International Journal of Applied Sciences: Current and Future Research Trends
(IJASCFRT)
ISSN: 2790-3990
https://ijascfrtjournal.isrra.org/index.php/Applied_Sciences_Journal

The Effects of Age and Some Vital Signs on Prostate-Specific Antigen Concerning Early Diagnosis of Prostate Cancer: A Multinomial Logistic Regression Approach

Chrysogonus Chinagorom Nwaigwe[a], Emmanuel Uchechukwu Oliwe[b], Desmond Chekwube Bartholomew[c*], Ugonma Winnie Dozie[d], Felix Chikereuba Akanno[e]

[a,c,e]*Department of Statistics, Federal University of Technology Owerri, Imo State, Nigeria*
[b]*Department Of Statistics, Federal University of Technology Owerri, Nigeria and Department of Statistics, Imo State Polytechnic Omuma, Nigeria*
[d]*Department of Public Health, Federal University of Technology Owerri, Imo State, Nigeria*
Corresponding author's email: desmond.bartholomew@futo.edu.ng (https://orcid.org/0000-0003-0541-1442)
[a]*Email: chrysogonus.nwaigwe@futo.edu.ng,* [b]*Email: emmaoliwe@gmail.com*
[c]*Email: desmond.bartholomew@futo.edu.ng,* [d]*Email: ugonma.dozie@futo.edu.ng,* [e]*Email: felix.akanno@futo.edu.ng*

Abstract

Prostate cancer, regarded as a health anomaly frequently experienced by males over the age of 45, has gained prominence among cancer disorders experienced by the entire human species in recent years. Discussions about the anomaly's management, treatment, and early diagnosis have also gained attention. There is a paucity of literature on the application of multinomial logistic regression (MLR) to model prostate-specific antigen (PSA) for early diagnosis of prostate cancer through the effects of age and some vital signs associated with fluctuations in PSA. In this study, multinomial logistic regression was applied to model changes in PSA under two different classifications of the PSA levels (four categories and five categories) with age, pulse rate, systolic blood pressure, and diastolic blood pressure as the predictor variables. In each classification, the procedure begins by only grouping the age predictor variable and finding the effects of the predictor variables on the categories of the PSA. The procedure is then repeated with age and pulse rate grouped.

Received: 9/25/2023
Accepted: 10/30/2023
Published: 11/10/2023

---

* Corresponding author.

The MLR for the two classifications were then compared based on prediction accuracy, no information rate, and kappa value. The results show that the model with the first classification of PSA was better than the second classification especially when the pulse rate is also grouped. Age and pulse rate significantly affect prostate-specific antigen (PSA) categories. The 45-55 age group is the most significant, while no-risk individuals have no significant difference in PSA levels. Increased pulse rates may reduce prostate cancer risk in males with PSA levels greater than 50.

Keywords: Risk groups; prostate cancer; pulse rate; age; significant.

1. Introduction

The most frequent type of cancer among Nigerian males is prostate cancer [1]. Prostate cancer is the second most frequent cancer in men, after only non-melanoma skin cancer, according to [2]. They stated that screening for prostate cancer is crucial to enhancing men's longevity and that the major technique of screening for this cancer is prostate-specific antigen (PSA). According to [3], early detection and treatment of prostate cancer improves the individual's chances of survival. The prostate gland produces prostate-specific antigen, which is an important tumor marker in the screening and diagnosis of prostate cancer since high PSA levels are frequently associated with prostate cancer [4]. Since 1994, prostate-specific antigen has been used to detect prostate cancer, and it has revolutionized the capacity to diagnose, treat, and follow up on prostate cancer individuals [5]. PSA screening, according to [5] is associated with a 75% reduction in the proportion of men presenting with metastatic disease and a 32.5% reduction in the age-adjusted prostate cancer mortality rate since 2003. References [5] further stated that, while PSA is not a perfect marker, it does have limited specificity for prostate cancer screening and its proper clinical application is still being debated. References [3] examined the probability distribution of prostate-specific antigens and observed that the burr distribution is an appropriate probability distribution for the description of prostate-specific antigens. They observed that the probability of observing a PSA greater than 4.0 ng/ml in adult males above the age of 45 years was seen to be moderately high (0.68) with a median PSA of 8.30 ng/ml and an interquartile range of 16.40. They obtained age-specific probabilities of having PSA greater than 4.0 ng/ml in adult males and concluded that the age-specific probabilities could be used to identify adult males with PSA greater than 4.0 ng/ml that would be subjected to needle biopsy. The study recommended that men with PSA higher than 4.0 ng/ml who are between the ages of 60 and 80 years should not be subjected to needle biopsies but referred directly for digital rectal examinations. References [6] studied 2779 men with prostate cancer and 1606 men without a cancer diagnosis who were recruited for research in New Zealand, the United States, and Taiwan. Multiple linear regression and univariate modeling were used to examine the relationship of PSA with demographic factors, clinical features (for individuals), and the aldo-ketreductase IC3 (AKRIC3) rs12529 genetic polymorphism. A multivariate analysis of pooled case data revealed that PSA was substantially associated with demographic lifestyle and clinical data, with an interaction between ethnicity and age that further affected the association with PSA level. According to them, separate case and control studies revealed that PSA-related characteristics were unique to each cohort. They went on to say that univariate analyses revealed significant age and PSA correlations among all cases and controls, except US-European cases, and that genetic stratification in cases revealed diversity in correlation.

They proposed that individual PSA cut-off values based on demographic, lifestyle, and genetic factors may be more appropriate for prostate cancer screening. References [7] used logistic regression to investigate the connection of PSA with age, ethnic groupings, digital rectal examination findings (presence or absence of a nodule), baseline PSA values, prostate volume by transrectal ultrasound (TRUS), PSA density (PSAD), and total number of prostate biopsies. The males in their study had initial PSA levels between 4.0 and 10.0 ng/ml and had undergone a 12-core TRUS-guided prostate biopsy between 2009 and 2016. The participants were 617 men from various ethnic backgrounds, including Chinese (63.5%), Malay (23.1%), and Indian (13.3%). They then calculated the prostate cancer detection rate and investigated the relevant risk variables. The study discovered a 14.3% overall cancer detection rate and discovered that prostate cancer diagnosis in Malaysia was considerably lower. In a multi-ethnic Malaysian population, the study recommended that ethnicity and PSA density be considered when proposing a first or repeat TRUS-guided prostate biopsy for prostate cancer detection. It has also been observed that men who are HIV positive or with a family history of cancer may not present elevated PSA levels at PSA screening tests [4]. According to [4] the prevalence of men with raised PSA (> 4 ng/ml) is higher among men in the 60-69 years age group. A series of integrated formulas have been developed by [8] to forecast prostate cancer by applying multinomial logistic regression to PSA, age, weight, and some other factors. According to [8] the developed formulas have higher capacities to detect prostate cancer than using only PSA level and PSA density. References [2] using Poisson regression found that a high level of PSA in elderly men is associated with years of study, race/ethnicity, family arrangement, health perception, systolic blood pressure, diastolic blood pressure, metabolic diseases, alcohol consumption, and sedentary behavior. In the present study, a multinomial logistic regression is used to further examine the influence of age, pulse, systolic blood pressure, and diastolic blood pressure on PSA. Unlike [2], the present study first classified the PSA of the men into four categories and later into five categories by subdividing the fourth category of the first classification for better and distinct inferences on the fourth and fifth categories. This present approach allows data collected on PSA to be treated as categorical data. However, previous studies as in the aforementioned opined that factors associated with PSA level were specific for each cohort, hence, to have a deeper understanding of the contribution of age to variations in PSA, the ages of the individuals were further divided into groups and the significance of each age group in influencing the PSA was examined. Thereafter, the examination was repeated when both the pulse rate and age variables were divided into groups. The remaining parts of this paper are divided into Sections: Section 2 – Material and Methods, Section 3 – Results and Discussions, Section 4 – Summary, and Section 5 – Conclusion, then finally the References.

