Remittances Review
August 2024,
Volume:9,No:4,pp. 3437-3451
ISSN:2059-6588(Print)|ISSN2059-6596(Online)
Received: 25 July 2024, Accepted: 28 August 2024
DOI: https://doi.org/10.47059/rr.vx9i2.256

Estimating the Prevalence of COVID-19 in Pakistan

Muhammad Umair Sohail[1], Nasreen Akhter[2], Ayesha Sultan[3], Fariha Sohil*[3], Eman[3], Maimona Tofeeq[3]

[1]Department of Statistics, Emerson University Multan, Multan, Pakistan
[2]Department of Special Education, The Islamia University Bahawalpur, Bahawalpur Pakistan
[3]Department of Education, The Women University Multan, Multan, Pakistan
[*]Corresponding Author: drfariha.6493@wum.edu.pk

Abstract

The repeated waves of COVID-19 and their various versions create the panic situation of how long we have to stay with COVIS-19. Estimating the prevalence of COVID-19 is the urgent need of time to rearrange available options (locally and globally) and control the widespread of the virus. In this scenario, time series models provide a precise picture to understand the prevalence of COVID-19 and engage all stakeholders to participate in this health emergency. The present study discusses the auto-regressive Integrated Moving Average (ARIMA) model developed to measure the predicted trend of COVID-19 in Pakistan. The prevalence COVID-19 data was collected from January, 1, 2021 to September, 30, 2021 from the National Command and Operation Center (NCOC)website. For better prediction, various ARIMA models are constructed at various parametric values of ARIMA. The model which has the lowest MAPE values is selected as the best model at ARIMA (1,1,3). Further, this investigation also suggested that the suggested ARIMA modelestimate the COVID-19 prevalence efficiently. The statistical finding also suggests better understanding the pattern, trend, and future values of COVID-19 patients inside Pakistan. Thus, it also facilitated the policy makers of the country and also our region to rearrange the available resources and establish the consensus between all stakeholders to participate significantly for the local and global epidemiological stability.

Keywords: ARIMA, COVID-19, Estimation, Prediction, Prevalence, Time series analysis.

1 Introduction

In the 21th century, the wide and repeated spreadness of COVID-19 create the health emergency and panic situation among public. In the late 2019, the spreadness of COVID-19 is started from the Wuhan City, China. Currently, there is no geographical zone on earth where the COVID-19 patients are not found. Due to the dynamic aspect of COVID-19 virus, it a challenging task to control/minimize the spreadness of novel virus. As, October, 3, 2021, approximately 219 million confirmed cases and 4.55 million deaths worldwide. The propensity of spreadness is varies worldwide due to geographical, cultural, and epidemiological diversity, but, no one is in the safe hands. This novel virus affected everyone beside the ethnicity, race, and geographical status. Recently, various COVID-19 vaccine such

Moderna, Pfizer, Sinovac, Sinophram, etc. are available in the market to control the widespread of COVID-19. Unfortunately, the low response and acceptability of public toward COVID-19 vaccine create the challenging environment to control the novel virus. So, its is need of the time to estimate the future prevalence of COVID-19 for the better arrangement and manipulation of available of option to stay with virus until everyone is not vaccinated. For this purpose, various statistical models facilitate us to estimate the future pattern of novel virus and helps epidemiologists for the timely and significantly manage healthcare infrastructure.

From the last few years, following Statistical models are used to predict the pattern of epidemiological cases, such as given in Table 1.

Table 1: Recently used statistical model for prediction

Sr. No.	Author(s)	Statistical Tool
1.	Kubalija et al. (2014)	Time Series Models
2.	Thomson et al. (2006)	Multivariate linear regression
3.	Wang et al. (2018a), Zhang et al. (2017), and Zhang et al. (2013)	Grey forecasting model
4.	Liu et al. (2019), Ren et al. (2013), and Zhang et al. (2013)	Backpropagation neural network
5.	Nsoesie et al. (2013) and Orbann et al. (2017)	Simulation model

The prevalence of epidemics is affected by many factors. Thus, the overall pattern is recognized by the tendency and their randomness. Therefore, the given statistical models are unable to capture the significant factors such as randomness and it's quite laborious to generalize.

The ARIMA model is commonly used in the field of epidemiology due to its simple mathematical computation andsignificant applicability along withcapacity to describe the randomness. Table 2, the importance and popularity of the ARIMA model is highlighted by citing the various authors who used it to estimate influenza diseases.

Table 2: Studies based on ARIMA model

Sr. No.	Author(s)	Statistical Tool
1.	Guan et al. (2004)	ARIMA and Artificial neural networks
2.	Earnest et al. (2005)	ARIMA
3.	Gaudart et al. (2009)	ARIMA
4.	Liu et al. (2011)	ARIMA
5.	Ren et al. (2013)	ARIMA and Backpropagation Neural Networks
6.	Zhang et al. (2015)	ARIMA
7.	Wu et al. (2015)	ARIMA, Generalized Regression Neural Networks, and Nonlinear Autoregressive Neural Networks

8.	Wei et al. (2016)	ARIMA and Generalized Regression Neural Networks
9.	Sun et al. (2018)	ARIMA, Negative Binomial Regression Model, and Generalized Additive Model
10.	Wang et al. (2018b)	Seasonal ARIMA and Nonlinear Autoregressive Network
11.	Wu et al. (2019)	ARIMA, ERNN, JNN,
12.	Chen et al. (2020)	Seasonal ARIMA
13.	Fang et al. (2020)	ARIMA/X Models and Random Forecast
14.	Polwiang et al. (2020)	ARIMA, Artificial neural networks, and Multivariate Poisson Regression
15.	Cao et al. (2020)	ARIMA

Thus, the ARIMA model is the most popular statistical package to estimate the time series prediction, temporal dependency, time series and random changes in time series pattern about the pandemic disease. Due to the simple mathematical background of the ARIMA model, it is very easy to understand, select a model and make a decision based on the best selected model.

