

## Respiratory Syncytial Virus Infection in Infants: A Comparative Study Using Discriminant, Probit and Logistic Regression Analysis

Orumie, Ukamaka Cynthia<sup>a</sup> and Desmond Chekwube Bartholomew<sup>b\*</sup>

<sup>a</sup> Department of Mathematics and Statistics, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria.

<sup>b</sup> Department of Statistics, Federal University of Technology Owerri, Imo State, Nigeria.

### Authors' contributions

*This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.*

### Article Information

DOI: 10.9734/AJPAS/2022/v18i230440

### Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://www.sdiarticle5.com/review-history/88607>

*Received 17 April 2022*

*Accepted 25 June 2022*

*Published 02 July 2022*

**Original Research Article**

## Abstract

In babies, respiratory syncytial virus (RSV) is the most common cause of lung inflammation (pneumonia) or bronchiolitis (inflammation of the lungs' airways). This virus comes with several symptoms such as congested or runny nose, dry cough, low-grade fever, sore throat, sneezing, headache, difficulty in breathing etc. The virus can cause death in babies if not properly managed and therefore calls for immediate investigations to reveal the significant causes. Several research works have been conducted but the idea of investigating more potential predictor variables and the application of both regression and classification models have been grossly understudied. Therefore, unpublished secondary data collected from three different hospitals in Port Harcourt, Rivers State, Nigeria on fifteen predictor variables which are potential causes of RSV are modeled using two categorical regression approaches – logistic and probit regression models and one classification model – discriminant function analysis. The models were compared using misclassification errors, Receiver Operating Characteristic (ROC) plot, concordance, sensitivity, specificity and pseudo R-square values. The linear discriminant function model outperformed both the logit and probit models. The results showed that paternal history of asthma, maternal history of asthma, mother's occupation, mother's smoking habit and mother's education level were the most important variables to linearly classify seropositive and seronegative RSV patients.

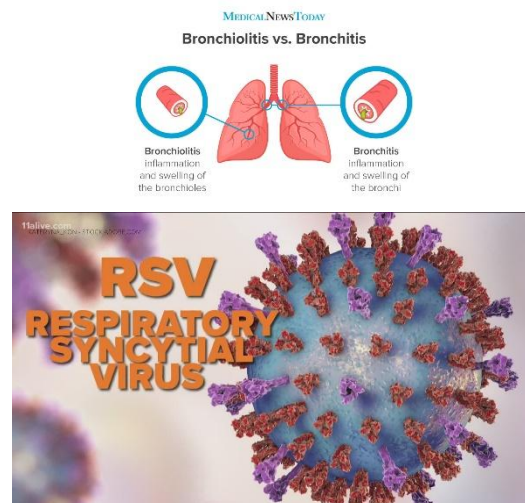
\*Corresponding author: Email: [desmond.bartholomew@futo.edu.ng](mailto:desmond.bartholomew@futo.edu.ng);

*Keywords: Probit model; binary logistic regression model; discriminant analysis; ROC plot; respiratory syncytial virus.*

## 1 Introduction

Respiratory syncytial virus (RSV) is a recognized major contributor to infant hospitalizations in the whole world Caroline et al. [1]. The symptoms of this respiratory virus are usually mild and cold-like; can be severe in infants and aged ones and could develop into bronchiolitis and pneumonia. According to the Center for Disease Control and Prevention (CDCP), RSV causes approximately 2.1 million outpatient visits (for babies < 5yrs), 58,000 hospitalizations (for babies < 5yrs), 177,000 hospitalizations among adults 65 and older and 14,000 deaths among adults 65 and older in the United States and other areas with similar climates in a year.

According to a report by National Centre for Immunization and Respiratory Diseases (NCIRD) in 2018, the symptoms take between 4 to 6 days to appear. These symptoms include Runny nose, loss of appetite, cough, sneezing, difficulty in breathing, fever, wheezing, etc. This RSV infection leads to severe ones such as an inflammation of the small airways in the lung and acute respiratory infection that affects the lungs Rha et al. [2]. Again, asthmatic individuals may have asthma episodes as a result of RSV infection and people with congestive heart failure may have more severe symptoms as a result of RSV infection Falsey et al. [3]. Acute lower respiratory infection (ALRI) is one of the primary causes of morbidity and mortality in babies Shi et al. [4] and human respiratory syncytial virus (RSV) is the most prevalent viral pathogen found in ALRI patients. Researchers are working on developing RSV vaccinations, but none are currently available Ramilo and Mejias [5]



<https://www.11alive.com/article/news/health/doctors-report-spike-in-rsv-cases-in-metro-atlanta-as-covid-restrictions-lift/85-37322ac6-1943-4efe-a492-b974ed1c136e>

