Original Research

Corresponding author:
George Kafatos;
Email: gkafatos@amgen.com

CAMBRIDGE
UNIVERSITY PRESS

A Curve-Fitting Approach for Generating Long-Term Projections of COVID-19 Mortality

George Kafatos PhD[1] (iD), George Seegan PhD[2], Bagmeet Behera PhD[3],
David Neasham PhD[1], Brian Bradbury PhD[2] and Neil Accortt PhD[2]

[1]Center for Observational Research, Amgen Ltd, Uxbridge, UK; [2]Center for Observational Research, Amgen Inc, Thousand Oaks, CA, USA and and [3]Center for Observational Research, Amgen Research Munich GmbH, Munich, Germany

Abstract

Objective: This study aims to develop a curve-fitting approach for long-term COVID-19 mortality projections and evaluate its effectiveness as a scalable, data-driven tool for pandemic forecasting.
Methods: The basic characteristics of a dynamic curve-fitting approach capable of generating long-term projections are described. To demonstrate its utility, the model was retrospectively applied using mortality data from the start of the pandemic, January to June 2020 (6-month data), to project into the period between June 2020 and April 2021 (11-month projections).
Results: For scenarios with the best fit, the difference between observed and projected total deaths varied in the projection period between 7.7% and 28.2%.
Discussion: When the COVID-19 pandemic started in early 2020, there was lack of understanding regarding its long-term impact. Available mathematical models were complex and typically provided short- and mid-term projections. The approach described generates long-term projections that are relatively easy to implement and can be enhanced to include other parameters such as vaccine impact or virus variants. The method could prove to be a valuable tool during a future pandemic.

Introduction

On 3rd of December 2019, the first COVID-19 cases were reported in Wuhan, China.[1] On 30th of January 2020, with 7,828 cases reported worldwide (99% in China), the WHO Director-General declared the novel coronavirus outbreak as a Public Health Emergency of International Concern.[2] On 11th of March 2020, with more than 118,000 cases and 4,291 deaths in 114 countries, the WHO declared COVID-19 a global pandemic.[3] During the early months of the COVID-19 pandemic, there was substantial uncertainty regarding the likely scale and trajectory of the outbreak, which made long-term projections particularly challenging.[4] The rapid escalation of the pandemic (that required rapid response), along with initial underestimations by some officials, highlights the need to have readily available models that can project the pandemic's long-term impact and guide policy-makers.

During the COVID-19 pandemic, mathematical models were used to assess the impact of different Public Health Intervention (PHI) measures or to predict the extent or duration of the disease.[4] Different types of models have been published projecting the COVID-19 mortality over time.[5] Most of the models presented in the literature were Susceptible-Infected-Recovered (SIR) or its extension Susceptible-Exposed-Infected-Recovered (SEIR) compartment-type models.[5–7] Other approaches included the empirical fitting of the cumulative curves (limited growth models) and simulating disease transmission in the community.[6–8]

Most infectious disease models generate short- and mid-term projections (up to 4-5 months). For example, information on demographics, social behaviors, and PHI measures can be used as input for the mathematical compartmental models to predict the peak or the end of an epidemic wave but it becomes more complex to project further into the future. Similarly, limited growth models rely on the "infection growth rate," which fluctuates over time, making it difficult to accurately estimate long-term. A review of COVID-19 mortality projection models, published in July 2020, found seven models that provided projections of more than 4 weeks, with a maximum forecasting period of 4 months.[5] A paper published on May 2021 used an approach combining the output of six models to project COVID-19 outcomes in the USA over a period of 6 months.[9] A 2022 publication describes a Bayesian dynamic model that can be used for long-term projections up to 12 months.[10]

While short-term projections were critical for managing immediate healthcare resource needs such as hospitalizations, ICU capacity, and healthcare workforce allocation, there was also an urgent need for long-term projections, spanning 6 months to a year or more.[4] Early in the pandemic, decision-makers in public health and government agencies required long-term forecasts to anticipate future waves, guide vaccine distribution planning, and develop sustainable

public health interventions. Within the private sector, organizations relied on these projections for workforce management, economic impact assessments, and operational continuity. Specifically, in the pharmaceutical industry, long-term projections were crucial for clinical trial operations, manufacturing and supply chain planning, and workforce logistics, ensuring the continued development and distribution of essential treatments and vaccines. Two early publications attempted to predict the overall impact of COVID-19 pandemic with some predicting a worst-case scenario of 2 million and 500,000 deaths in the USA and UK, respectively, by the end of the pandemic.[11,12] Based on assumptions on infection rate, mortality rates, and herd immunity level, it is possible to obtain crude estimates of the overall impact of a pandemic. However, it is challenging to provide mortality projections at specific timepoints. For example, at the start of the pandemic (early 2020), it was difficult to estimate the total number of COVID-19-related deaths that might occur in a given country by the end of 2021.

Due to the relatively limited global spread of previous coronavirus outbreaks (e.g., SARS and MERS), past influenza pandemics served as the primary reference point for understanding potential COVID-19 pandemic curves. Previous analyses of historical influenza pandemics suggested that pandemics are likely to manifest in multiple epidemic waves over time.[13] Similarities and differences between past influenza pandemics and the COVID-19 pandemic were discussed at the time with a reasonable assumption around COVID-19 epidemiology being the manifestation of epidemic waves.[13] Using, as an assumption, the occurrence of epidemic waves, the current work describes a dynamic curve-fitting model for projecting long-term mortality of the COVID-19 pandemic. The methodology is straightforward and can be easily applied across different countries and regions. It was later extended to incorporate assumptions such as the uptake and impact of COVID-19 vaccines and change in virus variants. Herein, the main framework of the approach is presented.

Methods

Data Source

Daily mortality statistics from the Institute for Health Metrics and Evaluation (IHME) were used as input data.[14] The IHME website included free-to-download data from multiple countries and was updated regularly. For the use case described within this paper, the 30th April 2021 update was used, with the data being truncated to 1st of June 2020 for the early projections.[14] Apart from providing longer-term projections, the IHME data were selected because they were not limited to hospitalized patients. Instead, they were derived from multiple data sources, including Johns Hopkins University mortality statistics and supplemented by government websites for a number of locations and subnational estimates.[15] A 2022 paper showed that compared with other data sources, the Johns Hopkins University COVID-19 data were among the most reliable.[16] This comprehensive and reliable measure of COVID-19-related mortality was particularly valuable, especially during the early pandemic phase where infection data were highly variable.

