The corona virus disease (COVID-19) has affected almost 210 countries in the world after its start from Wuhan, the city of China. Due to its rapid spread, the WHO declared this disease as pandemic. According to the latest data of WHO (up to 20th October 2020) 40881910 confirmed cases has been reported globally including 1126408 deaths. High fever, pain in body, respiratory problem and dry cough are the important indication of this disease stated by.[1] The incubation period of this disease varies from 6 to 12 days depending upon the immune system of the patient. The main cause of this transmission are respiratory droplets and physical gatherings. There are observed another kind of coronavirus patients who do not express such type of symptoms but still there are suffering from this dangerous disease stated by.[2] In the start of this pandemic, the death risk of this disease varies among Different countries and reported between 5.1 % to 8.1%. The most deaths due to COVID-19 were reported in USA, Spain, Italy and China. Like other countries, Pakistan is also declared as COVID affected country when first case of COVID patient has been reported on 26th February 2020 in Karachi. The covid positive patient return back from Iran. After that pandemic start spreading in all over Pakistan. In order to control the growth of coronavirus, the government of Pakistan decided to put the country under lockdown on 13th March 2020. After some days, this decision brought out some bad effects on all com-

Objectives: The aim of this study is to estimate the recovery rate and death rate of the Covid-19 patients in Pakistan both at provincial level and at national level.

Methods: In the current study simple linear regression model is used to estimate the case fatality rate (CFR) and recovery rate (RR). Also, the polynomial regression model has been employed to predict the infected patients of Covid-19 across Pakistan and for all its provinces.

Results: The results indicate that the provinces: Punjab, Sindh, KPK and ICT have been projected to be sensitive regarding COVID-19.

Conclusion: it is strongly suggested that prevention must be considered, and full attention must be provided by each citizen to follow the SOPs recommend by health department. If responsible attitude will not be expressed by individuals, then there is chance of increase of COVID patients again and put the country back in struggle.

Keywords: COVID-19, Pakistan, regression model, CFR, RR, Predictions
munities especially poor societies. The people were become just helpless to eat even a single day food. So, the lockdown was ended on 8th May 2020 and government also announced some facilities of 14000 rupees every month to each poor family. But the cases of coronavirus was started growing exponentially after opening the lockdown. So, government of Pakistan again put the country under the lockdown on 18th June 2020. The government also emphasized the Pakistan nation to follow the Standard Operating Procedures (SOPs) while going outside the home. The developed countries with advance health infrastructure are continuously struggling to understand the mechanism of this disease and trying to discover its vaccine but still no solution. The only solution of this pandemic is to wash hand, wearing mask and maintaining the social distance.

Even many advance countries facing very high mortality rate. In this situation, it is important to quantifying the disease burden by estimating the case fatality rates and calculating the reproduction number. This kind of epidemiological information plays very important role for healthcare planning and to control the disease efficiently. To investigate the pandemics, several studies have been carry out about the analysis of coronavirus situations in Pakistan, but there is a lack of reliable estimates of CFR using regression models in Pakistan especially province wise. Therefore the objective of this study is to estimate the case fatality rate and recovery rate in different provinces and in overall Pakistan. Moreover to predict the increasing and decreasing trend in the infected patients.

Methods

In order to meet the objectives of the study and analyze the COVID-19 situation in Pakistan, the number of confirmed cases, deaths and recoveries has been collected from http://covid.gov.pk/ up to 30th October 2020. In order to calculate the case fatality rate (CFR) the following regression model is used.

\[ Y = b_0 + b_1X \]

where \( Y \) is the number of deaths and \( X \) as number of confirmed cases. The term \( b_0 \) and \( b_1 \) represents the intercept and slope of the fitted line respectively. On the other hand, the recovery rate (RR) is calculated by using the below regression equation

\[ S = \lambda_0 + \lambda_1T \]

where \( S \) represents the number of recovered cases and \( T \) as number of confirmed cases. The term \( \lambda_0 \) is the intercept and \( \lambda_1 \) represent the slope of the fitted line. The CFR and RR will be estimated by using the slope of the fitted line 1 and 2 respectively.

