

Manuscript received May 27, 2023; revised June 20, 2023; accepted June 21, 2023; date of publication July 30, 2023  
Digital Object Identifier (DOI): <https://doi.org/10.35882/jeemi.v5i3.303>

Copyright © 2023 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License ([CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).

How to cite: Leader Lawanga Ontshick, Jean-Christophe Mulangu Sabue, Placide Mbala Kiangebeni, Olivier Tshiani Mbayi, Jean-Michel Nsengi Ntamabyaliro, Jean-jacque Muyembe Tamfumu and Rostin Mabela Makengo Matendo, "Comparison of the Performance of Linear Discriminant Analysis and Binary Logistic Regression Applied to Risk Factors for Mortality in Ebola Virus Disease Patients, vol. 5, no. 3, pp. 134–142, July 2023.

# Comparison of the Performance of Linear Discriminant Analysis and Binary Logistic Regression Applied to Risk Factors for Mortality in Ebola Virus Disease Patients

Leader Lawanga Ontshick<sup>1</sup>, Jean-Christophe Mulangu Sabue<sup>2</sup>, Placide Mbala Kiangebeni<sup>2</sup>, Olivier Tshiani Mbayi<sup>2</sup>, Jean-Michel Nsengi Ntamabyaliro<sup>3</sup>, Jean-jacque Muyembe Tamfumu<sup>4</sup> and Rostin Mabela Makengo Matendo<sup>1</sup>

<sup>1</sup> University of Kinshasa, Faculty of Science and Technology, Department of Mathematics, Statistics and Computer Science, Democratic Republic of Congo

<sup>2</sup> University of Kinshasa, Faculty of Medicine, Department of Clinical Biology, Democratic Republic of Congo

<sup>3</sup> University of Kinshasa, Faculty of Medicine, Department of Basic Science, Democratic Republic of Congo

<sup>4</sup> National Institute for Biomedical Research, Democratic Republic of Congo

Corresponding author: Leader Lawanga Ontshick (e-mail: [leader.ontshick@gmail.com](mailto:leader.ontshick@gmail.com))

**ABSTRACT** The aim of our study was to identify risk factors associated with mortality in patients with Ebola virus disease using binary logistic regression and linear discriminant analysis, to assess the predictive power of these two methods and to compare the performance of the models in terms of coefficients and predictions. Our study was a double-blind randomized controlled trial (observational study) conducted in 2018 during the 10th Ebola outbreak in eastern DRC. The study included 363 patients divided into two treatment arms, including 182 patients treated with MAB114 (Ebanga) and 181 patients treated with REGENERON (REGN-EB3). After an in-depth analysis of the data, the two statistical analysis methods selected the same set of variables (risk factors). For binary logistic regression we obtained: viral load 0.58 (0.5-0.67), creatinine 1.98 (1.58-0.67) and aspartate aminotransferase 0.99 (0.9-1); as for linear discriminant analysis, we have viral load (0.88), creatinine (0.94) and aspartate aminotransferase (0.78). We also find almost the same results when different prediction probabilities are evaluated. Logistic regression predicted a 36.5% mortality rate and linear discriminant analysis predicted a 38.8% mortality rate. With regard to the predictive power of the two models, we used the AUC (area under the curve) score and obtained a score of 0.935 for binary logistic regression and 0.6932 for linear discriminant analysis. According to the evaluation hypothesis, the two methods give the same risk factors (viral load (Ctnp), creatinine and alanine aminotransferase (ALT)) with a prediction probability of around 93%.

**INDEX TERMS** Ebola Virus, Ebola Virus Mortality, Linear Discriminant Analysis, Logistic Regression, Risk Factor.

## I. INTRODUCTION

Various predictive statistical methods have been proposed to estimate risk factors associated with disease mortality. The aim of this study was to identify various risk factors associated with mortality in Ebola virus patients using binomial logistic regression and linear discriminant analysis in a randomized trial. The performance of the models on coefficients and predictions was also compared. When the 10th Ebola epidemic broke out in eastern Democratic

Republic of Congo in 2018. Logistic regression measures the relationship between the occurrence of an event (qualitatively explained variables) and the factors that may affect it (explanatory variables). The selection of explanatory variables to be included in the logistic regression model is based on prior knowledge of the disease and the statistical association between the variables and the events as measured by odds ratios. The regression is based on maximum likelihood estimation. Discriminant analysis is both a

predictive technique (Linear Discriminant Analysis - LDA) and a descriptive technique (Discriminant Factor Analysis - DFA). It is designed to describe and predict the membership of an individual to a group (class). A categorical target variable from the Explanatory/Descriptive Variables collection. Mainly quantitative, but can be qualitative with adjustments. Discriminant analysis is based on least squares estimation, which corresponds to linear estimation [4], [8] [13], [14].

This study aims to evaluate the performance of two predictive data analysis methods: Linear Discriminant Analysis and Binary Logistic Regression. The authors of this study have specifically contributed to:

- Study the performance of Linear Discriminant Analysis applied to risk factors linked to Ebola virus disease mortality;
- Study the performance of Binary Logistic Regression on risk factors associated with Ebola virus disease mortality;
- To compare the performance of Linear Discriminant Analysis and Binary Logistic Regression;
- Show that we can achieve the same results with the two methods of predictive data analysis.

## II. METHODS

### A. STUDY POPULATION

The study was conducted in the provinces of North Kivu, South Kivu and Ituri in the Democratic Republic of Congo (DRC) during the 10<sup>ième</sup> Ebola outbreak in 2018. Our study was an observational study (a double-blind randomized clinical trial). In August 2018, an Ebola hemorrhagic fever (EHF) outbreak began in the provinces of North Kivu, South Kivu and Ituri in the Democratic Republic of Congo. This was the 10th confirmed Ebola outbreak in the country since Ebola was first reported in Zaire in 1976. After the end of the outbreak in West Africa, the World Health Organization (WHO) launched a series of discussions to advance research on Ebola. This included group work focusing on how experimental therapies should be evaluated in the context of the next Ebola outbreak. These discussions led to a consensus that, where possible, the most promising experimental treatments should be evaluated in randomized controlled trials in the event of a new epidemic. This infrastructure facilitated the unity of the international community and the leadership of the Democratic Republic of Congo in developing and conducting the clinical trials. The study compared two molecules: REGENERON and MAB114 (Ebanga) as these molecules have been shown to be effective in the management of patients [4], [13], [14], [15], [16]. The inclusion criteria for the study were a positive Ebola test (positive PCR). The study was jointly approved by the University of Kinshasa ethics committee and the National Institute of Allergy and Infectious Diseases (NIAID) review board, and is overseen by an independent data and safety monitoring committee. Written informed consent was obtained from all patients or their legal guardians and for children in accordance with local standards and requirements [16].

