November 2021

Infectious Disease Modeling and Intervention Planning

Shalome Hanisha Anand Tatapudi
University of South Florida

Infectious Disease Modeling and Intervention Planning

by

Shalome Hanisha Anand Tatapudi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Co-Major Professor: Tapas K. Das, Ph.D.
Co-Major Professor: Chaitra Gopalappa, Ph.D.
Jose Zayas Castro, Ph.D.
Ricardo Izurieta, M.D.
Ankit Shah, Ph.D.

Date of Approval:
November 10, 2021

Keywords: Impact assessment, pandemic preparedness, intervention analysis, epidemic mitigation, public health policy

Dedication

This dissertation is dedicated to my family, friends, advisors, and colleagues. I am very grateful to my parents Vijayalakshmi and Devanand, grandfather Ramabrahmam, and my sisters Aishwarya, Priyanka, and Devika whose love, motivation, and support inspired me to reach new heights.

I also dedicate this dissertation to many friends who have been my pillars of strength throughout my life. I would like to thank Arpita and Tanaya for constantly supporting me, listening to me, and advising me, and Vignesh who has been a dedicated companion and mentor. For Sandeep and Sydney for teaching me work life balance and for being my personal cheerleaders and advisors.

Thank you to my academic advisors Dr. Tapas Das and Dr. Chaitra Gopalappa for their patience, kindness, trust, and generosity, and to my graduate advisor Dr. Alex Savachkin who introduced me to the academic world. I also want to thank my PhD peers, both at USF and at UMass-Amherst, who taught me many invaluable lessons.

Finally, I would like to dedicate this work to God, for giving me opportunities I could never imagine and introducing me to life changing people.

Acknowledgments

I would like to thank my advisors Dr. Tapas Das and Dr. Chaitra Gopalappa. Dr. Das, thank you for teaching me almost everything valuable I have ever learnt, from professional to personal development and thank you also for believing in me when I doubted myself and for providing me opportunities to learn from and excel at. Dr. Gopalappa, thank you for being such a strong and powerful female role model, someone I aspire to be. Thank you also for further opening my eyes to the world of disease modeling and for teaching me how to create value and contribution with my research.

Thank you to my committee members, Dr. Jose Zayas Castro, Dr. Ricardo Izurieta, and Dr. Ankit Shah, for their invaluable insights and for going above and beyond to help me with my dissertation and career. I would like to acknowledge the staff and faculty of the IMSE department and college of engineering for helping me.

Thank you to Dr. Paul Farnham from the Centers for Disease Control and Prevention (CDC) for his insights on my research that significantly helped improve my dissertation.

I would like to acknowledge my source of funding, grant from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01AI127236.

Table of Contents

List of Tables

List of Figures

Abstract

Disease outbreaks caused by existing and emerging pathogens pose a serious threat to local and global communities in the form of epidemics and pandemics, respectively. The 2009 H1N1 pandemic and the 2019 COVID-19 pandemic are exemplars of how underprepared both the developed and developing nations are at mitigating pandemics.

Mathematical modeling and simulation of disease outbreaks has served as a powerful tool to understand disease transmission dynamics. They also aid in developing effective intervention strategies. However, existing models are usually particularized to a region, or a specific disease pathogen and interventions used in these models may not translate well to novel outbreaks or existing epidemics in another region. This could be due to the dynamic nature of evolving population demographics and disease parameters. Fully understanding the time-varying transmission dynamics of diseases, the effect of disease outbreaks on varying population demographics, the impact of effective interventions (both pharmaceutical and non-pharmaceutical), is imperative for epidemic and pandemic preparedness.

The goal of this dissertation is to propose frameworks and methodologies that model epidemics or pandemics and implement them to derive useful insights for intervention strategies that effectively mitigate disease burden in a population. Simulation models such as agent-based (AB) and compartmental approaches are implemented to closely follow the epidemic outcomes in regional outbreaks. The modeling frameworks presented in this dissertation have been used to simulate both novel disease outbreaks (COVID-19) and ongoing regional epidemics (HIV/AIDS). Combinations of varying levels of intervention measures are applied within the models to assess

the impact of the interventions, and the changes in trends of epidemic outcomes. Information and insights derived from these models aim to help policy makers with informed decision-making and improve public health metrics during an epidemic.

Chapter 1: Introduction

In the past, mathematical models have served as powerful tools for prediction of disease outbreak outcomes and public health decision making, especially in the wake of pandemics and epidemics [1]. In recent times, the COVID-19 pandemic has underscored the need for such models and their contribution in public health decision making, thus, there has been an explosive growth in literature for mathematical models that mimic epidemics and pandemics and develop intervention measures. Despite such plethora of studies, gaps in implementation of interventions derived from them to real-time policy has become increasingly evident. This could be caused by the manner in which interventions are designed in studies, type of interventions implemented, risk groups modeled, outcomes analyzed, socio-economic factors considered, among others [2]. Therefore, although existing models have filled a lot of existing research gaps, there appears to be a continuous need in developing innovative and insightful models that can adapt to changing dynamics of pathogens, populations, and interventions.

The objective of this dissertation is first, to develop mathematical simulation models that can mimic the transmission of infectious diseases in a population, and second, to develop effective intervention measures, and third, assess the impact of such interventions on the disease outcomes in a population. We developed AB models, and compartmental simulation models to evaluate the COVID-19 and HIV/AIDS epidemics in regions in the United States.

A granular AB model was developed to closely follow the COVID-19 disease outbreak outcomes (number of age-specific reported infections, hospitalizations, and deaths) in an urban region (considering Miami-Dade County, FL, U.S.A. as a case study). We implemented social

interventions such as full and partial stay at home orders, face mask usage, contact tracing and performed a sensitivity analysis on some interventions (face mask usage and contact tracing) and identified levels and combinations of interventions that would closely follow the epidemic outcomes and reduce the disease outbreak outcomes to statistically significant values. The methodology and results are explained in detail in Chapter 2.

The AB modeling framework was extended to implement school reopening (following intervention timelines implemented in the case-study region of Miami-Dade County, FL, U.S.A) at various levels of percentage return to school. A sensitivity analysis was performed on the transmission coefficient among students in a school (due to lack of data to calibrate transmission coefficient to age-specific reported infections in a school setting). The results of this study are presented in Chapter 3.

The AB model was recalibrated to follow outcomes (continuing to follow intervention timelines implemented in the case-study region of Miami-Dade County, FL, U.S.A) till early December 2020, and a vaccination framework is embedded. The model evaluated three different vaccine prioritization strategies (random strategy, age-specific strategy, and minor variant of CDC strategy). The epidemic outcomes (reported infections, hospitalizations, and deaths) from each of the three strategies are compared with no vaccination scenario to identify the most effective prioritization strategy. The results of this study are presented in Chapter 4.

A compartmental model was developed to simulate the COVID-19 epidemic in a residential university setting (University of Massachusetts-Amherst). The study analyzed combinations and levels of testing (symptom-based testing, mass testing, and tracing and testing) on thresholds of contact rate for non-essential workers, and population size on campus that would keep reported infections and deaths below three measures of tolerance (relaxed, medium, and

tight). The study also assessed combinations and levels of testing and relative levels of vaccine coverage and population size. The results of this study are presented in Chapter 5.

