Research Article
Implications of Delayed Reopening in Controlling the COVID-19 Surge in Southern and West-Central USA

Raj Dandekar,1 Emma Wang,2 George Barbastathis,3,4 and Chris Rackauckas5

1Department of Computational Science and Engineering, Massachusetts Institute of Technology, Cambridge, USA MA 02139
2Department of Electrical Engineering and Computer Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
3Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
4Singapore-MIT Alliance for Research and Technology (SMART) Centre, Singapore 138602
5Department of Applied Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Correspondence should be addressed to Raj Dandekar; raj.dandekar8@gmail.com

Received 16 November 2020; Accepted 12 September 2021; Published 28 October 2021

Copyright © 2021 Raj Dandekar et al. Exclusive Licensee Peking University Health Science Center. Distributed under a Creative Commons Attribution License (CC BY 4.0).

In the wake of the rapid surge in the COVID-19-infected cases seen in Southern and West-Central USA in the period of June-July 2020, there is an urgent need to develop robust, data-driven models to quantify the effect which early reopening had on the infected case count increase. In particular, it is imperative to address the question: How many infected cases could have been prevented, had the worst affected states not reopened early? To address this question, we have developed a novel COVID-19 model by augmenting the classical SIR epidemiological model with a neural network module. The model decomposes the contribution of quarantine strength to the infection time series, allowing us to quantify the role of quarantine control and the associated reopening policies in the US states which showed a major surge in infections. We show that the upsurge in the infected cases seen in these states is strongly correlated with a drop in the quarantine/lockdown strength diagnosed by our model. Further, our results demonstrate that in the event of a stricter lockdown without early reopening, the number of active infected cases recorded on 14 July could have been reduced by more than 40% in all states considered, with the actual number of infections reduced being more than 100,000 for the states of Florida and Texas. As we continue our fight against COVID-19, our proposed model can be used as a valuable asset to simulate the effect of several reopening strategies on the infected count evolution, for any region under consideration.

1. Background

The Coronavirus respiratory disease 2019 originating from the virus "SARS-CoV-2" [1, 2] has led to a global pandemic, leading to more than 50 million confirmed global cases in more than 200 countries as of November 13, 2020 [3]. In the United States, the first infections were detected in Washington State as early as January 20, 2020 [4], and now, it is being reported that the virus had been circulating undetected in New York City as early as mid-February [5]. As of September 21, 2020, the United States has ≈6.9 million infected cases since the virus began to spread.

Since the second week of June, a second surge of COVID-19 was seen in the United States [6], with rapidly increasing daily infected cases, hospitalization rates, and death rates [7, 8]. Initially driven by disastrous situations in the states of Arizona, South Carolina, Texas, Florida, and Georgia [6], the surge in cases was also later seen in several other Southern and West-Central states [9]. This surge can be seen in Figure 1 which shows the active infected cases over time as of July 14, 2020, with a 7-day moving average for 9 states. States which reopened early show a generally strong correlation with the rise in the infected cases over the 3-month period from late April to mid-July 2020 [9]. For example, states which opened before May 15 showed daily infected case increments as follows: Florida (1393%), Arizona (858%), South Carolina (999%), Alabama (547%), Oklahoma (477%), Tennessee (279%), Georgia (245%), Mississippi (215%), Nevada (697%), Texas (680%), and Utah (287%), while states which reopened after May 29 showed values as follows: Michigan (16%), Pennsylvania (~26%), New York (~52%), New Jersey (~32%), and Illinois (~54%). Thus, although early reopening seems to be correlated to the second surge of cases seen in the USA, there is a need...
Unbounded activation functions in particular, such as the rectified linear unit (ReLU), have been known to be effective in approximating nonlinear functions with a finite set of parameters [15–17]. Thus, a neural network solution is attractive to approximate quarantine effects in combination with analytical epidemiological models. The downside is that the internal workings of a neural network are difficult to interpret. The recently emerging field of scientific machine learning [18] exploits conservation principles within a universal differential equation [19], SIR in our case, to mitigate overfitting and other related machine learning risks.

In the present work, the neural network is trained from publicly available infection and population data for COVID-19 for each state under study.

