

Predicting the Number of COVID-19 Infections and Deaths in USA

Amarachukwu Felix Ebubeogu (✉ felixbosken@hotmail.com)

Universiti Malaya <https://orcid.org/0000-0001-8389-4860>

Chamberline Ekene Ozigbu

University of South Carolina Arnold School of Public Health

Donaldson F. Conserve

The George Washington University

Azizi Seixas

University of Miami

Paulinus Ofem

Universiti Malaya

Kholoud Maswadi

Jazan University

Research

Keywords: 2019 novel coronavirus, COVID-19, SARS-CoV-2, infection entropy, infection density

Posted Date: January 19th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1225501/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Globalization and Health on March 28th, 2022. See the published version at <https://doi.org/10.1186/s12992-022-00827-3>.

RESEARCH

Predicting the number of COVID-19 infections and Deaths in USA

Amarachukwu Felix Ebubeogu^{1*†}, Chamberline Ekene Ozigbu², Donaldson F Conserve³, Azizi Seixas⁴, Paulinus Ofem⁵ and Kholoud Maswadi⁶

*Correspondence:

felixbosken@hotmail.com

¹Department of Software Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

Full list of author information is available at the end of the article

[†]Equal contributor

Abstract

Background: Uncertainties surrounding the 2019 novel coronavirus (COVID-19) remain a major global health challenge and requires attention. Researchers and medical experts have made remarkable efforts to reduce the number of cases and prevent future outbreaks through vaccines and other measures. However, there is little evidence on how severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection entropy can be applied in predicting the possible number of infections and deaths. In addition, more studies on how the COVID-19 infection density contributes to the rise in infections are needed. This study demonstrates how the SARS-COV-2 daily infection entropy can be applied in predicting the number of infections within a given period. In addition, the infection density within a given population attributes to an increase in the number of COVID-19 cases and, consequently, the new variants.

Results: Using the COVID-19 initial data reported by Johns Hopkins University, World Health Organization and Global Initiative on Sharing All Influenza Data, the result shows that the original SAR-COV-2 strain has $R_0 < 1$ with an initial infection growth rate entropy of 9.11 *bits* for the United States. At close proximity the average infection time for an infected individual to infect others within a susceptible population is approximately 7 minutes. Assuming no vaccines were available, in the United States, the number of infections could range between 41,220,199 and 82,440,398 in late March 2022 with approximately, 1,211,036 deaths. However, with the available vaccines, nearly 48 Million COVID-19 cases and 706, 437 deaths have been prevented.

Conclusion: The proposed technique will contribute to the ongoing investigation of the COVID-19 pandemic and a blueprint to address the uncertainties surrounding the pandemic.

Keywords: 2019 novel coronavirus; COVID-19; SARS-CoV-2; infection entropy; infection density

Introduction

The COVID-19 outbreak has remained a universal health concern that requires urgent attention. Notably, Scientists have pointed out that COVID-19 is the newest coronavirus species with public health emergency [1], [2], [3]. Although several studies have been carried out aimed to provide useful information regarding the coronavirus pandemic, studies are still ongoing to uncover the root cause of the pandemic as well as solutions to address the outbreak. In December 2019, there were clusters of pneumonia cases in China. Later, investigations discovered that an unknown virus caused such clusters of pneumonia [4, 5]. The unknown virus is currently called the 2019 novel coronavirus. Coronaviruses are a large group of viruses that

consist of core genetic materials surrounded by specific protein spikes [6, 7]. Despite their unique nature, there are different types of coronaviruses that cause respiratory symptoms. These symptoms may range from the common cold to pneumonia, as in China's case, where it was first identified. These symptoms may be mild in most cases, whereas some cases are severe. For instance, fever, cough, and shortness of breath may be signs of mild symptoms. On the other hand, pneumonia, kidney failure, and death characterize severe cases. However, some kinds of coronaviruses are responsible for severe cases, such as severe acute respiratory syndrome coronavirus (SARS-CoV), first discovered in China in 2002-2003.

Another type of coronavirus that can also cause severe health damage is the Middle East respiratory syndrome-related coronavirus (MERS-CoV), identified in the Kingdom of Saudi Arabia in 2012 [8, 9]. Notably, COVID-19 poses a health threat and an economic threat across the globe [10]. Obviously, the gross domestic product (GDP) of almost all the affected countries has dropped tremendously, and as such, goods and services are affected along the supply chains. Besides, millions of schools from kindergarten to institutions of higher learning remain closed. Millions of both private and public companies, as well as their respective employers and employees, are under lockdown. Consequently, millions of workers have lost their jobs as a result of this pandemic.

Notwithstanding the remarkable achievements of the existing studies on providing useful information regarding the COVID-19 pandemic, the following gaps exist in the literature: (1) how the daily infection density attributes to increase in number of cases, (2) how daily infection entropy can be applied in predicting the number of cases and deaths and (3) the average time it may take for an infected individual to infect a susceptible population in close proximity.

Therefore, more studies are needed to support medical expert investigations as well as in their decision-making processes to uncover novel preventive measures to complement the available vaccines for the 2019 novel coronavirus. Hence, to better understand and characterize the initial behaviour of the virus, the current study aggregates the number of the initial reported COVID-19 cases before the emergence of the new variants and the corresponding numbers of deaths for March 2020 in the U.S. Based on the relevant data obtained, there is an indication that the daily infection entropy can be applied in predicting the likelihood of infection at a given period. In addition, there is a relationship between the daily infection density and the time of infection. The study therefore hypothesize that, with the emergence of the new variants, average time of infection in close proximity is <7 minutes.

The remainder of this paper is organized as follows. Section 2 presents the materials and methods. Section 3 presents our results. Section 4 presents discussion. Finally, Section 5 presents our conclusion and directions for future work.

