AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Application Notes

An interactive tool to forecast US hospital needs in the coronavirus 2019 pandemic

Kenneth J. Locey [iD],[1] Thomas A. Webb,[1] Jawad Khan,[2] Anuja K. Antony,[3] and Bala Hota[1,2,4]

[1]Center for Quality, Safety and Value Analytics, Rush University Medical Center, Chicago, Illinois, USA, [2]Knowledge Management Services, Rush University Medical Center, Chicago, Illinois, USA, [3]Division of Plastic and Reconstructive Surgery, Department of Surgery, Rush University Medical Center, Chicago, Illinois, USA and [4]Division of Infectious Diseases, Department of Internal Medicine, Rush Medical College, Chicago, Illinois, USA

Corresponding Author: Kenneth J. Locey, PhD, Center for Quality, Safety and Value Analytics, Rush University Medical Center, 600 S. Paulina St., Suite 544, Chicago, IL 60612, USA; kenneth_j_locey@rush.edu

ABSTRACT

Objective: We developed an application (https://rush-covid19.herokuapp.com/) to aid US hospitals in planning their response to the ongoing Coronavirus Disease 2019 (COVID-19) pandemic.

Materials and Methods: Our application forecasts hospital visits, admits, discharges, and needs for hospital beds, ventilators, and personal protective equipment by coupling COVID-19 predictions to models of time lags, patient carry-over, and length-of-stay. Users can choose from 7 COVID-19 models, customize 23 parameters, examine trends in testing and hospitalization, and download forecast data.

Results: Our application accurately predicts the spread of COVID-19 across states and territories. Its hospital-level forecasts are in continuous use by our home institution and others.

Discussion: Our application is versatile, easy-to-use, and can help hospitals plan their response to the changing dynamics of COVID-19, while providing a platform for deeper study.

Conclusion: Empowering healthcare responses to COVID-19 is as crucial as understanding the epidemiology of the disease. Our application will continue to evolve to meet this need.

Key words: COVID-19, epidemiology, healthcare, forecasting, modeling, SEIR

LAY SUMMARY

Hospitals have been continually faced with anticipating the resurgent spread of COVID-19 and its effects on visits, admissions, bed needs, and crucial supplies. However, few open-source tools are available to aid hospitals in planning. We developed a web application (https://rush-covid19.herokuapp.com/) for US states and territories to predict the spread of COVID-19 and to provide forecasts for hospital visits, admissions, discharges and to anticipate needs for intensive care unit (ICU) and non-ICU beds, ventilators, and personal protective equipment. Users can choose from a suite of models to predict the spread of COVID-19, some of which explain >99% of variation in COVID-19 cases within states. Users can modify a large set of inputs to obtain forecasts for their institution, examine variability in forecasts over time, download forecast data for further analysis, and explore trends in hospitalization and testing. We designed our application to be interactive, insightful, and easy to use for hospital leaders, healthcare workers, and government officials. However, specialists can use our models, open-source code, and aggregated data for deeper study. As the dynamics of COVID-19 change, our application will also change to meet emerging needs of the healthcare community.

INTRODUCTION

Healthcare enterprises are grappling with the challenges of preparing for the resurgence of Coronavirus Disease 2019 (COVID-19) and appropriation of resources needed to treat patients while protecting healthcare workers. We developed an open-source application that generates hospital-specific forecasts of COVID-19 patient needs and related supplies. Our application (https://rush-covid19.herokuapp.com/) allows users to employ a suite of models to predict the spread of COVID-19 and to generate forecasts for COVID-19 visits, admits, discharges, and needs for intensive care unit (ICU) and non-ICU beds, ventilators, and personal protective equipment (PPE). Users can also examine trends in hospitalization and testing.

We describe our application in detail, focusing on the data and models it uses, interactive inputs, and provided forecasts. We explain the use of our application and the caveats of our approach and discuss its value in aiding the needs of the healthcare community and its potential for generating novel insights. Our efforts were primarily aimed at addressing urgent healthcare demands and enabling informed decision-making. Our application uses open-source software and it source code and aggregated data are publicly available.