A compartmental model was developed to simulate the HIV/AIDS epidemic in the U.S. The study develops a jurisdictional model initialized to HIV epidemic data for 2018 and projects epidemic outcomes (incidence, prevalence, diagnoses, and deaths) till 2030. The model simulated geographical heterogeneity as outlined in the national HIV prevention plan: Ending the HIV Epidemic (EHE) plan, and applied interventions to reach the care continuum targets as per the EHE plan. The objective of the model was to evaluate if mixing between jurisdictions and jurisdictional heterogeneity is important in time varying outcomes and decisions. The results of this study are presented in Chapter 6.

1.1 Research Contribution

Individual research contribution for each study is highlighted in the respective chapters. In what follows, an overview of the research contribution and broader impacts are presented.

1.1.1 Developing a Granular Simulation Modeling Framework for Region-Specific Epidemics

Literature for COVID-19 modeling largely consisted of compartmental models that simulated SEIR (Susceptible-Exposed-Infected-Recovered/Removed) framework or relative extensions. Compartmental models are effective in modeling movement of aggregate populations across stages of disease but assume uniform mixing. Thus, they fail to capture special transmission dynamics (which could relate to occupation-level, age-specific, and gender-specific mixing) that may play a significant role for certain infectious diseases such as COVID-19. In the study presented in Chapter 2, a highly granular AB model was developed that considered comprehensive population demographics of a region and embedded it with the disease transmission model of COVID-19. The combination of a granular population structure combined with a disease

transmission model enabled the simulation to closely follow the epidemic outbreak in the case study region. The simulation model reports actual infected and reported infections, thus, allowing us to assess the true and relative impact of interventions in a population. The AB simulation model is also stochastic in nature, thus naturally modeling uncertainties of various input parameters, which is contrasting to the deterministic nature of the compartmental simulation model. Interventions implemented in the case study region can be applied to similar urban region that have similar population demographics. Further, the AB model can be particularized to other regions or other disease outbreaks that share disease dynamics of COVID-19. The model is built to be dynamic and thus can evaluate changes to disease transmission dynamics (changing transmissibility of virus, changing rate of hospitalization and deaths for specific age groups), interventions, and population demographics, and can apply real-time interventions applied to the case study region. The model was also used to identify effective COVID-19 vaccine prioritization strategies (presented in Chapter 4) which can be helpful for public health decision makers in regions which are still actively disseminating the vaccine among their population.

1.1.2 Developing Policies for Safe School/University Reopening During an Epidemic

When the COVID-19 pandemic hit the world in early 2020, school closure along with lockdown was implemented immediately across the U.S., (and the world). However, during Fall 2020, academic institutions debated on 1) how to safely reopen campuses, and 2) what would the impact of the reopening be on the epidemic. The compartmental model developed for a university setting (presented in Chapter 5) allows us to address the first question, i.e., how to safely reopen campuses and what levels of testing and vaccination resources (two key interventions) would be required. The AB model developed to evaluate impact of school reopening (presented in Chapter

3) would address the second question, i.e., what would be the impact of region wide school reopening on the epidemic outcomes.

1.1.3 Identifying Impact of Jurisdictional Mixing and Heterogeneity on National Epidemic

HIV/AIDS is a disease that is transmitted primarily by sexual interactions and injecting drug use. In the U.S. HIV is national epidemic and several models have evaluated either nationally aggregate model or individual level jurisdictional models. National level models are useful in projecting epidemic outcomes over time and developing nation-wide policies, but they do not consider the jurisdictional heterogeneity in population demographics and care, or the interactions between jurisdictions. Similarly, individual models specific to each jurisdiction are evaluated independently, which also ignore the interactions between jurisdictions. To address these gaps, a national HIV model was developed that split the model into individual jurisdictions and simulated mixing between these jurisdictions. Mixing of sexual partners across jurisdictions can have a significant impact on new infections and decisions taken to mitigate the disease burden. However, there are no models in literature that can evaluate mixing on a nation-wide level (especially for the U.S.). The compartmental model for HIV (presented in Chapter 6) can assess both, the impact of mixing, and changes to population demographics on an individual level for each jurisdiction on the epidemic. The model can also estimate levels of interventions to meet targets for each individual jurisdiction. This model can be replicated for other nations, or even evaluated to represent global mixing for HIV or other disease where mixing plays a role in transmission.

Chapter 2: Impact Assessment of Full and Partial Stay-at-Home Orders, Face Mask Usage, and Contact Tracing: An Agent-Based Simulation Study of COVID-19 for an Urban Region

The complete article titled "Impact assessment of full and partial stay-at-home orders, face mask usage, and contact tracing: An agent-based simulation study of COVID-19 for an urban region" [3] (published in Global Epidemiology) can be found in Appendix C. This article presents the framework for an agent-based model that simulates COVID-19 epidemic in an urban region in the U.S. (Miami-Dade County). The simulation model is particularized for the region of study with demographics obtained from U.S. census. The model closely follows the epidemic in the region of study and implements several non-pharmaceutical interventions implemented in the U.S., such as stay-at-home orders (or lockdowns), face mask usage, contact tracing, and phased reopening of businesses and workplaces. The study assesses the impact of these interventions and derives useful insights on how to significantly reduce the disease burden of the epidemic by implementing them.

2.1 Abstract

Social intervention strategies to mitigate COVID-19 are examined using an agent-based simulation model. Outbreak in a large urban region, Miami-Dade County, Florida, USA is used as a case study. Results are intended to serve as a planning guide for decision makers. The simulation model mimics daily social mixing behavior of the susceptible and infected generating the spread. Data representing demographics of the region, virus epidemiology, and

social interventions shapes model behavior. Results include daily values of infected, reported, hospitalized, and dead.

Results show that early implementation of complete stay-at-home order is effective in flattening and reversing the infection growth curve in a short period of time. Whereas, using Florida's Phase II plan alone could result in 75% infected and end of pandemic via herd immunity. Universal use of face masks reduced infected by 20%. A further reduction of 66% was achieved by adding contact tracing with a target of identifying 50% of the asymptomatic and pre-symptomatic.

In the absence of a vaccine, the strict stay-at-home order, though effective in curbing a pandemic outbreak, leaves a large proportion of the population susceptible. Hence, there should be a strong follow up plan of social distancing, use of face mask, contact tracing, testing, and isolation of infected to minimize the chances of large-scale resurgence of the disease. However, as the economic cost of the complete stay-at-home- order is very high, it can perhaps be used only as an emergency first response, and the authorities should be prepared to activate a strong follow up plan as soon as possible. The target level for contact tracing was shown to have a nonlinear impact on the reduction of the percentage of population infected. Increase in contact tracing target from 20% to 30% appeared to provide the largest incremental benefit.

Chapter 3: Impact of School Reopening on Pandemic Spread: A Case Study Using an Agent-Based Model for COVID-19

The complete article titled "Impact of school reopening on pandemic spread: A case study using an agent-based model for COVID-19" [4] (published in Infectious Disease Modelling) can be found in Appendix D. This article utilizes an agent-based model particularized to an urban region in the U.S. (Miami-Dade County, FL), to assess the impact of school reopening on the COVID-19 pandemic spread. The study performs a sensitivity analysis on two important factors that impact pandemic spread within schools: levels of student return to campuses, and levels of transmission probability. The study aims to assist policy makers in reopening schools in urban regions and assess the impact of varying levels of return and transmission probability on the pandemic outcomes of the region.