2. Materials and Methods

The data on the PSA levels of individuals were collected along with their corresponding age, diastolic, systolic, and pulse rates under the assumption that physiological and pathological states affecting blood pressure and pulse rate were kept under control. The PSA level was first classified into four categories according to previous studies. PSA levels $0 \leq PSA \leq 4.0$ were coded as 1 (no risk), PSA levels$4.0 < PSA \leq 10.0$ were coded as 2 (lower risk), PSA levels $10.0 < PSA \leq 20.0$ were coded as 3 (fair risk), and PSA levels greater than 20 were coded as 4 (high risk). The PSA levels were further classified into five categories to enable distinction between high PSA levels and very high PSA levels for better inference on the effects of extreme PSA levels. Thus, the

second classification consists of $0 \leq PSA \leq 4.0$(no risk),$4.0 < PSA \leq 10.0$(low risk), $10.0 < PSA \leq 20.0$(fair risk),$20.0 < PSA \leq 50.0$ (high risk) and$PSA > 50.0$(very high risk). After classifying the PSA levels, the categories were grouped as a factor. The results from the two distinct classifications were compared.

The data have four predictor variables {age, BP1 (diastolic), BP2 (systolic), and pulse rate}, which are all numerical data types used for predicting the categories of the individuals' PSA.

2.1 Multinomial Logistic Regression (MLR)

Logistic regression is a statistical model that employs a function other than the usual least-squares approach to model a binary dependent variable. The function is called the logistic function, and it models the probability of a certain dependent variable, class, or event, such as on or off, pass or fail, win or lose, healthy or sick. Logistic regression has been applied in infant disease modeling by [9], insurance coverage prediction by [10], infant birth weight determinant factors by [11], diabetes by [12], and other areas too. MLR generalizes logistic regression to more than two categories of problems. Multinomial logistic regression is simply a logistic model with more than two possible discrete outcomes for the outcome variable. A multinomial logistic regression model was applied to the data with the categorized PSA as the dependent and age, BP1 (systolic), BP2 (diastolic), and pulse rate as predictor variables.

According to [13] MLR uses a linear predictor function to predict the probability that an observation say l has an outcome k, of the following form:

$$f(k,l) = \beta_{0,k} + \sum_{m=1}^{k} \beta_{1,k} x_{m,i} + \sum_{m=1}^{k} \beta_{2,k} x_{m,i} + \beta_{3,k} x_{3,i} + \beta_{4,k} x_{4,i} + e_i ; \quad i = 1, 2, \ldots, n ; \quad l = 1, 2, \ldots, m; \tag{1}$$

where

$k = Categories\ of\ \mathrm{PSA}$

$f(k,l)$ is a linear prediction function that predicts the likelihood of an observation $x_{m,i}$ that has an outcome as k

$\beta_{0,k}$ is the intercept term of the linear prediction function

$\beta_{1,k}$ is the regression parameter of the age group at the k^{th} outcome.

$\beta_{2,k}$ is the regression parameter of pulse rate at the k^{th} outcome

$\beta_{3,k}$ is the regression parameter of diastolic blood pressure at the k[th] outcome

$\beta_{4,k}$ is the regression parameter of systolic blood pressure at the k[th] outcome

$x_{m,i}$ is the i^{th} observation of the m^{th} level of age and pulse rate variable

e_i is the i^{th} random error component associated with the observation i

In this paper,

$x_{3,i}$ is diastolic blood pressure

$x_{4,i}$ is systolic blood pressure

This paper uses 102 data points and each data point consists of a set of 4 predictor variables and an associated categorical outcome variable Y_i (PSA level). The Y_i was first coded into four categories (no risk, low risk, fair risk, and high risk) and later into five categories by subdividing the high-risk category into high risk and very high risk. In each case, the no-risk was chosen as the reference level in the dependent variable while the age group 66- 75 years and pulse rate group 76 - 85 were chosen as reference levels in the categorical predictor variables because they contain the mean values respectively. To arrive at the MLR model, for K possible outcomes of Y_i, running K-1 independent binary logistic regression models, in which one outcome is chosen as a "reference level" (no risk) and then the other K-1 outcomes are separately regressed against the reference level outcome.

If outcome K (no risk) is chosen as the pivot:

$$Ln\left(\frac{P_r(Y_i=1)}{P_r(Y_i=K)}\right)=\beta_1 X_{m,i} \tag{2}$$

$$Ln\left(\frac{P_r(Y_i=2)}{P_r(Y_i=K)}\right)=\beta_2 X_{m,i}$$

.

.

.

$$Ln\left(\frac{P_r(Y_i=k-1)}{P_r(Y_i=K)}\right)=\beta_{k-1} X_{m,i} \tag{3}$$

Exponentiating both sides and solving for the probabilities in equations (2) and (3), we get:

$$
\begin{aligned}
&P_r\left(Y_i = 1\right) = P_r\left(Y_i = K\right)\exp\left(\beta_1 X_{m,i}\right)\\
&P_r\left(Y_i = 2\right) = P_r\left(Y_i = K\right)\exp\left(\beta_2 X_{m,i}\right)\\
&.\\
&.\\
&.\\
&P_r\left(Y_i = K-1\right) = P_r\left(Y_i = K\right)\exp\left(\beta_{k-1} X_{m,i}\right)
\end{aligned}
\tag{4}
$$

The probability of any K possible childbirth weight can be expressed as:

$$
\begin{aligned}
P_r\left(Y_i = K\right) &= 1 - \sum_{k=1}^{k-1} P_r\left(Y_i = K\right)\\
&= 1 - \sum_{k=1}^{k-1} P_r\left(Y_i = K\right)\exp\left(\beta_k X_{m,i}\right)\\
&= \frac{1}{\sum_{k=1}^{k-1}\exp\left(\beta_k X_{m,i}\right)+1}
\end{aligned}
\tag{5}
$$

Other probabilities can be computed:

$$
P_r\left(Y_i = 1\right) = \frac{\exp\left(\beta_1 X_{m,i}\right)}{\sum_{k=1}^{k-1}\exp\left(\beta_k X_{m,i}\right)+1}
$$

$$
P_r\left(Y_i = 3\right) = \frac{\exp\left(\beta_3 X_{m,i}\right)}{\sum_{k=1}^{k-1}\exp\left(\beta_k X_{m,i}\right)+1}
$$

$$
\begin{aligned}
&.\\
&.\\
&.
\end{aligned}
$$

$$
P_r\left(Y_i = K-1\right) = \frac{\exp\left(\beta_{K-1} X_{m,i}\right)}{\sum_{k=1}^{k-1}\exp\left(\beta_k X_{m,i}\right)+1}
\tag{6}
$$

The unknown parameters in each regression parameter vector are typically jointly estimated by an extension of maximum likelihood using regularization of the weights. The solution is typically found using any of the following: iteratively reweighted least squares (IRLS) [14], gradient-based optimization algorithms such as L-BFGS[12], or specialized coordinate descent algorithms [15]. The iteratively reweighted least squares by [14] were implemented in this paper using the R statistical programming software, version 4.2.2.