2.2. Standard SIR Model. The SIR (Susceptible-Infected-Recovered) is governed by the following set of ODEs:

\[
\frac{dS}{dt} = -\frac{\beta S(t)I(t)}{N},
\]

\[
\frac{dI}{dt} = \frac{\beta S(t)I(t)}{N} - \gamma I(t),
\]

\[
\frac{dR}{dt} = \gamma I(t),
\]

where \( \beta \) and \( \gamma \) are the contact and recovery rates, respectively. We use this framework as our baseline model to be augmented with a neural network module. We do not consider the possibility of recovered individuals being reinfection [20]. We also do not consider the waning of immunity associated with COVID-19 as discovered in recent studies [21]. Here, \( \beta \) is the infection rate and \( \gamma \) is the recovery rate, and they are assumed to be constant in time. The total population \( N = S(t) + I(t) + R(t) \) is seen to remain constant as well; that is, births and deaths are neglected. The recovered population is to be interpreted as those who can no longer infect others, so it also includes individuals who are deceased due to the infection. The possibility of recovered individuals to become reinfected is accounted for by SEIS models [20], but we do not use this model here, as the reinfection rate for COVID-19 survivors is considered to be negligible as of now.

An important assumption of the SIR models is homogeneous mixing among the subpopulations. Therefore, this model cannot account for social distancing or social network effects. Additionally, the model assumes uniform susceptibility and disease progression for every individual, and that no spreading occurs through animals or other nonhuman means. Alternatively, the SIR model may be interpreted as quantifying the statistical expectations on the respective mean populations, while deviations from the model’s assumptions contribute to statistical fluctuations around the mean.

2.3. QSIR Model: ODE Formulation. The QSIR ODE model formulation is similar to the one studied previously [22] for robust, data-driven quantification of the effect of early reopening on the growth of infected count data. More importantly, it is of utmost importance to answer the question: How many infected cases could have been prevented, had the worst affected states not reopened early?

In an effort to address this question, we have developed a machine learning-aided epidemiological model. The novelty of our model arises from the fact that it allows us to decompose the contribution of quarantine/lockdown strength evolution to the infected data time series for the region under consideration. This enables us to simulate the effect of varying quarantine strength evolutions and hence varying reopening strategies on the infected count data. We define reopening as beginning when a state allows its stay-at-home order to expire or, in the case of states that never issued a stay-at-home order, when a state first starts allowing nonessential businesses, such as dine-in restaurants and hair salons, to reopen [10, 11]. The reopening details for the states considered in the study are shown in Table 1. Considering nine US states which showed a significant surge in cases since the last month, we demonstrate that our model shows a drop in the quarantine strength evolution when these states were reopened. Furthermore, we show that maintaining a strict lockdown without early reopening would have led to about 500,000 fewer infected cases in all these states combined.

2. Methods

2.1. QSIR Model. In general, neural networks with arbitrary activation functions are universal approximators [12–14].
and is briefly explained in this section. The equations governing the QSIR model are as follows:

\[
\frac{dS}{dt} = -\frac{\beta S(t)I(t)}{N},
\]

\[
\frac{dI}{dt} = \frac{\beta S(t)I(t)}{N} - (\gamma + Q(t))I(t)
= \frac{\beta S(t)I(t)}{N} - (\gamma + NN(W, U))I(t),
\]

\[
\frac{dR}{dt} = \gamma I(t) + \delta T(t),
\]

\[
\frac{dT}{dt} = Q(t)I(t) - \delta T(t) = NN(W, U)I(t) - \delta T(t).
\]

The SIR model is augmented by introducing a time-varying quarantine strength rate term \(Q(t)\) represented by a neural network [19] and a quarantined population \(T(t)\), which is prevented from having any further contact with the susceptible population. Thus, the term \(I(t)\) denotes the active infected population (Actively infected = Cumulative infected – Recovered) still having contact with the susceptibles, as done in the standard SIR model, while the term \(T(t)\) denotes the infected population who are effectively quarantined and isolated. Thus, we can write an expression for the quarantined infected population \(T(t)\) as

\[
\frac{dT}{dt} = Q(t)I(t) - \delta T(t).
\]

Since \(Q(t)\) does not follow from first principles and is highly dependent on local quarantine policies, we devised a neural network-based approach to approximate it. Recently, it has been shown that neural networks can be used as function approximators to recover unknown constitutive relationships in a system of coupled ordinary differential equations [19, 23]. Following this principle, we represent \(Q(t)\) as an \(n\) layer-deep neural network with weights \(W_1, W_2, \ldots, W_n\), activation function \(r\), and the input vector \(U = (S(t), I(t), R(t))\) as

\[
Q(t) = r(W_nr(W_{n-1} \cdots r(W_1U))).
\]

