



OPEN ACCESS

EDITED BY

Huibert Burger,
University Medical Center Groningen,
Netherlands

REVIEWED BY

Henk Groen,
University Medical Center Groningen,
Netherlands
Md. Kamrujjaman,
University of Dhaka, Bangladesh
Inger Fabris-Rotelli,
University of Pretoria, South Africa
Paolo Grigolini,
University of North Texas, United States

*CORRESPONDENCE

David Puelz
✉ david.puelz@mcombs.utexas.edu

RECEIVED 21 February 2024

ACCEPTED 27 May 2024

PUBLISHED 20 June 2024

CITATION

Sudhakar T, Bhansali A, Walkington J and
Puelz D (2024) The disutility of compartmental
model forecasts during the COVID-19
pandemic.
Front. Epidemiol. 4:1389617.
doi: 10.3389/fepid.2024.1389617

COPYRIGHT

© 2024 Sudhakar, Bhansali, Walkington and
Puelz. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

The disutility of compartmental model forecasts during the COVID-19 pandemic

Tarini Sudhakar, Ashna Bhansali, John Walkington and
David Puelz*

Salem Center for Policy, Department of Finance, & Department of Information, Risk, and Operations Management, Austin, TX, United States

During the COVID-19 pandemic, several forecasting models were released to predict the spread of the virus along variables vital for public health policymaking. Of these, the susceptible–infected–recovered (SIR) compartmental model was the most common. In this paper, we investigated the forecasting performance of The University of Texas COVID-19 Modeling Consortium SIR model. We considered the following daily outcomes: hospitalizations, ICU patients, and deaths. We evaluated the overall forecasting performance, highlighted some stark forecast biases, and considered forecast errors conditional on different pandemic regimes. We found that this model tends to *overforecast* over the longer horizons and when there is a surge in viral spread. We bolstered these findings by linking them to faults with the SIR framework itself.

KEYWORDS

compartmental models, SIR, SEIR, pandemic surveillance, COVID-19

1 Introduction

Forecasts of the spread of COVID-19 in the United States and across the world have played a significant role in informing policymakers. Employing different mathematical approaches, these forecasting models provide predictions that inform critical decisions such as healthcare administration, allocation of medical supplies, and business and school closures. Some models even attempted to estimate the impact of current and future policies on human behavior, COVID-19 transmission, and vaccinations (1–3). As the COVID-19 pandemic recedes, there is a significant opportunity to take stake of the predictions compartmental models generated. Indeed, this was the first major pandemic in the modern “computational era,” where simulations were cheap and accessible to scientists. Researchers quickly fit these models and advised world leaders on pandemic surveillance and decision-making. This paper presents a rigorous study of these compartmental model predictions, including the magnitude of forecasting mistakes and a novel investigation of the systematic bias induced by this technology.

For the Austin-Round Rock metropolitan statistical area, forecasts given by the University of Texas (UT) COVID-19 Modeling Consortium shaped COVID-19 policies to a significant extent such as the introduction of staged lockdowns (4). These had serious repercussions on the overall economy, spanning the closure of small businesses to setbacks in K-12 students due to hybrid or online education. Statewide, Texas is estimated to have suffered a GDP loss of \$106 million and job losses of 1.2 million (5). Texas also suffered unprecedented setbacks in student achievement in reading and mathematics due to the mass transition to remote learning. The Texas Education Agency cites a roughly 3-month setback in educational attainment relative to a

pre-COVID baseline. From 2019 to 2021, the percentage of Texas students performing at or above their grade level in math decreased from 50% to 35% (6). While we have moved on from such strict measures for now, some questions remain unanswered. Did we base these policies on well-sourced data? In the future, can we still rely on these forecasting models to aid our decision-making?

Research from the US COVID-19 Forecast Hub (1) demonstrated that standalone models tend to generate large prediction errors, especially when forecasting over a long time horizon. A main contribution of this paper is to utilize a rich set of predictions coupled with realized data to investigate the bias of these errors. The current state of literature reports error summary statistics, like mean absolute or squared deviation, as well as coverage of prediction intervals, as in Fox et al. (7). A related approach relies on visualizing the prediction errors (residuals), as in Kumar Ghosh et al. (8), where they assert that “residuals of the fit are randomly distributed, which imply that the fitness of the data with the model is overall good.” The lack of rigorous focus on forecast errors is a major gap in the epidemiological literature, especially since the sole task of these models is prediction. What remains understudied and obfuscated in these statistics and visualizations are the time dependency and systematic biases in these errors. Specifically, this paper probes the following set of questions: Are prediction errors from compartmental models systematically biased in positive or negative directions? If so, to what degree, why, and are these errors related to time? We analyzed an informative dataset of realized hospitalizations, deaths, and ICU patients coupled with forecasts of these three outcomes from the UT COVID-19 Modeling Consortium. The UT Austin model is an advanced compartmental model with frequent, long-horizon forecasts, making it an important one to study. Given its impact on COVID-19 policymaking, we investigate the UT model’s forecasting performance in light of realized data; discover that forecast errors are biased, predictable, and dependent on the pandemic regime; and propose paths forward and alternative approaches.

2 Overview of compartmental models

Most researchers adopted the susceptible–infected–recovered (SIR) model and its modifications to forecast COVID-19 metrics due to its simplicity and prevalence among academics and training of epidemiologists (9). The SIR modeling framework falls under the category of compartmental models that date back to the seminal work of Ross (10).

In these models, individuals within a closed population are separated into mutually exclusive groups, or compartments, based on their disease status. At any given time, each individual is considered to be in one compartment but can move to another compartment based on the model parameters. As per the model’s assumptions, a susceptible individual will become infected by coming into contact with an infected individual. The individual during the infected period is assumed to be contagious. After this, the individual advances to a non-contagious state, known as recovery. Recovery may also be death or effective isolation (11). The subgroups modeled by this framework are given by the following notation:

- S is the fraction of susceptible individuals who are able to contract the disease,
- I is the fraction of infectious individuals who can transmit the disease, and
- R is the fraction of recovered individuals who have become immune.

Once these groups are defined, the trajectory of this system, known as the classical SIR model where the subgroups are the state variables, is given by the following set of differential equations:

$$\begin{aligned}\frac{\partial S}{\partial t} &= -\beta IS, \\ \frac{\partial I}{\partial t} &= \beta IS - \gamma I, \\ \frac{\partial R}{\partial t} &= \gamma I.\end{aligned}\tag{1}$$

State variables S , I , and R are all functions of time. Model parameters β and γ are the rate of transmission and recovery, respectively. The first equation in Equation 1 models the fraction of people susceptible to the virus at a certain point, given the transmission rate and the fraction of infected individuals. The second equation in Equation 1 models the instantaneous fraction of infected individuals using susceptible individuals and the rate of recovery. The third equation in Equation 1 models the instantaneous fraction of recovered individuals using the rate of recovery.

Modified SIR models, such as susceptible–exposed–infected–recovered (SEIR), require a more complex set of equations and parameters. Due to the presence of little information and lack of reliable data at the beginning of the COVID-19 pandemic, many researchers relied on the classical SIR implementation (12). Nonetheless, the setup for this modified model formulation is an extension by the SIR framework by one additional compartment, denoted “exposed.” The state variable for this new compartment is E (13). Here, we assume that there are equal birth and death rates μ , α is the mean latency period for the virus, γ is the mean infectious period, the rate of transmission as before is β , and recovered individuals do not contract the disease again. This admits the following set of equations:

$$\begin{aligned}\frac{\partial S}{\partial t} &= \mu - \beta IS - \mu S, \\ \frac{\partial E}{\partial t} &= \beta IS - (\mu + \alpha)E, \\ \frac{\partial I}{\partial t} &= \alpha E - (\mu + \gamma)I.\end{aligned}$$

Since each variable is defined as a fraction of the entire population, we calculate R from the equation $S + E + I + R = 1$.