The RSV has been broadly studied and there exist some published works on findings relating to the virus. Light et al. [6] established a relationship between respiratory syncytial virus (RSV) and lower respiratory tract disease hospitalizations in Florida. Caroline et al. [1] reported that RSV infection in children in the United States is connected to considerable morbidity in both hospital and outpatient settings. Obodai et al. [7] in their study about RSV genotypes circulating in urban Ghana used a nested multiplex reverse transcriptase polymerase chain reaction (RT-PCR) method for genotyping RSV. The results revealed that RSV B infections accounted for 72 percent (23/32) of the 60.4 percent RSV infections whereas RSV A and B co-infections accounted for 28 percent (9/32) and age group of children between the ages of 2 and 12 months were the most affected. This finding was further verified by Aliyu et al. [8] which gave a statistical link between age and seropositivity. The findings revealed that seroprevalence was highest in the age group 49-60 months and lowest in the age group 0-12 months and that RSV infection was quite common among children aged 1-5 years in Zaria, Kaduna State, Nigeria. In 2017, Shi et al. [4] investigated the global, regional and national illness burden estimates of acute lower respiratory infections caused by RSV in children under the age of five. According to their findings, there

were roughly 6 million occurrences of RSV severe ALRI in infants in 2015 with 51 percent dying in hospitals and 49 percent dying outside of hospitals. They also discovered that there are around 118,000 serious newborn deaths in poor countries. Oluwadamilare et al. [9] conducted a study on RSV infections in children in Lagos State, Nigeria. According to their findings, RSV was most prevalent in children aged 2 to 6 months (32.4 percent). Further, Abdul et al. [10] named malignancy a risk factor for RSV infection and also revealed that cancer and immunosuppressive treatment were substantially linked with RSV detection. Ruimu & Jikui [11] looked at clinical characteristics and differences between infants with single Bordetella pertussis (*B. pertussis*) infection and those with RSV coinfection in their study on the Clinical impact of RSV infection on children hospitalized for Pertussis. They reported that sex, body weight, preterm birth history, pertussis vaccination, pneumonia presence and lymphocyte count did not differ significantly between the two groups. Using probabilistically linked perinatal, hospital and laboratory records of 321,825 children born in Western Australia (WA) from 2000 to 2012, Gebremedhin et al. [12] built a prediction model for RSV positivity in hospitalized children aged 5 years. RSV positivity was observed to be associated with a younger admission age, male gender, non-Aboriginal ethnicity, bronchiolitis diagnosis and a longer hospital stay.

The present work is different from a previous work that used a combination of discriminant and binary logistic regression and considered only three variables (Sex, Weight at birth and Weight after four weeks) from two hospitals Beki [13]. Also different from the work of Danbaba et al. [14] that investigated some variables considered from mother's aspect but used only logistic regression. Yakubu et al. [15] considered a combination of mother's and baby's independent variables and compared Discriminant Analysis and Logistic Regression models on Broncho-Pulmonary and not RSV. The present study is different from the rest because it attempted to compare three models - Discriminant Function Analysis (Linear and Quadratic score functions), Probit Regression and Logistic Regression models using data from three hospitals (University of Portharcourt Teaching Hospital, Aluu Health Center and Braitwait Memorial Specialist Hospital) with more predictor variables.

## 2 Materials and Methods

### 2.1 Logistic Regression Model

The linear regression that uses the ordinary least squares method to minimize sum of the squared deviations is often used for predicting continuous Y variables while logistic regression is used for categorical Y variable classification.

It is inappropriate to use linear regression to model a categorical (dichotomous) Y variable since the resulting model will give predicted Ys outside 0 and 1. Also, other numerous assumptions such as normality of the errors may also get violated.

The response variable RSV is categorical with two levels (seropositive or seronegative) and therefore suggests that the appropriate logit model is the binary logistic regression model. Since the response variable is dichotomous, the mathematical foundation and representation for modeling the log odds of the event

$\log_e \left( \frac{P}{1 - P} \right)$  where  $P$  is the probability of an event are described by Orumie et al. [16].

### 2.2 Probit Regression Model

When the conditional probability function is assumed to be linear, the probit or logit models are favoured over the standard least squares regression model. Again, the response variable in consideration must be categorical (dichotomous). While logistic regression utilizes a cumulative logistic function for the estimate model, probit regression uses a normal cumulative density function.

In Probit regression, the cumulative standard normal distribution function  $\phi(\cdot)$  is used to model the regression function, that is, we assume

$$E(Y | X) = P(Y = 1 | X) = \phi(\alpha + \beta X) \quad (1)$$

where

$\alpha + \beta X$  is regarded as a quantile  $Z$ .

$\alpha$  is the usual regression intercept term

$\beta$  is the change in  $Z$  associated with a one unit change in  $X$

Although the effect on  $Z$  of a change in  $X$  is linear, the link between  $Z$  and the response variable  $Z$  is nonlinear since  $\phi$  is a nonlinear function of  $X$ .

For a binary  $Y$  variable, the model

$$Y = \alpha + \sum_{i=1}^k \beta_i X_i + \varepsilon \quad (2)$$

with

$$P(Y = 1 | X_1, X_2, \dots, X_k) = \phi\left(\alpha + \sum_{i=1}^k \beta_i X_i\right) \quad (3)$$

Equation (3) is a population probit model with multiple regressors  $X_i$ ,  $i = 1, 2, 3, \dots, k$  and  $\phi(\cdot)$  is the cumulative standard normal distribution function.