Recently, several other statistical tools are used to predict the prevalence of COVID-19 in China, such as Li et al. (2020) use the time series analysis to estimate the running pattern and outbreak of COVID-19 in China. Fanelli and Piazza (2020) investigate the temporal dynamic of COVID-19 prevalencein China, Italy, and France. Anastassopoulou et al. (2020) discuss the critical analysis of epidemiological parameters and model the wide spreadness of COVID-19 in Hubei, China. Wang et al. (2020) proposed the patient's information-based algorithm to model the death rate by COVID-19 from the publicly available data.

The objective of this research is to estimate the wide spreadness of COVID-19 in Pakistan. The data analysis in this study is obtained corresponding to the period January, 1, 2021 to September, 30, 2021. The statistical analysis of the data set is carried by applying the ARIMA models. The model based simple quantitative models are used to provide a brief realistic description about the wide prevalence and intensity of pandemic. The statistical finding helps to predict the future need of the healthcare system and facilitate the patients on time.

2 Research Methodology

2.1 Data Collection

For this study, the data about the COVID-19 is obtained from the official NCOC website (https://ncoc.gov.pk). The data analysis is carried out by using the R-language statistical software. In Table 3, the descriptive statistics for the above said time period is defined.

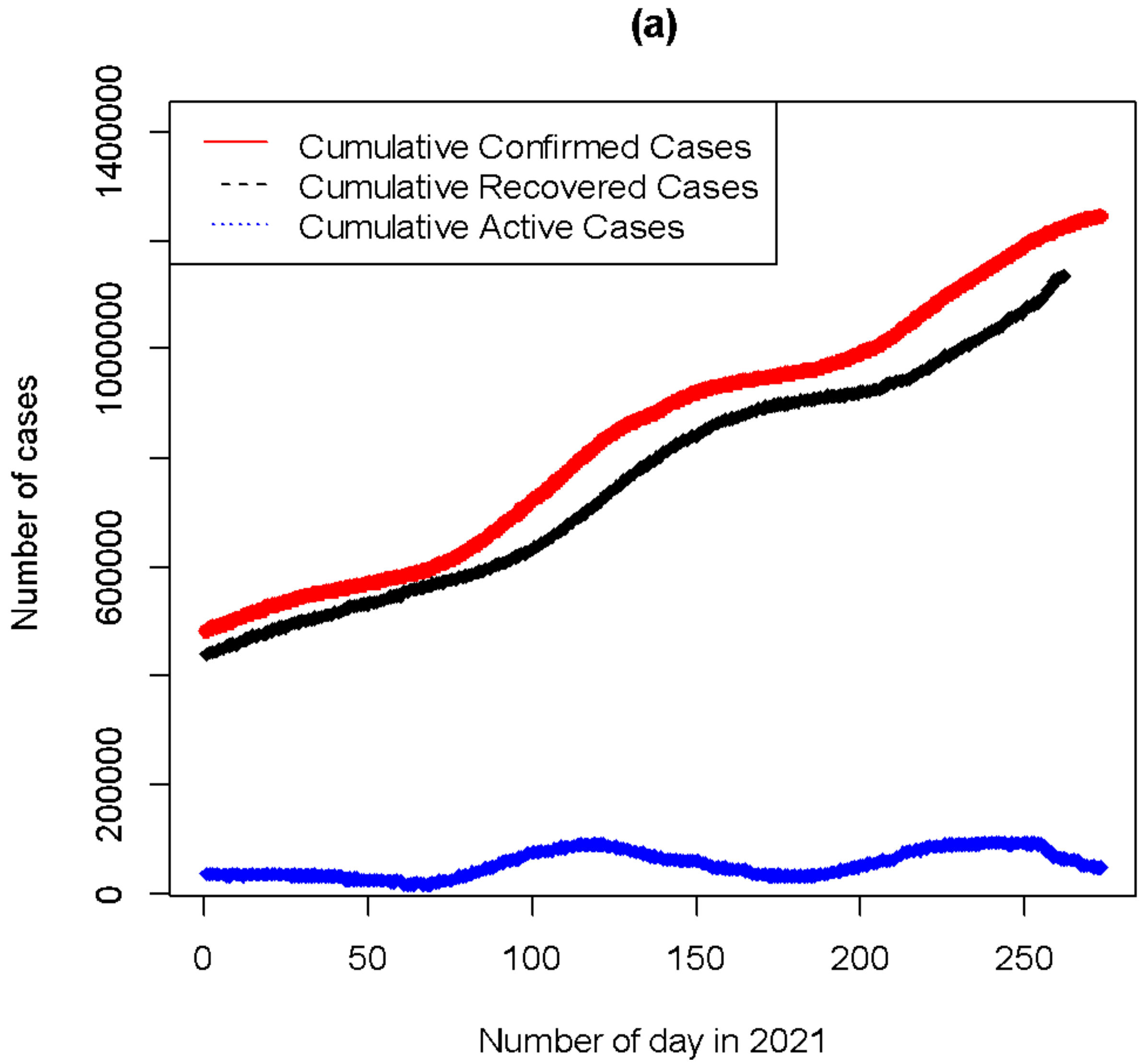
Table 3: Descriptive statistics of study variable

Variable of Interest	**Min.**	Q_1	**Median**	**Mean**	Q_3	**Max.**

Active Cases in last 24 hours	16349	33316	48708	54010	78553	94583
Deaths in last 24 hours	11.00	40.00	58.00	65.15	81.75	201.00
Recovered cases in last 24 hours	543	1388	2168	2663	3636	13716

In this study, the time series having 273 values is used to predict the COVID-19 prevalence in Pakistan with 80%-90% confidence interval.

