International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Variation in the CD4 Counts of HIV/AIDS Patients: A Discriminant Analysis Approach

Ijomah, Maxwell Azubuike and Biu, Oyinebifun Emmanuel

Dept. of Mathematics/Statistics, University of Port Harcourt, Nigeria
Email: zubikeijomahs[at]yahoo.com, biuemmanuel[at]yahoo.com

Abstract: ***A discriminant analysis is conducted in order to estimate a discriminant function to determine the expected CD4 counts of HIV patients at University teaching hospital in Nigeria. The explanatory variables are CD4 counts, blood group and genotype. Statistically significant differences are observed in the group means of the variables of the two groups: HIV-positive & HIV-negative patients. The log determinants are found approximately equal in size for the groups while the Box's M value shows that the assumption of the equal co-variances is met. Consequently, we proceeded to estimate the discriminant function. The estimated function is significant at 1 per cent level of significance and can explain 54.8 per cent of the variations in the group memberships. The structure matrix shows that the variables: CD4 count (0.990), Genotype (-0.205) are very important and the Blood group (0.034) is the least important determinants of the expected status of the patients. Finally, the prediction matrix of the holdout sample shows that 84 per cent of the cases are classified correctly.***

Keywords: Discriminant Analysis, linear discriminant function, Jackknife classification, confusion matrix

1. Introduction

CD4+ T-lymphocytes, also known as the helper T-cells, are the coordinators of the immune response which protects the body against microbial disease and some forms of cancer [1]. The measurement of CD4+ T-cell (CD4) counts is a strong predictor of progression to Human Immunodeficiency Syndrome (AIDS), as well as a means of monitoring antiretroviral therapy (ART) [2-5]. It is used to determine the immunological stage of HIV infection, for identifying when to start antiretroviral therapy (ART), to identify patients likely to benefit from co-trimoxazole prophylaxis and to recognize those most at risk of developing immune reconstitution syndrome. Measurement of CD4+ T lymphocytes in peripheral blood is therefore a critical laboratory parameter for the evaluation and monitoring of patients with HIV. Little wonder why the destruction of CD4+ T-lymphocytes by HIV is the main cause of the progressive weakening of the immune system in HIV infection, and leads ultimately to the acquired immune deficiency syndrome, AIDS. Low CD4 counts are associated with a greater risk of patients living with HIV developing opportunistic infections, which may then progress to advanced diseases and death [6]. HIV has particular tropism for CD4 cells and destruction of CD4+ cells results in many of the deleterious effects of HIV. In addition, CD4 may be used in evaluating the response to treatment and for recognizing treatment failure. WHO therefore encourages national programmes to increase access to CD4 measurement technologies to improve the quality of HIV prevention, care and treatment programmes. Possible increases or decreases in CD4 counts are directly related to HIV replication. The use of combinations of antiretroviral drugs (ART) generally results in the suppression of virus replication and hence increased levels of CD4. The success or failure in controlling levels in untreated patients or those on antiretroviral therapy may be associated with factors related to treatment adherence, habits, other correlated infections unrelated to HIV, cancer, immunosuppressive drugs (corticosteroids and chemotherapy), as well as socio-economic and psychosocial factors and access to healthcare [7-9].

Recent studies have shown that CD4 cell counts can vary a lot between people which could be as a result of different factors like as exercise, lack of sleep or smoking. However, these factors do not seem to make any difference to how well one's immune system can fight infections. Because of this variation in CD4 counts it is usually advisable for treatment decisions not be made on the basis of a single CD4 value. In addition, factors related to CD4 variations in patients living with HIV are multiple and complex and the rate of decline in CD4 count for an individual patient may be highly variable over time, making it difficult to predict at higher CD4 counts, the likely time course of an individual's progression to advanced disease. Furthermore, some genetic factors have been cited as contributors to HIV susceptibility or resistance, among them blood groups such as ABO/Rh and Pk [10-14]. Blood groups, therefore, appear to have a contribution to public health, at least in the area of infectious disease, which makes it imperative to synthesize available knowledge in an attempt to decipher the extent to which RBC antigens are involved in HIV epidemiology and to unveil avenues for future research. Underlying demographic and genetic factors, current exposure to infectious diseases and behavioural factors have also been associated with variations in CD4 cell counts in HIV-negative populations [15]. Healthy African and Asian populations typically have lower CD4 lymphocyte counts than their western European and Caucasian counterparts [15-17] but data from specific countries are limited. Underlying infectious diseases, such as pneumonia and tuberculosis (TB), have also been associated with decreased CD4 levels. Commercial sex workers (CSW), who are exposed typically to a wide variety of sexually transmitted infections, have somewhat lower lymphocyte counts than females who are not involved in the sex trade [18-19]. In western populations, black race, low body mass index (BMI) and injection drug use have also been associated with lower CD4 lymphocyte counts [20-21] and