3.1 Abstract

This article examines the impact of partial/full reopening of school/college campuses on the spread of a pandemic using COVID-19 as a case study. The study uses an agent-based simulation model that replicates community spread in an urban region of U.S.A. via daily social mixing of susceptible and infected individuals. Data representing population demographics, SARS-CoV-2 epidemiology, and social interventions guides the model's behavior, which is calibrated and validated using data reported by the government. The model indicates a modest but significant increase (8.15 %) in the total number of reported cases in the region for a complete (100%) reopening compared to keeping schools and colleges fully virtual. For partial returns of 75% and 50%, the percent increases in the number of reported cases are shown to be small (2.87%

and 1.26%, respectively) and statistically insignificant. The AB model also predicts that relaxing the stringency of the school safety protocol for sanitizing, use of mask, social distancing, testing, and quarantining and thus allowing the school transmission coefficient to double may result in a small increase in the number of reported infected cases (2.14%). Hence for pandemic outbreaks from viruses with similar characteristics as for SARS-CoV-2, keeping the schools and colleges open with a modest campus safety protocol and in-person attendance below a certain threshold may be advisable.

Chapter 4: Impact of Vaccine Prioritization Strategies on Mitigating COVID-19: An Agent-Based Simulation Study Using an Urban Region in the United States

The complete article titled "Impact of Vaccine Prioritization Strategies on Mitigating COVID-19: An Agent-Based Simulation Study using an Urban Region in the United States" [5] (submitted to BMC Medical Research Methodology) can be found in Appendix E. This article utilizes an agent-based model particularized to an urban region in the U.S. (Miami-Dade County, FL) and compares three vaccination prioritization strategies to evaluate which yields the most beneficial outcomes during a pandemic. The study aims to aid policy makers around the world (especially those belonging to urban regions) in identifying risk groups that need prioritizing during limited vaccine supply.

4.1 Abstract

Approval of novel vaccines for COVID-19 had brought hope and expectations, but not without additional challenges. One central challenge was understanding how to appropriately prioritize the use of limited supply of vaccines. This study examined the efficacy of the various vaccine prioritization strategies using the vaccination campaign underway in the U.S.

The study developed a granular agent-based simulation model for mimicking community spread of COVID-19 under various social interventions including full and partial closures, isolation and quarantine, use of face mask and contact tracing, and vaccination. The model was populated with parameters of disease natural history, as well as demographic and societal data for an urban community in the U.S. with 2.8 million residents. The model tracks daily numbers of infected, hospitalized, and deaths for all census age-groups. The model was calibrated using

parameters for viral transmission and level of community circulation of individuals. Published data from the Florida COVID-19 dashboard was used to validate the model. Vaccination strategies were compared using a hypothesis test for pairwise comparisons.

Three prioritization strategies were examined: a minor variant of CDC's recommendation, an age-stratified strategy, and a random strategy. The impact of vaccination was also contrasted with a no vaccination scenario. The study showed that the campaign against COVID-19 in the U.S. using vaccines developed by Pfizer/BioNTech and Moderna 1) reduced the cumulative number of infections by 10% and 2) helped the pandemic to subside below a small threshold of 100 daily new reported cases sooner by approximately a month when compared to no vaccination. A comparison of the prioritization strategies showed no significant difference in their impacts on pandemic mitigation.

Even though vaccines for COVID-19 were developed and approved much quicker than ever before, their impact on pandemic mitigation was small as the explosive spread of the virus had already infected a significant portion of the population, thus reducing the susceptible pool. A notable observation from the study is that instead of adhering strictly to a sequential prioritizing strategy, focus should perhaps be on distributing the vaccines among all eligible as quickly as possible, after providing for the most vulnerable. As much of the population worldwide is yet to be vaccinated, results from this study should aid public health decision makers in effectively allocating their limited vaccine supplies.

Chapter 5: Threshold Analyses on Combinations of Testing, Population Size, and Vaccine Coverage for COVID-19 Control in a University Setting

The complete article titled "Threshold analyses on combinations of testing, population size, and vaccine coverage for COVID-19 control in a university setting" [6] (published in PLOS One) can be found in Appendix G. This article presents the framework for a compartmental SEIR-type (Susceptible Exposed Infected Recovered/Removed) model that simulates the COVID-19 epidemic in a residential university setting in the U.S. (University of Massachusetts, Amherst, MA). The simulation model is particularized for the university using demographics obtained from census and literature. The study identifies threshold levels for combinations of testing (symptomatic testing, tracing and testing, mass testing), contact rates, vaccine coverage, and population sizes that can keep the epidemic in the university below certain levels of tolerance. Threshold levels were evaluated for three tolerances (relaxed, medium, and tight). The study also implemented varying coefficients of transmission and contact rates based on interventions (such as social distancing and face mask usage). The study derives useful insights into preparedness for residential campuses across the country when deciding to reopen amid a pandemic. It also provides trade-offs to policy makers to decide what levels of resources might be required to create safe reopening possible for students, faculty, and staff.

5.1 Abstract

We simulated epidemic projections of a potential COVID-19 outbreak in a residential university population in the United States under varying combinations of asymptomatic tests (5% to 33% per day), transmission rates (2.5% to 14%), and contact rates (1 to 25), to identify the

contact rate threshold that, if exceeded, would lead to exponential growth in infections. Using this, we extracted contact rate thresholds among non-essential workers, population size thresholds in the absence of vaccines, and vaccine coverage thresholds. We further stream-lined our analyses to transmission rates of 5 to 8%, to correspond to the reported levels of face-mask-use/physical-distancing during the 2020 pandemic.

Our results suggest that, in the absence of vaccines, testing alone without reducing population size would not be sufficient to control an outbreak. If the population size is lowered to 34% (or 44%) of the actual population size to maintain contact rates at 4 (or 7) among non-essential workers, mass tests at 25% (or 33%) per day would help control an outbreak. With the availability of vaccines, the campus can be kept at full population provided at least 95% are vaccinated. If vaccines are partially available such that the coverage is lower than 95%, keeping at full population would require asymptomatic testing, either mass tests at 25% per day if vaccine coverage is at 63-79%, or mass tests at 33% per day if vaccine coverage is at 53-68%. If vaccine coverage is below 53%, to control an outbreak, in addition to mass tests at 33% per day, it would also require lowering the population size to 90%, 75%, and 60%, if vaccine coverage is at 38-53%, 23-38%, and below 23%, respectively.

Threshold estimates from this study, interpolated over the range of transmission rates, can collectively help inform campus level preparedness plans for adoption of face mask/physical-distancing, testing, remote instructions, and personnel scheduling, during non-availability or partial-availability of vaccines, in the event of SARS-Cov2-type disease outbreaks.