In Figure 1, The COVID-19 prevalence is started in Pakistan at the end of February, 2020. In March, 2020, the government of Pakistan imposed the sudden lockdown which created panic and anxiety among the public about the spreadness of COVID-19. This is the first time in Pakistan, they have imposed a lockdown all over the country to control the prevalence of COVID-19. In February, 2020- September, 2021, overall 1246538 confirmed cases are reported while 1170590 are recovered with 94% recovery rate.



Volume:9,No:4,pp. 3437-3451
ISSN:2059-6588(Print)|ISSN2059-6596(Online)

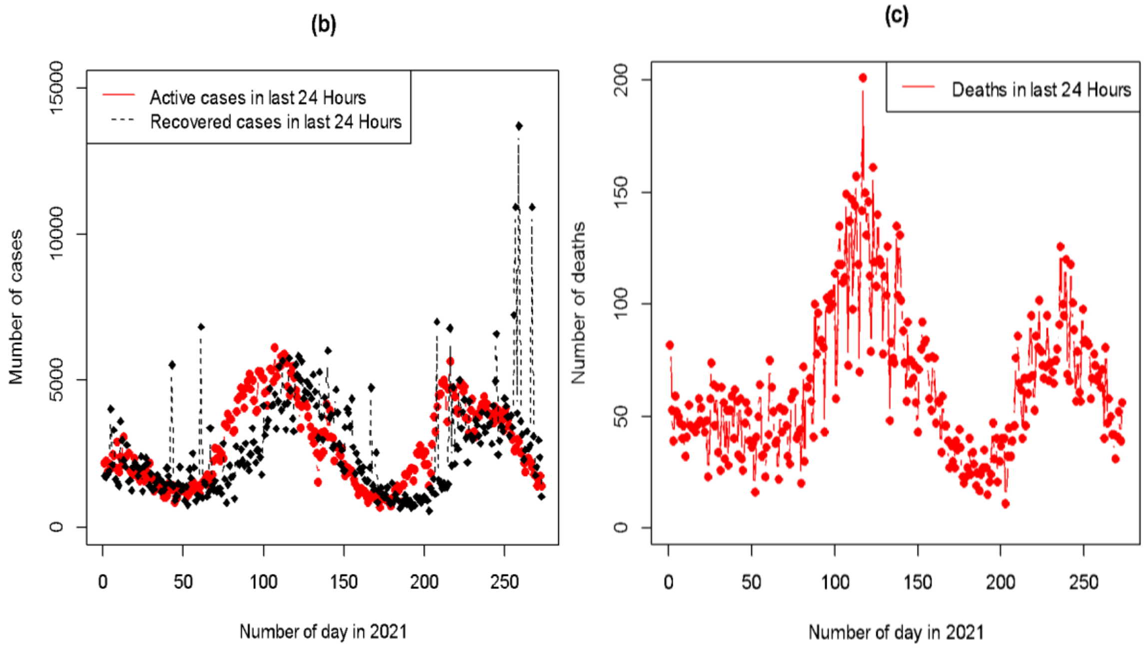


Figure 1: The prevalence of COVID-19, recovered, and deaths in 2021

2.2 ARIMA Model

Fanoodi et al. (2019) examine various statistical models, time series model is one them to order the data points with respect to time. Time series models aim to predict future response of the variable of interest by using the prior or historical response of it (Liu et al., 2011). Box and Jekins (1970) proposed the pioneer idea of ARIMA model to estimate the time series. The ARIMA is the one of the best models to capture the trend, periodic movements, and random disturbance is the time series. ARIMA model preferred in every data set which may include seasonality, trend, cycle fluctuations, and temporal dependency structures. The ARIMA models provide the clear picture of temporal dependency structure in the time series.

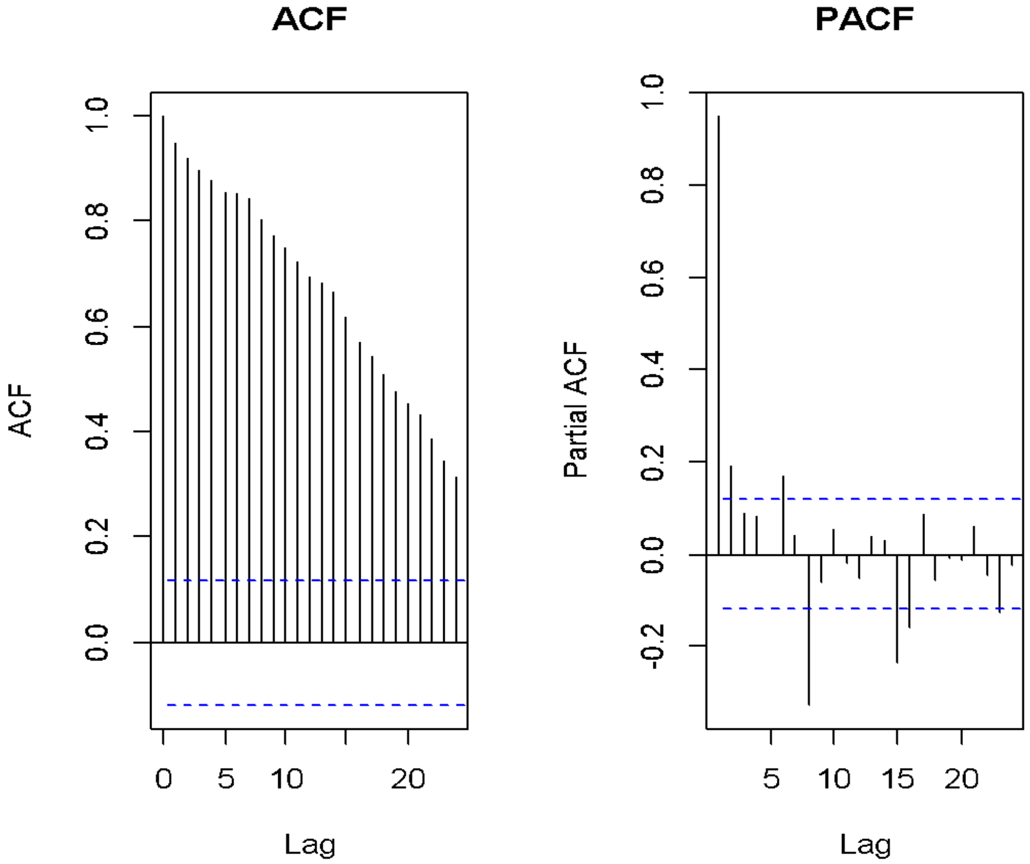


Figure 2: Estimated value of ACF and PACF

The ARIMA (p,d,q) model have three parameters such as order of autoregression (p), degree of difference (d), and the order of the moving average (q).

