Methodology

Last updated: May 07, 2020

covid19sim@mgh.harvard.edu

Massachusetts General Hospital Institute for Technology Assessment
Harvard Medical School
101 Merrimac St. STE 1010 • Boston, MA 02114 • United States

Model Overview

The *COVID-19 Simulator* uses a validated compartment model to simulate the trajectory of COVID-19 at the state level from March 15, 2020 to August 31, 2020 in the United States. Utilizing the most recent reported data for each state, the *COVID-19 Simulator* considers state-specific disease spread dynamics. Specifically, to reproduce the observed trends and project future cases of COVID-19, time-varying and state-specific effective reproductive numbers are estimated using curve fitting algorithms and fed as inputs into a compartment model. The compartment model is defined using Susceptible, Exposed, Infectious, and Recovered compartments (i.e., SEIR model) with continuous time progression. Model programming and analysis were performed in R (version 3.6.2), and the package deSolve was used to solve the ordinary differential equation system.

The *COVID-19 Simulator* evaluates the impact of different non-pharmaceutical intervention strategies to reduce the spread of COVID-19 under varying intensity and timing at the state and national level. For each selected strategy, the model projects and visualizes the total number of deaths from COVID-19, daily counts of cases, cumulative number of cases, number of active cases, and the number of hospital beds and intensive care unit (ICU) beds needed for COVID-19 patients.

Model Structure

For each state, we developed a system dynamics model, also known as a compartment model,[1] to project the trajectory of the COVID-19 pandemic. The model is defined using Susceptible, Exposed, Infectious, and Recovered compartments (i.e., SEIR model) as defined in Figure 1.

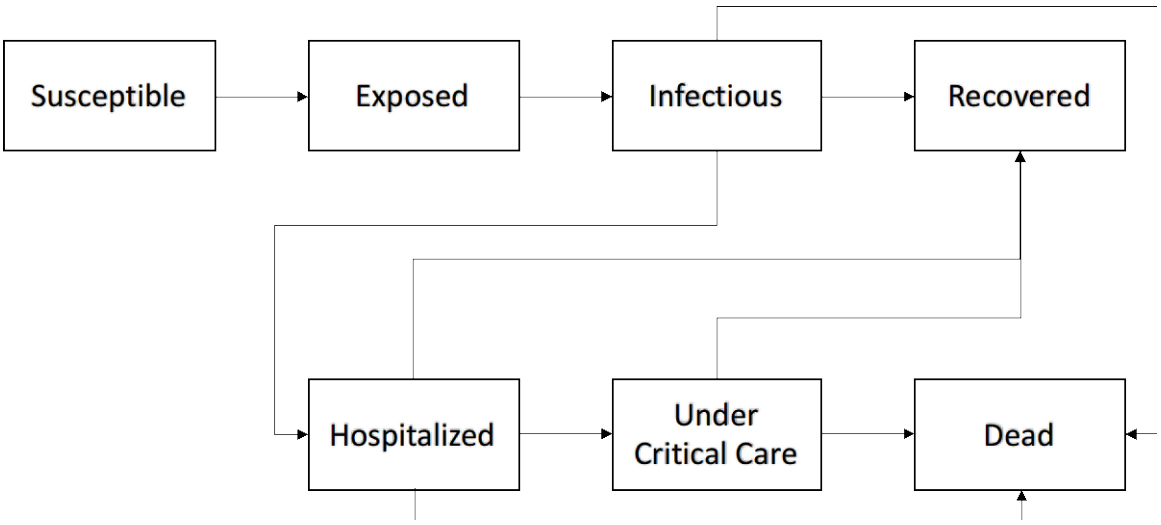

Figure 1. Schematic of SEIR model of COVID-19

The transmission force of the epidemic was controlled by a parameter called *effective reproductive number* (R_E) of the model. The effective reproduction number is defined as the average number of secondary cases from a single infected individual under a specific level of interventions. To reproduce the observed trends of COVID-19 cases in each state, we allowed the effective reproductive number to be a function of time to account for time-trends and the effects from various interventions (e.g., social distancing) over time on the spread of COVID-19. Specifically, we considered a stepwise function of time for the value of the reproduction number as defined below:

$$R_{\mathrm{E}}(t) = \begin{cases} R_0 & \text{if } 0 < t \leq \tau_1, \\ R_E^1 & \text{if } \tau_1 < t \leq \tau_2, \\ R_E^2 & \text{if } \tau_2 < t \leq \tau_3, \\ R_E^3 & \text{if } t > \tau_3, \end{cases}$$

where R_0 (a.k.a., basic reproductive number— the value under no implemented intervention), R_E^1, R_E^2 and R_E^3 are the time-varying values of $R_{\mathrm{E}}(t)$ over time and $\tau_1,\ \tau_2,\ \tau_3$ are the time points at which the reproduction number takes a new value. The values of τ_1, τ_2, τ_3 are determined using a data-driven approach based on a joinpoint regression method as implemented in the JPSurv package in R.[2]

For each state, the spread of COVID-19 is modeled based on the following set of differential equations:

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{R_E}{p_I} \cdot \frac{S \cdot I}{\mathrm{S}(0) - D} \\
\frac{dE}{dt} &= \frac{R_E}{p_I} \cdot \frac{S \cdot I}{\mathrm{S}(0) - D} - \frac{E}{p_E} \\
\frac{dI}{dt} &= \frac{E}{p_E} - \frac{I}{p_I} \\
\frac{dtoH}{dt} &= \mathrm{r_H} \cdot \frac{E}{p_E} - \frac{toH}{p_{toH}} \\
\frac{dHNCC}{dt} &= (1 - \mathrm{r_{CC}}) \cdot \frac{toH}{p_{toH}} - \frac{HNCC}{p_{HNCC}} \\
\frac{dHCC}{dt} &= \mathrm{r_{CC}} \cdot \frac{toH}{p_{toH}} - \frac{HCC}{p_{HCC}} \\
\frac{dCC}{dt} &= \mathrm{r_{CC}} \cdot \frac{HCC}{p_{HCC}} - \frac{HCC}{p_{CC}} \\
\frac{dD}{dt} &= \mathrm{CFR_{CC}} \cdot \frac{HCC}{p_{CC}} \\
\frac{dR}{dt} &= (1 - \mathrm{CFR_{CC}}) \cdot \frac{HCC}{p_{CC}} + (1 - \mathrm{r_H}) \cdot \frac{I}{p_I}
\end{aligned}$$

where:

$S(t)$: Number of susceptible individuals at time *t*
$S(0)$: Initial population of the state
$E(t)$: Number of exposed (latent) individuals at time *t*
$I(t)$: Number of diagnosed infectious individuals at time *t*
$toH(t)$: Number of individuals to be hospitalized at time *t*
$HNCC(t)$: Number of non-critical care individuals under hospitalization at time *t*

$HCC(t)$: Number of critical care individuals under hospitalization at time *t*
$CC(t)$: Number of individuals under critical care at time *t*
$D(t)$: Number of dead individuals at time *t*
R_E : Effective reproduction number
p_I : Duration of the infectiousness period
p_E : Duration of the exposed (latent) period
p_{toH} : Duration before hospitalization
p_{HNCC} : Duration of non-critical care hospitalization period
p_{HCC} : Duration of critical care hospitalization period
p_{CC} : Duration of critical care period
r_H : Rate of hospitalization
r_{CC} : Rate of critical care patients among hospitalized individuals
$\mathrm{CFR}_{\mathrm{CC}}(t)$: Case fatality rate of critical care patients at time *t*

Time progresses continuously in the model. Model programming and analysis were performed in R (version 3.6.2). We used the deSolve package to solve the ordinary differential equation system.[3]

Though there is uncertainty about the possibility of re-infection with COVID-19, we assume that immediate re-infection of COVID-19 is not feasible within the time frame of this study based on the opinions of several experts.[4]

