

A Review of Two Mathematical Models Describing COVID-19 Mortality Rates

Rice University
Department of Bioengineering
BIOE 391: Numerical Methods
May 9, 2021

Abstract

The COVID-19 pandemic has become a public global health crisis. To date, there have been over 156 million cases worldwide and 2.5 million deaths [1]. Mathematical models of COVID-19 spread are important tools in helping inform epidemiologists and public health officials. Understanding the trends and the nature of the pandemic's spread through human populations is essential in our ability to combat and control the pandemic in order to save numerous lives. This review analyzes two mathematical models that use two different classes of numerical methods to analyze the relationship between COVID-19 disease spread and death rates. The first model is named "Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis" by Yang et al. [2]. This paper investigates the case fatality rate (CFR) of Covid-19 using linear regression, which is an important mathematical tool to predict how lethal the virus is and compare death rates with positive cases between countries to better monitor the spread of the virus. The second paper, "A simulation of a COVID-19 epidemic based on a deterministic SEIR model" by Carcione et al. [3], uses Euler's forward divided difference method of solving ordinary differential equations (ODEs) to predict death rates from COVID-19 based on a number of parameters including infectious class, birth rate, and the number of exposed individuals. This review compares the two methods used to describe similar mortality consequences of the pandemic's spread by comparing the equations used in model formulation, the parameters taken into account in the model's execution, and the assumptions made to output key information to better inform policy aiming to curb the pandemic's deadly effects while maintaining economic stability. Based on the analysis of the differing numerical methods and their parameters, this paper evaluates the strengths and weaknesses of the models and identifies the applications of each method as a result of their tailoring to different needs at contrasting points in the pandemic.

Introduction

The spread of infectious diseases can be translated into mathematical models that are essential for informing public health policy. The history of mathematical models in epidemiology goes back to the eighteenth century, when Bernoulli used a mathematical model to evaluate the smallpox outbreak in 1766. His goal was to create a technique for analyzing the outbreak that could influence public health policy in order to make it more efficient and convince people to be inoculated against the virus [3-5]. He hypothesized that the spread of infectious diseases followed laws that could be formulated into a mathematical language [5]. It was not until the twentieth century, however, that the nonlinear dynamics of infectious disease transmission were really understood. In 1902, British doctor Sir Ronald Ross used mathematical modeling to investigate the effectiveness of intervention strategies for malaria. In 1906, the British scientist Hamer recognized that the diminishing density of susceptible populations could bring an epidemic to a halt [6]. His model is called Susceptible Infectious Recovered (SIR). In 1927, Kermack and McKendrick published a series of papers where they described the dynamics of disease transmission using a system of differential equations. By the end of the twentieth century, mathematical modeling in epidemiology began to be used to inform important aspects of public health policy. During the first two decades of the AIDS pandemic in 1981, modeling approaches were used to predict the further course of the epidemic and to identify the most effective prevention strategies. The later outbreak of the

SARS-CoV-1 virus in 2002 initiated the use of mathematical modeling for analyzing infectious disease outbreak in real time to assess the effectiveness of intervention measures. Models can highlight the differences between various diseases, which is important when creating a strategy to stop the further spread of the virus. For example, one disease that has been intensively analyzed is influenza A, and models have led to better understand the basic reproduction mechanisms of the virus. These models have implied that, in principle, an outbreak of influenza can be stopped with moderate levels of intervention, but such measures have to be taken rapidly and very early in the spread of the disease in order to be effective. In contrast, infectious diseases like measles have been found to need high levels of vaccination coverage for the elimination of the disease. These insights gained from mathematical models and analysis are essential for designing appropriate intervention strategies for the outbreak of such diseases [5].

In general, mathematical models used for diseases depend on the extent to which the disease has already spread. Before a pandemic starts, models can be used for identifying patterns to detect and respond to a virus in a timely manner to avoid further spread. Once the disease spreads further, these models help to identify where the pandemic started, how it spread in its earliest stages, and can give basic understanding of the pathogen and its epidemiological characteristics. Once a pandemic has spread inside of a country or to numerous regions, models are used for intervention and control strategies, forecasting the incidence rate, hospitalization rate, and mortality rate. These are also useful for allocating medical resources to different communities and understanding the change in collective behavior to public policies, such as lockdown and social distancing. Once a pandemic has slowed down, models are important to understanding the long term effects in social and economical sectors, as well as for developing strategies for economic recovery [7].

Due to its highly infectious and sometimes deadly nature, COVID-19 has quickly become a main focus of epidemiologists. Consequently, numerous models have been produced to better understand the spread of the disease and to mitigate its impacts [8]. The models developed to track the progress of this pandemic are used to supply information for the selection and implementation of public policies to reduce and contain the spread of the virus. COVID-19's global economic, social and health impacts continue to grow, making it increasingly important to find out which modeling techniques are most accurate [7]. Government officials can use these models to balance public goals of protecting health while limiting economic and social disruption [9]. Additionally, with the introduction of new COVID-19 vaccines, these models can be useful in determining what level of vaccination constitutes herd immunity, and what levels of vaccination are required to return to normalcy.