Time Period and Country Selection

For this study, we analyzed COVID-19 mortality data from 4th February 2020 to 1st June 2020 to project into the period 2nd June 2020 to 30th April 2021. The timeframes were chosen to cover the first major epidemic wave and validate projections over an extended

period before widespread vaccination significantly altered mortality trends. Data from France, Germany, Italy, Spain, the UK, and the USA were used as mortality data from these countries were initially released by IHME.

Curve-Fitting Methodology

The approach presented in this study focuses on projecting the timing and distribution of future pandemic waves while accounting for potential seasonal variations in transmission. A wave is defined as a period in which daily reported deaths rise, peak, and then decline to a local minimum before increasing again. Seasonal factors, such as increased indoor activities during colder months, can influence the frequency and intensity of the waves. A wave is considered to have concluded once daily deaths stabilize at a low point before a new surge begins. The curve-fitting approach for generating long-term mortality projections can be described as follows:

Step 1: Fitting a curve to the observed data
A curve can be fitted on an observed epidemic wave (W_i) based on mortality data.

A skew-normal distribution of the following form can be used:

$$f(x) = \frac{2}{s}\varphi\left(\frac{x-m}{s}\right)\phi\left(\lambda\frac{x-m}{s}\right), \qquad (1)$$

where $\hat{m}$, $\hat{s}$, $\hat{\lambda}$ are the location, scale, and skewness parameters, respectively. The standard normal density function is denoted by φ and the cumulative normal distribution function by Φ.[17] The resulting distribution can be weighted based on the total number of observed deaths N, i.e., $N \times f(x)$.

As an example, the Spain COVID-19 daily deaths were used. The time period covers the first deaths in early March 2020 until the 1st of June 2020.[14] The observed (IHME smoothed data) daily deaths and the fitted skew-normal distribution ($\hat{m} = 22 Mar 2020$, $\hat{s} = 25.9$, $\hat{\lambda} = 3.9$) are shown in Figure 1a. The skewness parameter of $\hat{\lambda} > 0$ shows that there is a steep increase in deaths and a slow decline in time.

Step 2: Distance between future epidemic waves
The timing of the peak of a future epidemic wave can be calculated based on the average distance between past epidemic peaks. Assuming there are $k = i-1$ observed waves, the timing of the future wave peak, $W_i\,peak$, can be estimated as:

$$W_i peak = W_{i-1} peak + \frac{1}{i-1}\sum_{k=1}^{i-1}(W_k peak - W_{k-1} peak) \qquad (2)$$

At the start of the pandemic, when there was limited past data, an assumption was made that the next epidemic wave will commence upon the completion of the previous wave, i.e., upon lifting the PHI measures. A simple way to express this mathematically is to limit the overlap between the wave distributions.

Following from Equation (1), the cumulative distribution can be expressed as:

$$F(x) = 2\int_{-\infty}^{0}\int_{-\infty}^{\frac{x-m}{s}}\phi_2(u_1, u_2)\,du_1\,du_2, \qquad (3)$$

where $\varphi_2(u_1, u_2)$ is a bivariate normal distribution with correlation $\rho = \frac{\lambda}{\sqrt{1+\lambda^2}}$.[17]
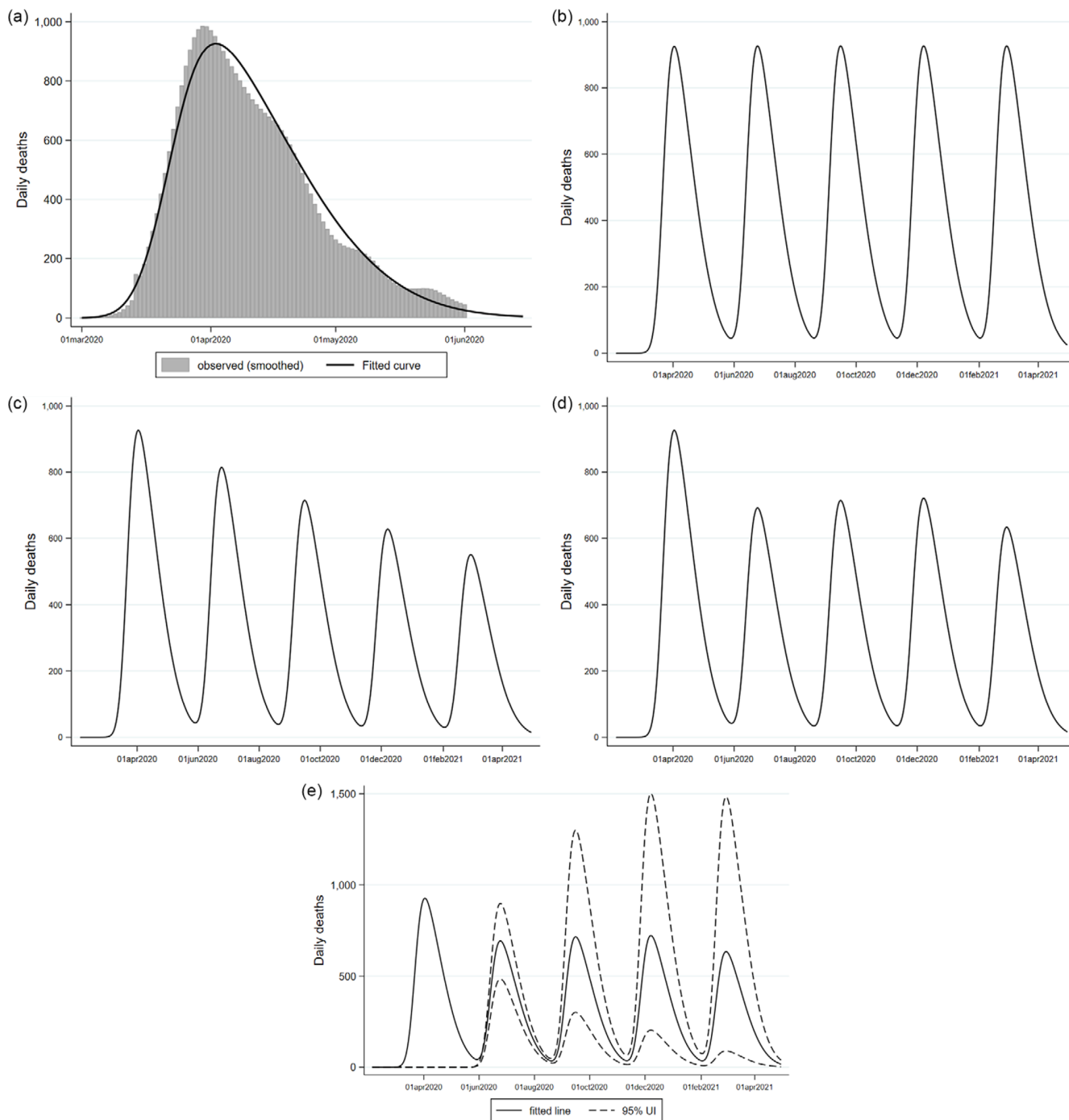
Figure 1. Step-by-step fitting COVID-19 long-term mortality projections.
1a Fitting curve to observed data (example of Spain daily deaths).
1b Adding future epidemic waves.
1c Accounting for change in the susceptible population.
1d Seasonality effect.
1e Uncertainty intervals.