women tend to have CD4 levels 1–200 cells/μl higher than men with comparable demographic and behavioural patterns. From Nigeria, some scattered studies have focused on relationship between CD4 counts and genotype, few studies have attempted to determine the variation in CD4 counts and symptoms. To date there are no comprehensive published data on variation in CD4 levels among HIV-positive and HIV-negative Nigerian. Against this backdrop was the study carried out using discriminant analysis to capture this variability. This study is therefore aimed to assess an explanatory model, using the statistical technique of discriminant analysis for the variations of CD4 counts between patients living with HIV using blood groupand genotype. The rest of the paper are as follows: Section 2 provides theoretical framework for discriminant analysis. In Section 3 we describe the material and method for analysis. Section 4 deals with the analysis of the data while concluding remark is in Section5.

2. Discriminant Analysis Theory

Discriminant analysis is a multiple regression technique that seeks to find the best linear weighting of predictor variables to maximize the differences among two or more groups. It tries to derive the linear combination of two or more independent variables that will discriminate best between a priori defined groups. It is a method used in statistics and pattern recognition to find a combination of features which characterize or separate two or more classes of objects or events. The resulting combination may be used as a classifier to assign objects to previously defined classes. This is achieved by the statistical decision rule of maximizing the between-group variance relative to the within group variance. This relationship is expressed as the ratio of between-group to within group variance. Variables that contribute most to the prediction of group membership in relation to other variables are given the highest weights. This permits the maximum prediction of group membership. In this analysis, we are only concerned with the case of two groups, η_1 and η_2, where $\eta_1 = MN(\mu_1 \sum_1)$ and $\eta_2 = MN(\mu_2 \sum_2)$ are two multivariate normal populations. We distinguish between our two groups based upon the values of our random variables $X^T = [X_1, X_2, ...X_p]$, where each group's values for each variable differ to some degree. Each group has a population consisting of the values of its variables defined by a probability density function $f_1(x)$. The above mentioned guidelines are developed via a training sample. Two regions are formed, R_1 and R_2. The training sample splits the majority of the original sample into two known or correctly classified (by characteristics) regions and then each region R_1 and R_2 is associated with the group, η_1 and η_2 respectively. The remaining sample, n minus the size of training sample, is called the test sample. This is used to test the validity of the classification rule formed by the training sample. Measurements of all objects of one class k are characterized by a probability density function $f_k(x)$ which is seldom known. And there might be some prior knowledge about the probability of observing a member of class k, the prior probability η_k with $\sum_{i=1}^{k} \eta_i = 1$ to estimate $f_k(X)$ and η_k, a training sample is used. Most often applied classification rules are based on the multivariate normal distribution.

$$f_k(X) = f(X/k) = \frac{1}{(2\pi)^{p/2}|\sum_k|^{1/2}} e^{-\frac{1}{2}(\bar{X}-\mu_k)^T \sum_k^{-1}(\bar{X}-\mu_k)} \quad (1)$$

where μ_k and $\sum_k$ are the class k population mean vector and covariance matrix. Under such assumption, the probability that one object with given vector $X = (X_1, X_2, ..., X_p)$ to belong to class k can be calculated by the formula below

$$P(k/X) = \frac{f(X/k)\eta(k)}{\sum_{k=1}^{k} \eta(k) f(X/k)} \alpha f(X/k)\eta(k) \quad (2)$$

