

Research Article

Epidemic Landscape and Forecasting of SARS-CoV-2 in India

Aravind Lathika Rajendrakumar^{1,*†}, Anand Thakarakkattil Narayanan Nair^{1,†}, Charvi Nangia¹, Prabal Kumar Chourasia², Mehul Kumar Chourasia¹, Mohammed Ghouse Syed¹, Anu Sasidharan Nair³, Arun B. Nair³, Muhammed Shaffi Fazaludeen Koya⁴

¹School of Medicine, University of Dundee, UK

²Hospitalist, Mary Washington Hospital, Fredericksburg, VA, USA

³Health System Research India Initiative, Kerala, India

⁴School of Public Health, Boston University, MA, USA

ARTICLE INFO

Article History

Received 24 May 2020

Accepted 15 August 2020

Keywords

COVID-19

India

R0

Rt

ABSTRACT

Background: India was one of the countries to institute strict measures for Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) control in the early phase. Since, then, the epidemic growth trajectory was slow before registering an explosion of cases due to local cluster transmissions.

Methods: We estimated the growth rate and doubling time of SARS-CoV-2 for India and high burden states using crowdsourced time series data. Further, we also estimated the Basic Reproductive Number (R0) and Time-dependent Reproductive number (Rt) using serial intervals from the data. We compared the R0 estimated from five different methods and R0 from SB was further used in the analysis. We modified standard Susceptible-Infectious-Recovered (SIR) models to SIR/Death (SIRD) model to accommodate deaths using R0 with the sequential Bayesian method for simulation in SIRD models.

Results: On average, 2.8 individuals were infected by an index case. The mean serial interval was 3.9 days. The R0 estimated from different methods ranged from 1.43 to 1.85. The mean time to recovery was 14 ± 5.3 days. The daily epidemic growth rate of India was 0.16 [95% CI; 0.14, 0.17] with a doubling time of 4.30 days [95% CI; 3.96, 4.70]. From the SIRD model, it can be deduced that the peak of SARS-CoV-2 in India will be around mid-July to early August 2020 with around 12.5% of the population likely to be infected at the peak time.

Conclusion: The pattern of spread of SARS-CoV-2 in India is suggestive of community transmission. There is a need to increase funds for infectious disease research and epidemiologic studies. All the current gains may be reversed if air travel and social mixing resume rapidly. For the time being, these must be resumed only in a phased manner and should be back to normal levels only after we are prepared to deal with the disease with efficient tools like vaccines or medicine.

KEY POINTS

Question: What are the estimates of infectious disease parameters of early phase of novel SARS-CoV-2 epidemic in India?

Findings: Incidence pattern SARS-CoV-2 shows possible evidence of community transmission. However, the estimated Basic Reproductive Number (R0) is relatively lower than those observed in high burden regions (range 1.43–1.85). Our simulation using susceptible-infectious-recovered/death model shows that peak of SARS-CoV-2 in India is farther than currently projected and is likely to affect around 12.5% of population.

Meaning: The lower estimated R0 is indicative of the effectiveness of early social distancing measures and lockdown. Premature relaxation of the current control measures may result in large numbers of cases in India.

© 2020 The Authors. Published by Atlantis Press International B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

SARS-CoV-2 originated in the Hubei province of China and quickly spread to several countries including Japan and South Korea [1]. The source of the infection was later traced back to the wet market in Wuhan, capital of Hubei [2]. A phylogenetic analysis

of virus in Italy confirmed likely import of disease from China [3]. Rapid traveling modes such as air transport further accelerated the epidemic by introducing new cases into more countries [4]. As of April 13, 2020, there were 1,922,891 reported cases of SARS-CoV-2 globally along with a huge death toll almost touching 120,000 [5]. There is increasing evidence that the virus may have jumped organisms to reach humans through pangolins or similar animals acting as intermediates [6].

Emerging and re-emerging infections of this scale disrupt health system functioning and cause massive losses to the economy.

*Corresponding author. Email: aravindlr13@gmail.com

†Contributed equally.