Chapter 6: Evaluating the Sensitivity of Jurisdictional Heterogeneity and Mixing in National-Level HIV Prevention Analyses: Context of the U.S. Ending the HIV Epidemic Goal

6.1 Introduction

In the United States (U.S.), there were an estimated 1.18 million people living with HIV (PWH) as of 2019, and an estimated average of 36,500 new infections each year between 2015 and 2019 [7]. Although HIV disease has no cure, consistent use of antiretroviral therapy treatment (ART) by infected persons can fully suppress viral load, thus preventing transmissions [8]. Further, pre-exposure prophylaxis (PrEP) for high-risk susceptible individuals can reduce HIV acquisition by 99% [9]. However, there are considerable gaps in administering these preventive tools. As of 2019, nationally, 87% of PWH were aware of their infection(proportion aware), but only 66% of those aware were on ART with viral load suppression (proportion with VLS) [10]. Among susceptible persons with PrEP eligibility, only 23% were administered PrEP. In addition, there is considerable heterogeneity in these care gaps by age, risk-group, and jurisdictions. Across geographical jurisdictions in the U.S., the proportion aware ranged from 80% to 96%, the proportion with VLS ranged from 49% to 83%, and PrEP coverage ranged from 6% to 93% [10]. Taking these jurisdictional disparities into consideration, the most recent U.S. national strategic plan, Ending the HIV Epidemic (EHE), in addition to continuing the age and risk-group focused efforts as in the previous national plan, also aims for jurisdictional focused efforts. Specifically, the EHE plan first aims to reduce national incidence by 75% by 2025 by focusing prevention efforts in 50 counties and 7 states (we will refer to these as the EHE jurisdictions), which had

accounted for more than 50% of nationwide diagnoses in 2017, and secondly reduce incidence by 90% by 2030 by expanding prevention efforts to the rest of the nation (we will refer to these as the non-EHE jurisdictions) [11]. The goal of the EHE plan is to reach care continuum targets of 95-95-95 (i.e., 95% awareness among PWH, 95% linkage to care among those aware, and 95% VLS among those aware) and PrEP target of 50% coverage among eligible, by 2025 in EHE jurisdictions, and by 2030 in all jurisdictions nation-wide [12].

Mathematical models play a key role in evaluating alternative combination interventions that are most-effective in achieving the EHE incidence goal. Recent literature includes multiple jurisdiction-specific models [6–13] and national-level models [21,22] that have conducted intervention analyses related to the U.S. EHE plan. The jurisdiction-specific models cover 6 EHE jurisdictions (Atlanta, Baltimore, Los Angeles, Miami, New York, and Seattle), however, each jurisdiction is evaluated independently, which ignores the interactions between jurisdictions. On the other hand, the national-level models do not consider the jurisdictional heterogeneity in population demographics and care, or the interactions between jurisdictions. In addition to the jurisdictional heterogeneity in care noted above, data show considerable jurisdictional differences in population size of men-who-have-sex-with-men (MSM) [23], who are a key high risk-group for HIV. Data also show significant partnership mixing between persons of different jurisdictions [24–27]. These data suggest that there is potential for strategies adopted in one jurisdiction to influence the nation-wide HIV incidence. However, these interactions between heterogeneous jurisdictions have not been studied, and thus, it is not clear if and to what extent it would alter the epidemic projections or decisions inferred through independent jurisdiction-specific models or generalized national-level models as those available in the current literature.

To address these gaps in literature, we developed a national HIV epidemic compartmental simulation model representative of the U.S. population and split into 96 jurisdictions. These jurisdictions represent 54 EHE jurisdictions (47 counties and 7 states) and the remaining represent 42 non-EHE jurisdictions (all 42 are states) (Table H.1). We did not model 3 of the 50 EHE counties and 1 non-EHE state due to data unavailability (see Table H.1). We simulated mixing between jurisdictions using varying assumptions for the magnitude of partnership mixing, which were informed by estimates from molecular network clusters that used nucleotide sequence data of persons with recent diagnoses of HIV to infer jurisdictional mixing [27]. Within each jurisdiction, the population was further split into risk-groups (HF, HM, and MSM) and age-groups (individuals from ages 13 to 100).

We used the simulation model to evaluate the following: 1) the sensitivity of jurisdictional-mixing and the sensitivity of jurisdictional heterogeneity in care (care-continuum and PrEP coverage) on the national HIV incidence estimates; 2) the sensitivity of jurisdictional-mixing on the projected national HIV incidence estimates when simulating the EHE plan, i.e., achieving the 95-95-95 care-continuum targets and 50% PrEP targets by 2025 in EHE-jurisdictions and by 2030 for all other jurisdictions; and 3) the sensitivity of jurisdictional-mixing on the intervention decisions inferred through simulated estimates, such as HIV-testing and retention-in-care necessary to achieve the EHE plan targets. These evaluations would help understand the implications of modeling jurisdictions independently, specifically, potential differences in model inferred decisions and corresponding incidence projections. Further, data for partnership mixing between persons of different jurisdictions are limited. Mixing data is indirectly inferred from molecular network clusters, which use nucleotide sequence data of persons with recent HIV diagnoses to infer close transmissions by creating pairwise links between persons with closely

related viral DNA sequence. Data from molecular network cluster analyses show that a significant number of pairwise links were formed between persons of different geographical jurisdictions [27]. However, links formed through nucleotide sequencing do not necessarily indicate direct transmissions. Therefore, the sensitivity analyses from this model can help inform future data collection related to jurisdictional heterogeneity and mixing, and subsequent model development. A model that can evaluate the national epidemic as a whole, with geographical heterogeneity in population demographics, HIV epidemic, and interventions, would help identify what jurisdiction-specific strategies to adopt, such as how often to test, what should be the aim for retention-in-care, and what should be the target for PrEP coverage to achieve the intended goals of the EHE.

6.2 Methods

6.2.1 Compartmental Model

We developed a compartmental simulation model of HIV stratified into three sexual risk-groups (HF, HM, MSM), eighty-eight age groups (individual ages from 13 to 100), four care continuum stages (Unaware, Aware no ART, ART no VLS, ART VLS), and five disease progression stages (Acute, CD4 > 500, CD4 350-500, CD4 200-350, CD4 <200). The flow diagram for care continuum and disease stage progression can be seen in Figure 6.1. We only simulated sexually transmitted cases of HIV and did not model injecting drug use. We developed two models, an aggregate National-Model, i.e., without geographical split (resulting in a total of 3 X 88 X 4 X 5 compartments), and a Jurisdictional-Model where the national population was further split into 96 geographical jurisdictions (resulting in a total of 3 X 88 X 4 X 5 X 96 compartments) as follows:

6.2.1.1 National-Model

The main purpose of the National-Model was, first, for calibration of sexual behavioral parameters specific to risk-group and age-group, for which data was more widely available at the national level, and second, as a comparison against the Jurisdictional-Model to evaluate the sensitivity of jurisdictional heterogeneity and mixing. We briefly discuss the model here and present the data used for model calibration in the Appendix H. We initialized the National-Model to the 2011 HIV epidemic, using data from the National HIV Surveillance System (NHSS) to distribute the population into the different compartments, and simulated the model for the period 2011 to 2018. We simulated using differential equations as typically done in compartmental modeling (see Appendix J).

For the rates of transitioning across disease stages (acute, and CD4-count stages) we used data from the literature (see Table H.2). For simulating the transitioning between the care continuum stages, we used data from literature for rate of achieving VLS when on ART and for rate of re-entry-to-care after dropping out, and data from the NHSS for the proportion linkage to care at diagnosis (see Table H.2).