Taking the logarithm of Equation (2) above, will lead to the discriminant function

$$d_k(X) = (X-\mu_k)^T \sum_k^{-1}(X-\mu_k) + \log|\sum_k| - 2\log\eta_k \quad (3)$$

and the classification rule

$$d_{\bar{k}}(X) = Mind_k(X) \Leftrightarrow MaxP(k/X) \quad (4)$$

The rule described in Equation (3) and Equation (4) is called Quadratic Discriminant Analysis (QDA). When a special case that all *k* class covariance matrices are identical $\sum_k = \sum$, the discriminant function can be simplified to Equation (5) which is called the Linear Discriminant Analysis (LDA).

$$d_k(X) = 2\mu_k^T \sum^{-1} X - \mu_k^T \sum^{-1} \mu_k - 2\log\eta(k) \quad (5)$$

Therefore, our first step in performing Discriminant Analysis is to check to see whether or not our covariance matrices, Σ_1 and Σ_2, from our two group model are equal. We check the equality of our covariance matrices in order to know if we could apply Linear Discriminant Analysis or Quadratic Discriminant Analysis. We check the equality of the covariance matrices by testing the null hypothesis H_0: $\Sigma_1 = \Sigma_2$ against H_1: $\Sigma_1 \neq \Sigma_2$. To test the null hypothesis, we evaluate the pooled unbiased estimate of the common covariance matrix under H_0, which is given by

$$S_p = \frac{1}{n_1 + n_2 - 2}\left\{\sum_{i=1}^{2}(n_i - 1)S_i\right\} \quad (6)$$

Where n_i is the sample size of group i and S_i is the i^{th} sample covariance matrix. After evaluating S_p, we calculate the test statistic for the equality of the covariance matrices, which has a chi-square (χ^2) distribution and is equal to M/c.

$$M = \left\{\sum_{i=1}^{2}(n_i - 1)\right\}\ln(\det(S_p)) - \left\{\sum_{i=1}^{2}(n_i - 1)\ln(\det(S_i))\right\}$$

$$1/c = 1 - \frac{2p^2 + 3p - 1}{6(p+1)}\left[\left(\sum_{i=1}^{2}\frac{1}{n_i - 1}\right) - \frac{1}{n_1 + n_2 - 2}\right]$$

Incorrect classification sometimes does occur in discriminant analysis due to the fact that the characteristics or variables of the two populations may not always be readily distinguishable. Some contributing factors to misclassification are incomplete knowledge of future performance, exhaustion of the object required for faultless information, and the event of information not being readily accessible. The guidelines followed for classification should minimize the frequency of a misclassification occurring. When determining guidelines one must look at factors such as prior probabilities and the cost of misclassification. To minimize the expected cost of misclassification (ECM) one would want the following to hold for each region:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{p_2}{p_1}\right)$$

Where c is the cost that an object is misclassified and p_1 and p_2 are the prior probabilities for η_1 and η_2. The left side of the inequalities is known as the density ratio. Under a multivariate normal population, the rule for assigning an object to either group becomes:

Allocate $\mathbf{x}_0$ to η_1 if

$$(\mu_1 - \mu_2)^T \Sigma^{-1} x_0 - \tfrac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 + \mu_2) \geq \ln\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{p_1}{p_2}\right)\right]$$

Allocate $\mathbf{x}_0$ to π_2 otherwise. Another method used to attain optimal classification would be to minimize the total probability of misclassification (TPM):

$$TPM = \alpha = 2\phi\left(-\tfrac{1}{2}\Delta p\right)$$

Where Φ(z) is the cumulative distribution function of the standard normal and

$$\Delta_p = \sqrt{(\underline{\mu_1} - \underline{\mu_2})^T \Sigma^{-1}(\underline{\mu_1}_1 - \underline{\mu_2})}$$

3. Materials and Methods

Data collection

The data for the study were obtained from the ARV Clinic of the University of Port Harcourt Teaching Hospital as well as from the departments of Chemical Pathology and Hematology. The data consist two groups and three variables: CD4 count, Blood group and Genotype from HIV-infected patients and those not infected. The idea is to discriminate among these variables, the most important variables when taking tests to determine the health status of HIV positive patients as well as comparing their results with those who are not positive. Data mining was then carried out to remove the statistics of those patients who, for one reason or the other or as a result of death, withdrew from the clinic or stopped attending the clinic after the 1st, 2nd or 3rd visits.