In addition to the point at which a model is implemented during a pandemic such as COVID-19, it is important to consider the local, regional, and cultural characteristics of the population in which the disease's spread is being modeled. This is because important parameters that influence disease transmission dynamics such as household and workplace design, interpersonal interactions, population age and spatial distribution, as well as other cultural and geographical factors vary both within and between countries [10]. For example, one study that searched for COVID-19 transmission clusters found that such transmission events tended to be found in different settings depending on the country or cultural region being studied: South Asian countries such as Singapore had prominent cluster transmission events in worker dormitories, China had problems with such cluster events mainly in households, and the United States and Western European countries had problems with transmission clusters in schools and meat packing plants early in the pandemic [11]. In this way, cultural and population demographic differences

must be accounted for with differing parameters and modifications to the same model, or by utilizing different models all together [10].

Moreover, the ways in which different countries respond to initial projections of pandemic urgency will vary and in turn require different modifications to the models that predict the pandemic's effects as a consequence of more urgent policies to curb such devastations. For example, China's extreme lockdown response to the pandemic spread of "no masks, no going outside" has differed greatly from the United States' more individualistic, less coordinated response [12]. Although the models analyzed in this review both serve the purpose of modeling pandemic spread and mortality prior to vaccine availability, the efficiency with which a given country obtains vaccines and implements their distribution also requires important modifications to such infection models since less of the population is susceptible to spreading the disease.

A closer inspection into the biochemical properties of COVID-19 explains the mechanisms by which the virus can spread so quickly within a given population, which is a primary reason why variability in model implementation is of such grave concern. COVID-19 is a disease spread through the virus SARS-CoV-2, which is a coronavirus that is closely related to the SARS-CoV-1 and MERS-CoV viruses that were the perpetrators behind the previously mentioned 2002 SARS pandemic in Guangdong, China and the 2012 endemic in the Middle East, respectively [13]. This class of respiratory viruses cause similar pneumonia-like symptoms such as acute respiratory syndrome and acute lung injury from alveolar damage, and SARS-CoV-2 has been shown to be more infectious than its aforementioned viral relatives [5,13]. Specifically, the World Health Organization (WHO) recorded 838 deaths due to MERS-CoV and 8437 infections with 813 deaths due to SARS-CoV-1, while in contrast, there have been over 156 million confirmed cases and 3.5 million deaths worldwide due to SARS-CoV-2 at the time of this paper's writing [1,14,15]. Albeit part of the reason for this large disparity in the infections and deaths due to these closely related viruses may be caused by different handling of their spreads by policymakers, SARS-CoV-2's higher infection rates are also attributed to its more rapidly transmissible nature and ability to spread asymptomatically [16]. The virus SARS-CoV-2 spreads primarily through respiratory droplets similar to other coronaviruses, meaning that breathing, coughing, and other forms of droplet ejection can transmit the virus through these droplets. The SARS-CoV-2 virus is more infectious than SARS-CoV-1, however, because it has biochemical properties that also allow it to be transmitted through aerosols and fomites on surfaces [17]. Furthermore, studies have shown that SARS-CoV-2 can remain infectious on surfaces for 5 to 10 hours following ejection from the body, allowing it to be transmissible through indirect contact [18]. Most worrisome, however, is the fact that SARS-CoV-2 has a 10-20 times greater affinity than SARS-CoV-1 for angiotensin converting enzyme 2 (ACE2), which is an enzyme that facilitates viral entry into human cells [19]. The virus' higher affinity for ACE2 means less viral material needs to be transmitted in order to infect an individual. All of these factors combine with the issue of the virus' novelty, which makes it so that human immune systems have little ability to recognize or defend against the virus, to create a huge problem with respect to worldwide spread [20].

Despite the multitude of determinants that render SARS-CoV-2 more infectious than other coronaviruses, it also fortunately exhibits lower mortality rates than its relatives [2,3]. Due to these differences between SARS-CoV-2 and related viruses, the models that track COVID-19's transmission and mortality rates must differ in the parameters used when compared to models that serve a similar purpose in simulating other infectious disease spreads. In particular, SARS-CoV-2's highly infectious

nature must be accounted for. For example, Carcione et al.'s ODE model uses the reproduction ratio R_0 , which measures the virus' ability to spread or reproduce, to incorporate the virus' infection ability. In order to spread to be a pandemic, R_0 must be greater than one. In comparison, Carcione et al. estimate an R_0 of 3 for a COVID-19 spread situation without any infection prevention measures. The specifics of this parameter will be explored in more detail in the model analysis section of the paper.

For models of fast-evolving pandemics such as COVID-19, therefore, the setting that the model describes in time, location, and societal response to disease spread is the most important consideration for the numerical methods and parameters used to produce useful results that inform further social policy measures. The two models detailed in this paper vary in both time and cultural setting of their target implementation, and therefore fulfill different needs despite modeling the same variable of interest in mortality due to the pandemic.

