ISSN: 2474-3658

Gadhvi. J Infect Dis Epidemiol 2020, 6:153
DOI: 10.23937/2474-3658/1510153
Volume 6 | Issue 4
Open Access

Journal of
Infectious Diseases and Epidemiology

REVIEW ARTICLE

Statistical Analyses of COVID-19 Cases in India

Neha K Gadhvi*

Technology, Operations and Decision Sciences, Ahmedabad University, Gujarat, India

***Corresponding author:** *Neha K Gadhvi, Technology, Operations and Decision Sciences, Ahmedabad University, Ahmedabad-380009, Gujarat, India*

Abstract

COVID-19 outbreak was first reported in Wuhan, China and has spread most of the countries of the world. The outbreak spreads are largely influenced by each country's policy and social responsibility. According to WHO, the estimate of attack the rate on 23, June, 2020 to be between 1.4 to 2.5. In India the situation is somehow controllable compared to the situations of developed countries. It will be worth interesting to know about the fact and figures of corona cases in India. The different types of data are available on worldometers. We have tried to analyse such data for India and made different kinds of prediction regarding affected rate, daily new cases, daily total finished cases etc. From our study we observed that the affected rate may decrease after 31-03-2020. Daily new cases will be under control and decreases from the end of April, 2020. After 23-06-2020 there may not be any pending effective cases.

Abbreviations

DTC: Data Consist Daily Total Cases; DNC: Daily New Cases; DRC: Daily Recovered Cases; DD: Daily Deaths; AR: Attack Rate; PAR: Predictive Attack Rate; TFC: Total Finished Cases; FR: Finished Rate; PTFC: Predictive Total Finished Cases

Introduction

The corona virus COVID-19 originated from Wuhan, the city in central China. The symptoms of corona virus are fever, cough and shortness of breath. It typically causes flu-like symptoms. In some patients particularly the elders and others with other chronic health conditions- these, symptoms can develop into pneumonia, with chest tightness, chest pain and shortness of breath. It seems to start with a fever, followed by a drug cough. After a week it can lead to shortness of breath, with about 20% patients requiring hospital treatment.

According to WHO-China Joint Mission published the following median time to clinical recovery:

- Mild cases: Approximately 2 weeks.

- Sever or critical disease: 3-6 weeks.

- Time from onset to the development of sever disease: 1 week.

The coronavirus COVID-19 is affecting 192 countries and territories around the world and one international conveyance (the Diamond Princess Cruise Ship harbored in Yokohama, Japan).

Wuhan, the city in Central China where the virus originated. The city on January 23, shutdown transport links. The attack rate (how rapidly the disease spreads) of a virus is indicated by its reproductive number, denoted by Ro. The reproductive rate represents the average number of people to which a single infected person will transmit the virus. According to WHO, the estimate of the rate on 23, June, 2020 to be between 1.4 to 2.5.

In India first case reported on 30, January 2020 from Kerala. The number of corona virus infected in India crossed the 200 number after a spike in cases over the past six days. The challenge for India versus its peers is starker if infections spread rapidly considering the higher density of population per capita and weaker health infrastructure. Day by day total number of corona virus cases have been increasing in folded. The centre and state governments have decided to completely shutdown 75 districts across the country where corona virus cases have been reported. It has also been decided to suspend the interstate bus services till March 31. Major train services are also suspended till March 31, including sub-urban rail services, exempting goods trains. All metro rails services are also suspended. At present it is

Citation: Gadhvi NK (2020) Statistical Analyses of COVID-19 Cases in India. J Infect Dis Epidemiol 6:153. doi.org/10.23937/2474-3658/1510153

Received: May 06, 2020: **Accepted:** September 01, 2020: **Published:** September 03, 2020

extended to 3, May 2020. Millions of people across the country stays indoors.

India is at second rank as per the area and the population density. Compared to other developing countries of the world, India possess a good control on the spread of corona virus. It will be worth interesting to know the attack rate, and rate of finished cases (recovered and death).

