000		

The chi-square test results presented in Table 7 collectively assess the association between gender, stress level, chest pain type, and heart attack risk levels. The analysis for gender ($\chi^2 = 0.561$, df = 2, p = 0.755) indicates no statistically significant relationship, implying that males and females have an equal likelihood of being classified into any of the heart attack risk categories—low, moderate, or high. This finding suggests that gender alone is not a meaningful predictor of heart attack risk within this dataset. Similarly, the test examining the association between stress level and heart attack risk produced a Pearson chi-square statistic of 7.305 (df = 4, p = 0.121), showing no overall significant relationship. However, the linear-by-linear association was statistically significant ($\chi^2 = 5.573$, p = 0.018), revealing a trend that as stress levels increase, the likelihood of falling into a higher heart attack risk category also rises. This trend supports the role of stress as a contributing factor in cardiovascular risk assessment. In contrast, the association between chest pain type and heart attack risk ($\chi^2 = 5.376$, df = 6, p = 0.497) was not statistically significant, and the linear-by-linear trend (p = 0.069) further confirmed the absence of a progressive relationship. Therefore, chest pain type does not appear to significantly differentiate individuals across the heart attack risk categories in this dataset. Overall, these results suggest that while gender and chest pain type show no significant association with heart attack risk, increasing stress levels may have a modest but meaningful influence on the likelihood of elevated risk.

3.6 Ordinal Logistics Regression

Prior to model fitting, a variable selection procedure was performed to identify the most relevant explanatory variables for predicting heart attack risk. The stepwise elimination method was applied, which starts with all candidate variables and iteratively removes the least significant ones based on statistical criteria. Through this process, Stress level was retained as the most importance explanatory variable. This selected variable was then used as the sole predictor in the ordinal logistic regression model to assess its effect on heart attack risk.

.

Table 8: Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	65.704			
Final	58.806	6.899	2	0.032

3.6.1 Test of goodness

Table 9: Table of Goodness of Fit

	Chi-Square	df	Sig.
Pearson	16.097	12	0.187
Deviance	16.131	12	0.185

Table 10: Pseudo R –square

Cox and Snell	0.000
Nagelkerke	0.000
McFadden	0.000

3.6.3 Test of Parallel line

Table 11: Table of test of parallel lines

Model	**-2 Log Likelihood**	**Chi-Square**	**Df**	**Sig.**
Null Hypothesis	58.806			
General	58.414	0.392	2	0.822

3.6.2 Test for the likelihood

The results presented in Tables 8 to 11 collectively evaluate the performance, adequacy, and validity of the model used to explain variations in heart attack risk. As shown in Table 6, the overall model fit was assessed by comparing the -2 Log-Likelihood of the intercept-only model with that of the final model. The Chi-square value of 6.899 with 2 degrees of freedom and a p-value of 0.034 ($p < 0.05$) indicates that the full model provides a statistically significant improvement over the intercept-only model. This suggests that the selected predictor makes a meaningful contribution to explaining variations in heart attack risk among the participants. Furthermore, the goodness-of-fit statistics in Table 9 (Pearson and Deviance $\chi^2 = 0.392$, df = 2, p = 0.822) reveal that the model fits the data well, as there is no significant difference between the observed and predicted values. However, Table 8 shows that the model explains virtually none of the variance in the dependent variable, indicating that the included predictors possess very low explanatory power in predicting heart attack risk. Finally, the test of parallel lines in Table 11 yielded a non-significant result ($\chi^2 = 0.392$, df = 2, p = 0.822), confirming that the proportional odds assumption is met. This implies that the effect of stress level on heart attack risk remains consistent across the different risk levels, thereby validating the suitability of the ordinal logistic regression model for this analysis.

3.6.4 Test of the explanatory variables contribution to Heart attack risk

Continuous variables: Age, BMI, Cholesterol, BP, Heart Rate, Max Heart Rate

Categorical variables: Gender, Chest Pain Types, Diabetes, Hypertension, Smoking, Alcohol, ECG, Physical Activity, etc.

Table 12: Parameters Estimates of the explanatory variables on Heart attack risk

		Estimate	Std. Error	Wald	Df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Heart Attack Risk = 0]	.019	.019	.993	1	0.319	-.019	.057
	[Heart Attack Risk = 1]	1.395	.020	4682.767	1	0.000	1.355	1.435
Location	[Stress Level=0]	.051	.024	4.339	1	.037	.003	.099
	[Stress Level=1]	.004	.022	.028	1	.867	-.040	.048
	[Stress Level=2]	0	.	.	0	.	.	.

