



Data analysis and mathematical model: control measures and prediction to prevent COVID-19 outbreak

Md. Kamrujjaman^{★1,2}, J. Jubyea³, Md Shafiqul Islam⁴

¹Department of Mathematics, University of Dhaka, Dhaka 1000, Bangladesh

²Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada

³Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh

⁴School of Mathematical and Computational Sciences, University of Prince Edward Island, Charlottetown, PE, Canada

Article Info

Keywords:

Coronavirus;
Wuhan;
COVID-19;
Data analysis;
Mathematical model

Received Apr 4, 2020

Revised May 1, 2020

Published May 13, 2020

*Corresponding author:

kamrujjaman@du.ac.bd

Abstract

Background. The recent COVID-19 outbreak is now an ongoing global health emergency. The world is dealing with the epidemic with no vaccine or medication to treat the disease, the only effectual measure for now is to implement quarantine methods. We designed a quarantine mathematical model with data analysis to predict the outcome of this pandemic.

Methods. We collected available online data of four different countries China, Italy, Spain and USA. First, we have analyzed the real-life data and abridged data. Then, fitting analysis of the data was done in comparison with the outcome of our mathematical results.

Results. It is found that disease progression in this model is determined by the basic reproductive ratio, R_0 . If $R_0 > 1$, the number of latently infected individuals grows exponentially; (in a case with enough public mobility). If $R_0 < 1$ then the infection rate decays exponentially i.e. government ensures the social isolation through quarantine. Data analysis of different countries show that the possible dynamics are growth, growth-decay and growth-decay-growth dynamics. After imposition of a quarantine on March 9, 2020 in Italy, within 13 days of lock-down, the maximum number of infection was observed after 42 days (from Feb 15, 2020) before decreasing. The quarantine model approximates that the disease in Italy could be under control by mid-May. Similar results present for Spain (growth-decay) and USA (growth only).

Conclusion. The R_0 of COVID-19 may vary from country to country. To control the pandemic, R_0 and incubation period play an important role in spreading and controlling of the disease.

1. Introduction

The Novel Coronavirus outbreak is now an ongoing epidemic affecting 204 countries and territories, with 1,018,920 ongoing confirmed cases and 53,292 deaths worldwide as on April 03, 2020 [1]. World Health Organization (WHO) has declared this outbreak a global emergency, as health care systems seem to collapse in almost every country [2]. In this study we will discuss the 2020 coronavirus pandemic of Italy, Spain, USA and China, simulate some control measures, and project estimated control time using a mathematical model.

The Coronavirus disease 2019 (COVID-19) was first reported in the last week of December 2019 in Wuhan, the capital city of Hubei province [2]. According to WHO coronaviruses are zoonotic, which means that they can be transmitted from animals to people [3]. Although the origin of COVID-19 is still unclear, but it is believed that the SARS-CoV-2 originated from a seafood market where wild animals are traded [4,5]. Researchers established that the interaction between the receptor-binding domain (RBD) of the coronavirus spike protein and the host receptor angiotensin-converting enzyme 2 (ACE2) influence disease transmission, resulting in Severe Acute Respiratory Syndrome (SARS) and COVID-19 [6]. The spike protein of COVID-19 attaches to ACE2 receptors on host cells, infects them, replicates, bursts open the cell, and the replicates infect new cells. Scientists modeled ACE2 receptor proteins with several species to see which ones are vulnerable to SARS-CoV-2 infection [7]. It was observed that SARS-CoV-2 have structural similarity of ACE2

receptors with pigs, ferrets, cats, orangutans, monkeys, some species of bats, and humans [8,9]. The current hypothesis is that outbreak started in bats, then moved to another unidentified species before being transmitted to humans [10,11].

As the world is dealing with a highly contagious disease without medication, the only effectual way for us to protect populations so far is prevention. Social and physical distancing, lock-down and testing are some measures prescribed by the World Health Organization to control the outbreak. Data analysis and mathematical modelling are some of the core components to study these undertaken polices for the best outcome. In mathematical modelling, some recent studies provided different guidelines by introducing basic reproduction number, education and socio-economic index and lock-down strategies [12-16].

