Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan

Muhammad Yousaf\textsuperscript{a}, Samiha Zahir\textsuperscript{a}, Muhammad Riaz\textsuperscript{a}, Sardar Muhammad Hussain\textsuperscript{b,}\textsuperscript{*}, Kamal Shah\textsuperscript{c}

\textsuperscript{a} Department of Statistics, Quaid-i-Azam University Islamabad 45320, Pakistan
\textsuperscript{b} Department of Mathematical Sciences, Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS), Quetta 87300, Pakistan
\textsuperscript{c} Department of Mathematics, University of Malakand, Chakdara Dir (I), KPK, Pakistan

\textbf{A R T I C L E   I N F O}

Article history:
Received 27 April 2020
Accepted 29 May 2020
Available online 25 May 2020

Keywords:
COVID-19 Pandemic
Confirmed Cases
Deaths
Recoveries
Forecast
ARIMA

\textbf{A B S T R A C T}

In this paper, we have conducted analysis based on data obtained from National Institute of Health (NIH) - Islamabad and produced a forecast of COVID-19 confirmed cases as well as the number of deaths and recoveries in Pakistan using the Auto-Regressive Integrated Moving Average Model (ARIMA). The fitted forecasting models revealed high exponential growth in the number of confirmed cases, deaths and recoveries in Pakistan. Based on our model prediction the number of confirmed cases will be increased by 2.7 times, 95% prediction interval for the number of cases at the end of May 2020 = (5681 to 33079). There could be up to 500 deaths, 95% prediction interval = (168 to 885) and there could be eightfold increase in the number of recoveries, 95% prediction interval = (2391 to 16126). The forecasting results of COVID-19 are alarming for May in Pakistan. The health officials and government should adopt new strategies to control the pandemic from further spread until a proper treatment or vaccine is developed.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

In the mid of December, a viral infection called coronavirus disease 2019 (COVID-19) was initially identified in Wuhan City of China [1] and it is believed that human contracted it from wild animals. It has caused thousands of deaths around the world and World Health Organization (WHO) declared it as pandemic on March 11 [2]. COVID-19 infection lead to respiratory illness and has sign like flu, fever, cough, myalgia, diarrhea, and dyspnea [3]. It is a highly contagious and is transmitted via bodily contacts and a respiratory bead from the infected patients, which is now the main source of transmission of the disease. The existence duration for the virus can be up to 12 hours or even two days on a contacted surface [4]. The fatality rate is higher among the young children and elderly aged group ≥ 60 years [5].

At present (as of April 14, 2020) the infected patient’s population worldwide is recorded as 1,925,877 (1.9 million) with 119,719 deaths and 452,333 recoveries [6], and it is expected that these statistics is increasing exponentially in the upcoming days.

In Pakistan the first two cases of COVID-19 appeared on 26th February 2020, within 48 hours three more cases appeared from different cities around the country and there was no linkage between these patients. Gradually, these cases increased exponentially until 14th April the cases were 5,716 with most cases of Punjab 2,826, 1452 cases, 800 cases, 232 cases, 131 cases, and 43 cases of Sindh, KPK, GB, Baluchistan, Islamabad and AJK respectively. With most effected cases the recoveries were recorded to be 1,378 with 96 deaths [7]. COVID-19 is pandemic and to control its spread, preventive measures be taken. For patients, all body fluids and electrolytes should be continuously checked with vital signs and to avoid further spread, they should be incubated with strict clinical measures under preventive guidelines [8]. The government need to find a strategy to fight this war in timely fashion, such as authorities took further measures of closing borders, suspending community services and schools, minimizing both domestic and international travels until further notice [9]. The purpose of these measures is to limit the chances of physical contacts among people, so that transmission of COVID-19 is controlled, as the incubation period for this virus is relatively longer than other viruses.

