

Original Article

Count Data Regression Modelling: An Application to Monkeypox Confirmed Cases

Divya Vijithaswan Nair<sup>1\*</sup>, Rujuta Hadaye<sup>2</sup>

<sup>1</sup>Department of Community Medicine, Assistant Professor in Biostatistics, Topiwala National Medical College & B.Y.L Nair Charitable Hospital, Mumbai Central, Maharashtra University of Health Science (Nashik), India.

<sup>2</sup>Department of Community Medicine, Professor and Head, Topiwala National Medical College & B.Y.L Nair Charitable Hospital, Mumbai Central, Maharashtra University of Health Science (Nashik), India.

ARTICLE INFO

ABSTRACT

Received 03.02.2023  
Revised 21.02.2023  
Accepted 17.03.2023  
Published 15.06.2023

**Key words:**

Negative binomial regression model;  
Poisson regression model;  
Monkeypox;  
Akaike information criterion (AIC);  
Bayesian information criterion (BIC).

**Introduction:** With the presence of COVID 19, some countries also faced an increase in number of cases due to Monkeypox virus.

The main aim of this research was to investigate whether it is possible to fit count data regression models to predict the daily incidence of Monkeypox confirmed cases.

**Methods:** In this study we have used two types of traditional count regression models like Poisson regression model and Negative binomial regression model using identity and logarithmic link function. Since our data was overdispersed, Negative binomial regression model with logarithmic link function fitted well as compared to other models. The parameters were estimated using SPSS, version 23.0.

**Results:** The Negative Binomial Regression model with logarithm function fits well to the data related to Monkeypox cases. Therefore, the model shows that majority of the countries like Brazil, Canada, France, Germany, Peru, Spain, United Kingdom and United States of America shows significant decrease in number of cases with respect to time. The prediction line was plotted using this model where the line predicts well about the daily Monkeypox cases reported by different countries.

**Conclusion:** From our study, we concluded that the count data regression model can be used widely to predict the incidence of any disease. The countries like Canada and Brazil have largest and smallest slope coefficient which shows maximum and minimum decrease in expected number of cases confirmed each day respectively.

**Introduction**

In past two decades, many countries faced the outbreak of viral diseases like Ebola, Chikungunya, H1N1, Bird flu, Nipah, Zika, SARS, MERS and COVID 19. With the

existence of COVID 19, in May 2022 an unprecedented case of Monkeypox virus has been reported in non-endemic countries.<sup>1</sup> Monkeypox is a viral zoonosis which is a virus that gets transmitted through either close contact with any infected person or any

\*.Corresponding Author: [divya.nair421@gmail.com](mailto:divya.nair421@gmail.com)



other materials which is contaminated with the virus. It is a member of Orthopoxvirus genus of the Poxviridae family. The symptoms of Monkeypox virus is similar to smallpox eradicated in 1980. However, Monkeypox is less infectious as compared to smallpox.<sup>2</sup> The incubation period of Monkeypox virus is usually from 6 to 13 days but can be extended upto 21 days. The invasion period which lasts between 0 to 5 days is distinguished by fever, muscle pain, severe headache, back pain, lack of energy and lymphadenopathy. Lymphadenopathy is a distinct characteristic of Monkeypox virus. Neither Chickenpox nor smallpox has such kind of characteristic. The skin eruption due to infection usually begins within 1 to 3 days of appearance of fever and it affects face more rapidly than other areas.

The first Human Monkeypox virus was reported in the year 1970, where maximum cases were found around Central Africa followed by the next significant Monkeypox case found in US in the year 2003 and was epidemiologically linked with imported exotic pets from Ghana who spread the disease to pet prairie dogs and from them it got transmitted to human beings.<sup>3</sup> In recent years, the number of cases reported by many countries has increased rapidly. The number of cases which increased throughout the countries also faced many deaths due to this infection, indicating global public health threat due to Monkeypox infection.<sup>4</sup> Many key characteristics of this newly reported Monkeypox remained unknown due to the new mode of transmission.<sup>5</sup>

The WHO claimed that Monkeypox outbreak was public health emergency of international concern and it was the highest level of public health alert issued by WHO on 23<sup>rd</sup> July, 2022.<sup>2</sup> The global effects of Monkeypox has led to

many countries like Brazil, Canada, Colombia, France, Germany, Mexico, Peru, Spain, United Kingdom and United States of America. Together, these countries reported 84.9% of Monkeypox cases. The case fatality ratio was around 3% – 6% in the general population and been higher among young children.<sup>6</sup>

When we deal with the number of newly confirmed cases related to Monkeypox virus which implies that the confirmed cases are nothing but count variable. Many studies used to convert the count variable in terms of binary variable and the standard logistic regression models were applied to analyze the entire data but such standard models will produce parameter estimates that are biased. However, these standard models make continuous predictions even though the outcomes are discrete. Most of the educational research or health research concentrates on count variable which indicates it is the number of times a particular event observed.<sup>7</sup> Count variable also had certain properties like the values are always positive and integers, the minimum value can be zero and they frequently appear to be positively skewed.