In estimating the CFR and RR using equation (1) and (2), the starting point will be considered as the day when the first death and first recovery was reported in Pakistan and its provinces. The starting days of pandemic period have been discarded when no death or recovery were reported. To determine the best fit of the model, the coefficient of determination \( R^2 \) is used. The standard error of the slope is used to estimate the 95% confidence interval (CI).

For the prediction of coronavirus patients in Pakistan and all of its provinces, the appropriate polynomial regression model will be applied. The number of infected individuals will be predicted up to 15th November 2020. The prediction analysis will be performed for all provinces separately. The regression model will be trained from 26th February 2020 to 23rd October 2020. The consequences of the analysis will be validated with rest of the 10% one week data (from 24th October to 30th October 2020). After that using the trained and validated model, the patients of coronavirus will be predicted from 31st October 2020 to 15th November 2020 for Pakistan and its provinces. Reduction or increasing rate for infected patients is also calculated, which tells that how much percentage the infected patients will be reduced. The number of patients up to 30th October 2020 will be taken the base.

Result and Discussion

The Estimation of CFR and RR in all over Pakistan

The scatter plots of estimated case fatality rate (CFR) and recovery rate (RR) of coronavirus in overall Pakistan are shown below in figure 1. The estimated regression line for CFR is written as

\[ Y = -22.2 + 0.021003X \]

The coefficient of determination \( R^2 \) is calculated to be 99.87% which shows a strong linear association between confirmed cases and number of deaths in Pakistan reported up to 30th October 2020 from the first death recorded (on 18 March 2020). The slope of the above regression line estimates the CFR which is calculated to be 2.1003% with 95% confidence interval (CI) 2.09% to 2.11%. The estimated regression line for RR is written as

\[ S = -24187 + 0.9528T \]

The \( R^2 \) for above line is calculated to be 93.73% that shows best association between the confirmed cases and recoveries in Pakistan up 30th October 2020 when first recovery is recorded (on 6 March 2020). The estimated RR is calculated from the slope of the above line which is 95.82% with 95% CI 92.64% to 98.99%.

The scatter plots of estimated case fatality rate (CFR) and recovery rate (RR) of coronavirus in province Punjab are shown below in figure 2. The estimated regression line for CFR is written as

\[ Y = -69.69 + 0.023385X \]

The coefficient of determination \( R^2 \) is calculated to be
99.73% which shows a strong linear association between confirmed cases and number of deaths in Punjab reported up to 30th October 2020 from the first death recorded (on 24th March 2020). The slope of the above regression line estimates the CFR which is calculated to be 2.33% with 95% CI 2.32% to 2.35%. The estimated regression line for RR is written as

\[ S = -10933 + 0.9710T \]

The coefficient of determination \((R^2)\) is calculated to be 88.71% that shows good association between confirmed cases and recovered cases in Punjab up to 30th October 2020 when first recovery is recorded (on 24th March 2020). The estimated RR is calculated from the slope of the above line which is 97.10% with 95% CI 92.50% to 1.011%.

The scatter plots of estimated case fatality rate (CFR) and recovery rate (RR) of coronavirus in province Sindh are shown below in Figure 3. The estimated regression line for CFR is written as

\[ Y = -51.97 + 0.018492X \]

The coefficient of determination \((R^2)\) is calculated to be 99.53% which shows a strong linear association between confirmed cases and number of deaths in Sindh reported up to 30th October 2020 from the first death recorded (on 12th April 2020). The slope of the above regression line estimates the CFR which is calculated to be 1.849% with 95% CI 1.83% to 1.86%. The estimated regression line for RR is written as

\[ S = -14529 + 1.0135T \]

The R2 for above line is calculated to be 95.37% that shows good association between confirmed cases and recovered cases in Sindh up to 30th October 2020 when first recovery is recorded (on 12th April 2020). The estimated RR is calculated from the slope of the above line which is 1.013% with 95% CI 98.25% to 104.4%.

