JAMIA Open, 4(3), 2021, 1–7
doi: 10.1093/jamiaopen/ooaa074
Application Notes

Application Notes

Population simulations of COVID-19 outbreaks provide tools for risk assessment and continuity planning

Bo Peng, Rowland W. Pettit, and Christopher I. Amos

Department of Medicine, Baylor College of Medicine, Houston, Texas, USA

Corresponding Author: Bo Peng, Department of Medicine, Baylor College of Medicine, Houston, TX, USA; Bo.Peng@bcm.edu.

Received 20 September 2020; Revised 17 December 2020; Editorial Decision 21 December 2020; Accepted 24 December 2020

ABSTRACT

Objectives: We developed COVID-19 Outbreak Simulator (https://ictr.github.io/covid19-outbreak-simulator/) to quantitatively estimate the effectiveness of preventative and interventive measures to prevent and battle COVID-19 outbreaks for specific populations.
Materials and methods: Our simulator simulates the entire course of infection and transmission of the virus among individuals in heterogeneous populations, subject to operations and influences, such as quarantine, testing, social distancing, and community infection. It provides command-line and Jupyter notebook interfaces and a plugin system for user-defined operations.
Results: The simulator provides quantitative estimates for COVID-19 outbreaks in a variety of scenarios and assists the development of public health policies, risk-reduction operations, and emergency response plans.
Discussion: Our simulator is powerful, flexible, and customizable, although successful applications require realistic estimation and robustness analysis of population-specific parameters.
Conclusion: Risk assessment and continuity planning for COVID-19 outbreaks are crucial for the continued operation of many organizations. Our simulator will be continuously expanded to meet this need.

Key words: COVID-19, Simulation, Risk assessment, Continuity planning

LAY SUMMARY

The effective deployment of mitigation strategies in response to the global COVID-19 pandemic relies on accurate and flexible modeling. A critical need exists for reliable models that can be meaningfully applied in diverse population and community scenarios: Underestimation of viral spread has clear detrimental consequences, but overestimation could lead to inappropriate allocation of resources and a scarcity of aid where it is most needed. To comprehensively address this need, we have developed a COVID-19 Outbreak Simulator (https://ictr.github.io/covid19-outbreak-simulator/) that allows flexible and accurate modeling of the spread of the SARS-CoV-2 virus based on intrinsic community realities and specific mitigation efforts. With over 15 command line parameters and 10 plugins available out of the box, this simulator is readily fit to population-specific requirements to provide tailored insight for the effectiveness of various quarantine, social distancing, and testing strategies subject to environmental factors, such as community infection rates. Our simulator will provide quantitative estimates of COVID-19 outbreaks that have assisted and will continue to assist in the development of public health policies, risk-reduction operations, and emergency response plans.

INTRODUCTION

Consequences of the current COVID-19 pandemic are far-reaching. The novel SARS-CoV-2 infection has produced life-threatening illness in diverse patient populations. As a direct result, previously accepted "normal" social and economic actions are now potentially hazardous. In light of generalized risk for viral transmission, dissimilar organizations are faced with difficult tasks of creating operation and management plans without adequate quantitative modeling assistance. Healthcare facilities, government agencies, factories, and businesses each immediately need robust tools to estimate the risks of various operational strategies. A clear need exists for tunable and deployable models of viral transmission that are publically available so that organizations may reliably compare options and minimize risks, and develop emergency plans to contain viral spread in the event of an outbreak.