Materials and methods

In this section, we present the measures applied to evaluate the current study. First, infection density β is defined as the ratio of the number of infections at constant population. In this study, the unit of measurement for β is number per population. Infection acceleration is defined as the change in daily infection velocity (v_i) over time. Note, the Infection acceleration, gain in virus momentum, rate of

infection, and increase in the number of infections represent the same measure. These measures indicate the change in behavior of the original SARS-COV-2 virus strain that may result in transmissible new variants. Other metrics include: entropy applied to determine the uncertainties in infections and deaths. The entropy is measured in *bits*.

In Figure 1, susceptible S , refers to the population who may be vulnerable to infection, infections (infected) I , are those who are infected by COVID-19, the recovered R is referred to those who with no symptom as a result of vaccines, antibodies or immune as well as those who may have died as a result of COVID-19. Considering the Susceptible Infections and Recovered (SIR) Model, we make the following assumptions:

1. A constant population with an increasing the number of infections.
2. Rate of infection in terms of infection acceleration influences the number of infections within a given population.
3. Increase in the number of infections is due to the rate of daily spread β .
4. The rate of daily spread and the infection density over a period depends on the rate at which the susceptible population is exposed to the virus and, consequently, gets infected. Hence, the rate of daily spread is equivalent to daily infection density.
5. A given population can easily be infected at close proximity with an infected individual.

Figures

Figure 1 SIR Model. Here β represents the infection density, v_0 represents the initial infection velocity and y_0 represents the initial phase (position) of infection.

The variables within the SIR model can be represented mathematically [11, 12, 13] as follows:

For the susceptible we have:

$$\frac{dS}{dt} = -\alpha SI \quad (1)$$

Assuming a decreasing number of susceptible at constant population, as S transits to I due to the rate of infections, the value of S decreases over time. Hence the value of α will remain (-ve) which shows a decrease in the number of susceptible [11, 12, 13].

For the infected we have:

$$\frac{dI}{dt} = \alpha SI - \beta I \quad (2)$$

Assuming an increase in the number of infections due to high contact rate. Hence the value of αSI remains (+ve).

However, if the ratio of daily reproduction rate and the rate of daily spread is greater than 1, (i.e., $\frac{\alpha}{\beta} > 1$), there is every possibility that the infection will rapidly

spread. On the other hand, if $\frac{\alpha}{\beta} < 1$, there could be spread with no exponential growth [11, 12, 13].

For the recovered we have:

$$\frac{dR}{dt} = \beta I \quad (3)$$

This indicates that those who have recovered from the infection due to antibodies or vaccines and may not be reinfected or even those who died. These values are excluded from the infection over time. Hence the $-\beta I$ in $\frac{dI}{dt}$ is regained as $+\beta I$ as presented in Equation 3.

where

S = the susceptible,

I = infected,

R = recovered

α = daily reproduction rate and

β = the rate of daily spread \equiv daily infection density.

Computation of α , β and R_0

Note, the daily reproduction rate is computed using the expression

$$\alpha = \frac{\text{No. of infections} - \text{No. of deaths}}{\text{population size}} \times 100$$

$\frac{\alpha}{\beta} = R_0$, which represents the basic reproduction number or the basic reproductive ratio, can be assumed to be the expected number of cases resulting from a single infection within a population where all individuals are susceptible. A higher value of R_0 means that the infection would be easily transmitted. $R_0 < 1$ means that the new cases will decrease over time, and ultimately, the outbreak will end on its own. $R_0 = 1$ means the cases may be stable over time, whereas $R_0 > 1$ indicates the virus may be autonomous, mutate into new variants, rapidly spread, and requires stringent and efficient control measures.

Looking back in March 2020, the virus's initial behavior shows how the infections will be over time with a high value of R_0 . Indicating a significant increase in positivity rate and new infection clusters, consequently increasing hospitalizations and deaths.

This implies that the daily infection density characterizes the daily reproduction rate of the SARS-CoV-2 virus. Hence,

$$f(\text{number of infections}) \Rightarrow \beta$$

The infection density β is influenced by the number of infections within the population, which typically depends on the rate of infections. The rate of infection in this context represents the acceleration of infection. At a constant population, an increase in infection rate will lead to an increase in infection density. Consequently, if the rate (acceleration of infection) decreases, it will lead to a decrease in infection density and a decrease in the rate of daily spread.

Thus,

$$f(\text{infection acceleration}) \Rightarrow \text{number of infections}$$

The above expression implies that the increase in the number of infections is a function of infection acceleration; hence, increasing the infection density. β .

Therefore,

$$f(\text{infection acceleration}) \Rightarrow \beta$$

The infection acceleration is defined as the change in infection velocity over the change in time as follows:

$$\text{infection acceleration} = \frac{\text{change in infection velocity } (v_i)}{\text{time}} \quad (4)$$

At a constant population, the amount of infections and infection density β within a population change as a result of infection acceleration. Thus,

$$\beta = \frac{\text{number of infections}}{\text{const. population}} \equiv \frac{f(\text{infection acceleration})}{\text{const. population}} \quad (5)$$

