2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Estimating the Continuously Evolving COVID-19 Case-Fatality Ratio in the United States using a Time-Delay Correcting Algorithm

Brett F. BuSha, *Senior Member*

***Abstract*—The COVID-19 pandemic has placed an extreme healthcare burden across the global community, and new population-based analyses are needed to identify successful mitigation and treatment efforts. The objective of this study was to design a computational algorithm to estimate the time-delay between a peak infection and associated death rate, and to estimate a measurement of the daily case-fatality ratio (D-CFR). Daily infection and death rates from January 22, 2020 through April 15, 2021 for the United States (US) were downloaded from the US Center for Disease Control COVID-19 website. A Savitzky-Golay filter estimated the moving time average of each data sequence with 5 different window-sizes. A locally-designed inflection point identification algorithm with a variable length line-fitting sub-routine identified peak infection and death rates, and quantified the time-delay between a peak infection and subsequent death rate. Although filter window-size did not affect the time-delay calculation (p = 0.99), there was a significant effect of fitting-line length ($p < 0.001$). A significant effect of time-delay length was found among three infection outbreaks ($p < 0.001$), and there was a significant difference between time-delay lengths ($p < 0.01$). A maximum D-CFR of approximately 7% occurred during the first infection outbreak; however, starting approximately 2.5 months after the first peak, a significant negative linear trend ($p < 0.001$) in the D-CFR continued until the end of the analyzed data. In conclusion, this research demonstrated a new method to quantify the time-delay between peak daily COVID-19 infection and death rates, and a new metric to approximate the continuous case-fatality ratio for the ongoing pandemic.**

I. Introduction

Since the first documented case of COVID-19 was identified in Wuhan, China in December 2019 [1], the SARS-CoV-2 virus has spread globally. By April 18, 2021, over 140 million infections and over 3 million deaths had been recorded [2]. Excluding AIDS-related mortality from HIV infections, the COVID-19 pandemic has resulted in the most virus-related deaths since the Spanish flu outbreak of 1918 [3].

Differential equation-based epidemiologic compartmental models that include symptomatic, infected, and recovered populations (SIR), and symptomatic, exposed, infected, and recovered populations (SEIR) have been implemented as attempts to predict the dynamics of the COVID-19 pandemic in the US [4, 5]. Although the SEIR model has an additional parameter to further refine the model, the added model complexity does not always provide an improved predictive ability [6]. Additionally, the a-priori predictive value of SIR and SEIR models is not conclusive; a review of the predictive performance of COVID-19 models found that approximately 30% of 242 published models over estimated disease induced fatality rates [7], and another investigation determined that following a new post-hoc analysis of the accuracy of COVID-19 models, over 35% of short-term predictions and 25% of longer-term predictions had error rates greater than 50% [8].

Since the SARS-CoV-2 is a novel virus in the human population, accurate estimates of many the common epidemiologic model parameters are unavailable, adding further variability to epidemiological estimates of the COVID-19 pandemic [9]. Also, there remains an unknown distribution of symptomatic versus asymptomatic, yet still contagious, cases of COVID-19. An early study that monitored individuals on a cruise line docked in Japan calculated that approximately 18% of infected individuals were asymptomatic [10]; while another study that tracked individuals admitted to a hospital found that 43% of their infected sample population was asymptomatic [11]. A modeling study that focused on cities in the US indicated that asymptomatic cases may outnumber reported cases [12]. Without a clearer understanding of the magnitude of asymptomatic population, the transmission potential of this group and the overall infection rates of the pandemic remain uncertain.

An alternative method to quantify the state of the ongoing COVID-19 pandemic, other than focusing on estimated infection rates, is to use measurements of disease associated fatalities. A common epidemiologic metric used to estimate the lethality of disease is the case fatality ratio (CFR); the number of deaths that resulted from infection divided by the number of infected individuals over a discrete time period. However, since during an ongoing epidemic or pandemic the total number of infection or infection-related deaths has not been achieved, CFR measurements are more easily biased relative to other predictive model estimates of disease progression [13]. Also, ongoing mitigation efforts, such as lock-downs, and treatments, such as new vaccines, are being introduced, which further effect the rate of infection-related deaths.

Although there are many tools available to quantify and predict the progression of the current pandemic, there is an urgent need for new analysis techniques to enhance the estimation of the disease progression and the impact of active treatment and mitigation efforts. The objective of this study was to design a computational algorithm to estimate the time-delay between the peak infection and death rates associated with waves, or outbreaks of infections, and to provide a daily estimate of the COVID-19 case-fatality ratio.

