

Original Research Article

Estimation of R_0 for COVID-19 in India through different mathematical model and their comparison

Nidhi Dwivedi¹, Sujata Gupta^{2*}, Archana Dwivedi³

¹Department of Community Medicine, North DMC Medical College and Hindu Rao Hospital, Delhi, India

²Department of Community Medicine, Government Medical College Kathua, Jammu, J and K, India

³Department of Neurology, IMS Banaras Hindu University, Lanka Varanasi, India

Received: 19 October 2020

Revised: 18 December 2020

Accepted: 29 December 2020

*Correspondence:

Dr. Sujata Gupta,

E-mail: sujata080583@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: The cases of novel coronavirus (COVID- 2019)-infected pneumonia started since the 19th of December, 2019, in Wuhan (Central China). A large scale outbreak of the disease resulted in a pandemic. This outbreak of the COVID -19 disease has spread on a wide scale. World health organization (WHO) has identified the ongoing outbreak of corona virus disease (COVID 2019) as pandemic on 11 March 2020. Basic reproduction number (R_0)- is one of the most important predictors of epidemic severity. It can help to understand the path of the epidemic and to assess the effectiveness of the various interventions to control the epidemic. The purpose of this study is to estimate R_0 by using five methods based on the Indian COVID-19 dataset and compare them.

Methods: We obtained data on daily confirmed, recovered and deaths cases from official site of ministry of health and family welfare. We implemented 5 mathematical methods to calculate R_0 . We estimated the number of active cases till 14th of April. We also compare these methods to find out the best method to predict R_0 .

Results: The estimated R_0 for the AR, EG, ML, TD, and gamma-distributed methods were 1.0004, 2.102, 1.895, 1.872 and 1.46 respectively. The computed R_0 in the TD method is closer to the actual R_0 and have a good fit on data as confirmed with MSE criterion.

Conclusions: Awareness of the basic reproduction number of COVID-19 is useful for controlling the spread of disease and for planning. It is therefore necessary to know the best method that has better performance.

Keywords: Corona virus, India, COVID-2019, Reproduction number, Pandemic

INTRODUCTION

The 2019 novel coronavirus (2019-nCoV) or the severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) as it is now called has spread from its origin in Wuhan City of Hubei province of china to the rest of the world.¹ The transmission route of 2019-nCoV is through the respiratory tract droplets. At present, the majority of cases are related to close contact and the population is generally susceptible.² According to the World health organization

(WHO) COVID 19 situation report 85 as of 14 April 2020, total of 18,44,863 confirmed deaths and 1,17, 021 worldwide reported deaths. Global spread has been rapid, with 209 countries reporting at least one case.³ In India, a total of 10,815 confirmed cases and 352 deaths have been reported by 14 April 2020.

R_0 is an important parameter for the epidemiological model. It represents the average number of cases that are transmitted during the disease period when all people are susceptible at the onset of the disease.⁵⁻⁷ When $R_0 > 1$, the

epidemic will erupt; otherwise the epidemic will die out. In the early stages of the outbreak, new outbreaks of infectious diseases will occur. In the case of newly emerging outbreaks of infectious diseases, the number of cases will in most cases show an exponential growth pattern in the early stages of the outbreak. The generation interval distribution determines the functional relationship between the exponential growth rate and the R_0 .

A number of studies have been conducted to estimate the epidemiological parameters of COVID-2019 worldwide. Read et al envisaged that the R_0 of 2019-nCoV would be between 3.6 and 4.0 with the use of the spread model.⁶ As for Tang et al it is necessary to reduce the transmission of the virus between 72% and 75% by means of various methods of prevention and control in the community in order to prevent the spread of the epidemic.⁷ Liu et al used 830 confirmed cases of 2019-nCoV as of 23 January 2020 and anticipated that the common incubation period would be 4.8 days.⁸ The estimated R_0 for EG (exponential growth) and ML (maximum likelihood) techniques was 2.90 (95 per cent CI: 2.32 ~ 3.63) and 2.92 (95 per cent CI: 2.28~3.67) respectively. Wu et al. used SEIR-based meta population version to simulate epidemics in predominant cities, and used Markov chain Luo technique estimates R_0 is 2.68 (95 per cent CI: 2.47~2.86).⁹

Under current and ongoing intervention measures, it is of great importance to analyze the evolutionary trends, epidemiological characteristics and dynamics of the epidemic on a continuous and timely basis, which can provide the basis for the scientific prevention and control of the COVID-19 epidemic.