This manuscript presents an analysis of different control mechanisms using mathematical models based on the ongoing viral epidemic of several countries. The objectives of this study are as follows:

- To work with real-life data to understand the cases and project the control of the infection.
- To make a prediction of controlling measure of COVID-19.
- To validate the dynamics of mathematical model to get better accuracy of results comparing with data analysis.

2. Materials and methods

2.1. Data management

In this study we considered four countries China, Italy, Spain and USA drastically infected by SARS-CoV-2 virus and collected available online

data in chronological order from Worldometer [1,2]. We considered China since it was the origin of COVID-19 and the other three countries were the most affected countries at the time of this study. We included the data from January 20 to April 01 (73 days) for China while for the other three countries data was available from February 15 to April 01 (47 days).

To ensure the curve fitting we did not include any additional data outside the time interval as declared above. For simplicity and to compare the results, zero (0) was used to denote the first day when a case was reported for each of four countries; January 20 (China) and February 15 (Italy, Spain and USA).

2.2. Mathematical model

Transmission of infectious diseases may occur through a variety of pathways. They can be transmitted in many ways and classified as person-to-person transmission, air-borne transmission, and food and waterborne transmission etc. Since transmission of COVID-19 is classified as a person-to-person transmitted disease [5, 7, 8], subsequent analysis focused on a person-to-person transmission model.

A mathematical model was used to develop a solution function based on SARS-CoV-2 virus. The function was subsequently used to model the current outbreak to predict future-trends of the outbreak. In the literature, for infectious disease modelling, the classical SIS, SIR, SID, SIRI and SEIR etc. models allow the determination of critical condition of disease development in the population with total population size [12, 15, 17, 18]. Since prevention is the only known solution to curb the spread of COVID-19, a quarantine model was developed based of the following assumptions:

- At the current stage, the main challenge was to reduce the infection and disease transmission, to describe the infected population we used the notation, I .
- Globally it was observed that most countries were locked down; there was no immigration or emigration meaning no flux on the boundary and we assumed that susceptible population was constant and defined by S_0 .
- For disease COVID-19, incubation period was a new factor and we defined it by τ .
- Compared to time t , the incubation period τ was very small and varies from patient to patient and could even vary from country to country.
- Recovered individuals were assumed to be immune and would not infect other individuals in the population.

Under these assumptions, we have designed the following ordinary 3. differential equation with partially delay (DDE) and defined as a quarantine model

$$(DDE) \quad I'(t) = S_0[\beta I(t) - \gamma I(t - \tau)], \quad I = I_0 > 0 \quad (2.1)$$

Here $I(t)$ is the number of individuals in a population in the infected compartment at time t with a day unit and constant susceptible population, S_0 . The $I(t - \tau)$ is the infected individuals undergoing the incubation period, τ . The parameter β denote the infection rate (disease transmission rate) for infected population and γ is the infection rate due to the infected individuals during incubation period only. The total population is defined by $N(t) = S_0 + I(t)$.

This equation resulted in the solution, $I(t) \equiv I_0 e^{mt}$ and after substituting to (2.1) output

$$m = S_0[\beta - \gamma e^{-m\tau}] \quad (2.2)$$

where m was the growth/decay rate of infection. Since the incubation period will end at time $t = \tau$ then $m = 0$ is the solution of (2.2) as

long as both transmission rates were equal, i.e. $\beta = \gamma$ and the solution was independent of parametric values.

Define a function $h(m) = S_0[\beta - \gamma e^{-m\tau}]$ which implies

$$h'(m) = \tau S_0 \gamma e^{-m\tau} = R_0 e^{-m\tau} \quad (2.3)$$

where $R_0 = \tau \gamma S_0$ was the basic reproduction number and plays the important role in the stability of solutions. If $h'(0) > 1$, then the function is increasing, and the solution was positive. The solution is decreasing and negative if $h'(0) < 1$. Finally, the solution of (2.1) was

$$I(t) \equiv I_0 e^{S_0[\beta - \gamma e^{-m\tau}]t} \quad (2.4)$$

2.2.4. Control measure

In this study, the basic reproduction number was defined as $R_0 = \tau \gamma S_0$. Compared to other classical epidemic mathematical models, here the meaning of R_0 was different because it depends on both incubation period and public mobility. During the incubation period, R_0 distinguished the total infection rate since τ can vary patient to patient. It is remarked that to handle basic reproduction number, we had no control on S_0 and τ since they characterized the existing population and the property of disease (incubation period), respectively [16,17]. Therefore, we had to control (reduce the value of γ) the disease transmission rate γ by the infected individuals; among which the symptoms were not visible yet.