### 2.3 The Discriminant Function Analysis (DFA)

This is another good model for application when the response variable is categorical but classification is the problem. The Logit and Probit models predict while the DFA classifies. It uses a linear or quadratic score function to classify observations into mutually exclusive groups within a categorical variable. It uses either continuous or categorical or both types of variables as the predictor variables. It can be a simple or multiple DFA. It is simple when only one predictor variable is used, otherwise it is multiple. As described by Egbo and Bartholomew [17], the discriminant function analysis uses any of the three known methods of assigning class probabilities: equal, arbitrary or estimated approaches to find the class probabilities while building the model. The estimated class probability approach was adopted in this study. The linear or quadratic discriminant function uses a linear or quadratic score function respectively as described by fisher [18].

### 2.4 Methods of Comparison

In this paper, the entire data is divided into two sets: training set and test set. The training set contains 70% of the entire sample size while the remaining is used as test set to validate the models. Simple methods like misclassification errors, ROC plot, concordance, sensitivity, specificity and pseudo r-square values are used for comparison purposes. At first, the probit and logit models is compared using the training set and the best model then compared with the best DFA type (linear or quadratic).

## 3 Empirical Results

The data used for this study are unpublished secondary data obtained from three hospitals mentioned earlier with the following sample sizes (UPTH = 263 records, AHC = 233 records and BMH = 257 records). These gave a total sample size of 753 records. The 120 rows with missing values were removed from the data leaving 633 valid cases as the total sample size. Out of the 633 valid cases, 285 are seronegative (absence of RSV) while 348 are seropositive (presence of RSV). The fifteen predictor variables for this study are : baby's gender (male or female), baby's weight at birth (low, normal or macrosoma), mother's smoking habit (yes or no), catarrh (yes or

no), running nose (yes or no), maternal history of ashma (yes or no), paternal history of ashma (yes or no), breastfeeding method (exclusive or nonexclusive), IgG antibody level (low, normal or high), mother's age, baby's weight 4week after, baby's age (grouped in months), maternal education level (primary, secondary or tertiary), mother's occupation (business, civil servant or house wife) and the hospital (UPTH, BMS or AHC).

### 3.1 Sampling Method

The simple random sampling method was used to randomly sample 70% of 285 (which is 199) and 70% of 348 (which is 243) to make up the training data set of size 442 (199+243). The sampling was necessary in order to solve the problem of class bias. Ideally, the proportion of events (seropositive) and non-events (seronegative) in the Y variable should approximately be the same but for the case of the data for this study, 285 is reasonably bigger than 348. The remaining 191 formed the sample size for the test set.

**Table 1. Logistic and Probit Regression model outputs on training set**

Training Data	Logit		Probit		Sig
	Estimate	Pr(> z )	Estimate	Pr(> z )	
Coefficients: (Intercept)	0.759444	0.739286	0.513519	0.701804	NS
<b>Gender (Male)</b>					
Female	-0.26296	0.288825	-0.14506	0.313336	NS
<b>Baby's weight at Birth (&lt; 2.5kg)</b>					
(2.5 - 4.2kg)	-0.11314	0.757531	-0.05052	0.812391	NS
(> 4.2kg)	0.237745	0.699562	0.143445	0.687549	NS
<b>Maternal smoking habit (No)</b>					
YES	0.122232	0.760897	0.08021	0.73347	NS
<b>Breastfeeding method (Exclusive)</b>					
Non Exclusive	-0.05525	0.849773	-0.04184	0.804631	NS
IgG-Low	0.264255	0.319554	0.154463	0.315354	NS
<b>IgG antibody level (Normal)</b>					
High	-1.52729	0.000336	-0.88224	0.000322	***
<b>Baby has cataarh? (No)</b>					
YES	-0.69299	0.468306	-0.41223	0.463152	NS
<b>Baby has running nose (No)</b>					
YES	-0.30858	0.852254	-0.18974	0.846627	NS
<b>Baby age (1-6 months)</b>					
7 - 12months	0.050546	0.875982	0.003332	0.985795	NS
13 - 18months	0.14371	0.670502	0.061412	0.754765	NS
19 - 24months	-0.00012	0.999772	-0.03571	0.88589	NS
25 - 30months	-1.04543	0.347422	-0.65232	0.317402	NS
<b>Maternal Edu. Level (No Educ.)</b>					
Secondary	0.827623	0.02351	0.506965	0.017888	*
Tertiary	0.668238	0.045781	0.406907	0.037828	*
<b>Maternal occupation (Business)</b>					
Civil Servant	-0.25699	0.449345	-0.06892	0.723597	NS
House Wife	1.982968	1.79E-11	1.204413	1.60E-12	***
<b>Maternal History of Asthma (No)</b>					
YES	1.862673	4.72E-10	1.088443	2.44E-10	***
<b>Paternal History of Asthma (No)</b>					
YES	3.654589	< 2e-16	2.149797	< 2e-16	***
<b>Hospital (UPTH)</b>					
Hospital-BMH	-0.1326	0.634778	-0.11135	0.493809	NS
Hospital-AHC	-0.14058	0.658759	-0.1142	0.536456	NS
<b>Maternal age</b>	-0.0415	0.080406	-0.02675	0.050559	NS
<b>Baby's weight 4weeks after</b>	-0.10234	0.614151	0.117893	0.596511	NS