Many studies concentrate on in vitro and in vivo models for Monkeypox virus<sup>3</sup> and also Artificial Intelligence have been used for the identification of disease clusters, forecasting the disease future outbreak and to develop follow up strategies.<sup>1</sup> From the literature it was evident that majority of the analyses concentrates on Clinical aspects of Monkeypox virus and very few concentrates on predicting the number of cases of Monkeypox virus.<sup>2</sup> A far and wide range of complicated Statistical models have already been applied to the analysis of Monkeypox cases. Those models are like SEIR epidemic models,

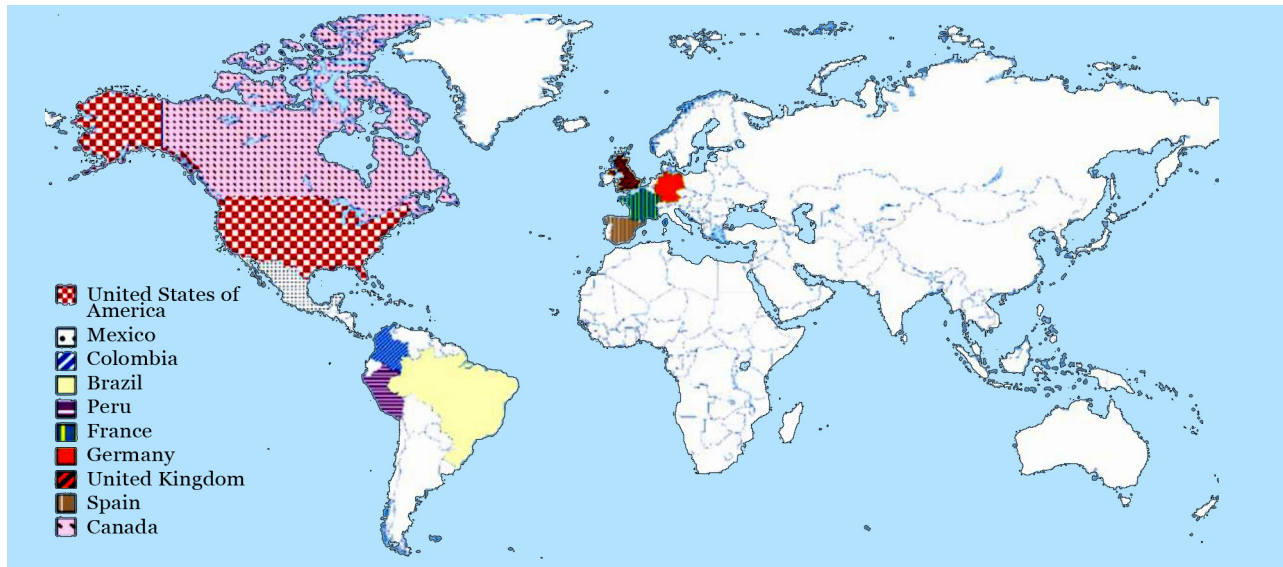


Figure 1. A world map of top ten countries reported cases of monkeypox virus

artificial intelligence, sexual behavior and clinical course of Monkeypox.<sup>8</sup> Many studies also looked after Clinical presentation and virological assessment of confirmed human Monkeypox virus cases.

## Data

The data analyzed consists of new cases on Monkeypox virus confirmed by 10 different countries worldwide (Brazil, Canada, Colombia, France, Germany, Mexico, Peru, Spain, United Kingdom and United States of America). The data was downloaded from the World Health Organization (WHO) website<sup>9</sup> reported from 02<sup>nd</sup> May 2022 to 05<sup>th</sup> February 2023.

The countries were selected on the basis of top 10 countries affected by Monkeypox virus listed in WHO website, thus we believe that the data selected gave a good representation of the main countries affected by Monkeypox virus. The analysis is limited to the data available at the time of writing.

## Methods

In the study of infectious disease, count data is widely used as the study is dependent on the daily incidence of an infectious disease which can be designed and fitted using various methods. In this study, we applied discrete count regression models with the aim of predicting the daily incidence of Monkeypox cases across the world. Such models are used whenever the variable is considered as discrete which is widely accepted and can be easily applicable to non-negative integers.<sup>10</sup> In count data modelling the distribution is often highly skewed or less skewed. We have explored the statistical models like Poisson Regression models and Negative Binomial Regression models.

## Objectives

Our main objective was to provide Statistical analysis to Monkeypox cases in top ten countries among the world. The main

contribution of the paper is:

- (i) To provide statistical analysis to Monkeypox confirmed cases.
- (ii) To investigate if it is feasible to fit count regression models for predicting the daily incidence cases of Monkeypox virus.
- (iii) To identify the trend of new confirmed Monkeypox cases.

**Poisson Regression Models**

For count data, the Poisson distribution model provides the probability of occurrence of number of events occurring within a specified time<sup>11</sup> which is preferred in maximum number of analyses. The Poisson Family of regression models provides ease to implement analysis on count data.<sup>12</sup> It is also a part of generalization of linear regression. Like other regression techniques,<sup>13</sup> the Poisson Regression model can also be used to predict the outcome based on the set of independent variables.

The model<sup>14-17</sup> can be defined using identity and logarithmic link function which is defined in equation I and II respectively

$$\text{Identity link: } \lambda = \beta_0 + \beta_1 t \quad [ I ]$$

$$\text{Natural log link: } \log(\lambda) = \beta_0 + \beta_1 t \quad [ II ]$$

where  $\lambda$  denotes the mean or expected value of the dependent variable, Days denoted as  $t$  is considered as the independent variable and  $\beta_0, \beta_1$  are the intercept and slope of regression equation.<sup>18</sup>

Let  $Y_t$  denotes number of newly confirmed cases of Monkeypox in day  $t$  where  $t = 1, 2, \dots, n$ . For each of the selected countries the following Poisson regression models are fitted with respect to their daily incidence cases.