Yang et al.'s regression model uses a more straightforward, basic linear regression approach to modeling COVID-19 because its main purpose is to quickly inform a wide range of policymakers about a relatively new virus and its implications in a general sense. The linear and exponential regressions used with relatively few parameters taken into consideration are important to provide a general trend of the virus spread to allow health officials to determine the urgency and severity of actions that need to be taken. Because Yang et al.'s model was published in February 2020, which was before the virus had spread outside of China in large numbers [2], there was still little data available off of which to base the model. Therefore, a model with too much specificity regarding its parameters would produce unnecessary error due to the speculation that would be necessary to develop the needed information to run more complex, detailed models. In this way, the implemented linear regression model serves its intended purpose of informing both China and the world of the general trend in the relationship between COVID-19 deaths and infections. The more data-driven, less speculative nature of the regression model warrants its use of case fatality rate (CFR) as a measure of COVID-19's lethality. This is because CFR measures only the rate of deaths to confirmed infections without extrapolation regarding non confirmed cases. Regression models produce poor extrapolation estimates, so the preliminary COVID-19 CFR estimate that Yang et al.'s model achieves is sufficient for the model's purposes.

Conversely, Carcione et al.'s model uses a more complex ODE system with more parameters, as it is more focused on the consequences of the pandemic for the specific region of Lombardy, Italy. This ODE model was published on May 5th, 2020, which was well after the virus had begun to spread throughout the world. Consequently, Carcione et al.'s model is less focused on providing a quick, speculative estimate of the trajectory of COVID-19 infections and deaths, and is instead more concerned with providing a detailed model whose parameters enable specificity to accurately simulate the effects of COVID-19 in Lombardy based on the region's culture, living style, and response to pandemic spread among other factors. The more detailed model coupled with the use of an ODE system allows for at least preliminary extrapolation to future time periods, which is essential for informing government and societal action to prevent later catastrophic casualties. Instead of CFR, Carcione's model uses infection fatality rate (IFR), which measures the death rate as a proportion of all infections, including those that are unconfirmed such as asymptomatic individuals. A calculation of IFR, then, requires extrapolation to account for these unrecorded cases. Because of its implementation at a later time in the pandemic, Carcione's ODE model is able to extrapolate the IFR more accurately as more data was available about asymptomatic and undiagnosed infections. The use of IFR is preferred in this case because it allows

greater specificity and accuracy regarding the total number of infected individuals and therefore a better ability to extrapolate COVID cases and deaths into the future.

Model Basis and Equations

The first model being analyzed by Yang et al. uses linear regression in order to gain more information about the COVID-19 pandemic [2]. The model was developed in February 2020 - towards the beginning of the pandemic - and focuses specifically on China. The main statistic which the model focuses on is CFR, which, as mentioned in the introduction, is defined as the proportion of people who die from a given disease out of all of those infected with it over a certain period of time [21]. This is a common measure used by clinicians and public health practitioners to evaluate the severity of a disease and how to implement effective methods of control.

This model was developed using data on the daily COVID-19 cases and deaths, which was collected from the Wuhan Municipal Health Commission between January 10, 2020 and February 3, 2020. The data is split into three main regions: Wuhan City, Hubei Province (except Wuhan), and mainland China (except Hubei). With this data, the model employs simple linear regression to illustrate the relationship between two variables using a line of the form $y = y_0 + bx$, where y_0 is the y-intercept, b is the slope, x is the independent/predictor variable, and y is the dependent/outcome variable. The specific equation of the line which best represents the data is found using the least-squares

method, in which the sum of the squares of the estimate residuals ($\sum_{i=1}^n (y_i - y_0 - bx)^2$) is minimized. By minimizing the sum of squares, the unique coefficients giving the line best representing the relationship for a given data set can be found. Then, the final equation can be used to predict the value of one variable based upon the other.

For the model being examined, the starting point is defined as the time at which the first death was reported in order to avoid skewing the data due to the initial time period when no deaths occurred. In the linear regression, the number of confirmed cases is used as the predictor (x) variable, while the cumulative deaths is the outcome variable (y). The slope of the line relating these two variables can be used to estimate the CFR. In order to account for the patients' incubation time, the disease duration, hospital time, and other variables, the intercept of the fitted line can be adjusted (without changing the slope). The model also implements nonlinear regression to apply an exponential fit to the data. This fitted curve can then be compared to the simple linear regression fit by comparing their R-squared values. The specifics of how these two curve fitting methods compare to each other will be discussed in the results section.

The second model by Carcione et al. is also used to examine COVID-19 deaths, but focuses on the Lombardy region of Italy later in the pandemic. Though this model also looks at the relationship between deaths and infections, it takes a very different modeling approach [3]. This model - known as a Susceptible-Exposed-Infectious-Recovered (SEIR) model - uses a system of ordinary differential equations to describe the spread of COVID-19 and compute the number of infected, recovered, and dead individuals. In order to do this, it takes into account the number of contacts individuals have with each other, the probability of disease transmission, the incubation period, the disease's recovery rate, and the

disease's fatality rate. The infected and dead individuals are computed as a function of time on a per-day basis.