Data Analysis

A predictive discriminant analysis was performed using the DISCRIMINANT subprogram [22-23] in SPSS for Windows 21.0 to determine the optimum weighting of the predictors used in Pegnato & Birch's study to distinguish group membership (HIV Negative and HIV positive). The METHOD=Direct option was to specify the criteria by which the independent variables would be included in the discriminant analysis. This procedure enters all the variables into a prediction equation simultaneously. Basically, the two groups Fisher Linear Discriminant Function [24] will be adopted in this study since it will discriminate between the two groups better than any other linear function [25]. Standardized discriminant function coefficients and structure coefficients (the correlation between the discriminant function and the predictor variables) were requested together with a jackknife classification analysis.

4. Results and Discussion

4.1 Group Means

Group means and standard deviations for each variable for HIV-negative (0) and HIV-positive (1) patients are calculated in table 1. Group mean provides an idea about whether the means of the variables differ between the groups. In addition, group means and group standard deviations can be used as characteristics profile for the two groups. The table 1 shows that the group means are different for the variables: CD4 counts, blood group and genotype.

Table 1: Summary Statistics

Status HIV-Negative (0) HIV-positive (1)	
Variables Mean Std. Deviation	Mean Std. Deviation
CD4 Counts 1023 435.14	289.95 229.46
Blood Group 2.373 1.17	2.289 1.099
Genotype 1.284 0.486	1.5301 0.5913

4.2 Tests of Equality of Group Means

The Wilk's lambda and the F ratio are used to test the equality of means of the groups for the same variable. The Wilk's lambda for each predictor is equal to the ratio of the within group sum of squares to the total sum of squares. It is estimated from one way analysis of variance by considering status variable as dependent variable and the predictor variables as independent variables. The Wilk's lambda is also known as U statistic. The range of Wilk's lambda value is 0 to 1. If a variable's Wilk's is less 0.95, it is revealed that the group means are significantly different. The larger the value, the smaller significance and the smaller the value, the larger significance is ensured. The Wilk's lambda and the transformation of its value to F is done as under and presented in the table 2.

Table 2: Test of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
CD4 count	0.458	175.3701	1	1998	0.000
Blood group	0.999	0.205	1	1998	0.651
Genotype	0.952	7.530	1	1998	0.007

The above table 2 'Test of Equality of Group Means' result shows that CD4 count and Genotype are significantly different for the two groups (HIV-positive and HIV-negative) while for blood group, there is no significant difference.

4.3Tests of the Total Variation

The canonical correlation coefficient, measuring the relation between the discriminant factorial coordinates and the grouping variable, shows that 54.8%, that is $(0.740)^2$, of the total variance accounts for the differences among the two groups through the first discriminant function. (See Table 3).

Table3: Canonical correlation

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.209	100	100	0.740

4.4 The Wilks' Lambda test

Wilks' lambda is a measure of how well each function separates cases into groups. Smaller values of Wilks' lambda indicate greater discriminatory ability of the function. The associated chi-square statistic tests the hypothesis that the means of the functions listed are equal across groups. The small significance value indicates that the discriminant function does better than chance at separating the groups.

Table 4: Wilks' Lambda test

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0.453	116.136	3	.000

Table 4 above shows that the discriminant function is statistically significant since p-value = 0.001 < 0.05, with Wilk's lambda being 0.453 which is closer to 0 than it is to 1. This indicates that the two groups; "HIV -positive" and "HIV-negative" patients seem to differentiate quite well.

4.5 Estimate the Discriminant Function Coefficients

Test of Equality of Covariance Matrices by Using Box's M: In ANOVA, an assumption is that the variances were equivalent for each group but in DA, the basic assumption is that the variance-co-variance matrices are equivalent. Box's M tests the null hypothesis that the covariance matrices do not differ between groups formed by the dependent. In our study, we conducted this test for making it not to be significant so that the null hypothesis that the groups do not differ can be retained. For this study to hold, the log determinants should be more or less equal, and in our study we got the same, which is a positive signal for the analysis. When tested by Box's M with a view for a non-significant M to show similarity and lack of significant differences. Our result shows that the log determinants appear similar and Box's M is 4.30 with F = 2.554 which is not significant as p > .005. Thus, we accept the null hypothesis which means the covariance matrices do not differ between groups formed by the dependent.