3. Results and Discussions

In this Section, the Multinomial Logistic Regression (MLR) model is applied to data collected on prostate-specific antigen (PSA), age, pulse rate, systolic blood pressure, and diastolic blood pressure of some men in the South-Eastern part of Nigeria. The results of the application of the model to the data are discussed. In the first instance, the PSA levels were coded into four distinct levels: no risk, low risk, fair risk, and high risk while in the second phase of the analysis reports, the PSA levels were further classified into no risk, low risk, fair risk, high risk and very high risk as stipulated in Section 2.

Table 1: Summary of PSA Readings of some 102 Men in South Eastern Nigeria.

PSA Group	Code	Frequency	Percentage frequency
No risk	1	28	27.5%
Low risk	2	29	28.4%
Fair risk	3	19	18.6%
High risk	4	16	15.69%
Very high Risk	5	10	9.81%

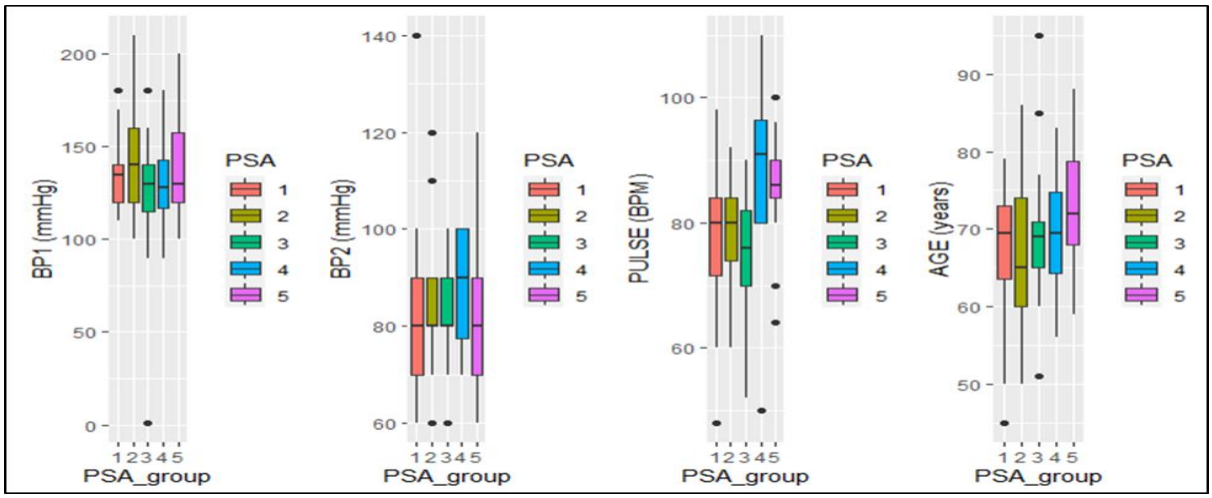

Figure 1: Distribution of PSA categories within each predictor.

The results in Table 1 show that most men who were included in the sample were in the low-risk PSA group (28.4%), followed by the no-risk group (27.5%), and the fair-risk group (18.6%). The very high-risk group recorded the lowest percentage (9.81%). The significant cause(s) of this variability in the PSA levels will be revealed by the multinomial logistic regression models that will be applied later in this paper. The graphical representation of the distribution of the PSA levels into categories is shown in Figure 1. The boxplot in Figure 1 suggests that pulse rate and age will have a more significant effect across the PSA categories than the other two predictor variables (BP1 and BP2). This is because a careful examination of the boxplots for pulse rate and age shows that the mean PSA values across the four groups are different from each other while those of diastolic (BP1) and systolic (BP2) are seemly the same. This evidence of the influence of age in the variation of PSA readings will be further investigated by grouping the age variable into five age groups as follows: 45-55 years, 56-65 years, 66-75 years, 76-85 years, and 86-95 years old. The number of men that fall into each of these

different age groups is given in Table 2. According to the results of Table 2, the majority of the men are between 66 – 75 years old, with only a few in 45 - 55 or 86 – 95 years. The mean age of the individuals is 68 years which falls into the 66–75 age group and will be chosen as the reference level for the age group variable.

Table 2: Frequency distribution for age groups.

Age group	Frequency
45 - 55 years	9
56 - 65 years	28
66 - 75 years	43
76 - 85 years	19
86 - 95 years	3

Further to the descriptive statistics in Table 2, the mean (M) and standard deviations (SD) of BP1 (diastolic), BP2 (systolic), and pulse rates about age groups of the individuals are shown in Tables 3, 4, and 5, respectively.

Table 3: Mean and standard deviations of BP1 (systolic) by age group.

Age group	M	SD
45 - 55 years	126.67	10.00
56 - 65 years	143.43	28.10
66 - 75 years	136.19	21.33
76 - 85 years	130.37	38.91
86 - 95 years	126.67	25.17

Table 3 shows close mean values for the diastolic levels of the individuals in the various age groups. This could be an indication that the mean diastolic level of the individual is not different due to their age groups. There are indications that the individuals are at risk for prehypertension given that the average systolic is between 120 - 139mmHg and the average diastolic is between 80 – 89mmHg (see Table 4) for the age groups.

Table 4: Mean and standard deviations of BP2 (diastolic) by age group.

Age group	M	SD
45 - 55 years	80.00	10.00
56 - 65 years	86.29	16.72
66 - 75 years	83.14	15.20
76 - 85 years	80.53	11.77
86 - 95 years	73.33	11.55

Table 5: Mean and standard deviations of Pulse rate by age group.

Age group	M	SD
45 - 55 years	73.56	7.20
56 - 65 years	83.50	10.19
66 - 75 years	76.26	11.06
76 - 85 years	83.58	8.04
86 - 95 years	82.67	12.06

The mean pulse rate for the entire sample is 79.55 bpm. The average calculated for each age group in Table 5 shows that the older individuals have higher pulse rates than the younger ones. Having described and visualized

the data, we proceed with the MLR modeling in Section 3.1.

3.1 Multinomial Logistic Regression with age groups, pulse rate, diastolic and systolic blood pressure as predictors when PSA is classified as no risk, low risk, fair risk, and high risk

In this Section, the MLR models were fitted using the following first classification of the PSA levels: PSA levels $0 \leq PSA \leq 4.0$ are coded as 1 (no risk), PSA levels$4.0 < PSA \leq 10.0$ are coded as 2 (lower risk), PSA levels $10.0 < PSA \leq 20.0$ are coded as 3 (fair risk), and PSA levels greater than 20 are coded as 4 (high risk), with the pulse rate in their raw scores. The first level of the PSA category (no risk) was selected as the reference level. The age group 66–75 years was chosen as the reference level in the age group variable because it contains the mean age, while No Risk (coded as 1) was selected as the reference level for PSA categories because the focus is to understand how the incidence rate of prostate cancer can be reduced. The predictors are age group, systolic, diastolic blood pressures, and pulse rate while the outcome variable is PSA categories. Further, the dataset was partitioned using sampling without replacement to get the training and test sets. The train set was 70% of the entire sample size of 102 individuals, while the test set is the remaining 30%. The multinomial logistic models were trained with the train set, while the test set was used to validate the classification ability of the models on a new data set. The R program randomly selected the observations for both the training and test sets to reduce class bias and better examine the prediction ability of the model. The results of the MLR model are presented in Table 6.

Table 6: Multinomial Logistic Model for PSA categories using the train set.