The initial population (defined as N_0) is split into four classes of people: susceptible ($S(t)$), exposed ($E(t)$), infected-infectious ($I(t)$), and recovered ($R(t)$), each of which are a function of time (t). The overall system of equations is illustrated by Equations 1-4.

$$\dot{S} = \Lambda - \mu S - \beta S \frac{I}{N}, \quad (1)$$

$$\dot{E} = \beta S \frac{I}{N} - (\mu + \epsilon) E, \quad (2)$$

$$\dot{I} = \epsilon E - (\gamma + \mu + \alpha) I, \quad (3)$$

$$\dot{R} = \gamma I - \mu R, \quad (4)$$

In this system, numerous parameters are used in order to estimate the desired variables. Λ represents the per-capita birth rate, μ represents the per-capita death rate, α represents the virus-induced average fatality rate, and β represents the probability of disease transmission per contact (which is multiplied by the number of contacts per unit time). In addition, ϵ represents the rate of progression from exposed to infectious, γ represents the rate of recovery of infectious individuals, and N represents the total number of humans alive at a given time. It is important to note that the reciprocal of ϵ is the incubation period of the disease, while the reciprocal of γ is the infectious period of the disease.

Based on this system of equations, the rate of susceptible individuals increases due to the birth rate, and decreases as death rate increases and as the rate of disease transmission increases. An increase in disease transmission then leads to an increase in the rate of exposed individuals, which decreases due to deaths and due to progression from exposed to infectious. As more exposed individuals become infected, the rate of infected-infectious individuals increases. This can decrease as individuals either die (either due to the disease or due to other causes) or recover. Finally, the recovered rate depends only on the rate of recovery and the rate of natural deaths. Figure 1 displays a flowchart which illustrates the relationship between each variable.

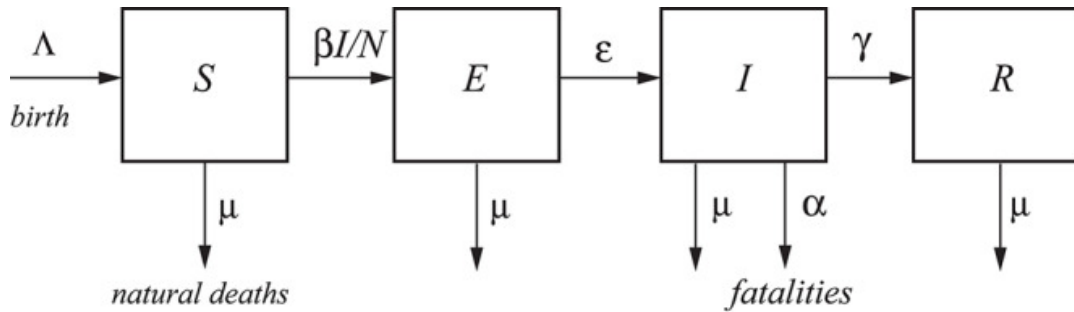


Figure 1. Flowchart Representing Mathematical Model

The parameters from the base system of differential equations can also be applied in order to estimate other statistics, most notably the rate of deaths (Equation 5) and the infection fatality rate (IFR) (Equation 6). Infection fatality rate, as mentioned in the introduction, is defined as the proportion of infected individuals who die due to the disease. It is used for similar purposes as the CFR calculated in the first model; however, IFR is always lower than CFR, because CFR only takes into account diagnosed individuals whereas IFR accounts for all infected individuals.

$$\dot{D}(t) = -\dot{N}(t) = -(\dot{S} + \dot{E} + \dot{I} + \dot{R})(t) \quad (5)$$

$$\text{IFR (\%)} = 100 \cdot \frac{\alpha I_\infty}{(\alpha + \gamma) I_\infty - \mu R_\infty} \approx 100 \cdot \frac{\alpha}{\alpha + \gamma} \approx 100 \cdot \frac{\alpha}{\gamma} \quad (6)$$

In order to solve the system of ordinary differential equations used in this model, Euler's Forward Difference scheme was applied. The general basis of Euler's forward difference method is that, in order to integrate a differential equation between two points (x_i and x_{i+1}), the formula

$y_{i+1} = y_i + f(x_i, y_i) h$ is used, where $f(x_i, y_i)$ represents the value of the differential equation function at the beginning of the interval. This value is used to estimate the slope, and it is multiplied by h , which is the selected step-size. This method assumes that the slope at the beginning of the interval (point x_i) can be used to approximate the slope throughout the whole interval of size h . The Euler approximation is then repeated for the next intervals until the entire range of the independent variable has been covered, providing function estimates for the entire desired data set [22].