Data availability statement: The data that support the findings of this study are openly available in COVID-19 Tracker, India at Available from: <https://www.covid19india.org/>.

Mathematical modeling has been used widely to understand the spread of disease in populations [7]. One of the aims of the modeling is to estimate parameters that are critical to the spread of diseases such as basic reproductive number or Reproductive Number (R0) and incubation period [8]. This is quite useful in the scenario of the early spread of disease to plan future strategies.

India was one of the countries to institute strict measures for SARS-CoV-2 control in the early phase [9]. The first case of SARS-CoV-2 in India was reported on January 30 2020 from a student airlifted from Wuhan and at the same time another 800 suspected individuals kept under observation [10]. Since then, the epidemic growth trajectory was slow before registering an explosion of cases due to local cluster transmissions. In this paper, we describe the current trend of SARS-CoV-2 transmission in India and estimate basic parameters such as basic reproductive number R0 and time reproductive number Rt, doubling time and future trend for India from the real-world datasets. Our findings will help to understand the effectiveness of government response and recalibrate suppression strategies for the epidemic.

2. METHODOLOGY

We used crowdsourced time series data available from the internet to estimate country-specific parameters for the epidemic [11]. As of now, this is the best available database for information on SARS-CoV-2 in India in the public domain. The most reliable information in this database is regarding the daily reported incidence of COVID cases. Also, it contains aggregated information on total confirmed cases, total death, and total recovery. We did not use variables with limited information. The data was scraped on to R software on 12th April 2020, cleaned and reshaped for the current analysis [12]. Package ggplot2 was mainly used to create figures along with features from base R [13]. We defined serial interval as the time difference in diagnosis of SARS-CoV-2 in infectee and infector [14]. We assumed serial interval and generation time to be same. The growth rate was computed from this incidence curve by fitting a log-linear model [15] and the basic reproduction number (R0) was obtained from the previously calculated serial interval. We also calculated the time-dependent reproductive number (Rt) to show the change in infectivity over time. State-wise growth was measured using the slope from a linear model. To estimate an R0 for projection purposes, in addition, to Log-Linear (LL), we also estimated R0 from other methods such as Exponential Growth (EG), Maximum likelihood estimation (ML), and Sequential Bayesian (SB) Method and Estimation of time-dependent reproduction numbers (TD) [16]. We selected Sequential Bayesian (SB) Method to account for stochasticity in incidence curves as the priors change with time and are drawn from posterior distribution from time-dependent informative priors [16].

Conceptually, susceptible individuals become infectious (i.e., move from the susceptible compartment to the infectious compartment), and then ultimately recover from the infection (i.e., move from the infectious compartment to the recovered compartment). The rates at which they move from one compartment to another depending on the proportion of the population in each of these compartments, as well as the transmission and recovery rates associated with the disease. We modified standard SIR models to accommodate an additional compartment for deaths by assuming transition probabilities, the Susceptible-Infectious-Recovered/Death (SIRD) model. Our model can be represented by the following ordinary differential equation

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - (1 - \alpha)\gamma I - \alpha \rho I$$

$$\frac{dR}{dt} = (1 - \alpha)\gamma I$$

$$\frac{dD}{dt} = \alpha \rho I$$

where β is the disease transmission rate, γ is the recovery rate, α is the case fatality rate and ρ rate at which death occurs. β is the product of the contact rate and transmission probability. As these parameters could not be estimated directly from the data, we back-calculated β from R0 which was computed previously using the Sequential Bayesian method. γ was calculated from the infectious period – (1/infectious period). Case fatality rate (α) can be defined as deaths out of confirmed cases. ρ was estimated as inverse of time to death. We did not have ρ from our data and hence used an estimated range of 2–8 weeks from the WHO report [17]. We used 5 weeks as the value of ρ for our study. The values of the parameters are as follows: R0 = 1.85, $b = 0.13$, $\gamma = 0.07$, $\alpha = 0.035$ and $\rho = 0.028$.

2.1. Ethics Approval

We used anonymized data available in the public domain for this analysis.