### B. DATA COLLECTION AND ANALYSIS

The data used in this study come from the results of a randomized, double-blind, controlled clinical trial conducted on a sample of 363 patients, including 182 patients with MAB114 (Ebanga) and 181 patients with REGENERON (REGN-EB3). The data were entered into REDCAP before being transferred to IBM SPSS version 21.0 and R version 4.2.1 for data analysis. The normality conditions of the variables were determined using the KOLMOGOROV test. The variables were found to be normally distributed (the data follow the normal distribution). The data were then classified by binary logistic regression and linear discriminant analysis. We then determined the risk factors and prediction probabilities for both methods.

### C. LOGISTIC REGRESSION

Logistic regression, developed in 1944 by Joseph Berkson (American Ph ysicist, Physician and Statistician, born in 1899 and died in 1982), allows the discrimination of a binary or polytomous response variable  $Y$  ( $K \geq 2$  classes) from a matrix of  $P$  explanatory variables  $X = (X^1, \dots, X^P)$ . The data can be a mixture of two different formats: continuous and qualitative. The strength of logistic regression lies in the form of the link function used (the logit or probit) and which allows a sigmoidal form of modelling, including the notion of slope influenced by the frequency of observations, in the form of weights by sector, when we move from one sector to another according to the class described by the response  $Y$  [1], [7].

Let  $Y$  a variable with values in  $\{0, 1\}$  to be explained by  $p$  explanatory variables  $X = (1, X_1, \dots, X_p)'$

The logistic model proposes a modeling of the law of  $Y/X = x$  by a Bernoulli distribution of parameter  $p_\beta(x) = P(Y = 1/X = x)$  as defined in equation (1) such that [1]:

$$\log \frac{p_\beta(x)}{1-p_\beta(x)} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = x' \beta \quad (1)$$

Or  $\text{logit } p_\beta(x) = x' \beta$ ,  $\text{logit}$  denoting the bijective and derivable function of  $]0, 1[$

In  $\mathbb{R} \mapsto \log(p/(1-p))$ . The equality in can be written, as explained in equation (2), such that [1]:

$$p_\beta(x) = P_\beta(Y = 1/X = x) = \frac{\exp(x' \beta)}{1 + \exp(x' \beta)} \quad (2)$$

In a logistic model, we make two (2) choices to define the model:

- The choice of a law for  $Y/X = x$ , here Bernoulli's law
- The choice of the  $P(Y = 1/X = x)$  by  $\text{logit } P_\beta(Y = 1/X = x) = x' \beta$ . The function  $\text{logit}$  is bijective. [1], [7] [10], [18], [26].

The parameters  $\beta^j$   $j = 1, \dots, K - 1$  are estimated by maximum likelihood. For an observation  $(x, y)$  is denoted by  $y_1, \dots, y_K$  a complete disjunctive coding of  $y$  i.e.,  $y_j = 1$  if

$y = j, 0$  otherwise. The likelihood is written [7], as explained in equation (3):

$$L(y, \beta) = p_{\beta}^1(x)^{y_1} \dots p_{\beta}^k(x)^{y_k} \quad (3)$$

The likelihood therefore follows a multinomial distribution  $\mathcal{M}(1, p_1(x), \dots, p_k(x))$ . This is why this model is also called a "polythomous multinomial model". The maximum likelihood estimators are again obtained by cancelling the partial derivatives with respect to the different parameters of the sample likelihood.

As in the dichotomous case, there are no explicit solutions for the estimators and numerical methods are used to calculate them. There is no real novelty compared to the binary case, the algorithm is simply more delicate to write because of the multiplication of the number of parameters.

The **odds ratios** do not generally appear in the software output for the multinomial model: they must therefore be calculated by hand, taking care to take into account the particular coding of the qualitative explanatory variables. We recall that for an individual  $x$ , the odds of an event  $Y = j$  is equal to the ratio  $P(Y = j/X = x)/P(Y \neq j/X = x)$ . In the case of the multinomial model, we define the odds of an event  $j_1$  against an event  $j_2$  by equation (4) [7]:

$$\text{odds}(x, Y = j_1 \text{ vs } Y = j_2) = \frac{P_{\beta}(Y=j_1/X=x)}{P_{\beta}(Y=j_2/X=x)} = \exp((\beta^{j_1} - \beta^{j_2})'x) \quad (4)$$

And for two individuals  $x_i$  and  $x_{i'}$ , then the odds ratio is defined in equation (5) [7]:

$$\begin{aligned} OR(x_i, x_{i'}, Y = Y = j_1 \text{ vs } Y = j_2) &= \frac{P_{\beta}(Y=j_1/X=x_i)/P_{\beta}(Y=j_2/X=x_i)}{P_{\beta}(Y=j_1/X=x_{i'})/P_{\beta}(Y=j_2/X=x_{i'})} \\ &= \exp((\beta^{j_1} - \beta^{j_2})'(x_i - x_{i'})) \end{aligned} \quad (5)$$

Thus, if the two individuals  $x_i$  and  $x_{i'}$  differ by only one unit for the variable  $\ell$  we have :

$$OR(x_i, x_{i'}, Y = Y = j_1 \text{ vs } Y = j_2) = \exp(\beta_{\ell}^{j_1} - \beta_{\ell}^{j_2}). \quad [1], [7],$$

#### D. SPECIFICATION AND INFERENCE TESTING

The various estimation methods presented above lead to asymptotically normal estimators when the number of observations tends to infinity. It is therefore easy to use these various estimators to construct test procedures, some of which will be asymptotically equivalent. We will present here the main test procedures based on the maximum likelihood estimation method which is the most frequently used. The most frequent tests are listed below:

- 1) Wald test
- 2) Likelihood Ratio Test (LRT)
- 3) Test of the score or Lagrange multiplier: LM (Lagrange Multiplier)

It should be remembered that these three tests are asymptotically equivalent, which implies that they can contradict each other on small samples. Moreover, since their distribution is only asymptotically valid, care should be taken when using them on small samples.

It is also known that the LRT test is the most powerful locally and should therefore be preferred a priori. We will only consider here the case of a two-way test on a coefficient or on a set of coefficients. (This is in order to have confidence intervals for the ORs.) [5], [7].

