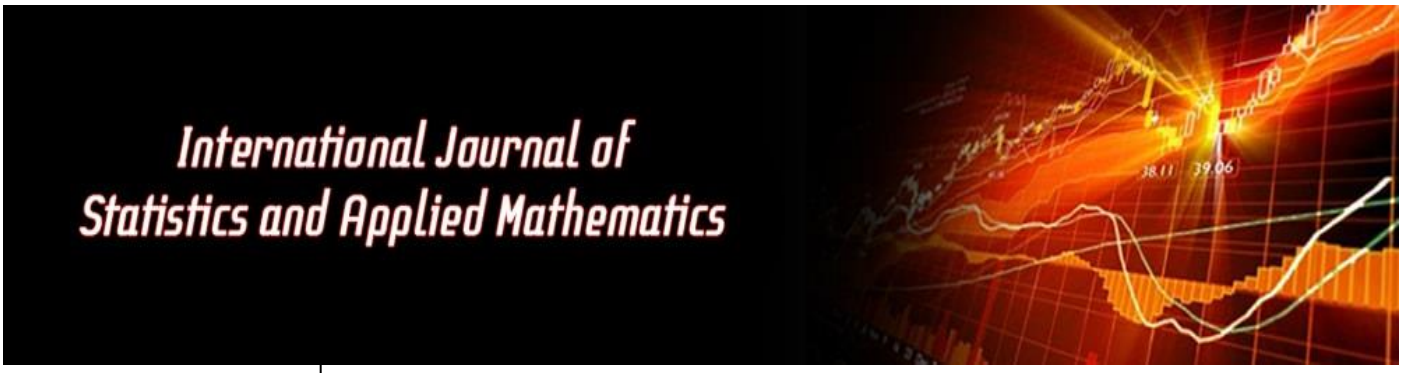
International Journal of Statistics and Applied Mathematics 2022; 7(3): 06-12



ISSN: 2456-1452
Maths 2022; 7(3): 06-12

www.mathsjournal.com
Received: 03-03-2022
Accepted: 07-04-2022

Mouhamed Amine NIANG
Département de Mathématiques, Université Alioune Diop, Bambey, Sénégal

Aba DIOP
Département de Mathématiques, Université Alioune Diop, Bambey, Sénégal

Cheikh Tidiane SECK
Département de Mathématiques, Université Alioune Diop, Bambey, Sénégal

Ibrahima FAYE
Département de Mathématiques, Université Alioune Diop, Bambey, Sénégal

Abdourahmane NDAO
Département de Mathématiques, Université Alioune Diop, Bambey, Sénégal

Idrissa SY
Département de Mathématiques, Université Alioune Diop, Bambey, Sénégal

Corresponding Author:
Mouhamed Amine NIANG
Département de Mathématiques, Université Alioune Diop, Bambey, Sénégal

A negative binomial regression analysis for COVID-19 death cases in Senegal

Mouhamed Amine NIANG, Aba DIOP, Cheikh Tidiane SECK, Ibrahima FAYE, Abdourahmane NDAO and Idrissa SY

DOI: https://doi.org/10.22271/maths.2022.v7.i3a.817

Abstract
We consider the negative binomial regression model (NBRM) to identify explanatory factors of the number of daily COVID-19 death cases in Senegal. The data are obtained from daily cases of COVID-19 following press releases from the Ministry of Health and Social Action (MSAS) during the period from February 24, 2021 to March 3, 2022. The extracted data were analysed using descriptive statistics and NBRM. The results of analysis revealed that the number of severe cases has a significant positive impact on the number of death cases of COVID-19 in Senegal. Furthermore, the number of vaccinated persons has a small and negative effect on the number of deaths.