B. F. BuSha is with the Department of Biomedical Engineering, The College of New Jersey, Ewing, NJ, 08628 USA (e-mail: busha@tcnj.edu).

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

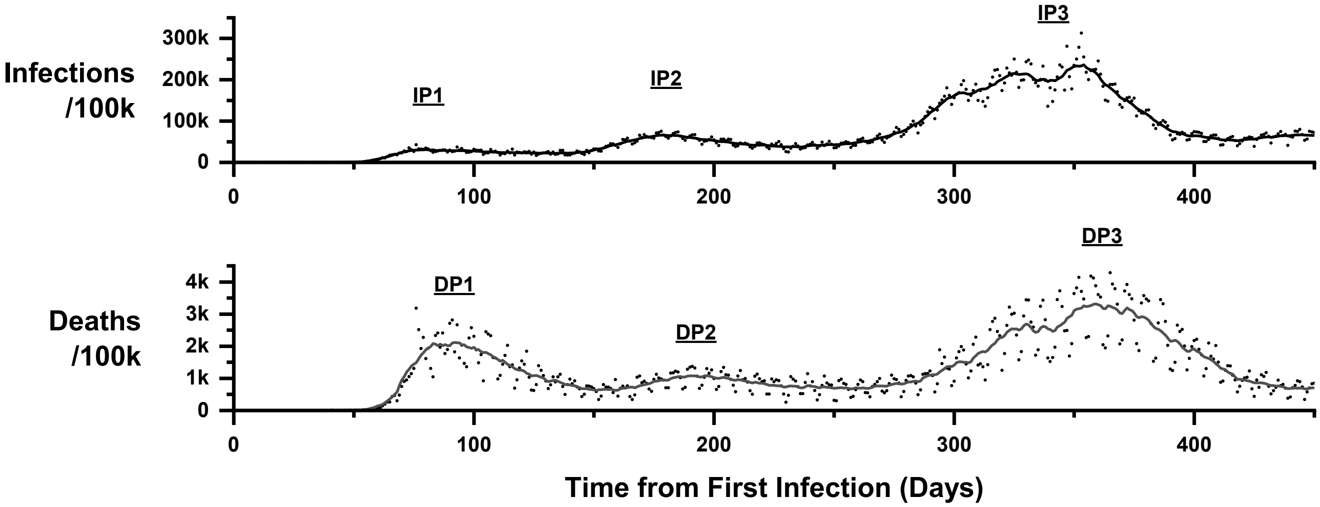

Figure 1. Daily Covid-19 infection and death rates, starting on January 22, 2020 and ending on April 15th, 2021. Individual marks represent daily counts, and the solid lines are the same data filtered with a 15-day window. The approximate location of the first, second, and third peak daily peak infection (IP1, IP2, IP3) and death (DP1, DP2, DP3) rates are identified.

II. Methods

A. Data Sourcing

Daily infection and death data from each of the fifty states within the United States and the District of Columbia, starting on January 22, 2020 and ending on April 15, 2021, were downloaded from the United Stated Center for Disease Control COVID-19 website on April 16, 2021 (https://covid.cdc.gov/covid-data-tracker/#datatracker-home).

B. Filtering and Identification of Peak Daily Infection and Death Rates

A Savitzky-Golay finite impulse response filter was implemented in MATLAB and used to compute the moving averages of the daily infection and death rate sequences without imposing a phase-delay. The polynomial order of the filter was set to one, and the moving averages were calculated with window-sizes of 7, 15, 21, 29, and 35.

The peak daily infection and death rates were identified from each of the five filtered sequences using a modified version of a locally designed inflection point identification algorithm that was implemented in MATLAB [14]. Briefly, an iterative linear regression algorithm calculated the sum of squares error (SSE) for each of five different length lines (57, 71, 85, 99, and 113 days) fit to each of the filtered daily infection and death sequence. Beginning with the fitting line aligned to the first day of a filtered sequence, the SSE was calculated. Next, the fitting-line was shifted forward by one day, and the SSE was recalculated. This process was repeated until the fitting line was shifted to the end of the sequence. In order to match the maximal SSE values to corresponding inflection points in the infection or death data, each SSE value was temporally aligned to the center point of the fitting line. This procedure was repeated across all the differently filtered data sequences. Since some of the fitting-line lengths were greater than the time between the start of the daily infection and death sequences, it was necessary to pad the beginning of the data with zeroes.

For each combination of the 5 filter window-sizes and 5 fitting line-lengths, a time-delay was calculated that described the difference between the peak daily infection and death rates, which resulted in 25 time delay measurements for each infection outbreak.