The infection density β , which depends on the infection acceleration at a contact population, can be expressed as shown in Equations (6) and (7):

$$\beta = \frac{f(\text{infection acceleration})}{\text{const. population}} \quad (6)$$

$$\beta = \frac{\text{change in infection velocity } (v_i)}{\text{time}} \quad (7)$$

$$v_i = \beta \times \text{time} \quad (8)$$

By integrating the infection velocity v_i over time, we obtain

$$\int v_i dt = \int \beta dt \quad (9)$$

$$\beta \int dt = \beta t + c \quad (10)$$

The initial infection velocity v_i before the outbreak equal to zero (i.e., $v_i = 0$). Therefore, replacing c with this initial infection velocity, we have

$$\beta \int dt = \beta t + (v_i = 0) \quad (11)$$

To determine the stage of infection in terms of its position y_i within the susceptible population over time, we apply the velocity relationship [14], expressed as

$$v = \frac{\text{change in position}(y_i)}{\text{time}(t)} \quad (12)$$

Hence, to obtain the target infection stage y_i with respect time (t), we integrate accordingly y_i with respect to time, we have

$$y_i(t) = \int v dt \quad (13)$$

Under the assumption that ($v = v_i$), by integrating the infection velocity over time, similar to Equation (9), yields

$$\int v dt = \int \beta t + v_0 dt \quad (14)$$

As we hope to determine the infection stage in terms of the position within a population, there is also a need to estimate the time these infections occur. From Equation (14) above, going further, we can determine time t of infection as

$$\int \beta t + v_0 dt = \int \beta t dt + \int v_0 dt \quad (15)$$

$$= \beta \int t dt + \int v_0 dt \quad (16)$$

$$= \beta \left(\frac{1}{2} t^2 \right) + v_0 t + c \quad (17)$$

Here, c which represents the initial infection stage before the outbreak is equal to zero (i.e., $c = y_i = 0$). In addition, at this early stage, the amount of infections as well as the infection velocity are both equal to zero.

$$y_i(t) = \beta \left(\frac{1}{2} t^2 \right) + v_0 t + (y_i = 0) \quad (18)$$

Hence, if $v = v_i$, we can now rewrite the above Equation as;

$$y_i(t) = \beta \left(\frac{1}{2} t^2 \right) + (v_i = 0)t + (y_i = 0) \quad (19)$$

Since before the outbreak, ($y_i = 0$ and $v_i = 0$), we have

$$y_i(t) = \beta \left(\frac{1}{2} t^2 \right) \quad (20)$$

To estimate the infection time t (i.e., the average time it take for an infected individual to transmit the virus daily to a susceptible population at proximity), from the resulting Equation (19), we have

$$t = \sqrt{\frac{2y_i}{\beta}} \quad (21)$$

where y_i represents the stage of daily infection within the population and β is the infection density.

Entropy

As applied in the current study, entropy can be referred to as the quantity of information uncertainties acquired from the information source measured in *bits* [15]. As the amount of uncertainties surrounding COVID-19 infection varies, the concept of entropy is applied to assess the initial daily infection uncertainties.

Thus, the entropy of an information source s on the daily infection growth rate (IGR) can be denoted by $IGR(s)$. To determine $IGR(s)$, we apply

$$IGR(s) = \sum_{i=1}^n p_i \log_2 \left(\frac{1}{p_i} \right) \quad (22)$$

We can rewrite the above equation as

$$IGR(s) = p_1 \log_2 \left(\frac{1}{p_1} \right) + \dots + p_n \log_2 \left(\frac{1}{p_n} \right) \quad (23)$$

where p_i represents the probability outcome for daily infection with respect to the uncertainties surrounding the information source and n is the number of sources of information.

Justification for choice of model

Despite remarkable achievements in the fight against the virus, there are still unknown factors surrounding the COVID-19 pandemic. However, the simplicity of our proposed model reveals how important it is to consider all possible parameters that might be responsible for the increase in the number of COVID-19 cases, consequently, the new variants. In addition, it shows how the daily infection density can be modeled via the SIR model as well as an easy to replicate approach. Notably, the current study does not involve any human or animal subjects. This study relied on the COVID-19 data reported by John Hopkins University, the World Health Organization and the Global Initiative on Sharing All Influenza Data. These datasets did not indicate the number of hospitalized persons or quarantined individuals but rather a generalized number of cases and deaths, respectively.

Data collection

The datasets applied in the current study are presented in Table 1. This study utilizes the initial COVID-19 records as reported by John Hopkins University [16], the World Health Organization [17] and the Global Initiative on Sharing All Influenza Data [18]. Table 1 presents the number of initially reported cases with respect to the original strain of SARS-CoV-2 virus in the United States.

Accumulated number of infections I_{cu}

The accumulated number of infections represents the possible number of COVID-19 cases over a given period (weeks or months) with respect to the average infection entropy. The accumulated number of infections is calculated as follows:

$$I_{cu(lowerlimit)} = N_{ti}(m * \left(\sum_{i=1}^n p_i \log_2 \left(\frac{1}{p_i} \right) \right)) \quad (24)$$

$$I_{cu(upperlimit)} = N_{ti}(m * (\sum_{i=1}^n p_i \log_2 (\frac{1}{p_i}))) * (2) \quad (25)$$

where

I_{cu} = the accumulated number of infections over a given period,

N_{ti} = the total number of infected cases at a given period,

m = number of months.

For example, if the total number of infections in the United States as of March 31, 2020, is 188,530 cases. To estimate the lower limit of possible number of infections in late March 2022(24 months apart), using Equation (24), we have:

$$I_{cu(lower\ limit)} = 188530(24(9.11)) = 41,220,199$$

To estimate the upper limit of infections using Equation (25), we have:

$$I_{cu(upper\ limit)} = 188530(24(9.11)) * 2 = 82,440,398$$

This number means that in late March 2022, the possible number of infections in the United States may be within the range of 41,220,199 and 82,440,398 but can be reduced if the necessary preventive guidelines are followed. Thus, the difference between the upper limit of infection and the possible number of cases prior to vaccine roll-out (i.e., $82,440,398 - 34,350,166 = 48,090,232$). This means that nearly 48 Million Americans have been prevented from COVID-19 infections and hospitalization since the vaccine rolled out. Note (34,350,166 is the upper limit of the predicted number of infections prior to the vaccines roll out in February 2021, 10 months apart).

