*epi*H

Epidemiology
and Health

HEALTH STATISTICS

Estimation of the reproduction number and early prediction of the COVID-19 outbreak in India using a statistical computing approach

Karthick Kanagarathinam[1], Kavaskar Sekar[2]

[1]Department of EEE, GMR Institute of Technology, Rajam, India; [2]Department of EEE, Panimalar Engineering College, Chennai, Tamilnadu, India

Coronavirus disease 2019 (COVID-19), which causes severe respiratory illness, has become a pandemic. The World Health Organization has declared it a public health crisis of international concern. We developed a susceptible, exposed, infected, recovered (SEIR) model for COVID-19 to show the importance of estimating the reproduction number (R_0). This work is focused on predicting the COVID-19 outbreak in its early stage in India based on an estimation of R_0. The developed model will help policy-makers to take active measures prior to the further spread of COVID-19. Data on daily newly infected cases in India from March 2, 2020 to April 2, 2020 were to estimate R_0 using the earlyR package. The maximum-likelihood approach was used to analyze the distribution of R_0 values, and the bootstrap strategy was applied for resampling to identify the most likely R_0 value. We estimated the median value of R_0 to be 1.471 (95% confidence interval [CI], 1.351 to 1.592) and predicted that the new case count may reach 39,382 (95% CI, 34,300 to 47,351) in 30 days.

KEY WORDS: Basic reproduction number, COVID-19, Forecasting, Statistical computing

INTRODUCTION

Coronavirus disease 2019 (COVID-19) has rapidly spread worldwide, with 896,450 confirmed total new cases and 45,526 deaths globally as of April 2, 2020 [1]. The disease emerged as 27 cases of pneumonia with an unknown cause in Wuhan, China. The first COVID-19 case in India was identified on January 30, 2020, and the total number of reported cases reached 2,322 as of April 3, 2020 [2]. On March 3, 2020, the Indian government suspended all new visas and visas issued to nationals of Iran, Italy, Japan, and Korea, and on the next day implemented compulsory screening of all international passengers. The Indian government

declared a countrywide lockdown for 21 days on March 24, 2020 as a measure to control the spread of COVID-19, which has developed into a pandemic. The transmission rate of COVID-19 has been relatively low in most countries, but with major outbreaks in a few countries, such as Iran, Italy, Japan, and Korea. Most countries have at least an early stage of COVID-19 spread before any mitigation measures have an impact [3]. Myers et al. [4] stated that accurate epidemic forecasting models would noticeably improve epidemic prevention and control capabilities. No vaccine is available for COVID-19, and vaccination is typically not a good option for stopping the spread of a new epidemic, as considerable time is required to develop a safe and effective vaccine (approximately 10 years) [5]. Li et al. [6] found that the COVID-19 incubation period was 5.2 days (95% confidence interval [CI], 4.1 to 7.0) and found indications that human-to-human transmission occurred among close contacts. India is the second most populated country, it is important to estimate the transmissibility of COVID-19 and to predict the total number of new cases, which will help direct focus towards this public health crisis. Mathematically based epidemic models, such as susceptible-infected-recovered (SIR) models [7], susceptible-infected-susceptible (SIS) models [8], susceptible-exposed-infected-recovered (SEIR) models [9], and susceptible-exposed-infected-recovered-susceptible

This article is available from: https://e-epih.org/

(SEIRS) models [10] are used to predict the trajectory of epidemics. Estimating the reproduction number (R_0) can be estimated statistically or empirically. In this work, we used the earlyR (https://cran.r-project.org/) package to estimate R_0 and predict the trajectory of the outbreak.

METHODS

Susceptible-exposed-infected-recovered-susceptible mathematical model

SEIR models can be used to predict the number of people infected based on R_0. We have given a SEIR model in this study to demonstrate the importance of estimating R_0 [11]. COVID-19 has an incubation period, also known as a latent period or latent delay (τ), of 2-14 days. The following assumptions were made for developing the mathematical model for COVID-19.