Assuming, for example, a small proportion of overlap α, the tails x_1 and x_2 of the Wave i distribution can be calculated by solving the formulas:

$$F(x_1) = \alpha/2$$

$$F(x_2) = 1 - \alpha/2$$

An alternative way to obtain estimates of x_1 and x_2 is to simply calculate the $(\alpha/2)$th and $(1 - \alpha/2)$th percentiles, respectively, of the skew-normal distribution.

Assuming there are $i = 1,..,n$ future waves, the location of the peak of a future wave, $W_i peak$, can be calculated as:

$$W_i peak = W_{i-1} peak + (x_2 - W_{i-1} peak) + (W_{i-1} peak - x_1) \quad (4)$$

Using Equations (3) and (4), future epidemic waves were projected for the period 1st of June 2020 until 30 April 2021 based on the Spain COVID-19 example above. Under the assumption of a 1% ($\alpha = 0.01$) overlap between distributions, there was an 83-day time period between wave peaks (Figure 1b).

Step 3: Accounting for change in the susceptible population

As the number of infections increases, the susceptible population will tend to decrease, resulting in smaller future epidemic waves (assuming no reinfection).

Assuming a proportion of the population, p_{HE}, will achieve herd immunity once $(100 \times p_{HE})\%$ of adults have been infected, the herd immunity threshold (HIT) can be calculated as:

$$HIT = \text{adult population} \times p_{HE} \qquad (5)$$

Assuming a COVID-19 mortality ratio (MR), the HIT will be achieved when the total deaths are:

$$HIT_{deaths} = HIT \times MR \qquad (6)$$

Assuming we want to predict the size of second wave, $W_2 size$, following the end of the first epidemic wave, a "susceptible ratio (SR)," SR_2, can be calculated as:

$$SR_2 = \frac{CW_1}{HIT_{death}}, \qquad (7)$$

where CW_1 denotes the cumulative deaths following the end of Wave 1.

The size of Wave 2 ($W_2 size$) can then be adjusted allowing for the change in the susceptible population as follows:

$$W_2 size = (1 - SR_2) \times W_2 size \qquad (8)$$

For $i = 1,..,n$ subsequent waves, Equation (8) can be generalized as:

$$W_{i+1} size = (1 - SR_{i+1}) \times W_{i+1} size, \qquad (9)$$

where $SR_{i+1} = \frac{CW_i}{HIT_{death}}$.

The impact of the reduced susceptible population following each projected wave can be viewed in Figure 1c assuming the Spanish population will achieve herd immunity once 80% of adults have been infected ($p_{HE} = 0.8$) with a mortality ratio of $MR = 0.01$.

Step 4: Seasonality effect

A seasonality effect can also be incorporated (assumes transmission of infectious disease will increase during winter months due to increased indoor activity). Assuming a seasonality effect expressed as a proportion p_{SE}, that would equate to an impact 0 to $2 \times p_{SE}$ on the epidemic waves.

For a subsequent Wave i, the size can be adjusted allowing for seasonality effect as follows:

$$W_i size = \begin{cases} (1 - p_{SE}) \times W_i size, & \text{if } 1\,Jun \leq W_i peak \leq 31\,Aug \\ (1 + p_{SE}) \times W_i size, & \text{if } 1\,Dec \leq W_i peak \leq 28\,Feb \end{cases} \qquad (10)$$

Equation (10) is only valid for countries within the Northern Hemisphere. For regions in the Southern Hemisphere, the seasonality effect should be adjusted by shifting the time frame of increased transmission to June–August, corresponding to the winter months in that hemisphere. Similarly, the lower transmission period should be adjusted to December–February, aligning with the summer season. A paper at the time predicted that COVID-19 may exhibit seasonal patterns similar to influenza, with higher transmission in winter months.[12]

Assuming a 15% (or $p_{SE} = 0.15$) seasonality effect based on model assumption published in early 2020, that would equate to an impact of 0% to 30% on the epidemic waves.[12] The impact of seasonality can be viewed in Figure 1d. The peak of Wave 2 occurs in June so the wave size is reduced by 15%, compared to Wave 1. Waves 4 and 5 fall during winter months, and the wave size is inflated by 15%. There is no change for Wave 3 as the peak occurs in September.

Step 5: Uncertainty intervals

Uncertainty intervals (UI) can be calculated by combining two sources of variation:

$$UI = UI_{pIR} \times UI_{pUI}, \qquad (11)$$

where UI_{pIR} denotes the uncertainty around infection rate, i.e., size of waves and UI_{pUI} the increase of uncertainty with time, say, i.e., for each subsequent wave.

Assuming the infection rate uncertainty can be quantified as p_{IR} then for $i-1$ observed waves:

$$UI_{pIR} = (1 \pm p_{IR}) \times W_i size \qquad (12)$$

The uncertainty increases for each projected wave, W_i (denoted by p_{UI}) as follows:

$$UI_{pUI} = [1 \pm (1 - i) \times p_{UI}] \times W_i size \qquad (13)$$

Using Spain as an example, the uncertainty in the infection was expressed based on a predefined parameter, $p_{IR} = 0.3$, which represents the expected variability in the infection rate due to factors such as changes in PHI measures and population behavior. Additionally, the uncertainty was assumed to increase with each wave, modeled as $p_{UI} = 0.2$ to reflect growing unpredictability in long-term projections. Figure 1e shows the combined UI around the projections.