COVID-19 average death growth rate

The average death growth rate represents the uncertainties in the number of deaths over a specified period. This metric allows us to keep track of the rate of deaths as a result of COVID-19 over time. The average death growth rate will also help us estimate the possible number of future COVID-19 death in the United States.

Accumulated number of deaths D_{cu}

The accumulated number of deaths represents the possible number of COVID-19 deaths cases over a given period (weeks or months) with respect to the average death entropy. The accumulated number of deaths is calculated as follows:

$$D_{cu} = N_{td}(m * (\sum_{i=1}^n p_i \log_2 (\frac{1}{p_i}))) \quad (26)$$

where

D_{cu} = accumulated number of deaths at given period,

N_{td} = the total number of deaths at a given period,

m = number of months

For example, if the number of deaths in the United States as of March 31, 2020, is 4053 deaths. To predict the possible number of deaths in late March 2022, using Equation (26), we have: $D_{cu} = 4053(24(12.45)) = 1,211,036$

This means with no vaccines available, in late March 2022, at least 1,211,036 deaths may be reported in the United States alone. However, with the available vaccines, nearly 706,437 deaths have been prevented.

Table 1 Numbers of reported COVID-19 cases and the corresponding numbers of deaths for March 2020 in the U.S. Here *IGR* represents the infection growth rate, *DGR* represents death growth rate, and β represents the infection density measured in (num/reported cases)

| March | No. of Reported Cases | Difference in the No. of Cases | IGR | No. of Deaths | Difference in the No. of Deaths | DGR | Entropy of IGR & DGR (<i>bits</i>) | Infection Density (β) | R_0 |
|-------|-----------------------|--------------------------------|---------------|-------------------|---------------------------------|------------------|--------------------------------------|-------------------------------|-------|
| 1 | 75 | 7 | 0.1 | 1 | 0 | 0 | 0.12 | 2.28e-7 | 0.985 |
| 2 | 100 | 25 | 0.33 | 6 | 5 | 5 | -11.2 | 3.04e-7 | 0.936 |
| 3 | 124 | 24 | 0.24 | 9 | 3 | 0.5 | 0.79 | 3.76e-7 | 0.928 |
| 4 | 158 | 34 | 0.27 | 11 | 2 | 0.22 | 0.65 | 4.80e-7 | 0.929 |
| 5 | 221 | 63 | 0.4 | 12 | 1 | 0.09 | 0.59 | 6.71e-7 | 0.945 |
| 6 | 319 | 98 | 0.44 | 15 | 3 | 0.25 | 0.75 | 9.68e-7 | 0.953 |
| 7 | 435 | 116 | 0.36 | 19 | 4 | 0.27 | 0.78 | 1.32e-6 | 0.957 |
| 8 | 541 | 106 | 0.24 | 22 | 3 | 0.16 | 0.51 | 1.64e-6 | 0.961 |
| 9 | 704 | 163 | 0.3 | 26 | 4 | 0.18 | 0.55 | 2.14e-6 | 0.962 |
| 10 | 994 | 290 | 0.41 | 30 | 4 | 0.15 | 0.67 | 3.02e-6 | 0.969 |
| 11 | 1301 | 307 | 0.31 | 38 | 8 | 0.27 | 0.68 | 3.95e-6 | 0.971 |
| 12 | 1630 | 329 | 0.25 | 41 | 3 | 0.08 | 0.41 | 4.95e-6 | 0.974 |
| 13 | 2183 | 553 | 0.34 | 48 | 7 | 0.17 | 0.61 | 6.63e-6 | 0.977 |
| 14 | 2770 | 587 | 0.27 | 57 | 9 | 0.19 | 0.56 | 8.41e-6 | 0.979 |
| 15 | 3613 | 843 | 0.3 | 69 | 12 | 0.21 | 0.64 | 1.09e-5 | 0.987 |
| 16 | 4596 | 983 | 0.27 | 87 | 18 | 0.26 | 0.71 | 1.39e-5 | 0.985 |
| 17 | 6344 | 1748 | 0.38 | 110 | 23 | 0.26 | 0.78 | 1.93e-5 | 0.980 |
| 18 | 9197 | 2853 | 0.45 | 150 | 40 | 0.36 | 0.90 | 2.79e-5 | 0.984 |
| 19 | 13779 | 4582 | 0.5 | 206 | 56 | 0.37 | 0.93 | 4.18e-5 | 0.986 |
| 20 | 19367 | 5588 | 0.41 | 255 | 49 | 0.24 | 0.74 | 5.88e-5 | 0.987 |
| 21 | 24192 | 4825 | 0.25 | 301 | 46 | 0.18 | 0.50 | 7.34e-5 | 0.988 |
| 22 | 33592 | 9400 | 0.39 | 414 | 113 | 0.38 | 0.83 | 1.02e-4 | 0.987 |
| 23 | 43781 | 10189 | 0.3 | 555 | 141 | 0.34 | 0.73 | 1.33e-4 | 0.986 |
| 24 | 54856 | 11075 | 0.25 | 780 | 225 | 0.41 | 0.73 | 1.67e-4 | 0.983 |
| 25 | 68211 | 13355 | 0.24 | 1027 | 247 | 0.32 | 0.69 | 2.07e-4 | 0.985 |
| 26 | 85435 | 17224 | 0.25 | 1295 | 268 | 0.26 | 0.64 | 2.59e-4 | 0.986 |
| 27 | 104126 | 18691 | 0.22 | 1695 | 400 | 0.31 | 0.63 | 3.16e-4 | 0.984 |
| 28 | 123578 | 19452 | 0.19 | 2220 | 525 | 0.31 | 0.54 | 3.75e-4 | 0.982 |
| 29 | 143491 | 19913 | 0.16 | 2583 | 363 | 0.16 | 0.44 | 4.36e-4 | 0.981 |
| 30 | 163788 | 20297 | 0.14 | 3141 | 558 | 0.26 | 0.59 | 4.97e-4 | 0.981 |
| 31 | 188530 (N_{ti}) | 24742 | 0.15 | 4053 (N_{td}) | 912 | 0.29 | 0.56 | 5.72e-4 | 0.978 |
| | | | $\sum = 9.11$ | | | $\sum_{12.45} =$ | $\sum = 8.05$ | | |

Note, (504, 599 is the predicted number of deaths prior to the vaccines roll out in February 2021).