The ARIMA model can be decomposed into different AR and MA models. In AR (p) model shows the linear association between the current (Y_t) and previous value $(Y_{t-1},Y_{t-2},Y_{t-3},...,Y_{t-p})$ of the observed variable. While in MA (q) refer the linear relationship between the current value (Y_t) and the previous residual series $(e_{t-1},e_{t-2},e_{t-3},...,e_{t-q})$. The generalized expression for AR (p) and MR (q) is given by

$$Y_t = \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + \lambda_3 Y_{t-3} + ... + \lambda_p Y_{t-p} + e_t \tag{1}$$

$$Y_t = \eta_1 e_{t-1} + \eta_2 e_{t-2} + \eta_3 e_{t-3} + ... + \eta_q e_{t-q} + e_t$$

where

$\lambda =$ auto-regressive parameter.

$\eta =$ moving average parameter.

$Y_t =$ observed value at time t.

$e_t =$ random error at time t.

The term e_t is independent identically distributed (IID) with mean 0 and variance σ^2. Further, the ARMA (p,q) is compose of AR (p) and MA (q), is given by

$$Y_t = \beta + \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + \lambda_3 Y_{t-3} + ... + \lambda_p Y_{t-p} + e_t - \eta_1 e_{t-1} - \eta_2 e_{t-2} - \eta_3 e_{t-3} - ... - \eta_q e_{t-q} \tag{3}$$

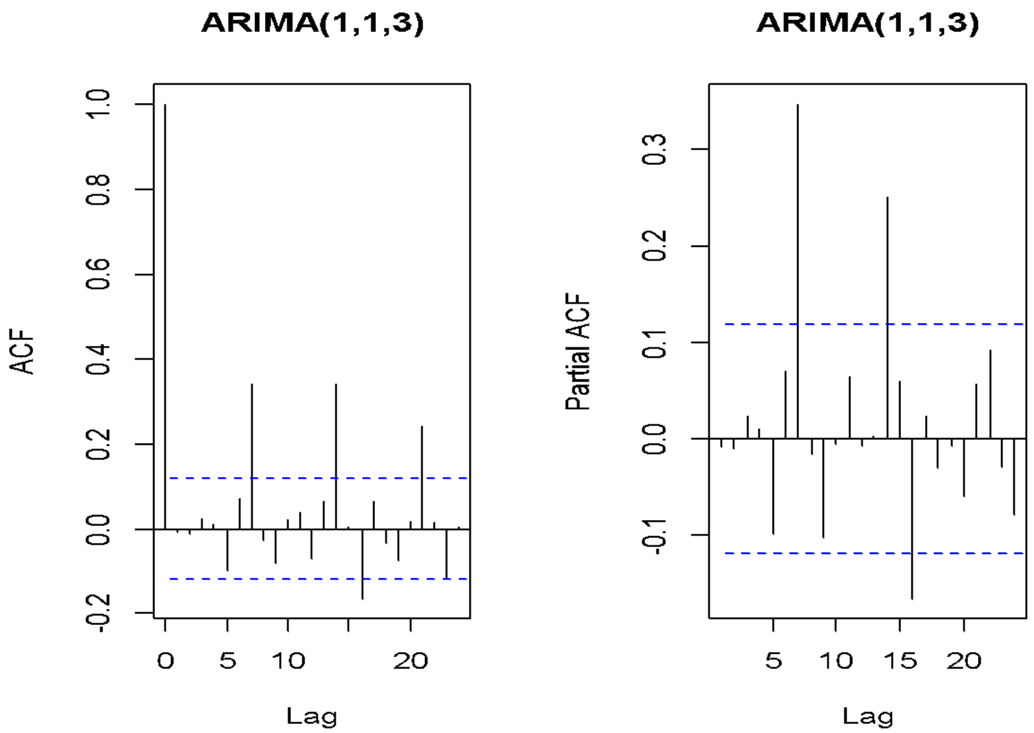
where

$\beta =$ constant value.

$e_{t-1} =$ previous random error term.

He and Tao (2018) discussed the different time series with the help of ARMA models to perform ARIMA models.

(a)