Hence the probability of occurrence of new cases using identity function is given in equation III and using logarithmic function is

given in equation IV.

$$P(Y_t = y_t) = \begin{cases} \frac{(\beta_0 + \beta_1 t)^{y_t}}{y_t!} * \exp(-(\beta_0 + \beta_1 t)) & y_t = 0, 1, 2, 3 \dots \\ 0 & \text{Otherwise} \end{cases} \quad [ III ]$$

$$P(Y_t = y_t) = \begin{cases} \frac{(\exp(\beta_0 + \beta_1 t))^{y_t}}{y_t!} * \exp[-\exp((\beta_0 + \beta_1 t))] & y_t = 0, 1, 2, 3 \dots \\ 0 & \text{Otherwise} \end{cases} \quad [ IV ]$$

Each of the models was fitted using maximum likelihood method. That is by maximizing the likelihood function with respect to  $\beta_0$ , and  $\beta_1$  as defined in equation V and VI.

$$L_1(\beta_0, \beta_1) = \prod_{t=1}^n \left\{ \frac{(\beta_0 + \beta_1 t)^{y_t}}{y_t!} * \exp(-(\beta_0 + \beta_1 t)) \right\} \quad [ V ]$$

$$L_2(\beta_0, \beta_1) = \prod_{t=1}^n \left\{ \frac{(\exp(\beta_0 + \beta_1 t))^{y_t}}{y_t!} * \exp[-\exp((\beta_0 + \beta_1 t))] \right\} \quad [ VI ]$$

We can denote the maximum likelihood estimates by  $\widehat{\beta}_0$  and  $\widehat{\beta}_1$  respectively.

The Poisson distribution assumes that mean and variance is always equal but in real count data examples it seems tedious task to make the parameters equal. Usually in any real count data examples, the variance exceeds mean which is often termed as “overdispersion”. For count data model, the standard overdispersion model is the negative binomial distribution.<sup>19</sup> The Negative binomial regression models is preferred over Poisson Regression models if the data is affected due to scatterdness.

**Negative Binomial Regression Models**

The Negative Binomial distribution provides

the occurrence of number of cases for overdispersed data. Let  $Y_t$  denotes number of newly confirmed cases of Monkeypox in day  $t$  where  $t = 1, 2, \dots, n$ . Hence, the probability of occurrence of new case using identity and Natural log function stated in equation [I] and [II] respectively is defined in equation [VII] and [VIII] respectively.

$$P(Y_t = y_t) = \begin{cases} \binom{y_t + \beta_0 + \beta_1 t - 1}{y_t} * p^{\beta_0 + \beta_1 t} * (1 - p)^{y_t} & y_t = 0, 1, 2, 3 \dots; 0 \leq p \leq 1 \\ 0 & \text{Otherwise} \end{cases} \quad \text{[VII]}$$

$$P(Y_t = y_t) = \begin{cases} \binom{y_t + \exp(\beta_0 + \beta_1 t) - 1}{y_t} * p^{\exp(\beta_0 + \beta_1 t)} * (1 - p)^{y_t} & y_t = 0, 1, 2, 3 \dots; 0 \leq p \leq 1 \\ 0 & \text{Otherwise} \end{cases} \quad \text{[VIII]}$$

Each of the models was fitted using maximum likelihood method. That is by maximizing the likelihood function with respect to  $\beta_0, \beta_1$  and  $p$  as defined in equation [IX] and [X].

$$L_3(\beta_0, \beta_1, p) = \prod_{t=1}^n \left\{ \binom{y_t + (\beta_0 + \beta_1 t) - 1}{y_t} * p^{\beta_0 + \beta_1 t} * (1 - p)^{y_t} \right\} \quad \text{[IX]}$$

$$L_4(\beta_0, \beta_1, p) = \prod_{t=1}^n \left\{ \binom{y_t + \exp(\beta_0 + \beta_1 t) - 1}{y_t} * p^{\exp(\beta_0 + \beta_1 t)} * (1 - p)^{y_t} \right\} \quad \text{[X]}$$

We can denote the maximum likelihood estimates by  $\widehat{\beta}_0, \widehat{\beta}_1$  and  $\widehat{p}$  respectively. In this study the maximum likelihood estimates for both Poisson and Negative Binomial Regression Models are generated using

Statistical Package for Social Sciences (SPSS, Version 23.0, 2015). The best fitted model was selected using the minimum value of Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

For each of the fitted models, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were computed 10 as defined in equation [XI] and [XII] respectively

$$\left. \begin{aligned} \text{AIC} &= 4 - 2 \log L_1(\beta_0, \beta_1) \\ \text{AIC} &= 4 - 2 \log L_2(\beta_0, \beta_1) \\ \text{AIC} &= 6 - 2 \log L_3(\beta_0, \beta_1, p) \\ \text{AIC} &= 6 - 2 \log L_4(\beta_0, \beta_1, p) \end{aligned} \right\} \text{[XI]}$$

and

$$\left. \begin{aligned} \text{BIC} &= 2 \log n - 2 \log L_1(\beta_0, \beta_1) \\ \text{BIC} &= 2 \log n - 2 \log L_2(\beta_0, \beta_1) \\ \text{BIC} &= 3 \log n - 2 \log L_3(\beta_0, \beta_1, p) \\ \text{BIC} &= 3 \log n - 2 \log L_4(\beta_0, \beta_1, p) \end{aligned} \right\} \text{[XII]}$$

The AIC and BIC values were computed using SPSS, version 23.0, 2015 which is represented in Table 1.