The scatter plots of estimated case fatality rate (CFR) and recovery rate (RR) of coronavirus in province KPK are shown below in Figure 4. The estimated regression line for CFR is written as

\[ Y = 64.41 + 0.032644X \]
The coefficient of determination ($R^2$) is calculated to be 99.16% which shows a strong linear association between confirmed cases and number of deaths in KPK reported up to 30th October 2020 from the first death recorded (on 18th March 2020). The slope of the above regression line estimates the CFR which is calculated to be 3.32% with 95% CI 3.22% to 3.30%. The estimated regression line for RR is written as

$$S=-4173+1.0135T$$

The $R^2$ for above line is calculated to be 92.58% that shows good association between confirmed cases and recovered cases in KPK up to 30th October 2020 when first recovery is recorded (on 25th March 2020). The estimated RR is calculated from the slope of the above line which is 98.58% with 95% CI 94.87% to 102.2%.

The scatter plots of estimated case fatality rate (CFR) and recovery rate (RR) of coronavirus in province Baluchistan are shown below in figure 5. The estimated regression line for CFR is written as

$$Y=-2.21+0.022572X$$

The coefficient of determination ($R^2$) is calculated to be 97.37% which shows a strong linear association between confirmed cases and number of deaths in Baluchistan reported up to 30th October 2020 from the first death recorded (on 22nd March 2020). The slope of the above regression line estimates the CFR which is calculated to be 1.004% with 95% CI 0.98% to 1.02%. The estimated regression line for RR is written as

$$S=-1798+0.9838T$$

The $R^2$ for above line is calculated to be 90.69% that shows good association between confirmed cases and recovered cases in Baluchistan up to 30th October 2020 when first recovery is recorded (on 28th March 2020). The estimated RR is calculated from the slope of the above line which is 98.38% with 95% CI 94.14% to 102.6%.

The scatter plots of estimated case fatality rate (CFR) and recovery rate (RR) of coronavirus in province Gilgit Baltistan are shown below in figure 6. The estimated regression line for CFR is written as

$$Y=-2.21+0.022572X$$
The coefficient of determination ($R^2$) is calculated to be 90.92% which shows a strong linear association between confirmed cases and number of deaths in Gilgit Baltistan reported up to 30th October 2020 from the first death recorded (on 22nd March 2020). The slope of the above regression line estimates the CFR which is calculated to be 2.25% with 95% CI 2.16% to 2.35%. The estimated regression line for RR is written as

$$S=-58.8+0.8590T$$

The $R^2$ for above line is calculated to be 91.13% that shows good association between confirmed cases and recovered cases in Gilgit Baltistan up to 30th October 2020 when first recovery is recorded (on 26th March 2020). The estimated RR is calculated from the slope of the above line which is 85.90% with 95% CI 82.31% to 89.49%.

The scatter plots of estimated case fatality rate (CFR) and recovery rate (RR) of coronavirus in province Azzad Jammu Kashmir are shown below in figure 7. The estimated regression line for CFR is written as

$$Y=0.718+0.025202X$$

The coefficient of determination ($R^2$) is calculated to be 97.64% which shows a strong linear association between confirmed cases and number of deaths in Azzad Jammu Kashmir reported up to 30th October 2020 from the first death recorded (on 10th May 2020). The slope of the above regression line estimates the CFR which is calculated to be 2.52% with 95% CI 2.46% to 2.57%. The estimated regression line for RR is written as

$$S=-96.2+0.8545T$$

The $R^2$ for above line is calculated to be 96.47% that shows good association between confirmed cases and recovered cases in Azzad Jammu Kashmir up to 30th October 2020 when first recovery is recorded (on 4th April 2020). The estimated RR is calculated from the slope of the above line which is 85.45% with 95% CI 83.21% to 87.68%.