DATA, MODELING, AND USE

Application development

We developed our application in response to the needs of our home institution—Rush University System for Health—in anticipation of COVID-19 cases and subsequent surges. A COVID-19 command center staffed by senior hospital administrators served as a real-time focus group for initial requirements and choice of user-defined variables. Feedback from the Illinois Hospital Association hospitals aided further refinement.

Data

Our application aggregates reports of cumulative cases across US states and territories from the Johns Hopkins University Center for Systems Science and Engineering repository,[1] state and territory population sizes based on US Census Bureau data (2010–2019), dates of COVID-19 arrival from state and territory health agencies, and testing and hospitalization levels from The Atlantic's COVID Tracking Project.[2]

Forecasting cumulative COVID-19 cases

We developed 7 models to forecast COVID-19 case volumes up to 60 days past the present day.

Exponential growth

Initial spread of infection may be limited primarily by inherent growth rate (r), which proceeds multiplicatively according to the function, $N_t = N_0 e^{rt}$. Here, N_0 is the initial infected population size, t is the change in time, and N_t is the infected population size at t. This model was widely used to characterize the spread of COVID-19 during initial weeks of infection.[3–5] Our application uses a regression on the log-linear transformation, $\log(N_t) = \log(N_0) + t \cdot r$, to predict the expected number (N) of cumulative cases.

Quadratic growth

Initial growth may be more rapid than expected from the exponential model and also characterized by a constant change in growth rate (Figure 1). In this case, growth may be quadratic. Early

COVID-19 studies implicated quadratic growth of COVID-19.[6,7] The quadratic function is a second-order polynomial that applies to population growth as $N_t = \beta_1 t^2 + \beta_2 t + N_0$. Our application predicts values for (N) using numerical optimization of β_1 and β_2. Like the exponential model, this model does not account for eventual leveling-off.

Logistic growth

The spread of a contagious disease must eventually slow as limitations are encountered (eg, immunity). The logistic model captures this dynamic and produces an *s*-shaped curve[8,9] (Figure 1). Early COVID-19 studies implicated logistic growth in the spread of the disease.[11,12] The logistic model takes a simple functional form, $N_t = \alpha/1 + e^{-rt}$, where α is the upper limit of N and r is the intrinsic rate of increase. Our application uses numerical optimization of α and r to find the best fit logistic function and predicted values for N_t.

Gaussian growth

The Gaussian (ie, normal) distribution can provide a fast approximation to some epidemiological models.[13] The Gaussian distribution has 2 parameters, mean (μ), standard deviation (σ) (Figure 1). When used to model the spread of disease, Gaussian curves are symmetrical around a climax day. Gaussian models have successful in approximating the spread of COVID-19.[14] Our application finds the best fit cumulative Gaussian function via numerical optimization of μ and σ.

SEIR-SD

COVID-19 modeling has often used refinements to the SEIR model,[4,15–18] which accounts for changes in fractions of susceptible persons (S), persons exposed but not exhibiting infection (E), infectious persons (I), and those recovered (R), where $S + E + I + R = 1$. These subpopulations are modeled in a set of differential equations:

$$\frac{dS}{dt} = \beta SI, \quad \frac{dE}{dt} = \beta SI - \alpha E, \quad \frac{dI}{dt} = \alpha E - \gamma I, \quad \frac{dR}{dt} = \gamma I.$$

Here, α is the inverse of the latent period, γ is the inverse of the mean infectious period, and β is the mean number of effective contacts (ie, contacts resulting in secondary infection). Our application imputes the initial value of β from a relationship between γ and the basic reproductive number (R_0), that is, $\beta = \gamma R_0$ (Table 1).[26–28] It then allows β to decrease in proportion to I, from which an implicit frequency-dependent degree social distancing emerges. Our model also simulates social distancing (λ) as an explicit response to public policy by including a time-iterative modification to β, that is, $\beta_{t+1} = \beta_t/(\lambda I + 1)$, where β remains unchanged when I or λ equal 0. When λ equals 1, the daily change in β is governed by the effect of I. The product of I and λ then determines the daily change in β: $\lambda I = (\beta_t - \beta_{t+1})/\beta_{t+1}$. Finally, our model captures lags in testing by allowing the apparent size of I to lag behind the actual magnitude via use of a logistic function: $I_t/1 + e^{-f(t)}$, where $f(t) \sim \tau t$.