Similar to the SIR model, this system describes the population evolution of the virus in terms of intuitive parameters that summarize its interaction with humans, like ease of transmission, incubation period, and length of recovery. Conveniently, it is easy to extend this framework to higher-dimensional systems.

The number of compartments can be extended, new parameters can be added, and time variation in parameters can be incorporated, as shown in Girardi and Gaetan (14) and Spannaus et al. (15).

As a final practical note, a useful metric derived from this system is called the basic reproduction number R_0 , which represents the average number of infections generated from an infected individual within a susceptible population. During the COVID-19 outbreak, this was a closely watched and debated number since it describes the evolution of the virus itself.

$$R_0 = \frac{\beta\alpha}{(\mu + \alpha)(\mu + \gamma)}.$$

The reproduction number (also called effective reproductive number) is a time-varying version of the basic reproduction number and denotes the evolving transmissibility of the disease. See Dharmaratne et al. (16) for a detailed discussion of these two important epidemiological concepts.

In the following subsections, we provide two examples of compartmental model use during the COVID-19 pandemic. The examples build on each other and demonstrate the flexibility of compartmental modeling. First, we summarize the work done by the Institute for Health Metrics. Second, we describe the additional features incorporated into the model developed by the UT COVID-19 Modeling Consortium. UT's model forecasts are the focus of this paper.

2.1 Example 1: Institute for Health Metrics and Evaluation

The Institute for Health Metrics and Evaluation (IHME) COVID-19 Forecasting Team located at the University of Washington created multiple models over the course of the pandemic. The initial model garnered much press and attention and used statistical curve-fitting to estimate hospital bed utilization, ICU admissions, ventilator use, and deaths from 25 March 2020 to 29 April 2020 (17).

In this model, the IHME team took a different approach to model death rates compared with the classical compartmental framework. They critiqued SEIR models for their assumption of random mixing between all individuals in the population because, under that assumption, millions of COVID-19-related deaths were predicted very early in the United States. Random mixing does not account for behavioral changes and government-mandated social distancing measures. Instead, the IHME team modeled actual COVID-19 death rates since they would indicate virus transmission and fatality rates. According to the authors, deaths were also more accurately reported than cases, especially in limited testing areas as those would allocate tests for severely ill patients first. They also assumed hospitalization and related services to be highly correlated with deaths.

Their next model took on a hybrid approach, estimating revised death rates and then fitting an SEIR model until 26 May 2020 (18, 19). The third model replaced curve-fitting with spline-fitting for the relationship between log cumulative deaths and log cumulative cases while retaining the SEIR model estimation (18, 20). When tested against other forecasting models for predictive accuracy, this third IHME model exhibited the best performance (out of seven models) (18).

2.2 Example 2: the UT COVID-19 Modeling Consortium

The UT COVID-19 Modeling Consortium developed a highly publicized and widely used model to predict trends in key pandemic variables, including hospitalizations, deaths, and ICU patients, in the Austin-Round Rock metropolitan statistical area (21).¹ Motivated by the IHME approach, the UT team developed an alternative curve-fitting method for forecasting COVID-19 outcome variables. While the underlying technology remained a compartmental model, the team layered in other sources of hierarchical information and complexity.

1. Mobile phone data to capture social distancing measures: To capture the effect of changing social distancing measures for each US state on individual-level mobility, the model used local data from mobile phone GPS traces from SafeGraph.²
2. A correction for the underestimation of uncertainty in the IHME forecasts: Using deaths as an example outcome, the IHME model estimated cumulative death rates using a least-squares-like procedure on the log scale and calculated confidence intervals based on large-sample statistical theory. For this method to produce valid uncertainty measures, consecutive model errors should be independent of each other. However, this assumption is violated in the IHME fitting procedure since today's cumulative death rate includes yesterday's death rate and an increment. This implies that these two death rates must be correlated. The UT model corrected this by fitting daily death rates using a mixed-effects negative-binomial generalized linear model, accounting for heteroskedasticity and correlation.

In the following sections, we investigate the forecasts generated from the UT model and describe telling features of the forecast errors.

¹The data are available on the UT team's [GitHub](#) directory. These forecasts are also visually available to the public on their [website](#).

²More information on SafeGraph can be found [here](#).

3 Methodology

We now focus exclusively on the forecasts generated from the UT COVID-19 Modeling Consortium. We first describe our methodology for analyzing the errors and then turn to our results. Our analysis began with forecast error computation and visualization. We then constructed a new variable defining “pandemic regime,” which describes whether viral spread is surging or waning. This allowed us to condition data on regime status and fit regression models that describe prediction error as a function of the regime status. In detail, our analysis of the UT forecasting model for hospitalizations, ICU patients, and deaths followed these steps:

1. Visualizing forecasts against realized data: We plotted realized data with forecasts generated by the UT model. While we have daily realized data on hospitalizations, ICU patients, and deaths, we did not have the same for forecasts as the model was not updated daily and had a different schedule for each outcome variable.
2. Mapping forecasting errors from 1 to 20 days ahead: We computed the forecasting errors out to 20 days ahead, showing the median forecasting error and interquartile range (IQR) of the error distribution. The forecast error for a given number of days out (denoted d) at a particular time point t is calculated as follows:

$$E_{t+d,t} = F_{t+d,t} - A_{t+d}, \quad (2)$$

where

- $E_{t+d,t}$ is the forecast error for the forecast made at time t for time $t + d$,
 - A_{t+d} is the actual (realized) value at time $t + d$, and
 - $F_{t+d,t}$ is the predicted value at time t for time $t + d$.
3. Mapping forecasting errors, conditional on whether the virus is spreading or waning: We computed the forecasting errors from 1 to 20 days out, conditional on whether there is a surge in the viral spread.

For each outcome of interest, we defined rising hospitalizations/ICU patients/deaths as surging and falling hospitalizations/ICU patients/deaths as waning. We constructed this new variable in two steps. First, we computed the daily percent change in the outcome of interest. Second, to smooth out noise in the realized data, we took a 14-day moving average of the realized outcome’s percent change. Our final definition of a rising (falling) point in time t is a day that has a 14-day average percent change of greater (less) than zero.

Using the realized data for a given outcome A_t , the percent change is given by

$$\dot{A}_t = \frac{A_t - A_{t-1}}{A_{t-1}} \times 100.$$

The smoothed data (over a 14-day rolling window) is then defined in the following way:

$$\dot{A}_t = \frac{1}{14} \sum_{k=t-14+1}^t \dot{A}_k.$$

With \dot{A}_t in hand, we can now define a new indicator variable that summarizes the state of the pandemic according to the smoothed growth or abatement of the outcome of interest. The new variable $\mathbf{I}_t^{\text{rising}}$ describes whether or not time t is a “rising” or “falling” regime in the following way:

$$\mathbf{I}_t^{\text{rising}} = \mathbb{1}(\dot{A}_t > 0) = \begin{cases} 1 & \text{if } \dot{A}_t > 0, \\ 0 & \text{if } \dot{A}_t \leq 0, \end{cases} \quad (3)$$

where $\mathbb{1}(\cdot)$ is the indicator function.