For the implementation, we choose a \(n = 2\)-layer densely connected neural network with 10 units in the hidden layer and the leaky ReLU activation function. This choice was because we found sigmoidal activation functions to stagnate. The final model was described by 54 tunable parameters. The neural network architecture schematic is shown in Figure 3(b). The governing coupled ordinary differential equations for the QSIR model are

\[
\frac{dS}{dt} = -\frac{\beta S(t)I(t)}{N},
\]

\[
\frac{dI}{dt} = \frac{\beta S(t)I(t)}{N} - (\gamma + Q(t))I(t)
= \frac{\beta S(t)I(t)}{N} - (\gamma + NN(W, U))I(t),
\]

\[
\frac{dR}{dt} = \gamma I(t) + \delta T(t),
\]

\[
\frac{dT}{dt} = Q(t)I(t) - \delta T(t) = NN(W, U)I(t) - \delta T(t).
\]
Figure 2: For the states of Arizona, Nevada, South Carolina, and Tennessee, the figure shows (a, d, g, and j) model recovery of infected and recovered case count as of 14 July 2020. (b, e, h, and k) Quarantine strength function as discovered by our trained model (with reopening). This is shown along with the quarantine strength function which we use to simulate strict quarantine without reopening after stay-at-home order was imposed. (c, f, i, and l) Estimated infected count if strict quarantine and lockdown measures were followed without reopening (5% and 95% quantiles are shown) as compared to the values corresponding to the actual early reopening scenario.
The starting number of recovered individuals was initialized from data (available here). Although the chemical master equation cannot be solved analytically, under certain conditions, it can be distilled down to a stochastic differential equation (SDE) which captures the fluctuations in species transfer as random walks. Such an SDE, also known as the chemical Langevin Equation, is thus based on the underlying ODE framework (macroscopic picture) and also includes stochastic effects reminiscent of microscopic modelling. In fact, in the Supplementary Information, we show that the microscopic simulation, macroscopic ODE formulation, and chemical Langevin equation (which acts as a bridge between the two) are all equivalent to each other.

The equivalent stochastic formulation or the chemical Langevin equation for the augmented SIR model is

\[
    dS = -\left[\frac{\beta S(t)I(t)}{N}\right]dt - \sqrt{\frac{\beta S(t)I(t)}{N}}dW_i(t),
\]

and \(\delta\) were optimized by minimizing the loss function given in (14). For states with a low recovered count: Arizona, Florida, Nevada, and Texas, we employed a two-stage optimization procedure to find the optimal \(W, \beta, \gamma, \text{ and } \delta\). In the first stage, (14) was minimized. For the second stage, we fix the optimal \(\gamma\) and \(\delta\) found in the first stage to optimize for the remaining parameters: \(W, \beta\) based on the loss function defined just on the infected count as \(L(W, \beta) = \|\log (I(t) + T(t)) - \log (I_{\text{data}}(t))\|^2\). Such an approach was found to be optimal for analyzing low recovered count data in previous studies [22].

In all states considered in the present study, we trained the model using data starting from the dates when the 500th infection was recorded in each region and up to July 14, 2020. For each state considered, \(Q(t)\) denotes the rate at which infected persons are effectively quarantined and isolated from the remaining population and thus gives composite information about (a) the effective testing rate of the infected population as the disease progresses and (b) the intensity of the enforced quarantine as a function of time.

This QSIR ODE framework applied on the infected and recovered data is used to estimate the quarantine strength function \(Q(t)\) in a particular state as shown in the first and second columns of Figure 2.