Pandemic Scenarios

Based on a review analyzing epidemiological trends from the 1918-1919, 1957-1958, and 2009-2010 influenza pandemics, three different pandemic wave scenarios were selected[13]:

- Scenario 1—Equal epidemic waves: Assuming that subsequent waves will follow patterns similar to those observed in the first wave, including social mixing behavior and PHI measures. Under this scenario, a series of approximately equal waves will occur in both size and duration.
- Scenario 2—Large second wave: Assuming PHI measures will be relaxed following the first wave or there will be reduced adherence to these measures, or both. In a similar pattern to what was observed during the 1889 and 1918 flu pandemics, a mild first wave is followed by a larger second wave and then smaller subsequent waves.
- Scenario 3—Small subsequent waves: Assuming strict PHI measures and public adherence to these measures. A large first wave is followed by small subsequent waves.[13,18]

Figure 2 shows the three pandemic wave scenarios for Spain. For Scenario 2, the second wave was set to be four times the size of the initial wave. For Scenario 3, the subsequent waves were assumed to be half the size of the first wave. These assumptions were based on an early paper modeling the impact of intervention methods on basic reproduction number R_0.[12]

Assessment of the Projecting Model

The projecting methodology and its assumptions were retrospectively validated by comparing estimates against the observed
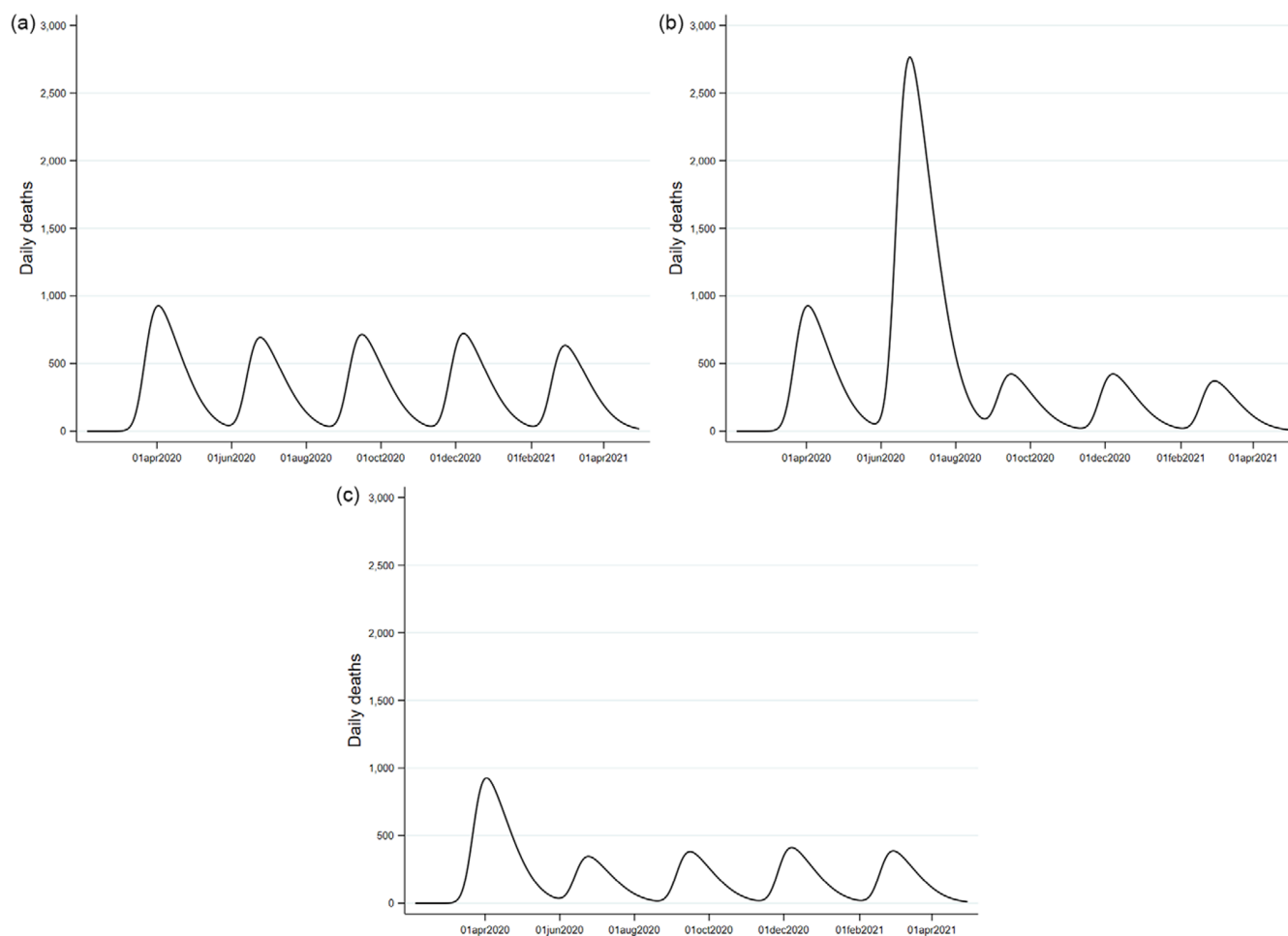
Figure 2. Different pandemic scenarios (based on Spain daily deaths).
2a Scenario 1—Equal waves.
2b Scenario 2—Large second wave.
2c Scenario 3—Small subsequent waves.

deaths occurring during the same period. COVID-19 mortality daily data were used from France, Germany, Italy, Spain, the UK, and the USA.[2] A skew-normal distribution was fitted to daily deaths occurring between 2nd February 2020 and 1st of June 2020 (first wave). The distance between future waves was set a priori as 1% overlap between fitted distributions. The proportion of adult population infected at the herd immunity level was set as $p_{HE} = 0.8$ and the COVID-19 mortality rate as $MR = 0.01$. A 15% seasonality effect was assumed. Three scenarios were considered as discussed above.

The model performance was assessed using the following outcomes: (1) the difference in the total number of deaths between projected and observed data during the 11-month period and (2) the number of projected epidemic waves (vs. actual waves that occurred). The assumption related to the distance between future epidemic waves was also assessed in a sensitivity analysis using wave distribution overlaps of 5% and 0.1%.

Software

The data manipulation and analysis were carried out using Stata 16.1 (StataCorp).