When applying the Euler's method formula to the COVID-19 model, the step-size h is replaced with a time-step dt , which is set to a value of .01 days. The value $f(x_i, y_i)$ is represented by the value of the rate equations (Equations 1-4) at point (x_i, y_i) , and y_i is represented by the value of S, E, I, or R at this point. The resulting system of equations can then be used to solve for the desired values of S, E, I, and R at each time point, and these values can be used to calculate the death rate at a given time (Equations 7-11). Finally, Equation 12 describes the total number of deaths at a given time.

$$S^{n+1} = S^n + dt \left(\Lambda - \mu S^n - \beta S^n \frac{I^n}{N^n} \right), \quad (7)$$

$$E^{n+1} = E^n + dt \left[\beta S^n \frac{I^n}{N^n} - (\mu + \epsilon) E^n \right], \quad (8)$$

$$I^{n+1} = I^n + dt \left[\epsilon E^n - (\gamma + \mu + \alpha) I^n \right], \quad (9)$$

$$R^{n+1} = R^n + dt \left(\gamma I^n - \mu R^n \right), \quad (10)$$

$$\dot{D}^n = -(\dot{S}^n + \dot{E}^n + \dot{I}^n + \dot{R}^n)(t), \quad (11)$$

$$D(t) = N_0 - N(t) \quad (12)$$

Results

Upon applying linear regression in the model by Yang et al., linear fits were generated for the number of cases vs. deaths in Wuhan, Hubei province (except Wuhan), mainland China (except Hubei), and all of mainland China. The resulting plots and equations are shown in Figure 2.

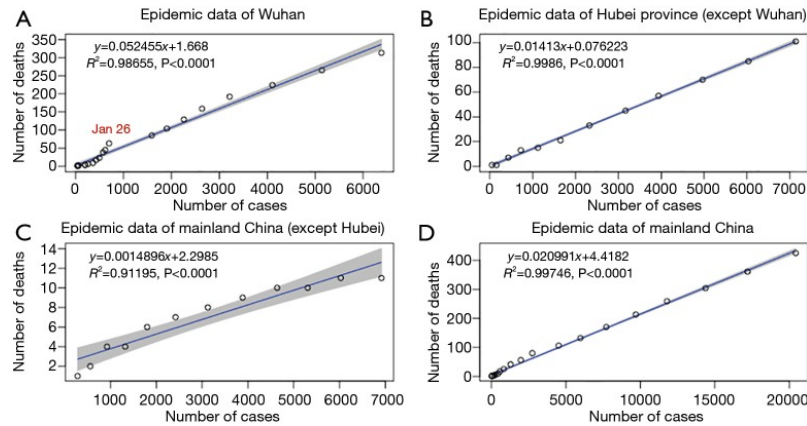


Figure 2. Linear regression results for each region of China

For each of these data sets, R-squared values, which show the goodness of fit of the model, were also computed. All of the R-squared values were between .912 and .999, meaning the simple linear regression produces a strong fit for every location. In the Wuhan region, although the overall R-squared of .98655 demonstrates a strong linear fit, there appeared to be two distinct trends on the graph: before January 26, the number of deaths is increasing with an increasing slope (and, thus, an increasing CFR), while after January 26 the number of deaths is increasing with a decreasing slope (a decreasing CFR). For this reason, the Wuhan model was modified with separate exponential models before and after January 26, producing R-squared values of 0.995 and 0.997 (Figure 3).

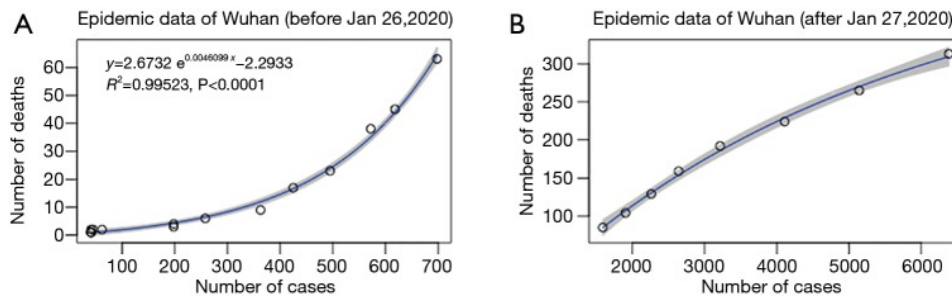


Figure 3. Exponential models of Wuhan deaths before and after January 26

In addition, the region of mainland China (except Hubei) has the lowest R-squared value of .91195, and is therefore also remodeled exponentially, increasing the R-squared to .99235 (Figure 4). The final two

regions of Hubei province (except Wuhan) and mainland China as a whole have very high R-squared values, and are therefore not remodeled with an exponential fit.

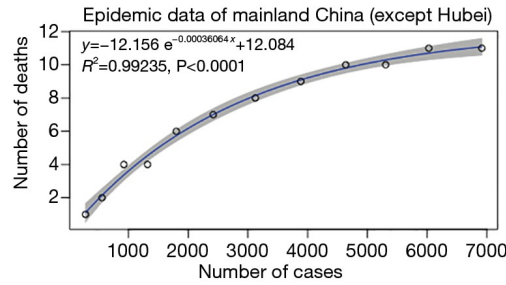


Figure 4. Exponential fit of mainland China (except Hubei)

Based on the slopes of the generated models, the CFR for the region of Wuhan was 5.25%, the CFR for Hubei (excluding Wuhan) is 1.41%, the CFR for mainland China (except Hubei) is .15%, and the CFR for mainland China as a whole is 2.10%. The high CFR in Wuhan makes sense, as the model was generated very early in the pandemic, and COVID-19 originated in Wuhan. In addition, the estimate for mainland China is significant because it suggests that the COVID-19 CFR in mainland China is significantly lower than that SARS-CoV-2 (which had a CFR of 6.4%). Although the data generated by the model is logical for the early stages of the pandemic, it is not necessarily representative of the CFR in China later in the pandemic; in March 2021, the case fatality rate for all of mainland China was 4.8%, which is over twice as large as the model's estimate [23].

