ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 15, Issue, 08, pp.25486-25490, August, 2023
DOI: https://doi.org/10.24941/ijcr.45714.08.2023

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

MODELING THE EFFECT OF INTERVENTION ON COVID-19 PREVALENCE IN KENYA USING LOGISTIC REGRESSION

***Odero Everlyne Akoth**

Masinde Muliro University of Science and Technology, Kenya

ARTICLE INFO

Article History:
Received 18th May, 2023
Received in revised form
08th June, 2023
Accepted 24th July, 2023
Published online 28th August, 2023

Key words:

Intervention, COVID-19 and Prevalence.

**Corresponding Author:*
Odero Everlyne Akoth

ABSTRACT

The COVID-19 pandemic has caused great disruption worldwide. In connection to this, over 43 million people have confirmed diagnosis of the disease, and over 1 million people have died from it. Though it is well documented that a number of factors have been linked to the reduction of the risk of COVID-19 infection ranging from social distancing, mask wearing and washing of hands, Intervention could fight against the spread of the COVID-19 virus. In this paper, we used maximum likelihood estimation (MLE) method, to fit logistic regression model; first to data set without intervention then to data set with effect of intervention incorporated. This was done with an aim of determining the effect of intervention on COVID-19prevalence in Kenya. Model parameters were estimated by MLE method. AKaike's information criteria (AIC) was used to compare the two models fitted to data so as to determine the better performing model. The study used data sets of 2019-2020(data obtained before vaccination intervention set in) and 2021-2023(data obtained after incorporation of vaccination intervention effect)COVID-19 data sets for Kenya (from the Centre for Disease Control and Prevention (CDC)).The model fitted to the data set which contained the effect of Vaccination intervention gave a lower AIC value of 1293 compared to the one fitted to the data set which did not contain the effect of Vaccination intervention which gave AIC value of 1389. The logistic regression model fitted to data set containing the effect of vaccination intervention therefore had a better performance as indicated by the AIC value. The results of this study revealed a high prevalence of COVID-19 in the data set which had no effect of vaccination intervention. Whereas, a low prevalence of COVID – 19 was realized with data set which had effect of vaccination intervention incorporated. This indicates that the existence of vaccination in Kenya has played a significant role in reducing the prevalence of COVID-19 disease. The results of this study would be of much benefit to the health sector in monitoring and sensitizing people on the effect of Intervention on COVID-19 prevalence.

Citation: ***Odero Everlyne Akoth.*** **2023.** "Modeling the effect of intervention on covid-19 prevalence in Kenya using logistic regression."". *International Journal of Current Research*, 15, (08), 25486-25490.

INTRODUCTION

The human coronavirus associated with severe acute respiratory syndrome was first identified in December 2019, in Wuhan, China (Fatima *et al*. 2021). Thereafter, coronavirus (COVID-19) infection spread globally and by February 2021, more than 106 million cases of COVID-19 had been confirmed and almost 2.5 million deaths had been reported (Kousi *et al*. 2021). The alarming rate of spread of the disease caused the World Health Organization to declare it a pandemic on 11 March 2020 (Kousi *et al*. 2021 and Cohen et.al 2020).Thereafter, the worldwide scientific community invested great efforts to find possible solutions to stop or to at least, mitigate the COVID-19 outbreak. The complex spreading patterns of COVID-19 and the various spread speed of its variants made its containment and mitigation a real challenge. In connection to this, several mathematical models have been developed in order to exhibit key epidemiological features, investigate transmission dynamics, and develop adequate control policies. Damiloba *et al*. (2021) developed a deterministic and Stochastic model for epidemic dynamics of COVID-19 in Wuhan China. They observed that both the exposed and infected classes play an important role in shaping the epidemic dynamics of COVID-19 in China. Rastko *et al.* (2021) observed that vaccination led to an increased number of infected cases and a higher reduction in mortality. Hence, simple models that contain the essential components and interactions are powerful tools to test different hypotheses and understand disease control for both short and long term effects (Algarni *et al*. 2022). The existing models vary in form and complexity, but the common objective is to provide important information for global health decision makers about the disease dynamics, Checks on the Prevalence rates of the disease is overlooked into. The first control measure was lockdown and then health authorities imposed mask wearing, social distancing and washing of hands (Algarni *et al*. 2022). Khajanchi *et al*. (2021) even observed that social distance is an important factor to reduce the CORONA virus spread. To enhance individual protection, it is critical for people to be fully vaccinated against COVID-19. Many Researchers have developed several vaccinations, and the World Health Organization (WHO) has urged people to get vaccinated to mitigate this pandemic.