Due to the novel nature of the virus, there is greater uncertainty around the decision on optimal time of disappearance of this disease. Therefore, short term forecasting is extremely important even in the slightest hint for predicting the upcoming month.
for the better management of the societal, economical, cultural and public health issues [10]. In the past few months researchers have developed or employed existing mathematical and statistical methods to predict the number of COVID-19 cases and related outcomes. Fractional time delay dynamic system (FTDD) reflects good forecast agreements to the public data [11]. Generalized logistic model (GLM) indicates that the epidemic growth was exponential in China [12]. Based on prediction, situation will be worsened in entire Europe and USA will become the epicenter of new cases during mid of April 2020 [13]. Around 2 million people will be infected by the beginning of May if no measures are taken [14]. Predictions/estimates help to strengthen the strategies in order to prevent the pandemic from worsening. In this research study, we used the available data to forecast the number of confirmed COVID-19 cases, deaths and recoveries in Pakistan for upcoming month. This forecasting is aimed to help government institutions and policy makers as well as public in adopting new strategies and strengthening the existing preventive measures against COVID-19 pandemic. In addition, this study may help in reliving current socio-economic and psychosocial distress caused by COVID-19 among public in Pakistan.

2. Methods

2.1. Data Source

We obtained data for the number of daily accumulated confirmed cases of COVID-19, deaths and recoveries from February 26, 2020 to April 12, 2020, from official reports of the National Institute of Health (NIH) - Islamabad, Pakistan [15]. NIH is an autonomous health research institute of the Ministry of National Health Services of Pakistan, situated in Islamabad; its primary responsibilities include biomedical and health-related research and vaccine manufacturing [15]. Since the 1st case of COVID-19 observed in Pakistan, NIH collects and publishes the daily reports on COVID-19 regularly [15]. The published data includes total number of confirmed cases in Pakistan (Province wise), deaths and recovery from COVID-2019. The analysis is based on aggregated numbers of confirmed cases, deaths and recoveries published online by NIH; therefore, ethical approval is not required for this study.

2.2. Statistical model used for the analysis

The available data is limited and is affected by fluctuations i.e., highly variable cases were reported day by day. As a result, cumulative data is used to predict the number of cases in Pakistan. The cumulative number of COVID-19 confirmed cases, deaths and recoveries are expected to show exponential growth over time. Therefore, we used the simple time series methods of Auto-Regressive Integrated Moving Average (ARIMA) Model [16] to forecast the number of cases, deaths and recoveries for upcoming month. The ARIMA model has higher fitting and forecasting accuracy than exponential smoothing [17]. It captures both the seasonal and non-seasonal forecasting trends. Due to the limited available data, we simply focus on non-seasonal models to describes the pattern (growth) over time. Hence, we assumed that the pattern of current cases will continue in the near future (at least a month). We believe that the ARIMA model, which is the combination of Autoregressive (AR) and Moving Average (MA) fits well to the nature of the available data and provide good forecasting for the short time series data. The forecasting and prediction intervals until the end of May is produced from the fitted model. In order to assess the model fit, parameters (p, d, q) are identified by Autocorrelation function (ACF) and Partial Autocorrelation function (PACF); whereas, p is the autoregressive term, d is the differencing order and q is the moving averages term. Furthermore, ARIMA (p, d, q) results are based upon Akaike information criterion (AIC) which is a goodness of fit test such that model with minimum AIC is considered best. All statistical analyses were conducted using the R-library “forecast, tseries and zoo” [18] developed for fitting ARIMA model.

3. Results

In Pakistan, the number of COVID-19 cases is now increasing exponentially, Fig. 1. As of 12 April 2020, the accumulative number of confirmed cases in Pakistan were 5230 with most cases occurred in Punjab 2464, and 1411, 744, 228, 224, 119, and 40 cases in Khyber Pakhtunkhwa (KP), Sindh, Baluchistan, Gilgit Baltistan(GB), Islamabad and Azad Jammu Kashmir (AJK), respectively, Table 1. There were 91 confirmed deaths due to COVID-19 and 13 cases were critical, while 1028 patients recovered.