#### E. DISCRIMINANT ANALYSIS

Discriminant Factor Analysis (DFA) or simply Discriminant Analysis is a statistical technique that aims to describe, explain and predict the membership of predefined groups (classes, modalities of the variable to be predicted...) of a set of observations (individuals, examples...) from a series of predictor variables (descriptors, exogenous variables...) it was discovered (Introduced) by Fisher in 1936 There are generally two main approaches to Discriminant Analysis (depending on the objectives):

#### F. DISCRIMINANT ANALYSIS CAN BE A DESCRIPTIVE TECHNIQUE.

This is known as discriminant factor analysis (or *descriptive discriminant analysis*). The objective is to propose a new system of representation, latent variables formed from linear combinations of the predictor variables, which make it possible to discern groups of individuals as far as possible. In this sense, it is close to factor analysis because it allows to propose a graphical representation in a reduced space, more particularly to the principal component analysis calculated on the conditional centers of gravity of the clouds of points with a particular metric. It is also known as canonical discriminant analysis, especially in Anglo-Saxon software.

#### G. DISCRIMINANT ANALYSIS CAN BE PREDICTIVE.

In this case, it is a question of constructing a classification function (assignment rule, etc.) which makes it possible to predict the group to which an individual belongs on the basis of the values taken by the predictive variables. In this sense, this technique is similar to supervised machine learning techniques such as decision trees, neural networks, etc. It is based on a probabilistic framework. The best known is certainly the multinomial distribution hypothesis (normal distribution). Added to the assumption of homoscedasticity, the conditional scatterplots have the same shape, we end up with linear discriminant analysis. It is very attractive in practice because the ranking function is expressed as a linear combination of the predictor variables, easy to analyze and interpret. Discriminant analysis is used in the context of modelling a qualitative variable  $Y$  à  $K$  categories (modalities) called variable to be explained (or endogenous variable or variable to be predicted), from quantitative  $p$  quantitative variables called explanatory variables (or exogenous variables or predictors). Discriminant analysis can be considered as an extension of the regression problem to the case where the variable to be explained is qualitative. We have  $n$  individuals (or observations) described by  $p$  variables and divided into  $K$  classes (groups) given by the qualitative variable  $Y$ . The  $K$  classes are known a priori. The

qualitative variable  $Y$  therefore has  $K$  modalities (it is the modalities of  $Y$  which define the classes).

Note:

$x_i^j$  the value of the  $j^{\text{th}}$  explanatory variable measured on the  $i^{\text{th}}$  individual.

$X_i = (x_i^1, \dots, x_i^p)' \in \mathbb{R}^p$  a line of  $X$  describing  $i^{\text{th}}$  individual.

$X^j = (x_1^j, \dots, x_n^j)' \in \mathbb{R}^n$  a column of  $X$  describing  $j^{\text{th}}$  individual.

$G_k$ : the group of individuals in the sample who have the modality  $k$

$n_k = \text{card}(G_k)$  the number of individuals who have the modality  $k$

The sets  $\{X_i \mid i \in \overline{1, n}\} \subset \mathbb{R}^p$  et  $\{X^j \mid j \in \overline{1, p}\} \subset \mathbb{R}^n$  denote the clouds of individuals and variables respectively.

The variable  $Y$  allows us to define, taking into account the  $K$  modalities, a partition of the set of individuals into  $K$  subsets  $G_1, G_2, \dots, G_K$  with the individual  $i$  belonging to  $G_k$  if it is the  $k^{\text{th}}$  modality of the qualitative variable which is realized. For our case we will present the discriminant analysis as a predictive technique (linear discriminant analysis).

Linear discriminant analysis

In the case where the variance-covariance matrices are identical  $\Sigma_1 = \dots = \Sigma_k = \Sigma$  (homoscedasticity hypothesis or equicovariance hypothesis), the calculations will be simplified. This assumption can be interpreted geometrically in terms of the shape and volume of the point clouds in the representation space: these clouds will have the same shape (and volume).

In this case, the Bayesian assignment rule is written in equation (6) [5],[7]:

$$G_{k^*} = \arg \max_{k=1, K} \left( x' \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k' \Sigma^{-1} \mu_k + \ln(\pi_k) \right) \quad (6)$$

Indeed, by developing the quantity in equation (7) [5],[6]:

$$D_k^2 = (x - \mu_k)' \Sigma^{-1} (x - \mu_k) - 2 \ln(\pi_k) + \ln|\Sigma| \quad (7)$$

There are in equation (8) [5],[6]:

$$D_k^2 = x' \Sigma^{-1} x - 2x' \Sigma^{-1} \mu_k + \mu_k' \Sigma^{-1} \mu_k - 2 \ln(\pi_k) + \ln|\Sigma| \quad (8)$$

So minimizing  $D_k^2(x)$  is equivalent to maximizing in equation (9) [5],[6]:

$$-\frac{1}{2} (-2x' \Sigma^{-1} \mu_k + \mu_k' \Sigma^{-1} \mu_k - 2 \ln(\pi_k)) \quad (9)$$

(For  $x' \Sigma^{-1} x$  and  $\ln|\Sigma|$  does not depend on  $k$ ). The maximum likelihood estimators are in equation (10) [5],[6]:

$$\widehat{\mu}_k = g_k = \frac{1}{n_k} \sum_{i \in G_k} x_i \quad (10)$$

$$\widehat{\Sigma} = W = \frac{1}{n} \sum_{k=1}^K \sum_{i \in G_k} (x_i - \mu_k)(x_i - \mu_k)' = \frac{1}{n} \sum_{k=1}^K n_k V_k$$

Or the unbiased version  $W = \frac{1}{n-K} \sum_{k=1}^K n_k V_k$

This gives the classification rule of linear discriminant analysis: in equation (11) [6],[7]:

$$G_{k^*} = \arg \max_{k=1, K} L_k(x) \quad (11)$$

Where  $L_k(x) = x' W^{-1} g_k - \frac{1}{2} g_k' W^{-1} g_k + \ln(\pi_k)$  is the linear discriminant function of the group  $G_k$  (also called the linear ranking function). Each linear discriminant function defines a score function and a new observation will be assigned to the group with the highest score.

## H. TESTS AND SELECTION OF DISCRIMINANT VARIABLES

### 1) HOMOSCEDASTICITY AND BOX TEST

The hypothesis of equality of the matrices  $\Sigma_k$  can be tested using the Box test. If the hypothesis  $\Sigma_1 = \Sigma_2 = \dots = \Sigma_k$  is true, the quantity: in equation (12) [5],[6]:

$$\left(1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\right) \left[ \left( \sum_k \frac{1}{n_{k-1}} - \frac{1}{n-K} \right) (n-K) \ln \left| \frac{n}{n-K} W \right| - \sum_k (n_k - 1) \ln \left| \frac{n_k}{n_{k-1}} V_k \right| \right] \quad (12)$$

This is followed by approximately one  $\chi^2$  à  $\frac{p(p+1)(K-1)}{2}$  degrees of freedom.