Various types of analysis have been made on COVID-19 data. Rachah and Torres [1] have done Analysis, simulation and optimal control of a SEIR model for Ebola virus and modelled with MCMC. Porter [2] used demographic effects such as birth and death rates in the SEIR to model equilibrium with vital dynamic. Hamzah, et al. [3] has considered Worldwide COVID-19 Outbreak data analysis and prediction *using SEIR models with* nonlinear ordinary differential equations. Rodriguez-Morales [4] have performed systematic review including mete analysis to assess clinical, laboratory imaging features and outcomes of COVID-19 confirmed cases.

Kumar, et al. [5] analyzed state wise data related to COVID-19 in India. Salgotra, et al. [6] reported time series model to forecast confirmed cases and death cases in India across the most affected states namely Maharashtra, Gujarat and Delhi and as a whole India using the genetic programming.

Ghosal, et al. [7] have determined the trend related to death counts observed at 5th and 6th week of the COVID-19 in India based on linear regression and multiple regression analysis. Schueller, et al. [8] have studied the potential impact of lockdown in India and estimated a basic reproduction rate, but in both the papers they have not predicted the number of cases.

Ranjan [9] estimated a final epidemic size of 13000 cases and predicted that India will enter in equilibrium position by the end of May 2020. The prediction has been made based on exponential, logistic and SIR models. In this article we have analyzed the corona cases data of India. The data is downloaded from the website: worldometers.info/coronavirus/country/India on 28-April 2020.

The data consist daily total cases, daily new cases, daily recovered cases and daily deaths. The attack rate for day t is calculated from the ratio of DNC of day t to DTC of day t-1.

This work was aimed to develop a model to predict attack rate and total finished cases for future and to find out the breakeven point for total cases and total finished cases. That means a time point after which there will not be any new active cases of corona virus in India under the assumption that the current environmental situations remains same.

Model and Methodology

We have used the data of COVID-19 cases in India from 1-3-2020 to 28-4-2020 time period. Based on the data regarding daily total cases, daily new cases, daily recovered cases and daily deaths the following simple mathematical equations are used to calculate attack rate, effective cases and finished rate.

AR_t = Daily new cases produce on day (t + 1)/total cases on day t

Effective cases on day t = Effective daily cases on (t-1) + daily new cases on day t - daily

finished cases on day t

FR_t = daily finished cases on day t/effective cases of day t

First of all we have plotted the daily attack rate against day t. The graph is shown below for 48 days (from 1-3-2020 to 17-4-2020). Here t = 1 for date 1-3-2020 and so on.

From the Figure 1 we observed that the attack rate was fluctuating up to 31 days (31-3-2020) but after that