COVID-19 epidemiological simulators have been developed. However, no model to date has been designed specifically for risk management and continuity planning, focusing on specific populations instead of management for generalized regions. In response to requests from various industries, including hospitals and government agencies, we developed a population-based viral outbreak simulator. Our COVID-19 Outbreak Simulator (https://ictr.github.io/covid19-outbreak-simulator/) models the outbreak of SARS-CoV-2 infection at an individual level and appropriately replicates the effect of preventative measures and postoutbreak containment strategies. Current implementations of our model are wide-ranging and have included shaping global industrial corporate policy, identifying safe practices in shipping commerce, as well as in aiding actors travel and scheduling in the film and other industries where close contact among employees is unavoidable. Further, our simulator has been deployed in schools for which a specified population with limited ingress is permitted, but for which controlling transmission is a critical requirement. Statistics summarized from simulations have been used to assist the development of public health policies, risk-reduction operations, and emergency response plans for diverse environments.

MATERIALS AND METHODS

We aimed to create a COVID-19 Outbreak Simulator to flexibly model viral transmission in site-specific settings. We estimate situation-specific viral transmission parameters and observe their effect on disease spread using forward-time simulations. Our model allows flexibility in establishing parameters that affect population-specific behaviors (i.e. interactions, transmission, interpopulation travel) so that the effect of intervention strategies may be observed.

Our simulator operates by modeling individuals in specific populations, including populations that may consist of multiple subpopulations with subpopulation-specific model parameters. After defining a virtual population space with site-specific probability of interaction parameters, the user can introduce infected individuals and observe the entire course of the ensuing infection and spread of the SARS-CoV-2 virus. Specifically, the simulator evolves the modeled heterogeneous population forward-in-time, based on the creation and handling of events of interest, such as quarantines.

For example, infection "events" can be introduced at the beginning of a simulation to specified carriers or en masse according to a specified incidence rate (Table 1), or during the simulation through community infection. The handling of infection events can trigger events, such as "shows symptoms," "quarantine," "recover," "removal" of the carrier, and multiple "infection" events of other individuals as the virus is transmitted to others. Operations, such as testing influence the simulation by, for example, quarantining carriers before they become infectious. A simulation stops after a prespecified time or after the exhaustion of events, which happens when no one or everyone becomes infected, and thus no new infections could happen. The simulator records details of all events and provides tools to extract various summary statistics from these records.

Table 1. System provided plugins and their features

Name	Features
init	Initialize population with a specified incidence rate and/or seroprevalence.
setparam	Change model parameters during the course of the simulation.
stat	Output population and subpopulation statistics, such as incidence rate.
sample	Calculate statistics from a sample drawn from the population.
insert	Addition of individuals to the population.
remove	Removal of individuals from the population.
move	Move individuals from one subpopulation to another.
quarantine	Quarantine all or infected individuals for a specified duration.
testing	Test all or selected individuals with a test with specified sensitivity, specificity, limit of detection, and turnaround time.
community_infection	Infect all or selected individuals with a specified probability.

The driving force of the simulation is the chain of reactions caused by infection events. The number and timing of secondary infections are simulated at the onset of infection. When the time for an infection event arrives, an "infectee" is selected randomly from the carrier's "social circle," which consists of randomly selected individuals from each subpopulation with whom the carrier can interact. An infection event can fail if the carrier or infectee is quarantined or if the infectee withstands an encounter due to previous infection or partial susceptibility that grants individuals certain levels of protection against viral infection.

The core of the simulator is a set of statistical models of transmissibility and viral load, which operate over the entire course of simulation for individuals (Supplementary Table S1). The precise nature of these submodels has and will continue to be updated as our collective understanding of the epidemiology of COVID-19 grows. Currently, we model viral transmissibility for an individual as a piecewise linear function; which has an initial period of noninfectivity, followed by a period of increasing infectivity, and ending with a period of decreasing infectivity (1). As a default model for a population without social distancing, symptomatic cases have an incubation period that follows a lognormal distribution with a mean of 5.5 days (2), a peak of transmissibility 1–3 days before symptom onset (3), and an infectivity period after the onset of symptoms that follows a shifted lognormal distribution. The transmissibility curves for asymptomatic carriers are similar, although the overall communicable periods are shorter than for symptomatic cases (Figure 1A) (4, 5). We draw a reproduction number (R_0) for each individual in the virtual population from a normal distribution with a 95% confi-