- The population growth of the region/country is exponential, and the COVID-19 epidemic is occurring in a sufficiently short period
- Infected individuals are assumed not to give birth
- Recovered individuals acquire permanent immunity with a probability $f (0 \leq f \leq 1)$ or die from the disease with a probability of $(1-f)$

With S referring to susceptible individuals, E to susceptible individuals that become exposed at time t-τ, I to individuals who are infected, and R to those who have recovered from COVID-19, the resulting differential equations are:

$$\frac{dS(t)}{dt} = b\, s(t) + bE(t) + bR(t) - \mu s(t) - \gamma \frac{I(t)\, S(t)}{N(t)} \qquad (1)$$

$$\frac{dE(t)}{dt} = \gamma \frac{I(t)S(t)}{N(t)} - \gamma \frac{I(t-\tau)\, S(t-\tau)}{N(t-\tau)} e^{-\mu\tau} - \mu\, E(t) \qquad (2)$$

$$\frac{dI(t)}{dt} = \gamma \frac{I(t-\tau)\, S(t-\tau)}{N(t-\tau)} e^{-\mu\tau} - \mu\, I(t) - \alpha\, I(t) \qquad (3)$$

$$\frac{dR(t)}{dt} = -\mu\, R(t) - f\, \alpha\, I(t) \qquad (4)$$

Where μ is the per capita death rate due to causes other than the disease, γ is the rate of contact (or) transmission rate (or) infection rate, α is the recovery rate, and b is the per capita birth rate (with $b > \mu$).

At any instant,

$$S(t) + E(t) + I(t) + R(t) = N(t) \qquad (5)$$

R_0 is defined as,

$$R_0 = \frac{\gamma\, e^{-b\tau}}{b+\alpha} \qquad (6)$$

This constant is extremely important in characterizing the spread of COVID-19. It reflects how many people contract the disease from an infectious individual. In general, If $R_0 > 1$, secondary infections will occur and the disease is spreading throughout the population. According to WHO information as of January 23, 2020, the R_0 of COVID-19 lies between 1.4 and 2.5. R_0 may vary considerably for different infectious diseases, but also for the same disease in different populations [12].

Data

All the data shown in Table 1 were collected from an Indian official website [2]. The epidemiological data from March 2, 2020 to April 2, 2020, as shown in Table 1, were utilized to estimate R_0. A higher R_0 indicates a higher likelihood of new infections.

Model development

The transmissibility of COVID-19 in India was evaluated using the earlyR package. It was assumed that interventions so far have had a minimal impact on COVID-19 transmission in India. The model used herein is a simplified version of the model introduced by Cori et al. [13]. Serial interval distributions (i.e., mean and standard deviation [SD]) are required to estimate R_0. We assumed that the mean and SD were 4.7 days and 2.9 days, respectively, based on existing research [14]. The maximum-likelihood (ML) approach was applied to obtain the distribution of R_0. The bootstrap strategy was applied for re-sampling 1,000 times to obtain likely R_0 values. The R package projection was used to predict the cumulative daily incidence [15]. We forecast the cumulative total new cases after 30 days. The daily incidence obeys a Poisson distribution determined by daily infectiousness, which is denoted as,

$$\lambda(t) = \sum_{k=1}^{t-1} X_k\, V\,(t-k) \qquad (7)$$

Where $V(t-k)$ the vector of the probability mass function and X_k is the real-time incidence at time k. The forecasting model depended on the present incidence and serial interval distributions. The projections were based on resampling and probability computations. The statistical analysis and model development were done using R version 3.6.3 (https://cran.r-project.org/bin/windows/base/old/3.6.3/).