Our application performs 200 000 iterations on combinations of SEIR parameter values, within reported ranges (Table 1). It then chooses the set of parameters that maximize the explained variation in observed data, which avoids the challenges of applying numerical optimizers to complex models.

Two-phase logistic growth

Many US states have experienced a resurgence in COVID-19, leading to increases that simple models and classic epidemiological mod-
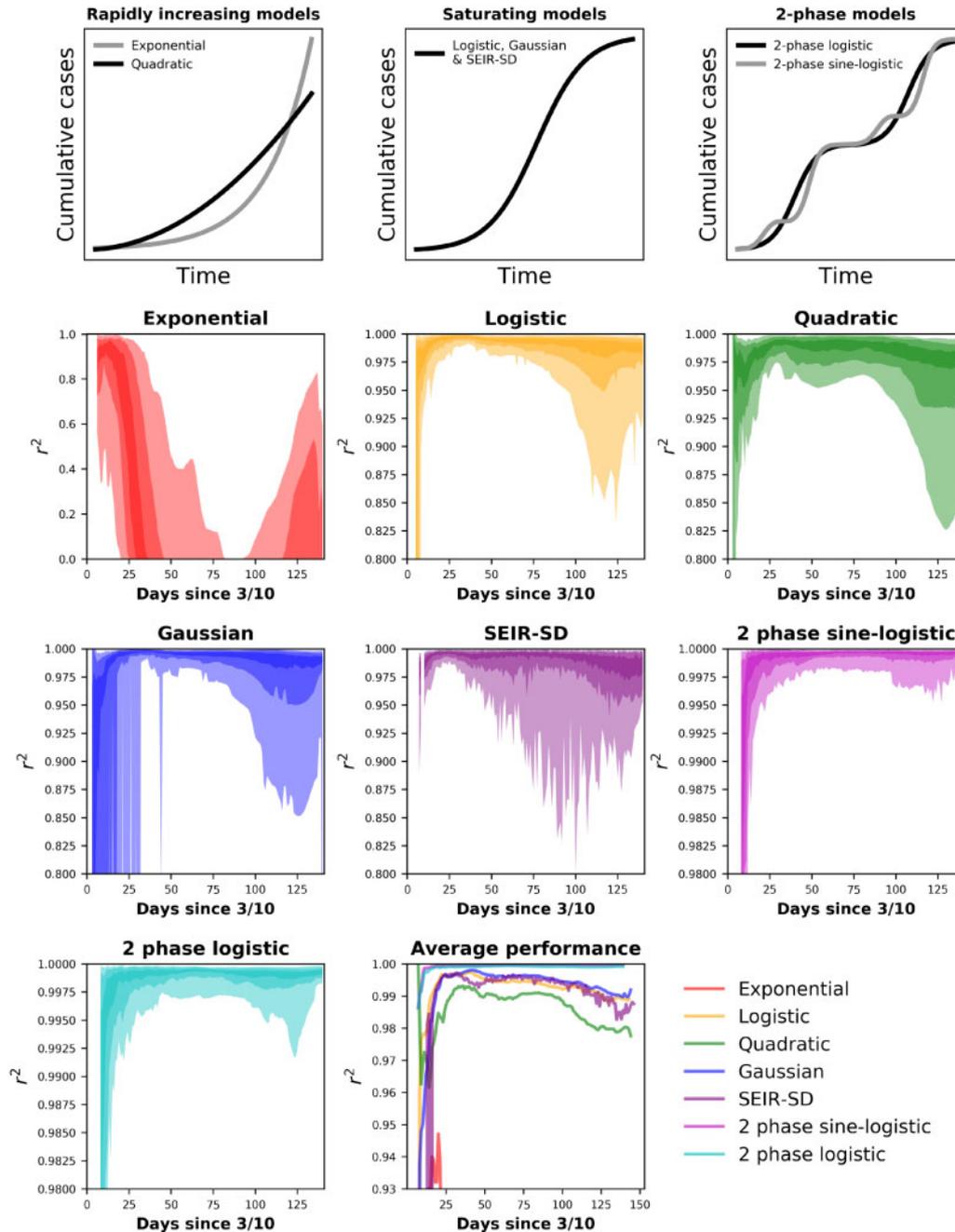
Figure 1. Top row: General forms of our application's models. Rows 2–4: 95, 75, and 65% confidence interval (CI) hulls for the performance of models across US states and territories from March 16 to August 1. The Johns Hopkins University Center for Systems Science and Engineering data shows most states beginning to report COVID-19 cases by March 10 and we allowed the models a week of data before fitting them. The x-axis is in days since the first reported case, occurring on or after March 10. Lightly colored hulls are 95% CI, hulls of intermediate darkness are 75%, and the darkest hulls are 65% CI. Performance is measured via a modified r^2 of observed versus predicted values where the y-intercept is forced through 0.[10] The y-axes are scaled to show the greatest resolution for each model and are not scaled similarly across all models. Lower right: Mean r^2 values across states for all models.