Results

Table 1a shows that four epidemic waves were projected for the six countries during the period June 2020 to April 2021. For Scenario 1, the highest projected COVID-19 mortality was in the USA (354,394; UI: [139,658, 652,270]) followed by the UK (145,214; UI: [57,670, 267,682]) and Spain (103,345; UI: [40,165, 193,496]). Germany had the lowest projected mortality (32,483; UI: [12,088, 61,745]). In Scenario 2, the projected deaths were higher than in Scenario 1, with as many as 569,654 deaths (UI: [308,738, 897,939]) estimated for the USA. Conversely, Scenario 3 projected lower deaths than Scenarios 1 and 2, with Germany having the smallest estimate at 16,510 deaths (UI: [6,096, 31,344]) (Table 1b).

For the period June 2020 to April 2021, the highest number of deaths was observed in the USA (461,302) followed by the UK (99,782). The lowest number of deaths was observed in Germany (72,966). Comparing with the period prior to June 2020 (timepoint from which projections were made), Germany reported 8.6 times more deaths due to a large second wave. Following a large first wave, the UK and Spain took stricter PHI measures and recorded 2 and 1.4 times more deaths, respectively, for the period June 2020 to April 2021 compared to the period prior to June 2020 (Table 1c).

Table 1. Assessment of projections for the period June 2020 to April 2021

(a) Number of waves

Countries	Observed	Projected
France	3	4
Germany	2	4
Italy	3	4
Spain	2	4
UK	2	4
USA	3	4

(b) Projected deaths (June 2020 to April 2021)

Countries	Scenario 1—Equal waves		Scenario 2—Large second wave		Scenario 3—Small subsequent waves	
	Projected	Observed—Projected n (%)	Projected	Observed—Projected n (%)	Predicted	Observed—Projected n (%)
France	**84,415 (32,412, 158,542)**	**−6,068 (7.7%)**	131,483 (70,592, 210,014)	−53,136 (67.8%)	44,451 (16,713, 83,790)	+33,897 (43.3%)
Germany	32,483 (12,088, 61,745)	+40,483 (55.5%)	**52,400 (26,578, 86,620)**	**+20,565 (28.2%)**	16,510 (6,096, 31,344)	+56,456 (77.4%)
Italy	**92,191 (39,045, 165,964)**	**−6,616 (7.7%)**	149,231 (86,797, 225,587)	−63,657 (74.4%)	49,358 (20,360, 89,299)	+36,217 (42.3%)
Spain	103,345 (40,165, 193,496)	−53,151 (105.9%)	148,618 (85,208, 227,553)	−98,424 (196.1%)	**57,291 (21,385, 108,634)**	**−7,098 (14.1%)**
UK	145,214 (57,670, 267,682)	−45,432 (45.5%)	212,486 (123,471, 320,386)	−112,703 (112.9%)	**80,771 (30,704, 150,209)**	**+19,012 (19.1%)**
USA	**354,394 (139,658, 652,270)**	**+106,908 (23.2%)**	569,654 (308,738, 897,939)	−108,352 (23.5%)	186,333 (71,670, 341,574)	+274,969 (59.6%)

Note: Smallest differences by country in bold.

(c) Observed deaths

Countries	Prior to Jun 2020	Jun 2020 to Apr 2021	Increase magnitude
France	24,985	78,347	3.1
Germany	8,481	72,966	8.6
Italy	33,557	85,575	2.6
Spain	34,678	50,194	1.4
UK	50,849	99,782	2.0
USA	108,337	461,302	4.3

The epidemic curve by country can be viewed in Supplementary material (Figure S1).

For Scenario 1, the smallest differences between projected and actual deaths were observed in France and Italy followed by the USA. For France, the projected deaths (84,415; UI: [32,412, 158,542]) for the period June 2020-April 2021 were 7.7% (6,068) higher than the observed deaths. For Italy, the projected deaths (92,191; UI: [39,045, 165,964]) were 7.7% (6,616) higher than the observed deaths. The projected deaths for the USA (354,394; UI: [139,658, 652,270]) were 23.2% (106,908) lower than observed deaths for the same period. For Scenario 2, there were 23.5% (108,352) higher projected deaths in the USA (569,654; UI: [308,738, 897,939]) compared to actual deaths. In Germany, the projected deaths (52,400; UI: [26,578, 86,620]) were 28.2% (20,565) lower than the observed deaths. For Scenario 3, the projected deaths for Spain (57,291; UI: [21,385, 108,634]) were 14.1% (7,098) higher than the observed deaths. In the UK, the projected deaths (80,771; UI: [30,704, 150,209]) for the period June 2020 to April 2021 were 19.1% (19,012) lower than the observed deaths for the same time period (Table 1b).

A sensitivity analysis examining the time interval between successive epidemic waves showed that for Scenarios 1 and 2, the 0.1% overlap between fitted wave distributions (representing the longest interval between waves) produced projections that were more closely aligned to the observed deaths. For Scenario 3, the

Table 2. Varying distance between waves (i.e., different levels of overlap)

(a) Scenario 1—Equal waves

Countries	Difference (Observed—Projected) n (%)		
	0.1% overlap	1% overlap	5% overlap
France	+15,718 (20.1%)	−6,068 (7.7%)	**−21,732 (27.7%)**
Germany	+49,552 (67.9%)	+40,483 (55.5%)	**+33,167 (45.5%)**
Italy	+9,174 (10.7%)	**−6,616 (7.7%)**	−29,046 (33.9%)
Spain	**−29,609 (59.0%)**	−53,151 (105.9%)	−71,565 (142.6%)
UK	**−11,615 (11.6%)**	−35,836 (35.9%)	−68,340 (68.5%)
USA	+193,340 (41.9%)	+106,908 (23.2%)	**+7,401 (1.6%)**

Note: Smallest differences by country in bold.

(b) Scenario 2—Large second wave

Countries	Difference (Observed—Projected) n (%)		
	0.1% overlap	1% overlap	5% overlap
France	**−35,676 (45.5%)**	−53,136 (67.8%)	−57,655 (73.6%)
Germany	+29,135 (39.9%)	+20,565 (28.2%)	**+15,008 (20.6%)**
Italy	**−52,785 (61.7%)**	−63,657 (74.4%)	−70,854 (82.8%)
Spain	**−84,849 (169.0%)**	−98,424 (196.1%)	−89,228 (177.8%)
UK	**−93,530 (93.7%)**	−107,267 (107.5%)	−120,701 (121.0%)
USA	**−36,969 (8.0%)**	−108,352 (23.5%)	−186,204 (40.4%)

Note: Smallest differences by country in bold.