Key: NS – not significant, \*\*\* - significant at 0.05 and 0.01 alpha values, \* significant at 0.05 alpha value

### 3.2 Data Structure

Out of the fifteen categorical variables, all except mother's age and baby's weight 4 weeks after are categorized. The following continuous variables were recoded: baby's weight at birth and baby's age (in months). Baby's weight at birth were categorized into low birth weight (less than 2.5kg), normal birth (between 2.5kg and 4.2kg) weight and macrosoma (over 4.2kg) as described in Bartholomew et al. [19] while the baby's age was grouped into 5 age groups : (1 - 6months, 7 - 12months, 13 - 18months, 19 - 24months and 25 - 30months). For the categorical variables, one of the levels is chosen as the reference level while for the outcome variable, the reference category was chosen as seropositive babies. The reference levels are enclosed in bracket in Table 1.

### 3.3 Comparison of Logit and Probit Model Results

The logit and probit models were implemented using the training data set and the results are summarized in Table 1. and Table 2.

**Table 2. Logit and Probit model validation results for test data**

Models	Comparison Criteria (Test Set)				
	AIC	Concordance	Sensitivity	Specificity	Pseudo R-square
Logit	489.97	0.7323	0.7143	0.6744	0.2985
Probit	474.56	0.7297	0.7048	0.6628	0.2988

*Confusion Matrix for Logit model*

	0	1
0	[ 58	30 ]
1	[ 28	75 ]

*Confusion Matrix for Probit model*

	0	1
0	[ 57	30 ]
1	[ 29	74 ]

**Table 3. Odd Ratio and its 95% Confidence Interval for Logit Model (sig. variables only)**

Coefficients	Odd Ratio	2.50%	97.50%
<b>IgG antibody level (Normal)</b>			
High	0.217124	0.09225	0.492684
<b>Maternal Edu. Level (No Educ.)</b>			
Secondary	2.287873	1.124857	4.726974
Tertiary	1.950798	1.018132	3.791326
<b>Maternal occupation (Business)</b>			
House Wife	7.26427	4.128636	13.15493
<b>Maternal History of Asthma (No)</b>			
YES	6.44093	3.636882	11.77849
<b>Paternal History of Asthma (No)</b>			
YES	38.65162	17.76243	91.88824

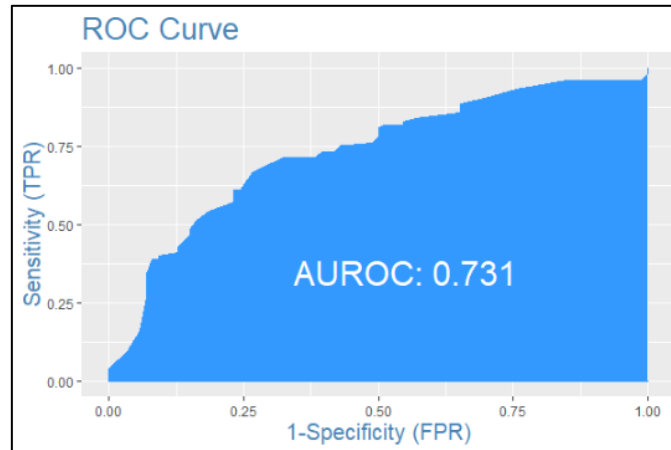
### 3.4 The Discriminant Function Analysis (DFA) Results

#### 3.4.1 Linear DFA Results

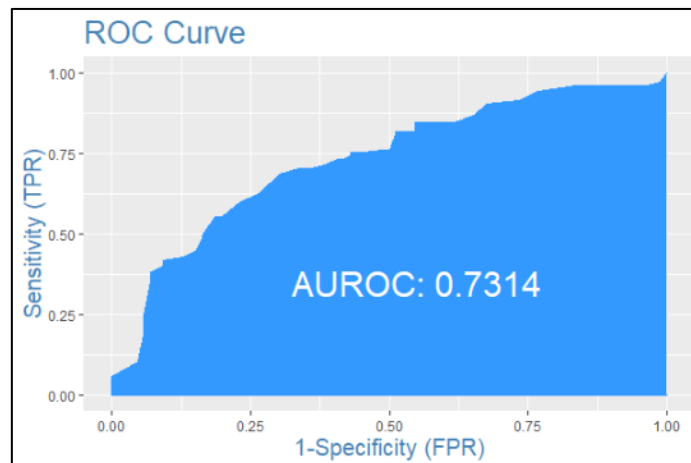
Because DFA presupposes multivariate normality, the data must be examined for significant deviations from normality before proceeding with the analysis. The first step, as always, is to plot the data to see if any outliers need to be removed or if any data transformations are required, particularly for the continuous predictor variables (baby's weight 4 weeks after birth and mother's age).