Our Question in this study is whether the Vaccination as an intervention mechanism to mitigate the spread of COVID-19 has had any significant effect in reducing prevalence rate of the disease in Kenya. In combination with preventive measures, intervention by way of vaccination could be beneficial for containing the COVID-19 pandemic.

Intervention Analysis: Intervention as described by Hardnett *et al.* (Hardnett, F. P *et al*. 2009), refers to a mechanism by which an independent variable causes change in a dependent variable. The independent variable causes change in the Intervention variable, which in turn cause change in the dependent variable, hence the effect of an independent variable is at least partially transmitted through an intervention variable to the dependent variable. Intervention analysis helps researchers to investigate how the intermediate variable, (V) influences the relationship between the independent variable, (X) and the dependent variable, (Y). Hayes et. al (Hayes, A. F. et.al,2009), describes total effect as the expected effect of a change in independent variable on dependent variable. Total effect can be decomposed into direct and indirect due to the effect of intervention variable, (Huberman *et al*. 2020). This study used partial intervention which maintains that the intervention variable accounts for some, but not all, of the relationship between the independent variable and dependent variable. This implies that there is both an indirect relationship between independent variable and the dependent variable trough an intervention and also a direct relationship between the independent and dependent variable.

Objectives of the study

Main Objective: To model the effect of vaccination on covid-19 prevalence in Kenya using logistic regression.

Specific Objective

- To fit logistic regression model to data set without intervention effect and determine its parameter estimates
- To fit logistic regression model to data set containing intervention effect and determine its parameter estimates
- To determine the effect of intervention on COVID_19 prevalence in Kenya.

Methods of study

Logistic regression: Logistic regression analysis studies the association between a categorical dependent variable and a set of independent variables. Logistic regression is suited for analyzing dichotomous outcome variables. We use the logistic regression equation to predict the probability of a dependent variable taking the dichotomy values 0 and 1. In logistic regression model, the log odds of the outcome is modeled as a linear combination of the predictor variables. The typical use of this model is predicting dependent variable given a set of predictors. The predictors can be continuous, categorical or a mix of both. to deal with It does not require many of the principle assumptions of linear regression models that are based on ordinary least squares method particularly regarding linearity of relationship between the dependent and independent variables, normality of the error distribution, homoscedasticity of the errors, and measurement level of the independent variables. Logistic regression can handle non-linear relationships between the dependent and independent variables, because it applies a non-linear log transformation of the linear regression (Park, 2013).

Maximum Likelihood Estimation: The Likelihood of a set of data refers to the probability of obtaining that particular set of data given the chosen probability model. Maximum likelihood thus begins with the mathematical expression called Likelihood function of the sample data, the expression which contains the unknown parameters. The values of the unknown parameter that maximizes the sample likelihood are the maximum likelihood estimates. This method aims and selecting the parameters θ that make the observed data the most likely or rather obtaining the parameters that maximizes the likelihood function. The approach uses differential calculus to determine the maximum of the likelihood function in order to determine the parameter estimates. Parameter estimation in MLE is done by partially differentiating the log of the likelihood function and equating the results to zero.

Akaike's Information Criterion (AIC)

The Akaike's Information Criterion (AIC), is defined by;

$$AIC = 2K - 2\log L(\beta)$$

Where;

K represents the number of parameters in the model and $L(\beta)$ denotes the maximum value of the likelihood function in the model (Sakamoto,1986). This approach refers to an estimator of prediction error and thereby relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. These statistical models represents the process that generated the data, which in so doing some information is lost in the process hence the representation can never be the same. The AIC approach was therefore used in this study to estimate the relative amount of information lost by the model with Intervention and by the model without intervention. The less information a model loses the higher the quality of the model, Chaba, (2011).

Fitting of Logistic Regression Model: Two data sets are used here; the 2019-2020 Kenya COVID-19 data set believed not to contain vaccination effect and then the 2021-2023 Kenya COVID-19 data set believed to contain vaccination effect. Logistic regression model is fitted to both the data sets with an aim of determining the effect of vaccination intervention on the COVID_19 prevalence.

Model variables

The dependent variable used in this study is COVID-19 prevalence (P).
The Intervention variable used is "vaccination"- whether vaccinated or not (Yes-1; No-0).
The independent variables are; social distancing (S), mask wearing (M) and hand washing (H)

Logistic Regression: The idea behind logistic regression modeling is that the response variable is the logarithm of the odds. Logistic regression therefore calculates the probability of an event occurring over the probability of an event not occurring.