(c) Scenario 3—Small subsequent waves

Countries	Difference (Observed—Projected) n (%)		
	0.1% overlap	1% overlap	5% overlap
France	+45,844 (58.5%)	+33,897 (43.3%)	**+24,591 (31.4%)**
Germany	+61,104 (83.7%)	+56,456 (77.4%)	**+52,640 (72.1%)**
Italy	+45,340 (53.0%)	+36,217 (42.3%)	**+22,129 (25.9%)**
Spain	+7,307 (14.6%)	**−7,098 (14.1%)**	−21,300 (42.4%)
UK	+39,629 (39.7%)	+24,9251 (25.0%)	**+2,689 (2.7%)**
USA	+321,672 (69.7%)	+274,969 (59.6%)	**+219,278 (47.5%)**

Note: Smallest differences by country in bold.

most accurate projections were produced for the shortest distance between waves (5% overlap) (Table 2).

Limitations

One key limitation of the method presented is the assumption of regularly/equally spaced pandemic waves. While the assumption of wave-like transmission patterns was based on historical pandemics, it was made early in the COVID-19 outbreak in the absence of long-term observational data. This limitation reflects the reality that many early modeling assumptions had to be made based on significant uncertainty and limited empirical evidence.

The model is based on three further assumptions: (1) minimal overlap between epidemic waves, (2) no reinfection, and (3) increased transmission during winter months. Our approach initially assumed minimal overlap, particularly when predicting the second wave of the pandemic, due to the limited data available at the time. However, in

principle, the degree of overlap between waves can be estimated more accurately when multiple past waves are available, allowing for a more data-driven approach to forecasting future wave patterns. Nonetheless, we acknowledge that epidemic waves can overlap, reinfections occur, and seasonality effects may vary by region. Future refinements of the model could incorporate reinfection dynamics, more complex wave-overlap patterns, and region-specific seasonal adjustments to improve accuracy.

In response to the urgent need within Amgen for long-term projections during the early stages of the COVID-19 pandemic, this methodology was developed in April 2020. From May 2020, dynamic projections were generated using this approach, which informed cross-functional decision-making. A visualization platform was developed to present updated projections across multiple countries and regions, with data refreshed regularly in alignment with IHME updates. This tool was used to support strategic planning, including manufacturing and supply chain

continuity, clinical trial site management and recruitment prioritization, and office reopening or closure decisions. The model's simplicity and adaptability made it especially valuable during the early phases of the pandemic, when timely insights were essential for operational readiness.

The model assumes a uniform approach across different countries, which does not fully account for the heterogeneity in socio-demographic, cultural, and behavioral factors, as well as variations in PHI measures. This simplification was necessary to ensure scalability and usability in data-limited settings. However, we acknowledge that country-specific adaptations, such as integrating PHI stringency levels or mobility data, could enhance model accuracy in future iterations.

In the examples provided within this paper, projections were produced for six countries. This method is most applicable to regions where public health policies and population characteristics are relatively uniform. For instance, while we produced single projections for the USA as a whole, differences in PHI measures across states suggest that a state-level modeling approach may be more appropriate in capturing localized variations in transmission and mortality trends.

The method was designed for ease of implementation, making it suitable for long-term mortality projections without requiring complex infectious disease modeling. However, this simplicity also presents a limitation, as the model relies on curve-fitting techniques, making it less adaptable to sudden shifts in pandemic behavior. Socio-demographic characteristics (e.g., age, education, ethnicity), social factors (e.g., population mixing, adherence to PHI measures), and PHI measures are implicitly captured via the infection rate or epidemic wave. While this makes the approach scalable, it limits the model's ability to account for sudden changes in transmission drivers.

The model accounts for a gradual reduction in the R_e over time due to increasing population immunity, which manifests to smaller subsequent waves. However, it does not explicitly capture increases in R_e driven by the emergence of more transmissible variants as observed during COVID-19 and in the 1918 influenza pandemic.[19] To incorporate these dynamics, the model includes alternative pandemic scenarios, such as Scenario 2, which reflects a substantially larger second wave potentially driven by relaxed public health interventions, reduced adherence, or variant emergence. In practice, during the COVID-19 pandemic, we adjusted the wave size parameters as new information about emerging variants became available. Although this scenario-based structure provides flexibility, future model enhancements could incorporate explicit parameters for variant-driven changes in transmissibility or virulence to improve long-term forecasting accuracy.

To better understand and model the complexity of the information that determines the infection rate, it is possible to use other types of models that can predict PHI and social mixing impact in the shorter term. For example, IHME was providing its 3-4 month projections based on a combination of curve-fitting and SEIR models.[8] These estimates were particularly useful, especially at the earlier stages of the pandemic, to better estimate the shape, size, and position of future waves.

A key limitation of the model is that it does not explicitly account for the decline in the effective reproduction number (R_t) across successive waves as the susceptible population decreases. In reality, this decline would result in a progressively slower rise in cases and deaths in later waves. While this effect could be incorporated by applying a multiplicative adjustment to the dispersion parameter, ensuring that the transmission dynamics reflect the shrinking susceptible population, this adjustment has not been implemented in the current model. As a result, the model may overestimate the growth rate of subsequent waves, particularly in later stages of the pandemic.

Discussion

In the early days of COVID-19 pandemic, organizational leaders within Amgen requested projections on the long-term course of the pandemic so they could assess its impact on company activities. These activities included, but were not limited to clinical trial operations, manufacturing supply chains, field staff engagement with healthcare professionals, and on-site vs. remote work options. In a similar way to Amgen, the broader pharmaceutical industry and private sector at large also sought projections and forecasts of COVID-19 progression to navigate its impact on various operations. Importantly, there was also a need for projections by public health decision makers so long-term planning could be made on hospital beds and ICU beds, hospital staff, vaccine supplies, etc.

Complex mathematical models were developed to guide policy during COVID-19 pandemic. These models accounted for population structure, social mixing patterns, and adherence to PHI measures among other factors and were typically designed to provide short- to medium-term forecasts up to a maximum of 3-4 months.[5] However, simpler mathematical models have proven valuable in capturing key epidemic dynamics, offering transparency, and allowing broader accessibility for decision-makers.[20] It would not be possible to use such models for long-term projections as they are sensitive to changes in parametric assumptions, which becomes challenging when projecting over long periods of time. In a similar fashion to weather/climate forecasting, in which there are different types of models for short-term forecasting of weather or longer-term modeling of climate,[21] we propose herein a simpler model for longer-term pandemic projections. Although a simpler mathematical model would not fully capture the complexity of a pandemic, it could be useful for producing long-term projections with corresponding uncertainty, as it may be less sensitive to changes in parametric assumptions over time. Compared to traditional infectious disease models such as SIR or SEIR, our curve-fitting approach offers a simpler and more scalable method for long-term projections, particularly in data-limited settings.