The purpose of this study is to use five different methods to estimate R_0 of nCoV-2019 based on the Indian COVID-19 dataset and compare them to find out the method best predicting R_0 closest to the actual value.

METHODS

Data were derived from the daily cumulative number of cases updated daily by the Ministry of Health and Family Welfare, as well as from the cases reported on the relevant websites and close contact groups. As of 14 April 2020, there were 10,815 confirmed cases reported in India. All the confirmed cases as par the site has been included in the study. Information was collected on three main components of interest, i.e. daily confirmed cases, daily recovered cases, and on a daily basis from 2 March to April 14 2020.

The reproduction number (R_0) was calculated by five different mathematical methods. The models used in this article included the attack rate (AR), exponential growth rate (EG), maximum likelihood (ML), time-dependent reproduction numbers (TD), gamma-distributed generation time (GT). The above mentioned methods were applied using R software (R_0 package and

programming) for R_0 calculation. After the calculation of R_0 by these methods we have also compare them to find out the model which gives value close to the actual R_0 at that particular time.

Formulae used for calculating R_0 are given below

The attack rate

The R_0 can be described by the AR with the following formula:

$$R_0 = -\frac{\log\left(\frac{1-AR}{S_0}\right)}{AR - S_0}$$

where AR defines the ratio of the people generating an infection disease and S_0 show the initial susceptible ratio.¹⁰

The exponential growth rate

The following formula was applied for computing the R:

$$\mu_t = R \left(\sum_{i=1}^t N_{t-i} w_i \right)$$

$$R = \frac{1}{M(-r)}$$

In this formula, M is the moment-generating function of the GT.¹¹ The parameter r is determined by the poisson regression. Furthermore, the parameter w is GT.

The maximum likelihood

Let N_0, N_1, \dots, N_T identify incident cases over sequential time. The log-likelihood function is:

$$LL(R) = \sum_{t=1}^T \frac{\exp(-\mu_t) \mu_t^{N_t}}{N_t!}$$

Where,

$$\mu_t = R \left(\sum_{i=1}^t N_{t-i} w_i \right)$$

R is the maximum value of the log-likelihood function.¹² Furthermore, the parameter w is estimated by maximizing log-likelihood is GT.

Time-dependent reproduction numbers

In this method, R_t is computed by averaging R_j , which is the mean of all transmission networks corresponding to the cases observed.¹³

$$R_t = \frac{1}{N_t} \sum_{\{t_j=t\}} R_j$$

Where,

$$R_j = \sum_i p_i$$

and

$$p_{ij} = \frac{N_i w(t_i - t_j)}{\sum_{i \neq k} N_i w(t_i - t_k)}$$

Consider that person i and person j are in times t_i and t_j , respectively, then displays the probability of infection transmission from person j to person i so R_t compute by averaging all R_j which is the mean of all transmission networks correspondent with the cases that observed.

The gamma-distributed generation time

Generation time is the time-gap between infection of a primary case and infection of a secondary case that is generated by the primary case.¹⁴

The number of cases on the day “t,” denoted by n_t in (t_1, t_2) grows exponentially where,

$$n_t = n_{t_1} \exp [r(t - t_1)]$$

$$r = \frac{\text{cov}(\{\log \log (n_t)\}_{t_n[t_1, t_2]}, [t_1, t_2])}{\text{var}([t_1, t_2])}$$

And

$$R = \left(1 + \frac{r}{b}\right)^a$$

The EG denotes by r. The mean and standard deviation of the GT are μ and σ , respectively, where $a = \mu^2 / \sigma^2$ and $b = \mu / \sigma^2$.¹⁵

Comparison of methods

In order to explore the proximity of the estimation of the above methods to the actual R_0 s and to compare them with each other, we used a 10,000-fold simulation for each formula based on the Covid-19 data from India. The epidemics were simulated with the following characteristics. The GT distribution was considered gamma with a mean of 5 and a standard deviation of 1.5. According to actual data (COVID-19 data), the duration of the epidemic is assumed to be 180 days. We therefore used a value equal to 100 for the peak value in the simulation command. The simulation of the basic reproduction number was made with the characteristics mentioned above and the MSE was calculated to evaluate

the performance of the models with the formula below. The lowest MSE value is the method that best fits the data.

$$MSE = \frac{\sum_{i=1}^n (r_{0i} - R_0)^2}{n - 1}$$

RESULTS

Based on available data as of 14 April 2020 (Ministry of health and family welfare, there were 10,815 confirmed cases and 352 deaths. Figure 1 shows the daily dataset of COVID-19 in India for the 44-days period in the study from 02 March, 2020 to 14 April 2020.