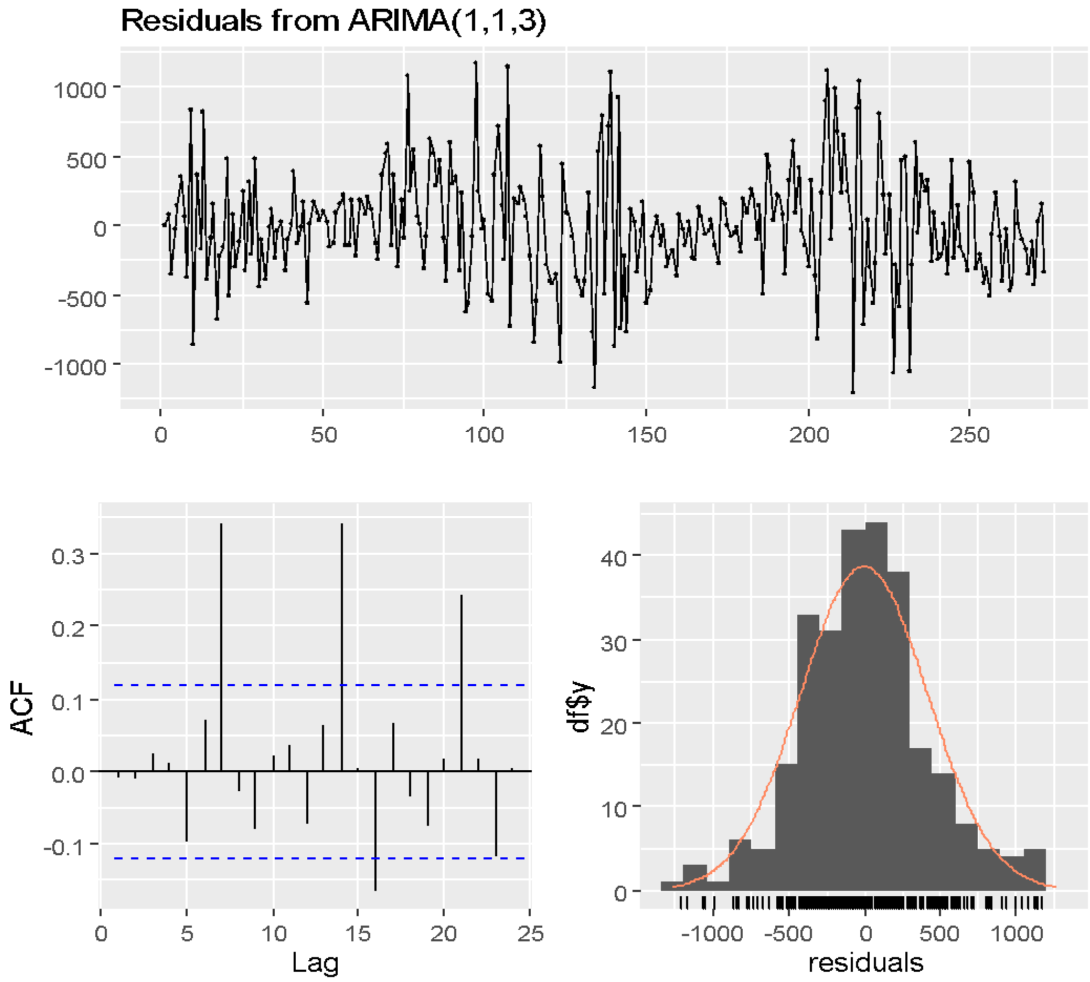


Figure 3: The estimated autocorrelation and partial autocorrelation illustration for the prevalence of COVID-19 in Pakistan

2.3 Selection of Model

The selection of the suitable model to predict the future is measured by comparing the observed values and predicted values. The predictive accuracy for ARIMA models is analyzed by Root Mean Square Error (RMSE), Mean Absolute Error, and Mean Absolute Percentage Error. These are given below as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)} \quad (3)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| \quad (4)$$

and

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \quad (5)$$

where

$Y_t =$ observed value

$\hat{Y}_t =$ estimated value

$n =$ total time points

That model which has the smaller value of RMSE, MAE, and MAPE has the highest value of predictive accuracy. The whole data analysis is performed by using the R-language (version 4.1.1) software along with $\alpha <$ 0.01 and 0.001.

Table 4: Related Assessment of ARIMA models

Model	**RMSE**	**MAE**	**MAPE**
ARIMA (0,1,0)	435.1012	326.4805	13.1193
ARIMA (0,1,2)	420.1082	322.2382	12.79715
ARIMA (0,1,1)	422.3702	325.5986	12.92182
ARIMA (1,1,2)	419.768	321.6475	12.7886
ARIMA (0,1,3)	419.5283	321.0077	12.7767
ARIMA (1,1,1)	419.8174	321.9527	12.80132
ARIMA (1,1,3)	418.401	319.5657	12.75483

Table 5: ARIMA (1, 1, 3) model parameters

Parameter	Coefficients	S. E	Z-value	P-value	C. I (2.5%)	C. I (97.5%)
AR (1)	-0.5450	0.2132	-2.5563	0.0106 *	-0.9630	-0.1272
MA (1)	0.3051	0.2133	1.4304	0.1526	-0.1129	0.7232
MA (2)	-0.2146	0.0755	-2.8421	0.0045**	-0.3626	-0.0666
MA (3)	-0.1423	0.0697	-2.0424	0.0411*	-0.0057	-0.0057
significance level: 0.001** and 0.01*						

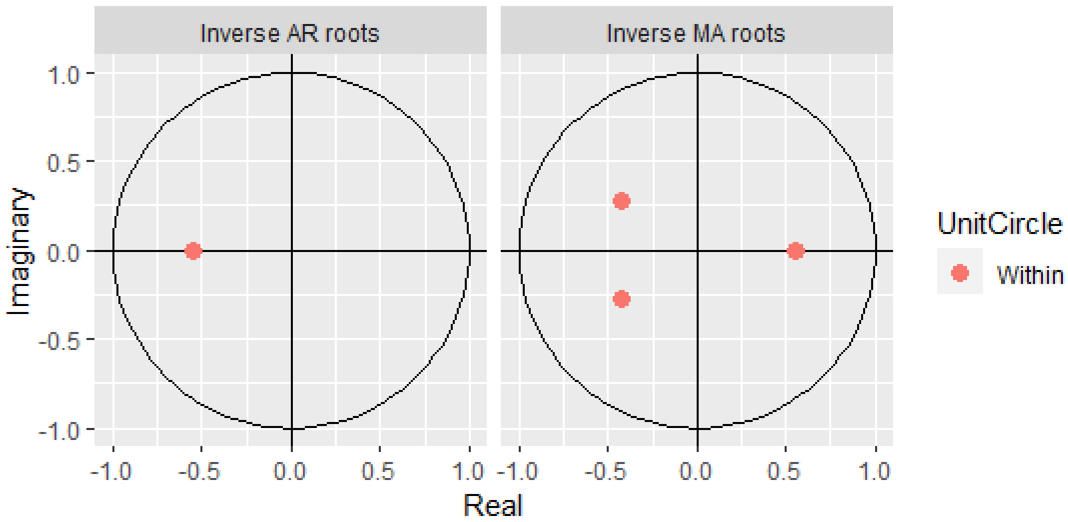


Figure 4: Inverse AR and MA roots for the best ARIMA model

3 Interpretation of Statistical Findings

3.1ARIMA model for the prediction COVID-19 prevalence

In ARIMA model, four iterative step are involved to for modeling the whole procedure such as:

i. Assessment
ii. Parameter estimation
iii. Diagnostic checking
iv. Prediction

The seasonality and stationary in the time series is accessed at the first phase in the ARIMA modeling, if the mean (μ), variance (σ^2), and autocorrelation (ρ_k) is constant over time then the time series is said to be stationary. In the ARIMA model, ACF and PACF plots are helpful to measure the existence of stationary and seasonality in the time series. The ACF graph checks the autocorrelation between the two successive values in the time series, while the PACF graph measures the degree of association between the variable and their respective lag which are not examined by the correlation over small value of lag (He and Tao, 2018). In Figure 2, the estimated time series is shown up to January, 01, 2021 to September, 30, 2021. The straight line of the plot shows the two standard deviation limits and measures the zero correlation. The lines beyond the limits show the significance of autocorrelation for COVID-19 data set. There is no seasonal variation in the used COVID-19 time series, as given in Figure 1 and 2. The ACF plot in Figure 2 shows that there is a lack of stationarity because the autocorrelation decreases slightly. Thus, the first lag difference is taken to stabilize the time series. In the second step, the parameter of the ARIMA model is determined with ACF and PACF plots. Figure 3 shows that the mean of the residual is equal to zero and there is no correlation among residuals. The histogram shows that the residuals are normally distributed and the forecast and prediction interval based on this model is quite good. For the selection of suitable ARIMA models, different models were created at the various parametric values. The validity and reliability of the model is measured with the help of RMSE, MAE, and MAPE. In Tab32100

le 4, the ARIMA (1,1,3) model was selected as the best model for predicting the COVID-19 prevalence. The model has the minimum RMSE = 418.401, MAE = 319.5657, and MAPE = 12.75483 (see Table 4 and Figure 3). Figure 4 indicates that the Inverse AR and MA root inside the circle, which mean the ARMA is stable and provide the precise estimates. The parameter coefficients for the best model are given in Table 4. The p-values of the given parameter is smaller than α, so we conclude that their values are considerably different from zero at the 99% and 99.9% confidence level, while the second parameter which has the value zero (insignificant). In Figure 4, the observed and predicted values are illustrated, respectively.

Table 6: Predictive case in next 24 hours with 80% and90% confidence interval

Days	Point	80% C. I	90% C.1

	Forecast	Low Limit	Upper Limit	Low Limit	Upper Limit
274	1453.205	916.0179	1990.392	763.73277	2142.677
275	1480.312	805.5809	2155.043	614.30402	2346.320
276	1513.225	746.9207	2279.530	529.68406	2496.766
277	1495.285	668.1425	2322.428	433.65911	2556.911
278	1505.064	610.9013	2399.226	357.4187	2652.709
279	1499.734	548.6029	2450.864	278.9706	2720.497
280	1502.639	495.0207	2510.257	209.3749	2795.903
281	1501.055	441.3798	2560.731	140.9765	2861.134
282	1501.919	391.8848	2611.952	77.20570	2926.631
283	1501.448	343.6315	2659.265	15.40654	2987.490

Table 6 indicates the next 10-days prediction of COVID-19 new cases in Pakistan. The minimum value is 1453-1513 new COVID-19 cases might be reported in the next 10 days. Figure 5, shows the forecast response for the next 10 days with 80% and 90% confidence interval.