Keywords: Negative binomial regression model, Regression Model, COVID-19 death cases, severe cases, vaccinated persons, Senegal

1. Introduction
The novel corona virus disease (COVID-19) is certainly the biggest health challenge that mankind has ever faced since the Spanish flu outbreak in 1918. It starts in December 2019 in Wuhan in the Hubei province of China (Wu *et al.* (2020)) [14]. COVID-19 is very contagious and has rapidly spread all around the world. Its transmission from person to person is very quick, because it is very difficult to avoid the contaminations sources such as: droplets, infected surfaces, interaction with people, and so on. COVID-19 is also a deadly disease; a considerable number of deaths (more than 6 millions) has been recorded by World Health Organization's statistics. In Senegal, the number of COVID-19 deaths reached 1964 and the total number of positive cases was 85,880 as of March 30, 2022. The first COVID-19 case in Senegal was reported on March 2, 2020 and the Senegalese government has built a response system following whose recommendations, with the aim of combating the disease. Also a vaccination campaign was launched on February 24, 2021, and is still taking place throughout the whole country. Despite all these government efforts, COVID-19 appears to be progressing with irregular peaks in the number of positives and death cases. In this paper we are interested in the factors influencing the number COVID-19 death cases in Senegal.
There is a quick and abundant literature about modeling the evolution of COVID-19 positive cases and deaths. Daniyal *et al.* (2020) [4] proposed three regression models (linear, logarithmic and quadratic) to predict the trend of COVID-19 deaths cases in Pakistan. They found that the quadratic model was more suitable than the two others for the studied data. Barcellos *et al.* (2021) [1] utilized a data mining approach to predict the evolution of COVID-19 positive cases in three metropolises in Brazilia. They also identified the environmental variables that are likely to be correlated with the COVID-19 confirmed cases and deaths. Khan *et al.* (2021) [11] developped a forecasting model in six countries: Pakistan, India, Afghanistan, Iran, Italy and China with data observed at the earlier of the pandemic; and predicted the number of COVID-19 related cases and deaths by using polynomial regression. Diop *et al.* (2021) [7] used an ARIMA model to study the evolution of COVID-19 positive cases in Senegal, and make short-term predictions for the number of positive cases, from daily data provided by the Senegalese Ministry of Health (MSAS). Motallebi *et al.* (2022) [13] examined the association between face-covering mandates and the COVID-19 mortality decline accross 44 countries in 2

continents, and found that the COVID-19 death rate is lower in countries with face-mask mandates than in countries without mandates. Lounis *et al.* (2021) [12] made a comparison of four models: Gompertz, Logistic, Bertalanffy and inverse artificial neural network (ANNi), with the aim of predicting the number of COVID-19 cases, deaths and recoveries in Algeria. For predicting death cases, they showed that the Gompertz model obtained the best precision with a minimum error of 1.14%. Diagne *et al.* (2021) [6] studied the impact of vaccination and treatment on the COVID-19 evolution and concluded that, despite the negative impact of both vaccination and treatment on the disease, other non-therapeutic measures should continue to be observed. Dengle and Nathan (2020) [5] studied the impact of the increase in COVID-19 confirmed cases on the number of deaths and its implications on food security in North Eastern of Nigeria.

In this paper we use the Negative binomial regression model to analyse factors influencing the COVID-19 related deaths in Senegal in the context of vaccination. The data are extracted from the daily COVID-19 statistics released by the Senegalese Ministry of health and social Action (MSAS), during the period from February 24, 2021 to March 3, 2022.

The paper is organized as follows. The proposed methodology is described in Section 2. More precisely, after describing the data set, we present the negative binomial regression model and the estimation procedure along with some key statistics for the model's goodness-of-fit. Section 3 presents and discusses the results. Finally, Section 4 concludes the paper.

2. Materials and Methods

Modeling count data is a common task in biomedical studies. The classical Poisson regression model is often of limited use in these disciplines because empirical count data sets typically exhibit overdispersion and/or an excess of number of zeros (see Cameron and Trivedi (1998) [1] for more details). The former issue can be addressed by extending the Poisson regression model in various directions. A more formal way is to use a Binomial negative regression model.