els fail to capture. To capture resurgence, we expanded the logistic model to include 2 phases. Using the functional form of the logistic model, our application iteratively searches across time series to find the point where 2 phases of logistic growth best explain the variation in cumulative cases.

Two-phase sine-logistic growth
An extension of the two-phase logistic, this model assumes periodic fluctuations in the increase of cumulative cases by including sine-

wave dynamics to the logistic equation: $N_t = \alpha/1 + e^{-rf(t)}$, where $f(t) \sim t + \text{sine}(t)$. As with the two-phase logistic model, our application searches across the time series to find the point where 2 phases of growth best explain variation in observed data.

Forecasting hospital visits and admits via time lags
Users can obtain interactive forecasts for hospital visits and admits by entering expected daily values for percent of state-wide COVID-19 cases visiting their hospital, percent of those admitted, and the

Table 1. Parameters for the SEIR-SD model

Parameter	Description	Reported range	References
Average latent period	Days during which an exposed individual cannot infect others	5–6 days	[18–22]
Average infectious period	Days over which a person remains infective	1.6–13 days	[18,23–25]
Basic reproductive number (R_0)	Average number of secondary infections generated by a primary infection in a totally susceptible population	2.1–6.5	[4,18,20,23,24]

Note: Our application attempts to optimize the following SEIR-SD parameters within ranges of reported values for the average latent period, the average infectious period, and the basic reproductive number.
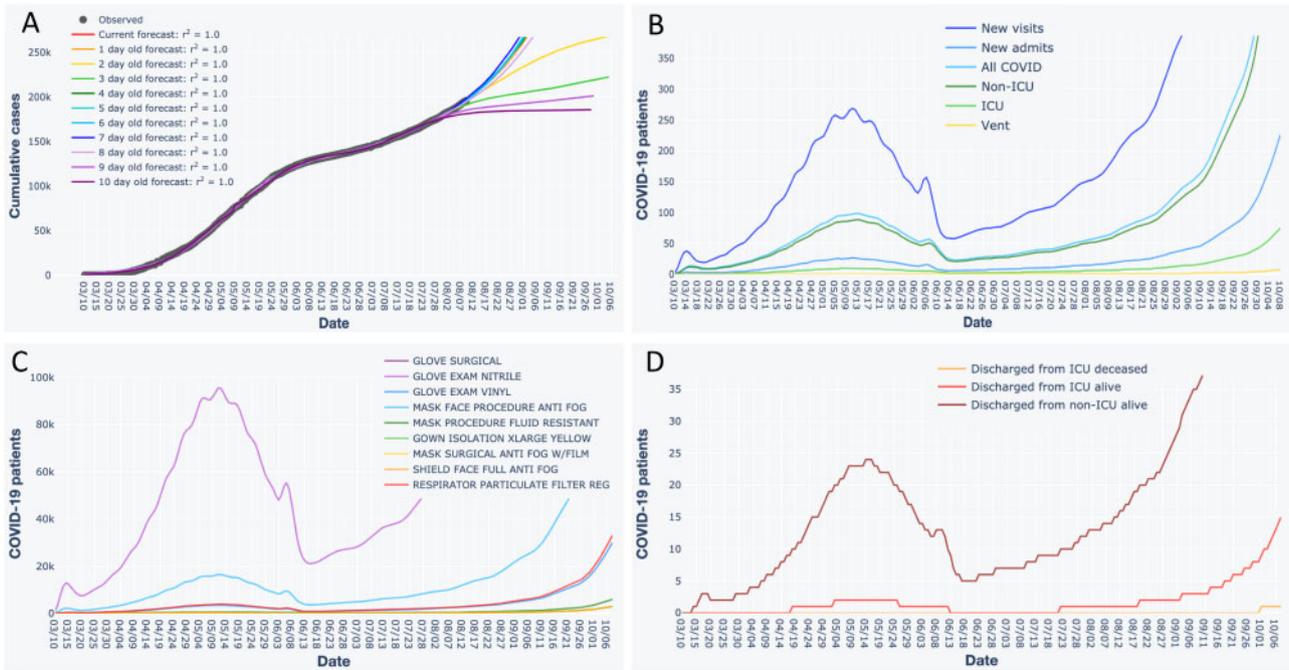
Figure 2. Interactive forecast plots generated by our application, using fits of the two-phase sine-logistic model to Illinois data as an example. (A) Forecast for cumulative COVID-19 cases. Black dots are observed data. Colored lines are forecasts. Values for coefficients of determination (r^2) pertain to observed versus predicted values, where the *y*-intercept is forced through 0.