### 2) WILKS' TESTS

Let the following assumptions apply:

Null hypothesis:  $H_0$  = conditional centres of gravity are merged: independence between  $X$  et  $Y$  ( $u_1 = u_2 = \dots = u_k$ ). Alternative hypothesis:  $H_1$  = there is at least one centre of gravity that deviates significantly from the others.

The test statistic is Wilks' lambda, its expression is as follows: in equation (13) [5],[6]:

$$\Lambda = \frac{|W|}{|V|} = \frac{|W|}{|W+B|} = \frac{|1|}{|W^{-1}B+I|} \quad (13)$$

It follows Wilks' law with parameter  $(p, n - k, k - 1)$  at  $H_0$  with  $|W|$  represents the determinant of the within-group variance-covariance matrix and  $|V|$  the determinant of the overall variance-covariance matrix.  $H_0$  is rejected if  $\Lambda$  calculated is less than  $\Lambda$  tabulated. This test can be expressed as a multidimensional generalization of the one-factor analysis of variance (ANOVA), in this case we speak of MANOVA (Multidimensional Analysis of Variance).

It is rare to find the Wilks' law table implemented under the various existing statistical software. Therefore, if  $n$  is large enough, we will use the following Bartlett approximation: in equation (14) [6],[7]:

$$\chi^2 = - \left[ n - \frac{1}{2}(p - K - 1) \right] \ln(\Lambda) \quad (14)$$

Which follows a law of  $\chi^2$  with  $P$  degrees of freedom.

In the case where  $K = 2$  we can use the Rao transformation which follows a Fisher distribution with parameter  $(p, n - p - 1)$  The formula for the test statistic then becomes [5], [6], [21]:

$$F = \frac{1-\Lambda}{\Lambda} \quad (15)$$

death_due_to_evd_x	Mab114 (Ebanga) (182)	REGEN (181)	ALL PATIENTS (363)
Not deceased	103(56.6)	116(64.1)	219(60.3)
Deceased	79(43.4)	65(35.9)	144(39.7)

### III. RESULTS

#### A. DESCRIPTIVE ANALYSES OF THE DATA

Demographic and biochemical characteristics of the study of the 363 subjects, 182 received MAb114 (EBANGA) and 181 received REGN-EB3. Most patients (66.1%) were aged  $\geq 18$  years, 15.2% aged 6-17 years, 9.4% aged  $\leq 5$  years, of which 9.1% were neonates ( $\leq 7$  days) (TABLE 1 and TABLE 2).

**TABLE 1**  
 Social demographic characteristics of the study

FEATURES	Mab114 (Ebanga) (182)	REGEN (181)	ALL PATIENTS (363)
Age	30 $\pm$ 19	30 $\pm$ 19	30.1 $\pm$ 18
Age_cat			
$\leq 5$ yrs	18(9.8)	15(8.3)	33(9.1)
$\geq 7$ days	13(7.1)	21(11.6)	34(9.4)
5yrs to 18yrs	28(15.4)	28(15.5)	56(15.4)
$\geq 18$ yrs	123(67.6)	117(64.6)	240(66.1)
dm_sex_x			
Female	111(51.4%)	105(48.6%)	216(59.5)
Male	71(48.3%)	76(51.7%)	147 (40.5)

**TABLE 2**  
 Biochemical parameters

FEATURES	Mab114 (Ebanga) (182)	REGEN (181)	ALL PATIENTS (363)
ctnp	23.1 $\pm$ 5.7	22.5 $\pm$ 5.0	22.7 $\pm$ 5.6
ev_experimental_vaccine_x			
Not vaccinated	92(50.5)	91(50.3)	183(50.4)
Vaccinated	90(49.5)	90(49.7)	180(49.6)
ch_malaria_result_x			
Negative	122(67.0)	115(63.5)	237(65.3)
Positive	25(13.7)	27(14.9)	59(14.3)
99	35(19.2)	39(21.5)	74(20.4)
ch_glucose_x	108.5 $\pm$ 52.4	110.5 $\pm$ 85.6	109.4 $\pm$ 70.8
ch_sodium_cu_x	124.2 $\pm$ 3.4	119.5 $\pm$ 37.7	121.8 $\pm$ 34.2
ch_ast_sgot_x	408.8 $\pm$ 621.4	469.6 $\pm$ 68.2	439.1 $\pm$ 652.3
ch_alt_sgpt_x	319.9 $\pm$ 38.4	322.5 $\pm$ 443.9	321.2 $\pm$ 415.9

#### B. MORTALITY BY STUDY ARM

Regarding the mortality of the study, overall there were almost 40% deaths distributed according to the study arms: 43.4% in the Mab114 (Ebanga) group and 35.9% in the REGENERON group (TABLE 3).

**TABLE 3**  
 Mortality by study arm

The initial nuclear protein Ct in this study was  $22.7 \pm 5.6$ , with 50.4% unvaccinated patients versus 49.6% vaccinated patients. For malaria alone, 14.3% of patients tested positive. The mean baseline blood glucose was  $109.4 \pm 70.8$  mg/deciliter, the mean aspartate aminotransferase was  $439.1 \pm 652.3$  U/liter and the mean alanine aminotransferase was  $321.2 \pm 415.9$ U/liter. Regarding mortality and various biochemical parameters, the variable ctnp in deceased patients indicated that the deceased patients had high viremia. The Ctnp values were lower than 22.  $19.39 \pm 3.92$  in patients receiving Mab114 (Ebanga) and  $18.97 \pm 3.33$  in patients receiving REGENERON. Patients who died had near-normal blood glucose levels.  $106.29 \pm 63.22$  in patients receiving Mab114 and  $105.72 \pm 93.65$  in patients receiving REGENERON. For creatinine, potassium and sodium, this table shows that different values are normal in two different groups of deceased patients. We note that the functional indices of hepatitis (aspartate aminotransferase (AST) and alanine aminotransferase (ALT)) in the deceased patients were above normal.

#### C. MULTIVARIATE ANALYSIS

##### 1) LOGISTIC REGRESSION

Our dependent variable is death and the different (independent) variables considered for the logistic regression are: age, viral load (ctnp), creatinine, glucose, potassium, sodium, aspartate aminotransferase (AST) and alanine aminotransferase (ALT).