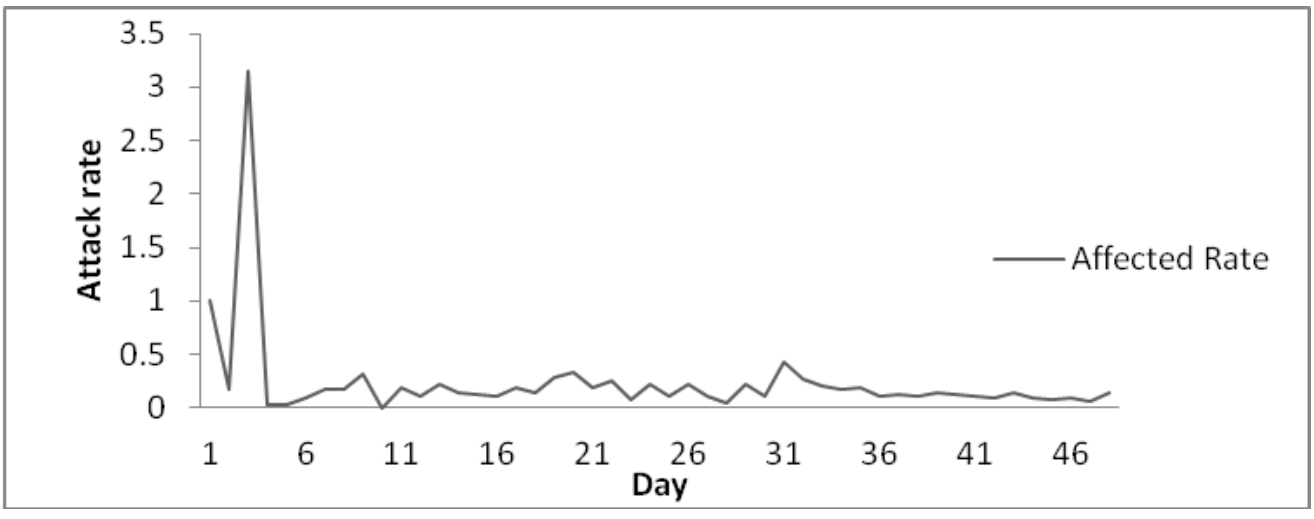

Figure 1: The graph of AR.

Table 1: Power curve model summary.

R	R Square	Adjusted R Square	Std. Error of the Estimate	p- value
0.816	0.666	0.645	0.273	0.000

Table 2: ANOVA of power curve.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	2.389	1	2.389	31.942	0.000
Residual	1.197	16	0.075		
Total	3.586	17			

Table 3: Coefficients for power curve.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
ln (t)	-2.740	0.485	-0.816	-5.652	0.000
(Constant)	3081.581	5482.579		0.562	0.582

Table 4: Cubic model summary.

R	R Square	Adjusted R Square	Std. Error of the Estimate	p-value
0.999	0.998	0.998	35.400	0.000

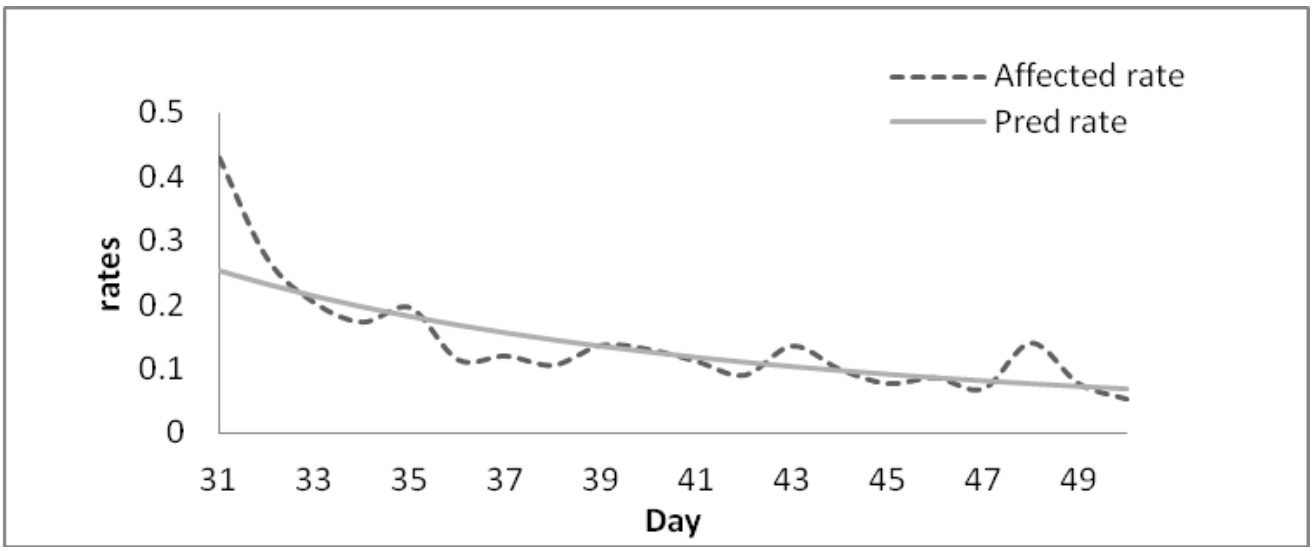

Figure 2: Graph of predicted AR.

it seems decreasing which shows the significant effect of lockdown period started form 23-3-2020. So we have used the attack rate from 31-3-2020 to build up a predictive model based on day time.

Power curve model is used for attack rate (rate of infection) against time (day) t based on the data 31-3-2020 to 17-4-2020. Table 1 shows that R-squared and adjusted R-squared are high with p-value 0.000 which shows the good capability of the model (Table 2).