[10] (B) Forecasted patient census. (C) Forecasted personal protective equipment supply needs. (D) Forecasted discharges. When using the application, users can pan across, zoom in, select and de-select data, hover over data points and lines for additional information, and download plots as a png files. The application also presents information in B–D as tables, which users can download as csv files.

number of transfer patients accepted (Figure 2). We accounted for the tendency of infected persons to not seek immediate attention by modeling time lags as Poisson random variables, whereby newly infected persons wait 0, 1, etc., days to visit the hospital. The Poisson model is a discrete probability distribution where the mean corresponds to the expected average time lag. This approach avoided overburdening users by not requiring parameter values for which data may not be readily available.

Forecasting hospital bed, ICU needs, and discharges via daily carry-over

Users can obtain interactive forecasts for hospital bed, ICU needs, and discharges by entering parameter values for the number of ICU and non-ICU beds in house and choose expected values for the percent of COVID-19 patients admitted to ICU, average length of stay (LOS) for ICU and non-ICU patients, percent of ICU patients on ventilators, and ICU mortality (Figure 2). These inputs drive calculations for newly admitted ICU, non-ICU, and ICU-ventilator patients expected each day, as well as numbers of discharges. Our application models the daily patient carry-over using a binomial model, a

probability distribution for binary outcomes (eg, patients leave or stay for an additional day) that requires 2 parameters (*p*, *n*). We set the value of *p* to 0.5 and set the value of *n* to be twice the LOS, which produces a probability mass function (pmf) with a mean equal to the LOS. This pmf is converted to a cumulative distribution function to obtain estimates for the fraction of 1-day, 2-day, etc., patients discharged on the present day. Patients not discharged are iteratively carried over to the following days. Like the Poisson, the binomial does not overburden the user with parameters based on data they may not readily have.