The remaining two care continuum transition parameters, i.e., rates of HIV-diagnosis and care-drop-out, are dependent on testing and retention-in-care interventions, which are likely to change over time. Thus, we estimated these rates in the simulation by fitting to the annual NHSS data on care continuum distributions (see Appendix K).

We estimated incidence using a Bernoulli transmission equation (see Appendix L), using data related to behavior, transmission risk due to infected PWH in varying care continuum and disease stages, and coverage of PrEP for susceptible population. For behavioral data, we used information from the literature for age-group and risk-group specific mixing, number of partners,

number and type of acts per partner, and condom use and effectiveness (see Tables H.8, H.9, H.10, and H.11). We used simulated data for proportion of infected in care continuum and disease stage, along with NHSS data for PrEP coverage among susceptible persons. Note, PrEP coverage was simulated for only 2017 and 2018 as PrEP was only recently initiated in the U.S., and data were available for only this period. We used literature data for transmission risk associated with care continuum and disease stages, and PrEP coverage. Using the data ranges from the literature, we calibrated the per act probability of transmission, age-group mixing, and average number of sexual acts per partner by fitting simulated incidence to the surveillance estimates of incidence over the period 2011 to 2018.

6.2.1.2 Jurisdictional-Model

In this model, we further split the population into 96 jurisdictions using census data [28] for the overall population sizes, NHSS data for HIV population by age and risk-groups [29], and estimates from the literature for the proportion of MSM among adult males, [23] using jurisdiction-specific data when available. We did not model jurisdictions that did not have prevalence data, i.e., where data was either unavailable or suppressed (see Table H.1 for the list of jurisdictions modeled and excluded).

For the rates of transitioning across disease stages (acute, and CD4-count stages) we used the same literature data as in the National-Model. For simulating the transitioning between the care continuum stages, we used the same data as that in the National-Model for the rate of achieving VLS when on ART and re-entry-to-care after dropping out (see Table H.2).

The remaining two care continuum transition parameters, i.e., rates of HIV-diagnosis and care-drop-out, are related to testing and retention-in-care interventions, and thus, likely to vary across jurisdictions and over time. Thus, we evaluated multiple scenarios to test the sensitivity of

jurisdictional variations. We evaluated scenarios that applied the same rates estimated by the National-Model to every jurisdiction. We also evaluated scenarios that estimated rates by fitting to jurisdiction-specific NHSS data on care continuum distributions. These scenarios are described in more detail under section Scenarios Modeled.

As HIV surveillance data by jurisdiction were only available for 2017, 2018, and 2019 at the time of this study, we initialized the Jurisdiction-Model to 2017. Further, as data for care continuum distributions specific to risk-group and for each jurisdiction were not available, we assumed that the ratio of risk-group specific metric to overall population metric observed at the national-level would be the same as the ratio at the jurisdictional-level. Specifically, we applied this simplified assumption to two metrics within each jurisdiction, proportion aware and proportion with ART VLS. We normalized the values to ensure the sum of the proportions across all care continuum stages (Unaware, Aware no ART, ART no VLS, ART VLS) is equal to 1 for each jurisdiction. These modified care continuum distributions were used in estimation of rates for HIV-diagnosis and care-drop-out specific to risk-group and jurisdiction.

We estimated incidence rates using a Bernoulli transmission equation, using the same behavioral data (including the calibrated values) as in the National-Model for every jurisdiction, except that we also modeled interactions between jurisdictions by using a jurisdictional-mixing matrix, and used jurisdiction-specific simulated data for disease and care continuum stage among infected persons. We evaluated the sensitivity of jurisdictional variations in access to PrEP by implementing scenarios that used the national PrEP coverage for every jurisdiction and jurisdiction-specific PrEP coverage. To test the sensitivity of jurisdictional-mixing, we evaluated scenarios with varying levels of mixing, using data from molecular cluster analyses studies that

used nucleotide sequence data to infer transmission clusters between people [27]. We discuss these scenarios under section Scenarios Modeled.

6.3 Scenarios Modeled

We used the Jurisdictional-Model to simulate the HIV epidemic for the period 2018 to 2030. The time-unit in the model was monthly. We evaluated 16 scenarios to analyze the sensitivity of jurisdictional heterogeneity in care continuum and the sensitivity of jurisdictional mixing. Details of each scenario are discussed below and summarized in Table 6.1, and their broad differences are as follows. Scenarios 1 to 8 assumed homogeneity in care across jurisdictions by using national-level estimates for HIV-diagnosis rate, care-drop-out rate, and PrEP coverage for every jurisdiction, and homogeneity in risk-group distribution by assuming national-level estimates for the proportion of the population who are MSM. While scenarios 9 to 16 assumed heterogeneity in care across jurisdictions by using jurisdiction-specific HIV-diagnosis rate, care-drop-out rate, and PrEP coverage, and heterogeneity in risk-group distribution by assuming jurisdiction-specific estimates for proportion MSM. Scenarios 1 to 4 and 9 to 12 assumed continuation of status-quo interventions by using baseline year (2018) estimates for HIV-diagnosis rate, care-drop-out rate, and PrEP coverage, and keeping it constant over the period 2019 to 2030. Scenarios 5 to 8 and 13 to 16 modeled the EHE plan by using time-varying values for HIV-diagnosis rate, care-drop-out rate, and PrEP coverage, estimated to reach the EHE targets (95-95-95 care targets and 50% PrEP coverage among eligible) by 2025 for EHE and by 2030 for non-EHE jurisdictions (see Appendix K). Scenarios 1, 5, 9, and 13 assumed no mixing between jurisdictions, Scenarios 2, 6, 10, and 14 assumed lower levels of mixing between jurisdictions within the same state (Level-1-mixing), Scenarios 3, 7, 11, and 15 assumed higher levels of mixing between jurisdictions within the same state (Level-2-mixing), and Scenarios 4, 8, 12, and 16

assumed higher levels of mixing between jurisdictions within and outside the same state (Level-3-mixing).

6.3.1 Jurisdictional Mixing Assumptions

As noted above, we evaluated no-mixing and three types of mixing scenarios. Though data for mixing between jurisdictions are not available, data from molecular analyses studies, which creates pairwise links between persons with closely related viral DNA sequence, provides a reference point. One study showed that, about 47% to 65% of links were between persons of same county, about 78% to 88% between persons of same state, and the remaining between persons of different states [27] (summarized in Table 6.1). The range corresponds to the differences across risk groups. Because links do not represent direct transmissions, they do not represent partnership links. As the results from the above levels of mixing were sensitive to model outputs, to further test the sensitivity we evaluated a scenario with higher values of mixing within county, 85% to 90%, which would represent a scenario with low levels of outside mixing.

6.3.1.1 No Mix (Scenarios 1, 5, 9, and 13 in Table 6.1)

For these scenarios we assumed partnership mixing was 100% within jurisdiction, and 0% outside jurisdiction.