##### 2) COMPOSITE TESTS OF MODEL COEFFICIENTS AND MODEL SUMMARY

The statistical indicator for the overall degree of association of the variables in the model is the chi-square. We can conclude that the 8 independent variables are globally associated with patient death (p-value=0.000). We can assess the quality of our regression by means of coefficients of determination: Cox and Snell R-two and Nagelkerke R-two. Our factors: age, viral load (ctnp), creatinine, glucose, potassium, sodium, ALT, AST influenced between 51 and 70% of the deaths of the patients in our study.

##### 3) EQUATION VARIABLES (INTERPRETATION OF OR RISK MEASURES)

The model provides us with three (3) factors related to patient deaths among others: viral load 0.58(0.5-0.67), creatinine 1.98 (1.58-2; 49) and aspartate aminotransferase 0.99(0.9-1) . We found that patients with high creatinine levels were twice as likely to die as patients without creatinine. Discriminant analysis is for the explanation and prediction of the membership of individuals in groups (classes), represented by a categorical target variable, by the set of explanatory/ descriptive tables, mainly quantitative but also qualitative upon examination [21].

4) MODEL VALIDATION TEST (BOX TEST AND WILKS' LAMBDA)

To confirm or refute a model in discriminant analysis, we use the Box test and Wilks' lambda. For the Box test, the null hypothesis is the equality of the variance matrices, in our case the probability is lower than the significance level (p-value=0.00), so the null hypothesis is rejected, there is inequality of variances between the different groups. The Wilks' lambda is 0.48 (p-value=0.00), we conclude that our model is valid for the rest of the analysis (TABLE 4).

TABLE 4  
 Equation variables (Interpretation of OR risk measures)

Variables	B	E. S	Wald	ddl	Sig.	Exp(B)	95% confidence interval for EXP(B)	
							Lower	Superior
AGE	0.02	0.01	3.16	1	0.08	1.02	0.99	1.04
ctnp	-0.54	0.07	53.93	1	0.00	0.58	0.50	0.67
ch_cre_x	0.68	0.12	35.56	1	0.00	1.99	1.59	2.49
ch_glu_x	0.00	0.01	0.01	1	0.92	1.00	0.99	1.04
ch_pot_x	0.05	0.09	0.22	1	0.64	1.05	0.86	1.26
ch_sod_cu_x	-0.01	0.01	1.85	1	0.17	0.99	0.98	1.04
ch_alt_sgpt_x	0.00	0.01	0.11	1	0.74	1.00	0.99	1.01
ch_ast_sgot_x	-0.00	0.00	13.64	1	0.00	0.99	0.99	1.00
Constant	11,08	1,60	47,76	1	0,00	65413,16		

TABLE 5  
 Discriminant function

Function	Eigenvalue	of variance	Cumulative	Canonical correlation
1	1,085 <sup>a</sup>	100,0	100,0	0,721

5) CANONICAL DISCRIMINANT FUNCTION

We note that the 100% discriminating power attributed to the eight (8) variables is attributed to the first discriminating function. The relatively strong canonical correlation (72.1%) testifies to the high utility of the first discriminant function (TABLE 5).

6) COEFFICIENTS OF STANDARDIZED CANONICAL DISCRIMINANT FUNCTIONS

The linear discriminant analysis model yields three (3) factors associated with death in Ebola virus disease patients.

The three factors are: viral load (0.88), creatinine (0.94) and AST (0.78) (TABLE 6).

7) EVALUATE THE PREDICTION PERFORMANCE (PREDICTION PROBABILITY) OF TWO METHODS BY THE ROC CURVE.

The logistic regression method predicted mortality for 132 out of 363 subjects, or 36.5%, while the linear discriminant analysis predicted mortality for 141 out of 363 subjects, or 38.8%. Receiver Operating Characteristic (ROC) curve analysis was used to assess the accuracy of predictions between one or more models. In our particular case, the two models (binary logistic regression and linear discriminant analysis) provide almost the same information, i.e. the area

TABLE 6  
 Coefficient of the standardized canonical discriminant functions

Variables	Function
	1
Age	-0.06
ctnp	0.88
ch_creatinine_x	0.95
ch_glucose_x	0.04
ch_potassium_x	-0.04
ch_sodium_cu_x	0.17
ch_alt_sgpt_x	-0.10
ch_ast_sgot_x	0.78

under the curve (AUC) is 0.935 for the binary logistic regression and 0.932 for the discriminant analysis (FIGURE 1).

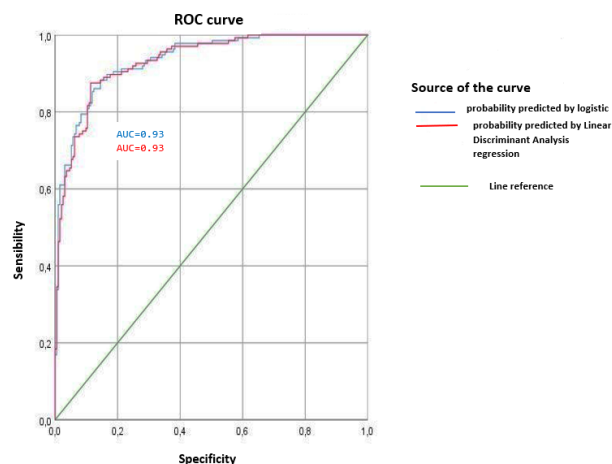


FIGURE 1. The ROC curve.

IV. DISCUSSIONS

To evaluate the performance of our study, we compared our results with those of the following authors: In the work by Rani, D. et al [21] entitled: A methodological comparison of

discriminant function analysis and binary logistic regression for estimating sex in forensic research and case-work, the aim of this study is to evaluate the accuracy of two multivariate statistical approaches for estimating sex from human external ear anthropometry, namely discriminant function analysis (DFA) and binary logistic regression (BLR). The predictive percentage of sex estimation calculated from the two models was approximately the same, i.e. 76.3% DFA and 76.2% BLR, for Altayeb Abdalla Ahmed et al [17], in their study of multidetector computed tomography: a comparison of discriminant function analysis and binary logistic regression. The study assessed sexual dimorphism in the Sudanese sterna using multidetector CT. This information was used to develop sex estimation models and to compare the accuracy of models based on discriminant function analysis (DFA) and binary logistic regression (BLR). Cross-validation sex estimation accuracies without a single deviation ranged from 60.4 to 88.9% for the DFA-based models and from 60.4 to 89.3% for the BLR-based models. Both studies show that discriminant analysis and linear logistic regression provide almost the same results. In our case, the study investigated the risk factors associated with mortality of patients with Ebola virus disease using two methods of analysis (logistic regression and linear discriminant analysis). We found the same risk factors influencing the mortality of the patients i.e. viral load, creatinine and aspartate aminotransferase and we evaluated the predictive accuracy of the two models (binary logistic regression and linear discriminant analysis) with the ROC curve, both methods provide almost the same information, i.e. the area under the curve (AUC) is 0.93 for the binary logistic regression and 0.93 for the discriminant analysis. The weakness of this study is that it only takes into account deterministic data to assess the risk factors of patients with Ebola virus, without considering fuzzy data.