2.1 Data description

The database contains 337 observations and 7 variables defined as: Number of death cases (Y), Number of tested cases (X_1), Number of positive cases (X_2), Number of community cases (X_3), Number of vaccinated persons (X_4), Number of severe cases (X_5) and Number of cured cases (X_6). Figure 1 shows the histogram of the dependent variable Y Number of daily COVID-19 death case. Clearly, the data are not in the form of a bell curve like a normal distribution. Figure 2 shows the linearity between the variables. We see a positive correlation between the number of deaths and the number of severe cases and a weak negative correlation between the number of vaccinated people and the number of deaths.

Tables 1 and 2 show respectively the descriptive statistics of the variables and the proportions in relation to the total number of performed tests. We note that the number of death cases is very low compared to the other variables. Thus we use the centered and reduced variables in the model to eliminate the problems of margin and variability.

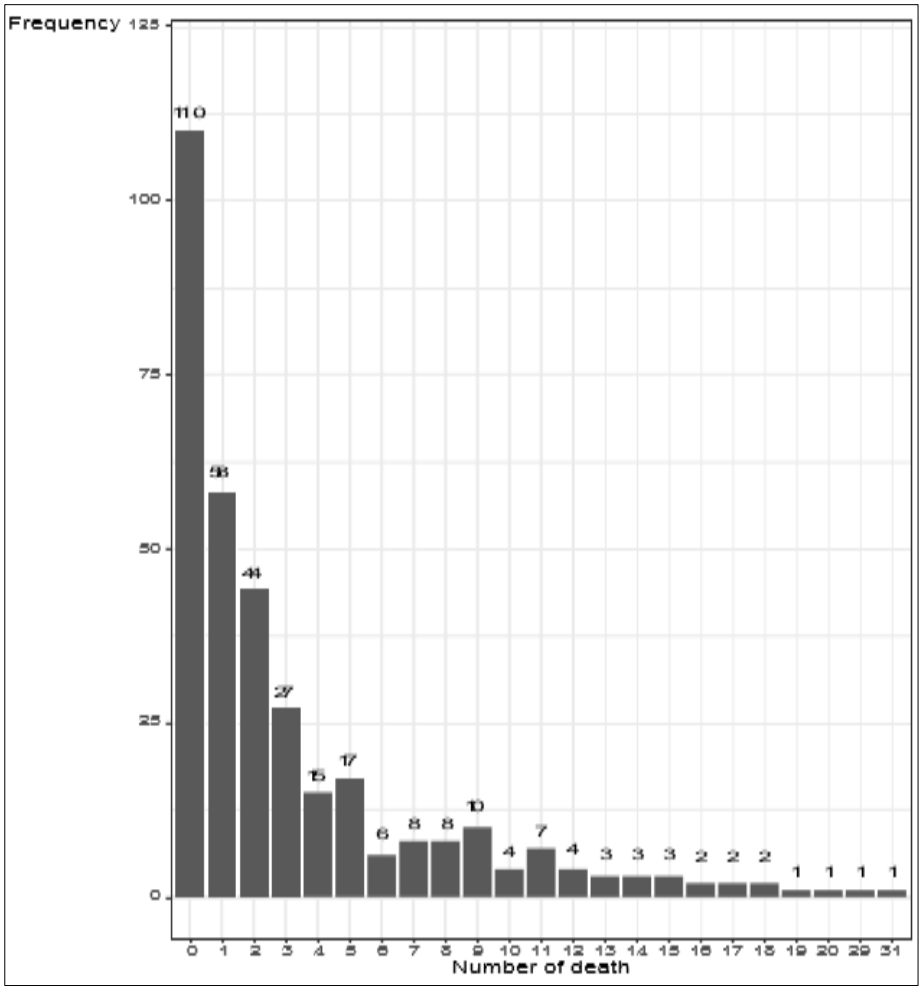


Fig 1: Distribution of daily COVID-19 death cases in Senegal

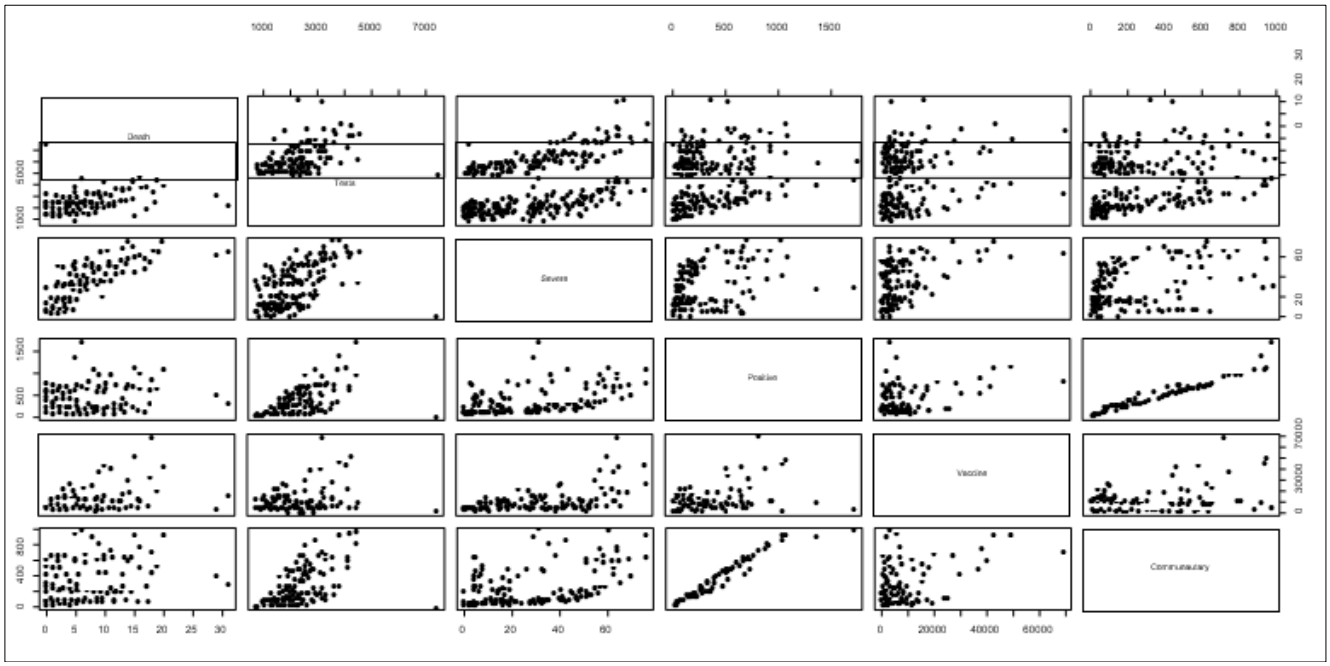


Fig 2: Correlation between COVID-19 death cases and potential factors.

Table 1: Data summary

Variable	Statistical characteristics					
	Min	**1st Qu.**	**Median**	**Mean**	**3rd-Qu.**	**Max**
Number of tested cases	692	1307	1642	1800	2082	7412
Number of positive cases	0	14	52	152.3	174	1722
Number of community cases	0	10	36	122.1	132	978
Number of vaccinated persons	82	1218	2447	4342	3943	69211