Case fatality rate

Since the true prevalence of COVID-19 is not known, mortality in the model was determined using the observed case fatality rate (CFR), i.e., the ratio of the confirmed deaths to the reported COVID-19 cases. To account for the duration between diagnosis and death, we assumed considered a lag time between the diagnosis of cases and the reported deaths.[5] We used the following to calculated CFR at time t:

$$CFR(t) = \frac{\text{Deaths at } t+16}{\text{Newly diagnosed COVID-19 cases at } t}$$

Furthermore, because COVID-19 testing rate has increased over time, which affects the diagnosis rate, we defined CFR as a function of time as below:

$$CFR(t) = CFR(0) \cdot \alpha^t$$

where $CFR(0)$ is the initial case fatality rate and α is the case fatality daily rate reduction factor. To calculate the reduction factor, we fitted an exponential regression model using time as the independent predictor of the case fatality rate. We estimated the national case fatality rate daily reduction factor (α) to be 95.5% (**Figure 2**). State-specific CFR projections based on most recent data are available in **Appendix Figure 1.**

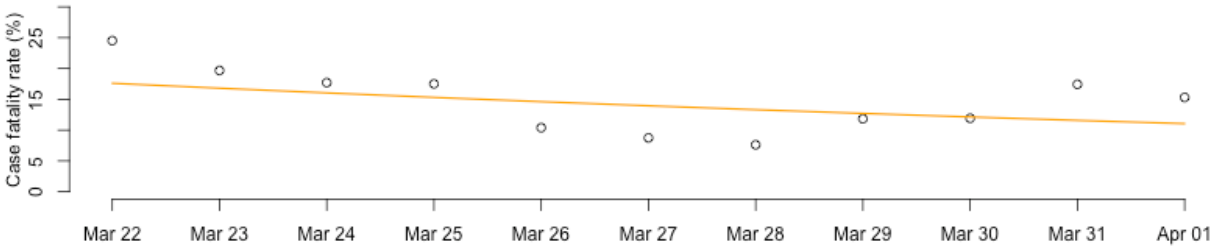

Figure *2. Estimation of case fatality rate as a function of time*

Calibration of unobserved parameters

Because several parameters in the model were not directly observable, we estimated the values of these parameters using a calibration approach. In particular, we calibrated the values of the following parameters by first defining clinically plausible ranges:

- Initial number of infections (range: 0–1,000 cases)
- Latent period duration (range: 4.5–5.8 days)
- Infectious period duration (range: 2.1–7 days)
- Basic reproductive number (R_0) (range: 2–2.4 cases)
- Other values used in the function of reproductive number (R_E^1, R_E^2, R_E^3) (range: 0.1–2.4 cases)
- Initial case fatality ratio (range: 0.01–0.4 per case)

We applied a directed search algorithm, using Generalized Simulated Annealing (package "GenSA"),[6] to identify the values of calibrated parameters such that model outcomes closely matched cumulative reported COVID-19 cases and deaths (i.e., calibration targets) defined by the weighted sum of absolute relative errors between model output and calibration target. To account for uncertainty in the calibrated parameter values, we repeated the calibration process 100 times, resulting in 100 unique sets of calibrated model parameters. We selected the top 100 sets of results for model projections and presented average outcome at each time point and 89% credible intervals.

Table 1. Key model parameters used in COVID-19 Simulator

Parameter	**Value or Range**	**Source**
Basic reproductive number (R_0)	2.4	[7]
Latent period duration	4.5–5.8	[8]
Infectiousness period duration	2.1–7	[9]
Effective reproduction number of *Current Intervention*	state specific	estimated
Effective reproduction number of *Public Awareness*	1.68	[10]
Effective reproduction number of *Lockdown*	0.3	[11]