3.6.5 Mathematical Model

(i) Proportional odds model:

$$\text{Log}\left(\frac{p(Y\leq k)}{P(Y>k)}\right) = \theta k - 0.051X1 - 0.004X2 \qquad (5)$$

Where:

Y = Heart attack risk level (ordered: Low = 0, Moderate = 1, High = 2)
k = threshold between adjacent outcome categories (e.g., between Low vs. Moderate/High and Low/Moderate vs. High)
X_1=1 if stress level is low, else 0
X_2=1 if stress level is moderate, else 0
The reference category is high stress (when both X_1=0 and X_2=0)

(ii) Ordinal logistics model:

$$P(Y \leq j|X) = \frac{e^{\theta_j-(0.051X_1+0.004X_2)}}{1+e^{\theta_j-(0.051X_1+0.004X_2)}} \qquad (6)$$

Where:

Y = Heart Attack Risk (Recoded: Low, Moderate, High)
X_1 = Indicator for Low Stress (vs. High Stress)
X_2= Indicator for Moderate Stress (vs. High Stress)
X_3= Indicator for High Stress (baseline category)
θ_1= 0.019 (threshold between Low vs. Moderate/High)
θ_2= 1.395 (threshold between Low/Moderate vs. High)

The model results reveal that individuals experiencing low stress levels have significantly greater odds (β = 0.051, p = 0.037) of being classified within a lower heart attack risk category compared to those with high stress levels. In contrast, individuals with moderate stress levels do not differ significantly from those with high stress, indicating that only low stress exhibits a meaningful protective effect against heart attack risk. Furthermore, other predictors including age, body mass index (BMI), cholesterol level, smoking status, hypertension, diabetes, physical activity, and electrocardiogram (ECG) results did not demonstrate statistically significant effects on heart attack risk levels. These findings suggest that, within this dataset, stress level is the only consistent and statistically reliable factor influencing the likelihood of elevated heart attack risk, highlighting its importance in cardiovascular health assessment and risk management.

3.7 Discussion

Table 3 presents the descriptive statistics of the study participants and indicates that ages ranged from 18 to 89 years, with a mean age of 53.4 years. This suggests that the sample is largely composed of middle-aged and older adults, a population group commonly associated with increased vulnerability to cardiovascular diseases. The Body Mass Index (BMI) ranged from 15.00 to 40.00, with an average value of 27.52, classifying the participants, on average, as overweight. This reflects a general tendency toward excess body weight, which is a known risk factor for cardiovascular complications. Cholesterol levels varied between 150.0 and 300.0 mg/dL, with a mean of 225.04 mg/dL, placing the average within the borderline-high category. This indicates that a considerable proportion of the population may be at elevated risk of heart disease due to dyslipidemia. Resting blood pressure values ranged from 90 to 179 mmHg, with a mean of 134.43 mmHg, corresponding to Stage 1 hypertension and suggesting a notable prevalence of high blood pressure among participants. Heart rate values ranged from 60 to 129 beats per minute (bpm), with an average of 94.53 bpm, which lies slightly above the normal resting range and may reflect cardiovascular strain or elevated stress levels among some individuals. The maximum heart rate achieved during exertion ranged from 100 to 199 bpm, with a mean of 149.31 bpm, which is consistent with expected physiological responses to physical activity. Overall, these descriptive statistics indicate that several biological indicators within the population point toward a moderate to high underlying cardiovascular risk.

The cross-tabulation analyses in Tables 4, 5, and 6 examined the distribution of heart attack risk across gender, stress level, and chest pain type. Table 4 shows a nearly identical distribution of heart attack risk categories among males and females, with approximately half of each gender classified as low risk and similar proportions observed in the moderate- and high-risk groups. This balanced distribution suggests that gender does not meaningfully differentiate heart attack risk in this dataset, a conclusion supported by the chi-square test result (p = 0.755). Similarly, Table 5 indicates that individuals across all stress levels were predominantly classified into the low- and moderate-risk categories, with moderate stress being the most prevalent level.

Although the raw distribution does not show a strong contrast across risk categories, it suggests a potential gradual pattern rather than a sharp difference. Table 6 further reveals that chest pain type does not substantially distinguish heart attack risk levels, as non-anginal, asymptomatic, typical, and atypical chest pain types show comparable proportions across all risk categories. The corresponding chi-square test ($p = 0.497$) confirms the absence of a statistically significant association.

