

Data model to predict prevalence of COVID-19 in Pakistan

Muhammad Qasim^{1*}(PhD), Waqas Ahmad^{2*}(MS), Shenghuan Zhang², Muhammad Yasir (PhD)²,
Muhammad Azhar³

1. Centre for Bioengineering & Nanomedicine (Dunedin), Department of Food Science, Division of Sciences. University Of Otago, Dunedin, PO Box 56, Dunedin 9054, New Zealand.
2. Department of Computer Science, University Of Otago, Dunedin, PO Box 56, Dunedin 9054, New Zealand.
3. Department of Education, Islamic International University Islamabad, Pakistan

‡ Equal Contribution:

*Corresponding author muhammadqasim.qasim@otago.ac.nz

Abstract

Coronavirus disease 2019 (COVID19) has spread to 181 countries and regions and leaving behind 1,133,788 confirmed cases and 62784 deaths worldwide. Countries with lower health services and facilities like Pakistan are on great risk. Pakistan so far has 3,277 confirmed cases and 50 reported deaths due to COVID19. Various mathematical models had presented to predict the global and regional size of pandemic. However, all those models have certain limitation due to their dependence on different variables and analyses are subject to potential bias. As each country has its own dimension therefore country specific model are required to develop accurate estimate. Here we present a data model to predict the size of COVID19 in Pakistan. In this mathematical data model, we used the time sequence mean weighting (TSMW) to estimate the expected future number of COVID19 cases in Pakistan until 29th April 2020. For future expected numbers of cases for long terms the data collection have to be maintained in real time.

Keywords: COVID19; Pakistan; Data model

Specification Tables

Subject	Infectious Diseases
Specific subject area	Time sequence mean weighting mathematical model applied to forecast the prevalence and incidence of COVID-2019 in Pakistan
Type of data	Graphs, Figure, Excel File
How data were acquired	Data was extracted from http://covid.gov.pk/
Data format	Data was in in raw format and have been arranged and analysed in Matlab. An Excel file with data has been uploaded.
Parameters for data collection	We collected daily reported cases and deaths for COVID19 in Pakistan
Description of data collection	The daily prevalence data of COVID-2019 from March 10, 2020 to April 1st, 2020 was collected from the official website of COVID-19 Health Advisory Platform by Ministry of National Health Services Regulations and Coordination, Pakistan http://covid.gov.pk/ . The Matlab was used to build associated mat files and data was analysed by time sequence mean weighting mathematical model to evaluate the daily reported cases and number of deaths due to COVID-19 in Pakistan. We have calculated the TSMW (η) and used the model equation $y_n =$

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

	$(y_{n-1} - d_{n-1})\left(\frac{1}{\eta} - 1\right) - R_{n-1}$ to predict the future state of the given sample space.
Data source location	University of Otago. Dunedin, New Zealand
Data accessibility	Raw data and its analysis can be retrieved from the https://github.com/qasimattock/COVID-19-Pakistan.git

Value of the Data

- This data will be helpful for Pakistan health authorities to prepare themselves to contain COVID-19.
- As Pakistan is very low in the Global Health Security (GHS) Index therefore they need to utilised their resources in targeted ways and this data will be important for them to take their strategic decisions to mitigate the COVID-19 pandemic in their country.
- Further our model can also use for monitoring and evaluation purpose of their own gown efforts and precautionary measures.

Data description

The daily prevalence data of COVID-2019 from March 10, 2020 to April 1st, 2020 was collected from the official website of COVID-19 Health Advisory Platform by Ministry of National Health Services Regulations and Coordination, Pakistan <http://covid.gov.pk/>.¹ The Matlab was used to build associated mat files and data was analysed by the time sequence mean weighting (TSMW) to predict the future state of the confirmed cases of COVID19 and deaths in Pakistan until 29th April, 2020 (Figure 1). The analysis predicted the expected count of patients 77,905 with minimum 8285 confirmed cases of COVID-19 while expected deaths count of 1382 with lower limit of 45 until 29th April 2020. Currently COVID19 has spread to 181 countries and regions, infected 1,133,788 individuals, and caused 62784 deaths worldwide.² Countries with lower health services and facilities like Pakistan are on great risk.³ Pakistan so far has 3,277 confirmed cases and 50 reported deaths due to COVID19.