Forecasting supply needs

Users can obtain forecasts for PPE needs after entering expected per patient daily values for surgical gloves, nitrile exam gloves, vinyl exam gloves, anti-fog procedural face masks, fluid resistant procedural face masks, extra-large yellow isolation gowns, anti-fog surgical face masks, anti-fog full face shields, and particulate filter respirators (Figure 2). Multiplying the expected PPE values by their respective patient type across the forecasted census produces PPE forecasts.
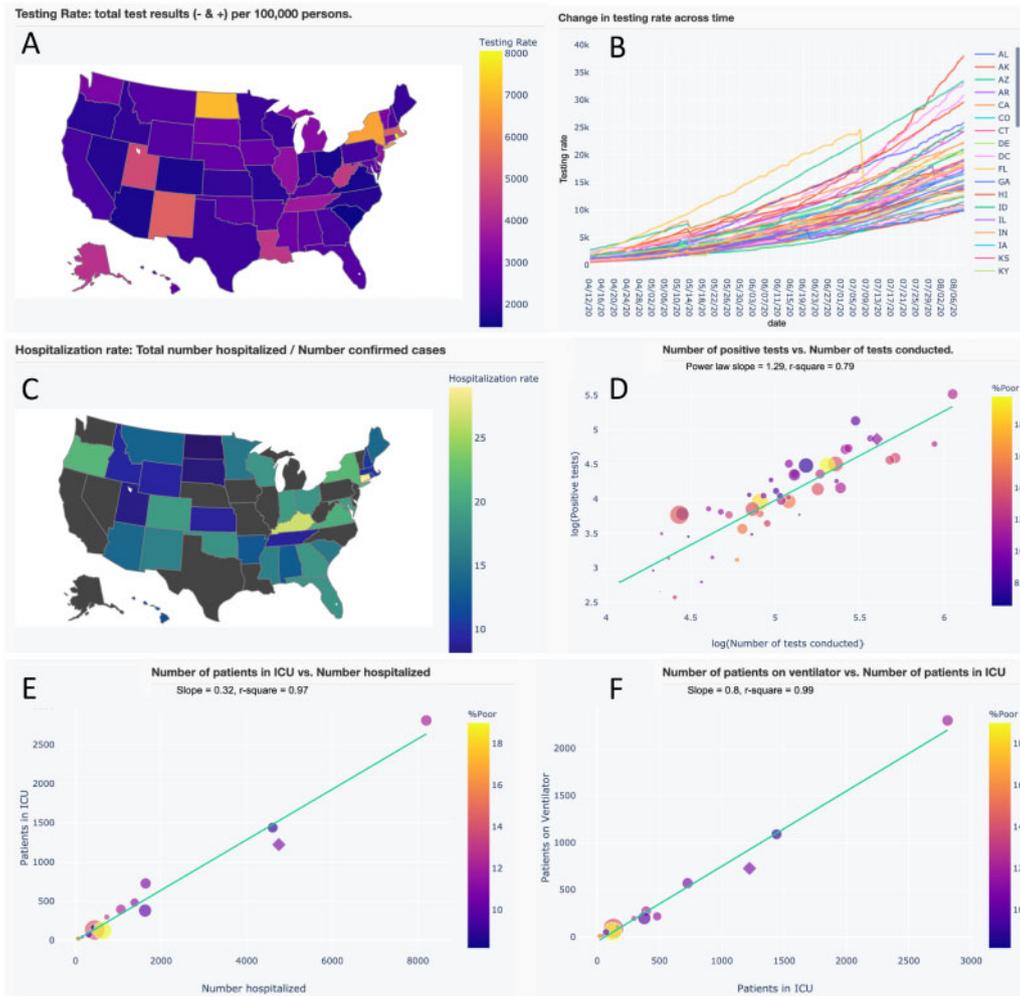
Figure 3. A subset of interactive maps and statistical relationships from our application's Trends in Hospitalization and Trend in Testing tabs. Users can pan across, zoom in, select and de-select data, and hover over data for additional information. (A) Map of testing rate (tests per 100 000 persons). (B) Time series of testing rates across US states and territories. (C) Map of hospitalization rate (total number of persons hospitalized/no. of confirmed cases). (D) A strong power-law ($y \sim x^z$) relationship between positive test results and number of tests conducted, where the slope (z) indicates that greater testing reveals disproportionately greater positive results. (E and F) Strong linear relationships between the number of patients in intensive care unit (ICU) versus the number hospitalized and the number of patients on ventilators versus those in ICU. These latter relationships may be used to aid in parameterizing forecast variables.

