

Modeling and Analysis of Influenza A H1N1 Outbreaks in India

Stavelin Abhinandithe K*, Sathya Velu R¹, Dr. Madhu B², Sahana S¹,
Sowmyavalli R¹, Bibin John¹, Dr. Balasubramanian S³

*Assistant Professor, Division of Medical Statistics, Faculty of Life Sciences,

¹Division of Medical Statistics, Faculty of Life Sciences,

²Associate Professor, Department of Community Medicine JSS Medical College,

³Dean & Director of Research, Department of Water & Health,

^{*,1,2,3}JSS AHER, Mysore, Karnataka, India

ABSTRACT

Influenza A (H1N1) incidence in India, which is a highly contagious acute respiratory disease in humans caused by type A influenza virus. This study employs retrospective comparative study of the data from National Centre for Disease Control (CDC) yearly reports from April-2010 to April-2019. The case fatality rates of Influenza A (H1N1) incidence in India was forecasted using autoregressive integrated moving average (ARIMA) models in order to build a predictive tool for Influenza A (H1N1) surveillance. Clearly from the study, lack of rainfall spread the virus more efficiently and Maharashtra stood first in total number of cases and deaths of Influenza A(H1N1) whereas Lakshadweep had no signs of disease. Further, number of Cases were reported in the year 2015 (i.e. 28%) and 25% of cases have been reported in 2017 when compared to ten years data. (ARIMA) (2, 1, 3) model was selected for its minimum value of normalized BIC, MAPE and good R-Square among all other models. The model forecasted the decrease in case fatality rate of Influenza A H1N1 for next 10years. Thus, results indicate that, ARIMA models provide a means to better understand Influenza A (H1N1) incidence yielding forecasts that can be used for public health planning at the national level.

KEYWORDS: ARIMA, case fatality rate (CFR), contagious, forecasts, Influenza A (H1N1), predictive tool

How to cite this paper: Stavelin Abhinandithe K | Sathya Velu R | Dr. Madhu B | Sahana S | Sowmyavalli R | Bibin John | Dr. Balasubramanian S "Modeling and Analysis of Influenza A H1N1 Outbreaks in India" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-3, April 2021, pp.1166-1170, URL: www.ijtsrd.com/papers/ijtsrd39862.pdf



IJTSRD39862

Copyright © 2021 by author (s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



I. INTRODUCTION

Influenza A (H1N1 or swine flu) is a highly contagious acute respiratory disease caused by type A influenza virus that regularly causes outbreaks of influenza in humans. As of June 2009, the CDC has identified that influenza viruses from different species infect pigs, thus the viruses can reassert (i.e. swap genes) and new viruses that are a mix of swine, human and/or avian influenza viruses can emerge. There are four main influenza types A virus subtypes that have been isolated in pigs: H1N1, H1N2, H3N2, and H3N1. Most of the recently isolated influenza viruses from pigs, however, have been H1N1viruses. Influenza viruses in human populations are usually classified as seasonal or pandemic. Seasonal influenza usually causes yearly epidemics in the old, the very young and those with underlying medical conditions, and is often associated with cold periods in temperate regions. An influenza virus becomes classified as pandemic when a novel strain emerges from an animal reservoir and is not only able to infect humans but also spreads quickly through the population across the globe irrespective of the season. The H1N1 influenza virus causes moderate to severe respiratory infections. Symptoms include fever, sore throat, cough, headache, chills and fatigue. Severe cases include Pneumonia bronchitis, sinus infections and an Increase in underlying conditions. It is contagious and can be contracted through

coughing, sneezing or even talking to someone carrying it. It can also be contracted through mouth or nose contact with something contaminated[1]. Of 18 500 laboratory-confirmed pandemic-associated deaths identified during April, 2009, to April, 2010, worldwide, less than 12% were reported from Africa and southeast Asia, although these regions are home to more than 38% of the world's population During the 2009 pandemic, India reported 27,236 laboratory confirmed cases of pandemic influenza A (H1N1) with 981 deaths. In 2015, India reportedly had over 31,000 people infected and 1,900 resulting deaths. Most of these cases were reported from the States of Rajasthan, Gujarat, Delhi, Jammu and Kashmir, Maharashtra, Madhya Pradesh, Telangana, Karnataka and Tamil Nadu with a major peak in rainy season (July-August), and a minor peak in winter-spring season (December-February)[2].