Table 5: Log Determinants

Status	Rank	Log Determinant
HIV negative	3	11.001
HIV positive	3	9.980
Pooled within-groups	3	10.667

Table 6: Test Results

Box's M		4.300
F	Approx.	0.589
	df1	6
	df2	140725.329
	Sig.	.340

This provides information on each of the discriminate functions (equations) produced. The maximum number of discriminant functions produced is the number of groups minus 1. We are only using two groups here, namely 'HIV negative' and 'HIV positive', so only one function is displayed**.** The significance value of 0.340 indicates that the data do not differ significantly from multivariate normal. This means one can proceed with the analysis.

4.6 The standardized canonical discriminant function coefficients table

The interpretation of the discriminant coefficients (or weights) is like that in multiple regressions. The table 6 below provides an index of the importance of each predictor like the standardized regression coefficients (beta's) did in multiple regression [25]. The sign indicates the direction of the relationship (here, only two of the predictors have positive sign, genotype showed a negative sign). CD4 count was the strongest predictor while Blood Group appears to be less significant in predicting HIV/AIDS status. Genotype score showed an inverse relationship.

Table 6: Standardized Canonical Discriminant Function Coefficients

	Function
	1
CD4 Count	.981
Blood Group	.048
Genotype	-.131

4.7 The structure matrix table

The table below provides another way of indicating the relative importance of the predictors and it can be seen below that the same pattern holds. Many researchers use the structure matrix correlations because they are considered more accurate than the Standardized Canonical Discriminant Function Coefficients. The structure matrix table shows the correlations of each variable with each discriminate function. These Pearson coefficients are structure coefficients or discriminant loadings. They serve like factor loadings in factor analysis. By identifying the largest loadings for each discriminate function, the researcher gains insight into how to name each function [26]. In the study, we found out that CD4 count as the function that discriminates between HIV negative and HIV positive. Generally, just like factor loadings, 0.30 or 0.50 is seen as the cut-off between important and less important variables. Blood group is clearly not loaded on the discriminant function, i.e. it is the weakest predictor and suggests that the Blood group is not associated with HIV status but is a function of other unassessed factors.

Table 7: Structure Matrix

	Function
	1
CD4 count	.990
Genotype	-.205
Bloodgroup	.034

4.8The canonical discriminant function coefficient table

These unstandardized coefficients (β) are used to create the discriminant function (equation). It operates just like a regression equation. In this study, we have: $D = -1.555 + 0.003CD4 + 0.042Bloodgroup - 0.240Genotype$. The discriminant function coefficients β or standardized form beta both indicate the partial contribution of each variable to the discriminate function controlling for all other variables in the equation. They can be used to assess each predictor's unique contribution to the discriminate function and therefore provide information on the relative importance of each variable. A closer look at the table 8 reveals that both CD4 count and Blood group has direct relationship with HIV-Status while Genotype has an indirect relationship with HIV- Status. This result also conforms standardized canonical discriminant function coefficients table above.

Table 8: Canonical Discriminant Function Coefficients

	Function
	1
CD4 count	.003
Bloodgroup	.042
Genotype	-.240
(Constant)	-1.555

4.9 Group Centroids table

A further way of interpreting discriminant analysis results is to describe each group in terms of its profile, using the group means of the predictor variables. These group means are called Centroids. These are displayed in the Group Centroids table 9 below. In our study, HIV-negative have a mean of 1.216 while HIV-positive produce a mean of –0.981. Cases with scores close to Centroids are predicted as belonging to that group. That means a respondent whose score tends to 1.216 is a HIV-negative and if the patient's score tends to -0.981 can be segregated as HIV-positive patient.