Results

The current study shows that gain in momentum of COVID-19 is influenced by the number of infections within a given population, consequently resulting in a higher daily reproduction number R_0 . At constant population, again in the momentum of infection will result in a gain in infection density. Therefore, if the growth in momentum decreases, it will result in a lower infection density and a decrease in the ratio of daily reproduction rate and the rate of daily spread, respectively.

Notably, the exponential increase in the number of infections on a daily basis in March 2020, is characterized by high ratio of daily reproduction rate and the rate of daily spread $R_0 \geq 0.9$ and $R_0 < 1$ for the original SAR-COV-2 virus strain. However,

On the average infection and death growth rates, *IGR* achieved an entropy of 9.11 *bits*, whereas *DGR* achieved an entropy of 12.45 *bits*, as presented in Table 1. These uncertainties in terms of information entropy are the determinants for future forecasts on the possible number of infections and deaths. Thus, assuming

no vaccines were available, in the United States, the number of infections could range between 41,220,199 and 82,440,398, in late March 2022, with approximately, 1,211,036 deaths. However, with vaccine roll-out, approximately 48 million COVID-19 cases and 706,437 deaths have been prevented. Furthermore, the current study shows that it takes approximately 7 minutes on average for an infected individual to infect others within a susceptible population in close proximity. Hence, from the initial characteristics of the SAR-CoV-2 virus, a single person with COVID-19 can infect approximately 9 people within 1 hour and 216 people in a single day in the United States.

Discussion

In this study, we demonstrated how the daily reproduction number of SARS-CoV-2 virus can be determined through the infection density within a given susceptible population. In addition, the current study also shows how the information entropy obtained during the early phase of the outbreak in the United States can be applied as determinants for predicting the number of infections and deaths. While numerous underlying but unknown factors surrounding the spread of COVID-19 still exist [19], these unknown factors may also avert the reliability of existing models in predicting and monitoring COVID-19 [20, 21, 22, 23]. As a result, the SARS-CoV-2 virus continues to gain momentum with high stakes on human lives. Hence, it will be necessary to formulate models that can access the gain in momentum of the SARS-CoV-2 virus, which enables its ability to spread, resulting in multiple new variants [24, 25, 26, 27]. As the need for early detection of COVID-19 infections arises, certain predictive models can be helpful in identifying potential cases [28]. For instance, a logistic model was used to predict the total number of infections to be 4 million during the outbreak in the United States [29]. Some of the existing models include but are not limited to the susceptible-infectious-susceptible (SIS) model, the susceptible-infected-recovered-deceased (SIRD) model alongside the SIR model, the infectious disease dynamics model and the time-dependent dynamic model previously applied in predicting the outcome of COVID-19 [19, 30, 31, 32, 33]. An infectious disease dynamic model (SEIR) model was applied to model and predict the number of COVID-19 cases in Wuhan, China [32]. The results of that study indicate unstable values of daily reproduction rates, which may lead to a continuous increase in cases in Wuhan if public health intervention is not implemented.

A time-dependent dynamic model has been applied as a measure of public health intervention enhancement strategy through self-isolation [33]. The results obtained using the time-dependent dynamic model indicated that the daily reproduction rate of coronavirus has fallen below 1. However, the virus will continue to spread within the susceptible population [33]. On the possible time to transmit the virus from an infected person to a susceptible population, a study further reported that it might take approximately 10 mins for an infected person to produce 6,000 particles of aerosol [26]. These particles may be potentially harmful and may not be seen easily with the human eye [34].

A Markovian stochastic framework has been proposed to analyze both the reproductive ratio and the entropy of COVID-19. These results indicated a significant but steady difference in the COVID-19 reproduction ratio and entropy, respectively,

with a clear indication of the uncertainties surrounding the pandemic [35, 36]. It is therefore important to note that information entropy can be useful in differentiating between severe and mild COVID-19 patients [37, 38, 39].

However, a variation in daily reproduction ratio may vary from one location to another based on certain parameters, such as the stage of the outbreak (i.e., the rate of infection) [40]. Therefore, it is important to determine the value of the daily reproduction number at every stage of infection [40]. One of the factors responsible for the rapid transmission of COVID-19 is the daily reproduction number. However, it may be challenging to accurately determine the daily reproduction number [41, 42].

A compartmental mathematical model was formulated to predict the evolution of the virus in Cameroon and to analyze the reported cases in Brazil [43, 44]. The results achieved using these compartmental mathematical models indicated that the dynamics of COVID-19 disease are influenced by variations in the value of R_0 .

During the early phase of the pandemic, the basic reproduction ratio was estimated to range between 4.02 to 1.51 and 4.22 ± 1.69 for the United States and some parts of Europe, respectively, indicating a variation in the daily reproduction ratio [45, 46].