Reality shows that the parameter γ was depended on the physical distance (not only social isolation as recent studies had reported) and the human mobility and hence it was possible to control. In the early stages of the disease development of COVID-19:

- If $R_0 > 1$ then each infected person infected on average more than one susceptible individual during the incubation time and the number of newly infected people would increase exponentially.
- While $\gamma < \gamma^*$, where γ^* =critical value that measures restricting potential contacts of infected and susceptible individuals, then forthwith infected individuals would decay exponentially.

Theoretically, we can summarize the results in the following theorem.

Theorem 1.

- When $R_0 < 1$, the disease-free solution of (2.1) is globally asymptotically stable;
- When $R_0 > 1$, the disease solution of (2.1) is globally asymptotically stable.

3. Results

3.1. Primary study based on discrete data

We summarize the officially reported data from Worldometer in Figure 1(a). There is an increasing-decreasing-increasing trend of daily new confirmations. Cumulative cases present in Figure 1(b) which is either increasing or flattened. Deaths were summarized in Figure 2 for all countries both daily and total cases. In Figure 1, considered the first reported case for all the countries as initial starting point 0 (day 1) to calculate frequency and time period. In the early days, China observed increasing daily reported cases, Figure 1(a), while Italy, Spain and USA for first two to three weeks seemed to have slower starting rate although USA experienced sharp increase in daily reported case than other countries. The trend of reported cases in Italy and Spain reflected that of China with a higher number of cases (January 20 was consider as 0 for China while for Italy and Spain it was February 15). Cumulative cases, Figure 1(b), of Italy, Spain and USA are exponentially increasing and China's cumulative cases seems to be stabilized.

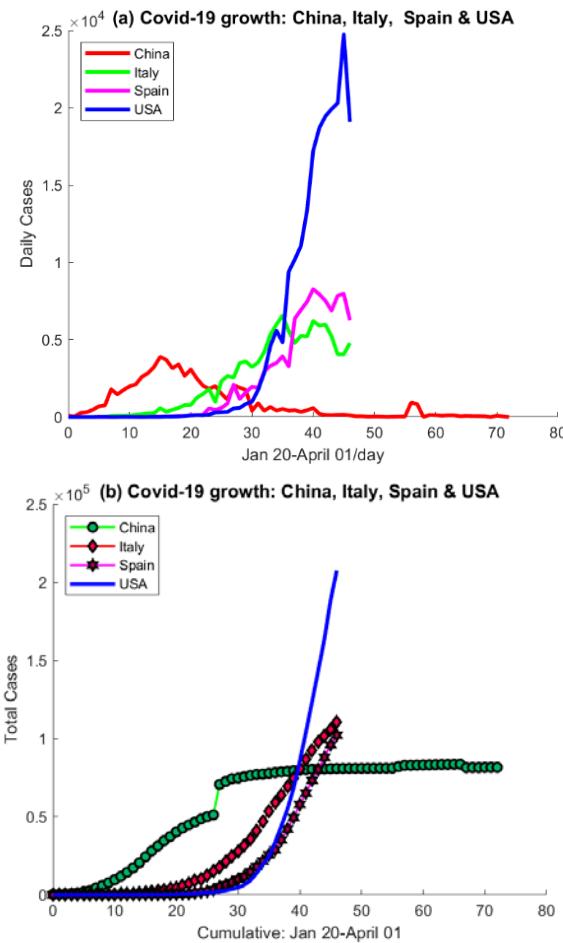


Figure 1: Discrete data curves of infections for (a) daily cases and (b) cumulative cases.