Experimental Design, Materials and Methods;

To derived this model, Let y_i represents the number patients reported at i^{th} day then we define the ratio x_i as

$$x_i = \frac{y_{i-1}}{y_{i-1} + y_i}$$

Where, $i - 1$ represents the previous day and the current day is denoted by i . The x_i is calculated for all $i \in 1, \dots, n$, where n is the total number of days for which estimation is performed. The mean of all x_i is calculated as⁴

$$\eta = \frac{\sum_{i=1}^n x_i}{n} \quad (2)$$

Let d_i represents the total deaths and R_i is the count of patients recovered until i days. Then the expected number of patients at day 1 was modelled as

$$y_1 = (y_0 - d_0)\left(\frac{1}{\eta} - 1\right) - R_0 \quad (3)$$

We have assumed the dead patients were eliminated from community without any further virus spreading capability. Similarly, the recovered patients were subtracted from the total number of patients. Extending the Eq. [3] for y_2, \dots, y_n as

$$y_2 = (y_1 - d_1) \left(\frac{1}{\eta} - 1 \right) - R_1$$

$$y_n = (y_{n-1} - d_{n-1}) \left(\frac{1}{\eta} - 1 \right) - R_{n-1} \quad (4)$$

Let P be the total population of the sample space then the spread will grow until difference between P and y_n is a positive integer defined as

$$P - y_n > 0 \quad (5)$$

When, $P - y_n \leq 0$, n is the expected number of days at which the whole population will be impacted. The proposed model is different from standard normal distribution as it continuously trains itself from the existing data to estimate the future state. Therefore, it can apply for future real time data.

Lower bound: Let ζ be the lower bound estimated from x_i as

$$\zeta = \max(x_i) \quad \forall x_i < 0.5$$

$$\zeta = \max(x_i) - \frac{\sum(x_i - \eta)^2}{i - 1} \quad \forall x_i \geq 0.5 \quad (6)$$

We use $\eta = : \zeta$ in Eq. [4] to find the lower bound on the estimated number of patients. The convergence of lower bound remains very close to the actual data set until the same situation prevails and subjected to change with new precautionary measures to contain pandemic.

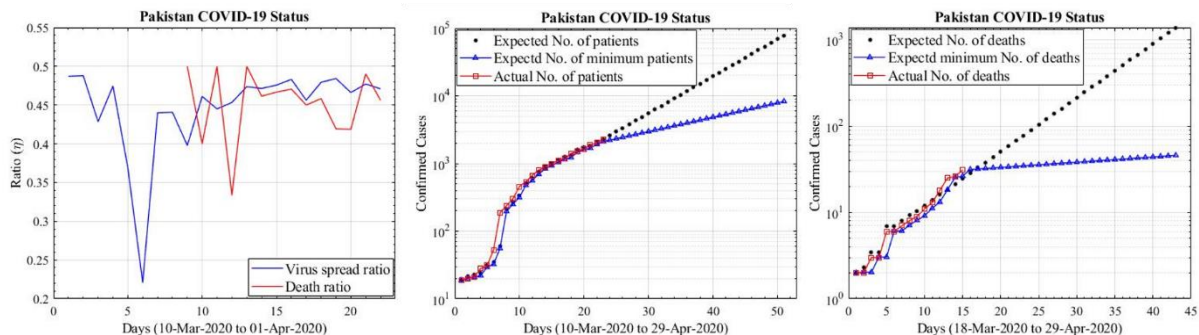


Figure: Spread ratio (η) for COVID-19 cases, expected number of cases and expected number of deaths in Pakistan.

Our models curves for COVID-19 prevalence and their future trend based on real time data, when new data will come it will further validate our model. If Pakistan does not act on complete lockdown and

strict precautionary measures due to its weak health system its whole population will be on risk.³ Our data shows that it will take roughly 90 days (1-July-2020) to affect the 95% population under the constraint that there is no isolation policy implemented among the population.

Conflict of Interest

Authors have no conflict of interest.

Appendix A.

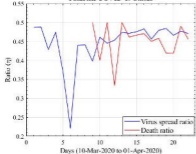
Supplementary data:

Supplementary data to this article can be found online at <https://github.com/qasimattock/COVID-19-Pakistan.git>

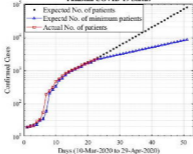
References:

- 1 COVID-19 Health Advisory Platform by Ministry of National Health Services Regulations and Coordination . Available at <http://covid.gov.pk/>. Accessed April 6, 2020, n.d.
- 2 Situation reports. Available at <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>. Accessed April 6, 2020, n.d.
- 3 Homepage -Global health Security GHS Index. Available at <https://www.ghsindex.org/>. Accessed March 19, 2020, n.d.
- 4 Qasim Muhammad, Ahmad Waqas, Yoshida Minami, Gould Maree, Yasir Muhammad. Analysis of the Worldwide Corona Virus (COVID-19) Pandemic Trend;A Modelling Study to Predict Its Spread. *MedRxiv* 2020:2020.03.30.20048215. Doi: 10.1101/2020.03.30.20048215.

Pakistan COVID-19 Status



Pakistan COVID-19 Status



Pakistan COVID-19 Status