This study also recognized some other approaches applied in uncovering useful information regarding the pandemic. For example, a study aims to properly identify critical information in an unprecedented situation such as this outbreak via a natural language processing approach to classify COVID-19-related information [47]. Such an approach enabled the extraction of certain predictor variables that can be used in predicting the amount of reposted information regarding COVID-19 on social media [47]. Similarly, a simple model constructed from the rate of social media posts can be used as a reliable prediction model when analyzing the uncertainties surrounding the pandemic [48]. Hence, accurate prediction models can be useful tools to model the outbreak of the pandemic as well as in diagnosis prediction [49, 50]. As the fight against the SAR-CoV-2 virus continues, more studies are needed to uncover useful information hidden as a result of the uncertainties surrounding the pandemic.

Despite noteworthy achievements of the existing models, there is a need to indicate how severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection entropy can be applied in predicting the possible number of infections and deaths. In addition, how infection density within a given population contributes to an increase in the number of cases as well as the average time it may take for an infected individual to infect a susceptible population in close proximity. Therefore, the current study is carried out to fill the gap identified above. Several limitations to this study are worth noting. These includes lack of human subject, the use of initial COVID-19 datasets and the integration of the SIR model into our proposed model.

Conclusion and Future Work

The available COVID-19 vaccines have saved so many lives in the United States and beyond. However, several health concerns and uncertainties that have arisen in the wake of the COVID-19 pandemic have yet to be fully resolved. This study presented certain estimation models to determine the possible number of COVID-19 cases and deaths before and after vaccine roll-out in the United States. The

proposed approach shows that a high daily reproduction number for SARS-CoV-2 virus is characterized by an increase in the infection density of the original variant.

This study also shows that COVID-19 infection density can be derived via the Susceptible Infection and Recovered (SIR) model and may be applicable to other infectious diseases such as HIV. The initial behaviour of the SARS-CoV-2 virus indicates a high R_0 . Such information can be useful in monitoring the behaviour of the virus within a given period as well as in predicting possible future variants. On the projections of the pandemic in late March 2022, using the the initial SAR-CoV-2 information entropy of both infection and death growth rates as the determinants for future forecasts. Assuming no vaccines available in the United States, the current study projects that the number of infections could range between 41,220,199 and 82,440,398, with approximately 1,211,036 deaths.

Furthermore, the current study shows that it takes approximately 7 minutes on average for an infected individual to infect others within a susceptible population in close proximity. Hence, from the initial characteristics of the SAR-CoV-2 virus, this study reports that a single person with COVID-19 can infect approximately 9 people within 1 hour. Consequently, infecting about 216 people in a single day.

The proposed approach can enable other researchers to investigate other unknown factors responsible for the rapid spread of COVID-19, resulting in the emergence of new variants. This study therefore, hypothesize that the new variants (Delta and Omicron) may have a higher daily reproduction number with less daily infection entropy, consequently, spread faster and contagious.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

Ebubeogu Amarachukwu Felix and Paulinus Ofem design the research and the proposed method, Chamberline Ekene Ozigbu, Donaldson F. Conserve and Khouloud Maswadi performed data curation, data preprocessing and formal analysis, respectively. Azizi Seixas secured the funding, supervision and corrections. Ebubeogu Amarachukwu Felix and Paulinus Ofem drafted the manuscript, and all co-authors reviewed the manuscript.

Acknowledgements

The authors would like to thank the anonymous reviewers for their time and thoughtful comments while reviewing our manuscript.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Not applicable

Funding

Not applicable

Author details

¹Department of Software Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia. ²Department of Health Services Policy and Management, Arnold School of Public Health, 29208 Columbia SC, United States. ³Department of Prevention and Community Health, Milken Institute School of Public Health, The George Washington University, 20052 United States. ⁴Department of Psychiatry and Behavioral Sciences, The University of Miami Miller School of Medicine, 33136 Miami, FL, United States. ⁵Department of Software Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia. ⁶Department of Management Information Systems, Jazan University, 45142 Jazan, Saudi Arabia.