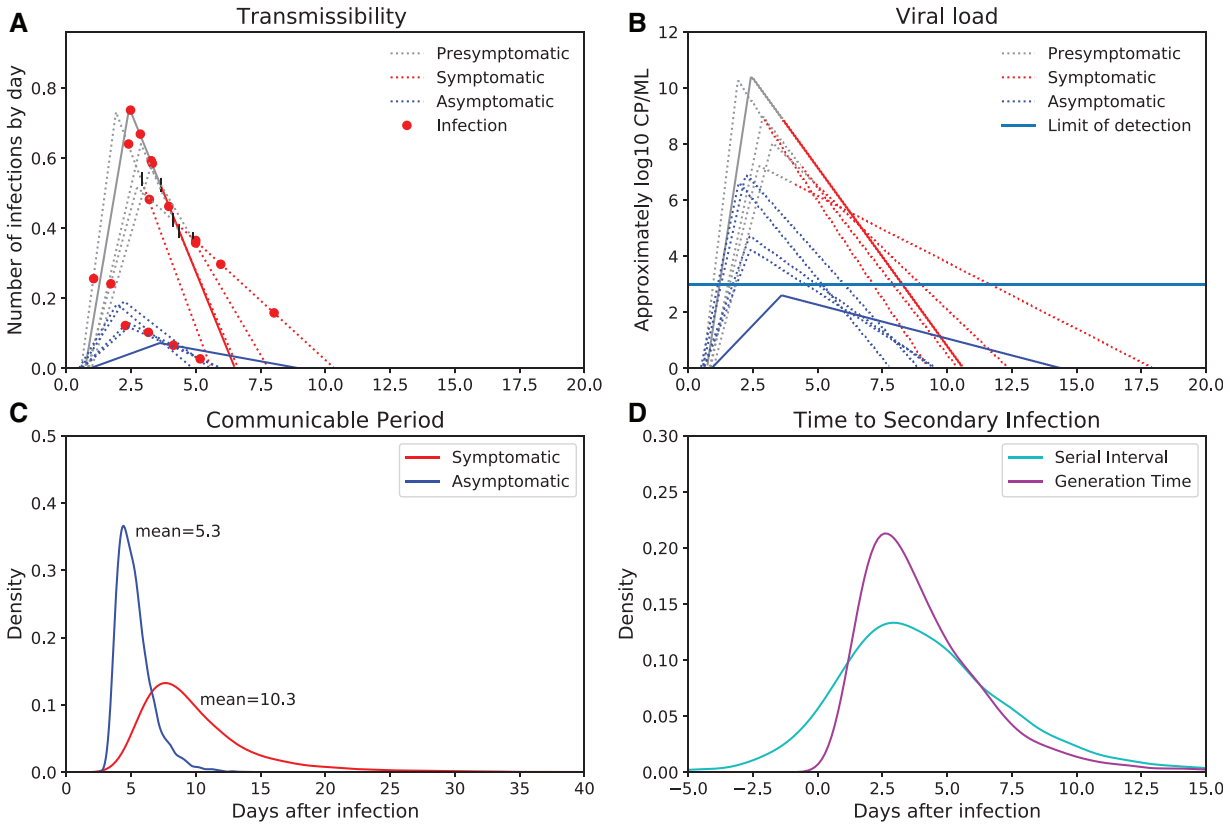


Figure 1. (A) Probability of transmission per day for a symptomatic case with an incubation period of 4 days (purple line) and for an asymptomatic case (green line). (B) Distribution of serial intervals of 10 000 simulated infector-infected individual pairs.