From Table 3, the outcomes of the fitted model summary, we build up the prediction equation for attack rate on day t as

$$AR_t = 3081.581411 t^{-2.739871} \quad (1)$$

The above fitted power curve model is used to predict daily attack rate. A very simple mathematical equation for daily new cases for time period after 31 days can be determined as follow:

Predicted daily new cases on day t = AR_t (total cases on day (t-1)) (2)

Hence the predictive cumulative total cases up to day t can be computed by adding predicted daily new cases on day t to previous day total cases.

The following graph shows the observed values and predicted values of AR (Figure 2).

To predict total finished cases (TFC) up to day t the cubic model is developed as below:

From Table 4 and Table 5 we observed a very strong value of R-squared and adjusted R-squared. Which rev-

Table 5: ANOVA of cubic model.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	9406749.640	2	4703374.820	3753.231	0.000
Residual	18797.305	15	1253.154		
Total	9425546.944	17			

Table 6: Coefficients for cubic model.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
t	-154.143	13.978	-1.105	-11.027	0.000
t** 3	0.061	0.003	2.089	20.841	0.000
(Constant)	3119.071	363.839		8.573	0.000

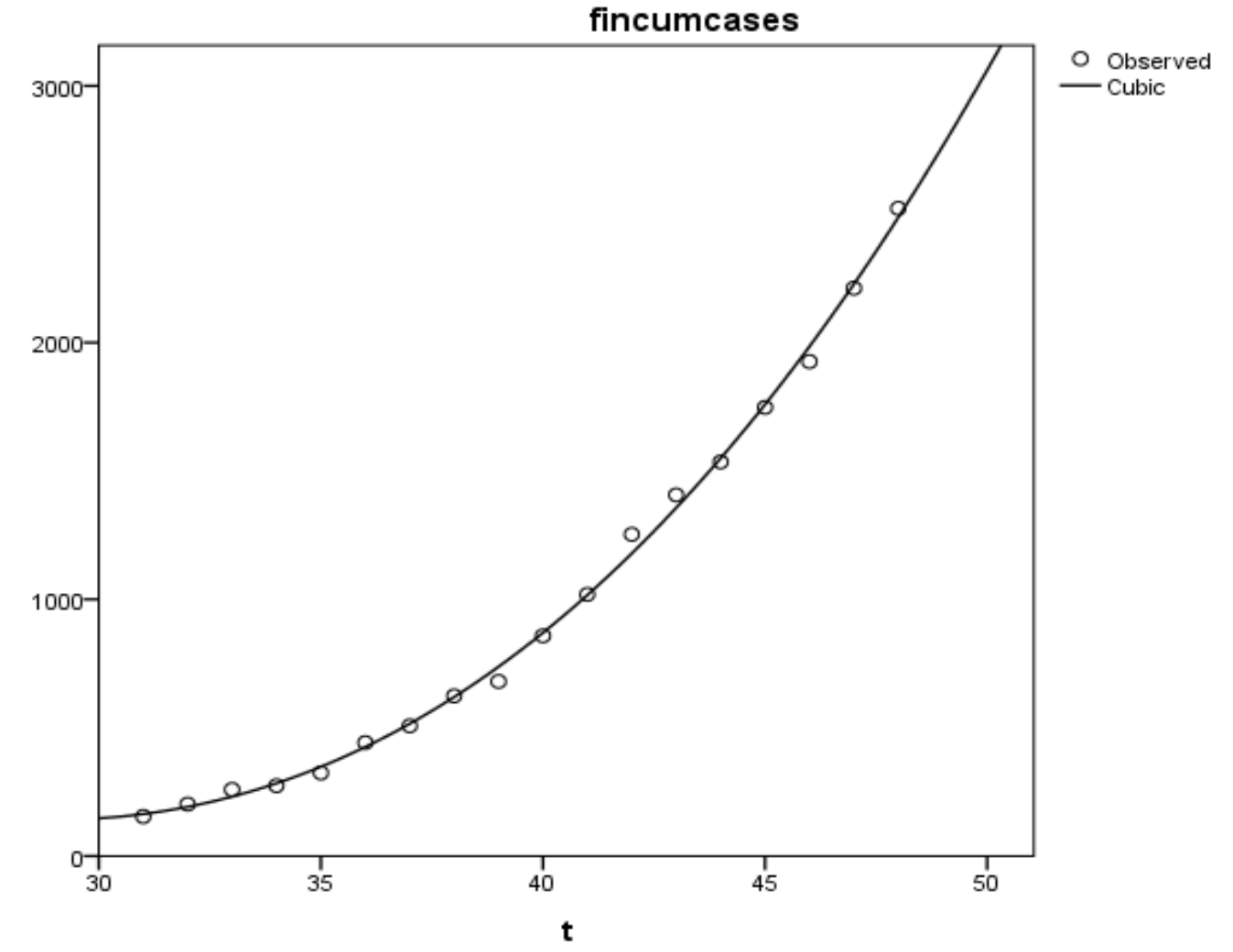

Figure 3: Graph of predictive total finished cases.

els the strong goodness of fit of the cubic model for TFC.

From the Table 6 we observed very high significant effect of time t and t^3 on total finished cases, since p-values are 0.000, which shows that the cubic model is highly reliable to predict total finished cases.

Thus, the cubic model for total (cumulative) finished cases up to day t is built up as

$$TFC = 3119.071 - 154.143t + 0.061t^3 \quad (3)$$

Predictive total finished cases can be obtained by the above model for future days.

From the Figure 2 and Figure 3 we say that the total (cumulative) finished cases will increases day by day with higher rate. And affect rate (infection rate) decreases day by day (after 31 days) may be due to strict lockdown effect and proper treatment to the affected cases.

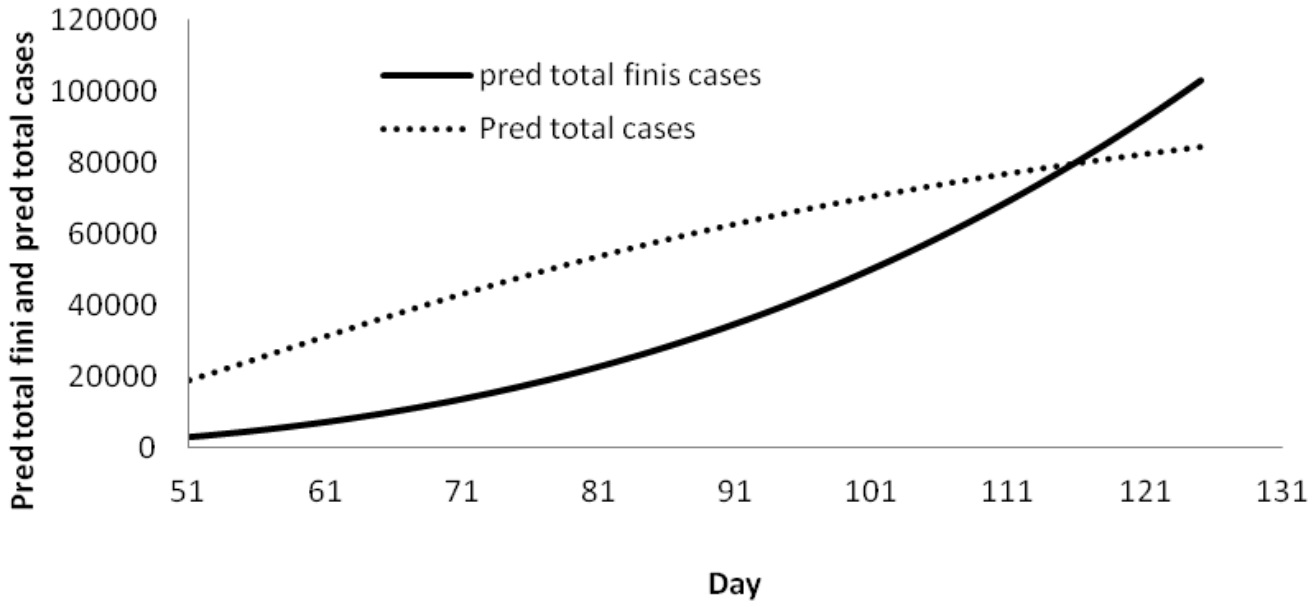

Figure 4: Graph of cumulative finished cases.