References

1. Adalja, A.A., Toner, E., Inglesby, T.V.: Priorities for the us health community responding to covid-19. *Jama* **323**(14), 1343–1344 (2020)
2. Wu, Y., Ho, W., Huang, Y., Jin, D.-Y., Li, S., Liu, S.-L., Liu, X., Qiu, J., Sang, Y., Wang, Q., *et al.*: Sars-cov-2 is an appropriate name for the new coronavirus. *The Lancet* **395**(10228), 949–950 (2020)
3. Chen, X., Liu, Z.: Early prediction of mortality risk among severe covid-19 patients using machine learning. *medRxiv*, 1–29 (2020)
4. Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O., Gu, J., Fan, Y., Zheng, C.: Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study. *The Lancet infectious diseases* **20**(4), 425–434 (2020)
5. Wei, X.-S., Wang, X.-R., Zhang, J.-C., Yang, W.-B., Ma, W.-L., Yang, B.-H., Jiang, N.-C., Gao, Z.-C., Shi, H.-Z., Zhou, Q.: A cluster of health care workers with covid-19 pneumonia caused by sars-cov-2. *Journal of Microbiology, Immunology and Infection* **54**(1), 54–60 (2021)
6. Alazawy, A., Arshad, S.-S., Bejo, M.-H., Omar, A.-R., Tengku Ibrahim, T.-A., Sharif, S., Bande, F., Awang-Isa, K.: Ultrastructure of felis catus whole fetus (fcwf-4) cell culture following infection with feline coronavirus. *Journal of electron microscopy* **60**(4), 275–282 (2011)
7. Hwa, K.-Y., Lin, W.M., Hou, Y.-I., Yeh, T.-M.: Molecular mimicry between sars coronavirus spike protein and human protein. In: 2007 *Frontiers in the Convergence of Bioscience and Information Technologies*, pp. 294–298 (2007). IEEE
8. Li, K., Wohlford-Lenane, C., Perlman, S., Zhao, J., Jewell, A.K., Reznikov, L.R., Gibson-Corley, K.N., Meyerholz, D.K., McCray Jr, P.B.: Middle east respiratory syndrome coronavirus causes multiple organ damage and lethal disease in mice transgenic for human dipeptidyl peptidase 4. *The Journal of infectious diseases* **213**(5), 712–722 (2016)
9. Chan, J.F.-W., Lau, S.K.-P., Woo, P.C.-Y.: The emerging novel middle east respiratory syndrome coronavirus: the “knowns” and “unknowns”. *Journal of the Formosan Medical Association* **112**(7), 372–381 (2013)
10. Baldwin, R., di Mauro, B.W.: Economics in the time of covid-19. A VoxEU.org Book, Centre for Economic Policy Research, London. Accessed **26** (2020)
11. Dallas, T.A., Carlson, C.J., Poisot, T.: Testing predictability of disease outbreaks with a simple model of pathogen biogeography. *Royal Society Open Science* **6**(11), 190883 (2019)
12. De Groot, M., Ogris, N.: Short-term forecasting of bark beetle outbreaks on two economically important conifer tree species. *Forest Ecology and Management* **450**, 117495 (2019)
13. Kelly, J.D., Park, J., Harrigan, R.J., Hoff, N.A., Lee, S.D., Wannier, R., Selo, B., Mossoko, M., Njoloko, B., Okitolonda-Wemakoy, E., *et al.*: Real-time predictions of the 2018–2019 ebola virus disease outbreak in the democratic republic of the congo using hawkes point process models. *Epidemics* **28**, 100354 (2019)
14. Felix, E.A., Lee, S.P.: Integrated approach to software defect prediction. *IEEE Access* **5**, 21524–21547 (2017)
15. Felix, E.A., Lee, S.P.: Predicting the number of defects in a new software version. *PloS one* **15**(3), 0229131 (2020)
16. Medicine, J.H.U.: Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). <https://www.coronavirus.jhu.edu/map.html>
17. Organization, W.H.: Coronavirus Disease (COVID-19) Outbreak Situation. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
18. Gostin, L.O.: hCoV-19 Tracking of Variants. <https://www.gisaid.org/hcov19-variants>
19. Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., Gloaguen, R.: Covid-19 pandemic prediction for hungary; a hybrid machine learning approach. *Mathematics* **8**(6), 890 (2020)
20. Rypdal, M., Sugihara, G.: Inter-outbreak stability reflects the size of the susceptible pool and forecasts magnitudes of seasonal epidemics. *Nature communications* **10**(1), 1–8 (2019)
21. Scarpino, S.V., Petri, G.: On the predictability of infectious disease outbreaks. *Nature communications* **10**(1), 1–8 (2019)
22. Zhan, Z., Dong, W., Lu, Y., Yang, P., Wang, Q., Jia, P.: Real-time forecasting of hand-foot-and-mouth disease outbreaks using the integrating compartment model and assimilation filtering. *Scientific reports* **9**(1), 1–9 (2019)
23. Luo, J.: Predictive monitoring of covid-19. *SUTD Data-Driven Innovation Lab* **446** (2020)
24. Kannan, S.R., Spratt, A.N., Cohen, A.R., Naqvi, S.H., Chand, H.S., Quinn, T.P., Lorson, C.L., Byrareddy, S.N., Singh, K.: Evolutionary analysis of the delta and delta plus variants of the sars-cov-2 viruses. *Journal of Autoimmunity* **124**, 102715 (2021)
25. Del Rio, C., Malani, P.N., Omer, S.B.: Confronting the delta variant of sars-cov-2, summer 2021. *Jama* **326**(11), 1001–1002 (2021)
26. Asadi, S., Bouvier, N., Wexler, A.S., Ristenpart, W.D.: The coronavirus pandemic and aerosols: Does COVID-19 transmit via expiratory particles? Taylor & Francis (2020)
27. Kupferschmidt, K., Wadman, M.: Delta variant triggers new phase in the pandemic. *American Association for the Advancement of Science* (2021)
28. Poletto, C., Scarpino, S.V., Volz, E.M.: Applications of predictive modelling early in the covid-19 epidemic. *The Lancet Digital Health* **2**(10), 498–499 (2020)
29. Bhardwaj, R.: A predictive model for the evolution of covid-19. *Transactions of the Indian National Academy of Engineering* **5**(2), 133–140 (2020)
30. Miller, J.C.: A note on the derivation of epidemic final sizes. *Bulletin of mathematical biology* **74**(9), 2125–2141 (2012)
31. Miller, J.C.: Mathematical models of sir disease spread with combined non-sexual and sexual transmission routes. *Infectious Disease Modelling* **2**(1), 35–55 (2017)
32. Wang, H., Wang, Z., Dong, Y., Chang, R., Xu, C., Yu, X., Zhang, S., Tsamslag, L., Shang, M., Huang, J., *et al.*: Phase-adjusted estimation of the number of coronavirus disease 2019 cases in wuhan, china. *Cell Discovery*