## V. CONCLUSION

This study focuses on the identification of risk factors associated with death in Ebola patients using binary logistic regression (LR) and linear discriminant analysis (LDA) methods in an observational study (randomized controlled trial). These two methods are most often used to determine the risk factors associated with the disease and sometimes estimate the probability of predicting disease-related mortality based on one or more factors. In the literature we know that some researchers have compared the performance of these two methods in terms of model coefficient, which gives the same results, i.e. the predictors (model variables) we find for logistic regression are the same as those we find in discriminant analysis [8], [17], [21], [22]. The main objective of our study was to quantify the factors (variables) influencing patient mortality and to predict the probability of death by using two predictive analysis methods (logistic regression and linear discriminant analysis) to evaluate the results produced by these two models. Note that the coefficients of the logistic and discriminant models give us the same information, i.e. the same risk factors. The factors

provided by logistic regression are: viral load 0.58 (0.5-0.67), creatinine 1.98 (1.58-2.49) and aspartate aminotransferase 0.99 (0.9-1). For linear discriminant analysis we have viral load (0.88), creatinine (0.94) and aspartate aminotransferase (0.78) [25]. We see the same results even when different prediction probabilities are evaluated. The logistic regression method predicted a mortality rate of 132 out of 363 or 36.5%, while the linear discriminant analysis predicted a mortality rate of 141 out of 363 or 38.8%. The ROC curve analysis allowed us to assess the nature of the models' predictions. The result of the area under the curve (AUC): 0.935 for the binary logistic regression and 0.932 for the discriminant analysis of the logistic regression [2], [8], [17], [22]. In the future, we will carry out another study comparing the two methods used in this paper with other existing predictive methods such as multinomial logistic regression, neural discriminant analysis, Bayesian networks, etc. We also plan to use fuzzy data to evaluate the performance of our study in a complex environment. We also plan to use fuzzy data to evaluate the performance of our study in a complex environment.

## REFERENCES

- [1] Afsa C. Statistical Methodology M 2016/01 The Logit Model Theory and Application [Internet]. Available from: <http://www.insee.fr> 35
- [2] Ahmed AA, Koko AO, Bahar ME. Estimation of sex based on the sterna of Sudanese adults using 47
- [3] Ali, A. (2021). Data mining methods in psychiatric epidemiology: application to the analysis of risk factors and markers of depressive symptomatology in adolescence (Doctoral dissertation, Université Paris-Saclay).
- [4] Bonita R, Kjellström Tord, Beaglehole R, World Health Organization. Elements 30 of epidemiology. World Health Organization; 2010. 233 p.
- [5] Bouveyron C, Girard S, Schmid C. High Dimensional Discriminant Analysis [Internet]. Available from <https://hal.inria.fr/inria-00071243/25>
- [6] Desbois D. An introduction to discriminant analysis with SPSS for Windows [Internet]. Available 39 from: <http://www.sm.u-bordeaux2.fr/~corsini/Pedagogy/>
- [7] Duyme F, Clautriaux JJ, Daudin JJ. REVUE DE STATISTIQUE APPLIQUÉE EGRESSION LOGISTIQUE BINAIRE [Internet]. Available from:<http://www.sfds.asso.fr/publicat/rsa.htm>
- [8] Feng C, Kephart G, Juarez-Colunga E. Predicting COVID-19 mortality risk in Toronto, Canada: a comparison of tree-based and regression-based machine learning methods. *BMC Med Res Methodol.* 2021 Dec 1;21(1).
- [9] Fourati, S. (2022). Multiomic analysis for the identification of predictive biomarkers of vaccine response.
- [10] Ghasemi, E., & Gholizadeh, H. (2019). Development of two empirical correlations for tunnel squeezing prediction using binary logistic regression and linear discriminant analysis. *Geotechnical and Geological Engineering*, 37, 3435-3446.
- [11] Graf, R., Zeldovich, M., & Friedrich, S. (2022). Comparing linear discriminant analysis and supervised learning algorithms for binary classification-A method comparison study. *Biometrical Journal*.
- [12] Hasan, M. N. (2019, December). A comparison of logistic regression and linear discriminant analysis in predicting of female students attrition from school in Bangladesh. In *2019 4th international conference on electrical information and communication technology (EICT)* (pp. 1-3). IEEE.