Forecasting the future is one the fundamental tasks of time series analysis. Although linear methods are useful, a prediction from a nonlinear point of view is one step closer to reality. The present study is concerned with the building of ARIMA models for forecasting case fatality rates of Influenza A (H1N1) patterns in India.

II. Materials and methods

A. Data sources:

We obtained the data of Influenza A cases from the National Centre For Disease Control (Formerly National Institute Of Communicable Diseases) which is freely available from the website: <https://ncdc.gov.in>. The data consists of seasonal influenza A (H1N1) – state/UT wise number of cases and deaths from 2010-2019. It is a retrospective study.

B. Statistical Analysis:

Correlation analysis was used to check the relationship or association between the cases the deaths of Influenza A (H1N1). Year-wise case fatality rate is calculated as the ratio of deaths occurring from influenza to the total number of cases due to the same cause. Case fatality rate measures the severity of the disease that causes death. Thematic maps were used to understand the patterns of the regional distributions of influenza A (H1N1) cases and deaths in India [3,4].

ARIMA models offer the statistical infrastructure for intervention analysis for the assessment of the clinical effect of therapeutic interventions .A ARIMA model introduced by Box and Jenkins (1976) was used to describe the different model structure with the obvious notations (p, d, q). In this process, the prime task is to identify an appropriate subclass of models from the general ARIMA family, and construct the ARIMA model for the data. The model is estimated using the following steps: Identification, estimation, diagnostic checking and forecasting. The first step in developing a Box-Jenkins model is to determine if the time series is stationary or non-stationary. The autocorrelation function (ACF) and partial correlation function (PACF) are used to detect the stationary or non-stationary of time series and a test of significant autocorrelation coefficients using Ljung-Box Chi-Square test (Q – test) Augmented dickey fuller (ADF) unit root test might be estimated to study the stationary of the series [5].

After achieving time-series stationary the identification model process begins, and we mean by the using of data about how they are generated by time-series. The identification of the model is the determination of the order of the respective processes of the ARIMA model which can be achieved using ACF and PACF plots.

The Method of least square is considered as an easiest way which is more widely used to estimate model parameters. The aim of Least Squares method is to choose parameter estimators that make the sum of residual squares between real values and the estimated values of the dependent variable possible minimum. Schwarz (1978) proposed Bayesian Information Criterion (BIC) for the selection of the model. Selection criterions such as R-Square, Stationary R-Square, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Bayesian Information Criterion (BIC) are used to select the best fitting ARIMA model [6,7,8]

After estimating the model parameters we will carry out the test of appropriate quality at this stage, this estimate and the test of diagnosed model will be carried out to find out the availability of its own conditions, by the test of random errors (residuals) [9]. Residuals are the difference between the value that occurs and the value that is forecasted for a given time period. The estimated model and the parameters are checked using certain techniques. The tools used for

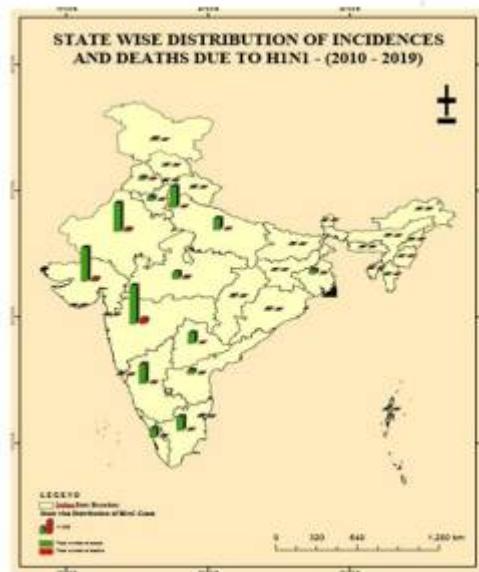
diagnostic checking are: over fitting, residuals of ACF and PACF.