Three scenarios were created based on literature from past pandemics. For the six countries included, there were two countries with a scenario of <10% difference between observed and projected deaths (France 7.7%; Italy 7.7%). There were two countries with a scenario of observed and projected differences 10-20% (Spain 14.1%; UK 19.1%) and two countries with a scenario of 20-30% difference (USA 23.2%; Germany 28.2%). The examples highlight the importance of incorporating multiple projection scenarios.

The amount of past information (duration, number, and timing of waves) was an important factor determining the accuracy of projections. Early in the pandemic, early 2020, there was considerable uncertainty surrounding the timing of the next wave. Therefore, a pragmatic approach was used setting a small overlap between waves (1%) based on the belief that a new epidemic wave will occur soon after the PHI measures are withdrawn. For the countries included, the number of projected waves was overestimated, e.g., Germany (Figure 3a and 3b). An alternative exemplar would include projections from a later date, for example, October 2020 (rather than June 2020). In October, the emergence of a second
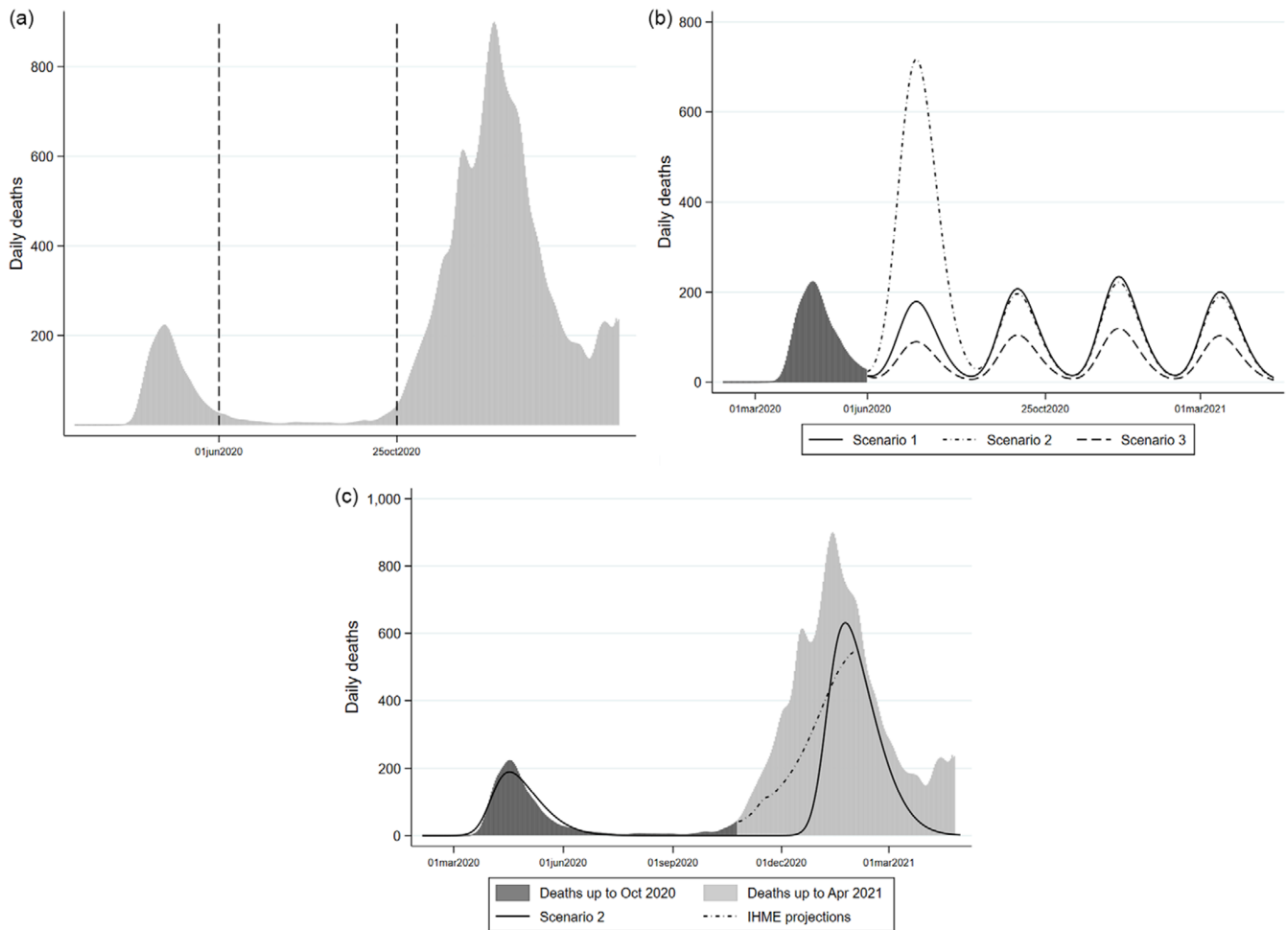
Figure 3. Example of Germany daily deaths.
3a Observed deaths for the period until April 2021.
3b Projected deaths for the period June 2020 to April 2021.
3c Scenario 2 projections for the period October 2020 to April 2021.

wave in Germany removed the need to "predict" the gap between waves, since by then it could be quantified. Instead, the position of the second wave can be estimated based on the observed data. Figure 3c shows the projected wave for Scenario 2 (large second wave) is close to the observed data. This example demonstrates that as time progresses, the model can be improved by including more time-varying information.

A skew-normal distribution was selected to model the distribution of the wave to allow for a steep increase in infection rate due to the highly infectious nature of the virus followed by a steady decrease after the introduction of strict PHI measures.[22] As the aim was to scale across multiple countries, the same distribution was used throughout. In practice, other types of distributions could be considered that could provide improved fit.

The relatively wide UI produced by this approach reflects the inherent challenges of generating long-term projections in a dynamic pandemic environment. These intervals capture multiple layers of uncertainty over time, including changes in PHI measures, public adherence, and the emergence of new variants. Rather than indicating a limitation, the width of the UI is a necessary trade-off for adopting a simple, scalable, and transparent modeling framework. While narrower intervals may be achievable with more complex, data-intensive models, this approach is designed to inform early-

stage scenario planning where data may be limited. In this context, the UI provides a useful tool for contingency planning, enabling decision makers to prepare for a range of potential outcomes.