**Table 4. Shapiro Wilk's normality test**

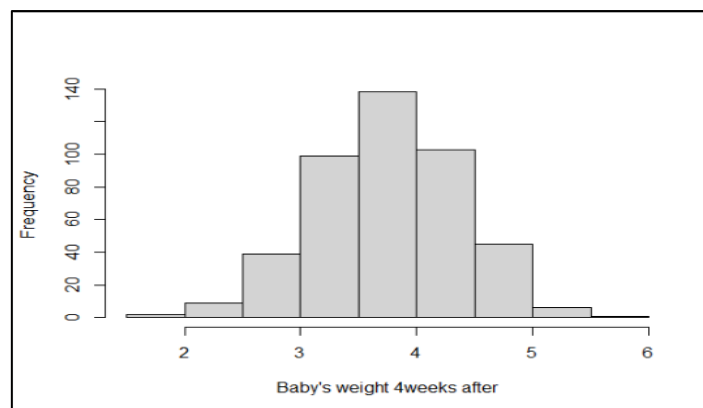
	Baby's weight 4weeks after	Mother's age
W	0.99474	0.99434
p-value	0.1356	0.1013



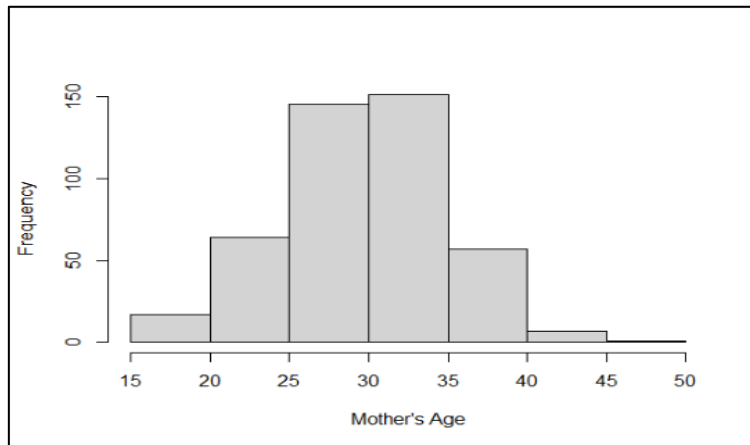
**Fig. 1. ROC curve for logit model (test data)**



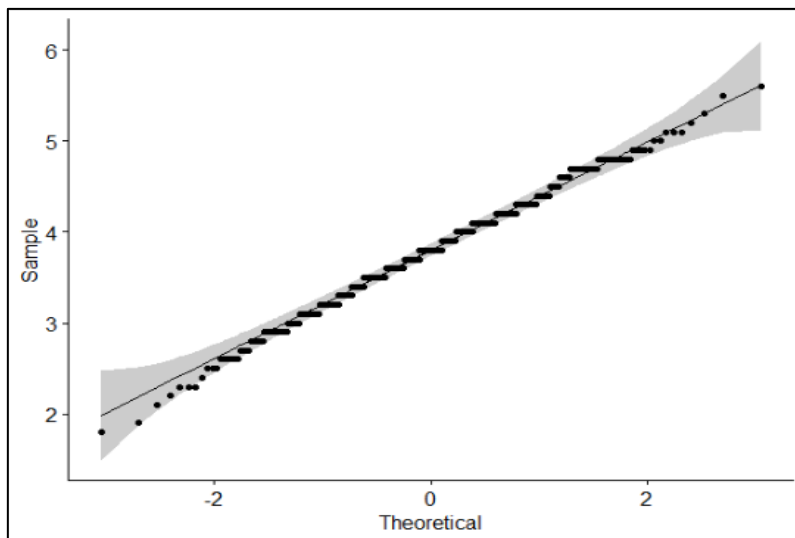
**Fig. 2. ROC curve for probit model (test data)**



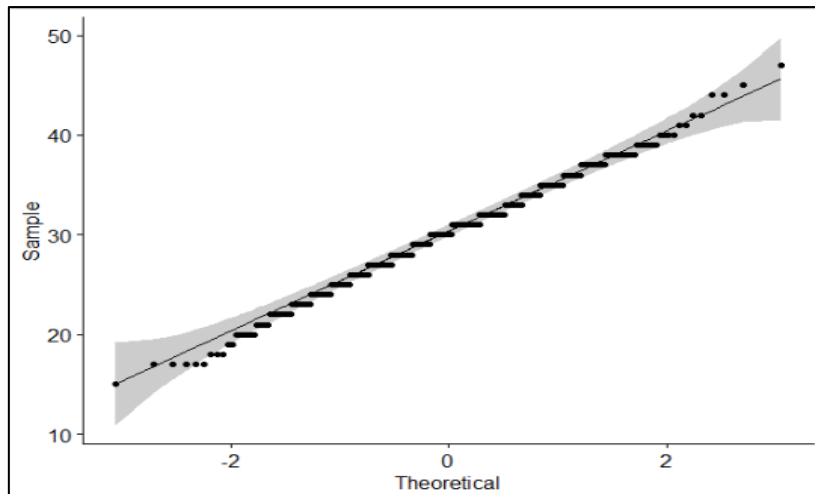
**Fig. 3. Histogram plot of baby's weight 4weeks after**