Forecasting is the last stage of study and analysis of time-series models, and is the main objective of the study. The optimum forecasting is the estimate from the resulting error is very small and its variance would be a least [10].

C. Statistical software's:

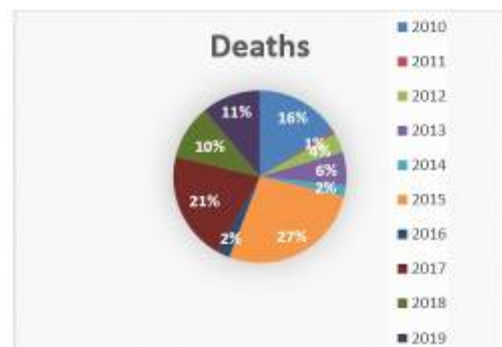
The above statistical analysis was carried out using the well-known software's: MS-Excel, IBM SPSS Statistics 22 and Arc Gis version 10.2 (Demo Version). MS-Excel was used for the calculation of case fatality rates and construction of bar and pie charts showing distribution of influenza cases and deaths. IBM SPSS Statistics 22 proved helpful in correlation analysis and building of ARIMA model. Outputs of thematic maps showing regional distribution of influenza cases and deaths have been driven from Arc Gis version 10.2 (Demo Version). A p-value>0.0 5 was considered statistically significant.

III. Results and Interpretation



A thematic map Showing the distribution of total number of cases and deaths with respect to state/UT wise from 2010-2019 of India

Here we can notice that Maharashtra stands first in total number of cases and deaths of Influenza A(H1N1), followed by Gujarat, Rajasthan, Madhya Pradesh and Karnataka. Whereas Nagaland, Tripura, Sikkim has few number of cases reported, but there is no traces of deaths due to Influenza A (H1N1) and Lakshadweep had no signs of disease. The states that are most affected by the virus experienced huge rainfall deficiencies. Long-term climate change has an impact on the spread of H1N1 virus in a unique way.



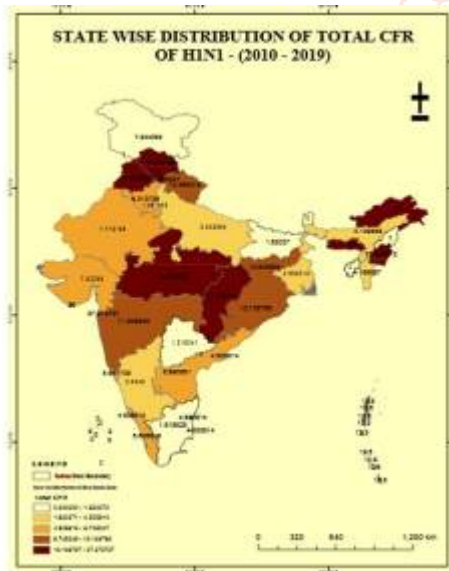
Pie chart showing the distribution of deaths of Influenza-A (H1N1)

We can see that more number of deaths were reported in the year 2015 (i.e. 27%) and 21% of deaths have been reported in 2017 when compared to ten years data. Further, there exists a significant (p value<0.05) strong positive correlation between cases and deaths (r=0.871). As number of cases increases the number of deaths also increases.

Table I: Calculated CFR for Influenza A (H1N1) year wise (2010-2019)

YEAR	CFR
2010	8.55
2011	12.43
2012	8.02
2013	13.3
2014	23.26
2015	7.02
2016	14.83
2017	5.84
2018	7.29
2019	3.71

From the table it is clear that proportion of deaths for a specified population containing H1N1 influenza cases is highest for the year 2014 (23.26) Further consecutive increase and decrease in the CFR in the upcoming years indicate clear fluctuations. However, CFR was found to be the lowest in the year 2019.