Different outcomes could potentially be used with this approach such as daily infections, hospital admissions, or occupants of ICU beds. Mortality data were used as they were considered more reliable than reported case numbers for cross-country comparisons and tracking epidemics. This is because case numbers were affected by variations in testing and lack of uniform criteria for confirming infections (particularly at the start of the pandemic).[8] If available, assumptions on case fatality rate or hospitalization fatality rates could be utilized to calculate these outcomes from the mortality results.

As the COVID-19 pandemic evolved, the emergence of vaccines and viral variants significantly influenced transmission dynamics. The impact of these factors was quantified by the anticipated changes in infection rate, which in turn affected the size of future epidemic waves. Vaccination reduces the susceptible population, thereby lowering overall transmission potential, while the effectiveness of vaccines lowers the mortality rate. The emergence of more transmissible or immune-evasive variants can increase the infection rate, potentially leading to larger or more prolonged waves. By adjusting the infection rate parameter within the model,

these evolving factors can be incorporated, allowing for more realistic long-term projection scenarios.

Conclusions

During the COVID-19 pandemic, there was a great need from public health leaders to understand its overall impact, both short- and long-term. During the pandemic, most mathematical models were focused on predicting mortality for time periods up to 3-4 months. The curve-fitting approach presented here was developed during the early months of 2020 to project the long-term impact of COVID-19 on population mortality. The scenarios selected had differing levels of accuracy in different countries, which can be attributed to the different social, political, economic, and public health effects influencing country decisions over PHI put in place during the course of the pandemic. The novel curve-fitting approach described is easy to implement and can be adapted and evolved over time. It could prove to be a valuable decision-making tool during a future pandemic.

Declarations

References

1. **Worobey M.** Dissecting the early COVID-19 cases in Wuhan. *Science.* Dec 3 2021;**374**(6572):1202–1204. doi:10.1126/science.abm4454
2. Archived: WHO timeline—COVID-19. Updated April 27, 2020. Accessed February 23, 2021. https://www.who.int/news/item/27-04-2020-who-timeline—covid-19
3. Prime Minister's statement on coronavirus (COVID-19): March 23, 2020. Accessed March 1, 2021. https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020
4. **Jewell NP, Lewnard JA, Jewell BL.** Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *JAMA.* May 19 2020;**323**(19):1893–1894. doi:10.1001/jama.2020.6585
5. **Friedman J, Liu P, Gakidou E**, IHME COVID-19 Model Comparison Team. Predictive performance of international COVID-19 mortality forecasting models. *medRxiv.* July 14, 2020. https://www.medrxiv.org/content/10.1101/2020.07.13.20151233v5
6. **Fernandez-Martinez JL, Fernandez-Muniz Z, Cernea A, Kloczkowski A.** Predictive mathematical models of the short-term and long-term growth of the COVID-19 pandemic. *Comput Math Methods Med.* 2021;2021:5556433. doi:10.1155/2021/5556433
7. **Tolles J, Luong T.** Modeling epidemics with compartmental models. *JAMA.* Jun 23 2020;**323**(24):2515–2516. doi:10.1001/jama.2020.8420
8. IHME COVID-19 Health Service Utilization Forecasting Team, Murray C. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *medRxiv.* doi:https://doi.org/10.1101/2020.03.27.20043752
9. **Borchering RK, Viboud C, Howerton E**, et al. Modeling of future COVID-19 cases, hospitalizations, and deaths, by vaccination rates and nonpharmaceutical intervention scenarios—United States, April-September 2021. *MMWR Morb Mortal Wkly Rep.* May 14 2021;**70**(19):719–724. doi:10.15585/mmwr.mm7019e3
10. **Friston KJ, Flandin G, Razi A.** Dynamic causal modelling of COVID-19 and its mitigations. *Sci Rep.* Jul 20 2022;**12**(1):12419. doi:10.1038/s41598-022-16799-8
11. **Ferguson NM, Laydon D, Nedjati-Gilani G**, et al. *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand.* 2020. March 16, 2020. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf
12. **Kissler S, Tedijanto C, Lipsitch M, Grad YH.** Social distancing strategies for curbing the COVID-19 epidemic. March 24, 2020. doi:https://www.medrxiv.org/content/10.1101/2020.03.22.20041079v1
13. **Moore K, Lipsitch M, Barry J, Osterholm M.** Part 1: the future of the COVID-19 pandemic: lessons learned from pandemic influenza. *COVID-19: The CIDRAP Viewpoint.* CIDRAP; 2020. April 30, 2020. https://www.cidrap.umn.edu/sites/default/files/public/downloads/cidrap-covid19-viewpoint-part1_0.pdf
14. **IHME.** COVID-19 estimate downloads—Archive. Updated November 04, 2021. http://www.healthdata.org/node/8787
15. **IHME.** COVID-19: our approach. Accessed March 9, 2025. https://www.healthdata.org/research-analysis/diseases-injuries/covid-our-approach
16. **Miller AR, Charepoo S, Yan E**, et al. Reliability of COVID-19 data: an evaluation and reflection. *PLoS One.* 2022;**17**(11):e0251470. doi:10.1371/journal.pone.0251470
17. **Azzalini A.** A class of distributions which includes the normal ones. *Scand J Stat.* 1985;**12**:171–178.
18. **Simonsen L, Viboud C, Chowell G**, et al. The need for interdisciplinary studies of historic pandemics. *Vaccine.* Jul 22 2011;**29** Suppl 2:B1–5. doi:10.1016/j.vaccine.2011.03.094
19. **Castro ESA, Bernardes AT, Barbosa EAG**, et al. Successive pandemic waves with different virulent strains and the effects of vaccination for SARS-CoV-2. *Vaccines (Basel).* Feb 22 2022;**10**(3). doi:10.3390/vaccines10030343
20. **Traulsen A, Gokhale CS, Shah, S, Uecker, H** The Covid-19 pandemic: basic insights from basic mathematical models version 1.0. *NAL-live.* 2022; 2022.3(01000).
21. **Harper L.** What are climate models and how accurate are they? Columbia Climate School. Accessed April 10, 2023. https://news.climate.columbia.edu/2018/05/18/climate-models-accuracy/#:~:text=Essentially%2C%20climate%20models%20are%20an,region%20over%20the%20coming%20decades
22. **Zhuang Z, Zhao S, Lin Q**, et al. Preliminary estimates of the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020. Int J Infect Dis. Jun 2020;**95**:308–310. doi:10.1016/j.ijid.2020.04.044