According to AIC and BIC values the best model fitted to the data was Negative Binomial regression with a logarithmic link function.

### Results

We applied Negative Binomial regression model as specified in Section 3.3 and fitted them into our data on the number of new daily cases of Monkeypox reported in ten different countries worldwide.

### Explanatory information on newly confirmed cases

The Table 2 provided an overall idea about

Table 1. AIC values (BIC values) for the four models fitted to the data from 02nd May 2022 to 05th February 2023.

Country	Models			
	Poisson (Identity function)	Poisson (logarithmic function)	Negative Binomial (Identity function)	Negative Binomial (logarithmic function)
Brazil	8218.530 (8221.641)	9204.720 (9207.830)	479.052.72 (483.718)	469.299 (473.965)
Canada	-	963.539 (966.650)	-	266.700 (271.366)
Colombia	5299.270 (5302.263)	5426.253 (5429.246)	-	377.851 (382.341)
France	-	5563.644 (5566.919)	388.839 (393.751)	369.621 (374.534)
Germany	-	3941.235 (3944.402)	-	347.960 (352.710)
Mexico	4145.811 (4148.978)	4334.114 (4337.281)	-	391.229 (395.979)
Peru	2138.506 (2141.437)	2574.494 (2577.425)	-	362.326 (366.723)
Spain	-	9387.581 (9390.803)	-	418.619 (423.452)
United Kingdom	-	4076.981 (4080.359)	398.211 (403.278)	373.304 (378.371)
United States of America	-	34874.640 (34877.807)	537.810 (542.560)	536.346 (541.096)

Table 2. Descriptive Statistics on newly confirmed cases

Country	Obs	Min	Max	Total No of cases	Mean	Standard deviation	Median	Interquartile Range (1st Quartile, 3rd Quartile)
Brazil	35	1	1022	10758	307.37	304.75	179	408.000 (63.00,471.00)
Canada	35	0	164	1461	41.74	49.72	26	65.000 (0.00,65.00)
Colombia	33	0	411	4072	123.39	144.07	58	241.500 (9.50,251.00)
France	38	0	541	4128	108.63	156.87	31	171.000 (2.75, 173.75)
Germany	36	0	478	3692	102.89	143.94	22	164.300 (2.00, 166.25)
Mexico	36	0	341	3768	104.67	108.39	71	166.500 (4.50, 171.00)
Peru	32	3	363	3727	116.47	105.05	78	182.000 (25.00, 207.00)
Spain	37	0	1234	7528	203.46	322.46	55	227.000 (11.50,238.50)
United Kingdom	40	0	402	3735	93.38	122.67	29	152.800 (2.00,154.75)
United States of America	36	0	3324	29933	831.47	1030.37	332	1299.500 (63.25, 1362.75)

Obs, Observations; Min, Minimum; Max, Maximum; No, Number;

the Descriptive Statistics of newly confirmed cases reported by ten different countries worldwide. It was observed that the maximum number of cases was reported by United states of America with a median number of cases as 332 and Interquartile range as 1299.5 followed by Brazil reported the second most country with a median number of cases as 179 and Interquartile range as 408. The United States of America reported 3324 cases on a particular

day followed by Brazil reported 1022 cases. From among the selected countries the least number of cases were reported by Canada with maximum 164 cases on a particular day.

**Assessing the number of new cases that occur in upcoming days**

Among the countries selected, the Table 3 gives estimates of intercept and slope parameter

Table 3. Parameter estimates and Standard errors with IRR and its 95% Confidence Interval for Negative Binomial Regression model with a logarithmic link function.

Country	Intercept	Standard Error	Slope	Standard Error	IRR	95% Confidence Interval for IRR	P value
Brazil	6.784	0.442	-0.010	0.003	0.990	0.984 – 0.997	0.004*
Canada	6.13	0.385	-0.029	0.003	0.972	0.966 – 0.978	<0.01*
Colombia	5.497	0.818	-0.006	0.007	0.994	0.980 – 1.007	0.368
France	7.167	0.518	-0.025	0.004	0.975	0.968 – 0.982	<0.01*
Germany	6.595	0.381	-0.023	0.003	0.977	0.972 – 0.983	<0.01*
Mexico	3.771	0.879	0.007	0.007	1.007	0.993 – 1.021	0.315
Peru	5.956	0.377	-0.013	0.003	0.987	0.981 – 0.994	<0.01*
Spain	7.433	0.386	-0.022	0.003	0.979	0.974 – 0.984	<0.01*
United Kingdom	7.213	0.493	-0.027	0.003	0.974	0.967 – 0.980	<0.01*
United States of America	8.536	0.502	-0.017	0.004	0.983	0.975 – 0.990	<0.01*

along with its standard error and the p values stating the significance of slope parameter. The Table 3 also provides the Incidence Rate Ratio (IRR) with its 95% confidence interval. It is clearly observed that the p values of all countries except Colombia and Mexico are less than 0.05 which indicates the significance of slope coefficient estimates at 5 % level of significance.