A thematic map of India showing average CFR of Influenza A (H1N1)

The above map shows that Meghalaya, Dadra & Nagar Haveli, Chhattisgarh, Manipur, Daman & Diu, Arunachal Pradesh, Punjab, Madhya Pradesh, Himachal Pradesh and Odisha have the highest distribution of total CFR, followed by Andaman & Nicobar, Maharashtra, Uttarakhand, Jharkhand, Chhattisgarh having moderate distribution. Tamil Nadu, Bihar, Delhi, Lakshadweep have lowest total CFR distribution among the other.

The case fatality rates calculated for 10 years are used for model development and to evaluate forecast performance. A graph is plotted to check the stationarity. The data is found to be no stationary, shown in Figure IV

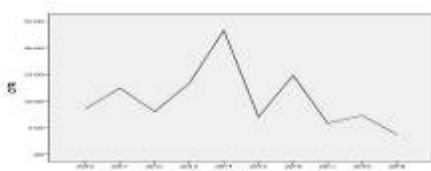
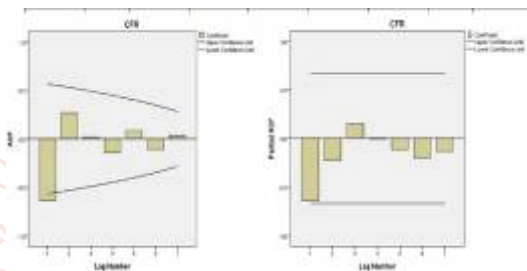


Figure IV: A sequence plot showing the trend line for the case fatality rate of the Influenza A (H1N1) from 2019. Therefore, the first order differencing of the series was carried out. The obtained differenced data was examined for stationarity in mean by computing the autocorrelation and partial autocorrelation coefficients (ACF and PACF). After examining Table II, Figure V and VI, it was concluded that it was stationary in mean.

Table II: ACF and PACF of CFR

lag	ACF	Std. error	Box-Ljung Statistic			lag	PACF	Std. error
			value	df	sig.			
1	-0.635	0.284	4.988	1	0.026	1	-0.635	0.333
2	0.27	0.266	6.018	2	0.049	2	-0.233	0.333
3	0.018	0.246	6.023	3	0.11	3	0.151	0.333
4	-0.142	0.225	6.422	4	0.17	4	-0.007	0.333
5	0.085	0.201	6.601	5	0.252	5	-0.112	0.333
6	-0.108	0.174	6.987	6	0.322	6	-0.196	0.333
7	0.035	0.142	7.05	7	0.424	7	-0.13	0.333



ACF and PACF of differenced CFR

Various orders of 'p' and 'q' were tried with the difference of one (d=1) and set 9 models with structure close to the tentative candidate model to allow the possible identification errors (Table III). This illustrates the importance of having many identification procedures so that the proper model is not missed just because one of the identification methods does not explicitly portray the best model to fit.

Table III: Suggested Models

ARIMA	stationary R-square	RMSE	R-square	MAPE	NBIC
(0,1,1)	0.432	6.86	-0.138	53.2	4.34
(0,1,2)	0.411	7.543	-0.18	51.111	4.774
(1,1,0)	0.408	7.002	-0.186	46.81	4.381
(1,1,1)	0.433	7.399	-0.135	49.425	4.735
(1,1,2)	0.433	8.107	-0.136	49.44	5.162
(2,1,0)	0.436	7.378	-0.129	49.927	4.729
(2,1,1)	0.451	7.975	-0.099	47.695	5.129
(2,1,2)	0.451	8.916	-0.099	48.534	5.596
(2,1,3)	0.509	9.737	0.17	40.665	4.017

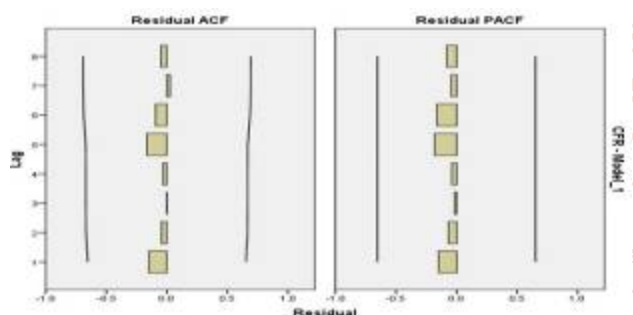
Parameter estimation: It is observed that the model structure of (2, 1, 3) have minimum value of normalized BIC, MAPE and good R-Square. Therefore we have selected ARIMA (2, 1, 3) model as a candidate model for the inflow series and parameters of selected model are given in Table IV.