6.3.1.2 Level-1-Mixing (Scenarios 2, 6, 10, and 14 in Table 6.1)

For these scenarios we assumed persons in a jurisdiction could have partnerships with persons in other jurisdictions within the same state but not with persons in other states. If the jurisdiction modeled is an EHE county or a non-EHE state with EHE counties within it, we used the following for proportion mixing-within-jurisdiction: 90% for HM and HF, and 85% for MSM and used 1 minus mixing-within-jurisdiction for the proportion mixing with the rest of the jurisdictions within the state. If there were multiple EHE counties within a state, we split the value

(1 minus mixing-within-jurisdiction) equally between the other EHE jurisdictions and the rest of the state. If the jurisdiction modeled is a state with no EHE counties within it, we assumed 100% mix within their state.

6.3.1.3 Level-2-Mixing (Scenarios 3, 7, 11, and 15 in Table 6.1)

For these scenarios, as above, we assumed persons in a jurisdiction could have partnerships with persons in other jurisdictions within the same state but not with persons in other states, but assumed higher levels of outside mixing. If the jurisdiction modeled is an EHE county or a non-EHE state with EHE counties within it, we used the following data from [27] for the proportion mixing-within-jurisdiction: 57% for HM, 65% for HF, and 47% for MSM and used 1 minus mixing-within-jurisdiction for the proportion mixing with the rest of the jurisdictions within the state. If there were multiple EHE counties within a state, we split the value (1 minus mixing-within-jurisdiction) equally between the other EHE jurisdictions and the rest of the state. If the jurisdiction modeled is a state with no EHE counties within it, we assumed 100% mix within their state.

6.3.1.4 Level-3-Mixing (Scenarios 4, 8, 12, and 16 in Table 6.1)

For these scenarios we assumed persons in a jurisdiction could have partnership with persons in any jurisdiction. If the jurisdiction modeled is an EHE county or a non-EHE state with EHE counties within it, we used the following data from [27] for the proportion mixing-within-jurisdiction: 57% for HM, 65% for HF, and 47% for MSM. We used mixing-within-state minus mixing-within-jurisdiction data for the proportion mixing within state but outside their own jurisdiction (28% for HM, 23% for HF, and 31% for MSM), and distributed it equally among the jurisdictions within state. We used 1 minus mixing-within-state for mixing-outside-state and distributed it across all other states weighing by distance to the state. If the jurisdiction modeled is

a state without EHE counties within, we used mixing-within-state data from [27] for mixing within jurisdiction and distributed the remaining across all other states weighting by the distance to that state. We used the Euclidean distance between the geographical co-ordinates (latitude and longitude) of two states as a proxy for the distance between jurisdictions.

6.3.2 Evaluating Sensitivity of Jurisdictional Mixing Keeping Homogeneity in Care

We used the Jurisdictional-Model to evaluate the sensitivity of jurisdictional mixing when keeping interventions at status-quo, i.e., HIV-diagnosis rate, care-drop-out rate, and PrEP coverage constant over the period 2018 to 2030 at 2018 baseline values (Scenarios 1, 2, 3, and 4), and when scaling-up interventions over time to meet the 95-95-95 care and 50% PrEP targets (Scenarios 5, 6, 7, and 8). In these scenarios, we initialized the care continuum distribution of each jurisdiction to be equal to the national level for year 2017 year-end. This is explained in more detail below.

6.3.2.1 Baseline National (Homogeneity in Care Across Jurisdictions) (Scenarios 1, 2, 3, and 4 in Table 6.1)

In these scenarios, we used the baseline estimates derived by the National-model for rates of HIV-diagnosis and care-drop-out, fitted to the national care continuum distribution in 2018. We kept these rates constant for the following years (i.e., 2019 to 2030) for all jurisdictions. We also kept PrEP-coverage constant at 2018 national level for all years and all jurisdictions. While we used jurisdiction-specific data for PWH in each risk group, we assumed that the proportion of the population who are MSM is the same for every jurisdiction and used national-level estimates from [23].

6.3.2.2 EHE National (Homogeneity in Care Across Jurisdictions) (Scenarios 5, 6, 7, and 8 in Table 6.1)

In these scenarios, we used national-level estimates for the scale-up in interventions (HIV-diagnosis rate, care-drop-out rate, and PrEP coverage) to meet the EHE targets. Specifically, we used the National-Model to estimate the HIV-diagnosis rate, care-drop-out rate, and PrEP coverage necessary to linearly scale-up care continuum proportions and PrEP coverage from its national baseline values in year 2018 to the EHE targets. For EHE jurisdictions, the interventions were linearly scaled- over the period 2019 to 2025 and kept constant thereafter, and for the non-EHE jurisdictions, the values were kept constant for the period 2018 to 2025 and linearly scaled-up over the period 2026 to 2030. To recollect, the EHE targets were 95-95-95 for the care continuum and 50% for PrEP coverage among those eligible. In the U.S., PrEP eligibility is based on specific indicators, such as a person's risk factor for acquiring HIV and recency in sexually transmitted infection [30]. In the model, we do not simulate these PrEP indicators, thus, we use the reported numbers for persons with prescribed PrEP among those eligible to determine the percentage of susceptible population who are eligible and take 50% of that percentage as the EHE target for PrEP coverage. Thus, a 50% coverage among those eligible would approximately be equal to 15% coverage among susceptible.

6.3.3 Evaluating Sensitivity of Jurisdictional Mixing and Jurisdictional Heterogeneity in Care

We used the Jurisdictional-Model to evaluate the sensitivity of jurisdictional heterogeneity in care when keeping HIV-diagnosis rate, care-drop-out rate, and PrEP coverage constant over the period 2018 to 2030 (Scenarios 9, 10, 11, and 12), and when scaling-up over time to meet the 95-95-95 targets (Scenarios 13, 14, 15, and 16). Contrary to the case above (Scenarios 1 to 8), here we initialized the model to jurisdiction-specific care data for 2017 year-end and estimated

jurisdiction-specific HIV-diagnosis rates, care-drop-out rates, and PrEP-coverage for the period 2019 to 2030. This is explained in more detail below.

6.3.3.1 Baseline Jurisdiction-Specific (Heterogeneity in Care Across Jurisdictions) (Scenarios 9, 10, 11, and 12 in Table 6.1)

In these scenarios, we used the jurisdiction-specific estimates for HIV-diagnosis rate, care-drop-out rate, and PrEP coverage. We derived jurisdiction-specific HIV-diagnosis rate and care-drop-out rate in the Jurisdictional-Model by using the 2018 jurisdiction-specific care continuum distributions, and kept it constant for the period 2019 to 2030. We also used 2018 jurisdiction-specific estimates of PrEP coverage and kept it constant for the period 2019 to 2030.

6.3.3.2 EHE Jurisdiction-Specific (Heterogeneity in Care Across Jurisdictions) (Scenarios 13, 14, 15, and 16 in Table 6.1)

In these scenarios, we used jurisdiction-specific estimates for the scale-up in interventions (HIV-diagnosis rate, care-drop-out rate, and PrEP coverage) to meet the EHE targets. Specifically, we used the Jurisdictional-Model to estimate jurisdiction-specific HIV-diagnosis rate, care-drop-out rate, and PrEP coverage necessary to linearly scale-up care continuum proportions and PrEP coverage from its jurisdiction-specific baseline values in year 2018 to the EHE targets. For EHE jurisdictions, the interventions were linearly scaled- over the period 2019 to 2025 and kept constant thereafter, and for the non-EHE jurisdictions, the values were kept constant for the period 2019 to 2025 and linearly scaled-up over the period 2026 to 2030. To recollect, the EHE targets were 95-95-95 for the care continuum and 50% for PrEP coverage among those eligible. PrEP eligibility was determined in the same manner as described above for EHE national scenarios.