2.6. QSIR Model: SDE Formulation. The ODE modelling framework described above is a deterministic approach to model transfer of species (here: people) from one compartment to another through different reaction channels. Such a deterministic approach ignores any random fluctuations during species transfer from one compartment to the other. To include such stochastic effects and thus get a measure of the model uncertainty, we note that the augmented SIR framework derives from the chemical master equation which describes the time evolution of the probability of such a system of interacting species to be in a given state at a given time (details in Supplementary Information (available here)). Although the chemical master equation cannot be solved analytically, under certain conditions, it can be distilled down to a stochastic differential equation (SDE) which captures the fluctuations in species transfer as random walks. Such an SDE, also known as the chemical Langevin Equation, is thus based on the underlying ODE framework (macroscopic picture) and also includes stochastic effects reminiscent of microscopic modelling. In fact, in the Supplementary Information, we show that the microscopic simulation, macroscopic ODE formulation, and chemical Langevin equation (which acts as a bridge between the two) are all equivalent to each other.
Table 2: Drop in quarantine strength function, $Q(t)$, after reopening as discovered by our trained model. $Q(t)$ trained by our model shows a significant drop for all Southern and West-Central states which showed a surge in cases from reopening, whereas the North-Eastern states which showed no surge do not see a drop in $Q(t)$.

<table>
<thead>
<tr>
<th>State</th>
<th>Reopening date</th>
<th>% increase in daily cases since reopening</th>
<th>Maximum % decrease in $Q(t)$ after reopening</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Arizona</td>
<td>May 15</td>
<td>+858</td>
<td>+22</td>
</tr>
<tr>
<td>(2) Florida</td>
<td>May 4</td>
<td>+1393</td>
<td>+10</td>
</tr>
<tr>
<td>(3) Louisiana</td>
<td>May 15</td>
<td>+193</td>
<td>+30</td>
</tr>
<tr>
<td>(4) Nevada</td>
<td>May 9</td>
<td>+697</td>
<td>+25</td>
</tr>
<tr>
<td>(5) Oklahoma</td>
<td>April 24</td>
<td>+477</td>
<td>+29</td>
</tr>
<tr>
<td>(6) South Carolina</td>
<td>May 4</td>
<td>+999</td>
<td>+71</td>
</tr>
<tr>
<td>(7) Tennessee</td>
<td>April 30</td>
<td>+279</td>
<td>+44</td>
</tr>
<tr>
<td>(8) Texas</td>
<td>May 1</td>
<td>+680</td>
<td>+29</td>
</tr>
<tr>
<td>(9) Utah</td>
<td>May 1</td>
<td>+287</td>
<td>+39</td>
</tr>
<tr>
<td>(10) New York</td>
<td>May 29</td>
<td>-52</td>
<td>-45</td>
</tr>
<tr>
<td>(11) New Jersey</td>
<td>June 9</td>
<td>-32</td>
<td>-60</td>
</tr>
<tr>
<td>(12) Illinois</td>
<td>May 29</td>
<td>-54</td>
<td>-8</td>
</tr>
</tbody>
</table>

\[
dI = \left[\frac{\beta S(t)I(t)}{N} - \gamma I(t) - Q(t)I(t)\right]dt + \sqrt{\frac{\beta S(t)I(t)}{N}}dW_1(t) - \sqrt{Q(t)I(t)}dW_2(t) - \sqrt{Q(t)I(t)}dW_3(t),
\]

(16)

\[
dR = [\gamma I(t) + \delta T(t)]dt + \sqrt{\gamma I(t)}dW_2(t) + \sqrt{\delta T(t)}dW_4(t),
\]

(17)

\[
dT = [Q(t)I(t) - \delta T(t)]dt + \sqrt{Q(t)I(t)}dW_3(t) - \sqrt{\delta T(t)}dW_4(t).
\]

(18)

In (15), $W_i(t) \sim N(0, t)$ is a normally distributed random variable with mean zero and variance $t$ or $dW_i(t) \sim N(0, dt)$. It should also be noted that each $W_i(t)$ represents an independent Brownian motion. The simulations were performed using the Catalyst.jl software in Julia using the LambdaEM algorithm based on [26]. 1000 trajectories were simulated for each state.

This QSIR SDE framework along with the simulated quarantine functions for no reopening is used to predict the new infected case count and hence estimate the reduction in the number of infected cases under the simulated no-reopening quarantine function. The results are shown as 5% and 95% quantiles in the third column of Figure 2.

2.7. Mean Absolute Percentage Error. The Mean Absolute Percentage Error (MAPE) is defined as

\[
MAPE = \frac{100}{N} \sum \frac{|I(t) + T(t) + R(t)| - |I_{data}(t) + R_{data}(t)|}{|I_{data}(t) + R_{data}(t)|},
\]

(19)

where $N$ is the number of observations.