dence interval (CI) between 1.4 and 2.8 for symptomatic cases and 0.28–0.56 for asymptomatic carriers (6). We set the probabilities of transmission over the course of infection to sum up to the reproduction number (R_0) of the infected individual. We assume 25% of carriers (95% CI from 10% to 40%) are asymptomatic. These parameters were calibrated to match epidemiological observations of SARS-CoV-2 infections. For example, the average communicable period of simulated SARS-CoV-2 carriers was 11.64 days for symptomatic cases and 5.35 for asymptomatic carriers (Figure 1C) (4, 7). The distribution of the serial interval between infector and infectee symptoms follows a normal distribution with ~12% of pairs showing negative serial intervals (the infected individuals show symptoms before the infector), which is consistent with nonsynthetic publicly reported observations (Figure 1D) (8). The proportions of asymptomatic, presymptomatic, and symptomatic transmissions are 6%, 48%, and 44%, respectively (9).

The viral load patterns follow individual transmissibility curves. We extend the tails of the distributions to reflect a longer viral shedding period than communicable period and adjust the distribution intensities so that symptomatic carriers and asymptomatic carriers have similar initial viral loads (Figure 1B) ([10–12]). We use a distancing factor to model social distancing and mask-wearing, which leads to different transmissibility probabilities for individuals with the same level of viral loads. Our model captures that SARS-CoV-2 tests may have fixed sensitivities or sensitivities that are determined by the limit of detection and therefore depend on an individual's viral load at the time of testing. Parameters may be modified by the user allowing a highly flexible approach for simulating populations and interventions.

This simulator is provided as a Python package with a command-line interface. The out of the box model provides 15 parameters, which can be tuned to appropriately fit site-specific constraints, including infectivity, susceptibility, social distancing and contact patterns, and actions taken upon people who show symptoms or test positive. Additionally, at the time of this publication, we offer 10 plugins to manipulate simulations as they progress, such as by removing or adding additional individuals into a population, moving between subpopulations, changing model parameters, and testing part or all individuals with SARS-CoV-2 tests with different sensitivity, specificity, the limit of detection, and turnaround time (Table 1). The simulator can run tens of thousands of simulations on multiple processors in a few minutes. The final performance of the simulator depends on the choice of parameters, especially the population size.

The main output of the simulator is a log file that records all events that happened during the simulation. The log files can be analyzed with any scripting language to extract statistics of interest. We provide an option in the simulator to generate summary statistics for models with particular parameters (–summarize-model), an option to generate a summary report with key statistics (–summary-report), and a number of scripts that extract and plot particular statistics. To facilitate usability, we provide Jupyter notebooks of real-world scenarios where our COVID-19 simulator has previously been deployed. These example notebooks include narratives, simulation commands, and analysis of results in the applications section of the homepage. We will delineate the results of 3 specific examples of simulator deployment, however, each analysis (and more) can be reproduced by executing these notebooks available online. Further,

we provide Docker containers to allow the execution of our simulator and notebooks without the need to install all relevant tools. Detailed documentation and instructions on the installation and execution of the simulator are provided on the project homepage.

RESULTS

Our COVID-19 Outbreak Simulator is capable of simulating heterogeneous populations with complex interactions with a relatively simple command-line interface. To illustrate with an example, the option "–popsize nurse = 10 room1 = 4 room2 = 4" models a nursing home with 10 nurses and 2 rooms each with 4 patients. Assuming that nurses are more "mobile" than patients and therefore an infected nurse can infect more people than patients, we can set the reproduction number for nurses to be higher than the default values, and patients to be less, with the option "–symptomatic-r0 nurse = 2 room1 = 0.8 room2 = 0.8." Further, our model can capture individual-specific "social circles" that may exist in a population. One could allow nurses to access all rooms, but patients can only interact with other patients in the same room with options, such as "–vicinity room1-! room1 = 0 room2-! room2 = 0." Negations (!) and wildcards (*) are allowed in parameter settings to facilitate the specification of complex interaction patterns across subpopulations. Figure 3 contains another example in which multiple plugins are used to construct a complex interventive strategy.