Our final methodological step was to empirically investigate the forecast errors through a regression model. We were interested in testing whether the errors are related to the pandemic regime, so we specified the following Gaussian linear regression model:

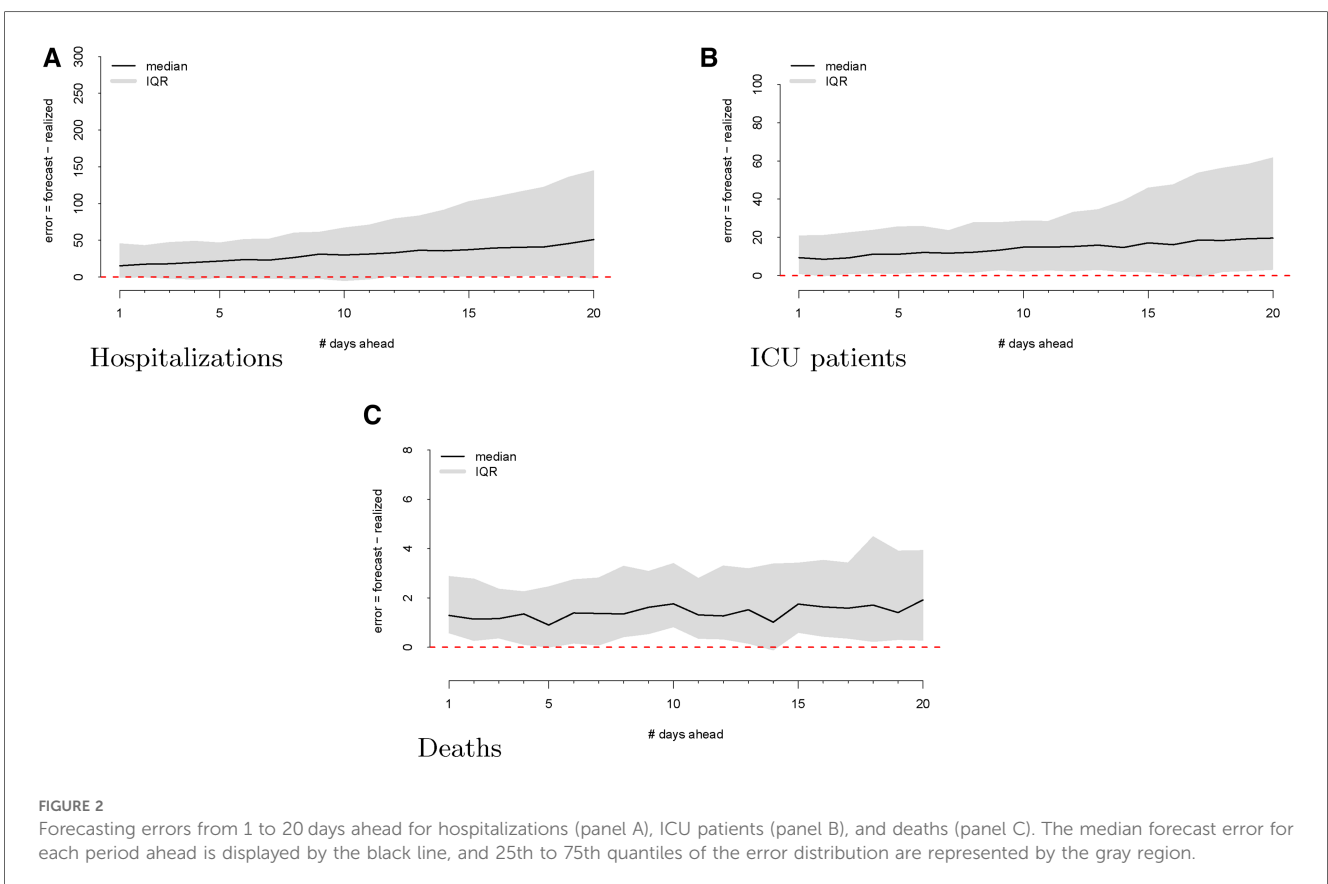
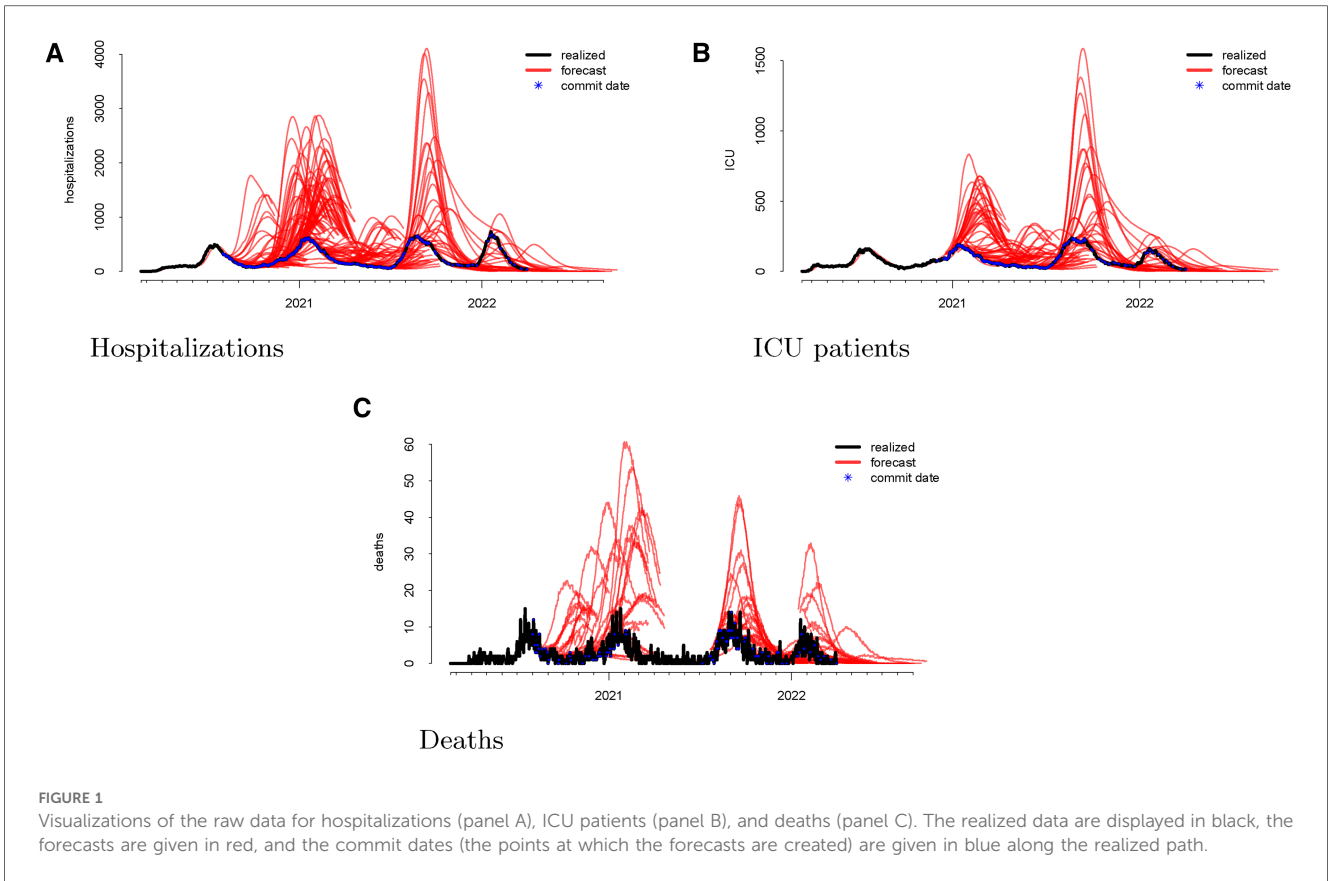
$$E_{t+d,t} = \eta + \theta \times \mathbf{I}_t^{\text{rising}} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \quad (4)$$

where η is the intercept, θ is the slope of the regression model, and $d = \{1, 5, 10, 20\}$ denotes the forward time horizons we considered. Moreover, we fitted 12 models, 4 forward time horizons for each of hospitalizations, deaths, and ICU capacity. The slope θ has an intuitive interpretation as the difference in average prediction error between rising and falling pandemic regimes. Therefore, the results of θ ’s inference are crucial to understanding the relationship between forecast mistakes and the “state of the pandemic.” The model is fit using ordinary least squares (OLS), and we report point estimates and standard errors to gauge statistical significance in the next section.