C. Time-Delay Correction for the estimated Daily Case-Fatality Ratio

In order to calculate the daily case-fatality ratio (D-CFR) relative to the first day the infection was diagnosed, a temporally shifted sequence of daily death rates was created. The average time-delay between the daily peak infection and death rates for each of the three outbreaks were used to time shift the daily death values into alignment with the corresponding dates of the initial infections. The daily time-delay values were linearly transformed between the first and second, and the second and third outbreaks to account for the differences in the average time-delays assigned to each outbreak. Since no time-delay could be estimated prior to the first outbreak, the first infection day (January 22, 2020) was assigned a time shift of zero, and was linearly increased on a day-to-day basis to match the average time-delay at the first outbreak. The time-delay shift remained constant from the third outbreak peak to the end of the data.

The daily case-fatality ratio (D-CFR) was calculated as the time-delay adjusted daily death rates divided by the corresponding daily infection rates. The D-CFR sequence started on the date of the first recorded death minus the estimated time-delay of the first outbreak.

D. Linear Trend Estimate of the Daily Case-Fatality Ratio Estimate

A linear regression was calculated for the final 290 points of the five D-CFR estimates. Each linear fit was generated with the daily infection and death rate data that was filtered with the same window-size.

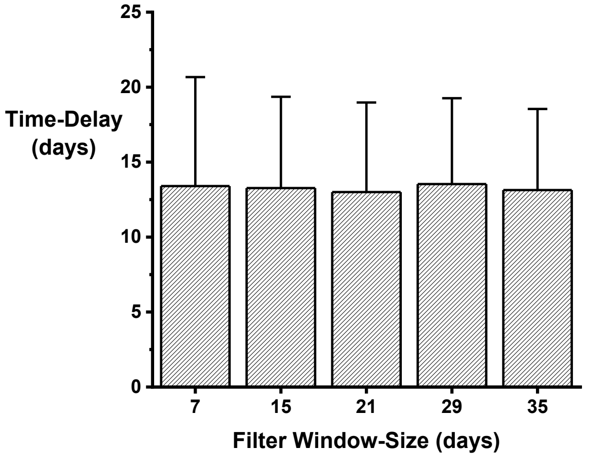

Figure 2. There was no effect of filter window-size on the average time-delay between consecutive peak daily infection and death rates.

E. Statistical analyses

The effects of the filter window-length, fitting-line length of the inflection point identification algorithm, and the three COVID-19 infection outbreaks in the US on the time-delay between the daily peak infection and consecutive peak death rate was evaluated with a three-way ANOVA. Where appropriate, multiple comparisons were assessed using Fisher's LSD. Statistical analyses were performed with OriginPro 2021 (OriginLab). Where appropriate, results are presented as the average plus/minus the standard deviation.

III. Results

The raw and filtered (15-day window size) daily infection and death rates of the US population, beginning on January 22, 2020 and ending on April 15, 2021 and normalized per 100k individuals, are illustrated in Fig. 1. Additionally, the first, second, and third peak infection outbreaks and consecutive peak death rates are identified in Fig 1.

A. Estimation of the Time-Delay between Peak Daily Infection and Death Rates

There was no effect of the filter window-size on the time-delay length between the peak infection and associated peak death rates, with average values ranging from 13.0 to 13.5 days (p = 0.99), illustrated in Fig. 2. However, there was a significant effect of the inflection fitting-line length on the time-delay, with average values ranging from 10 to 16 days ($p < 0.001$). There were significant differences between many of the time-delays associated with the fitting-line lengths, as shown in Fig. 3.

There was a significant effect of the outbreak number on the time-delay length ($p < 0.001$), and there were significant differences between all three time-delays ($p < 0.006$). The time-delays between identified daily peak infection and death rates of the three outbreaks were: 7 ± 1 day; 15 ± 2 days; and 18 ± 6 days, illustrated in Fig. 4.

B. Estimation of the Daily Case-Fatality Ratio

Accounting for the corresponding infection-to-death time-delays for each of the three outbreaks, an estimate of the continuous D-CFR is presented in Fig. 5. During the first months of the COVID-19 pandemic in the United States, the

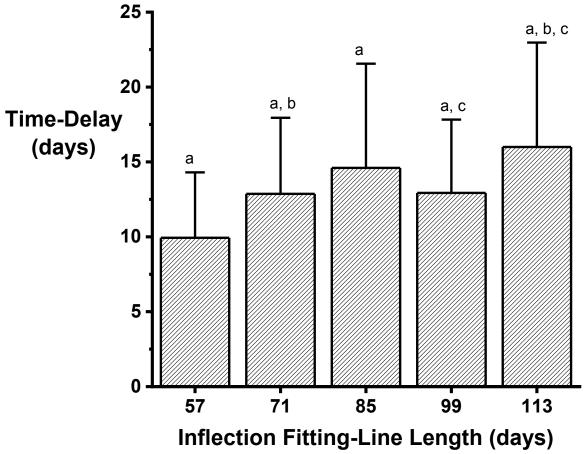

Figure 3. There was a significant effect of the fitting-line length used by the inflection point algorithm on the average time-delay ($p < 0.001$). There was a significant difference between the inflection-line lengths ($p < 0.02$); significant differences between lengths identified with 'a', 'b', and 'c'.