Among the countries selected where the model appears to be best fit, slope coefficient of majority of the country was negative indicating that the expected number of cases confirmed each day is likely to get reduced with respect to time whereas Mexico country appears with positive slope coefficient which indicates expected number of cases may rise with respect to time. However, the slope coefficient corresponding to Mexico country does not show any significant increase with respect to time. On similar line the slope coefficient corresponding to Columbia country does not show significant decrease with respect to time. Among the countries selected with significant slope coefficient, the countries like Canada and Brazil have largest and smallest

slope coefficient which shows maximum and minimum decrease in expected number of cases confirmed each day respectively.

The best fitted Negative binomial regression model also showed that the Incidence Rate Ratio with its 95% Confidence Interval was below 1.0, indicating the newly confirmed cases decreases with respect to time. The newly confirmed cases using IRR for Brazil got decreased by 1% (95% CI: 0.984 – 0.997), for Canada got decreased by 2.8% (95% CI: 0.966 – 0.978), for France got decreased by 2.5% (95% CI: 0.968 – 0.982), for Germany got decreased by 2.3% (95% CI: 0.972 – 0.983), for Peru got decreased by 1.3% (95% CI: 0.981 – 0.994), for Spain got decreased by 2.1% (95% CI: 0.974 – 0.984), for United Kingdom got decreased by 2.6% (95% CI: 0.967 – 0.980) and for United States of America got decreased by 1.7% (95% CI: 0.975 – 0.990) where Canada indicates maximum significant decrease in cases among other countries followed by the minimum significant decrease was observed in Brazil. However, the Mexico country were expected to experience 1.007 (95% CI: 0.993 – 1.021) times increase in cases per day

than other countries but the increase was not significant.

### Trend in newly confirmed cases of Monkeypox among ten different countries

The prediction of new cases is obtained using Negative Binomial regression model using log function where log of the outcome is predicted using linear combination of the predictors.<sup>20</sup>

$\log(\text{new cases}) = \text{Intercept} + \text{Slope} * \text{Days}$   
 This implies  $\text{New cases} = \exp(\text{Intercept} + \text{Slope} * \text{Days})$

The predicted lines as shown in Figure 2 and Figure 3 are plotted using Negative Binomial Regression model with logarithmic function. The number of cases corresponds to an epidemic curve representing newly confirmed cases of Monkeypox virus decreases with

respect to time. The prediction line was not perfectly fit in the beginning of the days but as time passes on, the prediction and the newly confirmed cases are almost similar. However, the model was unable to predict the number of cases in Mexico and Columbia as it neither shown any significant increase nor decrease with respect to the number cases every day.

### Discussion

It is clearly reported in this study that Monkeypox cases started decreasing significantly in many parts of the reported countries. It was mentioned in many literatures<sup>21</sup> that increase in Monkeypox cases was found more among the places where people are more in contact with each other or they travel internationally. The another factor was also stated that the routine

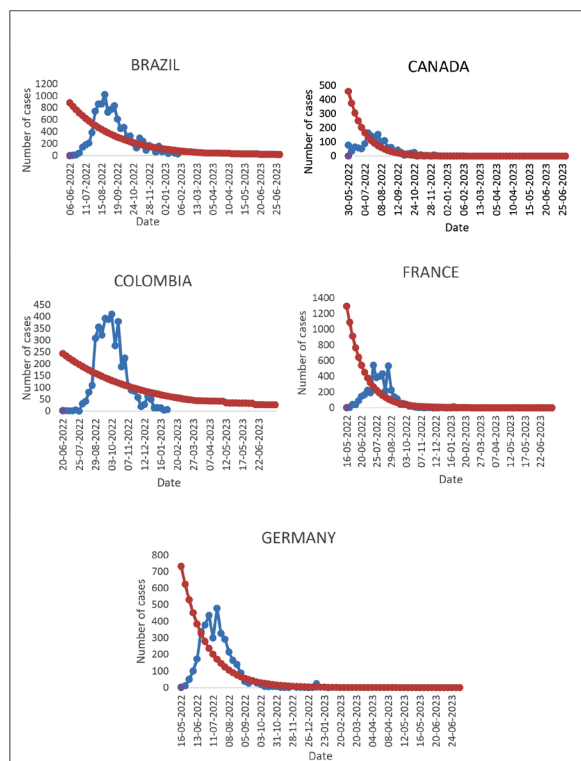


Figure 2. The number of cases versus Date of notification among countries like Brazil, Canada, Colombia, France and Germany. The blue line represents actual number of cases and red line represents the predicted number of cases on each day.

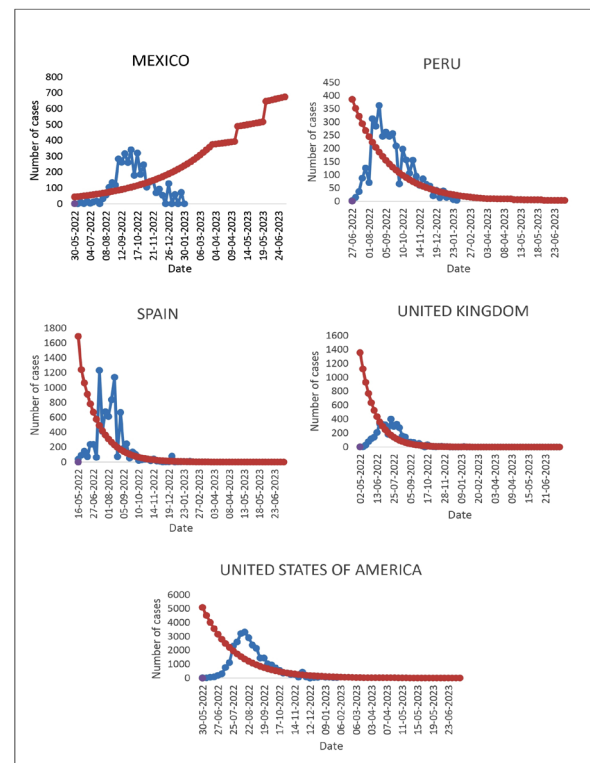


Figure 3. The number of cases versus Date of notification among countries like Mexico, Peru, Spain, United Kingdom and United States of America. The blue line represents actual number of cases and red line represents the predicted number of cases on each day.