The impact of independent variables is usually explained in terms of odds, which refers to the ratio of the probability that an event will occur to the probability that it will not occur.

$$odds(of\ event) = \frac{P}{1-P}$$

Binary Logistic Regression: This is a generalized linear model used in analyzing the relationship between the response variable that is binary in nature and a set of explanatory variables that are either continuous, categorical or a mix of both. In the current study, we considered the response variables as "the final result of COVID-19 test of an individual', which was either positive or negative, taking the values 1 and 0 respectively. According to McCullagh, (2019). Generalized linear model (GLM) is a lager class of models in which the response variable is assumed to follow an exponential family distribution with mean μi. This model has three components; Random component that specifies the probability distribution of the response variable E(Y) = P, the systematic component that specifies explanatory variables in the model and the link function g(μ) that specifies the link between the random and systematic components as $g(\mu) = \log it\ P = \log\left(\frac{P}{1-P}\right)$ for logistic regression.

In this paper we considered a logistic regression model for the binary response variable.

Given that P is the COVID-19 status of an individual .Y = 1 if individual is COVID-19 positive and Y = 0 if otherwise.

Let Xi = ($x_1, \cdots, x_n$) be a vector of all independent random variables.

The COVID-19 status of the individual is therefore a binary outcome in this study and hence follows a Bernoulli distribution Y∼ Ber(P) where P is the probability that individual is COVID-19 positive and 1 − P is the probability that individual is COVID-19 negative.
With logistic regression the mean of the response variable P in terms of an explanatory variable X is modeled by relating P and X through the equation

$$P = \beta_0 + \beta_r X'_i \qquad (1)$$

The extreme values of X_i may give values of $\beta_0 + \beta_r X'_i$ which does not fall between 0 and 1. Peng, et.al,(2002) provides a solution to this problem by transforming the odds using the natural logarithm; the natural log odds are modeled as linear function of the explanatory variables as follows;

$$\log it\ Y = \ln(odds) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_r X'_i \text{ r=0,1,... i=1,2,..} \qquad (2)$$

Extending the logic of the simple logistic regression to multiple predictors, a complex logistic regression model is formulated as

$$\log it\ P = \beta_0 + \beta_1 X_1 + ... + \beta_R X_R \qquad (3)$$

The Predicted Probabilities expressed as a natural logarithm (ln) of the odds ratio is given as follows

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + ... + \beta_R X_R \qquad (7)$$

The model parameters to be estimated include ($\beta_0, \beta_1, \cdots, \beta_R$)

Fitting logistic regression model to data set without intervention effect and parameter estimates

Letting; P represent COVID-19 Prevalence factor, S represent Social distancing factor,

M represent Mask wearing distancing factor and H represent Hand washing factor, Logistic regression model was then fitted as follows;

$$\log\left(\frac{p}{1-p}\right) = 0.06626 + 0.03236S + 0.05170M + 0.10731H$$

Table 1 Parameter estimates of the fitted Regression model to data set without Intervention

Coefficients			
Intercept	S	M	H
0.06626	0.03236	0.05170	0.10731
AIC: 1389			

From the Table 1;

From the Table 1, there is Positive correlation between social distancing factor and COVID-19 prevalence, that is, a unit increase in social distancing factor increases the log odds of COVID-19 prevalence by 0.03236. There is Positive correlation between Mask wearing factor and COVID-19 prevalence, that is, a unit increase in Mask wearing factor increases the log odds of COVID-19 prevalence by 0.05170.

There is Positive correlation between Hand washing factor and COVID-19 prevalence, that is, a unit increase in Hand washing factor increases the log odds of COVID-19 prevalence by 0.10731.

The value of AIC obtained while fitting the model was 1389

Fitting logistic regression model to data set containing intervention effect and parameter estimates
Similarly; P represent COVID-19 Prevalence factor, S represent Social distancing factor, M represent Mask wearing distancing factor and H represent Hand washing factor, Logistic regression model was then fitted as follows;

$$\log\left(\frac{p}{1-p}\right) = 0.02035 - 0.05596S - 0.02176M - 0.02592H$$

Table 2. Parameter estimates of the fitted Regression model to data set without Intervention

Coefficients			
Intercept	S	M	H
0.02035	-0.05596	-0.02176	-0.02592
AIC: 1293			

From the Table 2;

There is Negative correlation between social distancing factor and COVID-19 prevalence, that is, a unit increase in social distancing factor decreases the log odds of COVID-19 prevalence by 0.05596. There is Negative correlation between Mask wearing factor and COVID-19 prevalence, that is, a unit increase in Mask wearing factor decreases the log odds of COVID-19 prevalence by 0.02176. There is Negative correlation between Hand washing factor and COVID-19 prevalence, that is, a unit increase in Hand washing factor decreases the log odds of COVID-19 prevalence by 0.02592. The value of AIC obtained while fitting the model was 1293. It is worth noting that all the three parameters have posted a negative correlation with log odds of Covid–19 prevalence.