Number of severe cases	0	5	11	18.36	30	76
Number of cured cases	0	19	67	157.1	291	639

Table 2: Proportion of the number of cases in relation to the total number of performed tests

Variable	Total Proportion	(%)
Number of positive cases	51312	84.57
Number of community cases	41149	67.82
Number of severe cases	6189	10.20
Number of cured cases	52929	87.23
Number of death cases	1110	1.82

2.2 The Negative binomial regression model

Negative binomial regression is a generalization of Poisson regression which loosens the restrictive assumption that the variance is equal to the mean made by the Poisson model. Let $(Y_1, X_{1j}), \ldots, (Y_n, X_{nj})$ for $j = 1, \ldots, p$, be independent and identically distributed copies of the random vector (Y, X) defined on the probability space $(\Omega, \mathrm{A}, \mathrm{P})$. In negative binomial regression, the mean $\mu = \mathrm{E}(Y|X)$ of the response variable Y is generally determined by the exposure time t and a set of p repressor variables. Using the following equation

$$\log(\mu_i) = \log(t_i) + \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}, i = 1, \ldots, n \quad (1)$$

Remark: Note that the conditional mean μ_i in the case of the negative binomial distribution can be written as$\mu_i = \upsilon t_i$, where the parameter υ is the mean incidence rate of Y per unit of exposure and t_i represent the exposure for a particular observationi. Exposure may be time, space, distance, area, volume, or population size. When no exposure given, it is assumed to be one. The parameter υ may be interpreted as the risk of a new occurrence of the event during a specified exposure periodt.