In response to a request for assistance related to shipping, and as a real-world implementation of the model, we simulated a scenario where a single infected individual enters an enclosed environment without quarantine. In our model, we assumed that roughly 10%–40% of infected individuals never show COVID-19 symptoms, and those who show symptoms will be removed via quarantine from the enclosed population promptly. Under these parameters, assuming one infected individual enters the environment at the start of the simulation, we found that in 18.2% of simulations, the carrier will not infect anyone or show symptoms. In the remainder of cases, an outbreak will happen, and unless the carrier shows symptoms on the first day, there is a 47.3% probability the virus has already been transmitted to at least one other person by the time of detection. We concluded, therefore, that removal of symptomatic cases is insufficient to stop the outbreak. The probability that a second symptomatic case happens remains high until one week after the removal of the first symptomatic case occurs. Figure 2A illustrates the duration of the outbreak versus the remaining population size in this scenario, showing that a large percentage of outbreaks stop on the first day when the carrier, most likely an asymptomatic carrier, is not expected to infect anyone else. Based on the high probabilities of an outbreak under this scenario, better approaches to risk management were determined to require quarantine prior to the introduction of an infected individual to the population. A 7-day quarantine would avoid 82.1%, and a 14-day quarantine would avoid 99.3% of potential outbreaks (Figure 2A). The company mitigated risk by quarantining employees prior to travel.

A second real-world example resulted from our support of an academic institution that needed to estimate the effectiveness of periodic PCR-based tests in reducing within-lab infections. To ensure the safety of the lab environment, periodic testing of lab members was to be conducted to identify asymptomatic or presymptomatic individuals and quarantine them before infecting others. Using the simulator, we modeled lab environments of various sizes in which lab members are subject to infection from the public. We simulated an equal probability of infection for members ($p = 0.0022$) and tested the efficacy of PCR testing of all lab members every 3, 7, or 14 days. For each scenario, we performed 10 000 replicate simulations and compared outcomes, including the average lab size remaining after 90 days with an intervention. As can be appreciated in Table 2, implementing a routine test policy ensures fewer within lab-transmissions and greater retention of the lab worker pool.

Third, we present the result from our simulator being engaged to test the efficacy of test-assisted quarantine strategies. Several public recommendations of efficacious quarantine lengths have publically been reported, including a 14 day postexposure isolation period (13). Our model was engaged to see if this period could safely be shortened if viral testing was conducted during the quarantine win-