To solve the ODEs in the model by Carcione et al., the following base parameters were implemented in the system of equations: $N_0 = 10$ million, $\alpha = 0.006/\text{day}$, $\beta = 0.75/\text{day}$, $\gamma = 0.125/\text{day}$, $\epsilon = 0.333/\text{day}$. Furthermore, a balance between births and natural deaths was assumed: $\Lambda = \mu N$. This assumption is important because it implies that all deaths in the model are due to COVID-19. This assumption is reasonable due to the fast-moving nature of the pandemic. The initial conditions were set as: $S(0) = N_0 - E(0) - I(0)$, $E(0) = 20,000$, $I(0) = 1$ and $R(0) = 0$ [24]. Once Euler's Method was used to solve the system of ODEs (Equations 7-10), the number of individuals for each class of people was plotted against time (Figure 5).

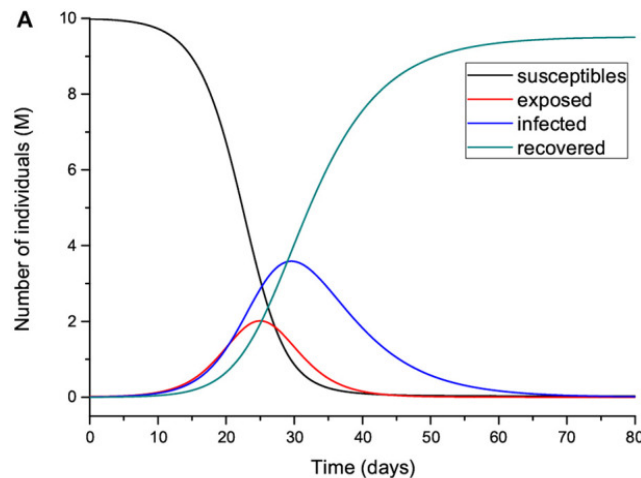


Figure 5. Number of individuals in the different classes (millions)

Finally, the rate of deaths is calculated using Equation 11, and the total number of deaths is calculated using Equation 12 (as detailed in the previous section). These results were also plotted against time (Figure 6).

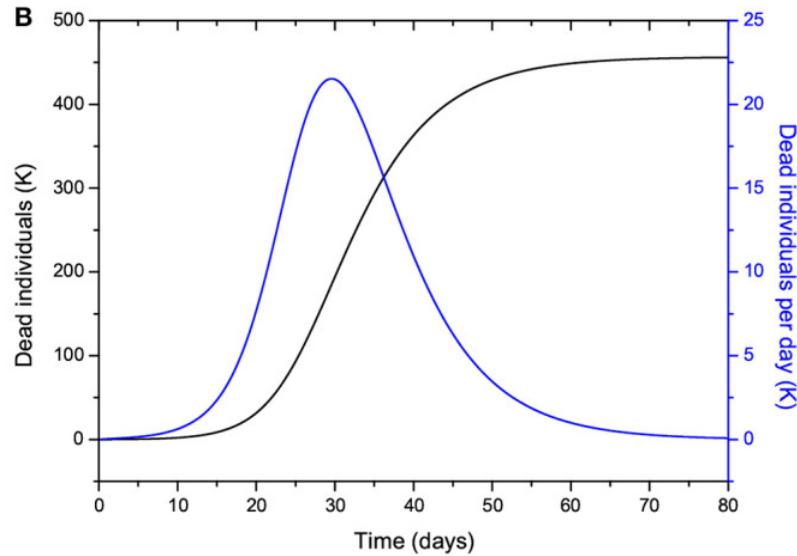


Figure 6. Total number of deaths and number of deaths over time (thousands)

Based on the generated solutions, the total number of deaths in Lombardy, Italy is estimated to be 15,600, and the IFR is approximately 0.57%. Like the linear regression model, due to the fact that the estimate of the total deaths was based on parameters in the beginning of the pandemic (before it had spread as widely and become as severe), the model significantly underestimates the total number of deaths in the Lombardy region. As of March 6, 2021, there had been 33,080 deaths due to COVID-19 in Lombardy, which is over twice as large as the model's prediction [25].

In the solutions producing the above plots, the value of R_0 , which is used to represent the degree of isolation ($\beta = (\gamma + \alpha) \times R_0$) is assumed to equal 5.72, representing imperfect isolation. However, the R_0 value can easily be adjusted to look at the impact of more complete isolation on the spread and deadliness of the pandemic. Decreasing R_0 over the course of the pandemic, representing a continually higher degree of isolation, generally causes a decrease in the number of infected and dead individuals. In addition, starting isolation (decreasing R_0) earlier also increases the total number of deaths.