Coefficients:

	(Intercept)	**Systolic**	**Diastolic**	**Pulse**	**45 - 55 years**
2	-0.8554232	-0.01533606	0.034120336	-0.004473137	0.3371984
3	2.2522265	-0.01912630	0.027992293	-0.027535877	-0.8344462
4	-8.9735584	-0.00634480	-0.001819418	0.122930996	-14.6837880

	56 - 65 years	**76 - 85 years**	**86 - 95 years**
2	0.7905723	0.2409674	-4.352525
3	-0.7088607	-1.0506887	16.086863
4	-1.0683781	-0.3423434	15.463565

Std. Errors:

	(Intercept)	**Systolic**	**Diastolic**	**Pulse**	**45 - 55 years**
2	2.960166	0.02657013	0.04157765	0.03650831	1.149166e+00
3	3.101567	0.03074986	0.04593973	0.03871353	1.324740e+00
4	4.298337	0.02816607	0.04397812	0.04813258	2.376110e-06

	56 - 65 years	**76 - 85 years**	**86 - 95 years**
2	0.8187713	0.9773914	1.181083e-09
3	0.9877598	1.2598577	1.096168e+00
4	1.0505192	0.9652984	1.096168e+00

Residual Deviance: 159.8151

AIC: 207.8151

Table 6 gives the log odds for the risk levels (categories) of PSA for the predictor variables. The calculation of the p-value begins by getting the z values (that is dividing the regression coefficients by the corresponding standard errors in Table 6), conducting a 2-tailed z test on the new values obtained thereafter, and then subtracting the standard normal table values from 1 to obtain the final results that are the p-values presented in

Table 7.

Table 7: The probability values of the multinomial logistic model.

	(Intercept)	Systolic	Diastolic	Pulse rate	45 - 55 years
2	0.77259812	0.5638099	0.4118506	0.90248420	0.7691945
3	0.46774211	0.5339443	0.5423081	0.47691522	0.5287637
4	0.03682659	0.8217740	0.9670002	0.01064908	0.0000000
	56 - 65 years	**76 - 85 years**	**86 - 95 years**		
2	0.3342647	0.8052632	0.000000		
3	0.4729763	0.4042956	0.000000		
4	0.3091534	0.7228515	0.000000		

The results in Table 7 show that pulse rate is significant in identifying an individual suffering from prostate cancer, most especially when the PSA reading of the individual falls under the high-risk category (Category 4, *p-value* =0.0106 < 0.05). The pulse rate for high-risk prostate cancer individuals when compared with the no-risk category (Category 1) is significant in increasing the odds of prostate cancer by 0.1229 units. The results of the significance analysis of age groups in identifying an individual suffering from prostate cancer using the PSA reading show that the age group 45–55 years is most significant (*p-value* = 0.0000000 < 0.05) when the PSA reading of the individual falls within the high-risk category when compared with the no-risk category. This implies that the PSA reading of an individual whose age falls in the age group 45 - 55 years should be carefully treated especially when it is above the 20- "high risk" category. The results also show that the age group 86–95 years is significant (*p-value*<0.05) in identifying an individual with prostate cancer using the PSA reading in all PSA risk categories when compared to the no-risk category. This means that an individual whose age falls into 86 - 95 years with a PSA reading greater than 4.0 ng/ml is a serious indication of prostate cancer. Other age groups: 56–65 and 76–85 years as well as systolic and diastolic blood pressures were not significant. Next is to calculate the odds ratio (relative risk) for prostate cancer based on the PSA level of the individuals. The relative risk is calculated by exponentiating the log odds (coefficients in Table 6). The relative risk values are shown in Table 8.

Table 8: Relative risk values of the multinomial logistic model.

	(Intercept)	Systolic	Diastolic	Pulse	45 - 55 years
2	0.4251032497	0.9847809	1.0347091	0.9955369	1.401017e+00
3	9.5088833876	0.9810555	1.0283878	0.9728398	4.341148e-01
4	0.0001267165	0.9936753	0.9981822	1.1308064	4.196738e-07
	56 - 65 years	**76 - 85 years**	**86 - 95 years**		
2	2.2046577	1.2724796	1.287426e-02		
3	0.4922046	0.3496968	9.692497e+06		
4	0.3435653	0.7101043	5.196858e+06		

3.1.1 Interpreting the significant age groups coefficient and relative risk values

From Tables 6 to 8, a one-unit increase in age (45–55 years) is associated with a decrease in the log-odds of an individual with no-risk vs. high-risk individual by 14.68 (see Table 6); this coefficient is significant at *p-value*<

0.05 (see Table 7). The relative risk as shown in Table 8 (exp(-14.68) = 4.196738e-07) implies that a one-unit increase in age for individuals in the age group (45–55 years) increases the odds of the individual being at high risk for prostate cancer by (1-4.196738e-07 = 1) times. This means that individuals within the age group (45–55 years) have a 100% chance of being at high risk for prostate cancer at their next birthday. Secondly, a one-unit increase in age (86–95 years) is associated with a decrease in the log-odds of an individual with no risk vs. low risk in the amount of 4.35 (see Table 6); this coefficient is significant at *p-value*< 0.05 (Table 7). The relative risk as shown in Table 8 (exp(-4.35) = 0.0129) implies that a one-unit increase in age for individuals in the age group (86–95 years) increases the odds of the individual being at low risk for prostate cancer by 0.9871 times. This means that individuals within the age group (86–95 years) have only 0.0129 chances of being at no risk for prostate cancer. A one-unit increase in age (86–95 years) is associated with a log-odd increase of 16.09 for an individual with no risk vs. fair risk (see Table 6); this coefficient is significant at *p-value*<0.05 (Table 7). The relative risk as shown in Table 8 (exp(16.09) = 9,692,497.00) implies that a one-unit increase in age for individuals in the age group (86–95 years) increases the odds of the individual being at fair risk for prostate cancer by 9,692,497 times. This means that individuals within the age group (86–95 years) are more likely to be at a fair risk for prostate cancer.Further, a one-unit increase in age (86–95 years) is associated with an increase in the log-odds of an individual with no-risk vs. high-risk by 15.46 units (see Table 6), this coefficient is significant at *p-value*< 0.05 (Table 7). The relative risk as shown in Table 8 (exp(15.46) = 5,196,858.00) implies that a one-unit increase in age for individuals in the age group (86–95 years) increases the odds of the individual being at high risk for prostate cancer by 5,196,858.00 times. This means that individuals within the age group (86–95 years) have a likelihood of having a high risk for prostate cancer.

3.1.2 Interpreting the significant pulse rate coefficient and relative risk values

Considering the pulse rate of the individuals, a one-unit increase in the individual's pulse rate is associated with an increase in the log-odds of an individual being at no risk vs. being at high risk in the amount of 0.123 (Table 6), this coefficient is significant as the *p-value*= 0.000 is less than 0.05 (Table 7). The relative risk (exp(0.12) = 1.13) in Table 8 indicates that a one-unit increase in pulse rate increases the individual's chances of having prostate cancer by (1.13 - 1.00 = 0.13)*100 = 13%.

3.2 Multinomial Logistic Regression with age groups, pulse rate groups, diastolic and systolic blood pressure as predictors when PSA is classified as no risk, low risk, fair risk, and high risk

In this Section, the pulse rate was grouped into six different categories as shown in Table 9. The essence of this grouping is to assess the effect of the pulse rate group on the PSA level of the individual.

The results in Table 9 show that most individuals in the sample for the study had pulse rates within the 70–80 pulse rate group, followed by the 81–91 pulse rate group. The pulse rate group of 103–113 recorded the lowest pulse rate. The mean (M) and standard deviation (SD) of the Pulse rate grouped by age were calculated and presented in Table 10.

Table 9: Frequency Distribution of Pulse Rate.

Pulse Group	Frequency
48 - 58 pulse rate	4
59 - 69 pulse rate	9
70 - 80 pulse rate	42
81 - 91 pulse rate	38
92 - 102 pulse rate	8
103 - 113 pulse rate	1

Table 10: Mean and Standard Deviation of Pulse Rates by Age of the Individuals.