4 Results

Following the methodological steps outlined in the previous section, we now investigate each of the three outcomes forecasted by the UT model. In the data visualizations, each figure panel corresponds to one of hospitalizations, deaths, or ICU patients. Figure 1 displays the raw data. The realized data, in black, are daily and starts on 19 February 2020 and ends on 1 April 2022. The forecasts, in red, start on 1 August 2020 and end on 1 April 2022. Although we have daily realized data, the model was not updated daily. Instead, there are 246 forecast dates, denoted “commit dates” in the plot and shown in blue. These are the dates when the entire model with new forecasts is updated, i.e., when each of the red forecasts begin.

As a first step, we computed the forecasting errors out to 20 days ahead, as displayed in Figure 2. The median forecasting error is shown in black, and the IQR of the error distribution is the shaded region. An immediately visible feature is the upward



bias of the error distribution, even via the median error, which is robust to egregiously large forecasts. This implies that the model systematically *overforecasts* hospitalizations.

We next considered visualizing the errors conditional on different stages of the pandemic. Our goal was to answer the question: Does the model systematically *over- or underforecast* when the virus is spreading more rapidly or is receding? In other words, is the model more likely to make forecasting errors during variant surges or when the viral spread is waning? To answer this question, we defined rising hospitalizations as surging and falling hospitalizations as waning. To smooth out noise in the realized data, we computed the percent change in the realized hospitalizations and smoothed with a 14-day rolling window. Our final definition of a rising (falling) regime is a day that has a 14-day average percent change of greater (less) than zero.

The regimes as defined above are displayed in Figure 3. The forecast errors conditional on these regimes (analogous to Figure 2) are displayed in Figures 4 (rising) and 5 (falling). When looking at these figures, an interesting feature emerges. The model makes large positive mistakes during surges (and with larger variance, Figure 4) compared to when the viral spread is waning (Figure 5) and hospitalizations are falling.

We then investigate the number of ICU patients and its corresponding forecasts. Figure 1 displays the raw data vs. the forecasts at each commit date. Similar to our work on hospitalizations, we calculated the forecasting error up to 20 days

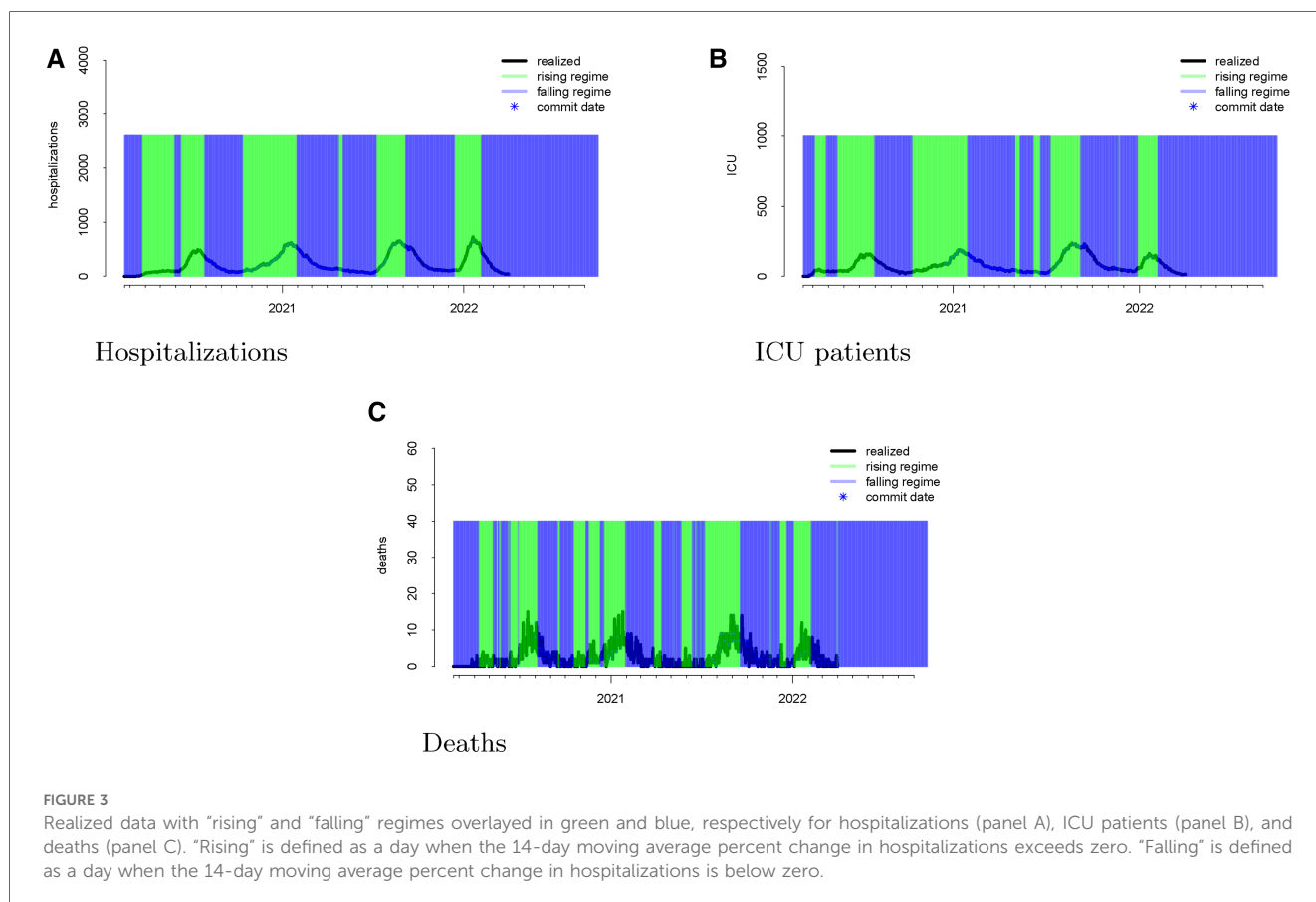
ahead, as shown in Figure 2. There is a clear upward bias, especially for predictions generated 15–20 days ahead. This implies again that the model systematically *overforecasts* the number of ICU patients.

We mapped the errors to the rising and falling regimes defined by the realized ICU admission data. As defined in Section 2.2, rising ICU admissions correspond to a surge in the viral spread, and falling ICU admissions denote waning viral spread. Figure 3 visualizes these rising and falling regimes. We see a distinct trend emerge in the rising regime. Errors in the model are larger and more positive when ICU admissions are surging compared to when the ICU admissions are falling.

Finally, we looked at the daily COVID-19 deaths. Figure 1 shows the model’s forecasts (in red), along with the actual values (in black). This is similar to the plots shown above for hospitalizations and ICU admissions, where the realized values are daily but aperiodic, and the forecasts lines begin on their respective commit dates.

Figure 2 shows the forecasting errors of the model computed with forecasts up to 20 days ahead. Similar to the previous error plots, the black line shows the median forecasting error, while the IQR is shaded in gray. Interestingly, deaths exhibit a weaker upward bias compared to ICU and hospitalization errors.