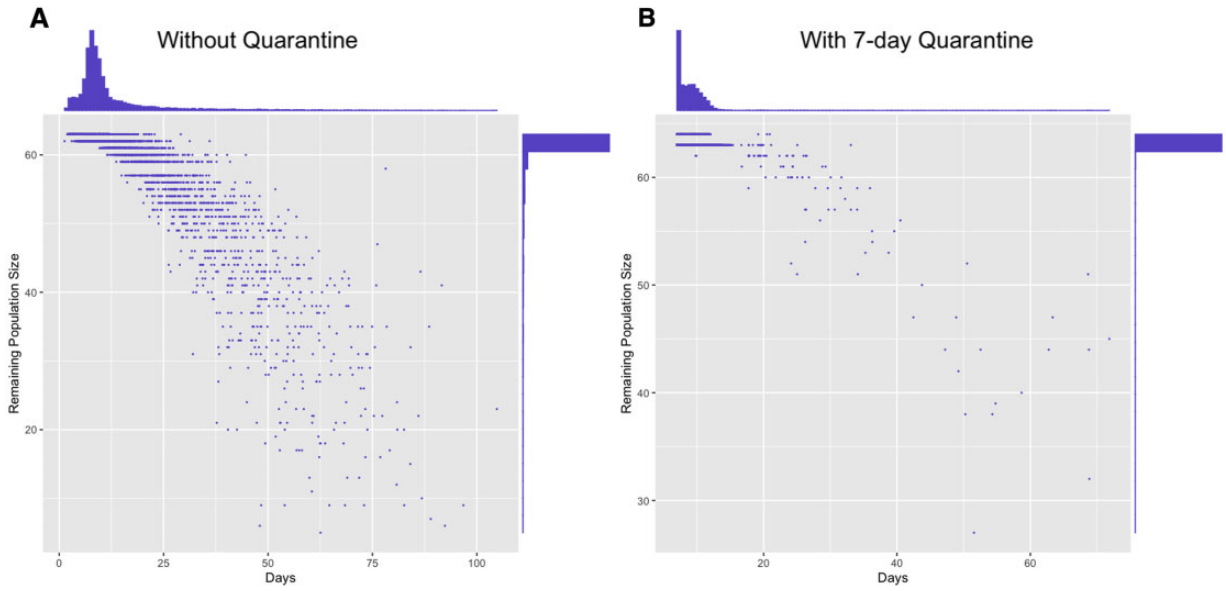
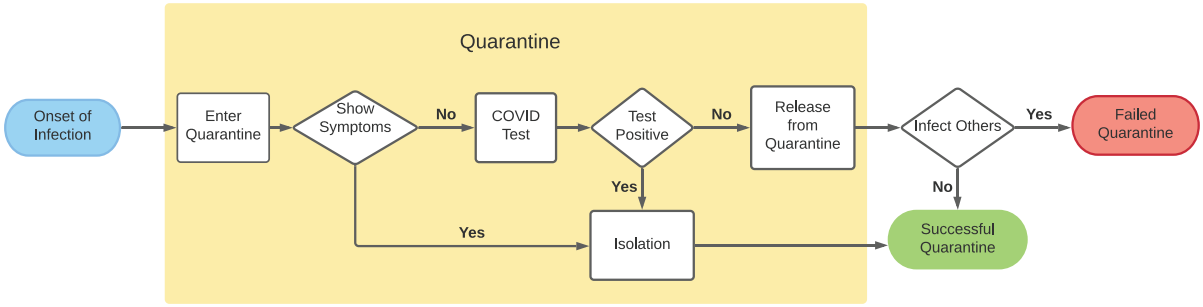


Figure 2. Duration versus remaining population size for 10 000 simulated outbreaks. (A) The virus carrier was introduced to the population as long as he or she did not show any symptoms. (B) The virus carrier was introduced to the population after 7-day quarantine.




```
outbreak_simulator --popsize 10000 --replicate 10000 --handle-symptomatic remove --stop-if 't>10' \
  --plugin init --incidence-rate 1 --leadtime asymptimatic \
  --plugin quarantine --at 0 --duration 10 --target all --proportion 1 \
  --plugin testing --proportion 1 --at 9  --turnaround-time 1 \
      --handle-positive remove --sensitivity 0.95 --specificity 0.99
```

Figure 3. Diagram of a quarantine process with one test performed before the end of quarantine, with the command to implement the simulation listed below. Briefly, (1) a population is initialized with an incidence rate of 1 so everyone is infected. A random lead time is used so that symptomatic cases are anywhere in their incubation period, and asymptomatic carriers are anywhere in their course of infections. (2) Everyone is quarantined for 10 days regardless of they have shown symptoms. (3) A SARS-CoV-2 test with a sensitivity of 95% and specificity of 99% is applied to everyone on day 9. The test results will be available at the end of quarantine due and people who test negative will be isolated. (4) People who show symptoms during quarantine will be isolated.