3. RESULTS

The first case of SARS-CoV-2 was reported in Kerala state. Since then as of 12 April 2020, there are 9212 confirmed cases, 248 deaths, and 1226 recovered in India from SARS-CoV-2. Maharashtra reports the highest number of cases (1982) and northeastern states have the lowest reported cases. The trend of the epidemic is shown in Figure 1. The daily growth rate of India fitted from a log incidence overtime is 0.16 [95% CI; 0.14, 0.17]. The doubling time for the epidemic is 4.30 days [95% CI; 3.96, 4.70]. The growth rate and doubling time for the five states with the highest-burden are given in Table 1 and the rates for the states with available data are given in Figure S3. Kerala, which had the highest number of cases have effectively contained the outbreak by prolonging the doubling time, 19.48 days [95% CI; 10.35, 164.60] and the present growth rate is 0.03 [95% CI; 0.004, 0.066].

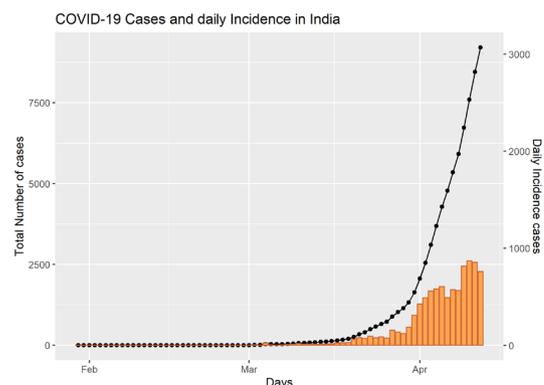


Figure 1 | Daily and cumulative incidence of SARS-CoV-2 cases in India (as on 12 April 2020).

The nature of the epidemic is changing as the transmission is now more prominent within local communities as opposed to imported cases from individuals with travel history from China, the Middle East, and Europe. The type of transmission for most cases was not available. However, those individuals whose source of infection is yet to be determined can be assumed to be of local origin as travel history of SARS-CoV-2 patients is well tracked and documented in India (Figure S2).

There were 145 index cases who were identified to have transmitted the disease. On average, 2.8 individuals were infected by an index case and the range varied from 1 to 20 individuals. The number of secondary cases arising from exposure to primary cases (colored bubbles) and index cases are given with their unique identifiers (x-axis) (Figure 2). We had information on 413 pairs of primary and secondary cases to generate serial intervals for computing Basic Reproductive Number or R0. Serial interval ranged from 0 to 19 days as the primary and secondary cases reported to the health facility on the same date. The mean of the serial interval was 3.9 days with a standard deviation of 2.85 computed from a pair of 324 pairs with non-zero values.

The lowest estimated R0 was from log-linear model 1.47 [95% CI; 1.43, 1.51] and highest from time-dependent model 1.89 [95% CI; 1.64–2.15] (Figure 3). The time-varying reproductive number (Rt) and SI distribution curve across the analysis time are provided in Figure S1. The recovery time was available only for 133 individuals. The mean time of recovery was 14 days (SD 5.3) ranging from 5 to 25 days.

From the compartmental models, it can be deduced that the peak of SARS-CoV-2 in India will be around mid-July to early August 2020. Around 12.5% of the susceptible individuals are likely to be infected during the peak. The R0 estimated from different methods ranged from 1.43 in the log-linear method to 1.85 in exponential growth methods. Among the five models that were used for computing R0: exponential, Time-dependent, and Sequential Bayesian

Table 1 | Growth rate and doubling curve from log-linear fit for five states with highest burden of SARS-CoV-2 in India

State	Growth rate (95% CI)	Doubling time in days (95% CI)
Tamil Nadu	0.173 (0.13, 0.20)	4.00 (3.32, 5.03)
Delhi	0.145 (0.12, 0.17)	4.78 (4.05, 5.83)
Maharashtra	0.144 (0.12, 0.16)	4.81 (4.23, 5.58)
Rajasthan	0.131 (0.11, 0.15)	5.27 (4.61, 6.17)
Gujarat	0.119 (0.07, 0.17)	5.77 (4.13, 9.59)