Hospitalization rate upon diagnosis	state specific	The COVID Tracking Project[12]
Rate of critical care after hospitalization	30.6%	[13]
Mean hospitalization duration for non-critical care patients	8 days	[14]
Mean hospitalization duration for critical care patient	16 days	[14]
Time until hospitalization	5 days	[14]
Case fatality rate reduction factor	95.1%	estimated
Threshold to suppress the epidemic	10 active cases per 1 million	assumption

Intervention Strategies

We simulate model outcomes until August 31, 2020 under different levels and duration of social-distancing interventions, as defined below:

1. **Minimal restrictions**: This strategy assumes that there is minimal social distancing in place to reduce the spread of COVID-19, but there is an assumed level of learned social awareness (handwashing, avoiding close contact when sick, etc.). We assume the $\boldsymbol{R_E}$ of this intervention will be 1.68, which is 30% lower than the basic reproduction number, R_0, of COVID-19.[10]
2. **Current intervention in each state**: For most states this is a stay at home order, where people are advised to stay at home except for essential needs such as grocery shopping and picking up prescriptions. The New York Times provides an updated list of the current interventions in each state. Considering current level of interventions and rising public awareness, we assumed that the transmission rate (controlled by the $\boldsymbol{R_E}$ parameter) will decrease over time. To estimate the reduction in the transmission, we first estimated $\boldsymbol{R_E}$ values of at the last 10 days for each state based on the daily count of COVID-19 cases using the R package EpiEstim by Cori et al.[15] Then, using these estimated $\boldsymbol{R_E}$ values, we build an exponential regression model using time as the independent predictor of the $\boldsymbol{R_E}$ values.

$$log\big(R_E(t)\big) = b_0 + b_1 t$$

We then projected the future $\boldsymbol{R_E}$ values using the fitted exponential regression model. We assume that under current intervention, that the projected $\boldsymbol{R_E}$ values would not go below 0.3, th**e** $\boldsymbol{R_E}$ value of *Lockdown* strategy. For a state where we observed an upward trend in $\boldsymbol{R_E}$ values, we assumed that the latest value of the fitted $\boldsymbol{R_E}$ values will determine the transmission rate of the epidemic under the current intervention. The estimation and the projection of the $\boldsymbol{R_E}$ values are dynamically updated with new data to better reflect the intervention dynamics in each state. **Appendix Table 1** presents the current $\boldsymbol{R_E}$ values using data up to April 30, 2020. Future trends in $\boldsymbol{R_E}$ are available in **Appendix Figure 2**.

3. **Lockdown**: there is a complete ban on travel, including cancelling flights and closing inter-state travel and local travel (except for limited time for essential needs such as grocery shopping and picking up prescriptions needs), as has been done in countries such as Italy, China, and India. We used the $\boldsymbol{R_E}$ of 0.3, as estimated in Wuhan after the lockdown of the region.[11]

We simulate different combinations of two sequential interventions, each of which could last for 1–8 weeks. After the interventions, the effective reproductive number is changed to that of the public awareness only scenario ($\boldsymbol{R_E}$ of 1.68). After each intervention, we considered a reproduction number changing phase of one week. During this phase, the effective reproductive number linearly interpolated from its last value to the effective reproduction value of the next intervention to eliminate abrupt changes in the projections. The first intervention starts being implemented on the following Sunday after the date of the latest observation in the dataset. After each intervention, we check if the epidemic is suppressed or not based on the total number of active (infectious and not recovered) cases. Suppression is defined as there being that there are less than 100 active cases.

Suppression of transmission

We assumed that once the number of active COVID-19 cases (infectious and not recovered) in a given state reaches below a threshold of 10 active cases per 1,000,000 people, all cases can be isolated, which will stop the transmission of coronavirus in the community.