The regression coefficients $\beta_0, \beta_1, \ldots, \beta_p$ are unknown parameters that have to be estimated from a dataset. Using this notation, the negative binomial regression model for an observation i is written as

$$\mathrm{P}(Y = y_i|\mu_i, \alpha) = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_i+1)} \left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}, i = 1, \ldots, n \quad (2)$$

Where α represents the number of failures before the $(1/\alpha)^{th}$ success (Hilbe (2011) [10]. The regression coefficients $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$ and α are estimated using the method of maximum likelihood. Cameron and Trivedi (1998) [1] gives the logarithm

of the likelihood function for the unknown $(p+1)$-dimensional parameter $\theta = (\beta, \alpha)$ from the independent sample $(y_1, x_1), \ldots, (y_n, x_n)$ is as follows:

$$L(\theta) = \sum_{i=1}^{n} \left[\sum_{j=0}^{y_i - 1} log\left(j + \frac{1}{\alpha}\right) - log\left(\Gamma(y_i + 1)\right) - \left(y_i + \frac{1}{\alpha}\right) log(1 + \alpha\mu_i) + y_i \log(\mu_i) + y_i \log(\alpha) \right] \quad (3)$$

We define the maximum likelihood estimator $\hat{\theta}_n$ of θ as the solution of the $(p+1)$-dimensional score equation

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \quad (4)$$

In general, however, we will use the iteratively-reweighted least squares (IRLS) algorithm (see (Hilbe (2011) [10]) to solve equation (4).

Cameron and Trivedi (1998) [1] gives the asymptotic distribution of the maximum likelihood estimates as follows: $\left(\hat{\theta}_n - \theta\right) \rightsquigarrow N(0, \Sigma), as\ n \to +\infty$, where the elements of the estimated variance-covariance matrix $\hat{\Sigma}$ of Σ given as follows:

$$\widehat{Var}\left(\hat{\beta}_n\right) = \left[\sum_{i=1}^{n} \frac{\hat{\mu}_i}{1 + \hat{\alpha}\hat{\mu}_i} x_i x_i' \right]^{-1}, Cov\left(\hat{\beta}_n, \hat{\alpha}_n\right) = 0, by\ estimators\ independence$$

$$\widehat{Var}(\hat{\alpha}_n) = \sum_{i=1}^{n} \left[\hat{\alpha}^{-4} \left(\log(1 + \hat{\alpha}\hat{\mu}_i) - \sum_{j=0}^{y_i - 1} \frac{1}{j + \hat{\alpha}^{-1}} \right)^2 + \frac{\hat{\mu}_i}{\hat{\alpha}^2 (1 + \hat{\alpha}\hat{\mu}_i)} \right]^{-1}$$

2.3 Goodness-of-fit criteria

Deviance: we have

$$D = 2 \sum_{i=1}^{n} \left[y_i log\left(\frac{y_i}{\mu_i}\right) - \left(y_i + \frac{1}{\alpha}\right) log\left(\frac{1 + \alpha y_i}{1 + \alpha\mu_i}\right) \right]$$

For large samples the distribution of the deviance is approximately a chi-squared with $n - p$ degrees of freedom, where n is the number of observations and p the number of parameters. Thus, the deviance can be used directly to test the goodness of fit of the model. The model fits well the data if the ratio between the deviance and its number of degrees of freedom is close to 1. In this case, we also say that the model handles well the over dispersion problem.