Table 2. The effectiveness of periodic PCR-based tests in reducing within-lab infections

Lab size	Test frequency	Average lab size after 90 days	Proportion of uninfected labs	Average number of community infections	Average number of within-lab transmissions
10	3	5.9	1.14%	3.6	0.5
10	7	5.2	1.14%	3.4	1.5
10	14	4.5	1.15%	3.1	2.4
10	No test	3.7	1.01%	2.9	3.5
20	3	11.8	0.02%	7.1	1.2
20	7	10.1	0.03%	6.6	3.3
20	14	8.3	0.00%	6.1	5.6
20	No test	6.2	0.00%	5.3	8.5
30	3	17.6	0.00%	10.6	1.9
30	7	15	0.00%	9.9	5.2
30	14	12	0.00%	9	9
30	No test	8.5	0.00%	7.7	13.8

All simulations assumed a daily probability of community infection of 0.0022. Results for each scenario are based on 10 000 replicate simulations.

dow. We tested the efficacy of several test-assisted quarantine strategies on their efficacy in preventing population viral spread. Methods tested included testing at various points during quarantine and at multiple times during quarantine. The simulations were precise, included modeling measured test parameters, such as the sensitivities and specificities of various COVID-19 testing strategies (Figure 3). Simulation revealed the efficacy of a 10-day, and other foreshortened, quarantine strategies. The command used for the simulation is listed under the flowchart (Figure 3), and for full details of this implementation, we refer the reader to Peng et al (14).

DISCUSSION

In response to the clear need for flexible, robust, and population-specific models of SARS-CoV-2 viral spread, we developed a COVID-19 outbreak simulator. This population and subpopulation specific Python-coded application simulates the community progression of viral infection via an event-based forward-time simulation. This approach affords the user the opportunity to directly model consequences for planned interventions. The insight gained from feed-forward modeling, in particular, allows for meaningful scenarios to be represented and subsequent informed decision-making to occur. Ultimately, our simulator provides quantitative estimates of COVID-19 outbreaks that have, and will continue to, assist in the development of public health policies, risk-reduction operations, and emergency response plans.

Our COVID-19 Outbreak Simulator has met several user demands, demonstrating a robust capacity to model diverse scenarios. In our second simulation, we detail a real example of lab member testing, extensive simulation allowed for the benefit of routine viral testing to be quantified. Exact numbers of tests needed to be administered, and their average effects on transmission prevention in the form of lab member retention were estimated. In the immediate context, this information allowed for testing strategy selection decisions to be appropriately informed, adding justification (or not) to the expenses between testing frequency strategies. In our third simulation example of test-assisted quarantine strategies, our model was the first to comprehensively demonstrate the efficacy of a quar-

antine scenario where a SARS-CoV-2 test is administrated at the end of a 10-day quarantine. This scenario demonstrated novel insight into the contingent efficacy of 10-day, or other foreshortened, quarantine strategies. In the film industry, our simulator has been implemented to test the efficacy of pre-film quarantine measures. By comparing outcomes, best practices were determined and acted upon to safely reduce the cost of film production while ensuring safety for all set members. We present these specific results and the corresponding jupyter notebooks which generated them online for free, and aim for similar applications to be generalized from them by the public.

Strengths of this simulator include that we constantly update the baseline model according to the most recent discoveries about the epidemiology of COVID-19 outbreaks, and we allow for the specification of distributions instead of fixed values for many parameters to account for variations of models. We perform a large number of simulations with different parameters to obtain predictions associated with a parameter space and provide results under the best and worst scenarios. During this process, we are able to not only assess the robustness of our predictions but also identify parameters that have the most impact on outcomes (e.g. the severity of the outbreak), which by itself were valuable for policy-making purposes.

A limitation of the simulator is that it does not simulate physical locations, so the transmission of the virus is performed by probability, not by actual physical vicinity or contact. Although "social circles" can be used to model the number of individuals from each subpopulation a carrier can interact with, the target of infections will be chosen randomly. This design avoids a large number of model assumptions, such as population boundaries and individual movement patterns and allows much more efficient simulations of large populations. It, however, makes it currently infeasible to simulate the geographic-based transmission of the virus in a region. Despite this limitation, the current version of the software is useful for studying the behavior of outbreaks affecting many environments in which individuals are well mixed. Successful applications have occurred in a wide variety of industries, including shipping, mining, actors, and laboratory environments, for which a limited population may interact regularly. Our simulator is effective for identifying how risk can be managed and what kinds of mitigation provide acceptable control for these situations.