Table IV: Model statistics of ARIMA

Model	Number of Predictors	Model Fit statistics				
		Stationary R-squared	R-squared	RMSE	MAPE	Normalized BIC
CFR-Model 1	0	0.509	0.017	9.74	40.7	4.017

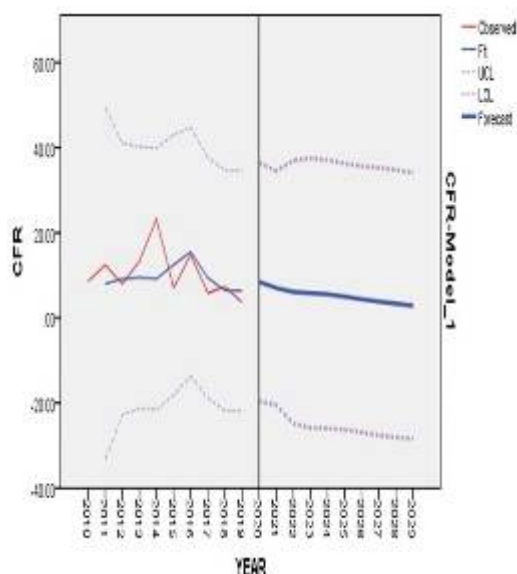
Table V: Predicted values.

Year	Actual values	Predicted values	LCL	UCL
2010	8.55	-	-	-
2011	12.43	8.03	-33.39	49.5
2012	8.02	9.11	-22.83	41.1
2013	13.3	9.45	-21.28	40.2
2014	23.26	9.18	-21.55	39.9
2015	7.02	12.46	-18.13	43.1
2016	14.83	15.52	-13.69	44.7
2017	5.84	9.41	-18.87	37.7
2018	7.29	6.42	-21.85	34.7
2019	3.71	6.39	-21.86	34.7
2020		8.43	-19.61	36.5
2021		7.03	-20.45	34.5
2022		6.08	-24.84	37
2023		5.82	-25.88	37.5
2024		5.56	-25.98	37.1
2025		5	-26.29	36.3
2026		4.36	-26.92	35.6
2027		3.82	-27.63	35.3
2028		3.35	-28.06	34.8



Showing residuals of ACF and PACF

The mean CFR forecasted values for next 10 years have been calculated and the results are presented in Table 7, along with the observed records. From the figure 15, the projection shows a decreasing trend of case fatality rates of Influenza A (H1N1) after 2019 till 2029, and it is observed that the forecasted values are closer to that of the actual values and fall within the confidential limits (95%).



Forecast model of Case Fatality Rate

IV. Discussion

The respiratory infection popularly known as swine flu is caused by an influenza virus. An H1N1 subtype pandemic