immunization was stopped in the early 1970 in places like United States of America and United Kingdom. Many people who born after this year was not at all immune because small pox vaccines can prevent the rapid spread of Monkeypox virus. The Monkeypox cases in India were quite less as compared to other countries because India already witnessed a similar situation due to COVID 19 outbreak where the entire Health care workers had an impact of this outbreak on their mental and physical health. So India was prior prepared to face the outbreak of Monkeypox cases and contact tracing was carried out with strict implementation of appropriate prevention and control measures to restrict the further increase of Monkeypox cases in India. Since the countries were already facing the outbreak of COVID 19 so due to vaccinations and due to the precautions taken before hand the spread of Monkeypox was very less and the casualties reported due to Monkeypox was also quite less as compared with the outbreak of COVID 19. In the recent meeting<sup>22</sup> of the International Health Regulations (2005), the Committee reported the significant decline in the global spread of Monkeypox cases in many countries. The Director-General also noted that even though there is a downward trend globally, some communities still face the transmission of virus from one person to another person. The Committee also expressed concern towards lack of availability of vaccines, medicines, and diagnostic testing centers in many countries. In this article, the best fitted model was considered to be Negative Binomial Regression Model for modelling the number of newly confirmed cases of Monkeypox virus in different countries. However, the model cannot predict any significant increase

in Monkeypox cases in Mexico as the cases in Mexico are completely zero nowadays and some days more than 10.<sup>23</sup> Earlier, it was significantly increasing in Mexico as the MPOX vaccine was not authorized to use in Mexico Country.<sup>24</sup> The Colombia and Mexico were the two countries which showed an increase in Monkeypox cases at the last moment as compared with other countries so may be because of that it doesn't show any significant increase or decrease in this study. Since Monkeypox has no treatment<sup>25</sup> it has been authorized that one medication, Tecovirimat (TPOXXTM) can be given for the treatment of small pox in the United States, Canada and Europe from May 2022. In this study, the data collected from WHO website does not contain values with no cases hence Zero Order Inflated Model or Zero Order Hurdle Model were not used as the surplus zeros (no cases) where not at all affecting the dispersion in data. Such Zero Inflated models<sup>26</sup> are used in many areas like insurance,<sup>27</sup> education,<sup>28</sup> health care<sup>29</sup> and transport.<sup>30</sup>

In this era, where variety of diseases spread out across the countries, it becomes very essential to get involved with Statistical models to predict the number of newly infected cases.<sup>31</sup> In most of the studies, the number of newly infected cases are determined using Negative Binomial Regression model as this confirms over dispersion is due to unobserved heterogeneity. In health services, the count data modelling can be used widely as it depends on discrete data.<sup>32</sup> There are several studies related to count data Regression models<sup>31</sup> which concentrates on predicting the number of involved nodes in Breast Cancer patients.<sup>33</sup> The prediction works on similar line as compared to actual number of confirmed cases after collecting data of at least

4 months. Usually whenever it is a categorical variable, Binary Logistic Regression model is used but several studies<sup>34</sup> states that Negative Binomial Regression predicts effectively than Binary logistic regression model.

The Negative Binomial Regression model fits the data better than Poisson Regression model because the predicted values are fairly close enough to the observed values at the two tails but Poisson Regression model underestimates very low and high frequencies.<sup>18</sup> This study truly ignores the over dispersion and provides an unbiased result. This model may be more useful in predicting the new daily cases when the trend is significantly increasing or decreasing. It gives improper result when there is no significant increase or decrease observed in the data. In addition, such predictive models could be very useful in contributing to making health policy decisions or government intervention, but most importantly it should be compared with all other models<sup>35</sup> which are suitably fit for epidemiology.

Due to availability of technologies nowadays Machine learning have shown potential for predicting kidney allograft outcomes.<sup>36</sup> The model based on Machine Learning are developed for kidney allograft survival and it was compared with Cox Based Prognostication system. Artificial intelligence is widely used by many researchers to predict the outcome or to predict the pattern of increase and decrease in cases. Artificial Intelligence<sup>37</sup> can be used to analyse a wide variety of biopsy related data. Nowadays using Neural networks<sup>38</sup> infant cries are examined to support inexperienced parents to make an early analysis of any medical disorder exists. Nowadays, the Artificial Intelligence<sup>39-41</sup> are widely used for all kind of predictions. But such models are bit

difficult and require lot of time to understand the workings of such models.

## Conclusion

The study was investigated using Negative binomial regression model with logarithmic link function as it fitted well to the data. The model was fitted and analyzed using top ten countries who severely affected due to Monkeypox cases. Given these results, it was suggested that this model is very useful for predicting the number of cases provided the significance should exists. Among the selected countries that fit well, the slope estimates found to be negative which indicates that the number of cases decreases with time. The countries like Canada and Brazil have largest and smallest slope coefficient which shows maximum and minimum decrease in expected number of cases confirmed each day respectively. However, the model could not predict any significant decrease or increase in the number of Monkeypox cases reported by Columbia and Mexico.