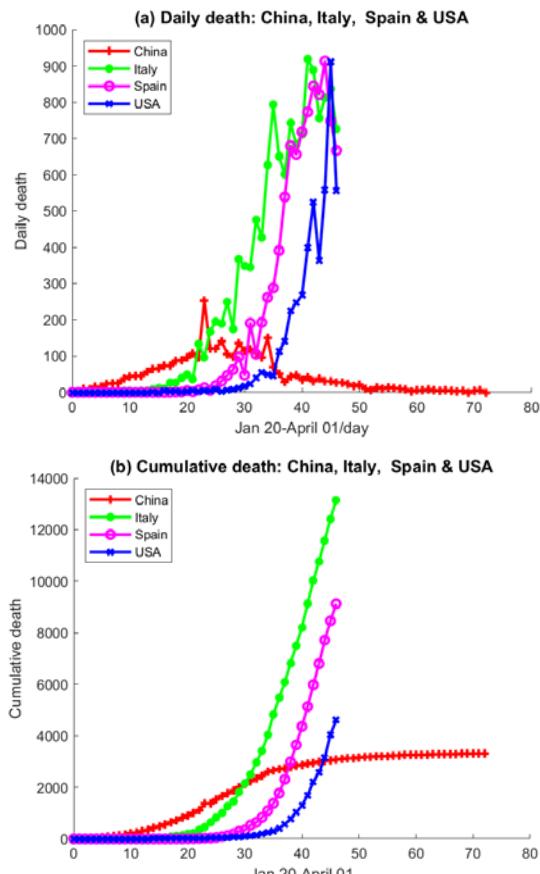


Figure 2: Discrete data curves of infections for (a) daily cases and (b) cumulative cases.

Figure 2 referred to daily death and cumulative death for respective countries. In both cases USA was showing the worst-case scenario with 1049 death on April 1, 2020 [1, 2], Figure 2(a). During the time interval February 15 to April 01, 2020, USA death was maximum on April 01 compared to Italy and Spain. It is important to note that daily cases of death in USA was continuously rising with small fluctuation so the cumulative death might follow Italy's pattern. Although Spain appeared to be experiencing some fluctuation throughout the outbreak cumulative deaths might rise concern as the cumulative death curve was growing exponentially as depicted in Figure 2(b). Visibly it was seen that sequentially the total death in China turns to a straight line, e.g. in control. These observed trends were extrapolated using mathematical models.

3.2. Applications and predictions

To test our mathematical model and for curve fitting with real life data, we considered the initial time from the day where cumulative cases tended to go up.

- Italy, $I_0 = 229$ Infected Individuals at February 24, 2020 [1].
- Spain, $I_0 = 58$ Infected Individuals at February 29, 2020 [1].
- USA, $I_0 = 124$ Infected Individuals at March 3, 2020 [1].

Studies indicate that median incubation period for COVID-19 is 5.1 days (4.5-5.8 days) [19]. In this paper, we considered the incubation period $\tau = 5$; some studies found that incubation time interval can be any time between 2-14 days to 2-27 days [20,21]. It was also noted that, sometimes individuals might not show any symptoms in that case incubation period was not finite such that susceptible individuals are infected by the patients drastically.

Figure 3 illustrates discrete Data map vs DDE growth and DDE decay for both (a) Italy and (b) Spain. For Italy March 21, 2020 (42 days) had the highest reported case 6557 and we considered this as peak of growth. Although there were 6203 and 5975 reported cases of March 26 and 28, respectively with some fluctuation but we considered the overall trend indicate decay (decline in number of cases). In Figure 3(a), DDE growth map coincides with data map with growth rate $m = 0.15$ and DDE decay map seems to follow the downward pattern of data decay map with decay rate $m = -0.13$ (see equations (2.2) and (2.4) to draw the solution curves). To generate the decay curve, we considered the number of initial latent infected population using the approximate sum of everyday cumulative individuals. If there was no sudden change in growth or decay in the ongoing outbreak, then we assumed that Italy might be able to control of the epidemic between second and third week of May 2020. While $\tau = 5$, we found the reproductive ratio as $R_0 = 1.42$ for growth and $R_0 = 0.709$ to decay and therefore theoretical statement in Theorem 1 is valid. A drastic spread of COVID-19 would be expected if patients had no symptoms or even the incubation period was too high; for example, $\tau = 27$ would result in an infection rate $R_0 = 4.12$ indicating that one individual in a population could infect on average 4 individuals [20, 21].