The scatter plots of estimated case fatality rate (CFR) and recovery rate (RR) of coronavirus in province Islamabad Capital Territory are shown below in figure 8. The estimated regression line for CFR is written as

$$Y=-3.08+0.011107X.$$
The coefficient of determination ($R^2$) is calculated to be 99.49% which shows a strong linear association between confirmed cases and number of deaths in Islamabad Capital Territory reported up to 30th October 2020 from the first death recorded (on 6th April 2020). The slope of the above regression line estimates the CFR which is calculated to be 1.117% with 95% CI 1.09% to 1.12%. The estimated regression line for RR is written as

$$S = -1172 + 0.9364T$$

The $R^2$ for above line is calculated to be 93.38% that shows good association between confirmed cases and recovered cases in Islamabad Capital Territory up to 30th October 2020.

### Table 1. A Brief Summary to the Estimation of CFR and RR of Overall Pakistan and its Provinces

<table>
<thead>
<tr>
<th>States</th>
<th>Case Fatality Rate</th>
<th>Recovery Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CFR (%)</td>
<td>95 % CI</td>
</tr>
<tr>
<td>Pakistan</td>
<td>2.10</td>
<td>2.09–2.11</td>
</tr>
<tr>
<td>Punjab</td>
<td>2.31</td>
<td>2.32–2.35</td>
</tr>
<tr>
<td>Sindh</td>
<td>1.84</td>
<td>1.83–1.86</td>
</tr>
<tr>
<td>Khyber Pakhtunkhwa</td>
<td>3.26</td>
<td>3.22–3.30</td>
</tr>
<tr>
<td>Baluchistan</td>
<td>1.00</td>
<td>0.98–1.02</td>
</tr>
<tr>
<td>Gilgit Baltist</td>
<td>2.25</td>
<td>2.16–2.35</td>
</tr>
<tr>
<td>Azad Jammu Kashmir</td>
<td>2.52</td>
<td>2.46–2.57</td>
</tr>
<tr>
<td>Islamabad Capital Territory</td>
<td>1.11</td>
<td>1.09–1.12</td>
</tr>
</tbody>
</table>
when first recovery is recorded (on 16th March 2020). The estimated RR is calculated from the slope of the above line which is 93.64% with 95% CI 90.38% to 96.90% (Table 1).

Now, the prediction analysis will be performed by using polynomial regression model. The number of patients of coronavirus will be predicted for Pakistan and its provinces up to 15th November 2020. The first case of this pandemic was reported in Pakistan on 26th February 2020. The significant rise (above one lac cases) in the coronavirus patients in Pakistan was recorded after 6th June 2020.

The scatter plot between confirmed cases vs. days and prediction of the patients in overall Pakistan can be seen in the figure 9. The data from 7th June 2020 to 23rd October 2020 are utilized to train the regression model. The least square third-degree polynomial is used to estimate the best fit for available data of overall Pakistan. The coefficient of determination $R^2$ is estimated to be 99.7% shown in the figure 9. The model is eventually tested with the data (7 days) from 24th October 2020 to 30th October 2020. The patients predicted from 31st October 2020 to 15th November 2020 are plotted in the figure 9. The number of patients up to 15th November 2020 is predicted to be 396809.