Once again, we display the realized data during the “rising” and “falling” regimes in Figure 3. To assess whether the model has different forecasting errors during a “rising” regime or a “falling”



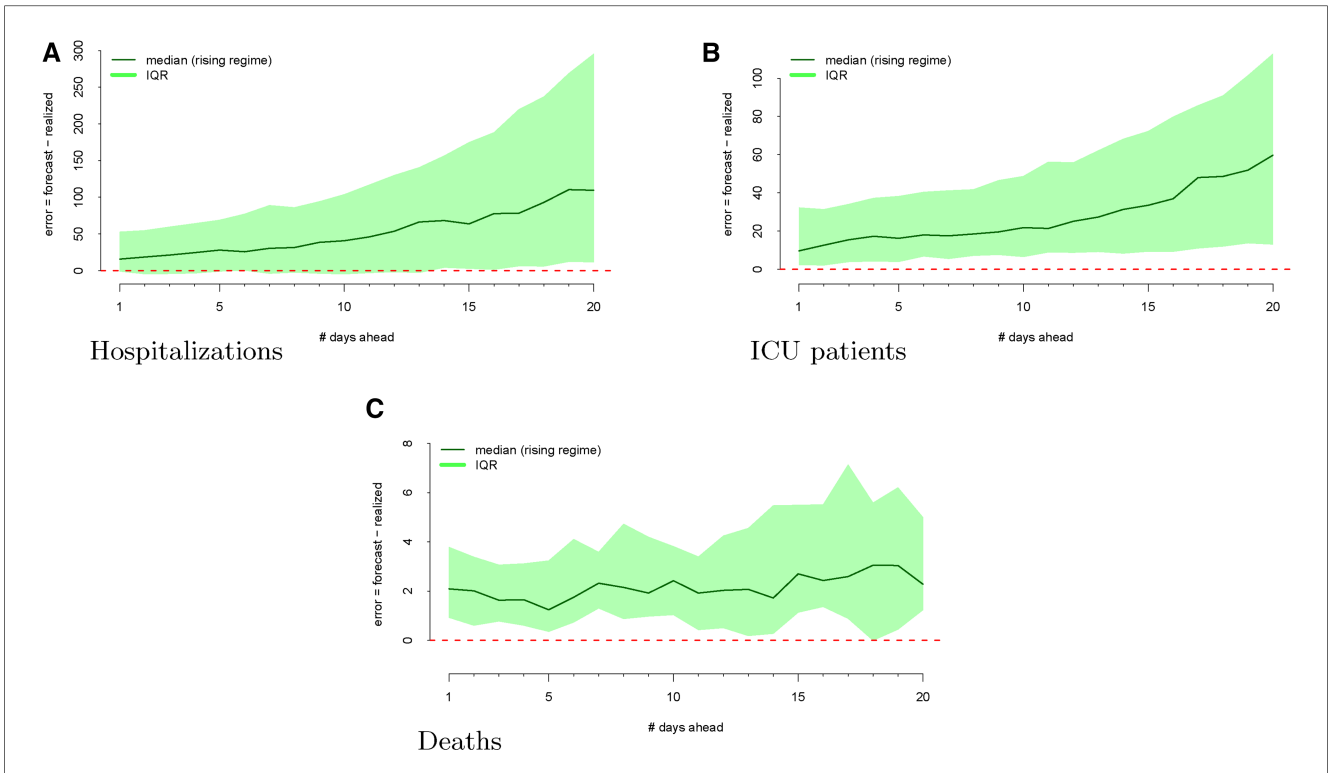


FIGURE 4 Forecasting errors from 1 to 20 days ahead, conditional on forecasting in a rising regime for hospitalizations (panel A), ICU patients (panel B), and deaths (panel C). The median forecast error for each period ahead is displayed by the bold line, and 25th to 75th quantiles of the error distribution are represented by the shaded region.

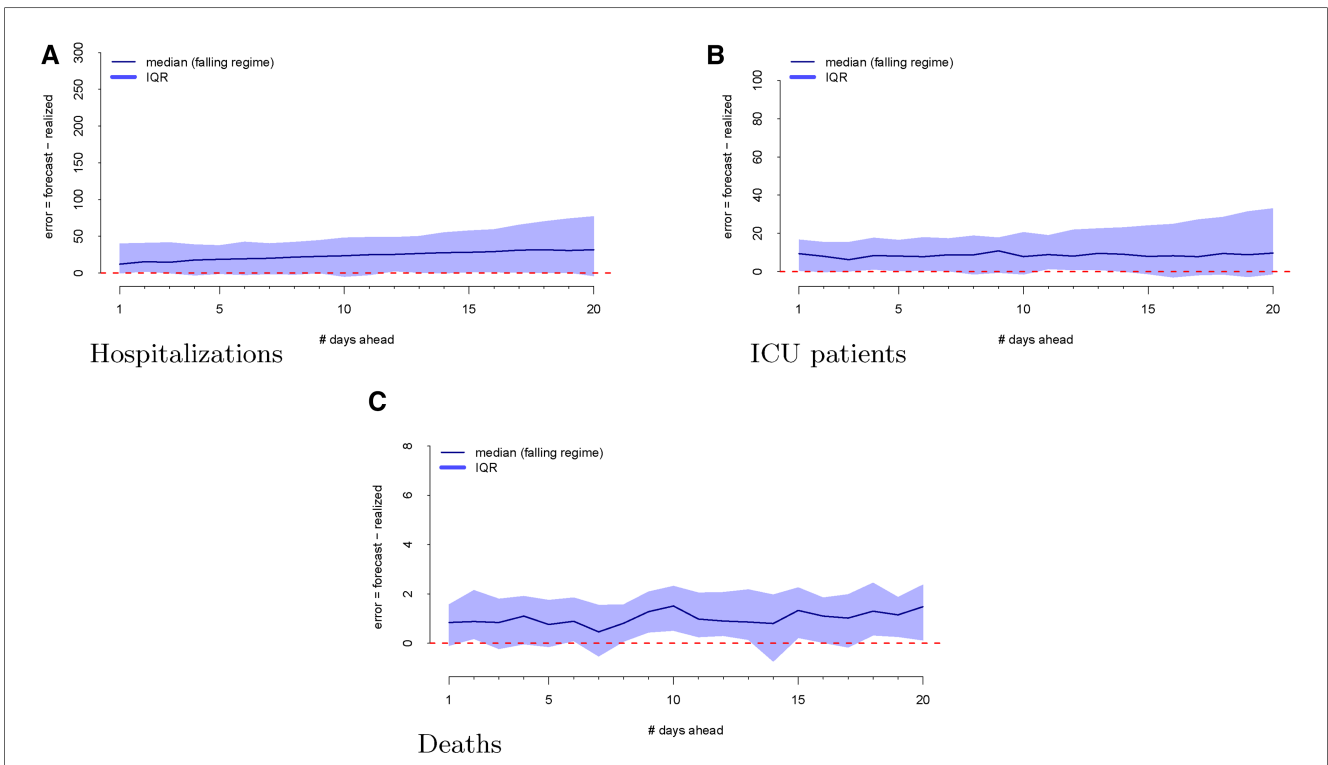


FIGURE 5 Forecasting errors from 1 to 20 days ahead, conditional on forecasting in a falling regime for hospitalizations (panel A), ICU patients (panel B), and deaths (panel C). The median forecast error for each period ahead is displayed by the bold line, and 25th to 75th quantiles of the error distribution are represented by the shaded region.

one, we plotted the errors conditioned on the regime when these values were forecast.

Figure 4 shows the model's errors during a "rising" regime, whereas Figure 5 shows the model's errors during a "falling" regime. Unlike with hospitalization and ICU forecasts, the errors are similar and more attenuated in both regimes.

4.1 Empirical analysis of model errors: are the errors predictable?

We conclude our analysis of the SIR model errors by fitting several regression models relating the forecast error to the viral regime. Specifically, we regressed the forecast error at 1, 5, 10, and 20 days ahead on the indicator variable of whether or not the forecast date was a "rising regime," as defined in Equation 3. The unit of analysis for these regressions is the commit date for each outcome since this is when the forecasts were generated; the outcome is the forecasting error d days ahead, as given by Equation 2. The regression model Equation 4 was fitted using OLS for 12 outcomes: 4 forward time horizons for each of hospitalization, deaths, and ICU patients.

Tables 1–3 correspond to the regression fits for hospitalization, ICU, and death forecast errors, respectively. Each table shows the intercept and coefficient estimates for each regression in a column, as well as standard errors in parentheses and p -values noted in asterisks. Each coefficient on the indicator variable represents the average difference in error between rising and falling regime forecasts.