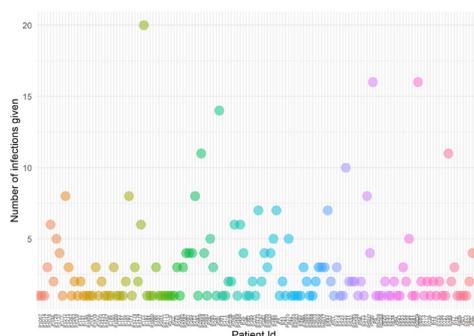


Figure 2 | Number of secondary cases resulting from contact with each index case.

methods provided similar estimates ($R_0 > 1.80$). We chose the Sequential Bayesian method (SB) over others for modeling as it was similar to the time-dependent R0 estimate and due to its flexibility to accommodate changes in transmission dynamics by incorporating prior distributions (Figure 4).

4. DISCUSSION

We estimated several parameters of SARS-CoV-2 infection in India, from crowdsourced data, and used a minimal set of assumptions in our forecasts. Models are useful in identifying the transmission dynamics if parameter inputs are based on real-world data in the early phase of the epidemic. Social distancing and travel ban to reduce community transmission seems to have an impact on infection transmission.

There is an urgent need to estimate setting specific parameters for future planning. For instance, we did not have information on death rate and recovery rate and used estimates from published studies in the SIRD model. We also could not compute the incubation period as the database contained information only on linked cases and their admissions. Studies have shown that this could be approximately 5–8 days [18,19]. Also, it is difficult to separate the serial interval time from generation time overlaps as

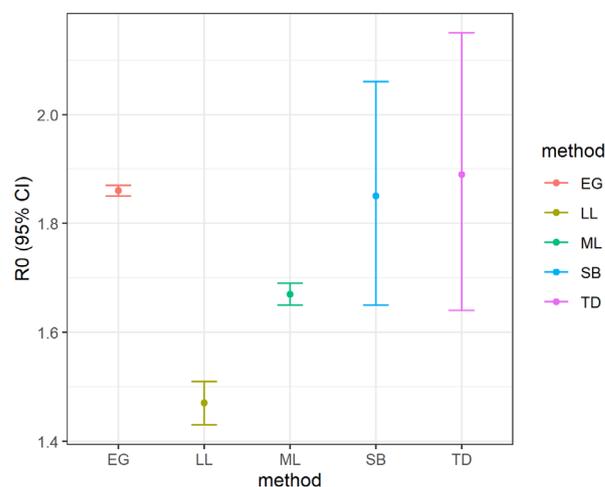


Figure 3 | Basic reproductive number (R0) calculated from different estimation methods.

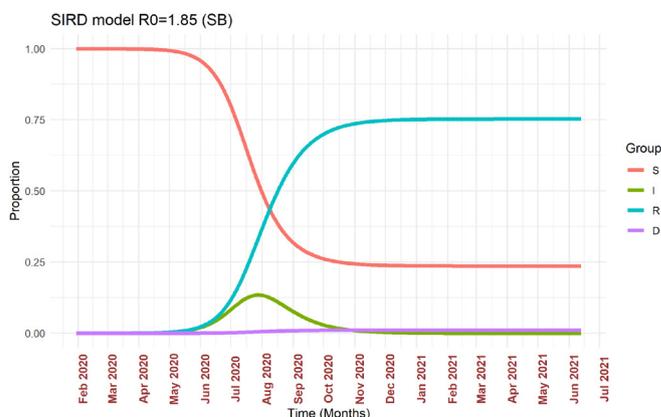


Figure 4 | SIRD models based on R0 derived from Sequential Bayesian (SB) method.

the information of cases is only available from reporting to the health authorities. Studies have suggested that SARS-CoV-2 can spread before the symptomatic phase which precludes a full understanding of the transmission pattern and underlying characteristics.