Table 1. Actual coronavirus disease 2019 daily new confirmed cases in India

Date in 2020	New confirmed cases (n)	Date in 2020	New confirmed cases (N)
Mar 2	2	Mar 18	14
Mar 3	1	Mar 19	22
Mar 4	22	Mar 20	50
Mar 5	2	Mar 21	60
Mar 6	1	Mar 22	77
Mar 7	3	Mar 23	74
Mar 8	5	Mar 24	85
Mar 9	5	Mar 25	87
Mar 10	6	Mar 26	88
Mar 11	10	Mar 27	140
Mar 12	13	Mar 28	84
Mar 13	8	Mar 29	106
Mar 14	16	Mar 30	227
Mar 15	10	Mar 31	146
Mar 16	11	Apr 1	437
Mar 17	19	Apr 2	235

Ethics statement

The analysis in the article is based on data which is open to public. The article does not require the ethical committee approval.

RESULTS AND DISCUSSION

Figure 1 shows the daily incidence of COVID-19 in India from March 2, 2020 to April 2, 2020. Figure 2 shows the distribution of likely values of the R_0 of COVID-19 in India. We estimated the ML value of R_0 as 1.471 (95% CI, 1.351 to 1.592) for COVID-19 in the early stage in India. Figure 3 shows a histogram of R_0 values using the bootstrap strategy with 1,000 likely samples.

Figure 4 shows the global spread of COVID-19 during the same period. The vertical gray bars indicate the presence of cases and black dots denote the dates of symptom onset. The dashed vertical blue line indicates the current date (April 3, 2020). The vertical scale in Figure 4 shows the relative scale of infections. Figure 5 shows the predicted cumulative cases in next 30 days.

We computed that the cumulative number of new cases may reach 39,382 (95% CI, 34,300 to 47,351) in the next 30 days. The R_0 data were estimated based on the existing COVID-19 data from March 2, 2020 to April 2, 2020. The Indian government has already announced a nationwide lockdown. As per the WHO information on January 23, 2020, the R_0 of COVID-19 lies between 1.4 and 2.5. Our estimation indicates that for India, the median R_0 value of 1.471 (95% CI, 1.351 to 1.592) is in the lower range. However, various studies have indicated that precisely estimating R_0 is challenging, because R_0 depends on environmental conditions, demography, and the modeling method. In our method, the accuracy of R_0 depended on the premise that all cases of
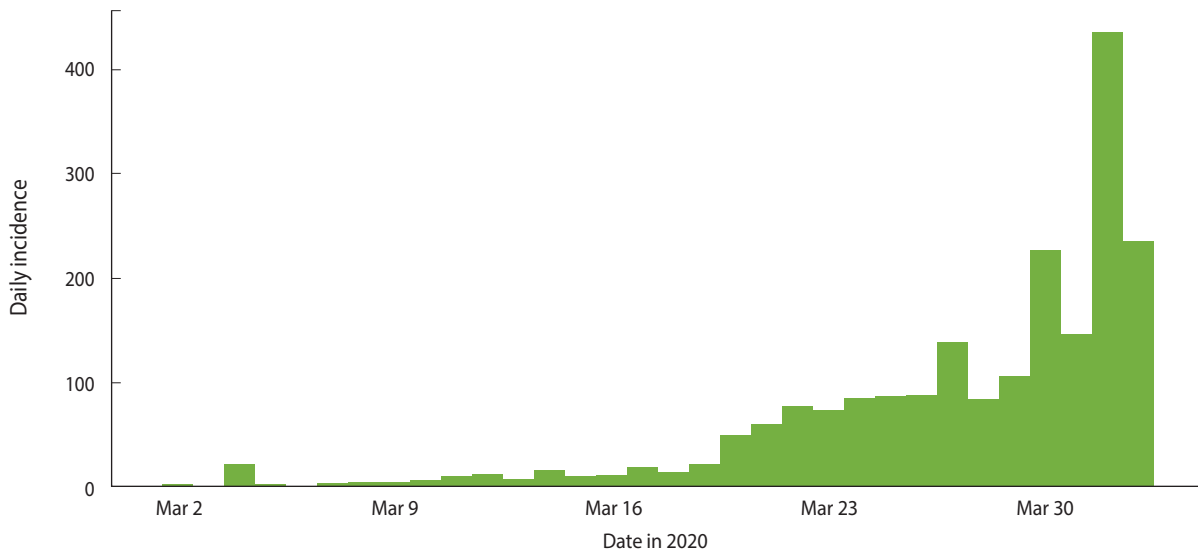