3. Results

The first stage of our analysis is using our model [22], called the QSIR model to diagnose the underlying quarantine strength evolution $Q(t)$ in the regions under consideration. By applying the QSIR model to more than 70 countries globally, we have established the validity of $Q(t)$ in accurately diagnosing the on-the-ground quarantine situation in majorly affected European, South American, and Asian countries [22]. A slow growth of $Q(t)$ without a significant increase indicates relaxed quarantine policies, a sharp transition point in $Q(t)$ is indicative of a sudden ramp-up of quarantine measures, and an inflection point corresponds to the time when the quarantine response was the most rapid in the region under consideration. The results of our model applied globally to all continents are hosted publicly at http://covid19ml.org.

In this study, to perform the quarantine diagnosis to analyze the implications of delayed reopening, we applied the QSIR model to 9 US states which showed a significant surge in the infected case count in the last month: Arizona, Florida, Louisiana, Nevada, Oklahoma, South Carolina, Tennessee, Texas, and Utah. Figure 2 shows representative results for Arizona, Nevada, South Carolina, and Tennessee. The plots for the remaining states are provided in the Supplementary Information. Figures 2(a), 2(d), 2(g), and 2(j) show the comparison of the infected and recovered count estimated by our model with the actual data. A reasonable agreement is seen for all states, with the model being able to capture the rise in infections seen in the tail end of the time series. The QSIR model details are provided in Methods; Mean Absolute Percentage Error (MAPE) values for the model along with the epochs required for convergence for each state are provided in Supplementary Information.

Figures 2(b), 2(e), 2(h), and 2(k) show the quarantine strength evolution $Q(t)$ as learnt by the neural network module, which shows a decline whose starting point corresponds well to the time when these states began reopening, as seen from Table 2 and the green dotted line in
or because of the time delay for population-level changes corresponding to the Phase 2 or Phase 3 of reopening (Table 2) in Eastern states of New York, New Jersey, and Illinois, which showed a surge in cases last month, whereas the North-Eastern states of New York, New Jersey, and Illinois, which reopened late and showed no surge in infections, did not show a drop in $Q(t)$ (Table 3 and figures in Supplementary Information). Thus, the upsurge in the infected cases seen in these states is strongly correlated with a drop in the quarantine/lockdown strength $Q(t)$ diagnosed by our model. This is indicative of two things: (a) the Southern and West-Central states reopened early, which led to a relaxed imposition of quarantine/lockdown measures in these states and consequently a surge in infections was seen, and (b) the North-Eastern states of New York, New Jersey, and Illinois reopened late, and even after reopening, a relatively low contact rate was maintained among the population, leading to a relatively high magnitude of the imposed quarantine strength, which prevented a surge of infections in these states. The percentage decrease in quarantine strength observed after reopening for all states considered is shown in Table 3. It should be noted that for North-Eastern states which did not show a surge of infections last month, such as New York and New Jersey, such a drop in $Q(t)$ is not seen (figures in Supplementary Information). This indicates that the surge in infections, predominantly seen in the Southern and West-Central states, was caused by an early reopening which led to a relaxed imposition of quarantine/lockdown measures in these states.

To further demonstrate the validity of our model in capturing the actual quarantine policy evolution in a particular region, the model has been applied to 70 countries globally. The quarantine strength behaviour learnt from the model accurately mimics the on-the-ground situation in majorly affected European, South American, and Asian countries. The results of our model applied globally to all continents are hosted publicly at http://covid19mll.org.

After confirming that our model is able to accurately depict the correlation between the surge in infections and early reopening in these states through the diagnosed $Q(t)$, we proceed to the second stage of our analysis. In the second stage, we use the diagnosed $Q(t)$ to address the question: How many infected cases would have been reduced, had the worst affected states not reopened early? To answer this question, we simulate the “no-reopening” strategy by assuming that $Q(t)$ is maintained at the value it was before reopening, without decreasing. This simulated $Q(t)$ is shown in Figures 2(b), 2(e), 2(h), and 2(k). The flexibility of our model allows us to run our model with this simulated $Q(t)$ for all states considered. To quantify the aleatory uncertainty resulting from random fluctuations in the model, we utilized the chemical Langevin equation extension to the QSIR model whose definition and justification are described in Methods and Supplemental Information. This allows us to estimate bootstrapped confidence intervals resulting from 1000 simulations of such a stochastic model and thus quantify the effect of such a “no-reopening policy” on the epidemic spread. The infected count evolution for the simulated $Q(t)$ without reopening is shown in Figures 2(c), 2(f), 2(i), and 2(l) (% and 95% quantiles are shown). We can see that, for all these states, instead of seeing a spike in infections, we would have seen a plateau in the infected case count evolution. The number and the percentage of infected cases that would have been prevented by July 14 had these states not reopened are shown in Table 3. It is evident that the number of infections could have been reduced by more than 40% in all states considered, with the actual number of infections reduced being more than 100,000 for the states of Florida and Texas. Even the less populated states of Louisiana, South Carolina, and Tennessee show mean infected case reduction values of 44%, 84%, and 47%, respectively, which correspond to 36,000, 51,000, and 31,000 infected cases reduced.