**Fig. 4. Histogram plot of mother's age**



**Fig. 5. Quantile-Quantile plot of baby's weight 4 weeks after**



**Fig. 6. Quantile-Quantile plot of mother's age**



**Table 5. LDA Result for the training data**

<b>Prior probabilities of groups:</b>												
0	1											
0.45	0.55											
<b>Group means:</b>												
	Gender	BWB	MSH	BBM	IgG	Cataarh	BA(Months)	MEL	MO	MAH	PAH	Hospital
0	1.56	1.91	0.10	1.21	1.65	0.99	2.34	2.35	1.68	0.18	0.07	1.91
1	1.53	1.95	0.12	1.24	1.58	0.97	2.22	2.37	0.12	0.37	0.35	1.86
	MA	BW4W	RN									
0	30.43	3.82	0.99									
1	29.84	3.76	1.00									
<b>Coefficients of linear discriminants:</b>												
												LD1
<b>Gender</b>												-0.20
BWB (Baby's weight at birth)												0.10
MSH (Maternal smoking habit)												0.18
BBM (Baby's breastfeeding method)												-0.01
IgG (IgG antibody level)												-0.32
RN (Baby has running nose?)												-0.42
Cataarh (Baby has catarrh?)												-0.65
`BA(Months)` (Baby age in months)												-0.05
MEL (Maternal education level)												0.15
MO (Maternal occupation)												0.66
MAH (Maternal asthma history)												1.17
PAH (Paternal asthma history)												2.28
Hospital (Hospital admitted)												-0.05
MA (Maternal age)												-0.03
BW4W (Baby's weight 4weeks after)												-0.08

**Table 6. QDA Result for the training data**

<b>Prior probabilities of groups:</b>												
0	1											
0.45	0.55											
<b>Group means:</b>												
	Gender	BWB	MSH	BBM	IgG	Cataarh	BA(Months)	MEL	MO	MAH	PAH	Hospital
0	1.56	1.91	0.10	1.21	1.65	0.99	2.34	2.35	1.68	0.18	0.07	1.91
1	1.53	1.95	0.12	1.24	1.58	0.97	2.22	2.37	0.12	0.37	0.35	1.86
	MA	BW4W	RN									
0	30.43	3.82	0.99									
1	29.84	3.76	1.00									
<b>Coefficients of Quadratic discriminants:</b>												
	QD											
Gender	0.10											
BWB (Baby's weight at birth)	0.23											
MSH (Maternal smoking habit)	-0.03											
BBM (Baby's breastfeeding method)	0.18											
IgG (IgG antibody level)	0.09											
RN (Baby has running nose?)	1.70											
Cataarh (Baby has catarrh?)	-0.52											
`BA(Months)` (Baby age in months)	-0.11											
MEL (Maternal education level)	0.09											
MO (Maternal occupation)	0.14											
MAH (Maternal asthma history)	-0.05											
PAH (Paternal asthma history)	0.02											
Hospital (Hospital admitted)	0.15											
MA (Maternal age)	0.00											
BW4W (Baby's weight 4weeks after)	1.73											

**Table 7. Comparison of LDA and QDA on test set**

	<b>LDA</b>	<b>QDA</b>
Model Accuracy	0.63	0.62
Actual number of observations in group 1	86	86
Number classified	81	93
Actual number of observations in group 2	105	105
Number classified	110	98

### 3.5 Comparison of LDA, Logit and Probit Models

**Table 8. Comparison of LDA, Logit and Probit Models on test set**

	<b>LDA</b>	<b>Logit</b>	<b>Probit</b>
Model Accuracy	0.63	0.71	0.70
Actual number of observations in group 1	86	86	86
Number classified	81	58	57
Actual number of observations in group 2	105	105	105
Number classified	110	75	74

## 4 Discussion of Results

The Logit and Probit models were fitted using the training set and the results were displayed in Table 1. Both models behaved alike. Out of the fifteen predictor variables considered in this study, both models identified IgG antibody level, Mother's education level, Mother's occupation, maternal history of ashma and Paternal history of ashma as the significant predictor variables for seropositive RSV babies. The two models were further compared based on the comparison criteria listed in this study in section 2.4. The values shown in Table 2 suggests that Logit and Probit models performed equally but the Logit model performed slightly better as seen in the confusion matrix. The higher the values on the diagonal of the confusion matrix, the better the model. As the prediction probability cutoff is lowered from 1 to 0, the Receiver Operating Characteristics Curve (ROC) tracks the percentage of true positives accurately predicted by a given logit model. As the cutoff is lowered, a successful model should mark more genuine 1s as positives and fewer actual 0s as 1s. When a result, for a decent model, the curve should rise steeply, suggesting that as the cutoff score decreases, the TPR (Y-Axis) climbs faster than the FPR (X-Axis). The greater the area under the ROC curve, the higher the model's prediction performance. The ROC curves for both models in Fig. 1 and Fig. 2 are nearly identical. Though, the probit model appears to be a better fit for the training set (Table 2. Probit AIC value of 474.56 is smaller than Logit AIC value of 489.97) but the logit model has more predict power (Table 2. Sensitivity value of logit is higher than that of probit), therefore the logit model is chosen as the best model and was used on the test set. The odd ratio and the corresponding 95% confidence intervals were displayed in Table 3. According to Orumie et al. [16], our results are interpreted as follow:

- Those with low IgG antibody level when compared with those with high IgG antibody levels are 0.217 times likely to test positive for RSV. That is about  $(1 - 0.217 = 0.783)$  78.3% more likely to be seropositive for RSV.
- A baby whose mother's highest education qualification is primary school when compared with another baby whose mother's highest education level is secondary school and Tertiary education level is 2.28 and 1.95 times more likely to be seropositive to RSV respectively. That is  $(2.28 - 1.00 = 1.28)$  128% and  $(1.95 - 1.00 = 0.95)$  95% more likely to be attacked by RSV respectively. This is in line with the findings of Bartholomew et al. [19], who found that the mother's educational level has a significant impact on the likelihood and/or productivity of health investment, as well as the financial resources available to the child, both directly and indirectly, through the choice of partner, fertility timing, and number of offspring.
- A baby whose mother is a House wife when compared with another baby whose mother is either a civil servant or business woman is 6.44 times more likely to be seropositive to RSV.
- With odds ratios of 6.44 and 38.7, maternal and paternal history of ashma has relatively the highest odds for RSV. This suggests that the variables maternal and paternal history of ashma were the most

significant variables in predicting the incidence of RSV in the three hospitals. This finding is in agreement with the finding of Abdul et al. [10].

The Linear Discriminant Function (LDF) and Quadratic Discriminant function (QDF) were fitted on the training set using the estimated priors to estimate class probabilities and the best discriminant model was obtained before comparing with logistic regression model and the results were displayed from section 3.4. The continuous predictor variables were first tested for normality assumptions since that is one of the basic assumptions of the DF model. Both graphical methods (Fig. 3 through Fig. 6) and Shapiro wilk's normality test (Table 4.0, p-values 0.1356 and 0.1013 are less than 0.05) suggested that baby's weight 4weeks after and mother's age came from a normal distribution. LDA and QDA results were shown in Table 5 and Table 6 respectively. LDA performed slightly better than the QDA (Table 7, model accuracy) on the test set and was compared with Logit model in Table 8. The LDA performed better than both logit and probit models in predicting the incidence of RSV in the three hospitals (Table 8, more number of observations correctly classified). This finding is in agreement with Sule [20]. Each discriminant function should be scanned for the largest loadings, positive or negative, indicating the variables that contribute most to that discriminant function. Therefore, the variables that contributed as shown in Table 5 under coefficients of linear discriminants are Paternal history of ashma, Maternal history of ashma, mother's occupation(in agreement with Yakubu et al. [15], mother's smoking habit (in agreement with Ting et al. [21] and mother's education level.

## 5 Conclusion and Recommendations

On the basis of the above findings, it is concluded that the logit model outperformed the probit model while the LDA outperformed QDA in predicting the incidence of RSV in the three hospitals considered. The three models (logit, probit and LDA) were compared and LDA was chosen as the best model over the logit model. However, the three models has similar results as, maternal history of ashma, paternal history of ashma, mother's education level and mother's occupation were important contributors to a baby having RSV. Thus, it is recommended that

- i. Babies whose either or both parents has(have) history of ashma should be quickly immunized to increase the IgG antibody level in the baby to fight RSV.
- ii. House wives should improve their attention to health care practices for their babies and always seek medical attention from health practitioners. They should try to avoid self-medical practices.
- iii. Mothers who smoke should stop as this study revealed that 38 out of the 348 (that is 11%) babies that were seropositive had smoking mothers.
- iv. Also mothers should embrace education as educated mothers usually change their perceptions regarding how best to allocate resources for the betterment of children's health and fewer but healthier children.

## Competing Interests

Authors have declared that no competing interests exist.

## References

- [1] Caroline Breese Hall MD, Geoffrey A, Weinberg MD, Marika KI, Aaron KB, Kathryn MS, Edwards M, et al. The Burden of Respiratory Syncytial Virus Infection in Young Children, the New England Journal of Medicine. 2009;360:588-598.
- [2] Rha B, Curns AT, Lively JY, et al. Respiratory Syncytial Virus–Associated Hospitalizations among Young Children: 2015–2016. *Pediatrics*. 2020; 146(1).
- [3] Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE. Respiratory syncytial virus infection in elderly and high-risk adultsexternal icon. *New England Journal of Medicine*. 2005;352(17):49-59.
- [4] Shi T, Angeline Denouel, Anna K Tietjen, Iain Campbell, Emily Moran, Xue Li, Harry Campbell, Clarisse Demont, Bryan O Nyawanda, Helen Y Chu, Sonia K Stoszek, Anand Krishnan, Peter Openshaw,

Ann R Falsey, Harish Nair. Global, regional and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study external icon. *Lancet. Journal*. 2017;390(10098):946-985.