Across all outcomes, there are strongly significant differences in forecast errors depending on the viral regime, especially for forecasts 10 and 20 days ahead. All estimated indicator coefficients are positive. For example, the hospitalization model errors are 125.99 (Table 1) greater in rising regimes than in falling regimes. This predictability of errors is shocking since forecast errors should be random.

TABLE 1 Hospitalization forecast errors regressed on rising regime indicators.

Outcome	1 day ahead	5 days ahead	10 days ahead	20 days ahead
Intercept	20.24 (3.29)***	20.64 (4.47)***	24.96 (7.18)***	42.72 (14.99)**
ℓ (rising regime)	8.80 (4.87)	18.75 (6.61)**	37.58 (10.62)***	125.99 (22.12)***
R ²	0.01	0.03	0.05	0.12
Adj. R ²	0.01	0.03	0.05	0.11

*** $p < 0.001$; ** $P < 0.01$; * $P < 0.05$.

TABLE 2 ICU forecast errors regressed on rising regime indicators.

Outcome	1 day ahead	5 days ahead	10 days ahead	20 days ahead
Intercept	11.97 (2.08)***	10.89 (2.47)***	11.04 (3.31)**	15.29 (6.70)*
ℓ (rising regime)	4.55 (3.09)	10.91 (3.67)**	18.82 (4.92)***	58.50 (9.93)***
R ²	0.01	0.05	0.09	0.18
Adj. R ²	0.01	0.05	0.08	0.18

*** $p < 0.001$; ** $P < 0.01$; * $P < 0.05$.

TABLE 3 Death forecast errors regressed on rising regime indicators.

Outcome	1 day ahead	5 days ahead	10 days ahead	20 days ahead
Intercept	0.95 (0.42)*	0.77 (0.51)	1.50 (0.46)**	1.73 (0.71)*
ℓ (rising regime)	1.82 (0.59)**	1.03 (0.72)	1.26 (0.64)	2.22 (0.98)*
R ²	0.10	0.02	0.04	0.06
Adj. R ²	0.09	0.01	0.03	0.05

*** $p < 0.001$; ** $P < 0.01$; * $P < 0.05$.

5 Discussion: modeling challenges

The SIR model fails to properly model certain aspects of disease spread. Melikechi et al. (22) pointed out that over the last century, many have modified the SIR model to incorporate different compartments for various subpopulations or added new terms that identify unique pathogen transmissions. However, adding too many features may lead to overfitting when making inferences of parameters early on in an epidemic (23). At the beginning of an epidemic or pandemic, there is a risk of noisy observations. Melikechi et al. (22) raised the concept of *practical identifiability*, which refers to the ability to "discern different parameter values based on noisy observations." Most SIR models employ Monte Carlo simulations, where the model is simulated with predetermined parameters. Noise is added to the simulated data, and then, a fitting procedure is used on the noisy data. With increase in the magnitude of noise, Monte Carlo parameter estimates often display large difference in values, leading to huge uncertainty in the parameters and unreliable inferences.

Lemoine (24) dissected missteps of a popular SEIR model built by Flaxman et al. (25), in which they analyzed the effect of non-pharmaceutical interventions on deaths due to the virus. Flaxman et al. (25) used partial pooling of information between countries, with both individual and shared effects on the time-varying reproduction number. According to them, pooling allowed for more information to be used and helped overcome country-specific idiosyncrasies in the data to enable more timely estimates. The authors argued that government lockdowns were the interventions that made a bulk of the impact on controlling the viral spread. Lemoine, however, pointed out that the overall effect of government interventions on deaths in Sweden was similar to that in other countries. Why is this important? Sweden was the one country in the analysis that did not have a full lockdown. When he reproduced their analysis, Lemoine found that the country-specific effect for Sweden that the model ignored was almost as large as the effect of a full lockdown, a feature that the authors failed to present in their findings. Based on this, Lemoine argued that such SIR models have been unable to deliver useful inputs for policymakers. However, more importantly, epidemiologists have failed to acknowledge this fact, by not ascribing the failure of their models to the right causes.

There are three major critiques for the SIR models: assumption of homogeneous mixing, assumption of closed population, and latency period of infection.

5.1 Homogeneous mixing

SIR models assume random mixing between all individuals in a given population. Based on this assumption, each individual has the same amount of contact as everyone else. Such a model would not be able to account for a higher contact rate at hospitals or a lower contact rate for quarantined individuals (26). There is a lack of adequate inclusion of individual behavioral and social influence in SIR models. Infectious disease epidemics have a substantial social aspect and public health implication. Homogeneous mixing assumes an equal probability of transmission between two people regardless of their age or location (27). This can fail to take into account age-dependent or location-dependent risks. We need to include varying degrees of interventions such as social distancing, stay-at-home, and shelter-in-place orders at different times and across different regions. The assumption of homogeneous mixing of S with I state individuals in the SIR model is therefore invalid during COVID-19.

This flaw of SIR models may explain why the errors tend to be so high with the UT Austin model. While homogeneous mixing can be helpful for projecting the number of cases, it can lead to large errors both in the early stages of the epidemic and in calculating the final epidemic size (27). Homogeneous mixing often overestimates the epidemic's size, and can lead to more interventions than needed. While the UT model accounts for different age groups and risk factors, it still models disease transmission through an SIR framework where the base assumption remains homogeneous mixing. That is, all individuals within each group would have the same susceptibility to infection, and all individuals within each infection status compartment would have the same infectiousness.

Even with more spatially explicit metapopulation models, homogeneous mixing at smaller scales, such as within a state, county, or city, is still questionable, as some people can stay home and some are essential. In addition, the regional variability of individual sentiment and behavior, for example, whether to obey or enforce these orders, is essential to determine to predict the trajectory of the COVID-19 pandemic, but it is generally not included in the SIR models (28).

Assumption of homogeneous mixing can lead to overestimating health service needs by not accounting for behavioral changes and government-mandated actions. In Wuhan, strict social distancing was instituted on 23 January 2020, and by the time new infections reached 1 or fewer a day (15 March 2020), less than 0.5% of the population was infected. At the time, SIR models generally suggested that 25%–70% of the population to be infected (17).

Since most SIR models consider a single R_0 value, they miss unexpected social behavior changes and are unable to follow the alterations. For instance, social gatherings have a great impact on disease spread. A religious event in Malaysia, held from 27 February to 3 March, was supposed to be the source of viral spread in India and Pakistan (12, 29).

Chen et al. (28) argued that since SIR models are formulated at the population level, we face an important discrepancy between patient-level data and population-level modeling. Exposed (E) and

infectious (I) compartments characterize the disease spread at the population level, ignoring individual clinical variations in patients. Due to its broadness, the E compartment assumes that everyone exposed is unable to infect others and the I compartment does not account for varying levels of severity among patients such as asymptomatic, mild, and severe stages. Given these assumptions, SIR models cannot pin down and quantify the impact of superspreaders who can lead to a disproportionately large number of new cases. Superspreading can be due to individual clinical characteristics such as supershedding of virus or behavioral aspects like supercontacting. Neither are addressed well by SIR models built at a population level. This flaw in the SIR models may also explain some of the inaccuracy in the UT model, as not accounting for patient-level differences may cause them to overlook individual behaviors and *overforecast*.

5.2 Closed population

The focus of SIR models is often placed on the estimation of the basic reproduction number R_0 (30). However, what should be addressed is the assumption of a closed population in SIR models. Most regions do not follow complete isolation, making them vulnerable to changes in the neighboring communities. SIR models also consider recovered individuals to be immunized. This assumption contrasts with the possibility of the reactivation of the virus or reinfection of previously infected individuals (12). Similarly, it does not account for asymptomatic individuals. With the closed population assumption, Ding et al. (31) argued that standard SIR models miss out on the fact that presymptomatic and asymptomatic cases can spread the disease between populations through travel. Researchers have addressed this at multiple levels: within state, country, and even globally.