4. Conclusion

In this study, we have developed a novel methodology to quantify the effect of early reopening on the infected case count surge seen during the period of June-July 2020. We have proposed a machine learning model, called the QSIR model, rooted firmly in fundamental epidemiology.

Table 3: Infected count reduction by 14 July 2020, if states had not reopened early, as estimated by our model.

<table>
<thead>
<tr>
<th>State</th>
<th>% decrease (5%-95% quantiles)</th>
<th>Mean % decrease</th>
<th>Case reduction</th>
<th>Mean case reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Arizona</td>
<td>35-62</td>
<td>49</td>
<td>44000-79000</td>
<td>63000</td>
</tr>
<tr>
<td>(2) Florida</td>
<td>20-75</td>
<td>49</td>
<td>57000-218000</td>
<td>144000</td>
</tr>
<tr>
<td>(3) Louisiana</td>
<td>37-50</td>
<td>44</td>
<td>31000-41000</td>
<td>36000</td>
</tr>
<tr>
<td>(4) Nevada</td>
<td>32-68</td>
<td>51</td>
<td>100000-20000</td>
<td>15000</td>
</tr>
<tr>
<td>(5) Oklahoma</td>
<td>46-69</td>
<td>58</td>
<td>100000-15000</td>
<td>13000</td>
</tr>
<tr>
<td>(6) South Carolina</td>
<td>83-86</td>
<td>84</td>
<td>50000-52000</td>
<td>51000</td>
</tr>
<tr>
<td>(7) Tennessee</td>
<td>41-53</td>
<td>47</td>
<td>27000-36000</td>
<td>31000</td>
</tr>
<tr>
<td>(8) Texas</td>
<td>41-51</td>
<td>46</td>
<td>115000-143000</td>
<td>129000</td>
</tr>
<tr>
<td>(9) Utah</td>
<td>35-47</td>
<td>41</td>
<td>110000-14000</td>
<td>12000</td>
</tr>
</tbody>
</table>
principles which has the following attributes: (a) it is highly
interpretable with few free parameters rooted in an epidemi-
ological model, (b) it relies on only COVID-19 data and not
on previous epidemics, and (c) it can decompose the
infected time-series data to reveal the quarantine strength/
variation, $Q(t)$, in the region under consideration.
To demonstrate the validity of our model in capturing the
actual quarantine policy evolution in a particular region,
the model has been applied to 70 countries globally. The
quarantine strength behaviour learnt from the model ac-
curately mimics the on-the-ground situation in majorly
affected European, South American, and Asian continents.
The results for this global analysis are hosted at http://
covid19ml.org [22].

After confirming our belief in the model through a global
analysis, we apply the model to the Southern and West-
Central US states which have shown a massive surge in
COVID-19-infected cases since June 2020. We demonstrate
that the $Q(t)$ extracted by our model shows a significant drop
in value for the Southern and West-Central states which reo-
pened early and showed a surge in infections. The time at
which $Q(t)$ starts to decline generally agrees well with the
reopening date for the states considered. Since the decline in
$Q(t)$ is strongly correlated to the surge of infections and also
the reopening date for states which reopened early, we can
then simulate the effect of “no-reopening” by maintaining
the $Q(t)$ at a constant level after reopening, instead of declin-
ing. We show that maintaining a steady imposition of quaran-
tine/lockdown control would have played a massive role in
bring down the infected count by more than 40% in all
states considered, with the infections reduced reaching more
than 100,000 for the states of Florida and Texas.