The chi-square test results summarized in Table 7 provide further insight into these relationships. Gender was not significantly associated with heart attack risk ($\chi^2 = 0.561$, df = 2, $p = 0.755$), indicating that males and females have similar probabilities of falling into low, moderate, or high risk categories. The association between stress level and heart attack risk was also not statistically significant at the overall level ($\chi^2 = 7.305$, df = 4, $p = 0.121$). However, the significant linear-by-linear association ($\chi^2 = 5.573$, $p = 0.018$) suggests a monotonic trend whereby increasing stress levels are associated with a higher likelihood of elevated heart attack risk. This indicates that stress may exert a cumulative effect rather than producing sharp categorical differences. In contrast, chest pain type showed no significant overall or linear association with heart attack risk, confirming that it does not progressively influence risk classification within this dataset. Prior to fitting the ordinal logistic regression model, a stepwise variable selection procedure was employed to identify the most relevant predictors of heart attack risk. Through iterative elimination of non-significant variables, stress level emerged as the most important explanatory variable and was therefore retained as the sole predictor in the final model.

The results from Tables 8 to 11 collectively demonstrate that the ordinal logistic regression model is statistically adequate and valid. The likelihood ratio test comparing the intercept-only model to the final model yielded a chi-square value of 6.899 (df = 2, $p = 0.034$), indicating that the inclusion of stress level significantly improved model fit. The goodness-of-fit statistics (Pearson and Deviance tests, $p = 0.822$) suggest no significant discrepancy between observed and predicted values, confirming an acceptable model fit. Additionally, the test of parallel lines was non-significant ($p = 0.822$), indicating that the proportional odds assumption holds and validating the use of the ordinal logistic regression framework. However, the pseudo R-square values were approximately zero, implying that the model explains only a very small proportion of the variation in heart attack risk.

The parameter estimates reveal that individuals with low stress levels have significantly higher odds ($\beta = 0.051$, $p = 0.037$) of being classified into lower heart attack risk categories compared to those experiencing high stress. In contrast, individuals with moderate stress levels did not differ significantly from those with high stress, indicating that only low stress confers a meaningful protective effect. Other conventional cardiovascular risk factors including age, BMI, cholesterol level, smoking status, hypertension, diabetes, physical activity, and ECG results were not statistically significant predictors in the model. These findings suggest that, within this dataset, stress level is the only consistent and statistically reliable factor influencing heart attack risk, underscoring its importance in cardiovascular risk assessment and management.

In conclusion, the findings of this study underscore the importance of stress as a significant determinant of heart attack risk within the analyzed population. However, the model's low explanatory power emphasizes the need for more comprehensive approaches to cardiovascular risk prediction. Future research should incorporate multidimensional frameworks that integrate clinical, behavioral, psychosocial, and biological factors using longitudinal designs. Such an approach would provide a more holistic understanding of the intricate pathways linking stress and other determinants to heart attack risk, thereby improving prevention and management strategies for cardiovascular disease.

4 Summary, General Discussion and Conclusion

4.1 Summary

This study examined factors influencing heart attack risk among individuals using ordinal logistic regression analysis on a dataset of 50,000 respondents. The outcome variable, Heart Attack Risk Level, was categorized into low**,** moderate**,** and high risk. Independent variables included both continuous predictors (e.g., age, BMI, cholesterol level, blood pressure) and categorical factors (e.g., stress level, chest pain type, smoking, diabetes, hypertension, physical activity, and family history of heart disease).

Descriptive statistics revealed that the average age was 53.4 years, with a mean BMI of 27.52 (classified as overweight), and a mean cholesterol level of 225.04 mg/dL (borderline high). Among respondents, 29.8% smoked, 19.9% had diabetes, and 29.7% had hypertension. Chi-square tests of association indicated that stress level showed a significant linear trend with heart attack risk (p = 0.018), while gender showed no significant relationship with heart attack risk (p=0.755). However, In the multivariate ordinal logistic regression model, only low stress level emerged as a statistically significant predictor (p = 0.036), indicating a modest association with reduced heart attack risk. Other conventional predictors—including age, BMI, cholesterol, smoking, diabetes, hypertension, physical activity, and chest pain type—were not statistically significant. Although the model satisfied key assumptions, particularly the proportional odds assumption (Test of Parallel Lines, p = 0.822), and showed good model fit (Goodness-of-fit Pearson and Deviance p = 0.822), its predictive power was very weak, with pseudo R-square values (Cox & Snell, Nagelkerke, McFadden) all near 0.000. The overall model improvement over the intercept-only model was not statistically significant (χ^2 = 23.559, p = 0.487), indicating that the included predictors explained little variance in heart attack risk.