6.4 Model Verification and Output Metrics

As the sexual behavioral parameters in the National-Model were calibrated to the national incidence between 2011 and 2018, and because these parameters were then used in the Jurisdictional-Model, we first verified that the annual risk-group specific incidence simulated by the National-Model compares well with NHSS estimates for years 2011 to 2018 (Figure 6.2). Results suggest overall good fit to NHSS incidence for all three risk-groups.

For each of the 16 scenarios simulated in the Jurisdictional-Model, for each jurisdiction, we extract the following metrics for the period 2018 to 2030: incidence, prevalence, HIV-testing interval (using the inverse of the HIV-diagnosis rate as proxy), and retention-in-care rates (using 1 minus the care-drop-out rate as proxy). HIV testing intervals and retention-in-care rates serve as decision metrics to inform HIV testing and retention-in-care intervention programs, and incidence and prevalence projections serve as expected outcomes from implementing those decisions. We compare these metrics across the 16 scenarios to infer the sensitivity of model outputs to jurisdictional mixing and jurisdictional heterogeneity in care.

6.5 Results

While the risk-specific incidence estimates from the National-Model were within the range of NHSS estimates (as mentioned in Model Verification), the incidence estimates from the Jurisdictional-Model were not always within the range (Figure 6.2), which is expected as the mixing assumptions were national aggregated estimates and the jurisdictional model excluded some counties and states due to data unavailability. However, the magnitude of the estimates are close to the ranges, providing verification that the Jurisdictional-Model, which simulated local HIV epidemics in 96 jurisdictions can collectively generate results close to the overall national estimates.

For years 2018 and 2019, risk-specific incidence estimates simulated by the Jurisdictional-Model were sensitive to jurisdictional mixing (see Figure 6.2) (comparing no-mix Scenarios 1, 5, 9 and 13, with Level-1-mixing Scenarios 2, 6, 10, and 14, Level-2-mixing Scenarios 3, 7, 11, and 15, and Level-3-mixing Scenarios 4, 8, 12, and 16). Keeping care metrics (HIV-diagnosis rate, care-drop-out rate, and PrEP coverage) fixed at the 2018 baseline values, incidence projections for the period 2020 to 2030 were sensitive to jurisdictional mixing, both when assuming jurisdictional homogeneity in care (Scenario 1 compared to 2, 3 and 4) and jurisdictional heterogeneity in care (Scenario 9 compared to 10, 11, and 12) (Figure 6.2). Similar observations can be made for time varying estimates of overall incidence (Figure 6.3, Table 6.2). Although overall estimates of incidence were sensitive to jurisdictional mixing, the differences in incidence were much lower with jurisdictional heterogeneity in care, i.e., differences between Scenarios 9, 10, 11, and 12 were lower than differences between Scenarios 1, 2, 3, and 4. Compared to Scenario 9, the overall incidence in Scenarios 10, 11, and 12 changed by 2 to 2%, 7 to 5%, and 8 to 9%, respectively, the range corresponding to years 2018 to 2030 (Table 6.2), whereas, compared to Scenario 1, the overall incidence in Scenarios 2, 3, and 4 changed by 7 to 11%, 24 to 22%, and 24 to 26%, respectively (Table 6.2). In scenarios that scaled-up care metrics to meet EHE targets, differences in incidence were identical in year 2018, i.e., differences between Scenarios 5 to 8 (jurisdictional homogeneity) were identical to differences between Scenarios 1 to 4, and differences between Scenarios 13 to 16 (jurisdictional heterogeneity) were identical to difference between Scenarios 9 to 12, which is as expected as they start with the same baseline in 2018 (Table 6.3). However, as incidence significantly decreased over the period 2019 to 2030 because of scale-up of care, the differences between Scenarios 5 to 8, and differences between Scenarios 9 to 12 diminished (Figure 6.3 Table 6.3).

The differences in incidence estimates between no mixing and different levels of mixing assumptions observed in Figures 6.2 and 6.3 predominantly resulted from the non-EHE jurisdictions (see Figure 6.4, and are summarized in Tables 6.2 and 6.3). When assuming jurisdictional homogeneity in care, compared to no-mixing Scenario 1, incidence in Scenarios 2, 3, and 4 changed by 19 to 28%, 67 to 55%, and 60 to 60%, respectively, for non-EHE jurisdictions, whereas, it changed by -3 to -4%, -10 to -8%, and -5 to -4%, respectively, for EHE jurisdictions (Table 6.2), the range corresponding to years 2018 to 2030. Similarly, when assuming jurisdictional heterogeneity in care, compared to no-mixing Scenario 9, incidence in Scenarios 10, 11, and 12 changed by 15 to 7%, 18 to 15%, and 19 to 21%, respectively, for non-EHE jurisdictions, whereas, it changed by -1 to -1%, -2 to -2%, and 0 to 0%, respectively, for EHE jurisdictions (Table 6.2). As with overall incidence, differences in incidence split by EHE and non-EHE jurisdictions between Scenarios 5 to 8 were identical to differences between Scenarios 1 to 4, and differences between Scenarios 13 to 16 were identical to difference between Scenarios 9 to 12, which is as expected as they start with the same baseline in 2018 (Table 6.3).

In baseline year, 2018, though differences in incidence between mixing assumptions were minimal in most cases (overall or split by EHE and non-EHE) when assuming heterogeneity in care, i.e., between Scenarios 9 to 12 (Table 6.2), or between Scenarios 13 to 16 (Table 6.3), the differences at the individual jurisdictions varied over a wide range. Taking differences in incidence within each jurisdiction, compared to Scenario 13, incidence in Scenarios 14, 15, and 16 changed by -9% to 30%, -31% to 109%, and -27% to 94%, respectively, the range corresponds to data across individual jurisdictions (Table 6.4). Further, taking only EHE jurisdictions, compared to Scenario 13, incidence in Scenarios 14, 15, and 16, changed by -9% to 11%, -31% to 39%, and -27% to 46%, respectively (see Figure 6.5.a, Table 6.4). Considering only non-EHE jurisdictions,

compared to Scenario 13, incidence in Scenarios 14, 15, and 16, changed by -5% to 30%, -18% to 109%, and -11% to 94%, respectively (see Figure 6.5.b, Table 6.4). Differences in incidence for Scenario 13 compared to Scenario 14, 15, and 16, for EHE and non-EHE jurisdictions specific to risk groups are presented in Figures I.1.a and I.1.b for HM, Figures I.2.a and I.2.b for HF, and Figures I.3.a and I.3.b for MSM in Appendix I.

A consequence of the differences in the baseline jurisdictional-level incidence estimates is that the jurisdiction-level decisions inferred from the model would vary based on our mixing assumptions. We evaluate % change in testing interval across different mixing assumptions compared to the no-mixing scenario for risk group MSM (see Figures 6.6.a and 6.6.b, and Table 6.5). We grouped the testing interval into two cohorts: interval <2 years, and interval between 2-4 years.