Pulse rate	M	SD
45 - 55 years	73.56	7.20
56 - 65 years	83.50	10.19
66 - 75 years	76.26	11.06
76 - 85 years	83.58	8.04
86 - 95 years	82.67	12.06

Table 11: Multinomial Logistic Regression Model with both Age and Pulse Rate grouped.

Coefficients:

	(Intercept)	**Systolic**	**Diastolic**	**(45 - 55 years)**	**(56 - 65 years)**
2	-1.8273478	-0.008267194	0.037382261	0.4002885	0.7565882
3	1.3986307	-0.013496729	0.006333204	-0.4013890	-0.5694593
4	0.5939208	-0.006205480	-0.004311006	-26.9529606	-0.9330766

	(76 - 85 years)	**(86 - 95 years)**	**(103 - 113 pulse rate)**	**(48 - 58 pulse rate)**
2	0.07175445	-9.190242	-17.822358	-23.9124573
3	-1.22642090	27.605753	-7.890063	0.4571013
4	-0.09395476	14.386004	42.533430	-22.1805626

	(59 - 69 pulse rate)	**(81 - 91 pulse rate)**	**(92 - 102 pulse rate)**
2	-0.7374031	-1.3209258	-0.6350679
3	-1.3058340	-0.1163394	-29.8414159
4	-0.3343435	1.1037969	1.5587123

Std. Errors:

	(Intercept)	**Systolic**	**Diastolic**	**(45 - 55 years)**	**(56 - 65 years)**
2	2.408675	0.02927776	0.04354572	1.197559e+00	0.8534141
3	2.777194	0.03127749	0.04766813	1.393588e+00	1.0180436
4	2.652097	0.03114078	0.04614963	5.132112e-12	1.0942295

	(76 - 85 years)	**(86 - 95 years)**	**(103 - 113 pulse rate)**	**(48 - 58 pulse rate)**
2	1.0085312	5.910423e-11	9.062410e-16	2.554614e-11
3	1.2684610	3.595981e-06	2.913045e-16	1.399521e+00
4	0.9566614	3.447707e-06	4.285014e-18	4.490979e-10

	(59 - 69 pulse rate)	**(81 - 91 pulse rate)**	**(92 - 102 pulse rate)**
2	0.9881211	0.9483316	1.643650e+00
3	1.3003341	0.9617092	1.671374e-07
4	1.4053753	0.9613772	1.573961e+00

Residual Deviance: 153.632
AIC: 225.632

The results show that the age group 76–85 years recorded the highest mean pulse rate of 83.58 with a standard deviation of 8.04. This was followed by the age group of 56–65 years, which had a mean pulse rate of 83.50 with a standard deviation of 10.19. The age group of 45–55 years recorded the lowest mean pulse rate of 73.56 with a standard deviation of 7.20. It can therefore be inferred that those younger men under the age of 55

recorded lower pulse than their older counterparts. In addition to the reference group for age chosen earlier, the pulse rate group 76–85 years was chosen as the reference level in this section of the analysis because the group contained the mean pulse rate. The MLR model in this section was applied to the same train set and the results are presented in Table 11. The results in Table 11 show the coefficients (log odds) of the multinomial logistic regression model and their standard errors when both age and pulse rate were grouped. The coefficients and standard deviations are given for each of the PSA risk categories. To identify the significant coefficients associated with each predictor, *the p-value* of each coefficient was evaluated as illustrated earlier and presented in Table 12. The results in Table 12 show that the age group of 45–55 years is significant (*p-value*<0.05) in the high-risk PSA category when both age and pulse rate were grouped. The age group of 86–95 years is significant for all the PSA risk categories (low risk, Fair risk, and high risk), with the no-risk category as the reference group.

Table 12: Probability values of the Multinomial Logistic Regression Model with Age and Pulse Rate grouped.

	(Intercept)	Systolic	Diastolic	(45 - 55 years)	(56 - 65 years)
2	0.4480603	0.7776589	0.3906385	0.7381882	0.3753250
3	0.6145334	0.6660934	0.8943038	0.7733271	0.5759117
4	0.8228010	0.8420501	0.9255749	0.0000000	0.3938119
	(76 - 85 years)	**(86 - 95 years)**	**(103 - 113 pulse rate)**	**(48 - 58 pulse rate)**	
2	0.9432804	0.00000	0.00000	0.0000000	
3	0.3336153	0.00000	0.00000	0.7439609	
4	0.9217647	0.00000	0.00000	0.0000000	
	(59 - 69 pulse rate)		**(81 - 91 pulse rate)**	**(92 - 102 pulse rate)**	
2	0.4555056		0.1636517	0.6992178	
3	0.3152679		0.9037136	0.0000000	
4	0.8119561		0.2509102	0.3220218	

Table 13: The Relative Risk (odds Ratio) for Prostate Cancer Disease when both Age and Pulse Rate were Grouped.

	(Intercept)	Systolic	Diastolic	age group (45 - 55 years)
2	0.1608396	0.9917669	1.0380898	1.492255e+00
3	4.0496510	0.9865939	1.0063533	6.693896e-01
4	1.8110754	0.9938137	0.9956983	1.970053e-12
	(56 - 65 years)	**(76 - 85 years)**	**(86 - 95 years)**	
2	2.1309934	1.0743915	1.020301e-04	
3	0.5658313	0.2933406	9.750484e+11	
4	0.3933417	0.9103239	1.769139e+06	
	(103 - 113 pulse rate)	**(48 - 58 pulse rate)**		
2	1.819066e-08	4.120517e-11		
3	3.744459e-04	1.579489e+00		
4	2.965064e+18	2.328649e-10		
	(59 - 69 pulse rate)	**(81 - 91 pulse rate)**		
2	0.4783545	0.2668881		
3	0.2709465	0.8901730		
4	0.7158079	3.0155942		
	(92 - 102 pulse rate)			
2	5.298995e-01			
3	1.096574e-13			
4	4.752697e+00			

On the other hand, the pulse rate group of 103–113 is significant (*p-value*<0.05) with all the PSA risk categories

(low risk, fair risk, and high risk), with the no-risk category as the reference group. This means that an individual whose pulse rate is 103-113 and whose PSA reading falls into the low-risk, fair-risk, or high-risk categories has the likelihood of prostate cancer. Similarly, the pulse rate group of 48–58bpm was significant with the low-risk PSA category and the high-risk PSA category only. This means that an individual with a pulse rate between 48 and 58 bpm and a PSA reading of low or high risk should be considered to have prostate cancer.

The pulse rate group of 92–102 is significant (*p-value*<0.05) only with the fair-risk PSA category, implying that a pulse rate under 92–102 with a PSA reading in the fair-risk PSA category is a serious biomarker for prostate cancer. The relative risk (odds ratio) for prostate cancer disease is given in Table 13.

Table 13 shows that the age group 45 – 55 years when compared to 66 – 75 years has a relatively high risk for prostate cancer when compared to the risk PSA category by 1.49 units.

3.3 Multinomial Logistic Regression with age groups, pulse rate, diastolic and systolic blood pressure as predictors when PSA is classified as no risk, low risk, fair risk, high risk, and very high risk

In this section, the MLR models were fitted using the following second classification of the PSA levels: $0 \leq PSA \leq 4.0$ (no risk), $4.0 < PSA \leq 10.0$ (low risk), $10.0 < PSA \leq 20.0$ (fair risk), $20.0 < PSA \leq 50.0$ (high risk) and $PSA > 50.0$ (very high risk), with the pulse rate in their raw scores. The new classification is important so that the new model (train) and performance (test) can be compared with the MLR model for the initial four PSA classifications. Again, the first level of the PSA category (no risk) was selected as the reference level. The results of the MLR model are presented in Table 14.