novelty of the pathogen and resulting lack of medical treatment options resulted in a relatively rapid increase and decrease in the D-CFR. The largest percentage of fatalities resulting from COVID-19 infection occurred during the first outbreak, at approximately 80 days after the first recorded US-based infection (January 22, 2020).

C. Linear Regression Analysis of Daily CFR Estimates

After excluding the D-CFR data associated with the first infection peak, a significant negative linear relationship was identified for the last 290 days of the D-CFR estimates calculated with the unfiltered or any of the paired filtered daily infection and death rate data ($p < 0.001$ for all trends), individual results for the slope, y-intercept, and p-value are presented in Table 1.

IV. Discussion

Following the first COVID-19 outbreak in the US in early April 2020, which had an estimated time delay of 7 days between the peak daily infection and death rates, there were

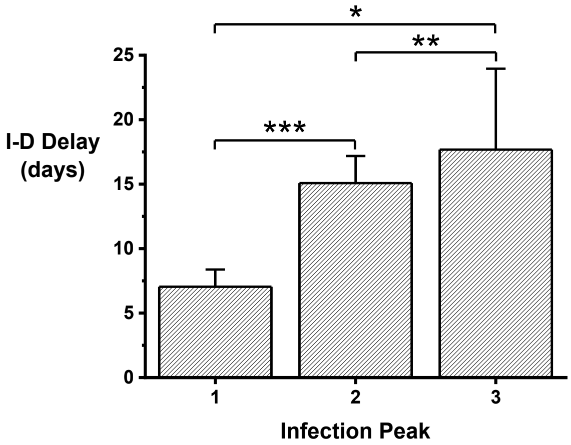

Figure 4. There was a significant effect of the outbreak number on time-delay length (p < 0.001), and there were significant differences between the all three time-delays (*, $p < 0.001$; **, $p = 0.006$; ***, $p < 0.001$).

TABLE I.

Filter Window-Length	Slope	Y-Intercept	*p* value
n.a. (raw data)	-1.46E-03	1.83	< 0.001
7	-1.57E-03	1.80	< 0.001
15	-1.66E-03	1.81	< 0.001
21	-1.66E-03	1.81	< 0.001
29	-1.71E-03	1.82	< 0.001
35	-1.73E-03	1.82	< 0.001

significant increases in the time-delay between the peak infection and death rates across the two successive outbreaks. Although during the first outbreak the estimated D-CFR approached a maximum of 7%, the D-CFR estimate quickly reduced to below 2% and maintained a significant negative linear decrease through April 2021. Additionally, although there was a relatively large increase in the D-CFR estimate during the first outbreak in April 2020, subsequent outbreaks in July 2020 and January 2021 did not produce similar increases in the D-CFR.

Since in the beginning of the pandemic the SARS-CoV-2 was a novel virus to the population, there were no known effective treatments. Nonpharmaceutical interventions (NPIs), such as lockdowns and facemask wearing, were the only methods known to control the pandemic. Although some research suggests that the US population may not have adhered to the government issued NPI recommendations as well as other countries [13], these data from this study suggests that there was sufficient participation to significantly decrease COVID-19 mortality, prior to the emergency approval of any pharmaceutical treatments.

There was relatively little variability in the estimate of the time-delay between peak daily infection and death rates as a function of filter window-length; however, there was a significant effect of the fitting-line length of the inflection point identification algorithm. Depending on the filter window-length, not all peak infection and consecutive death rates had definitive single peaks, an example is IP3 in Fig. 1. Although maintaining higher filter window-lengths and fitting-line lengths would increase the likelihood of identifying individual peaks, there would also be a higher likelihood of an adjacent, but independent peak, influencing the time-delay calculation. This may explain the relatively lower time-delay value associated with the shortest inflection fitting-line length.

V. Conclusion

These data suggest that infection mitigation efforts, such as social distancing and mask wearing, and treatment modalities, including antibody therapies and vaccines, implemented in the US following the first outbreak in April 2020 provided a beneficial impact on the local pandemic as early as the spring of 2020, and continued to show positive benefits through the early spring of 2021.