The first case of this coronavirus was reported in Punjab on 15th March 2020. The substantial increase (above ten thousand cases) in the coronavirus patients in Punjab was recorded after 6th May 2020. The scatter plot between confirmed cases vs. days and prediction of the patients in Punjab can be seen in the figure 10. The data from 7th May 2020 to 23rd October 2020 are utilized to train the regression model. The least square third-degree polynomial is used to estimate the best fit for available data of Punjab. The coefficient of determination $R^2$ is estimated to be 97.7% shown in the figure 10. The number of patients up to 15th November 2020 is predicted to be 101265. The first case of this coronavirus was reported in Sindh on 26th February 2020. The remarkable growth (above one lac cases) in the coronavirus patients in Sindh was recorded after 8th July 2020. The scatter plot between confirmed cases vs. days and prediction of the patients in Sindh can be seen in the figure 11. The data from 9th July 2020 to 23rd October 2020 are utilized to train the regression model.
The least square third-degree polynomial is used to estimate the best fit for available data of Sindh. The coefficient of determination $R^2$ is estimated to be 99.4% shown in the figure 11. The number of patients up to 15th November 2020 is predicted to be 162978.

The first case of this coronavirus was reported in KPK on 16th March 2020. The significant increase (above ten thousand cases) in the coronavirus patients in KPK was recorded after 30th May 2020. The scatter plot between confirmed cases vs. days and prediction of the patients in KPK can be seen in the figure 12. The data from 31st May 2020 to 23rd October 2020 are utilized to train the regression model.

The least square third-degree polynomial is used to estimate the best fit for available data of KPK. The coefficient of determination $R^2$ is estimated to be 99.8% shown in the figure 12. The number of patients up to 15th November 2020 is predicted to be 45206.

The first case of this coronavirus was reported in Baluchistan on 10th March 2020. The remarkable increase (above ten thousand cases) in the coronavirus patients in Baluchistan was recorded after 25th June 2020. The scatter plot between confirmed cases vs. days and prediction of the patients in Baluchistan can be seen in the figure 13. The data from 26th June 2020 to 23rd October 2020 are utilized to train the regression model.

The least square third-degree polynomial is used to estimate the best fit for available data of Baluchistan. The coefficient of determination $R^2$ is estimated to be 96.3% shown in the figure 13. The number of patients up to 15th November 2020 is predicted to be 17184.

The first case of this coronavirus was reported in GB on 3rd March 2020. The substantial increase (above one thousand cases) in the coronavirus patients in GB was recorded after 9th June 2020. The scatter plot between confirmed cases vs. days and prediction of the patients in GB can be seen in the figure 14. The data from 10th June 2020 to 23rd October 2020 are utilized to train the regression model.

The least square third-degree polynomial is used to estimate the best fit for available data of GB. The coefficient of determination $R^2$ is estimated to be 76.9% shown in the figure 14. The number of patients up to 15th November 2020 is predicted to be 17184.
The number of patients up to 15th November 2020 is predicted to be 5015. The first case of this coronavirus was reported in AJK on 18th March 2020. The substantial growth (above one thousand cases) in the coronavirus patients in AJK was recorded after 25th June 2020. The scatter plot between confirmed cases vs. days and prediction of the patients in AJK can be seen in the figure 15. The data from 26th June 2020 to 23rd October 2020 are utilized to train the regression model. The least square third-degree polynomial is used to estimate the best fit for available data of AJK. The coefficient of determination R^2 is estimated to be 99.8% shown in the figure 15. The number of patients up to 15th November 2020 is predicted to be 5831.
The first case of this disease was reported in ICT on 10\textsuperscript{th} March 2020. The substantial increase (above ten thousand cases) in the coronavirus patients in ICT was recorded after 18\textsuperscript{th} June 2020. The data from 19th June 2020 to 23rd October 2020 are utilized to train the regression model. Third-degree polynomial is used to estimate the best fit for available data of ICT. R\textsuperscript{2} is estimated to be 99.4\% shown in the figure 16. The number of patients up to 15\textsuperscript{th} November 2020 is predicted to be 23514 (Table 2).