- 6(1), 1–8 (2020)
33. Tang, B., Bragazzi, N.L., Li, Q., Tang, S., Xiao, Y., Wu, J.: An updated estimation of the risk of transmission of the novel coronavirus (2019-ncov). *Infectious disease modelling* **5**, 248–255 (2020)
 34. Asadi, S., Wexler, A.S., Cappa, C.D., Barreda, S., Bouvier, N.M., Ristenpart, W.D.: Aerosol emission and superemission during human speech increase with voice loudness. *Scientific reports* **9**(1), 1–10 (2019)
 35. Wang, Z., Broccardo, M., Mignan, A., Sornette, D.: The dynamics of entropy in the covid-19 outbreaks. *Nonlinear Dynamics* **101**(3), 1847–1869 (2020)
 36. Bandt, C.: Entropy ratio and entropy concentration coefficient, with application to the covid-19 pandemic. *Entropy* **22**(11), 1315 (2020)
 37. Bajić, D., ajić, V., Milovanović, B.: Entropy analysis of covid-19 cardiovascular signals. *Entropy* **23**(1), 87 (2021)
 38. Albahri, A.S., Hamid, R.A., Albahri, O.S., Zaidan, A.: Detection-based prioritisation: Framework of multi-laboratory characteristics for asymptomatic covid-19 carriers based on integrated entropy–topsis methods. *Artificial intelligence in medicine* **111**, 101983 (2021)
 39. Hasan, A.M., Al-Jawad, M.M., Jalab, H.A., Shaiba, H., Ibrahim, R.W., AL-Shamasneh, A.R.: Classification of covid-19 coronavirus, pneumonia and healthy lungs in ct scans using q-deformed entropy and deep learning features. *Entropy* **22**(5), 517 (2020)
 40. Liu, Y., Gayle, A.A., Wilder-Smith, A., Rocklöv, J.: The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine* (2020)
 41. Viceconte, G., Petrosillo, N.: COVID-19 R0: Magic number or conundrum? *Multidisciplinary Digital Publishing Institute* (2020)
 42. Tao, Y.: Maximum entropy method for estimating the reproduction number: An investigation for covid-19 in china and the united states. *Physical Review E* **102**(3), 032136 (2020)
 43. Nabi, K.N., Abboubakar, H., Kumar, P.: Forecasting of covid-19 pandemic: From integer derivatives to fractional derivatives. *Chaos, Solitons & Fractals* **141**, 110283 (2020)
 44. Kumar, P., Erturk, V.S., Abboubakar, H., Nisar, K.S.: Prediction studies of the epidemic peak of coronavirus disease in brazil via new generalised caputo type fractional derivatives. *Alexandria Engineering Journal* **60**(3), 3189–3204 (2021)
 45. Gunzler, D., Sehgal, A.R.: Time-varying covid-19 reproduction number in the united states. *medRxiv* (2020)
 46. Linka, K., Peirlinck, M., Kuhl, E.: The reproduction number of covid-19 and its correlation with public health interventions. *Computational Mechanics* **66**(4), 1035–1050 (2020)
 47. Li, L., Zhang, Q., Wang, X., Zhang, J., Wang, T., Gao, T., Duan, W., Tsoi, K.K., Wang, F.: Characterizing the propagation of situational information in social media during covid-19 epidemic: A case study on weibo. *IEEE Transactions on Computational Social Systems* **7**(2), 556–562 (2020)
 48. Asur, S., Huberman, B.A.: Predicting the future with social media. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499 (2010). IEEE
 49. Ardabili, S.F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A.R., Reuter, U., Rabczuk, T., Atkinson, P.M.: Covid-19 outbreak prediction with machine learning. Available at SSRN 3580188, 1–38 (2020)
 50. Zoabi, Y., Shomron, N.: Covid-19 diagnosis prediction by symptoms of tested individuals: a machine learning approach. *medRxiv*, 1–9

Figures

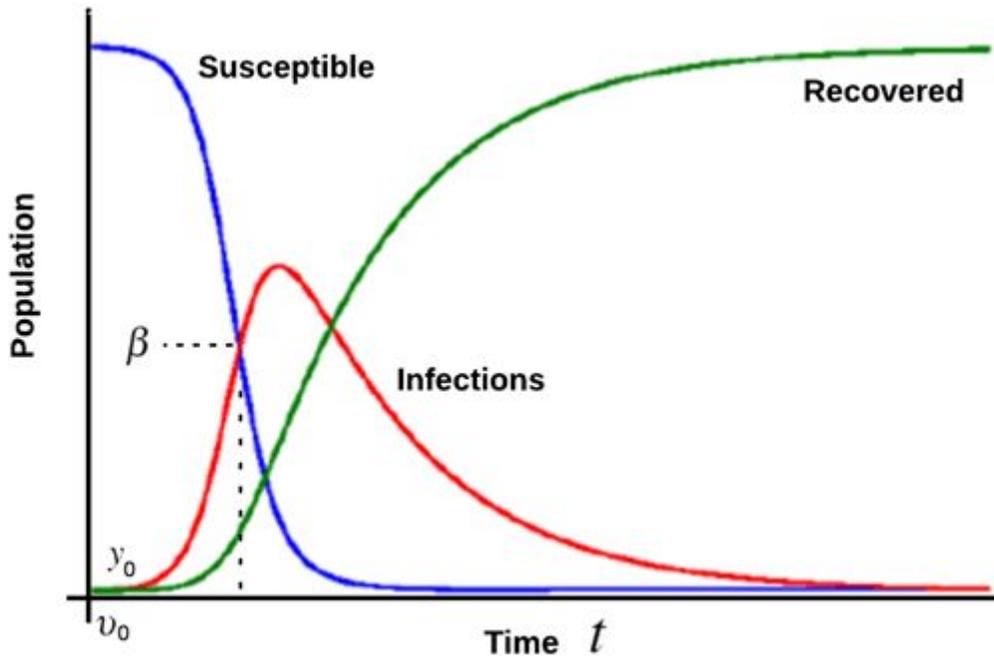


Figure 1

SIR Model. Here β represents the infection density, u_0 represents the initial infection velocity and y_0 represents the initial phase (position) of infection.