Table 7: Data Downloaded from Worldometer.

Date March-2020	Day (t)	DTC	DNC	TFC	AR	Date March/April -2020	Day (t)	DTC	DNC	TFC	AR
1	1	3	3	0	1	26	26	727	70	61	0.2200825
2	2	6	3	0	0.1666667	27	27	887	160	89	0.1127396
3	3	7	1	0	3.1428571	28	28	987	100	104	0.0374873
4	4	29	22	0	0.0344828	29	29	1024	37	118	0.2216797
5	5	30	1	0	0.0333333	30	30	1251	227	130	0.1167066
6	6	31	1	0	0.0967742	31	31	1397	146	154	0.4302076
7	7	34	3	0	0.1764706	1	32	1998	601	202	0.2727728
8	8	40	6	0	0.1750000	2	33	2543	545	259	0.2029099
9	9	47	7	0	0.3191489	3	34	3059	516	274	0.1729323
10	10	62	15	0	0.0000000	4	35	3588	529	324	0.1953735
11	11	62	0	0	0.1935484	5	36	4289	701	442	0.1140126
12	12	74	12	1	0.1081081	6	37	4778	489	507	0.1199247
13	13	82	8	8	0.2195122	7	38	5351	573	624	0.1055877
14	14	100	18	8	0.1400000	8	39	5916	565	680	0.1367478
15	15	114	14	11	0.1315789	9	40	6725	809	858	0.1301115
16	16	129	15	11	0.1085271	10	41	7600	875	1019	0.1113158
17	17	143	14	13	0.1818182	11	42	8446	846	1253	0.0898650
18	18	169	26	13	0.1479290	12	43	9205	759	1407	0.1355785
19	19	194	25	20	0.2835052	13	44	10453	1248	1535	0.0989190
20	20	249	55	24	0.3333333	14	45	11487	1034	1748	0.0768695
21	21	332	83	24	0.1927711	15	46	12370	883	1926	0.0856912
22	22	396	64	27	0.2601010	16	47	13430	1060	2212	0.0686523
23	23	499	103	40	0.0741483	17	48	14352	922	2523	0.1402592
24	24	536	37	46	0.2257463	18	49	16365	2013	2980	0.0763825
25	25	657	121	51	0.1065449	19	50	17615	1250	3409	0.0524553

The predictive effective cases up to day t is computed by subtracting predicted total finished cases from predicted total cases up to day t.

To know the break-even point the graph of predictive total cases and predictive total finished cases are plotted on the same graph as Figure 4.

Table 8: Table of Predictive Results.

Date	Day (t)	AR	PAR	TFC	PTFC	PDC	DTC	PDTC	PEC
20/04/2020	51	0.0831221	0.0646027	3861	3349	1198	18539	18812	15463
21/04/2020	52	0.0642430	0.0612555	4616	3681	1215	20080	20028	16347
22/04/2020	53	0.0781001	0.0581406	5047	4031	1227	21370	21255	17224
23/04/2020	54	0.0611138	0.0552379	5729	4401	1236	23039	22491	18090
24/04/2020	55	0.0751012	0.0525295	6272	4790	1242	24447	23733	18943
25/04/2020	56	0.0611422	0.0499992	6760	5200	1247	26283	24980	19780
26/04/2020	57	0.0559699	0.0476324	7400	5630	1249	27890	26229	20599
27/04/2020	58	0.0635972	0.0454159	8072	6081	1249	29451	27478	21397
28/04/2020	59	0.0554846	0.0433378	8751	6553	1248	31324	28726	22173
29/04/2020	60	0.0544734	0.0413874	9512	7046	1245	33062	29970	22924
30/04/2020	61		0.0395548		7562	1240		31211	23649
1/05/2020	62		0.0378313		8100	1235		32446	24346
2/05/2020	63		0.0362086		8661	1227		33673	25012
3/05/2020	64		0.0346795		9245	1219		34893	25648
4/05/2020	65		0.0332372		9852	1210		36103	26251
5/05/2020	66		0.0318755		10483	1200		37303	26820
6/05/2020	67		0.0305889		11138	1189		38491	27353
7/05/2020	68		0.0293721		11818	1177		39669	27851
8/05/2020	69		0.0282204		12522	1165		40834	28312
9/05/2020	70		0.0271295		13252	1152		41986	28734
10/05/2020	71		0.0260954		14007	1139		43125	29118