Numerical Methods Used: Pitfalls and Successes

The simplicity of the linear regression model is both an advantage and a pitfall: it doesn't require a lot of parameters which makes it easy to use, efficient, and easy to find the CFR. However, there are many assumptions made in this model, such as the variables exhibiting a linear relationship, the error terms are independent, the error terms have a common variance and the error terms are normally distributed [26]. However, the simplicity makes the model less accurate, and it is difficult to extrapolate

the model's results over the course of the pandemic. In fact, while linear regression proved to be a relatively accurate way to model deaths early in the pandemic, as the disease continued to spread, the trend became less linear and showed sub-exponential growth in some regions [27]. The paper was completely data-driven in its extrapolation, and did not take into account other factors that may occur as the disease spreads that might change the shape of the curve. The lack of precision in the use of the linear regression model for extrapolation can be demonstrated by the fact that the CFR in mainland China in March 2021 (4.8%) was almost twice the value calculated by the model (2.1%) [23]. Also, the comparison between simple linear fitting and exponential curve fitting showed that the exponential curve fit the data better. Specifically, for the relationship between the number of cases and the number of deaths in mainland China excluding Hubei, the R-squared value for the linear fit was 0.912 while the R-squared value for the exponential fit of the same data was 0.992.

The ODE model uses more parameters compared to the linear regression model, which can make it more accurate. These parameters can be easily adjusted. However, the addition of these parameters makes the model complex and more difficult to use. Some of these parameters can be difficult to estimate, and may also change over the course of the pandemic, making extrapolation using the model inaccurate. This is demonstrated by the fact that the model predicted the total number of deaths in Lombardy, Italy at the end of the pandemic to be 15,600; however, by March 2021, 33,080 people had died due to COVID-19 in the region [25]. Specifically, the authors mention that they lacked confidence in their estimation of compliance to social distancing measures over time, the initial number of exposed and infected people in the region, the infectious period of the disease, and the fatality rate. Thus, each additional parameter introduces an additional place for error. The ODE model also makes several assumptions, including assuming birth and natural death numbers are the same, that all changes in population are due to deaths from disease, that recovered individuals cannot become reinfected, and that slope in the beginning of interval can be used to estimate the entire step.

Another drawback of the ODE model is its use of Euler's method. One major downside to using this method is the error associated with it. While truncation error can be reduced by decreasing the step size, continuing to reduce the step size will cause roundoff error to dominate. Furthermore, any errors experienced for each given step can accumulate and be propagated over the interval of interest to increase global error. However, the global error can also be reduced by cancellation of local error based on the error's sign over the interval. A possible alternative to Euler's method is to use the midpoint method or Heun's iterative method. Although these two alternatives might be slightly more complicated, they hold many benefits over Euler's method. For the midpoint method, rather than saying the slope at the beginning of the interval is representative of the whole interval as in Euler's method, the slope at the middle of the interval is used. Heun's iterative method allows the error to become even smaller by using iterations to converge to a solution with a specified error [22].

A final drawback of these papers is that both use data and parameters from specific regions. The paper that used a linear regression model used data from China while the paper that specified an ODE model used parameters that described the state of the pandemic in Northern Italy. In other words, the two models analyzed in this review use localized data rather than global data to calculate their results. As such, it would be difficult and dangerous to directly apply the conclusions from these two papers in a global context.

Conclusion

Mathematical models have been used since the late eighteenth century to evaluate the outbreaks of diseases. Through time, these models have progressed and become more precise when evaluating a specific virus. They have been used extensively for creating strategies relating to public health policy and mitigating the impact of these diseases. Over the last year, COVID-19 is the disease that has been modeled the most extensively, as it has caused a global pandemic that has immensely affected social and economical structures. Having an accurate model of the spread of the disease has been critical in helping government officials make decisions to reduce transmission of the virus and stabilize the economy.

The models by Yang et al. and Carcione et al. both utilize numerical methods in order to evaluate deaths as a consequence of infections caused by the COVID-19 pandemic, but each takes a very different approach to doing so. Yang et al.'s model primarily utilizes linear regression on existing data, in which the slope of the line relating the number of cases to the number of deaths represents the COVID-19 case fatality rate. This curve fitting method, while not necessarily the most accurate method of analyzing the pandemic, is very simple and straightforward. On the other hand, Carcione et al.'s model solves a system of ordinary differential equations using Euler's method in order to find the infection fatality rate and total number of deaths due to the pandemic. This method is more complex, but it is also more accurate and utilizes scientific assumptions about the nature of the virus. Thus, it can be used to model the impacts of a wider variety of factors. The major downside to both numerical methods is their lack of ability to accurately predict the COVID-19 pandemic in the future, although Carcione et al.'s model has a more theoretical ability to do this; based on more recent data, the models underpredict the CFR and total deaths.