Using 47 days data from 26 February 2020 – 12 April 2020 and ARIMA model, we forecasted the data up to 31st of May. As, we were dealing with timeseries and non-stationary data, it is observed that mean and variance of data is variable in nature. Therefore, double differencing is used, in order to stabilize (made stationary) the data by removing trends. ARIMA (0,2,1), ARIMA (2,2,0) and ARIMA (1,2,1) is applied to produced plots for the number of confirmed cases, recoveries and deaths over time (days) as shown in 2, 3, and 4 respectively of Fig. 1. Results from the model revealed that the number of confirmed cases show a rapid exponential growth which may increase by 2.7 times compared to current cases until end of May 2020. The 95% prediction interval for confirmed cases is from 5681 to 33079 which are at much higher growth, Fig. 2. The results from the forecasting model for deaths
Table 1
Number of confirmed cases, death and recoveries by date and region from 26 February to 12 April 2020.

| Date   | Punjab | Sindh | KP   | Baluchistan | GB   | AJK  | Islamabad | Total Cases | Total Deaths | Total Recoveries |
|--------|--------|-------|------|-------------|------|------|------------|-------------|--------------|-----------------|
| 26/2/20| 0      | 2     | 0    | 0           | 0    | 0    | 0          | 2           | 0            | 0               |
| 04/3/20| 0      | 4     | 0    | 0           | 0    | 0    | 0          | 5           | 0            | 0               |
| 11/3/20| 0      | 9     | 0    | 1           | 3    | 0    | 2          | 20          | 0            | 2               |
| 18/3/20| 33     | 308   | 19   | 23          | 13   | 1    | 5          | 302         | 2            | 5               |
| 25/3/20| 323    | 413   | 80   | 131         | 83   | 1    | 25         | 1057        | 8            | 21              |
| 01/4/20| 845    | 743   | 276  | 169         | 187  | 9    | 62         | 2291        | 31           | 107             |
| 12/4/20| 2464   | 1411  | 744  | 228         | 224  | 40   | 119        | 5230        | 91           | 1028            |

4. Discussion and Conclusion

This is a first formal study to make a short term forecast about COVID-19 confirmed cases as well as the number of related deaths and recoveries. The results of this study revealed that there could be a three-fold increase in the number of confirmed cases at the end of May if the current trend continues. Currently, the number of deaths is quite low in Pakistan but results from the model revealed that it may increase up to 500 at the end of May. On the other hand, recoveries from COVID-19 related complications were slow in the beginning, but it is now increasing exponentially and the forecasting model for it shows that there could be eightfold increase in the number of recoveries. However, the number of confirmed cases is increasing in Pakistan at higher rate as compared to the number of recoveries as the disease is spreading to a wider region of the country.

Similar kind of forecasts has been conducted by other researchers such as [10] but with slightly different method of the forecasting models and presentation of the results is not delivered in a simple way to be understood by a layman, while others are now planning to conduct or continue to conduct such a kind of analysis [19]. It is important to note that majority of the research studies are modeling the preparedness scenarios to inform planning rather predictions [20], that is they informed the actions to be taken to slow the spread and prepare the health system to respond to the pandemic. In addition, their analysis has focused on countries in the South East Asia and Western Pacific regions and ignored other countries such as Pakistan [20] situated in the neighbor of China.