Table 14: Multinomial Logistic Model for second PSA Classification using the train set.

Coefficients:

	(Intercept)	Systolic	Diastolic	45 - 55 years	56 - 65 years
2	-1.140452	-0.017376248	0.03971769	0.3379234	0.7962202
3	2.139488	-0.019043570	0.02825710	-0.8395729	-0.7270420
4	-20.424025	0.002114708	0.04576104	-13.7052009	-0.6807863
5	-2.125487	-0.003437838	-0.07043212	-16.8076968	-1.8779029

	76 - 85 years	86 - 95 years	PULSE
2	0.2393216	-3.528621	-0.003275782
3	-1.0802089	14.252374	-0.026313658
4	1.1246086	16.387209	0.175617099
5	-1.2606306	-6.042683	0.097736295

Std. Errors:

	(Intercept)	Systolic	Diastolic	45 - 55 years	56 - 65 years
2	3.053467	0.02737020	0.04448917	1.152932e+00	0.8159890
3	3.223214	0.03143583	0.04925825	1.325603e+00	0.9846219
4	8.248979	0.03589792	0.05426117	6.706518e-06	1.4712491
5	5.117276	0.03901431	0.06534695	3.942321e-07	1.5135378

	76 - 85 years	86 - 95 years	PULSE
2	0.9726413	2.241064e-07	0.03692149
3	1.2561603	1.897151e+00	0.03897810
4	1.3501583	1.897142e+00	0.07798192
5	1.3128030	3.652927e-08	0.05629202

Residual Deviance: 169.0011

AIC: 233.0011

In Table 14, the log odds of the various independent variables with their corresponding standard errors were displayed. The AIC value is 233.0011 and will be compared with the AIC value achieved in Table 6. Only log-odds values whose *p-value* is less than 0.05 are concluded to be statistically significant in Table 15.

Table 15: The probability values of the multinomial logistic model.

	(Intercept)	Systolic	Diastolic	45 - 55 years
2	0.70878064	0.5255197	0.3719912	0.7694464
3	0.50683441	0.5446530	0.5662032	0.5265041
4	0.01328837	0.9530247	0.3990339	0.0000000
5	0.67788193	0.9297834	0.2811150	0.0000000

	56 - 65 years	76 - 85 years	86 - 95 years	PULSE
2	0.3291769	0.8056409	0.000000e+00	0.92930213
3	0.4602731	0.3898281	5.795364e-14	0.49961973
4	0.6435603	0.4048753	0.000000e+00	0.02432080
5	0.2147028	0.3369250	0.000000e+00	0.08252189

Table 15 shows that the age group 45 - 55 years is statistically significant in explaining the difference between the PSA levels of no-risk individuals against high-risk and very high-risk individuals. This means that the PSA level of the no-risk individuals that are 45 – 55 years old is statistically significantly different from the PSA level of those of high and very high risk. Another significant age group is the 86 – 95 years which is statistically significant in explaining the difference between the PSA level of no-risk individuals and low, high, and very high-risk individuals. The pulse rate of no-risk individuals is significantly different from the pulse rate of high-risk individuals. However, it can be observed that the pulse rate of no-risk individuals is not significantly different from the pulse rate of very high-risk individuals for the second classification of the PSA levels. This particular result shows that males with PSA levels greater than 50 who do not belong to the age group 45 – 55 years or 86 – 95 years age group face less risk of developing prostate cancer if their pulse rates increase.

Table 16: Relative risk values of the multinomial logistic model.

	(Intercept)	Systolic	Diastolic	45 - 55 years	56 - 65 years
2	3.196744e-01	0.9827738	1.040517	1.402033e+00	2.2171447
3	8.495090e+00	0.9811366	1.028660	4.318949e-01	0.4833366
4	1.348834e-09	1.0021169	1.046824	**1.116624e-06**	0.5062188
5	1.193748e-01	0.9965681	0.931991	**5.017762e-08**	0.1529104

	76 - 85 years	86 - 95 years	PULSE
2	1.2703871	**2.934534e-02**	0.9967296
3	0.3395246	1.547845e+06	0.9740295
4	3.0790115	**1.308803e+07**	**1.1919816**
5	0.2834752	**2.375178e-03**	1.1026720

Tables 14 and 16 show that being in the age group 45 – 55 years reduces the odds of prostate cancer for those at no risk by 13.7052009 units relative to those at high risk. The probability that someone in the 45 – 55 years age group with a PSA level less than or equal to 4.0 will have prostate cancer compared to another individual with a PSA level between 20 and 50 is 0.00000111. This means that the likelihood is rarely possible. Also, for very high-risk individuals in the same age group, the probability is 0.00000005 with a corresponding -16.8076968 log odds value. This means that being in the 45 – 55-year-old age group reduces the odds of prostate cancer by 16.81 units by 99% (1 – 0.00000005). Individuals that are between 86 – 95years of age will have their PSA

level that is between 20 and 50 increased by 16.387209 units (see Table 14) while those whose PSA level is above 50 will be reduced by 6.042683 units (see Table 14). An increase in the pulse rate will decrease the PSA level of the individuals whose initial PSA level was between 30 and 50 by 1.87 (see Table 14).

3.4 Multinomial Logistic Regression with age groups, pulse rate groups, and diastolic and systolic blood pressure as predictors for when PSA is classified as no risk, low risk, fair risk, high risk, and very high risk

In this Section, the pulse rate grouping in Table 9 was used in place of the pulse rate raw scores. The multinomial logistic regression model is presented in Table 17.

Table 17: Multinomial Logistic Model for PSA categories using the train set.

Coefficients:

	(Intercept)	Systolic	Diastolic	45 - 55 years	56 - 65 years
2	-1.898003	-0.009265326	0.039878783	0.3954829	0.7418927
3	1.379640	-0.014383037	0.008042308	-0.4070588	-0.5697559
4	-4.502469	-0.010209564	0.042233575	-26.7466108	-0.1157438
5	3.364008	-0.001396714	-0.048181856	-33.7015241	-1.7655733

	76 - 85 years	86 - 95 years	103 - 113 pulse rate
2	0.07832842	-11.01331	-17.7268179
3	-1.21985257	0.19806	-8.6548562
4	1.26880466	38.78916	46.2441966
5	-1.41983499	-24.58293	-0.5983745

	48 - 58 pulse rate	59 - 69 pulse rate	81 - 91 pulse rate	92 - 102 pulse rate
2	-32.2778507	-0.7531949	-1.2889137	-0.7170943
3	0.4494706	-1.2877551	-0.1336885	-47.3308843
4	-17.2136150	-20.9804830	1.4540416	-19.0072710
5	-36.5676915	-0.5298246	0.6899050	2.1910110

Std. Errors:

	(Intercept)	Systolic	Diastolic	45 - 55 years	56 - 65 years
2	2.451988	0.02962992	0.04538067	1.199827e+00	0.852243
3	2.877047	0.03160610	0.04979414	1.395611e+00	1.018068
4	3.943872	0.03883491	0.05553154	5.156108e-12	1.508394
5	3.544159	0.04256017	0.06780303	3.855373e-15	1.592467

	76 - 85 years	86 - 95 years	103 - 113 pulse rate
2	1.009056	2.070198e-12	1.342688e-14
3	1.267763	1.068892e-05	0.241684e-14
4	1.289463	1.093803e-05	1.847024e-19
5	1.428909	6.157461e-17	8.931096e-21

	48 - 58 pulse rate	59 - 69 pulse rate	81 - 91 pulse rate	92 - 102 pulse rate
2	6.342688e-14	9.885499e-01	0.9421150	1.677597e+00
3	1.401567e+00	1.299023e+00	0.9616349	4.001931e-06
4	4.193128e-08	1.878959e-09	1.3347944	4.265582e-06
5	2.805113e-16	1.566995e+00	1.2011784	1.903082e+00

Residual Deviance: 162.4537
AIC: 258.4537

Table 18: The probability values of the multinomial logistic model.