From the Figure 4 we observed that both the curves intersect between day 115 and 116. That means the rate of predictive finish cases would be more than that of predictive total cases after the day 115. Hence there will not be any pending effective cases after 115 days (i.e. after 23-6-2020) (Table 7 and Table 8).

Conclusions

The spread of COVID-19 epidemic has been slow in India compared to other countries like USA and Italy due to the social distancing measures taken by the government of India. In our study we observed that due to lockdown measure the affect rate was decreasing, though the rate is slow but it has also a positive effect on total cases. We also observed that the total finished cases increases, which is the outcome of the good efforts taken the government and medical staffs. On the basis of the current environment and situations we have tried to fit two models (i) Power curve model for attack rate and (ii) Cubic model for total finished cases. The following conclusions have been formulated on the basis of the proposed analysis.

1. The proposed power curve model is highly reliable in predicting attack rate in India.
2. The R values and adjusted R values both are high with p-value 0.000, which shows the higher chance of reliable prediction about the attack rate.
3. The cubic model used to predict total finished cases found highly reliable for prediction since R-squared and R-squared adjusted values are very much high and significant. Also the time variables t and t^3 both are significant with p-values 0.000.

   We also observed that total finished cases increases heavily day by day. This shows the good efforts done by central and state governments and medical staffs.
4. Daily new cases decrease from the end of April-2020.
5. As total cases decreases and finished cases increases, a break-even time is observed on 23-06-2020. That means after 23-06-2020 there will be not be any actual affected cases left. Hence the spread of COVID-19 cases would be under control.

Thus overall, it can be said that the power curve model and cubic models are highly reliable and can be used as predictive model for predicting attack rate and total finished cases in India. According to our analysis, if situation continue in present state; after 6-05-2020 the attack rate decreases to 3 per day per 100 total cases and will be 1 per 100 total cases after 9-06-2020. We suggest that the government should take more effective actions to control the spreads of COVID-19 cases. The outcomes of our analysis would help planning for the future course of action to the government and the medical team associated with treatment to COVID-19 cases. The projections are based on the current trend and may fail if an unforeseen development occurs.

Funding

None.

Declaration of Competing Interest

None.

Acknowledgements

The author is grateful to the Editor and the anonymous referee for making many helpful comments and suggestions on an earlier version of this paper.

References

1. Rachah A, Torres DFM (2018) Analysis, simulation and optimal control of a SEIR model for Ebola virus with demographic effects. Commun Fac Sac Univ Ank Series 67: 179-197.
2. Porter AT (2012) A path-specific approach to SEIR modeling. University of Iowa.
3. Binti-Hamzah FA, Lau C, Nazri H, Ligot DV, Lee G, et al. (2020) CoronaTracker: Worldwide COVID-19 outbreak data analysis and prediction.
4. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, et al. (2020) Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Medicine and Infectious Disease 34: 101623.
5. Kumar SU, Kumar DT, Christopher BP, Doss C (2020) The rise and impact of COVID-19 in India. Frontiers in Medicine 7: 250.
6. Salgotra R, Gandomi M, Gandomi AH (2020) Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. Chaos, Solitons, and Fractals 138: 109945.
7. Ghosal S, Sengupta S, Majumder M, Sinha B (2020) Linear regression analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases - March 14th 2020). Diabetes & Metabolic Syndrome 14: 311-315.
8. Schueller E, Klein E, Lin G, Tseng K, Balasubramanian R, et al. (2020) COVID 19 in India: Potential impact of the lockdown and other long term policies. CDDEP.
9. Ranjan R (2020) Predictions for COVID-19 outbreak in India using epidemiological models. BMJ Yale.