Downloadable forecast data

Users can download data pertaining to each forecast. These csv files are dynamically updated upon any changes to their associated graphs or tables.

Source code and Data

We developed our application using open-source software (python 3.7, plotly dash) and the Heroku web-hosting service. Source code, data, and README file for our application and for reproducing analyses herein are available at https://github.com/Rush-Quality-Analytics/SupplyDemand and via the Dryad Digital Repository at https://doi.org/10.5061/dryad.1ns1rn8rx.[29]

RESULTS

COVID-19 predictions

The exponential model explained >90% of variation in cases during initial weeks of infection but quickly began to fail (Figure 1). Its performance then increased with the onset of resurgence. The logistic,

quadratic, Gaussian, and SEIR-SD models explained >95% of variation in cases among states but suffered with the onset of resurgence (Figure 1). Because of their design to capture resurgence, the two-phase logistic and sine-logistic models generally explained >99% of variation in cases (Figure 1).

Use upon release

Our application has been in continuous use by our home institution to forecast our patient census and PPE needs or to examine observed trends against forecasts. Our application was also used on a daily basis by a NY hospital, whose frequent feedback led to several modifications.

DISCUSSION

We expect forecasting tools to continue informing operational responses until a vaccine is available and the spread of COVID-19 subsides. Because we intend for the usefulness of our application to extend throughout the pandemic, we recently implemented accurate

resurgence models, analyses of testing and hospitalization, and numerous modifications from user feedback. Because different hospitals in different states may experience resurgence at different times, the demand for our application at individual hospitals will vary as local areas experience peaks and valleys in COVID-19 infections.

Our application also provides insights beyond its primary use. First, few SEIR models account for social distancing as a combination of an emergent social response and public policy, or account for time lags in testing. Second, our two-phase resurgence models offer novel extensions of the popular logistic model. Both can be extended to any number of phases and the two-phase sine-logistic represents a novel combination of logistic growth, resurgence, and periodic fluctuations. Third, our application revealed strong statistical relationships in testing and hospitalization that have otherwise received little attention (Figure 3). Finally, researchers could conduct future studies using the downloadable data from our application, its aggregated source data, or by modifying the source code.

Hospitals should weigh the accuracy of our tool against the knowledge of their system's needs to avoid over/under allocation of resources. The following should also be considered when using our application: first, we used a popular COVID-19 dataset that may not reflect true prevalence; second, values of input parameters such as LOS and PPE needs likely vary across time; third, LOS may not necessarily be binomially distributed and time lags in hospital visits may not necessarily be Poisson distributed; finally, our application does not account for differences in susceptibility to COVID-19 with respect to age, comorbidities, and sociodemographic factors.

CONCLUSION

The worst pandemic in a century has coincided with a revolution in open-source tools and data science and has been met with numerous public-facing applications. Continued efforts should leverage a variety of models and deliver actionable predictions to prevent overreliance on a single model and to couple the regional spread of COVID-19 to local healthcare needs.