In Figure 3(b), DDE growth map nearly overlaps with Data map with growth rate $m = 0.17$. Until April 1, 2020 we did not noticed any significant downward trend in Spain's daily reported cases, so there was no available data to match our DDE decay map. We illustrated a DDE decay map with $m = -0.13$ (approximately), which suggested that Spain may have gained control of the pandemic by the end of May or first week of June. In the case of Spain, the disease spread with a $R_0 = 1.48$ and if incubation period increased then the infection rate would be higher; for instance, $\tau = 14$ implies $R_0 = 2.62$. For Spain to experience an exponential decay of cases with control mechanisms in place, the basic reproductive number would have to be less than 1 i.e. $R_0 = 0.71$ when $\tau = 5$. The illustrated results validated Theorem 1 in this study.

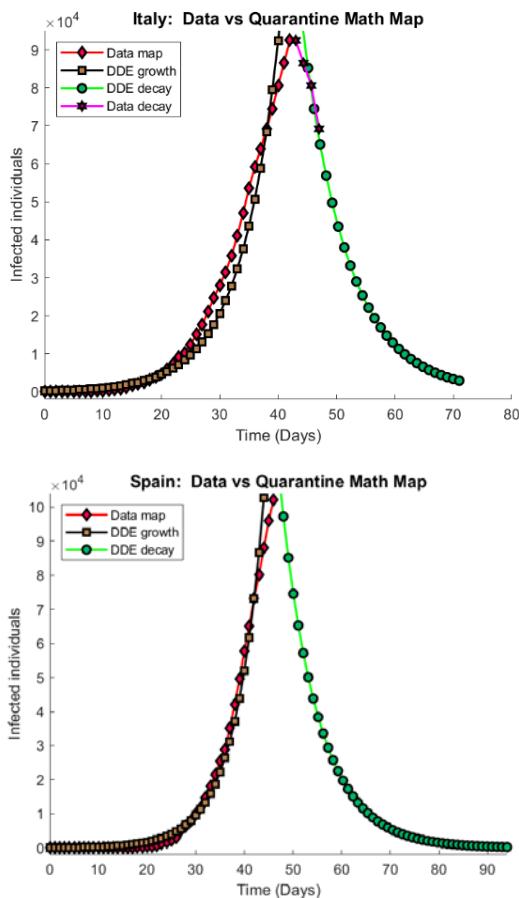


Figure 3: Discrete data curve vs DDE solution of infections for (a) Italy, and (b) Spain.

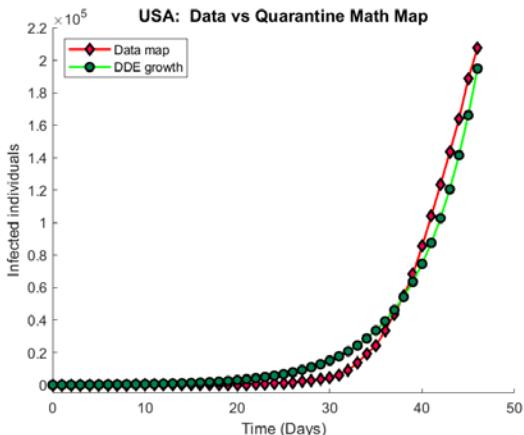


Figure 4: Discrete data curve vs DDE solution of infections for USA.

Figure 4 describes an exponentially growing data map corresponding to the DDE map. We observed that for USA, $m = 0.16$ for our DDE growth curve. As for exponential increases in daily cases in USA (until April 1, 2020), there was a good agreement between two solutions (data and model solutions). At the time of our study, it seemed unrealistic to draw DDE decay map for USA unless the number of daily cases tend to decay asymptotically for a near future period of time. In all models, the R_0 was greater than 1 and still the number of infected individuals increased exponentially, Theorem 1. The possible infection rate varied with the incubation period. For example, we observed the following results

$$(\tau, R_0) \equiv \{(5, 1.45), (9, 1.88), (14, 2.5), (27, 4.37)\}$$

The authors felt that the growth and decay curves are not properly satisfied by the available data from China [1, 2]; this might be due to the data collection methodology or any missing or unreported data. As such, those results were not presented in this manuscript.

4. Discussion

In this study, theoretical results established and presented in terms of basic reproduction number. We have disease free solution for $R_0 < 1$ and endemic solution while $R_0 > 1$. Besides data analysis, the proposed quarantine model showed that human population can control and protect the spread of infectious diseases by creating social isolation, hospitalization, lock-down and physical distancing. Data analysis and model results predicted the time boundary to control the epidemic in Italy and Spain. The model ensured the efficiency of the approach to intercept the disease by the limitation of contacts between the individuals through quarantine of infected individuals.