- [13] Jaspard, M., Mulangu, S., Juchet, S., Serra, B., Dicko, I., Lang, H. J., ... & Malvy, D. (2022). Development of the PREDS score to predict in-hospital mortality of patients with Ebola virus disease under advanced supportive care: Results from the EVISTA cohort in the Democratic Republic of the Congo. *Eclinicalmedicine*, 5(4), 101699.
- [14] Jaspard, M., Mulangu, S., Juchet, S., Serra, B., Dicko, I., Lang, H. J., ... & Malvy, D. (2022). Development of the PREDS score to predict in-hospital mortality of patients with Ebola virus disease under advanced supportive care: Results from the EVISTA cohort in the Democratic Republic of the Congo. *Eclinicalmedicine*, 5(4), 101699.
- [15] Lacroix, A., Mbala Kingebeni, P., Ndimbo Kumugo, S. P., Lempu, G., Butel, C., Serrano, L., ... & Ahuka Mundeke, S. (2021). Investigating the circulation of Ebola viruses in bats during the Ebola virus disease outbreaks in the Equateur and North Kivu Provinces of the Democratic Republic of Congo from 2018. *Pathogens*, 10(5), 557.
- [16] Mulangu S, Dodd LE, Davey RT, Tshiani Mbaya O, Proschan M, Mukadi D, et al. A Randomized, 9 Controlled Trial of Ebola Virus Disease Therapeutics. *New England Journal of Medicine*. 2019 Dec 12;381(24):2293-303.
- [17] Altayeb Abdalla Ahmed, Alaa Osman Koko, Mustafa Elnour Bahar, multidetector computed tomography: a comparison of discriminant function analysis and binary logistic regression. *Homo*. 2021 Mar 21;72(1):41-51.
- [18] Pham, B. T., & Prakash, I. (2019). Evaluation and comparison of LogitBoost Ensemble, Fisher's Linear Discriminant Analysis, logistic regression and support vector machines methods for landslide susceptibility mapping. *Geocarto International*, 34(3), 316-333.
- [19] Pietri, M., & Bonnet, A. (2017, February). Alexithymia, emotional intensity and anxiety/depressive symptomatology: Explanatory dimensions of smoking. In *Annales Médico-psychologiques, revue psychiatrique* (Vol. 175, No. 2, pp. 146-152). Elsevier Masson.
- [20] Rani D, Krishan K, Kanchan T. A methodological comparison of discriminant function analysis and 43 binary logistic regression for estimating sex in forensic research and case-work. *Med Sci Law*. 2022 Nov 10;002580242211366
- [21] Rani, D., Krishan, K., & Kanchan, T. (2022). A methodological comparison of discriminant function analysis and binary logistic regression for estimating sex in forensic research and case-work. *Medicine, Science and the Law*, 00258024221136687.
- [22] Santos F, Guyomarc'h P, Bruzek J. Statistical sex determination from craniometrics: Comparison of 14 linear discriminant analysis, logistic regression, and support vector machines. *Forensic Sci Int*. 2014 15 Dec 1;245:204.e1-204.e8.
- [23] Steffen, I., Lu, K., Yamamoto, L. K., Hoff, N. A., Mulembakani, P., Wemakoy, E. O., ... & Simmons, G. (2019). Serologic prevalence of Ebola virus in equatorial Africa. *Emerging Infectious Diseases*, 25(5), 911.
- [24] Traoré, B. B. (2018). Information modeling and knowledge extraction for crisis management (Doctoral dissertation).
- [25] Tshomba, A. O., Mukadi-Bamuleka, D. R., De Weggheleire, A., Tshiani, O. M., Kitenge, R. O., Kayembe, C. T., ... & Mulangu, S. (2022). Development of Ebola virus disease prediction scores: Screening tools for Ebola suspects at the triage-point during an outbreak. *Plos one*, 17(12), e0278678.
- [26] Yan, D., Chi, G., & Lai, K. K. (2020). Financial distress prediction and feature selection in multiple periods by lassoing unconstrained distributed lag nonlinear models. *Mathematics*, 8(8), 1275.



**Leader Lawanga Ontshick, Msc. (ORCID: 0009-0005-5970-519X)**

Leader Lawanga Ontshick holds a master's degree in Applied Mathematics, multidimensional statistical option at the Department of Mathematics, Statistics and Computer Science of the University of Kinshasa, Faculty of Science and Technology. He has more than 7

years of experience in the field of statistics (Data analysis, design of quantitative and qualitative studies as well as clinical trials, design of study protocol and data analysis plan of a study). Since 2014, he has worked as a consultant Senior Statistician at the African Development Bank to support the National Institute of Statistics in the project on poverty (1-2-3 surveys in the Democratic Republic of Congo), Survey by cluster indicators of SDGs (Sustainable Development Goals and Indicators in the DRC), National Strategy for Environmental and Forest Statistics in the DRC with the World Bank, he worked as a senior statistician consultant to oversee the general business census (RGE) in the Democratic Republic of Congo. Researcher and Assistant in the Mathematics, Statistics and Computer Science Department of the University of Kinshasa in the Decisional Mathematics Unit. Biostatistician in the Department of Epidemiology and Global Disease and at the Clinical Research Center of the National Institute for Biomedical Research (INRB). With the National Institute of Disease (NIH-USA), he works as a biostatistician in the PALM project in collaboration with the INRB to set up (carry out) observational studies (clinical trials) finally to find solutions on emerging diseases (Ebola Virus Disease, Monkeypox, COVID-19, Measles...)

**Jean Christophe Mulangu Sabue, MD, DTMH, PhD. (ORCID :0000-0002-2722-2637)**

Jean Christophe Mulangu Sabue is Associate Professor at the Faculty of Medicine of the University of Kinshasa and Head of the Department of Clinical Immunology at the National Institute for Biomedical Research. For more than 20 years, he has been involved in several collaborative research projects in the filoviruses area. I have spent a decade (2006-2017) of research and training in the intramural program of NIH working with world-renowned scientists of whom the research laboratories have been visited by the President of the United States during the West Africa Ebola outbreaks. He had the opportunity to build an extensive expertise in research fields related to Ebola and Marburg viruses including surveillance, control, epidemiology, screening, ecology, immuno-vaccinology and clinical trials. During the 10th Ebola outbreak in DRC, He implemented on August 10th for the first time the Expanded Access Protocol of Ebola therapeutics in DRC and in November 2018, He led the first NIH sponsored RCT on Ebola treatments in DRC which has resulted to the identification of two proven Ebola therapeutics for the first time; Regeneron and Mab114. He is listed as one of the inventors of Mab114 which is anti-Ebola monoclonal antibody. Our research participation and efforts have led to publications in very high impact journals as highlighted below as well as to the preparation of my PhD thesis which I defended in March 2016 at the UNIKIN. Since April 2018, he has been promoted as Associate Professor at the Microbiology Unit, Department of Medical Biology, Teaching Hospital at the UNIKIN. Since July 2017, I led the Clinical Immunology Laboratory at National Institute for Biomedical Research. He conducts clinical immunology research and I teach immunology to post graduate students.



Honors, awards and invention

- 2017: Vaccine Research Center (VRC): For outstanding contributions of major importance to Vaccine Research Center work (VRC)
- 2015: NIH Director's Recognition Award: For exceptional dedication, skill, and leadership in accelerating the advancement of novel and groundbreaking Ebolavirus vaccines to address a global public health emergency.
- 2015: Listed as inventor on the patent application for mAb 114, US Application No.62/087, 087(PCT Application No. PCT/US2015/060733) related to anti-Ebola virus antibodies and their use.