AUTHOR CONTRIBUTIONS

KJL developed the application's models and source code, was responsible for deploying and maintaining the application, and wrote the manuscript. TAW guided application development and deployment and wrote the manuscript. JK assisted with website architecture and assisted with writing the manuscript. AKA provided input on the modeling and assisted with writing the manuscript. BH guided application development and wrote the manuscript.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Johns Hopkins University Center for Systems Science and Engineering, 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository. https://github.com/CSSEGISandData/COVID-19 Accessed March 21, 2020.
2. The Atlantic. The COVID Tracking Project. 2020. https://covidtracking.com/ Accessed June 1, 2020.
3. Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *JAMA* 2020; 323 (16): 1545.
4. Liu Y, Gayle AA, Wilder-Smith A, *et al.* The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med* 2020; 27 (2): 1–4.
5. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet* 2020; 395 (10231): 1225–8.
6. Brandenburg A. Quadratic growth during the 2019 novel coronavirus epidemic. arXiv.org: arXiv:2002.03638 Accessed April 6, 2020.
7. Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis. *J Med Virol* 2020; 92 (6): 645–59.
8. Maynard-Smith J. *Models in Ecology*. Cambridge: Cambridge University Press; 1978: 16–59.
9. Bacaër N. *A Short History of Mathematical Population Dynamics*. London: Springer Science & Business Media; 2011: 35–40.
10. Locey KJ, White EP. How species richness and total abundance constrain the distribution of abundance. *Ecol Lett* 2013; 16 (9): 1177–85.
11. Roosa K, Lee Y, Luo R, *et al.* Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect Dis Model* 2020; 5: 256–63.
12. Wu K, Darcet D, Wang Q, *et al.* Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. arxiv.org: arXiv:2003.05681 Accessed March 24, 2020.
13. Buckingham-Jeffery E, Isham V, House T. Gaussian process approximations for fast inference from infectious disease data. *Mathematical Biosciences* 2018; 301: 111–20.
14. Schlickeiser R, Schlickeiser F. A Gaussian model for the time development of the Sars-Cov-2 corona pandemic disease. Predictions for Germany made on March 30, 2020. doi: 10.1101/2020.03.31.20048942. medrxiv.org Accessed April 6, 2020.
15. Hethcote HW, van den Driessche P. Some epidemiological models with nonlinear incidence. *J Math Biol* 1991; 29 (3): 271–87.
16. Boldog P, Tekeli T, Vizi Z, Dénes A, Bartha FA, Röst G. Risk assessment of novel coronavirus COVID-19 outbreaks outside China. *J Clin Med* 2020; 9 (2): 571.
17. Peng L, Yang W, Zhang D. Epidemic analysis of COVID-19 in China by dynamical modeling. arxiv.org: arXiv:2002.06563 Accessed March 24, 2020.
18. Wang H, Wang Z, Dong Y, *et al.* Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. *Cell Discov* 2020; 6 (1): 1–8.
19. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 2020; 395 (10228): 931–4.
20. Pedersen MG, Meneghini M. Quantifying undetected COVID-19 cases and effects of containment measures in Italy. doi: 10.13140/RG.2.2.11753.85600. researchgate.net Accessed April 1, 2020.
21. Kucharski AJ, Russell TW, Diamond C, *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 2020; 20 (5): 553–8.
22. Cao Z, Zhang Q, Lu X, *et al.* Estimating the effective reproduction number of the 2019-nCoV in China. 10.1101/2020.01.27.20018952 Accessed April 1, 2020.
23. You C, Deng Y, Hu W, *et al.* Estimation of the time-varying reproduction number of COVID-19 outbreak in China. doi: 10.1101/2020.02.08.20021253. medrxiv.org Accessed March 28, 2020.

24. Maier BF, Brockmann D. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China. *Science* 2020; 368 (6492): 742–6.

25. Read JM, Bridgen JR, Cummings DA, *et al*. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. doi: 10.1101/2020.01.23.20018549. arxiv.org. Accessed March 28, 2020.

26. Ridenhour B, Kowalik JM, Shay DK. Unraveling R0: considerations for public health applications. *Am J Public Health* 2018; 108 (S6): S445–54.

27. Rocklöv J, Sjödin H, Wilder-Smith A. COVID-19 outbreak on the Diamond Princess Cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. *J Travel Med* 2020; 27 (3): taaa030.

28. Radulescu A, Cavanagh K. Management strategies in a SEIR model of COVID 19 community spread. 2020. arxiv.org: arXiv:2003.11150 Accessed April 1, 2020.

29. Locey KJ, Webb TA, Khan J, Antony AK, Hota B. Data from: An Interactive Tool to Forecast US Hospital Needs inthe Coronavirus 2019 Pandemic. *Dryad Digital Repository* 2020. doi: 10.5061/dryad.1ns1rn8rx.