Effect of intervention on covid–19 prevalence: Determination of the effect of Intervention on COVID_19 prevalence was done by comparing the two fitted Logistic regression models: one to data set before vaccination intervention was incorporated and the other, is the data set after incorporation of vaccination Intervention. From Table 3, the Logistic regression model fitted in the data set with intervention effect had AIC value of 1293 which was lower than the AIC value of 1389 obtained as a result of Logistic regression model being fitted in the data set without intervention effect . This showed that the amount of information lost in fitting the model with intervention was less compared to amount of information lost in fitting the model without intervention, hence the intervention has effect of reducing prevalence of Covid-19 prevalence.

Table 3. Comparison of Performance of fitted Logistic regression model to data without intervention and to data with intervention effect

	Parameter estimates	AIC Value
Data set without intervention effect	0.03236	1389
	0.05170	
	0.10731	
Data set without intervention effect	-0.05596	1293
	-0.02176	
	-0.02592	

CONCLUSION

The main objective of the study was to model the effect of vaccination on covid-19 prevalence in Kenya using logistic regression. To achieve this, two data sets were used; Kenya covid-19 dataset of 2019-2020 which were collected before vaccination intervention set in and also the Kenya covid-19 dataset of 2021-2023 which were collected after vaccination intervention had set in. In both cases, Logistic regression model was fitted. The model fitted to the data set which contained the effect of Vaccination intervention gave a lower AIC value of 1293 compared to the one fitted to the data set which did not contain the effect of Vaccination intervention which gave AIC value of 1389. The logistic regression model fitted to data set containing the effect of vaccination intervention therefore had a better performance as indicated by the AIC value.
The results of the study revealed a high prevalence of COVID-19 in the data set which had no effect of vaccination intervention. Whereas, a low prevalence of COVID – 19 was realized with data set which had effect of vaccination intervention incorporated. This indicates that the existence of vaccination in Kenya has played a significant role in reducing the prevalence of COVID-19 disease.

REFERENCES

Algarni AD, Ben Hamed A, Hamdi M, Elmannai H, Meshoul S. 2022. Mathematical COVID-19 model with vaccination: a case study in Saudi Arabia. Peer J Comput. Sci. 8:e959 http://doi.org/10.7717/peerj-cs.959

Chaba, L. A. 2011. Modeling of sti prevalence among hiv-infected adults in hiv care programs in Kenya using logistic regression (Doctoral dissertation).

Cohen, J., Normile, D. New SARS-like virus in China triggers alarm. Science 2020, 367, 234–235. Maleki, M., Mahmoudi, M. R., Wraith, D., & Pho, K. H. 2020. Time series modelling to forecast the confirmed and recovered cases of COVID-19. Travel medicine and infectious disease, 37, 101742.

Damiloba, O. et.al 2021. Deterministic and Stochastic models for the epidemic dynamics of COVOD – 19 in Wuhan, China.Mathematical Biosciences and Engineering Journal.18(1):950-967

Fatima, M., O'Keefe, K.J., Wei, W., Arshad, S., Gruebner, O. Geospatial Analysis of COVID-19: A Scoping Review. Int. J. Environ. Res. Public Health 2021, 18, 2336.

Hardnett, F. P., Pals, S. L., Borkowf, C. B., Parsons, J., Gomez, C., & O'Leary, A. 2009. Assessing HIVintervention studies. Public Health Reports, 124 (2), 288-294.

Hayes, A. F. et.al 2009. Beyond Baron and Kenny: Statistical analysis in the new millennium. Communication monographs, 76 (4), 408-420. 52

Huberman, D. B., *et al*. 2020. Estimating the drivers of species distributions with opportunistic data. Ecosphere, 11 (6), e03165.

Khajanchi S, Sarkar K, Banerjee S. 2022. Modeling the dynamics of COVID-19 pandemic with implementation of intervention strategies. The European Physical Journal Plus 137(1):129

Kousi, T., Mitsi, L.-C., Simos, J. The Early Stage of COVID-19 Outbreak in Greece: A Review of the National Response and the Socioeconomic Impact. Int. J. Environ. Res. Public Health 2021, 18, 322.

McCullagh, P. 2019. Generalized linear models. Routledge

Park, H. A. 2013. An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. Journal of Korean Academy of Nursing, 43(2), 154-164.

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. 2002. An introduction to logistic regression analysis and reporting. The journal of educational research, 96(1), 3-14.

Rastko J. *et al*. 2021. Modelling Voluntary General Population Vaccination Strategies during COVID-19 Outbreak. International Journal of Environmental Research and Public Health .6217:1-18

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. 1986. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel, 81(10.5555), 26853.

****