	(Intercept)	Systolic	Diastolic	45 - 55 years
2	0.4388911	0.7545073	0.3795306	0.7416896
3	0.6315593	0.6490577	0.8716909	0.7705384
4	0.2536051	0.7926303	0.4469360	0.0000000
5	0.3425343	0.9738202	0.4773228	0.0000000

	56 - 65 years	76 - 85 years	86 - 95 years	103 - 113 pulse rate
2	0.3840174	0.9381260	0	0.678012
3	0.5757221	0.3359447	0	0.56112
4	0.9388359	0.3251257	0	0
5	0.2675582	0.3203933	0	0

	48 - 58 pulse rate	59 - 69 pulse rate	81 - 91 pulse rate
2	0.0000000	0.4461084	0.1712787
3	0.7484441	0.3215265	0.8894327
4	0.0000000	0.0000000	0.2760051
5	0.0000000	0.7352766	0.5657264

	92 - 102 pulse rate
2	0.6690492
3	0.0000000
4	0.0000000
5	0.2496105

Table 18 shows that 45 – 58 pulse rate significantly reduced the relative risk for prostate cancer when those with no risk are compared with those with low risk by 32.2778507 units, no risk compared with high-risk individuals by 17.2136150 and no risk compared with very high-risk individuals by 36.5676915. Pulse rate group 59 – 69 influences significantly the PSA level of individuals at high risk but is insignificant on the PSA level of individuals at very high risk while Pulse rate group 92 – 102 influences significantly the PSA level of fair risk and high-risk individuals but also is insignificant on PSA level of individuals at very high risk. Age groups 45 – 55 years and 86 –95 years are also significant factor levels. The relative risk for these significant factor levels is given in Table 19.

Table 19: Relative risk values of the multinomial logistic model.

	(Intercept)	Systolic	Diastolic	45 - 55 years
2	0.1498676	0.9907775	1.0406846	1.485101e+00
3	3.9734700	0.9857199	1.0080747	6.656051e-01
4	0.0110816	0.9898424	1.0431381	2.421556e-12
5	28.9048107	0.9986043	0.9529605	2.310011e-15

	56 - 65 years	76 - 85 years	86 - 95 years	103 - 113 pulse rate
2	2.0999063	1.0814778	1.648080e-05	2.001432e-08
3	0.5656635	0.2952737	6.320338e+2	11.742785e-04
4	0.8907034	3.5565987	7.013239e+16	1.212270e+20
5	0.1710887	0.2417539	2.107500e-11	5.497045e-01

	48 - 58 pulse rate	59 - 69 pulse rate	81 - 91 pulse rate
2	9.591965e-15	4.708598e-01	0.2755700
3	1.567482e+00	2.758894e-01	0.8748625
4	3.343659e-08	7.732002e-10	4.2803794
5	1.314781e-16	5.887082e-01	1.9935262

	92 - 102 pulse rate
2	4.881687e-01
3	2.782647e-21
4	5.562206e-09
5	8.944251e+00

3.5 Model Performance Comparison

Table 20: Comparison of MLR model with age group, pulse rate for four PSA groups and MLR model with age group, pulse rate for five PSA categories.

	Train Set		**Test Set**	
Metric	Model with age group, pulse rate, and four PSA groups	Model with age group, pulse rate, and five PSA groups	Model with age group, pulse rate, and four PSA groups	Model with age group, pulse rate, and five PSA groups
AIC	207.8151	233.0011		
Prediction accuracy	0.4559 (45.59%)	0.5 (50%)	0.4412 (44.12%)	0.3235(32.4%)
No Information Rate	0.2941	0.2941	0.3235	0.3235
P-Value [Acc > NIR]	0.003438	0.0002859	0.1016	0.5642
Kappa	0.2695	0.3469	0.2291	0.1032

Consider Table 20, any model with a smaller AIC value, higher prediction accuracy, significant *p-value* for the Accuracy greater than No information rate, and higher kappa value in the train set is the best model for fitting the train set. The same applies to the test set except that AIC is not available for the test set. Table 20 shows that the model with age group, pulse rate, and five PSA groups is better than the model with age group, pulse rate, and four PSA groups in the train set. However, the model with age group, pulse rate, and four PSA categories outperformed the model with age group, pulse rate, and five PSA categories in a new dataset (test set). Therefore, the model with age group, pulse rate, and four PSA categories is better since it has higher prediction accuracy (44%) and kappa (0.2291) in the test set.

Table 21: Comparison of MLR model with age group, pulse rate group for four PSA categories and MLR model with age group, pulse rate group for five PSA categories.

	Train Set		**Test Set**	
Metric	Model with age group, pulse rate group, and four PSA groups	Model with age group, pulse rate group, and five PSA groups	Model with age group, pulse rate group, and four PSA groups	Model with age group, pulse rate group, and five PSA groups
AIC	225.632	258.4537		
Prediction accuracy	0.5 (50%)	0.4706 (47.06%)	0.2647 (26.47%)	0.1471 (14.71%)
No Information Rate	0.2941	0.2941	0.3235	0.3235
P-Value [Acc > NIR]	0.0002859	0.001585	0.8193	0.9945
Kappa	0.3329	0.3124	0.0047	-0.1079

Table 21 shows that the model with age group, pulse rate group, and four PSA categories is better than the model with age group, pulse rate group, and five PSA categories in the train set (50% > 47.06%). Also, the model with age group, pulse rate group, and four PSA categories outperformed the model with age group, pulse rate group, and five PSA categories in the test set (26.47% > 14.71%). Therefore, the model with age group, pulse rate group, and four PSA groups is better since it has higher prediction accuracy (26.47%) and kappa

(0.0047) in the test set. Nonetheless, the model with five categories of PSA levels showed a distinction in the effect of pulse rate on the high-risk category and very high-risk category of PSA levels. Generally, the model with an age group, pulse rate group, and four PSA categories is better than the model with an age group, ungrouped pulse rate, and four PSA categories. Therefore, it is better to group both the age and pulse rate variables.

4. Summary

From the foregoing, a multinomial logistic regression model was applied to data collected on prostate-specific antigen (PSA), age, pulse rate, systolic blood pressure, and diastolic blood pressure of some men attending a specialist hospital on prostate cancer in the eastern part of Nigeria. The prostate-specific antigen was treated as a categorical dependent variable, while the other variables were treated as predictors, with the initial grouping of age and, subsequently, the grouping of pulse rate. The PSA level was first classified into four categories according to previous studies. PSA levels $0 \leq PSA \leq 4.0$ are coded as 1 (no risk), PSA levels$4.0 < PSA \leq 10.0$ are coded as 2 (lower risk), PSA levels $10.0 < PSA \leq 20.0$ are coded as 3 (fair risk), and PSA levels greater than 20 are coded as 4 (high risk). The PSA levels were further classified into five to enable distinction between high PSA levels and very high PSA levels for better inference on the effects of extreme PSA levels. Thus, the second classification consists of $0 \leq PSA \leq 4.0$ (no risk),$4.0 < PSA \leq 10.0$ (low risk), $10.0 < PSA \leq 20.0$ (fair risk),$20.0 < PSA \leq 50.0$ (high risk) and $PSA > 50.0$ (very high risk). Investigation of the results obtained showed that both age and pulse rate had significant effects on the categories of the prostate-specific antigen (PSA). The results by age group showed that the age group, of 45–55 years is most significant for the PSA high-risk category compared with the no-risk group. The age range of 86-95 years was found to be significant in all PSA risk categories, with the no-risk category serving as the control group. The results on the pulse rate showed that the pulse rate group of 45–55 was significant with only the low- and high-risk PSA categories when compared with the no-risk category, while the pulse rate group of 103–113 was found to be significant with all the PSA risk categories. On the other hand, the pulse rate group of 92–102 was observed to be significant only in the fair risk category. Additionally, the pulse rate of no-risk individuals is not significantly different from the pulse rate of very high-risk individuals for the second classification of the PSA levels. This particular result shows that males with PSA levels greater than 50 who do not belong to the age group 45 – 55 years or 86 – 95 years age group face less risk of developing prostate cancer if their pulse rates increase. Though the model with five categories of PSA levels was not a better model based on the criteria used for the comparison, it however, showed a distinction on the effects of pulse rate on the high-risk category and very high-risk category of PSA levels.