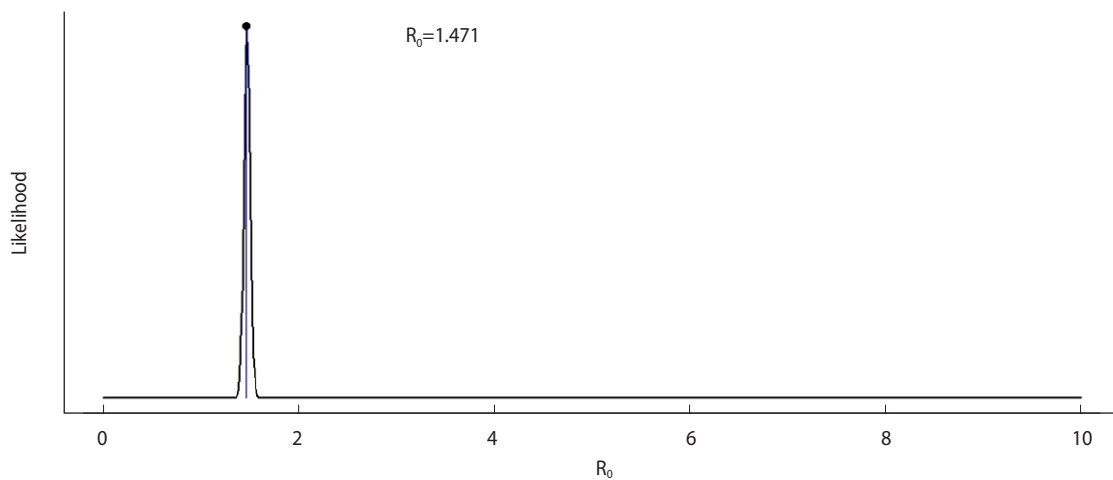**Figure 1.** Actual daily incidence of coronavirus disease 2019 in India.

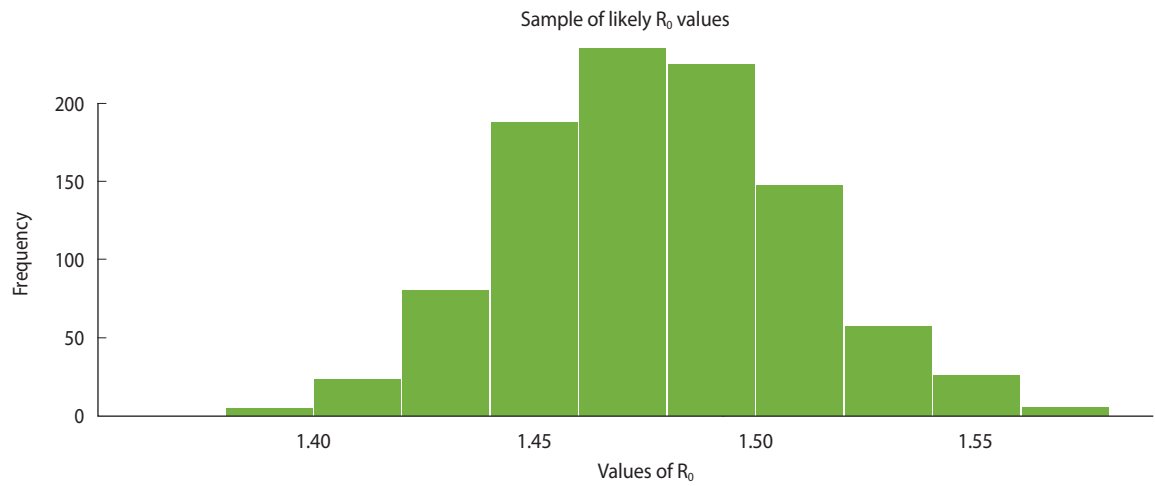**Figure 2.** Maximum-likelihood value of reproduction number (R_0).