occurred in 1918, causing millions of deaths. H1N1 reappeared in humans in 1976, and more recently in 2009. It was first recognized in spring 2009, near the end of the usual Northern Hemisphere flu season. The new virus, 2009 H1N1, spreads quickly and easily. A few months after the first cases were reported, rates of confirmed H1N1-related illness were increasing in almost all parts of the world. WHO declares on June 11, 2009, that H1N1 (Swine-influenza A) is pandemic. There have been nearly 30,000 confirmed H1N1 cases across 74 countries [11]. It has been noted that around 2% of infected individuals in some countries have developed severe disease, and that this has occurred most commonly in those aged 30–50 years. In March 2009, novel 2009 influenza A (H1N1) was first reported in the south-western United States and Mexico. The population and health care system in Mexico City experienced the first and greatest early burden of critical illness. Critical illness occurred in 58 of 899 patients (6.5%) admitted to the hospital with confirmed, probable, or suspected 2009 influenza (A) H1N1. Critical illness from 2009 influenza A (H1N1) in Mexico occurred in young individuals, was associated with severe acute respiratory distress syndrome and shock, and had a high case-fatality rate. Around 12, 3397 people have been tested in India as on February 1, 2010. In India, 23.3% of people who have tested for swine flu are found suffering from swine flu. Also around 4% of people who have tested positive for swine flu have died and could not be saved in India. Maharashtra with 317 deaths has seen maximum deaths from swine flu in India. Gujarat also has 242 deaths from swine flu till February 1, 2010. Indian states suffering from are Rajasthan (176 deaths), Karnataka (141 deaths), and Delhi (93 deaths). Delhi has the highest number of patients who are tested positive for swine flu, with 9,652 people tested positive. Maharashtra has 5,116 cases of swine flu affected patients and Rajasthan has 2,101 patients who were tested positive for swine flu till February 1, 2010. The states which are severely affected by swine flu have high exposure to foreigners¹. The new influenza A (H1N1) virus is genetically related to swine flu viruses, and differs substantially from circulating human influenza viruses. It is able to spread from person to person. Because it is a completely new virus, there is probably little immunity in the population [12]. This pandemic of influenza A (H1N1) has shown mild to moderate severity. Overall case-fatality (CFR) estimates ranged from 0.054–0.086 per 1000 persons with nH1N1-associated ILI and were highest for persons ≥65 years (0.094–0.147 per 1000) and lowest for those 0–17 (0.008–0.012). Hygiene measures and administration of antiviral drugs to patients and their contacts may slow the spread [12].

Forecasting the dynamics of influenza outbreaks could be useful for decision-making regarding the allocation of public health resources. Reliable forecasts could also aid in the selection and implementation of interventions to reduce morbidity and mortality due to influenza illness [13]. For diseases that show recurrent seasonal patterns or occur in cyclic patterns, time series models are the most widely used statistical models by health researchers for forecasting [14]. Time series forecasting is commonly performed using autoregressive integrated moving average (ARIMA) models¹⁴ that can accommodate both trend and seasonal variations. ARIMA models are typically selected by maximizing some measure of predictive accuracy. A study was conducted to identify circulating influenza virus strains and vulnerable population groups and investigate the distribution and

seasonality of influenza viruses in Ningbo, China. Then, an autoregressive integrated moving average (ARIMA) model for prediction was established. Influenza surveillance data for 2006–2014 were obtained for cases of influenza-like illness (ILI) ($n = 129,528$) from the municipal Centres for Disease Control and virus surveillance systems of Ningbo, China. The ARIMA model was proposed to predict the expected morbidity cases from January 2015 to December 2015. Of the 13,294 specimens, influenza virus was detected in 1148 (8.64%) samples, including 951 (82.84%) influenza type A and 197 (17.16%) influenza type B viruses. ARIMA (1, 1, 1) (1, 1, 0)₁₂ model could be used to predict the ILI incidence in Ningbo. The seasonal pattern of influenza activity in Ningbo tended to peak during the rainy season and winter. Given those results, the model we established could effectively predict the trend of influenza-related morbidity, providing a methodological basis for future influenza monitoring and control strategies in the study area [15].

4.1. Limitations

Missing observations are as frequent as variable time intervals between measurements. While regular time intervals can be accomplished by adaption of methods, missing values have to be bridged through interpolation or by the application of a Kalman filter. A limited number of data points can mean a serious restriction to the applicability of time series analysis. The number of available measurements can be compromised both by the critical state of the respective patient and the thus limited time for measurement, and by methodological and organizational problems (e.g. laboratory). From a conservative point of view, more than 50 observations are needed for a reliable estimation of an ARIMA model. But more recent simulation experiments show that a number of 20 data points can be sufficient.