The R_0 estimated across many studies may not reflect the actual transmission potential of the virus as these were measured in dynamic cohorts with some sort of mitigation measures. Estimates from studies among closed populations such as in Diamond Princess and Wuhan may be more reliable in this regard [20]. Riou et al. [21] estimated a reliable range for R_0 which was between 1.4 and 3.8 by simulating Wuhan epidemic incidence trajectories in a cluster environment. We estimated the range of R_0 in India to be between 1.42 and 1.84 by comparing estimates from five different methods. Each of the R_0 estimation methods employed in this analysis has its distinct advantages. The Time-dependent (TD) and Exponential growth methods are less prone to bias and over dispersion respectively. Furthermore, Sequential Bayesian uses prior information to update the estimates. Hence considering the similarity of R_0 from all the above methods and the advantages provided by Bayesian updating of estimates from SB we decided to use it for SIRD modelling.

It can be therefore assumed that the real R_0 in India is close to 1.85. The upper limit of the uncertainty of R_0 in our analysis exceeds 2 which may be indicative of the spread of infection at the local level and is comparable to the European scenario [22]. This is much lower than R_0 in most countries and may be indicative of the effectiveness of measures taken at policy levels such as lockdown, cancellation of flights, social distancing, and sealing of identified hotspots.

India has done well to contain population-level spread to a great extent and to keep low mortality rates. India has one of the lowest attack rates of 0.47 at this stage of the epidemic which is several times lower than developed countries [23]. However, the epidemic is on the rise in India and creates new challenges such as optimal use of resources including swabs and testing kits besides identifying the high-risk groups. A recent study by Prem et al. [24] suggests that mitigation measures may be helpful only in delaying the infection from reaching its peak. Thus, it implies that the measures would likely prevent immediate flooding of hospitals preventing people from obtaining adequate care. The epidemic is yet to reach its peak in India which we expect to be around mid-July and early August. Our models suggest that 12.5% of the susceptible individuals are likely to be infected during the peak. This means that there are still around 2–3 months to prepare for the worst phase of the epidemic.

Control measures of SARS-CoV-2 are less effective in populations with high density [20]. India has a huge population and the potential of SARS-CoV-2 for sustained transmission is well established [21,25]. We relied solely on estimates from the compartmental model which generates estimates from the infection parameters than depending on forecasts based on-trend. Our findings are based on simulations and further measures at policy level can alter the trajectory of the disease. Lipsitch et al. [26] recommends a strategy combining epidemiological studies, testing and lab studies as a more effective strategy in generating evidence for suppressing the epidemic.

Our analysis has certain limitations. First, there may be underreporting of cases and a lack of validated data for research purposes available to the public. However, we compared the incidence from our

dataset to the aggregated numbers elsewhere like European Disease Prevention and Control [23] and Worldometer [5] and the figures were similar. Another issue in estimating the parameters is the presence of super spreaders and asymptomatic cases. The estimated proportion of asymptomatic cases can be as high as 10% of the population [27]. The virus can be spread around during the asymptomatic stage and delays from onset to reporting or treatment can be as high as 11 days [28]. Nonetheless, pooling estimates across several studies will aid in computing values that are closer to the true estimates.

In conclusion, our study provides information on India specific estimates of SARS-CoV-2 transmission parameters using real-world data for the first time and shows that measures taken to date have been effective in reducing the spread of disease. However, the rising incidence and pattern of spread is suggestive of community transmission and is likely to increase cases in the future. The availability of individual-level data is critical to assess the effectiveness of ongoing measures and plan future strategies. There is a need to increasingly fund infectious disease research and epidemiologic studies and make that data available in the public domain. Future studies could focus on studying genetic variations in asymptomatic individuals and those index cases with higher disease transmission rates. There needs to be a renewed thinking of health systems particularly regarding emergency preparedness and optimal utilization of scarce resources. Epidemiologists need to be considered in decision making and developing disease control strategies. Furthermore, there needs to be concerted action at the global level to contain the virus including joint research and developing vaccines and drugs. All the current gains may be reversed rapidly if air travel and social mixing resume rapidly. For the time being, these must be resumed only in a phased manner and should be back to normal levels only after we are prepared to deal with the disease with efficient tools like vaccines or medicine.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

AUTHORS' CONTRIBUTION

ABN, ALR, ATN and MSFK conceptualized this study. CN, PKC, MKC and SMG provided technical assistance and critical inputs to the study. ASN, ABN and MSFK supervised the project. ATN, ALR and ABN conducted statistical analysis and prepared the manuscript. All authors provided revisions and approved the manuscript for submission.