## Limitation and Future Research

This study is limited to secondary data collected from WHO website. This method has its own advantages of being simple and easy to understand but yet it is difficult to handle. The conclusions drawn using this model should be done with proper care. The prediction does not give proper fit in the beginning days. The direct extension of this study should include the mortality rate due to Monkeypox virus. The other model with less dispersed can be utilized for mortality rate due to Monkeypox virus.

## Acknowledgment

The authors are thankful to World Health Organization (WHO) website for providing us with the data to do further analysis. The authors thank the reviewers for their insightful comment which helped a lot to improve this article.

## Conflicts of Interest

The Authors declare no conflicts of interests.

## Funding

This research work received no grant from any funding agencies.

## References

- Patel M, Surti M, Adnan M. Artificial intelligence (AI) in Monkeypox infection prevention. *J Biomol Struct Dyn*. 2022 Oct 11;1–5.
- Wei F, Peng Z, Jin Z, Wang J, Xu X, Zhang X, et al. Study and prediction of the 2022 global monkeypox epidemic. *J Biosaf Biosecurity*. 2022 Dec;4(2):158–62.
- Rosa RB, Ferreira de Castro E, Vieira da Silva M, Paiva Ferreira DC, Jardim ACG, Santos IA, et al. In vitro and in vivo models for monkeypox. *iScience*. 2023 Jan;26(1):105702.
- Besombes C, Mbrengra F, Schaeffer L, Malaka C, Gonofio E, Landier J, et al. National Monkeypox Surveillance, Central African Republic, 2001–2021. *Emerg Infect Dis*. 2022 Dec;28(12):2435–45.
- Miura F, Backer JA, van Rijckevorsel G, Bavalia R, Raven S, Pettrignani M, et al. Time scales of human mpox transmission in the Netherlands [Internet]. *Infectious Diseases (except HIV/AIDS)*; 2022 Dec [cited 2023 Feb 27]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2022.12.03.22283056>
- Kozlov M. Monkeypox goes global: why scientists are on alert. *Nature*. 2022 Jun 2;606(7912):15–6.
- Beaujean AA, Grant MB. Tutorial on Using Regression Models with Count Outcomes using R. [cited 2023 Feb 11]; Available from: <https://scholarworks.umass.edu/pare/vol21/iss1/2/>
- Ogoina D. Sexual behaviours and clinical course of human monkeypox in Spain. *The Lancet*. 2022 Aug;400(10353):636–7.
- 2022-23 Mpox (Monkeypox) Outbreak: Global Trends [Internet]. [cited 2023 May 2]. Available from: [https://worldhealthorg.shinyapps.io/mpox\\_global/#35\\_Symptomatology](https://worldhealthorg.shinyapps.io/mpox_global/#35_Symptomatology)
- Chan S, Chu J, Zhang Y, Nadarajah S. Count regression models for COVID-19. *Phys Stat Mech Its Appl*. 2021 Feb;563:125460.
- Du J, Park YT, Theera-Ampornpant N, McCullough JS, Speedie SM. The use of count data models in biomedical informatics evaluation research. *J Am Med Inform Assoc*. 2012 Jan;19(1):39–44.
- Coxe S, West SG, Aiken LS. *The Analysis of Count Data: A Gentle Introduction*

- to Poisson Regression and Its Alternatives. *J Pers Assess.* 2009 Feb 17;91(2):121–36.
13. Vetter TR, Schober P. Regression: The Apple Does Not Fall Far From the Tree. *Anesth Analg.* 2018 Jul;127(1):277–83.
14. Christou V, Fokianos K. Estimation and testing linearity for non-linear mixed poisson autoregressions. *Electron J Stat [Internet].* 2015 Jan 1 [cited 2023 Feb 2];9(1). Available from: <https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-9/issue-1/Estimation-and-testing-linearity-for-non-linear-mixed-poisson-autoregressions/10.1214/15-EJS1044.full>
15. Christou V, Fokianos K. QUASI-LIKELIHOOD INFERENCE FOR NEGATIVE BINOMIAL TIME SERIES MODELS: NEGATIVE BINOMIAL TIME SERIES. *J Time Ser Anal.* 2014 Jan;35(1):55–78.
16. Fokianos K, Fried R. Interventions in log-linear Poisson autoregression. *Stat Model.* 2012 Aug;12(4):299–322.
17. Kitromilidou S, Fokianos K. Mallows' quasi-likelihood estimation for log-linear Poisson autoregressions. *Stat Inference Stoch Process.* 2016 Oct;19(3):337–61.
18. Moksony F, Hegedűs R. The use of Poisson regression in the sociological study of suicide. *Corvinus J Sociol Soc Policy.* 2014 Dec 20;5(2):97–114.
19. Lindsey JK. On the Use of Corrections for Overdispersion. *J R Stat Soc Ser C Appl Stat.* 1999 Dec 1;48(4):553–61.
20. Dupont WD. *Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data [Internet].* 2nd ed. Cambridge University Press; 2009 [cited 2023 May 10]. Available from: <https://www.cambridge.org/core/product/identifier/9780511575884/type/book>
21. Mohapatra RK, Tuli HS, Sarangi AK, Chakraborty S, Chandran D, Chakraborty C, et al. Unexpected sudden rise of human monkeypox cases in multiple non-endemic countries amid COVID-19 pandemic and salient counteracting strategies: Another potential global threat? *Int J Surg.* 2022 Jul;103:106705.
22. Fifth Meeting of the International Health Regulations (2005) (IHR) Emergency Committee on the Multi-Country Outbreak of mpox (monkeypox) [Internet]. [cited 2023 May 13]. Available from: [https://www.who.int/news/item/11-05-2023-fifth-meeting-of-the-international-health-regulations-\(2005\)-\(ihr\)-emergency-committee-on-the-multi-country-outbreak-of-monkeypox-\(mpox\)](https://www.who.int/news/item/11-05-2023-fifth-meeting-of-the-international-health-regulations-(2005)-(ihr)-emergency-committee-on-the-multi-country-outbreak-of-monkeypox-(mpox))
23. Saavedra J, Faviero GF, Baruch R, Pinzon A. MPOX vaccines needed in Mexico. *Lancet Reg Health - Am.* 2023 Apr;20:100457.
24. Gordon NC, Rampling T. The 2022 monkeypox outbreak: the need for clinical curiosity. *Lancet Infect Dis.* 2023 Feb;23(2):133–4.
25. Aljabali AAa, Obeid MA, Nusair MB, Hmedat A, Tambuwala MM. Monkeypox