References

- [1] World Health Organization, 2021. WHO Coronavirus (COVID-19) Dashboard. [online] Covid19.who.int. [Link](#).
- [2] S. Yang et al., “Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis,” *Ann. Transl. Med.*, vol. 8, no. 4, p. 128, 2020. [Link](#).
- [3] J. M. Carcione, J. E. Santos, C. Bagaini, and J. Ba, “A simulation of a COVID-19 epidemic based on a deterministic SEIR model,” *Front. Public Health*, vol. 8, p. 230, 2020. [Link](#).
- [4] Blower S, Bernoulli D. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. 1766. *Rev Med Virol.* 2004;14(5):275–88.
- [5] M. Kretzschmar and J. Wallinga, “Mathematical Models in Infectious Disease Epidemiology,” *Modern Infectious Disease Epidemiology*, pp. 209–221, 2009.
- [6] P. Rohani and A. A. King, “Never mind the length, feel the quality: the impact of long-term epidemiological data sets on theory, application and policy,” *Trends in Ecology & Evolution*, vol. 25, no. 10, pp. 611–618, 2010.
- [7] A. Adiga, D. Dubhashi, B. Lewis, M. Marathe, S. Venkatramanan, and A. Vullikanti, “Mathematical Models for COVID-19 Pandemic: A Comparative Analysis,” *Journal of the Indian Institute of Science*, vol. 100, no. 4, pp. 793–807, 2020.
- [8] E. K. Grillo Ardila, L. E. Bravo Ocaña, R. Guerrero, and J. Santaella-Tenorio, “Mathematical models and the coronavirus, COVID-19,” *Colombia Medica*, 2020.
- [9] L. P. James, J. A. Salomon, C. O. Buckee, and N. A. Menzies, “The Use and Misuse of Mathematical Modeling for Infectious Disease Policymaking: Lessons for the COVID-19 Pandemic,” *Medical Decision Making*, vol. 41, no. 4, pp. 379–385, 2021.
- [10] R. M. Eggo, J. Dawa, A. J. Kucharski, and Z. M. Cucunuba, “The importance of local context in COVID-19 models,” *Nature Computational Science*, vol. 1, no. 1, pp. 6–8, 2021.
- [11] Q. J. Leclerc, N. M. Fuller, L. E. Knight, S. Funk, and G. M. Knight, “What settings have been linked to SARS-CoV-2 transmission clusters?,” *Wellcome Open Research*, vol. 5, p. 83, 2020.
- [12] X. Yu and R. Yang, “COVID-19 transmission through asymptomatic carriers is a challenge to containment,” *Influenza and Other Respiratory Viruses*, vol. 14, no. 4, pp. 474–475, 2020.
- [13] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, “COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses,” *Journal of Advanced Research*, vol. 24, pp. 91–98, 2020.
- [14] “The American Journal of Public Health (AJPH) from the American Public Health Association (APHA) publications,” American Public Health Association (APHA) publications, 07-Aug-2019. [Link](#)
- [15] N. S. Zhong, B. J. Zheng, Y. M. Li, L. L. M. Poon, Z. H. Xie, K. H. Chan, P. H. Li, S. Y. Tan, Q. Chang, J. P. Xie, X. Q. Liu, J. Xu, D. X. Li, K. Y. Yuen, J. S. M. Peiris, and Y. Guan, “Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003,” *The Lancet*, vol. 362, no. 9393, pp. 1353–1358, 2003.
- [16] “COVID-19, MERS & SARS,” National Institute of Allergy and Infectious Diseases. [Link](#)

- [17] T. Galbadage, B. M. Peterson, and R. S. Gunasekera, "Does COVID-19 spread through droplets alone?," *Front. Public Health*, vol. 8, 2020.
- [18] N. van Doremalen et al., "Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1," *N. Engl. J. Med.*, vol. 382, no. 16, pp. 1564–1567, 2020.
- [19] M. Hoffmann et al., "SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor," *Cell*, vol. 181, no. 2, pp. 271-280.e8, 2020.
- [20] U. Katie Kerwin McCrimmon, "Viruses 101: Why the new coronavirus is so contagious," Uchealth.org, 17-Mar-2020. [Link](#)
- [21] R. A. Harrington, "Case fatality rate," Encyclopedia Britannica. 05-May-2020.
- [22] S. C. Chapra and R. P. Canale, Numerical methods for engineers, 7th ed. New York, NY: McGraw-Hill Professional, 2014.
- [23] "China: novel coronavirus case death rate trend 2021," Statista.com. [Link](#)
- [24] Chowell G, Fenimore PW, Castillo-Garsow MA, Castillo-Chavez *CJ Theor Biol.* 2003 Sep 7; 224(1):1-8.
- [25] "Italy: coronavirus deaths by region," Statista.com. [Link](#)
- [26] H.-W. Cho and C. Chu, "A tale of two fields: mathematical and statistical modeling of infectious diseases," *Osong Public Health Res. Perspect.*, vol. 2, no. 2, pp. 73–74, 2011.
- [27] J. Chu, "A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain," *PLoS One*, vol. 16, no. 3, p. e0249037, 2021.