CONCLUSION

The COVID19 Outbreak Simulator simulates the spread of the SARS-CoV-2 virus in environments with heterogeneous but well-mixed populations that are subject to operations, such as quarantine, testing, social distancing, and influences, such as community infection and incomplete detection. The simulator provides quantitative measures to assess the risks of an outbreak associated with several operations and assist the development of emergency plans in response to an outbreak. The simulator has been successfully used for applications in industries, schools, and government agencies. As the number of notebooks with simulated scenarios for particular environments grows, analyzing particular scenarios with realistic assumptions will become increasingly easier. Comparisons between simulation contexts will allow for shared assumptions and quick applications of our simulator to new contexts. This simulator is suited to provide actionable, forward mapping guidance to epidemiologists and policymakers in diverse fields.

AUTHOR CONTRIBUTIONS

CIA conceptualized the idea of the simulator and applied it to real-world applications. BP implemented the simulator, performed all the simulations and data analyses, and drafted the manuscript. RWP contributed to the applications of the simulator and revision of the manuscript. All authors reviewed and approved the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

The authors appreciated feedbacks from users of this simulator, and Benjamin Wu, a summer student, for his contribution to the data analysis part of the project.

CONFLICT OF INTEREST STATEMENT

None declared.

AVAILABILITY OF DATA AND MATERIALS

The COVID-19 Outbreak Simulator is released through the Python Package Index under the name covid19-outbreak-simulator. All data generated and/or analyzed during the current study, in particular applications in the format of Jupyter notebooks, are publicly available through https://github.com/ictr/covid19-outbreak-simulator/.

REFERENCES

1. Larremore DB, Wilder B, Lester E, *et al.* Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci Adv* 2020; 7 (1): eabd5393.
2. 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. *Ann Intern Med* 2020; 172 (9): 577–82.
3. Wolfel R, Corman VM, Guggemos W, *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* 2020; 581 (7809): 465–9.
4. Zhang Z, Xiao T, Wang Y, *et al.* Early viral clearance and antibody kinetics of COVID-19 among asymptomatic carriers. medRxiv. 2020; 2020.04.28.20083139.
5. Yang R, Gui X, Xiong Y. Comparison of clinical characteristics of patients with asymptomatic vs symptomatic coronavirus disease 2019 in Wuhan, China. *JAMA Netw Open* 2020; 3 (5): e2010182.
6. Ferretti L, Wymant C, Kendall M, *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 2020; 368 (6491): eabb6936.
7. Hu Z, Song C, Xu C, *et al.* Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci* 2020; 63 (5): 706–11.
8. Du Z, Xu X, Wu Y, *et al.* Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg Infect Dis* 2020; 26 (6): 1341–3.
9. Ferretti L, Wymant C, Kendall M, *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 2020; 368 (6491): eabb6936.
10. Le TQM, Takemura T, Moi ML, *et al.* Severe acute respiratory syndrome coronavirus 2 shedding by travelers, Vietnam, 2020. *Emerg Infect Dis* 2020; 26 (7): 1624–6.

11. Lavezzo E, Franchin E, Ciavarella C, *et al.* Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo′. *Nature* 2020; 584 (7821): 425–9.
12. Corcorran MA, Olin S, Rani G, *et al.* Prolonged persistence of PCR-detectable virus during an outbreak of SARS-CoV-2 in an inpatient geriatric psychiatry unit in King County, Washington. *Am J Infect Control* 2020. Epublished on August 2020. doi: 10.1016/j.ajic.2020.08.025.
13. 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. *Ann Intern Med* 2020; 172 (9): 577–82.
14. Peng B, Zhou W, Pettit RW, *et al.* Optimal test-assisted quarantine strategies for COVID-19. *medRxiv* 2020; 2020.11.06.20222398.