FUNDING

No financial support was provided.

ACKNOWLEDGMENTS

We acknowledge Dr. Mike Lonergan, University of Dundee, UK for reviewing and providing critical inputs for the manuscripts. We thank the team managing COVID19 open data source which was used for this analysis.

SUPPLEMENTARY MATERIALS

Supplementary data related to this article can be found at <https://doi.org/10.2991/jegh.k.200823.001>.

REFERENCES

- [1] Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A novel coronavirus emerging in China — key questions for impact assessment. *N Engl J Med* 2020;382;692–4.
- [2] Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. *J Adv Res* 2020;24;91–8.
- [3] Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M. The first two cases of 2019-nCoV in Italy: Where they come from? *J Med Virol* 2020;92;518–21.
- [4] Arima Y, Shimada T, Suzuki M, Suzuki T, Kobayashi Y, Tsuchihashi Y, et al. Severe acute respiratory syndrome coronavirus 2 infection among returnees to Japan from Wuhan, China, 2020. *Emerg Infect Dis* 2020;26;1596–600.
- [5] Worldometer. Coronavirus cases [Internet]. Worldometer. 2020, pp. 1–22. Available from: <https://www.worldometers.info/coronavirus/coronavirus-cases/#daily-cases> [cited April 10, 2020].
- [6] Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein structure and sequence reanalysis of 2019-nCoV genome refutes snakes as its intermediate host and the unique similarity between its spike protein insertions and HIV-1. *J Proteome Res* 2020;19;1351–60.
- [7] Siettos CI, Russo L. Mathematical modeling of infectious disease dynamics. *Virulence* 2013;4;295–306.
- [8] Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 2020;20;553–8.
- [9] Coronavirus Disease (COVID-19). Available from: [https://www.who.int/india/emergencies/coronavirus-disease-\(covid-19\)](https://www.who.int/india/emergencies/coronavirus-disease-(covid-19)).
- [10] Coronavirus: India confirms first case, with 800 others under observation [Internet]. Available from: <https://www.telegraph.co.uk/global-health/science-and-disease/coronavirus-india-confirms-first-case-800-others-observation/> [cited April 10, 2020].
- [11] COVID-19 Tracker. India. Available from: <https://www.covid-19india.org/>.
- [12] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
- [13] Wickham H. ggplot2: elegant graphics for data analysis using the grammar of graphics. Springer-Verlag: New York; 2016. Available from: <https://ggplot2.tidyverse.org>.
- [14] Moghadas S, Milwid R. Glossary of terms for infectious disease modelling: a proposal for consistent language. Winnipeg, MB: NCCID; 2016, pp. 1–3.
- [15] Jombart T, Cori A, Kamvar ZN, Schumaker D. Small helpers and tricks for epidemics analysis. Package ‘epitrix.’ 2019. Available from: <https://cran.r-project.org/web/packages/epitrix/epitrix.pdf>.
- [16] Obadia T, Haneef R, Boëlle PY. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Med Inform Decis Mak* 2012;12;147.
- [17] WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). 2020. Available from: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf> (accessed April 13, 2020).
- [18] Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Euro Surveill* 2020;25;2000062.
- [19] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 2020;172;577–82.
- [20] Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. *J Travel Med* 2020;27;taaa038.
- [21] Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill* 2020;25;2000058.
- [22] Yuan J, Li M, Lv G, Lu ZK. Monitoring transmissibility and mortality of COVID-19 in Europe. *Int J Infect Dis* 2020;95;311–15.
- [23] European Centre for Disease Prevention and Control. COVID-19. Available from: <https://qap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html> [cited April 12, 2020].
- [24] Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* 2020;5:e261–e70.
- [25] Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res* 2020;5;67.
- [26] Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19 — studies needed. *N Engl J Med* 2020;382;1194–6.
- [27] Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health* 2020;8:e488–e96.
- [28] 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;1199–207.