Studies such as those by Kucharski et al. and Wu et al. (32, 33) estimated cases in Wuhan, China, by considering the movement people in and out of the city. Kucharski et al. (32) did so by assuming that once exposed, a part of the population would travel internationally. To account for international travelers, they used the number of outbound travelers (assuming 3,300 per day before travel restrictions were imposed on 23 January 2023, and zero afterward), relative connectivity of different countries, and relative probability of reporting a case outside Wuhan vs. within Wuhan and internationally.

Wu et al. (33) first inferred the R_0 , the basic reproduction rate of the virus, of COVID-19 and the outbreak size in Wuhan from 1 December 2019 to 25 January 2020, on the basis of confirmed cases exported from Wuhan to cities outside of mainland China, where symptom onset date was reported to range from 25 December 2019 to 19 January 2020. They also forecasted the spread of COVID-19 within and outside Mainland China, taking into account public health interventions and outbound travelers by air, train, and road during the Spring Festival. While they assumed that travel behavior was not affected by the disease and, therefore, international case exportation occurred according to a non-homogeneous process, their work still addresses mobility across cities and countries for modeling COVID-19.

Chinazzi et al. (34) used a global epidemic and mobility model (GLEAM) to model the international spread of COVID-19, considering varying transmissibility and air traffic reductions. The model uses a metapopulation network approach with real-world data, where the world is divided into subpopulations centered around major transportation hubs such as airports. The subpopulations are connected to each other by individuals traveling daily through them. COVID-19 transmission within each subpopulation is modeled through a susceptible–latent–infectious–recovered compartmental framework.

Ding et al. (31) differentiated their study by incorporating granular changes in air traffic and simulate varying travel restrictions. They focused on data from Canada, which showed a large number of flights going to and from the country despite travel restrictions and reduced air traffic. They proposed a modified SIR model that considers a dynamic flight network, by estimating imported cases using air traffic volume and positive testing rates. Their model operates in an “open population setting,” where people are free to travel in and out of the population.

Not accounting for a lack of closed population leads to poor estimation of forecasts. Depending on the assumptions and circumstances of the model, this assumption can lead to both underestimation and overestimation. For instance, Chowell and Nishiura (35) revealed that in the case of the Ebola virus, variations in the R_0 number in an SEIR model were due to different assumptions regarding the international or domestic spread of the virus and the lack of high-quality data.

In our analysis of the UT Austin model, we captured the model’s systematic tendency to *overforecast*. UT’s initial R_0 values in 2020 were “best guesses,” as seen in the study by Wand et al. (36). They assumed R_0 to be 2.2 but did not provide a clear source for how they arrived at this number.³ Tec et al. (37) estimated R_0 using the basic infectiousness of the disease, the number of people susceptible to infection, and the impact of social distancing, mask wearing, and other measures to slow transmission. One potential reason why the UT model *overforecasts* could be due to their social distancing data. They used mobility trends from SafeGraph data and regressed the transmission rate of the virus on the first two principal components derived from a principal component analysis (PCA) on eight independent mobility variables, such as home dwell time and visits to universities, bars, grocery stores, museums and parks, medical facilities, schools, and restaurants. If the PCA components did not appropriately capture the variation in the mobility, such as inter-Austin movement, then the transmission rate would lead to a poor estimation of the viral spread.

5.3 Latency period of infection

The SIR model also does not incorporate the latent period between when an individual is exposed to a disease and when

that individual becomes infected and contagious. This is because the only categories in the model are susceptible, infectious, and recovered. The S(E)IR model tries to account for this parameter by creating a category for people who are exposed but not yet contagious. However, even S(E)IR models are oversimplified, and they will need to model other time-dependent factors, such as the introduction of community mitigation strategies (11).

Changes in model parameters E , I , or R at time t is dependent on a fraction of E and I at time t . As per Liu (38), this means that after being exposed to the virus on a particular day, an individual may become contagious or recover on that same day. However, in reality, an exposed individual will become infectious only after a latent period and recover after an infectious period (38). This timing issue with the compartmental models can cause the forecasts to be inaccurate, especially for further days out.

SEIR models account for this latency period using an “exposed” compartment, but even this model feature is too structured and simplistic. If the latency period is not calibrated precisely, the model may *over/underforecast*. A study using COVID-19 data in Tennessee found that the “optimal” latency period is 2.40 days, which is close to the mean latency period of 2.52 days, which was estimated from the data of seven countries (39). If UT Austin’s SEIR model did not correctly model this latency period, this could explain some of the model’s errors.

6 Conclusion

Given the wide use of compartmental models to describe the transmission dynamics of COVID-19 and other diseases, we must carefully consider their limitations when using them to inform public health interventions. In particular, homogeneity assumptions underlying these models do not accurately reflect heterogeneity of the population, and estimates of key parameters such as R_0 are often noisy and unreliable. In addition, these models do not account for the impact of non-pharmaceutical interventions on disease transmission or capture the complex interactions between the virus, people, and the environment.

Historically, compartmental models were used as descriptive tools instead of real-time models for prediction and decision-making. This change in “use-case” partially explains their significant failure. Chen et al. (28) pointed out that in the initial papers describing the SIR approach, the model was applied after the epidemic had ended. However, SIR models have little room for new evidence without modifying the model structure and estimation of R_0 . In addition, it was difficult to model the initial part of COVID-19 spread using SIR models, compared to other models, since we had limited or no information on aspects such as asymptomatic transmission, superspreaders, and unreported cases. Moreover, constructing a long-term forecasting model is a major challenge because of the lack of data. The SIR model has a tendency to underestimate peak infection rates and substantially overestimate the persistence of the epidemic after the peak has passed (40).

Intuitively, prediction accuracy can be increased by synthesizing the forecasts of many unique models. This represents one important area of future research to improve the

³The authors linked a source, but it is not open access.

usefulness of compartmental models. Cramer et al. (1) found that an ensemble model provides more accurate short-term forecasts of hospital and ICU admissions than the individual models alone. The ensemble model also has a lower prediction error and better calibration than the individual models, suggesting that it could be more effective in real-time decision-making for healthcare systems.

In this paper, we investigated predictive failures of the compartmental model framework and its derivative models. While our empirical analysis utilized forecasts from the model constructed at UT Austin, our conclusions apply to all compartmental models generated from sets of differential equations. The epidemiological community should scrutinize this framework and promote new and innovative techniques that go beyond differential equations and adopt solutions from economics, statistics, machine learning, and beyond. A pandemic forecasting model can always be marketed as unique, but if it relies on differential equations and compartmental modeling, this structural bias will outweigh most novel modifications. We hope this manuscript contributes to a rebirth of research in epidemiological modeling so the best techniques can be elevated and utilized in future pandemics.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