Overall, the findings reinforce the role of psychological stress as a relevant cardiovascular risk factor, while also highlighting the limited utility of traditional clinical variables alone in predicting heart attack risk within this dataset.

4.2 General Discussion

This study examined demographic, lifestyle, and clinical factors associated with heart attack risk using ordinal logistic regression on a large secondary dataset of 50,000 individuals. The analysis provided important insights into the determinants of heart attack risk by categorizing individuals into low, moderate, and high-risk groups. While several established cardiovascular risk indicators were prevalent within the study population, stress level emerged as the only statistically significant predictor in the multivariate model, underscoring the complex and multifactorial nature of cardiovascular disease.

Descriptive findings indicated that the study population was predominantly middle-aged and older, with average BMI and cholesterol levels falling within the overweight and borderline-high ranges, respectively. These characteristics are widely recognized in the literature as important contributors to cardiovascular disease risk. Similarly, the observed mean resting blood pressure corresponded to stage 1 hypertension, suggesting that a substantial proportion of participants had elevated cardiovascular risk profiles. Despite the presence of these conventional risk factors, they did not demonstrate independent statistical significance in the ordinal logistic regression model. This unexpected outcome may reflect the influence of risk factor clustering, potential measurement limitations, or the categorical nature of the outcome variable, which may obscure subtle effects of individual predictors.

Gender did not show a significant association with heart attack risk in this study. This finding contrasts with earlier epidemiological studies that report higher cardiovascular risk among men, particularly at younger ages. However, the balanced distribution of risk across genders observed here may be attributable to the large sample size, broad age range, or the omission of sex-specific biological variables such as hormonal status. This result suggests that gender alone may be insufficient to explain variations in heart attack risk without accounting for interacting clinical and behavioral factors.

Stress level emerged as a key determinant of heart attack risk. Although the overall association between stress and risk category was not significant in the chi-square analysis, the significant linear-by-linear trend and the ordinal logistic regression results indicate that increasing stress levels are associated with a higher likelihood of elevated heart attack risk. Specifically, individuals reporting low stress were significantly more likely to be classified into lower risk categories compared to those with high stress. This finding aligns with existing evidence linking psychosocial stress to adverse cardiovascular outcomes through mechanisms such as chronic inflammation, autonomic nervous system imbalance, hormonal dysregulation, and stress-induced unhealthy behaviors.

Despite the statistical significance of stress level, the ordinal logistic regression model exhibited very low explanatory power, as reflected by near-zero pseudo R-square values. This suggests that heart attack risk cannot be adequately explained by a limited set of predictors and highlights the inherent complexity of cardiovascular disease etiology. Heart attack risk is influenced by dynamic interactions among genetic predisposition, metabolic factors, psychosocial stressors, lifestyle behaviors, and environmental conditions. The weak predictive performance of the model emphasizes the need for more comprehensive approaches that incorporate longitudinal data, biochemical markers, genetic information, and detailed psychosocial assessments.

Overall, the findings of this study reinforce the importance of psychological stress as a modifiable risk factor in cardiovascular health, while also demonstrating the limitations of traditional clinical indicators when analyzed in isolation. The results suggest that effective heart attack risk assessment and prevention strategies should integrate mental health considerations alongside conventional cardiovascular risk monitoring. Future research should adopt multidimensional modeling frameworks and longitudinal designs to better capture the complex pathways through which stress and other factors influence heart attack risk.

4.3 Conclusion

This study applied ordinal logistic regression to examine factors influencing heart attack risk using a large secondary dataset. The results indicate that stress level is the only variable with a statistically significant and consistent association with heart attack risk, with low stress serving as a protective factor. Other commonly cited cardiovascular risk factors did not significantly predict risk levels within the model. Although the fitted model was statistically valid, its low explanatory power underscores the complexity of cardiovascular disease and the limitations of using a narrow set of predictors. The findings suggest that effective heart attack prevention strategies should place greater emphasis on stress management alongside traditional clinical monitoring. Future studies should adopt more comprehensive modeling frameworks that integrate psychosocial, biological, and longitudinal data to enhance predictive accuracy and clinical relevance.