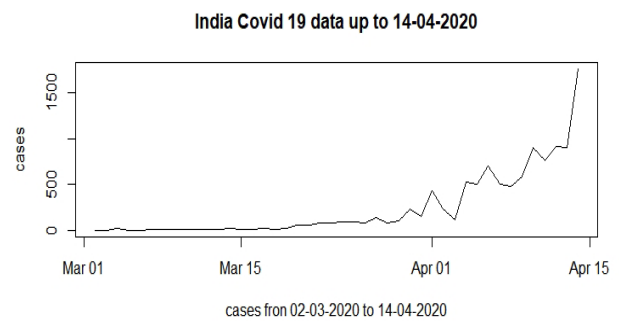


Figure 1: The incidence case counts COVID 19 data of India during 02 March 2020–14 April 2020.

All dates of the Indian data are based on month/day form 02 March, 2020 to 14 April, 2020. Moreover, the number of infected people is plotted as frequency.

In order to demonstrate the difference in modeling with various formulas, the results of the five models are presented in Table 1 in form of R_0 (95% confidence interval [CI]). R_0 (95% confidence interval [CI]) using the EG model is 2.102 [2.072-2.132] which highest among all of the considered models that means every person infected 2.102 other people on average during the infection period. R_0 (95%CI) estimated by ML is 1.895 [1.844-1.945] and TD is 1.872 [1.869-2.013] which are also quite high. The Gamma-distributed GT method estimate R_0 as 1.46 [1.39-1.57]. The estimated R_0 (95% CI) by the AR is 1.0004 [1.0003-1.0004] which is minimum computed R_0 (95% CI).

In order to compare the mentioned models to find the formula with better fit to the actual values, we conducted a simulation with R software and calculated R_0 based on the five models reported in Table 2. We used gamma distribution for the GT with the mean of 5 and standard deviation of 1.5. As the number of cases in India is increasing but the actual Value of R_0 is continuously decreasing [3-1.4] we are assuming that the length of epidemic may be 180 days. Then, using the above parameters, the simulation is implemented and R_0 was computed for each method. The simulation results for

comparing the quality of the five methods are represented in Table 2. In order to carry out the simulation, the number of runs to achieve the R 0 is 10000.

The results, given in Table 2, indicated that there were differences between the actual and simulated R 0; however, the TD method had the closest value to the R0 calculated from the simulation compared to the other methods. Variation in all of the simulated values of R0 by all the method has been considered for small actual values of R0 as compared to high values. For smaller values of R0 all the simulated values for all the methods are closer to actual values but the difference between the actual and simulated values of R0 is high as actual value of R0 is high. The computed R 0 by the AR methods seemed likely to reflect stability for all R 0 s.

Table 1: The reproduction number estimation by the different methods for the COVID-19 data (India up to 14, April 2020).

S. no	Method	R0 (95% CI for R0)
1.	AR	1.0004 [1.0003-1.0004]
2.	EG	2.102[2.072-2.132]
3.	ML	1.895 [1.844-1.945]
4.	TD	1.872 [1.869-2.013]
5.	Gamma-distributed generation time	1.46 [1.39-1.57]

AR based on incidence (n=10,742), R0: Reproduction number; TD=Time-dependent reproduction numbers; ML=Maximum likelihood; EG=Exponential growth rate; AR=Attack rate, CI=Confidence interval

Table 2: The simulated R0s and their 95% confidence interval for each method.