- virus: An emerging epidemic. *Microb Pathog*. 2022 Dec;173:105794.
26. Sidumo B, Sonono E, Takaidza I. Count Regression and Machine Learning Techniques for Zero-Inflated Overdispersed Count Data: Application to Ecological Data. *Ann Data Sci* [Internet]. 2023 Apr 13 [cited 2023 May 4]; Available from: <https://link.springer.com/10.1007/s40745-023-00464-6>
  27. Desjardins CD. Modeling Zero-Inflated and Overdispersed Count Data: An Empirical Study of School Suspensions. *J Exp Educ*. 2016 Jul 2;84(3):449–72.
  28. Araújo EG, Vasconcelos JCS, Dos Santos DP, Ortega EMM, De Souza D, Zanetoni JPF. The Zero-Inflated Negative Binomial Semiparametric Regression Model: Application to Number of Failing Grades Data. *Ann Data Sci* [Internet]. 2021 Jun 30 [cited 2023 May 4]; Available from: <https://link.springer.com/10.1007/s40745-021-00350-z>
  29. Bekalo DB, Kebede DT. Zero-Inflated Models for Count Data: An Application to Number of Antenatal Care Service Visits. *Ann Data Sci*. 2021 Dec;8(4):683–708.
  30. Mehdizadeh A, Cai M, Hu Q, Alamdar Yazdi MA, Mohabbati-Kalejahi N, Vinel A, et al. A Review of Data Analytic Applications in Road Traffic Safety. Part 1: Descriptive and Predictive Modeling. *Sensors*. 2020 Feb 18;20(4):1107.
  31. Dwivedi AK, Dwivedi SN, Deo S, Shukla R. Statistical models for predicting number of involved nodes in breast cancer patients. *Health (N Y)*. 2010;02(07):641–51.
  32. Salinas-Rodríguez A, Manrique-Espinoza B, Sosa-Rubí SG. Análisis estadístico para datos de conteo: aplicaciones para el uso de los servicios de salud. *Salud Pública México* [Internet]. 2009 Oct [cited 2023 May 4];51(5). Available from: [http://www.scielo.org/scielo.php?script=sci\\_arttext&pid=S0036-36342009000500007&lng=es&nrm=iso&tlng=es](http://www.scielo.org/scielo.php?script=sci_arttext&pid=S0036-36342009000500007&lng=es&nrm=iso&tlng=es)
  33. Long JS, Freese J. Regression models for categorical dependent variables using Stata. College Station, Tex: Stata Press; 2001. 341 p.
  34. Yu L. Using Negative Binomial Regression Analysis to Predict Software Faults: A Study of Apache Ant. *Int J Inf Technol Comput Sci*. 2012 Jul 16;4(8):63–70.
  35. Roques L, Klein E, Papaix J, Soubeyrand S. Mechanistic-statistical SIR modelling for early estimation of the actual number of cases and mortality rate from COVID-19.
  36. Truchot A, Raynaud M, Kamar N, Naesens M, Legendre C, Delahousse M, et al. Machine learning does not outperform traditional statistical modelling for kidney allograft failure prediction. *Kidney Int*. 2023 May 1;103(5):936–48.
  37. Becker JU, Mayerich D, Padmanabhan M, Barratt J, Ernst A, Boor P, et al. Artificial intelligence and machine learning in nephropathology. *Kidney Int*. 2020 Jul

1;98(1):65–75.

38. Ozseven T. Infant cry classification by using different deep neural network models and hand-crafted features. *Biomed Signal Process Control*. 2023 May 1;83:104648.

39. Churyumov AY, Kazakova AA. Prediction of True Stress at Hot Deformation of High Manganese Steel by Artificial Neural Network Modeling. *Materials*. 2023 Jan;16(3):1083.

40. Churyumov AY, Pozdniakov AV, Mondoloni B, Prosviryakov AS. Effect of boron concentration on hot deformation behavior of stainless steel. *Results Phys*. 2019 Jun 1;13:102340.

41. Arun Babu K, Prithiv TS, Gupta A, Mandal S. Modeling and simulation of dynamic recrystallization in super austenitic stainless steel employing combined cellular automaton, artificial neural network and finite element method. *Comput Mater Sci*. 2021 Jul 1;195:110482.