4.3 Recommendations

From this study and the results gotten from the analysis, the following are the recommendations made;

i. Patients should monitor blood pressure, cholesterol, BMI, and blood sugar regularly, even in the absence of symptoms, to facilitate early detection of risk.
ii. Patients should engage in relaxation techniques like meditation, exercise, or therapy to reduce stress, which may affect heart health.
iii. Federal government and health authorizes should formally recognize stress as a controllable cardiovascular risk factor**.** Programs for stress education, workplace mental health, and resilience training should be integrated into public health campaigns.
iv. Patients should report any symptoms like chest discomfort, fatigue, or shortness of breath early even if they seem mild, as early medical intervention can be life-saving.
v. Federal government promote and support the use of health technologies (e.g., mobile heart health apps, wearable monitors) that allow individuals to track not only heart rate or blood pressure, but also psychological well-being and lifestyle behaviors.
vi. Healthcare providers should refer patients to nutritionists, mental health counselors, and fitness experts when necessary.

References

[1]American Heart Association. (2021). *Heart disease and stroke statistics 2021 update.* Circulation, 143(8), e254-e743.

[2] Benjamin, E. J. (2019). "Heart Disease and Stroke Statistics 2019 Update." *Circulation*, 139(10), e56-e528. https://www.sciencedirect.com/science/article/pii/S0002934314008894

[3] D'Agostino Sr, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., and Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*, *117*(6), 743-753. DOI: 10.1161/CIRCULATIONAHA.107.699579

[4] Einarson, T. R. (2018). "Prevalence of Cardiovascular Disease in Type 2 Diabetes." *Diabetes Care*, 41(4), 573-580.

[5] Grundy, S. M.(2019). "2018 AHA/ACC Guidelines on the Management of Blood Cholesterol." *Journal of the American College of Cardiology*, 73(24), e285-e350.

[6] Khera, A. V. (2016). "Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease." *New England Journal of Medicine*, 375(24), 2349-2358. DOI: 10.1056/NEJMoa1605086

[7] Lavie, C. J., Ozemek, C., Carbone, S., Katzmarzyk, Lavie, C. J., Ozemek, C., Carbone, S., Katzmarzyk, P. T., and Blair, S. N. (2019). Sedentary behavior, exercise, and cardiovascular health. Circulation research, 124(5), 799-815. DOI: 10.1161/CIRCRESAHA.118.312669.

[8] Lloyd-Jones, D., Adams, R. J., Brown, T. M., Carnethon, M., Dai, S. and Wylie-Rosett, J. (2010). Heart disease and stroke statistics 2010 update: a report from the American Heart Association. *Circulation*, *121*(7), e46-e215. https://doi.org/10.1161/CIRCULATIONAHA.109.192667

[9] Olisah, C. C., Smith, L., and Smith, M. (2022). Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Computer Methods and Programs in Biomedicine*, *220*, 106773. https://doi.org/10.1016/j.cmpb.2022.106773

[10] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., and Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ digital medicine*, *1*(1), 18. doi:10.1038/s41746-018-0029-1

[11] Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Adams, R. J., Berry, J. D., Brown, T. M and Wylie-Rosett, J. (2011). Heart disease and stroke statistics 2011 update: a report from the American Heart Association. *Circulation*, *123*(4), e18-e209. DOI: 10.1161/CIR.0b013e3182009701

[12] Shickel, B., Tighe, P. J., Bihorac, A., and Rashidi, P. (2017). Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE journal of biomedical and health informatics*, *22*(5), 1589-1604. doi:10.1109/JBHI.2017.2767063.

[13] Tiwari, A., Shah, P. C., Kumar, H., Borse, T., Arun, A. R., Chekragari, M., and Mylavarapu, M. (2025). The Role of Artificial Intelligence in Cardiovascular Disease Risk Prediction: An Updated Review on Current Understanding and Future Research. *Current Cardiology Reviews*, *21*(6), e1573403X351048. https://doi.org/10.2174/011573403X351048250329170744

[14] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., and Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. PloS one, 12(4), e0174944. https://doi.org/10.1371/journal.pone.0174944

[15] Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Collins, K. J., Dennison Himmelfarb, C., and Wright, J. T. (2018). 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Journal of the American College of Cardiology*, *71*(19), e127-e248. https://doi.org/10.1016/j.jacc.2017.11.006

[16] World Health Organization (WHO). (2021). Cardiovascular diseases (CVDs). Retrieved from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[17] Yusuf, S, *et al.,* "Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. "*The Lancet* 395.10226 (2020): 795-808.