**Conclusion**

This study mainly focused to predict the fatality of coronavirus in Pakistan and across each of its provinces; using the regression model the CFR and RR have been estimated up to 30\textsuperscript{th} October 2020. The estimated CFR is 2\% with 95\% confidence interval of 2.09\% to 2.11\% and RR is estimated as 93\% with 95\% CI having limits from 92.64\% to 98.99\% across Pakistan. The estimated CFR for Punjab, Sindh, KPK, Baluchistan, GB, AJK and ICT have been estimated as 2\%, 2\%, 3\%, 1\%, 2\%, 3\% and 1\% respectively and the estimated RR for Punjab, Sindh, KPK, Baluchistan, GB, AJK and ICT have been found as 97\%, 101\%, 98\%, 98\%, 86\%, 85\% and 93\% respectively.

The number of infected patients have been predicted using polynomial regression model. Number of patients have been predicted for from 31\textsuperscript{st} October 2020 to 15\textsuperscript{th} November 2020. For the purpose of validation, one week test data (from 24 October 2020 to 30 October 2020) were used. The predicted patients likely to be 400,000 and reduction rate will be 19\% across Pakistan up to 31st October 2020. The reduction rate for respective provinces: Punjab, Sindh, KPK, Baluchistan, GB, AJK and ICT is estimated as 3, 12, 15, 8, 18, 43 and 19 respectively. The Province of Azad Jammu Kashmir (AJK) contained highest reduction rate as 43\% and the province Punjab contained the lowest reduction rate as 3\%. According to the polynomial regression analysis, the provinces Punjab, Sindh, Kyber Pakhtunkhwa and ICT are predicted to be sensitive regions regarding coronavirus pandemic because the number of patients predicted in these states have been reported more than the remaining regions of Pakistan. It is suggested that people have to follow the standard protocol (SOPs) as directed by the Government of Pakistan.

But according to the latest information up to 12\textsuperscript{th} December 2020, it can be seen that the no. of patients has been increased. Ignorance of SOPs by the people has become

<table>
<thead>
<tr>
<th>States</th>
<th>Degree of polynomial</th>
<th>R\textsuperscript{2} (%)</th>
<th>Predicted Patients up to 31 Oct 2020</th>
<th>Reduction (%)</th>
<th>Latest no. of patients up to 12 Dec 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pakistan</td>
<td>3</td>
<td>99.7</td>
<td>396809</td>
<td>19</td>
<td>438425</td>
</tr>
<tr>
<td>Punjab</td>
<td>3</td>
<td>97.7</td>
<td>101265</td>
<td>3</td>
<td>127212</td>
</tr>
<tr>
<td>Sindh</td>
<td>3</td>
<td>99.4</td>
<td>162978</td>
<td>12</td>
<td>194359</td>
</tr>
<tr>
<td>Khyber Phaktunkhwa</td>
<td>3</td>
<td>99.8</td>
<td>45206</td>
<td>15</td>
<td>52092</td>
</tr>
<tr>
<td>Baluchistan</td>
<td>3</td>
<td>96.3</td>
<td>17184</td>
<td>8</td>
<td>17737</td>
</tr>
<tr>
<td>Gilgit Baltistan</td>
<td>3</td>
<td>76.9</td>
<td>5015</td>
<td>18</td>
<td>4783</td>
</tr>
<tr>
<td>Azzad Jammu Kashmir</td>
<td>3</td>
<td>99.8</td>
<td>5831</td>
<td>43</td>
<td>7663</td>
</tr>
<tr>
<td>Islamabad Capital Territory</td>
<td>3</td>
<td>99.4</td>
<td>23514</td>
<td>19</td>
<td>34579</td>
</tr>
</tbody>
</table>
the reason of this increment. The patients in overall Pakistan have been increased up to 434425. In the same way, it can be reported as the no. of covid-19 cases have been extended or increased up to 127212, 194359, 52092, 17737, 7663, 34579 in Punjab, Sindh, KPK, Blouchistan, AJK, and ICT. Only the single province Gilgit Baltistan is found where the number of patients have been decreased and observed to be 4783 on 12th Dec 2020 while 5015 patients were predicted up to 31st Oct 2020.

Disclosures

Ethics Committee Approval: The study was approved by the Local Ethics Committee.

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.


References