Pakistan is expecting dire consequence of the COVID-19 full-fledged global pandemic such as (12.3 to 18.53) millions of layoffs out of the employees in different sectors of economy in the aftermath of partial or complete shutdown due to the countrywide outbreak and lockdowns. It has formulated/taken drastic precautionary measures to contain COVID-19 including [21] establishing the National Coordination Committee for Covid-19 and National Disaster Management Authority [22], closure of all educational institutions, sealing borders with neighboring countries, travel ban and screening of travelers, social distancing and ban on public gathering, and comprehensive food security. These precautionary measures seem to be working well in the containment of COVID-19 as compared to other nations data worldwide [23,24,25]. Yet, the results of this study suggest an increasing trend of COVID-19 cases and deaths for the upcoming month and we recommend continuation of the above or more stringent measures to contain COVID-19. We believe that the forecasts established by this study is useful for Pakistani government and public in making informed decisions and taking appropriate steps to prevent further spread of COVID-19 disease.
We assume that the analysis for this study is based on an accurate data recorded by NIH [26] in Pakistan and we used appropriate forecasting methods (timeseries modelling-ARIMA). The modeling strategies is based on current trends and non-seasonal time-series variations, following the patterns shown in Fig. 1; assuming the data is accurate, and the trends will continue in the upcoming month of May. We used conventional statistical approaches AIC [27] for model assessment and selection. However, we acknowledge that our analysis is based on the assumptions and if the assumptions are not true, it may lead to an inaccurate forecast. Furthermore, forecasting with timeseries modeling, requires enough historical data, which is not the case with our analysis, and there is always uncertainty associated with prediction as current patterns in the data may not be continued to future.

As Pakistan is a developing country, having a lack of medical facilities which resultantly is affecting the situation further no vaccine or medicine is developed yet to prevent or cure the COVID-19 pandemic permanently. The public health officials and government should take hard decisions to control the rapid increase of the COVID-19. Besides officials, the general public should keep social distancing and use precautions to ensure their safety and control the disease from further spreading, Fig. 3.

Declaration of Competing Interest

We declare that none of the authors has the competing or conflict of interest.

CRediT authorship contribution statement

Muhammad Yousaf: Conceptualization, Methodology. Formal analysis, Project administration, Software, Visualization, Writing - original draft. Samiha Zahir: Data curation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. Muhammad Riaz: Conceptualization, Methodology, Supervision, Visualization, Validation, Writing - original draft, Writing - review & editing. Kamal Shah: Software, Visualization, Validation.