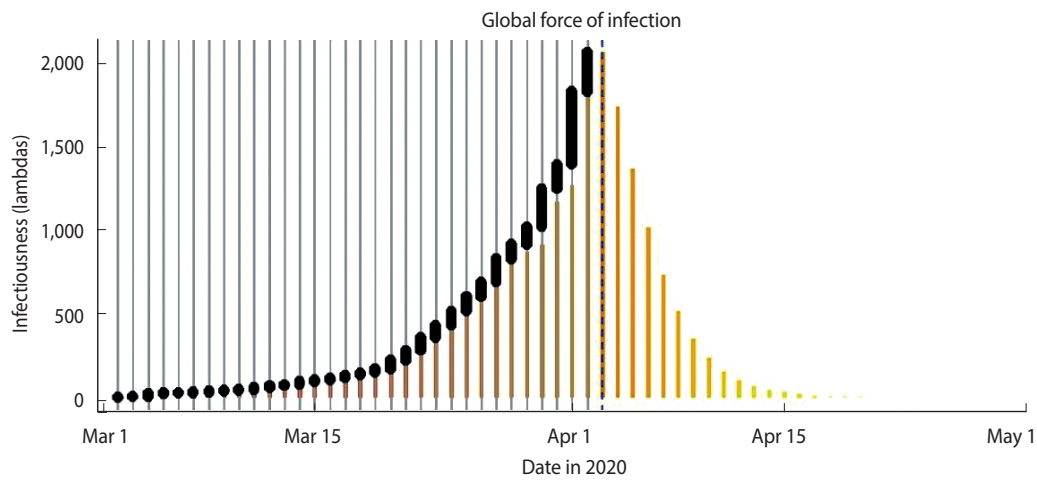**Figure 3.** Sample of likely values of reproduction number (R_0).

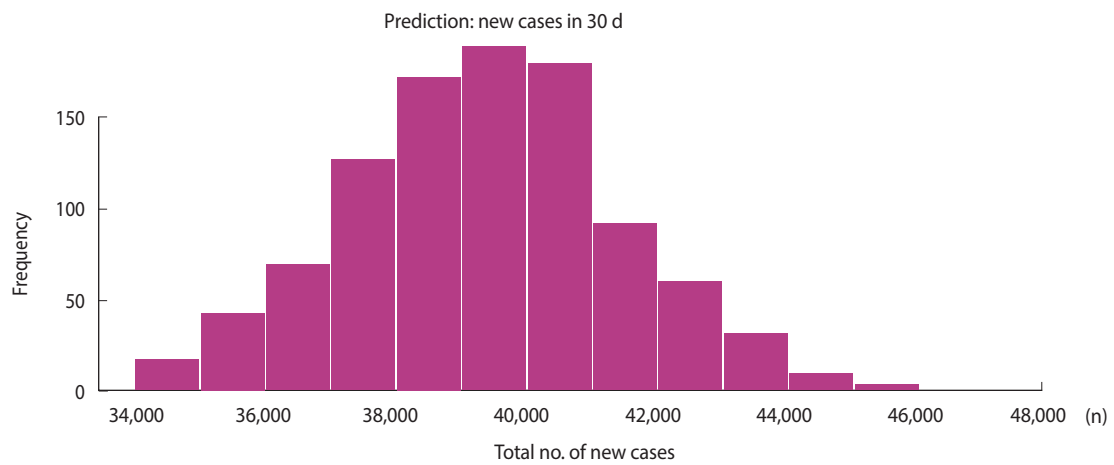**Figure 4.** Global spread of infections.