Pearson Residuals: The Pearson residual corrects for the unequal variance in the residuals by dividing by the standard deviation of y. The formula for the Pearson residual is

$$P = \sum_{i=1}^{n} \left[\frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + \alpha\hat{\mu}_i^2}} \right]$$

In large samples the distribution of Pearson's statistic is also approximately chi-squared with $n - p$ degrees of freedom. The model fits well the data (Hypothesis H_0) at the risk threshold $\alpha \in]0;1[$ if the p-value of the test $\left(p - value = \mathrm{P}\left(\chi_{n-p} > \chi_{obs} | H_0\right)\right)$ is above the threshold α.

Mean Absolute Error (MAE): MAE is a very good key performance indicator to measure forecast accuracy. As the name implies, it is the mean of the absolute error $\hat{\epsilon}_i = y_i - \hat{y}_i$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{\epsilon}_i|$$

3. Results and Discussion

In this section, we deal with the data set of daily reported COVID-19 cases in Senegal during the period from February 24, 2021 to February 3, 2022 (data from the Ministry of Health and Social Action of Senegal). In this study, we aim to establish a relationship between the number of daily COVID-19 death cases and the vector of explanatory variables $(X_1, \ldots, X_6)$. We first run the Poisson regression model:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6}, i = 1, \ldots, n \quad (6)$$

Note that in model (6), the exposure time t_i is the same for all patients and is assumed to be 1($\log(t_i)$=0, for i=1,…,n). The results of model (8) are presented in the Table 3.

Table 3: Poisson regression model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	0.6795	0.0434	15.64	0.0000
Tests	-0.0158	0.0459	-0.34	0.7309
Positive cases	0.2053	0.1345	1.53	0.1269
Community cases	-0.1972	0.1404	-1.40	0.1603
Severe cases	0.9336	0.0389	23.98	0.0000
Vaccinated persons	-0.0521	0.0234	-2.23	0.0261

We perform the Cameron-Trivedi dispersion test (Cameron and Trivedi (1998)) [5, 1] for the model (8) to study the overdispersion property. The null hypothesis of this test is the absence of overdispersion. The computed p-value of this test is equal to 0.00000173, which leads to the presence of overdispersion at 5% significance level.
We thus propose to use the negative binomial regression model to overcome this problem of overdispersion. Now, we run the following negative binomial regression model:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_6 X_{i6}, i = 1, \ldots, n \quad (7)$$

Table 4 shows the first results obtained from the negative binomial model (7).

Table 4: Negative binomial regression model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.6390	0.0497	12.86	0.0000
Tests	-0.0271	0.0631	-0.43	0.6673
Positive cases	0.1814	0.1761	1.03	0.3029
Community cases	-0.1365	0.1852	-0.74	0.4612
Severe cases	0.9994	0.0520	19.22	0.0000
Vaccinated persons	-0.0722	0.0359	-2.01	0.0446

Then, we use the backward regression to select the best model. Finally we obtain the following model presented in Table 5.

Table 5: Final negative binomial regression model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.6431	0.0493	13.05	0.0000
Severe cases	0.9983	0.0410	24.34	0.0000
Vaccinated persons	-0.0668	0.0326	-2.05	0.0402

The adjusted equation of the model is given as follows

$$\log(\mu) = 0.6431 + 0.9983 \times Severe.Cases - 0.0668 \times Vaccinated.Persons$$

We can see from the results that the ratio of the residual deviance (377.02) to its degree of freedom (334) is almost 1.128, which suggests that the negative binomial model has taken account of the overdispersion problem. The p-value of the Pearson goodness-of-fit test on the residuals is equal to 0.6620. This means that the final model is well adapted to the data. The Pseudo R^2 of Cragg and Uhler (see Cragg and Uhler (1970)) [3] is equal to 0.67 and the mean absolute percentage error is equal to 1.08%. These values indicate a good predictive capacity of the model in Table 5.
Figure 3 shows that the final negative binomial regression model almost overlap with the observed data, which is a good performance.