We have proposed a novel machine learning methodol-
yogy, rooted in fundamental epidemiological models, which
is able to recover the real-time quarantine strength evolution
for any region under consideration. As the pandemic evol-
ues and we continue our fight against COVID-19, and for future
outbreaks, our globally applicable methodology can be a
valuable asset for researchers and policymakers to simulate
several reopening strategies and counterfactual scenarios and
analyze their impact on the infected count evolution.
Our findings highlight that as we continue the fight against
COVID-19, it is imperative to reduce the contact between suscep-
tible and infected individuals in public places by for-
mulating robust safety guidelines. Such guidelines imple-
mented and maintained in the affected states would ensure
a high level of quarantine strength associated with that state
and can prevent a future surge or wave in the COVID-19-
infected count time series.

Validation of the model robustness and parameter iden-
tifiability have been mentioned in the Supplementary Infor-
mation. We have also compared an equivalent of the
effective reproduction number called the COVID spread
parameter in our study, with other studies to further validate
the results of our modelling approach. The COVID spread
parameter is defined by (a) the infected individuals and (b)
the recovered individuals from both the infected and the
quarantined states, since both of those effectively do not fur-
ther contribute to the infection spread [22].

The results of our model should be taken in the context
of its assumptions. Ideally, one needs to consider the shifting
US testing policies for the time period under consideration.
Since the testing efforts did not show a significant increase
during and after the reopening in the US states in the time
period considered within the present study [27, 28] and we
did not want to burden our model with additional parame-
ters to fit, testing compartments have not been included in
the present study. Additionally, several studies in literature
[29–32] have attempted to incorporate underreporting of
infected/recovered cases in their modelling paradigm. Most
of these studies use previously known estimates of testing
data, serology data, or Infection-Fatality-Rate (IFR). In these
studies involving multiple parameters, a number of param-
ters are assumed to be fixed at the start of the simulation
from prior studies. These parameters include and are not
limited to time between onset of infections and symptoms,
transmission duration, rate at which hospitalized patients
recover [32], mean duration from symptom onset to recov-
ery [29], or even the IFR ratio [29]. A second class of studies
uses antibody testing from collected serum samples to esti-
mate the actual number of infected cases [33].

As the pandemic unfolds and starts spreading, the first
information available is the number of infected, recovered,
and deaths (for example, the Johns Hopkins public reposi-
tory for COVID-19 tracking). Unless we have serum sample
data information or we can confidently rely on prior studies
for assessment of certain parameters, accurate information
of the underreporting factor is difficult to obtain in real time.
One of the goals of the present modelling methodology is to
assist researchers and policymakers with quarantine diagno-
sis information in real time, with no reliance on parameters
derived from prior studies.

Finally, the model is based on the SIR framework, which
assumes a constant, age-independent contact and recovery
rate between the infected and susceptible populations. Ad-
ditionally, we do not consider the spatial heterogeneity in the
infected count within a particular state and assume the gov-
erning dynamics to be only time-dependent. Consideration
of these second-order aspects would further refine the model
and would be the subject of future studies.

Determining the optimal reopening policy for different
states is a composite challenge depending on a wide range
of social, economic, and political factors beyond the scope
of the present study. Our results show that irrespective of
these factors and their role in influencing the reopening pol-
icy, it is imperative to reduce the contact rate between
infected and susceptible individuals, thereby maintaining
or increasing the quarantine strength. When a state reopens
its public spaces like restaurants, bars, schools, and cinema
halls, the state reduces its quarantine strength, and even a
small drop in this number can be enough to lead to a mas-
sive surge in the infected count. When a state has to reopen
due to socioeconomic or political factors, it should do so
with the utmost care and with detailed guidelines for reduc-
ing the contact rate as much as possible in schools, child care
programs, offices, restaurants, bars, and vehicles of mass
transit. This aligns well with the COVID-related safety
guidelines issued by the CDC [34].
Data Availability

Data for the infected and recovered case count in all regions was obtained from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. All code files and results are publicly available at https://github.com/RajDandekar/Reopening_ImpactSimulator_US_States.

Additional Points

Resource Availability. Lead contact is Raj Dandekar, MIT: Email: rajd@mit.edu.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors’ Contributions


Acknowledgments

This effort was partially funded by the Intelligence Advanced Research Projects Activity (IARPA). We are grateful to Haluk Akay, Hyungseok Kim, and Wujie Wang for helpful discussions and suggestions.

Supplementary Materials


References