The current viral epidemic pushed every medical system to its breaking point without any permanent solution to the disease. So far, there has been minimal advancement towards the use of medication or vaccination to control the spread of COVID-19. As such, the most effective mitigation measure currently available is physical distancing to reduce interaction between infected individuals and susceptible populations. The efficiency of quarantine methods determines our protection against the virus.

In this paper, we have studied a basic quarantine model where the dynamics are determined by the threshold levels of R_0 , the basic reproduction number. Both theoretically and numerically with data analysis we have demonstrated that the infected population will be decrease asymptotically as long as R_0 less than 1. The proposed model confirmed the efficiency of the approach to reduce and gradually stop the disease by the limitation of contacts between the individuals through quarantine of infected individuals. In the model results, we compared outbreak data from Italy, Spain and USA with the quarantine model to calculate the growth and decay of the current outbreak. We observed that both growth and decay patterns were exponential; Figures 3 and 4. For USA, there was an exponential growth of infected cases. The growth exponent of USA (growth factor) was higher than Italy's, ($m = 0.16$) and ($m = 0.15$) respectively assuming the same incubation period of the disease $\tau = 5$.

In Italy, lock-down (quarantine) was initiated on 9 March-2020, at that time the total number of reported cases was 9,172. The maximum number of reported cases was observed on 21 March-2020; which means that the number of infections was on an upward trend for only 13 days. After that day, the number of daily reported cases was observed to decrease asymptotically after which the health authorities appeared to gain control of the situation, which is an the most important interpretation of our quarantine model; there was strict limitations to interact population with each other or much gathering. In Spain, a state of alarm (quarantine) was declared on 13 March-2020 and was affected on the following day. Like Italy, Spain also experience an increase in reported cases for 13 days of quarantine enforcement (March 14-26, 2020) and from the 14th day, daily reported cases started to decline albeit with some fluctuation.

In a scenario where there were no symptoms or the incubation period was too high meaning that the infected individual was asymptomatic for longer; for example, $\tau = 27$ the infection rate (R_0) would be 4.12 which means one individual in a population would infect on average 4 susceptible individuals [20, 21]. As such, this infection rate would accelerate through persons in asymptotic incubation period (patients who had no symptoms or symptoms show after a long period) with or without detectable virus or in non-severe symptomatic period with the presence of virus; if the infected individual was not separated from the susceptible population [22, 23].

These values can be used by policy makers and other professionals to make decisions on how certain interventions such as distancing and isolation can be effectively implemented. Although, we only used DDE maps to predict estimated impact of disease control in Italy and Spain,

similar scenarios can be observed in several European countries such as France, Germany and UK.

In this study there were some limitations, for example: the model took into account the beginning of the disease development only and the suspected individuals were considered as constant. We ignored the spatial distribution of population densities and just consider the case when infection rate was changing with time. It will give a good approximation if the disease expansion was controlled when relatively small populations are infected compared to the total population. We had only single incubation period $\tau = 5$, although we made some predictions for different incubation periods. According to the data and methodology, it's possible that the incubation periods varies from zone (country) to zone and patient to patient. Due to these limitations, the accuracy of the existing results and prediction in this paper could vary for a short period of time. Hypothetically if we think there was no limitations on this model (study), it was possible to predict the control measurement of COVID-19 in most of the countries in the world. For this type of pandemic, the standard modelling is the partial delay differential equations (PDDEs) which concludes both time-space distributions and incubation period at a time.

The final conclusion about modelling is that all mathematical models have some limitations and is not perfect for all countries. However, in recent times, COVID-19 reinfection is catching people's attention. The World Health Organization (WHO) is investigating reports of COVID-19 reinfection and the organization is working with clinical experts to get more information on this matter [24]. In South Korea 91 patients who had tested positive, and declared negative after two consecutive tests, tested positive again while 14% of recovered coronavirus patients in China have tested positive again [24, 25]. The reinfection aspect was not considered during the modelling process in this study. However, this provides an opportunity for further research and model simulations.

Acknowledgment

The authors are acknowledged to the anonymous reviewers for their constructive suggestions to improve the quality of the manuscript. The author M. Karmrujjaman research was partially supported by TWAS grant: 2019_19-169 RG/MATHS/AS_I.