VI. References

[1] H. Nishiura, S.-m. Jung, N. M. Linton, R. Kinoshita, Y. Yang, K. Hayashi, T. Kobayashi, B. Yuan and A. R. Akhmetzhanov. "The extent of transmission of novel coronavirus in Wuhan, China, 2020," *J. Clin. Med.*, vol. 9 (2), 330, https://doi.org/10.3390/jcm9020330, 2020.

[2] World Health Organization (WHO), COVID-19 Weekly Epidemiological Update - 18 April 2021, WHO, 2021.

[3] J. Feehan and V. Apostolopoulos. "Is COVID-19 the worst pandemic?" *Maturitas*, https://doi.org/10.1016/j.maturitas.2021.02.001, 2021.

[4] M. Liu, R. Thomadsen and S. Yao. "Forecasting the spread of COVID-19 under different reopening strategies," *Sci. Rep.*, vol. 10, 20367, https://doi.org/10.1038/s41598-020-77292-8, 2020.

[5] R. C. Reiner, Jr., et al. "Modeling COVID-19 scenarios for the United States," *Nat. Med.*, vol. 27, pp. 94-105, 2021.

[6] W. C. Roda, M. B. Varughese, D. Han and M. Y. Li. "Why is it difficult to accurately predict the COVID-19 epidemic?," *Infect. Dis Model.*, vol. 6, pp. 258-272, 2021.

[7] J. E. Gnavni, K. V. Salako, G. B. Kotanmi and R. G. Kakaï. "On the reliability of predictions on COVID-19 dynamics: A systematic and critical review of modelling techniques," *Infect. Dis. Model*, vol. 6, pp. 258-272, 2021.

[8] J. Guan, Y. Wei, Y. Zhao and F. Chen. "Modeling the transmission dynamics of COVID-19 epidemic: a systematic review," *J. Biomed. Res.*, vol. 34(6), pp. 422-430, 2020.

[9] I. Holmdahl and C. Buckee. "Wrong but useful — What COVID-19 epidemiologic models can and cannot tell us," *N. Engl. J. Med.*, vol. 383 (4), pp. 303-305, 2020.

[10] K. Mizumoto, K. Kagaya, A Zarebski and G Chowell. "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020," *Euro. Surveill.*, vol. 25(10), https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180, 2020.

[11] N. V. V. Chau, et al. "The natural history and transmission potential of asymptomatic severe acute respiratory syndrome coronavirus 2 infection," *Clin. Infect. Dis.*, vol. 71, pp. 2679-2687, 2020.

[12] H. M. Dobrovolny. "Modeling the role of asymptomatics in infection spread with application to SARS-Cov-2," *PLoS ONE*, 15(8): e0236976. https://doi.org/10.1371/journal.pone.0236976, 2020.

[13] Garske T, Legrand J, Donnelly CA, et al. "Assessing the severity of the novel influenza A/H1N1 pandemic," *B.M.J.,* vol. 339, doi: https://doi.org/10.1136/bmj.b2840, 2009.

[14] B. F. BuSha and M. H. Stella, "State and chemical drive modulate respiratory variability," *J. Appl. Physiol.*, vol. 93, pp. 685-696, 2002.

[15] J. Margraf, J. Brailovskaia and S. Schneider. "Adherence to behavioral Covid-19 mitigation measures strongly predicts mortality," *PLoS ONE*, vol. 16(3): e0249392, https://doi.org/10.1371/journal.pone.0249392. 2021.

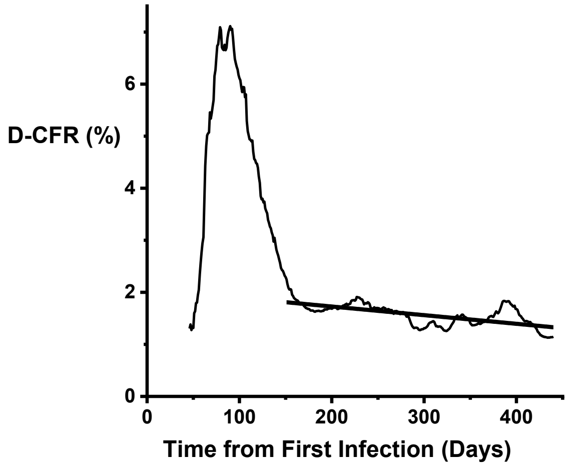

Figure 5. The daily case-fatality ratio (D-CFR) of the Covid-19 pandemic in the United States, using the filtered (21-day window-length) infection and time-delay corrected death data. In order to exclude the effect of the first outbreak, the linear fit was applied to the last 290 days of the data.