Actual R0	R0 (95% CI)				
	AR	ML	EG	TD	Generation Time
1.36	1.000003 [1.000003-1.000004]	1.502 [1.391-1.681]	1.526 [1.460-1.532]	1.449 [1.213-1.604]	1.481 [1.212- 1.758]
1.55	1.000004 [1.000003-1.000005]	1.670 [1.583-1.889]	1.730 [1.691-1.750]	1.649 [1.407-1.728]	1.543 [1.277- 1.816]
1.83	1.000005 [1.000004-1.000006]	1.893 [1.804-1.976]	2.001 [1.268-2.302]	1.873 [1.641-1.986]	1.647 [1.380- 1.917]
2.8	1.000009 [1.000008-1.000009]	1.693 [1.588-1.907]	1.721 [1.659-1.734]	1.689 [1.674-1.733]	1.622 [1.353- 1.894]
3.2	1.000003 [1.000003-1.000004]	1.26 [1.014-1.352]	1.338 [1.281-1.365]	1.224 [1.186-1.245]	1.560 [1.293- 1.821]

Table 3: Mean squared error of reproduction number estimation for each method

Actual R0	Method				
	AR	ML	EG	TD	Generation Time
1.36	0.187	0.029	0.063	0.028	0.046
1.55	0.204	0.027	0.057	0.022	0.040
1.83	0.550	0.041	0.051	0.019	0.022
2.8	2.268	1.354	0.867	0.149	0.778
3.2	4.064	3.101	2.450	0.318	2.096

For evaluating the performance of models, we computed MSE for all methods (Table 3). The TD method had the lowest MSE value in comparison to other methods. The MSE of AR method was very varied. As noted above, the TD introduced the approach with the nearest estimation to the actual R 0 based on MSE criterion.

DISCUSSION

Five methods (ML, EG, TD, AR, gamma distributed) were considered for the estimation of the R0 which is a key parameter for understanding the dynamics of an epidemic. This was based on the Indian COVID-19 dataset (from 02 March 2020 to 14 April 2020). The estimated R0 for the AR, EG, ML, TD and gamma distributed methods was 1.0004, 2.102, 1.895, 1.872 and

1.46 respectively. In most cases, the R0 was greater than that of unity but less in comparison with the values of the SARS epidemics (R0 = 4.91) in Beijing, China, in 2003 16, and MERS in Jeddah (R0=3.5–6.7) and Riyadh (R0=2.0–2.8), Kingdom of Saudi Arabia, in 2014.¹⁷

Furthermore, we noted that more recently published studies based on datasets during periods comparable with ours reported higher values of basic reproduction number. Such as Read et al using deterministic SEIR model based on confirmed cases of 2019-nCoV in Chinese cities estimated R0 as 3.8 (95% CI, 3.6-4) which is higher than our result.⁶ Tang et al estimated the mean control reproduction number as 6.47 (95% CI, 5.71-7.23)⁷ which is much higher than our estimated reproduction number by any of the method. Liu et al estimated R0 for EG

(exponential growth) and ML (maximum likelihood) techniques was 2.90 (95 per cent CI: 2.32~3.63) and 2.92 (95 per cent CI: 2.28~3.67) respectively.⁸ Wu et al. used SEIR-based metapopulation version to simulate epidemics in predominant cities, and used Markov chain Monte Carlo technique estimates R_0 is 2.68 (95 per cent CI: 2.47 ~ 2.86).⁹ Zhao et al estimated the mean R_0 of 2019-nCoV 2.24 (95%CI: 1.96-2.55).¹⁸ Result of those studies are very much similar to our estimated value of R_0 .

As estimated value of R_0 is greater than unity so preventive measures, such as the progressive tightening of international travel, the issue of public notices, the maintenance of quarantine facilities and various social distancing measures, have been taken by the Government of India to suppress the spread of disease. In most cases, the R_0 was greater than that of unity. Preventive measures, such as the progressive tightening of international travel, the issue of public notices, the maintenance of quarantine facilities and various social distancing measures, have been taken by the Government of India to suppress the spread of disease. However, as far as our results are concerned, R_0 is still above 1 so more holistic effort, including extensive testing, contact tracing and isolation of COVID positive, with emphasis on health education and awareness-raising activities of the general public, are essential to reduce R_0 to below zero.

Moreover, we performed a simulation using R software for several R_0 and obtained their estimates based Indian COVID-19 data for the five methods. The computed R_0 in the TD method is closer to the actual R_0 and have a good fit on data as confirmed with MSE criterion. It has the minimum MSE than other considered methods. For the most actual R_0 , the simulated R_0 by the AR method is equal to one. Whereas these type of modelling approaches are not able to differentiate between various R_0 . We believe that this may correspond to the small number of the infected cases compared to the susceptible cases. Finally, a more comprehensive study for COVID-19 using new method such as Bayesian is needed that we are going to do in the future research.