5. Conclusion

In this paper, unlike [2] that used Poisson regression to study PSA levels in elderly men, we studied PSA levels in both elderly and younger men. The results of [2] revealed that a high level of PSA is associated with the systolic blood pressure and diastolic blood pressure of elderly men, but our results did not identify systolic blood pressure and diastolic blood pressure as significant indicators of high PSA levels compared with low PSA

levels for both elderly and younger people. However, in agreement with the results of [4, 6], age group is a significant variable for high levels of PSA. Further, we have also established those different categories of pulse rate are significant to high PSA levels compared to low PSA level individuals.

Hence, we recommend that older men (86–95 years) be checked regularly for early diagnosis of prostate cancer and immediate medical attention to forestall further deaths recorded as a result of this ailment.

Acknowledgment

We thank the St. Jones Clinic, 28 Chukwuma Nwoha Street, Ikenegbu, Owerri, Imo State, Nigeria for providing us with the data for this research.

6. Funding

This research did not receive any kind of funding

7. Conflict of Interests/Competing Interests

The authors declare that there was no conflict of interest.

8. Author Contributions

Conceptualization, Chrysogonus Nwaigwe, and Emmanuel Oliwe; Data curation, Desmond Bartholomew and Ugonma Dozie; Formal analysis, Desmond Bartholomew; Investigation, Emmanuel Oliwe, and Ugonma Dozie; Methodology, Desmond Bartholomew and Chrysogonus Nwaigwe; Project administration, Chrysogonus Nwaigwe; Resources, Emmanuel Oliwe, and Felix Akanno; Software, Desmond Bartholomew; Supervision, Chrysogonus Nwaigwe; Validation, Ugonma Dozie; Writing – original draft, Desmond Bartholomew, Chrysogonus Nwaigwe and Emmanuel Oliwe; Writing – review & editing, Desmond Bartholomew.

9. Data Availability

Yes (https://figshare.com/articles/dataset/data_xlsx/22087256)

References

[1]. E.N. Afogu, I. Sunday-Adeoye, K.C. Ekwedigwe, M.E. Isikhuemen, S.C. Okenwa, S.A. Popoola, M.O.Eliboh, and I.C. Amamilo. *Prostate Specific Antigen Screening among Men in Abakaliki, South East Nigeria*. *Open Jour. of Urology*, 7(2017),79–85. https://doi.org/10.4236/oju.2017.75011

[2]. L.L. Galvão, S. Tribess, T.G. Silva, C.G. Santa Rosa, C. G. Pereira, C. G., R.R. Silva, J.E. Sasaki, J.S. Virtuoso Junior, C.A.B. de Lira, and D.A.T. Santos. (2020). *Prevalence and Factors Associated with High Concentration of Prostate-Specific Antigen: ELSIA Study*.9(2020). Biology. https://doi.org/10.3390/biology9100329

[3]. C.M. Bishop. *Pattern Recognition and Machine Learning*, (2006)

https://doi.org/10.1007/b9479810.1007/978-0-387-45528-0

[4]. E.N. Afogu, I. Sunday-Adeoye, K.C. Ekwedigwe, M.E. Isikhuemen, S.C. Okenwa, S.A. Popoola, M.O. Eliboh, and I.C. Amamilo. *Prostate Specific Antigen Screening among Men in Abakaliki, South East Nigeria*. Open Jour. of Urology, 7(2017),79–85. https://doi.org/10.4236/oju.2017.75011

[5]. L. Nogueira, *Prostate-specific antigen for prostate cancer detection*, Int. Brazil Jour. of Urology, 35(2009), 521-531.

[6]. N. Karunasinghe,T.Z. Minas, B.Y. Bao, A. Lee, A. Wang, S. Zhu, J. Masters, M. Goudie, S.P. Huang, F.J. Jenkins, and L.R. Ferguson, L. R. *Assessment of factors associated with PSA level in prostate cancer cases and controls from three geographical regions*, Scie. Reports. 12 (2022). https://doi.org/10.1038/s41598-021-04116-8

[7]. R.S.L. Yii, J. Lim, S.Sothilingam, W.S. Yeoh, A.N.Fadzli, T.A. Ong, S. Kuppusamy, and A.H. Abdul Razack.*Predictive factors of prostate cancer diagnosis with PSA 4.0–10.0 ng/ml in a multi-ethnic Asian population, Malaysia*, Asian Jour. of Sur., 43(2020), pp. 87–94. https://doi.org/10.1016/j.asjsur.2019.02.014

[8]. H. Wang, S. Tai, L. Zhang, and C. Liang. Integrated formulas to forecast prostate cancer: the parameters of influencing the prostate-specific antigen level as an adjunct to prostate-specific antigen and multi-parametric MRI to predict prostate cancer before biopsy.6(2017), *Translational Cancer Research*,1180–1187. https://doi.org/10.21037/tcr.2017.11.17

[9]. U.C. Orumie and D.C. Bartholomew. *Respiratory Syncytial Virus Infection in Infants: A comparative study using Discriminant, Probit and Logistic Regression Analysis*", Asian Journal of Probability and Statistics, 18(2022), 18-31.

[10]. U.C. Orumie, D.C. Bartholomew, C.P. Obite and L.C. Kiwu. 2021). *Likelihood of Insurance Coverage on Damages Due To Level of Insecurity In Nigeria: Logistic Modeling Approach*, 7(2021), Financial Risk and Management Reviews, pp. 50-59.

[11]. D.C. Bartholomew, E.O. Biu, and D. Enegesele. *Impact of Maternal Education and Age on Weight of Child at Birth: Use of Multinomial Logistic Model*, Asian Journal of Probability and Statistics, 18(2022), 46-59. Article no.AJPAS.88848

[12]. C.C. Nwaigwe, A.C.OnyekaandS.N. Nwanneako. *Comparison of Logit and Probit Models in the Analysis of Severity of Diabetes*", EC Diabetes and Metabolic Research, 4(2020), 1 – 11.

[13]. G.U. Ugwuanyim, C.O. Osuchukwu, D.C. Bartholomew and C.P. Obite. *Medical Choices for a Wealthy Nation – A Multinomial Logistic Regression Model*, Asian Journal of Probability and Statistics, 6(2020),1-12. Article no.AJPAS.54245.

[14]. C.M. Bishop. *Pattern Recognition and Machine Learning*, (2006) https://doi.org/10.1007/b9479810.1007/978-0-387-45528-0

[15]. H.F. Yu and F.L. Huang (2011). *Dual coordinate descent methods for logistic regression and maximum entropy models*, Machine Learning. 85(2011), 41–75. doi:10.1007/s10994-010-5221-8.