Figure 5. Predicted cumulative new cases in the next 30 days.

COVID-19 in India were identified in the study period. If the same scenario continues, we predict that the cumulative number of new cases may reach 39,382 (95% CI, 34,300 to 47,351) in next 30 days. We believe that our forecasting numbers may help in various aspects, such as developing the required medical infrastructure and focusing efforts on mitigating the economic impact of the pandemic. Our findings were derived based on a limited time frame, and the results may change after the occurrence of a considerable number of additional cases. The R_0 value corresponding to the spread of COVID-19 can be controlled by strictly following social distancing in daily life, wearing masks, frequent hand-washing with soap or sanitizers, quarantining infected people, identifying cases using rapid diagnostic methods, and so on.

CONCLUSION

We estimated the median value of R_0 to be 1.471 (95% CI, 1.351 to 1.592) and predicted that the cumulative number of new cases may reach 39,382 (95% CI, 34,300 to 47,351) in the next 30 days. The predicted size largely depends on changes in R_0. Effective measures against COVID-19 will help to reduce R_0. The presence of numerous unidentified cases in the study period may result uncertainties in the estimated value of R_0 used in the developed forecasting model.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare for this study.

FUNDING

None.

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTIONS

Conceptualization: KK. Data curation: KS. Formal analysis: KS. Funding acquisition: None. Methodology: KK. Writing – original draft: KK. Writing – review & editing: KK, KS.

ORCID

Karthick Kanagarathinam: *https://orcid.org/0000-0001-7755-5715*; Kavaskar Sekar: *http://orcid.org/0000-0003-4735-0537*

REFERENCES

1. World Health Organization. Coronavirus disease 2019 (COVID-19) situation report-73 [cited 2020 Apr 3]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200402-sitrep-73-covid-19.pdf?sfvrsn=5ae25bc7_2.
2. Ministry of Health and Family Welfare, Government of India. COVID-19 India [cited 2020 Apr 3]. Available from: http://www.mohfw.gov.in/index.html#.
3. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020;395:931-934.
4. Myers MF, Rogers DJ, Cox J, Flahault A, Hay SI. Forecasting disease risk for increased epidemic preparedness in public health. Adv Parasitol 2000;47:309-330.
5. Pronker ES, Weenen TC, Commandeur H, Claassen EH, Osterhaus AD. Risk in vaccine research and development quantified. PLoS One 2013;8:e57755.
6. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020;382:1199-1207.
7. Eksin C, Paarporn K, Weitz JS. Systematic biases in disease forecasting - the role of behavior change. Epidemics 2019;27:96-105.
8. Pinsent A, Liu F, Deiner M, Emerson P, Bhaktiari A, Porco TC, et al. Probabilistic forecasts of trachoma transmission at the district level: a statistical model comparison. Epidemics 2017;18:48-55.
9. Funk S, Camacho A, Kucharski AJ, Eggo RM, Edmunds WJ. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. Epidemics 2018;22:56-61.
10. Khan MA, Badshah Q, Islam S, Khan I, Shafie S, Khan SA. Global dynamics of SEIRS epidemic model with non-linear generalized incidences and preventive vaccination. Adv Differ Equ 2015:88.
11. Yan P, Liu S. SEIR epidemic model with delay. ANZIAM J 2006;48:119-134.
12. Dietz K. The estimation of the basic reproduction number for infectious diseases. Stat Methods Med Res 1993;2:23-41.
13. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 2013;178:1505-1512.
14. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020;93:284-286.
15. Jombart T, Nouvellet P, Bhatia S, Kamvar ZN. Projections: project future case incidence; 2018 [cited 2020 Apr 3]. Available from: https://cran.r-project.org/web/packages/projections/index.html.