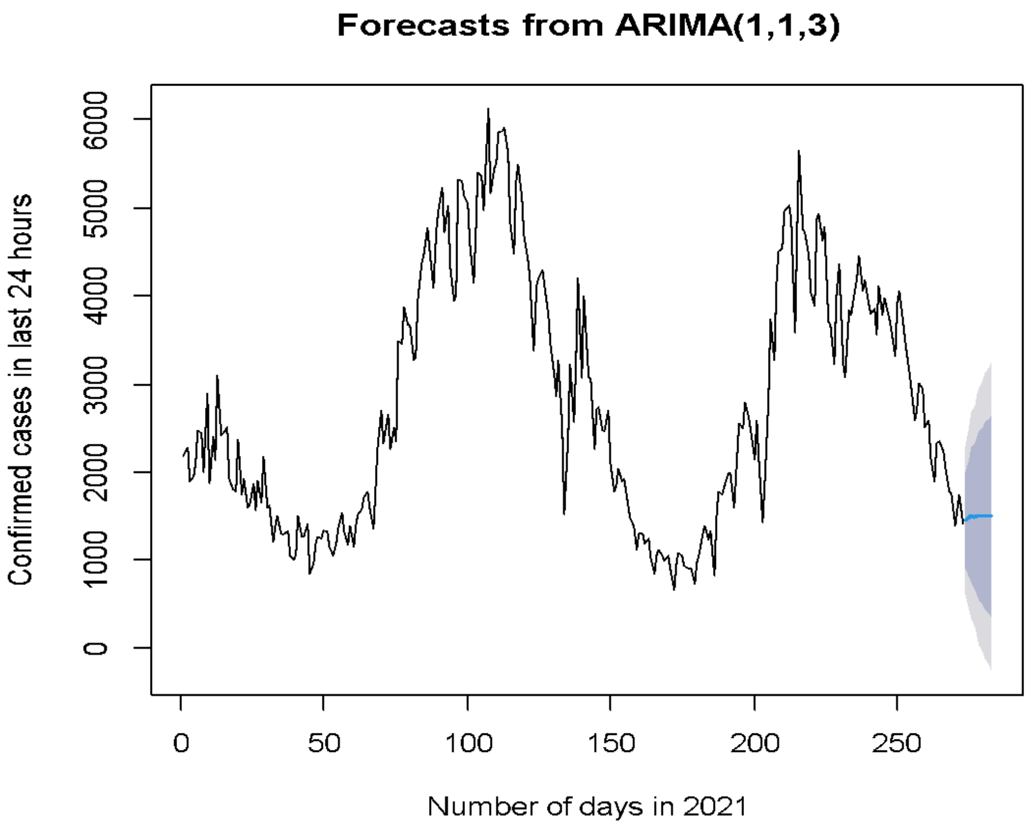


Figure 5: Time Series plot for forecast values

4. Discussion

The only possible ways to control the high prevalence of COVID-19 and repeated waves of pandemic demands are effective planning, administrative support, rearrangement of the available resources, and active participation of stakeholders to restore the economic, social, cultural, and religious activities normal again. Thus, that is the need of time to develop a reliable forecasting model to predict the future response of COVID-19 and help out the policy maker to arrange their options accordingly. In this scenario, time series models are more reliable to estimate the predictive value more precisely by capturing the dynamic aspects of the given phenomena.

In the present study, the prevalence of COVID-19 is investigated with ARIMA (1,1,3) model. According to our knowledge, the presentstudy is the pioneer to estimate the COVID-19 prevalence in Pakistan from South-Asia by ARIMA model. Currently, the healthcare infrastructure in Pakistan is struggling to facilitate the COVID-19 patients. From the last few days, the gradual decrease in the prevalence of COVID-19 is observed because of lockdown in the past month. As we observed the time series of the first nine months, the prevalence of COVID-19 is increasing as the authorities remove the lockdown. Currently, various COVID-19 vaccines are available in the market but the low acceptability rate of COVID-19 vaccine again creates the problem for the authorities. The local authorities tried to ensure the vaccination of everyone beside their religious, geographical and provincial affiliation. Vaccination and follow of COVID-19 precaution measure is the only option to control the pandemic.

5. Conclusion

Due to limited healthcare facilities and low gross domestic production (GDP) growth rate, forecasting the COVID-19 prevalence is very important in low income countries to optimize their available resources and healthcare infrastructure. For the precise prediction, time series tools are best to predict the response of variables of interest. In this investigation, ARIMA models are used to estimate the widespread of pandemic in Pakistan and to facilitate the policy makers and NCOC officials to reopen social activities or impose lockdown, manage healthcare facilities, and restore economic activities over the coming weeks. For the better forecast about the prevalence of pandemic more update time series might be used to understand the future response of novel virus.

Reference

Anastassopoulou, C., Russo, L., Tsakris, A., Siettos, C., 2020. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS One 15, e0230405. https://doi.org/ 10.1371/journal.pone.0230405.

Chen, Y., Leng, K., Lu, Y., Wen, L., Qi, Y., Gao, W., Chen, H., Bai, L., An, X., Sun, B., Wang, P., Dong, J., 2020. Epidemiological features and time-series analysis of influenza incidence in urban and rural areas of Shenyang, China, 2010-2018. Epidemiol. Infect. 148, e29. https://doi.org/10.1017/S0950268820000151.

Earnest, A., Chen, M.I., Ng, D., Leo, Y.S., 2005. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. BMC Health Serv. Res. 5, 1–8. https://doi.org/10.1186/1472-6963-5-36.

Fanelli, D., Piazza, F., 2020. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons and Fractals 134, 1–12. https://doi.org/10.1016/j.chaos.2020.109761.

Fang, X., Liu, W., Ai, J., He, M., Wu, Y., Shi, Y., Shen, W., Bao, C., 2020. Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. BMC Infect. Dis. 20, 1–8. https://doi.org/10.1186/s12879-020-4930-2.

Fanoodi, B., Malmir, B., Firouzi, F., 2019. Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. Comput. Biol. Med. 113, 103415. https://doi.org/10.1016/j.compbiomed.2019.103415.

Gaudart, J., Touré, O., Dessay, N., Dicko, A.L., Ranque, S., Forest, L., Demongeot, J., Doumbo, O.K., 2009. Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali. Malar. J. 8. https://doi.org/10.1186/1475-2875-8-61.

Guan, P., Huang, D.S., Zhou, B. Sen, 2004. Forecasting model for the incidence of hepatitis a based on artificial neural network. World J. Gastroenterol. 10, 3579–3582. https://doi.org/10.3748/wjg.v10.i24.3579.

He, Z., Tao, H., 2018. International Journal of Infectious Diseases Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: a nineyear retrospective study. Int. J. Infect. Dis. 74, 61–70. https://doi.org/10.1016/j.ijid.2018.07.003.

Kurbalija, V., Ivanovi, M., Schmidt, D., Lindemann, G., Trzebiatowski, V., Burkhard, H., Hinrichs, C., 2014. Time-series analysis in the medical domain: a study of Tacrolimus administration and influence on kidney graft function. Comput. Biol. Med. 50, 19–31. https://doi.org/10.1016/j.compbiomed.2014.04.007.

Li, X., Zhang, C., Zhang, B., Liu, K., 2019. A comparative time series analysis and modeling of aerosols in the contiguous United States and China. Sci. Total Environ. 690, 799–811. https://doi.org/10.1016/j.scitotenv.2019.07.072.

Li, Q., Feng, W., Quan, Y.H., 2020. Trend and forecasting of the COVID-19 outbreak in China. J. Inf. Secur. 80, 469–496. https://doi.org/10.1016/j.jinf.2020.02.014. Liu, Q., Liu, X., Jiang, B., Yang, W., 2011. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC Infect. Dis. https://doi.org/10.1186/1471-2334-11-218.