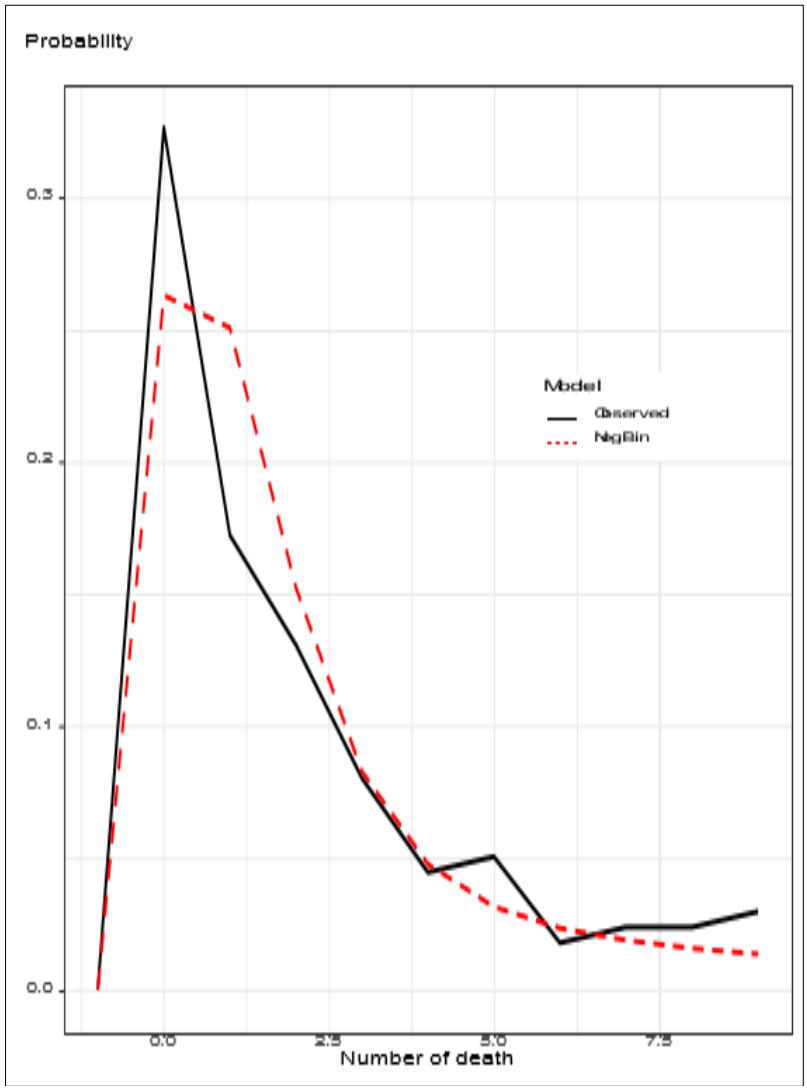


Fig 3: Predicted probabilities vs Observed

Table 6: Odd-Ratios (OR) results

Variable	OR	[2.5% − 97.5%]		Pr (>|z|)
(Intercept)	1.90	[1.73 -	2.10]	0.0000
Severe cases	2.71	[2.50 -	2.94]	0.0000
Vaccinated persons	0.94	[0.88	- 1.0]	0.0402

Table 6 shows the odd-ratios of the variables ”Number of severe cases” and ”Number of vaccinated persons” that have a significant impact on the number of deaths due to COVID-19. We see that the number of severe cases is a risk factor because it has a positive impact (OR = 2.71 with a confidence interval of [2.50−2.94]) on the number of death cases. Then a one unit change in the number of severe cases increases the number of death cases by 0.9983 units. In contrast, the number of vaccinated persons has a negative impact on the number of death cases, and is therefore a protective factor (OR = 0.94 with a confidence interval of [0.88 − 1.0]). Then a one unit change in the number of vaccinated persons decreases the number of death cases by 0.0668 units.