Limitations

Our model necessarily makes a number of assumptions. Our estimates of the reproductive number of this novel coronavirus are suitable for specific time period and for data analyzed here. This measure may change substantially over the course of this outbreak and as additional data arrives. Additionally, the models do not include the effect of recent events like lockdown and quarantine etc.

CONCLUSION

Awareness of the basic reproduction number of Covid-19 is useful for controlling disease spread and planning. It is therefore necessary to know the best method that has

better performance. Our study shows that the TD method should be preferred in the calculation of the basic reproduction number. One advantage of the TD method compared to the other methods is that it is useful for computing the R_0 with respect to the actual cumulative case count data and does not require a lot of parameters to calculate the basic reproduction number. We therefore recommend that this method be used to estimate the basic reproduction number.

Funding: No funding sources

Conflict of interest: None declared

Ethical approval: Not required

REFERENCES

1. Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J Pediatr*. 2020(87):281–6
2. Tua H, Tua S, Gao S, Shao A, Shenga J. Epidemiological parameter estimation of early infectious diseases of new coronavirus pneumonia. *Chinese J Epidemiol*. 2020 (81):1-9
3. World Health Organization. Coronavirus disease 2019 (COVID-19) situation report–85. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200414-sitrep-85-covid-19.pdf?sfvrsn=7b8629bb_4. Accessed on 16 April 2020.
4. Home Ministry of Health and Family Welfare GOI. Available at: www.mohfw.gov.in. Accessed on 15 April 2020.
5. Li J, Blakeley D. The failure of R_0 . *Computational and Mathematical Methods in Medicine*. 2011.
6. Read JM, Bridgen JRE, Cummings DAT, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions [J]. *Med Rxiv*. 2020.
7. Tang B, Wang X, Li Q, Bragazzi N L, Tang S, Xiao Y et al. Estimation of the transmission risk of 2019-nCoV and its implication for public health interventions. *J Clin Med*. 2020;9:462
8. Liu T, Hu JX, Xiao JP, He G, Knag M, Rong Z et al. Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China. *Bio Rxiv*, 2020.
9. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, 2020;395(10225):689-97
10. Chowell G, Bettencourt LM, Johnson N, Alonso WJ, Viboud C. The 1918-1919 influenza pandemic in England and Wales: Spatial patterns in transmissibility and mortality impact. *Proc Biol Sci*. 2008; 275:501-9.
11. Chowell G, Diaz-Dueñas P, Miller JC, Alcazar-Velazco A, Hyman JM, Fenimore PW, et al. Estimation of the reproduction number of dengue fever from spatial epidemic data. *Math Biosci* 2007;208:571-89.

12. Chowell G, Nishiura H, Bettencourt LM. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J R Soc Interface*. 2007;4:155-66.
13. Chowell G, Simonsen L, Towers S, Miller MA, Viboud C. Transmission potential of influenza A/H7N9, February to May 2013, China. *Bio Med Cent Med*. 2013;11:214.
14. Chowell G, Ammon CE, Hengartner NW, Hyman JM. Estimating the reproduction number from the initial phase of the Spanish flu pandemic waves in Geneva, Switzerland. *Math Biosci Eng*. 2007;4:457-70.
15. Chowell G, Towers S, Viboud C, Fuentes R, Sotomayor V, Simonsen L, et al. The influence of climatic conditions on the transmission dynamics of the 2009 A/H1N1 influenza pandemic in Chile. *BMC Infect Dis*. 2012;12:298.
16. Gumel AB, Ruan S, Day T, Watmough J, Brauer F, Van den Driessche P, et al. Modelling strategies for controlling SARS outbreaks. *Proceedings of the Royal Society of London. Series B: Biologic Sci*. 2004;271(1554):2223-32.
17. Majumder, M.S.; Rivers, C.; Lofgren, E.; Fisman, D. Estimation of MERS-Coronavirus reproductive number and case fatality rate for the Spring 2014 Saudi Arabia outbreak: Insights from publicly available data. *PLoS Curr*. 2014;18:6.
18. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International journal of infectious diseases*. 2020;92:214-7.

Cite this article as: Dwivedi N, Gupta S, Dwivedi A. Estimation of R0 for COVID-19 in India through different mathematical model and their comparison. *Int J Community Med Public Health* 2021;8:660-5.