Model outcomes

For each state, the model generated the following outcomes from March 15, 2020 to August 31, 2020:

- Total number of deaths from COVID-19
- Daily counts of COVID-19 cases
- Cumulative number of COVID-19 cases
- Number of active COVID-19 cases
- Number of hospital beds needed for COVID-19 patients
- Number of ICU beds needed for COVID-19 patients

Hospital beds Capacity

Data on hospital beds and capacity were extracted from the annual cost reports (fiscal years 2016 through 2019) that hospitals file to the Centers for Medicare & Medicaid Services (CMS). The data from these reports is then made available through CMS's Healthcare Cost Report Information System (HCRIS).[16] Data were analyzed over a period of years to allow for corrections of both missing and inaccurate data. Hospitals that were deemed unlikely to be able to assist greatly in a pandemic were not counted in this analysis (alcohol and drug treatment hospitals, psychiatric hospitals, community mental health hospitals, hospice, religious non-medical hospitals, and skilled nursing facilities and homecare). For Intensive Care Unit (ICU) beds, we also included beds in similar units that could be repurposed as general intensive care in the event of a pandemic (cardiac critical care, burn ICU, and surgical ICU units).

To get the estimated number of beds available to COVID-19 patients, we calculated the average number of available beds (hospital beds or ICU beds) in each hospital on a single day. This was done using reported bed days and reported inpatient days for each type of bed. If the hospital reported bed numbers

but did not report bed utilization numbers, we used the state average occupancy rate (calculated from all states that provided this data) to calculate the estimated number of beds available to coronavirus patients.

References

1. Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities. *American journal of public health.* 2006;96(3):452-458.
2. Shao Y. JPSurv: Methods for population-based cancer survival analysis [R package]. 2015.
3. Soetaert KE, Petzoldt T, Setzer RW. Solving differential equations in R: package deSolve. *Journal of Statistical Software.* 2010;33.
4. Leung H. Can You Be Re-Infected After Recovering From Coronavirus? Here's What We Know About COVID-19 Immunity. Time. April 3, 2020. 2020.
5. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. *The Lancet Infectious Diseases.*
6. Xiang Y, Gubian S, Suomela B, Hoeng J. Generalized Simulated Annealing for Global Optimization: The GenSA Package. *R Journal.* 2013;5(1).
7. Guan W-j, Ni Z-y, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine.* 2020.
8. Lauer SA, Grantz KH, Bi Q, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine.* 2020;172(9):577-582.
9. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine.* 2020;382(13):1199-1207.
10. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet.* 2020;395(10228):931-934.
11. Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA.* 2020.
12. The COVID Tracking Project (https://covidtracking.com/)
13. CDC COVID-19 Response Team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(12):343-346.
14. Ferguson N, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team. Preprint at Spiral https://doi.org/10.25561/77482 (2020). In:2020.
15. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol.* 2013;178(9):1505-1512.
16. Centers for Medicaid and Medicare Services. Cost Reports. In:2019.

Appendix Table 1. Current effective reproduction number (R_E) of each state using data up to April 30, 2020.

State	R_E	State	R_E	State	R_E
Alabama	0.93	Maine	1.01	Oregon	0.97
Alaska	0.71	Maryland	1.09	Pennsylvania	1.01
Arizona	1.11	Massachusetts	1.03	Rhode Island	0.97
Arkansas	0.91	Michigan	0.98	South Carolina	1.06
California	0.99	Minnesota	1.45	South Dakota	0.88
Colorado	1.08	Mississippi	1.04	Tennessee	1.10
Connecticut	0.80	Missouri	0.97	Texas	1.05
Delaware	1.13	Montana	0.56	Utah	1.06
Florida	0.86	Nebraska	1.32	Vermont	0.77
Georgia	0.90	Nevada	0.98	Virginia	1.12
Hawaii	0.57	New Hampshire	1.02	Washington	0.95
Idaho	0.77	New Jersey	0.89	West Virginia	0.86
Illinois	1.17	New Mexico	1.17	Wisconsin	1.16
Indiana	1.16	New York	0.92	Wyoming	0.72
Iowa	1.26	North Carolina	1.14	District of Columbia	0.97
Kansas	1.23	North Dakota	1.17	Puerto Rico	0.88
Kentucky	1.06	Ohio	0.80	US	0.99
Louisiana	0.80	Oklahoma	0.94		