References

[1] 'Worldometer', COVID-19 CORONAVIRUS/CASES, 2020.

[2] World Health Organization (WHO), 2020.

[3] Ahmad, T., Khan, M., et al. 2020. COVID-19: Zoonotic aspects. *Travel Medicine and Infectious Disease*, ELSEVIER.

[4] Zhao, S., Musa, S.S., Lin, Q., ... & Wang, M. H., 2020. Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak. *J. Clin. Med.*, 9(2): 388.

[5] Chen, N., Zhou, M., Dong, X., ... & Zhang, L., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, 395(10223): 507–513.

[6] Wan, Y., Shang, J., et al. 2020. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of Virology*, 94(7), e00127-20.

[7] Lu, R., Zhao, X., Li, X., ... & Tan, W., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224): 565–574.

[8] Ji, W., Wang, W., Zhao, X., Zai, J., Li, X., 2020. Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. *Journal of Medical Virology*, 92(4): 1–29.

[9] Wang, M., Jing, H. Q., Xu, H. F., ... & Yan, M. Y., 2005. Surveillance on severe acute respiratory syndrome associated coronavirus in animals at a live animal market of Guangzhou in 2004. *Zhonghua liuxingbingxue zazhi*, 26(2), 84–87.

[10] Wang, M., Yan, M., Xu, H., ... & Wang, H., 2005. SARS-CoV infection in a restaurant from palm civet. *Emerging Infectious Diseases*, 11(12), 1860–1865.

[11] Kamrujjaman, M., Hossain, M. A., 2020. Spreading reason of 2019-nCoV and the message of Almighty God (Allah): A justification for the protection of mankind. *Arabian Journal of Medical Sciences*. 3(1): 8–10.

[12] Murray J. 2002. *Mathematical Biology I*. third edition, Springer-Verlag, Heidelberg. [13] Kermack, W. O., McKendrick, A. G., 1927. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society*. 115(772): 700–722.

[14] Castillo-Chavez, C., Feng, Z., 1996. Mathematical models for the disease dynamics of tuberculosis. *Advances in Mathematical Population Dynamics-Molecules, Cells and Man*. 1–28.

[15] Kamrujjaman, M., Mahmud, M. S. & Islam, S., 2020. Coronavirus outbreak and the mathematical growth map of Covid-19. *Annual Research & Review in Biology*. 35(1), 72-78.

[16] Mahmud, M. S., Kamrujjaman, M., Jubrey, J., Islam, M. S., 2020. Mathematical Modelling of Social Consciousness to Control the Outbreak of COVID-19. preprint, DOI: 10.20944/preprints202004.0196.v1.

[17] Volpert, V., Banerjee, M., Petrovskii, S., 2020. On a Quarantine Model of Coronaviruses Infection and Data Analysis. *Mathematical Modelling of Natural Phenomena*. 15(24): 1–6.

[18] Lin, Q., Zhao, S., Gao, D., ... & He, D., 2020. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases*. 63: 211–216.

[19] Lauer, S. A., Grantz, K. H., Bi, Q., et al. 2020. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application, *Ann Intern Med*. DOI: 10.7326/M20-0504.

[20] Guan, W. J., Ni, Z. Y., Hu, Y., et al. 2020. Clinical characteristics of 2019 novel coronavirus infection in China. *The New England Journal of Medicine*. DOI: 10.1056/NEJMoa2002032.

[21] Bai, Y., Yao, L., Wei, T., et al. 2020. Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA*. doi:10.1001/jama.2020.2565.

[22] Bai, Y., Yao, L., Wei, T., et al. 2020. Presumed Asymptomatic Carrier Transmission of COVID-19, *JAMA*. 323(14):1406-1407. doi:10.1001/jama.2020.2565.

[23] Shi, Y., Wang, Y., Shao, C. et al. 2020. COVID-19 infection: the perspectives on immune responses. *Cell Death Differ*. <https://doi.org/10.1038/s41418-020-0530-3>.

[24] https://www.upi.com/Top_News/World-News/2020/04/11/WHO-investigating-reports-of-COVID-19-reinfection/4511586628287/.

[25] <https://7news.com.au/lifestyle/healthmedicine/china-reports-deeply-disturbing-coronavirus-development-c-765460>.