Table 8: Canonical Discriminant Function Coefficients

HIV Status	Function
	1
Negative	1.216
Positive	-0.981

4.10 Classification table

In the classification table below, the rows are the observed categories of the dependent and the columns are the predicted categories. When prediction is perfect, all cases will lie on the diagonal. The percentage of cases on the diagonal is the percentage of correct classifications. The cross validated set of data is a more honest presentation of the power of the discriminant function than that provided by the original classifications and often produces a poorer outcome. The cross validation is often termed a 'jack-knife' classification, in that it successively classifies all cases but one to develop a discriminant function and then categorizes the case that was left out. This process is repeated with each case left out in turn. This cross validation produces a more reliable function. The classification results reveal that 83.0% of respondents were classified correctly into 'HIV-negative' or 'HIV-positive' groups. This overall predictive accuracy of the discriminant function is called the 'hit ratio'. HIV-negative and HIV-positive were predicted in the same accuracy in the study that is 83% which is on the higher side as it tends to 100%.

Table 4.10a: Classification Function Coefficients

	HIV Status	
	Negative	Positive
CD4 count	.010	.003
Blood group	2.080	1.988
Genotype	4.995	5.523
(Constant)	-11.353	-7.679

Table 4.10b: Prior Probabilities for Groups

HIV Status	Prior	Cases Used in Analysis	
		Unweighted	Weighted
Negative	.500	67	67.000
Positive	.500	83	83.000
Total	1.000	150	150.000

Prior Probabilities are used in classification. The default is using observed group sizes. In your sample to determine the prior probabilities of membership in the groups formed by the dependent, and this is necessary if you have different group sizes. If each group is of the same size, as

an alternative you could specify equal prior probabilities for all groups.

5. Conclusion

A discriminant analysis was conducted to predict whether variation in CD4 counts can be used to determine HIV/AIDS status of patients. Predictor variables were CD4 count, Blood group and Genotype. Significant mean differences were observed for CD4 count and Genotype. The Wilk's Lambda test was significant indicating that the two groups HIV -positive" and "HIV-negative" patients seem to differentiate quite well. The test of equality of group means shows that CD4 count and Genotype are significantly different for the two groups (HIV-positive and HIV-negative) while for blood group, there is no significant difference. Also, the log determinants were quite similar and Box's M also indicated that the assumption of equality of covariance was also accepted. The discriminate function revealed a significant association between groups and all predictors, accounting for 54.8% of between group variability, although closer analysis of the structure matrix revealed only two significant predictors, namely CD4 counts and Genotype while blood group showed a poor predictor. The structure matrix shows that the variables: CD4 count (0.990), Genotype (-0.205) are very important and the Blood group (0.034) is the least important determinants of the expected status of the patients. The classification results reveal that 84.0% of respondents were classified correctly into 'HIV-negative' or 'HIV-positive' groups.

Acknowledgement

We thank the department of ARV Clinic and Chemical Pathology/Hematology of the University of Port Harcourt Teaching Hospital for their assistance during the data collection.

Reference

[1] Peeling RW, Sollis KA, Glover S, Crowe SM, Landay AL, Cheng B, et al. (2015) CD4 Enumeration Technologies: A Systematic Review of Test Performance for Determining Eligibility for Antiretroviral Therapy. PLoS ONE 10(3): e0115019. doi:10.1371/journal.pone.0115019.

[2] Koenig S, Fausi A.(1991) Immunology of human immunodeficiency virus infection. In: Holmes KK, Mardh P, Sparling PA, Lemon SM, editors. Sexually transmitted diseases. 3. New York: McGraw-Hill : 231–49.

[3] Tsegaye A, Wolday D, Otto S, et al. (2003) Immunophenotyping of blood lymphocytes at birth, during childhood, and during adulthood in HIV-1-uninfected Ethiopians. ClinImmunol. 109:338–46.

[4] Giorgi, J. V. (1993) Characterization of T lymphocyte subset alterations by flow cytometry in HIV disease. Ann. New York Acad. Sci. 677, 126–137.

[5] Sterling T.R., Chaisson R.E., Moore R.D. (2001) HIV-1 RNA, CD4 T-lymphocytes, and clinical response to highly active antiretroviral therapy. AIDS. 15: 2251–2257.

[6] Langford SE, Ananworanich J, Cooper DA (2007) Predictors of disease progression in HIV infection: a review. AIDS Res Ther 4:11.