4. Conclusion

In this work, we used a negative binomial regression model to study the impact of certain factors on the number of daily deaths from COVID-19 in Senegal. The statistical analysis revealed that the number of severe cases of COVID-19 and the number of vaccinated people have a significant impact on the number of deaths. Increasing the number of severe cases increases the number of death cases and increasing the number of vaccinated persons decreases the number of deaths. However, the number of severe cases is more significant than the number of vaccinated people.

Therefore, public health policy makers need to support health workers to manage positive cases earlier so that their illnesses do not progress to severe cases, and to encourage people to go for vaccination to limit the spread of the virus.

However, this study do not take into account the environmental and socio-demographic variables such as temperature, age and co-morbidity of patients, health care capacity.

Based on the goodness of fit of the model, we may envisage in the future to study the predictive quality of the model. Decision-makers could thus use this model to better understand the number of deaths from COVID-19 in Senegal.

5. References

1. Barcellos DD, Fernandes GMK, de Souza FT. Data based model for predicting COVID-19 morbidity and mortality

in metropolis. Sci Rep. 2021,29;11(1):24491. DOI: 10.1038/s41598-021-04029-6.
2. Cameron AC, Trivedi PK. Regression Analysis of Count Data. Cambridge University Press, Cambridge; c1998.
3. Cragg SG, Uhler R. The demand for automobiles. Canadian Journal of Economics. 1970 Aug 1;3(3):386-406.
4. Daniyal M, Ogundokun RO, Abid K, Khan MD, Ogundokun OE. Predictive modeling of COVID-19 death cases in Pakistan. Infect Dis Model. 2020;5:897-904. DOI: 10.1016/j.idm.2020.10.011
5. Dengle GY, Nathan N. A Poisson Regression Analysis of COVID-19 Pandemic: Implication on Food Security in North Eastern Nigeria; c2020. Available at SSRN: https://ssrn.com/abstract=3690886 or http://dx.doi.org/10.2139/ssrn.3690886.
6. Diagne ML, Rwezaura H, Tchoumi SY, Tchuenche JM. A Mathematical Model of COVID-19 with Vaccination and Treatment. Computational and Mathematical Methods in Medicine. 2021;2021:16p. Article ID 1250129. https://doi.org/10.1155/2021/1250129
7. Diop A, Ndao A, Seck CT, Faye I, Niang MA. Short-term prediction model for daily COVID-19 reported positive cases in Senegal. Afr. J Appl. Stat. 2021;8(1):111-1126. DOI: 10.16929/ajas/2021.1111.260.
8. Forsido SF, Abddisa FM, Ejeta TT, Lemessa F. COVID-19 Probable Impacts on Ethiopian Agriculture and Potential Mitigation and Adaptation Measures: No Food-No Health-No Life; c2020. DOI:10.13140/RG.2.2.29988.42888. Project: food supply chain.
9. Giroh YD, Nachandiya N. A Poisson Regression Analysis of COVID-19 Pandemic: Implication on Food Security in North Eastern Nigeria; p 2020. http://dx.doi.org/10.2139/ssrn.3690886.
10. Hilbe J. Negative Binomial Regression, 2nd ed., New York: Cambridge University Press; c2011.
11. Khan N, Arshad A, Azam M, Al-marshadi AA, Aslam M. Modeling and forecasting the total number of cases and deaths due to pandemic. Journal of Med. 2022 Apr;94(4):1592-605. Vir. https://doi.org/10.1002/jmv.27506.
12. Lounis M, Torrealba-Rodriguez O, Conde-Gutierrez RA. Predictive models for COVID-19 cases, deaths and recoveries in Algeria. Results in Physics. 2021;30:104845. https://doi.org/10.1016/j.rinp.2021.104845.
13. Motallebi S, Cheung RCY, Mohit B, Shahabi S, Tabriz AA, Moattari S. Modeling COVID-19 Mortality Across 44 Countries: Face Covering May Reduce Deaths. Am J Prev Med. 2022;62(4):483-491. doi:10.1016/j.amepre.2021.09.019.
14. Wu YC, Chen CS, Chan YJ. The outbreak of COVID-19: An overview. Journal of the Chinese Medical Association. 2020;83(3):217-220.