- [5] Ramilo O, Mejias AP. Measuring the Burden of RSV in Children to Precisely Assess the Impact of Preventive Strategies. 2020;146(1):e20201727. DOI: 10.1542/peds.2020-1727. Epub 2020 Jun 16.PMID: 32546584 No abstract available.
- [6] Light M, Bauman J, Mavunda K, Malinoski F, Eggleston M. Correlation between respiratory syncytial virus (RSV) test data and hospitalization of children for RSV lower respiratory tract illness in Florida. *Pediatr Infect Dis J*. 2008;27(6):2-8.
- [7] Obodai, Evangeline, Richard Asmah, Isaac Boamah, Bamenla Goka, John Kofi Odoom, and Theophilus Adiku. Respiratory syncytial virus genotypes circulating in urban Ghana: February to November 2006. *The Pan African Medical Journal*. 2014;19.
- [8] Aliyu A, Detalle L, Stohr T, Palomo C, Piedra PA, Gilbert BE, Mas V, et al. Generation and characterization of ALX-0171, a potent novel therapeutic nanobody for the treatment of respiratory syncytial virus infection. *Antimicrobial agents and chemotherapy*. 2015;60(1):6-13.
- [9] Oluwadamilare AO, Bamidele WM, Abimbola A. Respiratory Syncytial Virus Infection among Children in Lagos, Nigeria. *Journal of Clinical Immunology & Microbiology*. 2021;2(1).
- [10] Abdul MK, Obianuju GA, Waliyah M, Nazia A, Saima J, Ayub K, Hanya MQ, Inci Y, Fauzia AM, Saad BO. Respiratory Syncytial Virus-Associated Mortality Among Young Infants in Karachi, Pakistan: A Prospective Postmortem Surveillance Study. *Clinical Infectious Diseases*. 2021;73(3):203–209.
- [11] Ruimu Z, Jikui D. Clinical impact of respiratory syncytial virus infection on children hospitalized for pertussis *BMC Infectious Diseases* volume 21, Article number. 2021;21:161.
- [12] Gebremedhin AT, Hogan AB, Blyth CC, et al. Developing a prediction model to estimate the true burden of respiratory syncytial virus (RSV) in hospitalised children in Western Australia. *Sci Rep*. 2022;12:332.
- [13] Beki OD. The Use of Discriminant Model and Logistic Regression for Tracking the Incidence of Broncho-Pulmonary Dysplasia among Infants. An unpublished Dissertation Submitted to Usman Danfodio University, Sokoto Nigeria; 2012.
- [14] Danbaba A, Audu M, Mahmud AM. Investigation of risk factors of low birth weight using multivariate logistic regression analysis. *J. of Hum, Sci. Technology*. A publication of Niger State Polytechnic. 2013;3:59-77
- [15] Yakubu Y, Ahmed SS, Audu I. Binary Logistic Regression Methods for Modeling Broncho-Pneumonia Status in Infants from Tertiary Health Institutions in North Central Nigeria IUSMAN, A *J. Appl. Sci. Environ. Manage*. 2019;23(8): 1607-1614.
- [16] Orumie UC, Bartholomew DC, Obite CP, Kiwu LC. Likelihood of Insurance Coverage on Damages Due To Level of Insecurity In Nigeria: Logistic Modeling Approach, *Financial Risk and Management Reviews*. 2021;7(1):50-59.
- [17] Egbo MN, Bartholomew DC. A Discriminant Function Analysis Approach to Country's Economy Status, *Journal of Advanced Statistics*. 2017;2(4):4-9.
- [18] Fisher RA. The Use of Multiple Measurements *annals of Eugenics*. 1936;7:179–188.
- [19] Bartholomew DC, Oyinebifun EB, Arimie CO. Hierarchical Regression Modeling of some Factors Affecting Weight of Child at Birth, *Asian Research Journal of Mathematics*. 2021;17(12):11-27.

- [20] Sule AS. Modeling Broncho-Pneumonia Status in Infants Using Discriminant and Logistic Regression Analysis; 2013.  
Available:<https://eduproject.com.ng/statistics/modeling-bronchopneumonia-status-in-infants-using-discriminant-and-logistic-regression-analysis/index.html>.
- [21] Ting S, Evelyn B, Elizabeth W, Rosalyn S, Zeba AR, Heather JZ, Barbara AR, Shabir AM, Stuart C, Linda CV, Lisa RB, Elizabeth DT, Whitney B, Christian H, Harry C, Harish N. Risk factors for respiratory syncytial virus associated with acute lower respiratory infection in children under five years: Systematic review and meta-analysis, J Glob Health. 2015;5(2):020416.  
DOI: 10.7189/jogh.05.020416.

---

© 2022 Cynthia and Bartholomew; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Peer-review history:**

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<https://www.sdiarticle5.com/review-history/88607>