Liu, Q., Li, Z., Ji, Y., Martinez, L., Zia, U.H., Javaid, A., Lu, W., Wang, J., 2019. Forecasting the seasonality and trend of pulmonary tuberculosis in Jiangsu Province of China using advanced statistical time-series analyses. Infect. Drug Resist. 12, 2311–2322. https://doi.org/10.2147/IDR.S207809.

Nsoesie, E.O., Beckman, R.J., Shashaani, S., Nagaraj, K.S., Marathe, M.V., 2013. A simulation optimization approach to epidemic forecasting. PLoS One 8. https://doi.org/10.1371/ journal.pone.0067164.

Orbann, C., Sattenspiel, L., Miller, E., Dimka, J., 2017. Defining epidemics in computer simulation models: how do definitions influence conclusions? Epidemics 19, 24–32. https://doi.org/10.1016/j.epidem.2016.12.001.

Polwiang, S., 2020. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003-2017). BMC Infect. Dis. 20, 208. https://doi. org/10.1186/s12879-020-4902-6.

Ren, H., Li, J., Yuan, Z.A., Hu, J.Y., Yu, Y., Lu, Y.H., 2013. The development of a combined mathematical model to forecast the incidence of hepatitis E in Shanghai, China. BMC Infect. Dis. 13, 1–6. https://doi.org/10.1186/1471-2334-13-421.

Roda, W.C., Varughese, M.B., Han, D., Li, M.Y., 2020. Why is it difficult to accurately predict the COVID-19 epidemic? Infect. Dis. Model. 5, 271–281.

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J.M., Yan, P., Chowell, G., 2020. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect. Dis. Model. 5, 256–263. https://doi.org/10.1016/j. idm.2020.02.002.

Sun, J.M., Lu, L., Liu, K.K., Yang, J., Wu, H.X., Liu, Q.Y., 2018. Forecast of severe fever with thrombocytopenia syndrome incidence with meteorological factors. Sci. Total Environ. 626, 1188–1192. https://doi.org/10.1016/j.scitotenv.2018.01.196.

Thomson, M.C., Molesworth, A.M., Djingarey, M.H., Yameogo, K.R., Belanger, F., Cuevas, L.E., 2006. Potential of environmental models to predict meningitis epidemics in Africa. Trop. Med. Int. Heal. 11, 781–788. https://doi.org/10.1111/j.1365-3156.2006.01630.x.

Wang, Y. wen, Shen, Z. zhou, Jiang, Y., 2018a. Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. PLoS One 13, 1–11. https://doi.org/10.1371/ journal.pone.0201987.

Wang, Y., Xu, C., Wang, Z., Zhang, S., Zhu, Y., Yuan, J., 2018b. Time series modeling of pertussis incidence in China from 2004 to 2018 with a novel wavelet based SARIMA-NAR hybrid model. PLoS One 13, 1–23. https://doi.org/10.1371/journal.pone.0208404.

Wang, L., Li, J., Guo, S., Xie, N., Yao, L., Day, S.W., Howard, S.C., Graff, J.C., Gu, T., 2020. J. ur of. Sci. Total Environ., 138394 https://doi.org/10.1016/j.scitotenv.2020.138394. Wei, W., Jiang, J., Liang, H., Gao, L., Liang, B., Huang, J., Zang, N., Liao, Y., Yu, J., Lai, J., Qin, F., Su, J., Ye, L., Chen, H., 2016. Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng County, China. PLoS One 11, e0156768. https://doi.org/10.1371/journal.pone.0156768.

Wu, W., Guo, J., An, S., Guan, P., Ren, Y., Xia, L., Zhou, B., 2015. Comparison of two hybrid models for forecasting the incidence of hemorrhagic fever with renal syndrome in Jiangsu Province, China. PLoS One 10, 1–13. https://doi.org/10.1371/journal.pone.0135492.

Wu, W., An, S.Y., Guan, P., Huang, D.S., Zhou, B. Sen, 2019. Time series analysis of human brucellosis in mainland China by using Elman and Jordan recurrent neural networks. BMC Infect. Dis. 19, 1–11. https://doi.org/10.1186/s12879-019-4028-x.

Wu, J.T., Leung, K., Leung, G.M., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 395, 689–697. https://doi.org/10.1016/S0140-6736(20)30260-9.

Zeng, Q., Li, D., Huang, G., Xia, J., Wang, X., Zhang, Y., Tang, W., Zhou, H., 2016. Time series analysis of temporal trends in the pertussis incidence in Mainland China from 2005 to 2016. Sci. Rep. 6, 1–8. https://doi.org/10.1038/srep32367.

Zhang, X., Liu, Y., Yang, M., Zhang, T., Young, A.A., Li, X., 2013. Comparative study of four time series methods in forecasting typhoid fever incidence in China. PLoS One 8. https://doi.org/10.1371/journal.pone.0063116.

Zhang, L., Wang, L., Zheng, Yanling, Wang, K., Zhang, X., Zheng, Yujian, 2017. Time prediction models for echinococcosis based on gray system theory and epidemic dynamics. Int. J. Environ. Res. Public Health 14. https://doi.org/10.3390/ijerph14030262.

Zhang, S., Diao, M., Yu, W., Pei, L., Lin, Z., Chen, D., 2020. International Journal of Infectious Diseases Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: a data-driven analysis. Int. J. Infect. Dis. 93, 201–204. https://doi.org/10.1016/j.ijid.2020.02.033.

Zheng, Y.L., Zhang, L.P., Zhang, X.L., Wang, K., Zheng, Y.J., 2015. Forecast model analysis for the morbidity of tuberculosis in Xinjiang, China. PLoS One 10, 1–13. https://doi.org/ 10.1371/journa-l.pone.0116832