V. Conclusion

This study was able to analyse the variations comprising the case fatality rates of the Influenza A (H1N1). The formulation, estimation and forecasting of time series ARIMA model was built from a pragmatically point of view based on the available Influenza A (H1N1) data. The results indicate that the proposed ARIMA (2, 1, 3) model can forecast with a satisfactory goodness of fit and performance with the precision. The model was selected for its minimum value of normalized BIC, MAPE and good R-Square among all other models. Application of this model, could successfully forecast the decrease in the case fatality rate of Influenza A H1N1 for next 10 years. This model is recommended to the metropolitan health directorate and re-searchers who would want to monitor the incidence of Influenza A (H1N1) outbreaks in the metropolis and other parts of the world.

Conflict of interest

The authors do not have any conflict of interest with other entities or researchers.

Acknowledgement

The authors sincerely thank Head of the Department, Department of Water and Health, JSS Academy of Higher Education and Research, Mysore, for having granted the use of SPSS software for performing the statistical analysis.

References:

[1] Dandagi GL, Byahatti SM. An insight into the swine-influenza A (H1N1) virus infection in humans. *Lung*

India: Official Organ of Indian Chest Society. 2011 Jan; 28(1):34.

- [2] Murhekar M, Mehendale S. The 2015 influenza A (H1N1) pdm09 outbreak in India. *The Indian journal of medical research*. 2016 Jun; 143(6):821.
- [3] Case fatality rate (<http://www.britannica.com/science/case-fatality-rate>). Accessed May 2019.
- [4] Thematic Map (http://www.researchgate.net/publication/255398560_Thematic_Map). Accessed May 2019.
- [5] Kis M. Analysis of the time series for some causes of death. *Studies in health technology and informatics*. 2002; 90:439-43.
- [6] Petukhova T, Ojkic D, McEwen B, Deardon R, Poljak Z. Assessment of autoregressive integrated moving average (ARIMA), generalized linear autoregressive moving average (GLARMA), and random forest (RF) time series regression models for predicting influenza A virus frequency in swine in Ontario, Canada. *PloS one*. 2018;13(6)
- [7] Anwar MY, Lewnard JA, Parikh S, Pitzer VE. Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malaria journal*. 2016 Dec;15(1):566 (7)
- [8] Kucharski AJ, Edmunds WJ. Case fatality rate for Ebola virus disease in West Africa. *The Lancet*. 2014 Oct 4; 384(9950):1260.
- [9] Paul S, Mgbere O, Arafat R, Yang B, Santos E. Modeling and forecasting influenza-like illness (ILI) in Houston, Texas using three surveillance data capture mechanisms. *Online journal of public health informatics*. 2017; 9(2).
- [10] Singh M, Sharma S. An epidemiological study of recent outbreak of influenza A H1N1 (Swine Flu) in Western Rajasthan region of India. *Journal of Medical & Allied Sciences*. 2013 Aug 31; 3(2):48.
- [11] Khan MS, Akbar SM, Hossain ST, Mahtab MA. Swine influenza (H1N1) pandemic: developing countries' perspective.
- [12] Koopmans MP, Meijer A, Boucher C, Fouchier RA, Osterhaus AD, Timen A. Combating the new influenza A (H1N1) virus. I. Overview of the relevant virological aspects. *Nederlandstijdschriftvoorgeneeskunde*. 2009; 153:A770.
- [13] Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and other respiratory viruses*. 2014 May;8(3):309-16.
- [14] Sym D, Patel PN, Ei-Chaar GM. Seasonal, avian, and novel H1N1 influenza: prevention and treatment modalities. *Annals of pharmacotherapy*. 2009 Dec; 43(12):2001-11.
- [15] Wang C, Li Y, Feng W, Liu K, Zhang S, Hu F, Jiao S, Lao X, Ni H, Xu G. Epidemiological features and forecast model analysis for the morbidity of influenza in Ningbo, China, 2006–2014. *International journal of environmental research and public health*. 2017 Jun;14(6):559