[7] Langford SE, Ananworanich J, Cooper DA (2007) Predictors of disease progression in HIV infection: a review. AIDS Res Ther 4: 11.

[8] Hoffman J., Griensven J.V. andMcKella M., (2010) Role of the CD4 count in HIV management. HIV Ther 4: 27–39.

[9] Hoffman R.M., Jamieson B.D., Bosch R.J., Currier J., Kitchen M.R.C., et al. (2011) Baseline immune phenotypes and CD4+ T lymphocyte responses to antiretroviral therapy in younger versus older HIV-infected individuals. J ClinImmunol 31: 873–881.

[10] Neil S. J., Magre S., McKnight A., and Weiss R.A.,(2003)A-B-O blood groups and HIV-1 infection, Antiviral Therapy, vol. 8, abstract 134.

[11] Sayal S.K., Das A.L., and Nema S.K.,(1996)Study of blood groups in HIV seropositive patients, Indian Journal of Dermatology, Venereology and Leprology, 62: 295–297.

[12] Arendrup M., Hansen J.E.S., Clausen H., Nielsen C., Mathiesen L.R., and Nielsen J.O, (1991) Antibody to histo-blood group A antigen neutralizes HIV produced by lymphocytes from blood group A donors but not from blood group B or O donors," AIDS, 5: 441–444.

[13] Abdulazeez A.A., Alo E.A., and Rebecca S.N., (2008) Carriage rate of human immunodeficiency virus (HIV) infection among different ABO and Rhesus blood groups in Adamawa state, Nigeria," Biomedical Research, 19: 41–44 .

[14] Dirisu J.O., Alli T.O., Adegoke A.O., and Osazuwa F., (2011) A survey of prevalence of serum antibodies to human immunodeficiency deficiency virus (HIV), hepatitis B virus (HBV) and hepatitis C virus (HCV) among blood donors," North American Journal of Medical Sciences, 3:35–38.

[15] Clerici M., Butto S., Lukwiya M., et al.(2000) Immune activation in Africa is environmentally-driven and is associated with upregulation of CCR5. Italian–Ugandan AIDS Project.14:2083–92.

[16] Howard R.R., Fasano C.S., Frey L., Miller C.H., (1996) Reference intervals of CD3, CD4, CD8, CD4/CD8, and absolute CD4 values in Asian and non-Asian. Populations Cytometry, 26:231–2.

[17] Messele T., Abdulkadir M., Fontanet A.L., et al.(1999) Reduced naive and increased activated CD4 and CD8 cells in healthy adult Ethiopians compared with their Dutch counterparts. ClinExp Immunol.115:443–50.

[18] Messele T., Rinke de Wit T.F., Brouwer M., et al.(2001) No difference in *in vitro* susceptibility to HIV type 1 between high-risk HIV-negative Ethiopian commercial sex workers and low-risk control subjects. AIDS Res Hum Retroviruses,17:433–41.

[19] Babu P.G., Pramilabai A., Sripriya G., Damodharan S., John T.J.,(1997) Immunologic profiles of HIV-infected and uninfected commercial sex workers in the Vellore region of Southern India. J Acquir Immune DeficSyndr Hum Retrovirol, 16:357–61.

[20] Aldrich J., Gross R., Adler M., King K., MacGregor R.R., Gluckman S.J.,(2000) The effect of acute severe illness on CD4+ lymphocyte counts in nonimmunocompromised patients. Arch Intern Med, 160:715–16.
[21] Tollerud DJ, Clark JW, Brown LM, et al.(1989), The influence of age, race, and gender on peripheral blood mononuclear-cell subsets in healthy nonsmokers. J ClinImmunol, 9:214–232.
[22] Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. Psychological Bulletin. 95, 156-171.
[23] Klecka, W. R. (1980), Discriminant Analysis, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-019, Beverly Hills, CA.
[24] Anderson, T. W. (1968). An Introduction to Multivariate Statistical Analysis. New York: Wiley.
[25] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179–188.
[26] Burns, R. and Burns, R. (2008), Business Research Methods and Statistics Using SPSS, Sage Publications LTD, available at: http://www.uk.sagepub.com/burns/website%20material/ Chapter%2025%20-20Discriminant%20Analysis.pdf