**Placide Mbala-Kingebeni, MD, PhD.**

Placide Mbala-Kingebeni is an Associate Professor at the University of Kinshasa, School of Medicine; he is the Head of the Epidemiology and Global Health Division and Director of the Clinical Research Center at the National Institute of Biomedical Research (INRB in



French) in the Democratic Republic of the Congo. Dr. Mbala holds a PhD in virology; he is a medical doctor with over 10 years of experience leading research projects on viral zoonotic diseases such as Ebola and Mpox in collaboration with USAMRIID, Africa CDC, US CDC, UCLA, NIH, ITM-Antwerp, IRD-Montpellier, UNMC, Gates Foundation, etc., focusing on genomics, clinical research studies, passive and active surveillance as well as rapid response to outbreaks caused by these diseases. Placide Mbala-Kingebeni est professeur agrégé à l'Université de Kinshasa, dans la faculté de médecine ; il est Directeur de la Division Epidémiologie et Santé Globale et Directeur du Centre de Recherche Clinique à l'Institut National de Recherche Biomédicale (INRB) en République Démocratique du Congo. Le Dr Mbala est titulaire d'un doctorat en virologie ; il est médecin et compte plus de 10 ans d'expérience dans la direction de projets de recherche sur les zoonoses virales telles que Ebola et Mpox en collaboration avec USAMRIID, Africa CDC, US CDC, UCLA, NIH, ITM-Anvers, IRD-Montpellier, UNMC, Gates Foundation , etc., en se concentrant sur la génomique, les études de recherche clinique, la surveillance passive et active ainsi que la réponse rapide aux épidémies causées par ces maladies.

#### Olivier Tshiani Mbaya, MD, MTM, MSc

Olivier Tshiani Mbaya is a Senior Lecturer at the Faculty of Medicine of the University of Kinshasa and Deputy Head of the Department of Clinical Immunology at the National Institute for Biomedical Research. He is a medical Doctor with an expertise of more than 15 years participating in public health related work in resource-limited settings and research work in collaboration with international research teams focusing on emerging and re-emerging diseases. He has a considerable expertise on field of epidemiology, medical microbiology, immuno-vaccinology, and clinical research. He participated in several investigations of Ebola outbreak under World Health Organization (WHO) consultancy and contributed to panel of WHO Target Product Profiles for Mpox Therapeutics. He has more than 10 years conducting research in the intramural program of NIH under the leadership of world-renowned scientists of whom the President of the United States visited the research laboratories during the outbreak of Ebola in West Africa in 2014. During this largest and unprecedented West Africa's Ebola outbreak, his work contributed to accelerate the immunogenicity assessment of the Chimp-Adenovirus 3 vaccine (ChAd3), one of the first Ebola vaccine introduced in clinical trial and other fast track candidate vaccines such as the recombinant Vesicular Stomatitis Vaccine (rVSV) for dose determination. During the second largest Ebola outbreak in North-Kivu/DRC, he substantially contributed, as the clinical coordinator of the PALM study, to implement, execute and complete the first randomized trial, supported by the NIH, on investigational countermeasures against Ebola virus disease (EVD). He holds a master's degree in Tropical Medicine from the University of Nagasaki, where he is currently pursuing his PhD degree. Awards and Honors:

- 2017: Vaccine Research Center (VRC): For outstanding contributions of major importance to Vaccine Research Center work (VRC/NIH)
- 2015: NIH Director's Recognition Award for exceptional dedication, skill, and leadership in accelerating the advancement of novel and ground-breaking Ebolavirus vaccines to address a global public health emergency.
- 2015, 2014: Kelly Government Solutions Recognition Award for outstanding contributions of major importance to Vaccine Research Center work (VRC/NIH)
- 2008: Japan International Cooperation Agency (JICA): Scholarship Award for outstanding contribution in the response against Ebola outbreak in DRC (2008/09)

#### Dr Nsengi



Tuberculosis,

#### Ntamabyaliro, MD, MSc, PhD. (ORCID: 0000-0003-2929-9301)

Nsengi Ntamabyaliro is a Medical Doctor. Since 2007, he has worked as Principal investigator, investigator, clinical monitor, medical monitor in many interventional and observational study on Malaria, Monkeypox, HIV/AIDS, Human African

Trypanosomiasis, and Ebola Virus Disease. He is a senior lecturer at the Clinical Pharmacology Unit of the University of Kinshasa, Faculty of Medicine. He holds a PhD in Pharmacology from the University of Kinshasa and a Master's degree in Pharmacovigilance and Pharmacoepidemiology from the European Program in Pharmacovigilance and Pharmacoepidemiology (Eu2P). Dr. Nsengi Ntamabyaliro also works in the following fields:



- Clinical pharmacology
- Rational use of anti-infective drugs and prevention of microbial resistance
- Pharmacovigilance and management of drug induced diseases

#### Jean Jacques MUYEMBE TAMFUM, MD, PhD.

Professor of Microbiology, Faculty of Medicine, University of Kinshasa; Director of the National Institute of Biomedical Research (INRB); President of the Congolese Academy of Sciences (ACCS).

- Doctor of Medicine, Surgery and Childbirth, Lovanium University of Kinshasa in 1969;
- Co-discoverer of the Ebola virus in 1976; development of response strategies against Ebola Virus Disease epidemics (Kikwit 1995), co-inventor of the monoclonal antibody "mAb114", approved by FDA as an Ebola treatment under the name "EBANGA" in December 2020 ;
- Winner of the Christophe Mérieux Prize, 2015 (Paris), for his pioneering work on Ebola hemorrhagic fever in Africa;
- Winner of the 3rd Hideyo Noguchi Africa Prize (HNAP), 2019 (Tokyo);
- Named by Nature's among the 10 personalities who mattered in science in 2018 and 2019 ;
- Named by Time Magazine as one of the 100 most influential people in 2020 ;
- Winner of the "CPHIA 2021" Excellence Award, presented by the African Union and the African Centers for Disease Control and Prevention (Africa CDC), December 2021.
- Global Health Leaders Awards (WHO), May 2023.

#### Rostin Mabela Matendo Makengo Phd

Rostin Mabela is a PhD in Mathematical Sciences since 2008 and a full professor at the University of Kinshasa since 2021. He has been alternately Junior Assistant, Senior Assistant, Associate Professor, Full Professor. Currently, He is Head of Department of Mathematics, Statistics and Computer Science at the Faculty of Science, University of Kinshasa. He is also the General Director of the Centre of Research in Mathematic Education at the Ministry of Scientific Research and Technological Innovation. In the past, he had to lead the Backbone Center of the University of Kinshasa for the management of the Internet network. His current research focuses on the analysis of queueing systems in a fuzzy environment. He also conducts research in data mining in classical and fuzzy environments. He is a member of the African Mathematical Union. He has already directed four PhD theses in applied mathematics and co-directed other theses. He has participated in many scientific conferences at both national and international levels.