References

[1] Paulus CI, Marston HD, Fauci AS. Coronavirus infections-more than just the common Cold. JAMA 2020;323(8):707–8. doi:10.1001/jama.2020.0757.
[2] WHO | World Health Organization. https://www.who.int, Accessed April 14, 2020.
[3] COVID-19 outbreak and probiotics: facts and information from BioGaia | BioGaia. https://www.biogaia.com/other-news/covid-19-outbreak-and-probiotics-facts/, Accessed April 14, 2020.
[4] Fong SJ, Li G, Dey N, Crespo RC, Herrera-Viedma E. Finding an accurate early forecasting model from small dataset: a case of 2019-nCoV novel coronavirus outbreak. Int J Interact Multim Med Intell 2020;6(1):132. doi:10.9781/jiimi.2020.02.002.
[5] [https://static1.squarespace.com/static/5e6d5df1ff954d5b7b139463/t/6e77d975e60d4b666d4d2f9/1584912703215/1/Covid_Worldmap.png](https://static1.squarespace.com/static/5e6d5df1ff954d5b7b139463/t/6e77d975e60d4b666d4d2f9/1584912703215/1/Covid_Worldmap.png)
[6] Worldometer. Coronavirus cases. Worldometer. doi:10.1101/2020.01.23. 20003849v2.
[7] COVID-19 Health advisory platform by Ministry of National Health Services Regulations and Coordination. http://covd.gov.pk/stats/pakistan. Accessed April 14, 2020.
[8] Raza S, Rashid MA, Rashid MK. Transmission potential and severity of COVID-19 in Pakistan. April 2020. doi:10.20944/preprints202004.0004v1.
[9] COVID-19: Pakistan extends border closure for 2 weeks. https://www. aa.com.tr/en/latest-on-coronavirus-outbreak/covid-19-pakistan-extends-border-closure-for-2-weeks/1802952. Accessed April 14, 2020.
[10] Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19: Braunstein LA, ed.. PLOS ONE 2020;15(3):e0231236. doi:10.1371/journal.pone.0231236.
[11] Chen Y, Cheng J, Jiang X, Xu X. The reconstruction and prediction algorithm of the fractional TDD for the local outbreak of COVID-19. February 2020. https://arxiv.org/abs/2002.10032, Accessed April 14, 2020.
[12] Roosa K, Lee Y, Luo R, et al. Real-time forecasts of the COVID-19 epidemic in china from February 5th to February 24th, 2020. Infect Disease Modell 2020;5:256–63. doi:10.1016/j.idm.2020.02.002.
[13] Kumar P, Kalita H, Patainiya S, et al. Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020 through ARIMA Model with machine learning approach. medRxiv. April 2020:2020.30.20046227. doi: 10.1101/2020.03.30.20046227.
[14] Pais RJ, Taveira N. Predicting the evolution and control of COVID-19 pandemic in Portugal. https://www.medrxiv.org/content/10.1101/2020.03.28.20046250v2, doi:10.1101/2020.03.28.20046250.
[15] National Institute of Health Islamabad Pakistan. https://www.nih.org.pk/?cfc_hl=ischl_1k&_cfctx=5a58d904a5e9259b5c5dc3520ebe326a95-1962314558-D- AUC9C8W37v7vpJ3OzA5tALI-F6gqOz501L-8E0M4jnkJtbbxvPEx-K71jophYd14jYLYL 62AM59JFYoOj3j1tRkUxUWxRbA7tYxWwCjulaAYDVjdjA5tAJnIjPVRHnMuyLoop6e DGFQ-z2BvVj1PEcOmNuf. Accessed April 14, 2020.
[16] Chatfield C. The analysis of time series: an introduction.; 1984. doi:10.1007/ 978-1-4899-2921-1.
[17] Chen P, Yuan H, Shi X. Forecasting crime using the ARIMA model. In: Proceedings - 5th international conference on fuzzy systems and knowledge discovery; 5. FSKD. 2008. p. 627–30. 2008. doi:10.1604/FSKD.2008.222.
[18] Quick-R: Time Series. https://www.statmethods.net/advsats/series.html. Accessed April 14, 2020.
[19] Australian Health Protection Principal Committee (AHPPC) coronavirus (COVID-19) statement on 16 April 2020. Australian Government Department of Health; 2020. https://www.health.gov.au/news/australian-health-protection-principal-committee-ahppc-coronavirus-covid-19-statement-on-16-april-2020 Accessed April 17.
[20] Shearer FM, Walker J, Telliggio N, et al. Assessing the risk of spread of COVID-19 to the Asia Pacific region.
[21] 10 steps Pakistan is taking to contain coronavirus | Pakistan – Gulf News. https://gulfnews.com/world/asia/pakistan/10-steps-pakistan-is-taking-to-contain-coronavirus-1-70403640, Accessed April 17, 2020.
[22] NDMA National Disaster Management Authority Pakistan. http://www.ndma. gov.pk/, Accessed April 17, 2020.
[23] Pakistan has world’s best National Response Program against COVID-19: WHO official. https://www.thenews.com.pk/latest/628662-pakistan-has-worlds-best-national-response-program-against-covid-19-who-official. Accessed April 17, 2020.
[24] COVID-19 Health Advisory Program by Ministry of National Health Services Regulations and Coordination. https://covd.gov.pk/, Accessed April 17, 2020.
[25] WHO COVID-19 Dashboard. https://covid19.who.int/, Accessed April 17, 2020.
[26] COVID-19 - National Institute of Health Islamabad. https://www.nih.org.pk/ novel-coronavirus-2019-nov/, Accessed April 17, 2020.
[27] Using AIC to Test ARIMA Models – CoolStatsBlog. https://coolstatsblog.com/ 2013/08/14/using-aic-to-test-arima-models-2/, Accessed April 17, 2020.