References

- Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, Rivadeneira AJC, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proc Natl Acad Sci U S A*. (2022) 119(15): e2113561119. doi: 10.1073/pnas.2113561119
- Hassan MN, Mahmud MS, Nipa KF, Kamrujjaman M. Mathematical modeling and COVID-19 forecast in Texas, USA: a prediction model analysis and the probability of disease outbreak. *Disaster Med Public Health Prep*. (2023) 17:e19. doi: 10.1017/dmp.2021.151
- Mahmud MS, Kamrujjaman M, Adan MMIY, Hossain MA, Rahman MM, Islam MS, et al. Vaccine efficacy and SARS-CoV-2 control in California and US during the session 2020–2026: a modeling study. *Infect Dis Model*. (2022) 7(1):62–81. doi: 10.1016/j.idm.2021.11.002
- Yang H, Sürer Ö, Duque D, Morton DP, Singh B, Fox SJ, et al. Design of COVID-19 staged alert systems to ensure healthcare capacity with minimal closures. *Nat Commun*. (2021) 12(1):3767. doi: 10.1038/s41467-021-23989-x
- Liu L, Shwiff SS, Shwiff SA, Miller M. Impact of COVID-19 on the US and Texas economy: a general equilibrium approach (2021). Available online at: https://digitalcommons.unl.edu/icwdm_usdanwrc/2545/ (Accessed June 12, 2024).
- Texas Education Agency. Impacts of COVID-19 and accountability updates for 2022 and beyond. In *2021 Texas Assessment Conference; 2021 Nov 1–3; Round Rock, TX*. Texas Assessment Conference (2021).
- Fox SJ, Lachmann M, Tec M, Pasco R, Woody S, Du Z, et al. Real-time pandemic surveillance using hospital admissions and mobility data. *Proc Natl Acad Sci U S A*. (2022) 119(7):e2111870119. doi: 10.1073/pnas.2111870119
- Kumar Ghosh J, Saha P, Kamrujjaman M, Ghosh U. Transmission dynamics of COVID-19 with saturated treatment: a case study of Spain. *Braz J Phys*. (2023) 53(3):54. doi: 10.1007/s13538-023-01267-z
- Rahimi I, Chen F, Gandomi AH. A review on COVID-19 forecasting models. *Neural Comput Appl*. (2023) 35(33):23671–81. doi: 10.1007/s00521-020-05626-8
- Ross R. An application of the theory of probabilities to the study of a priori pathometry—part I. *Proc R Soc London Ser A*. (1916) 92(638):204–30. doi: 10.1098/rspa.1916.0007
- Tolles J, Luong T. Modeling epidemics with compartmental models. *JAMA*. (2020) 323(24):2515–6. doi: 10.1001/jama.2020.8420
- Moien S, Nickaeen N, Roointan A, Borhani N, Heidary Z, Javanmard SH, et al. Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. *Sci Rep*. (2021) 11(1):4725. doi: 10.1038/s41598-021-84055-6
- Brauer F, Castillo-Chavez C, Feng Z. *Mathematical Models in Epidemiology*. Vol. 32. New York: Springer (2019).
- Girardi P, Gaetan C. An SEIR model with time-varying coefficients for analyzing the SARS-CoV-2 epidemic. *Risk Anal*. (2023) 43(1):144–55. doi: 10.1111/risa.13858
- Spannaus A, Papamarkou T, Erwin S, Christian JB. Inferring the spread of COVID-19: the role of time-varying reporting rate in epidemiological modelling. *Sci Rep*. (2022) 12(1):10761. doi: 10.1038/s41598-022-14979-0
- Dharmaratne S, Sudaraka S, Abeygunawardena I, Manchanayake K, Kothalawala M, Gunathunga W. Estimation of the basic reproduction number (R0) for the novel coronavirus disease in Sri Lanka. *Virology*. (2020) 17:1–7. doi: 10.1186/s12985-020-01411-0
- IHME COVID-19 Health Service Utilization Forecasting Team and Murray, Christopher JL. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by us state in the next 4 months. *medRxiv* [Preprint]. (March 2020).
- Friedman J, Liu P, Troeger CE, Carter A, Reiner Jr RC, Barber RM, et al. Predictive performance of international COVID-19 mortality forecasting models. *Nat Commun*. (2021) 12(1):2609. doi: 10.1038/s41467-021-22457-w
- IHME. COVID-19: what's new for May 4, 2020: main updates on IHME COVID-19 predictions since April 29, 2020 (2020). Available online at: <https://www.healthdata.org/research-analysis/diseases-injuries/covid> (Accessed June 12, 2024).

Author contributions

TS: Writing – original draft, Writing – review & editing. AB: Writing – original draft, Writing – review & editing. JW: Writing – original draft, Writing – review & editing. DP: Writing – original draft, Writing – review & editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

20. IHME. COVID-19: what's new for May 29, 2020: main updates on IHME COVID-19 predictions since May 26, 2020 (2020). Available online at: <https://www.healthdata.org/research-analysis/diseases-injuries/covid> (Accessed June 12, 2024).
21. Spencer Woody MT, Dahan M, Gaither K, Lachmann M, Fox SJ, Meyers LA, et al. Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones (2020). Available online at: <https://covid-19.tacc.utexas.edu> (Accessed June 12, 2024).
22. Melikechi O, Young AL, Tang T, Bowman T, Dunson D, Johndrow J. Limits of epidemic prediction using SIR models. *J Math Biol.* (2022) 85(4):36. doi: 10.1007/s00285-022-01804-5
23. Britton T, Tomba GS. Estimation in emerging epidemics: biases and remedies. *J R Soc Interface.* (2019) 16(150):20180670. doi: 10.1098/rsif.2018.0670
24. Lemoine P, *The Case Against Lockdowns*. Austin, TX: Center for the Study of Partisanship and Ideology (2021).
25. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.* (2020) 584(7820):257–61. doi: 10.1038/s41586-020-2405-7
26. Meyers LA, Dimitrov N. *Mathematical Approaches to Infectious Disease Prediction and Control*. INFORMS, Tutorials in Operations Research (2010).
27. Del Valle SY, Hyman JM, Chitnis N. Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases. *Math Biosci Eng.* (2013) 10:1475. doi: 10.3934/mbe.2013.10.1475
28. Chen S, Robinson P, Janies D, Dulin M. Four challenges associated with current mathematical modeling paradigm of infectious diseases and call for a shift. *Open Forum Infect Dis.* (2020) 7(8):ofaa333. doi: 10.1093/ofid/ofaa333
29. Dimitrov NB, Meyers LA. Mathematical approaches to infectious disease prediction and control. In *Risk and Optimization in an Uncertain World*. INFORMS (2010). p. 1–25. doi: 10.1287/educ.1100.0075
30. Katul GG, Mrad A, Bonetti S, Manoli G, Parolari AJ. Global convergence of COVID-19 basic reproduction number and estimation from early-time SIR dynamics. *PLoS One.* (2020) 15(9):e0239800. doi: 10.1371/journal.pone.0239800
31. Ding X, Huang S, Leung A, Rabbany R. Incorporating dynamic flight network in SEIR to model mobility between populations. *Appl Network Sci.* (2021) 6(1):1–24. doi: 10.1007/s41109-021-00378-3
32. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis.* (2020) 20(5):553–8. doi: 10.1016/S1473-3099(20)30144-4
33. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet.* (2020) 395(10225):689–97. doi: 10.1016/S0140-6736(20)30260-9
34. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science.* (2020) 368(6489):395–400. doi: 10.1126/science.aba9757
35. Chowell G, Nishiura H. Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC Med.* (2014) 12:1–17. doi: 10.1186/s12916-014-0196-0
36. Wang X, Pasco R, Pierce K, Du Z, Fox S, Meyers LA. *COVID-19 Healthcare Demand Projections: Austin, Texas*. 19 Modeling Consortium (2020).
37. Tec M, Lachmann M, Fox SJ, Pasco R, Woody S, Starling J, et al. Austin COVID-19 transmission estimates and healthcare projections (2020). Available online at: <https://covid-19.tacc.utexas.edu> (Accessed June 12, 2024).
38. Liu X. Analytical solution of a new SEIR model based on latent period-infectious period chronological order. *medRxiv* [Preprint]. (2021).
39. Patterson B, Wang J. How does the latency period impact the modeling of COVID-19 transmission dynamics? *Math Appl Sci Eng.* (2022) 3(1):60–85. doi: 10.5206/mase/14537
40. Pereira IG, Silva Jr AGD, Aragão DP, de Oliveira EV, Bezerra AA, Pereira FDA, et al. Epidemiology forecasting of COVID-19 using AI—a survey. In *Computational Intelligence for COVID-19 and Future Pandemics: Emerging Applications and Strategies*. Springer (2022). p. 89–120.