In year 2022 (mid-intervention for EHE jurisdictions), when testing interval for no-mixing (Scenario 13) was < 2 years, interval for Scenario 14 was between 14% lower and 9% higher, Scenario 15 was between 32% lower and 25% higher, and Scenario 16 was between 35% lower and 20% higher. When testing interval for no-mixing (Scenario 13) was 2-4 years, interval for Scenario 14 was between 9% lower and 10% higher, Scenario 15 was between 23% lower and 36% higher, and Scenario 16 was between 20% lower and 34% higher.

In year 2025 (year reaching target for EHE jurisdictions), when testing interval for no-mixing (Scenario 13) was < 2 years, interval for Scenario 14 was between 23% lower and 5% higher, Scenario 15 was between 44% lower and 13% higher, and Scenario 16 was between 45% lower and 12% higher. When testing interval for no-mixing (Scenario 13) was 2-4 years, interval for Scenario 14 was between 2% higher and 10% higher, Scenario 15 was between 4% higher and 34% higher, and Scenario 16 was between 6% higher and 30% higher. The levels of HIV-testing

interval for HM are presented in Figures I.4.a (for EHE) and I.4.b (for non-EHE) and for HF are presented in Figures I.5.a (for EHE) and I.5.b (for non-EHE).

In year 2028 (mid-intervention for non-EHE jurisdictions), when testing interval for no-mixing (Scenario 13) was <2 years, interval for Scenario 14 was between 2% lower and 15% higher, Scenario 15 was between 0% lower and 53% higher, and Scenario 16 was between 16% lower and 48% higher. When testing interval for no-mixing (Scenario 13) was 2-4 years, interval for Scenario 14 was between 10% higher and 15% higher, Scenario 15 was between 34% higher and 60% higher, and Scenario 16 was between 36% higher and 37% higher.

In year 2030 (year reaching target for non-EHE jurisdictions), when testing interval for no-mixing (Scenario 13) was < 2 years, interval for Scenario 14 was between 4% lower and 13% higher, Scenario 15 was between 7% lower and 50% higher, and Scenario 16 was between 14% lower and 37% higher. Testing interval for no-mixing (Scenario 13) did not exceed 2 years for non-EHE jurisdictions. These differences show us that on an individual jurisdictional level, across mixing assumptions, there could be significant change in both outcomes (such as incidence) and future decisions (such as testing interval).

The estimated levels of retention-in-care were similar across all scenarios, and high, likely because of the high rates necessary to achieve the care continuum targets (Figures I.6.a and I.6.b for HM, Figures I.7.a and I.7.b for HF, and Figures I.8.a and I.8.b for MSM).

Comparing the estimates for overall reduction in incidence between EHE target scenarios and baselines scenarios within each mixing and care-continuum assumption, the differences were similar except for non-EHE jurisdictions (Table 6.6). While the expected incidence reduction in no-mixing assumption was 5% when assuming jurisdictional homogeneity in care (and 9% when assuming jurisdictional heterogeneity in care), the incidence reductions in level-1 mixing was 14%

(and 15%), level-2 mixing was 24% (and 24%), and level-3 mixing was 23% (and 25%) (Table 6.6).

We also compare prevalence (number of PWH) estimates from the Jurisdictional-Model for years 2018 and 2019 with the surveillance (NHSS) estimates in Figure 6.7. For the surveillance estimates, we plot both national estimates ('NHSS-National') for years 2017-2019 and sum of the 96 jurisdictions for 2017 ('NHSS-Sum of jurisdictions'). The 'NHSS-Sum of jurisdictions' were lower than the 'NHSS-National' estimates because of data suppression or unavailability in some jurisdictions. Simulated estimates of prevalence match close to the surveillance estimates for years 2017 to 2019, however, similar to the differences in incidence as seen in Figure 6.2, the projected estimates for years 2020 to 2030 are sensitive to jurisdictional mixing (no-mixing Scenarios 1, 5, 9 and 13, compared to mixing, i.e., Level-1-mixing Scenarios 2, 6, 10, and 14, Level-2-mixing Scenarios 3, 7, 11, and 15, and Level-3-mixing Scenarios 4, 8, 12, and 16). Simulated prevalence estimates were also sensitive to jurisdictional heterogeneity in care parameters (Scenarios 1 to 8, compared to Scenarios 9 to 16). Projections of PWH estimates for EHE and non-EHE jurisdictions are presented in Figure I.9 in Appendix I. Non-EHE jurisdictions have lower PWH than EHE jurisdictions for all scenarios. When considering national estimates of care metrics (Scenarios 1-8) projections of PWH for non-EHE jurisdictions are sensitive to mixing assumptions, while EHE jurisdictions are not significantly different. When considering jurisdiction-specific estimates of care metrics (Scenarios 9-16) projections of PWH for non-EHE jurisdictions are sensitive to mixing assumptions, while EHE jurisdictions are not significantly different for scenarios that fixed care metrics at baseline level (2018) (Scenarios 9-12). For scenarios that scaled-up care metrics to reach 95-95-95 targets (Scenarios 13-16), both, EHE and non-EHE jurisdictions do not see significant differences in PWH projections for mixing assumptions.

6.6 Conclusions

In this study, we developed a compartmental model to project the national HIV epidemic in the U.S. We first developed a National-Model, which calibrated risk-group and age-group specific sexual behavioral parameters at the national level for years 2011-2018. The national model was not split into individual jurisdictions. We then developed a Jurisdictional-Model, which split the population into 96 jurisdictions and simulated the national HIV epidemic in the U.S. from 2018-2030. The purpose of the Jurisdictional-Model was to evaluate the sensitivity of jurisdictional-mixing and the sensitivity of jurisdictional heterogeneity in care (care-continuum distributions and PrEP coverage) on national HIV incidence estimates, PWH estimates, evaluate the sensitivity of jurisdictional-mixing on the projected national HIV incidence estimates when simulating the EHE plan, , and evaluate the sensitivity of jurisdictional-mixing on the simulated estimates of HIV-testing interval and retention-in-care rates necessary to achieve the targets in the EHE plan.

We believe this is the first model that simulates the U.S. national HIV epidemic through simulating interacting individual sub-geographical jurisdictions. Our results suggest that incidence and PWH estimates are sensitive to jurisdictional-mixing assumptions and heterogeneity in care metrics, more so in earlier years when incidence was high and intervention was low. Intervention implemented as scaled-up care metrics to reach targets minimize the differences in mixing assumptions over time. However, decisions related to HIV-testing inferred from the model are sensitive to jurisdictional-mixing assumptions and heterogeneity in care metrics. As the level of partnership-mixing between jurisdictions increases, the differences increased. Another important observation from our study is that individual level differences in mixing may be significantly different for each jurisdiction. These results suggest that when modeling jurisdictions

independently, understanding the magnitude and accounting for the mixing outside jurisdiction could lead to better decisions.

Our model has limitations. Our model only simulates sexual partnership and does not model transmission due to injecting drug use, which are a high-risk demographic for HIV. Assumptions of mixing and jurisdiction-specific care continuum bear a significant impact on incidence and prevalence estimates; thus, our model is currently limited to be used as a tool to evaluate trends of decisions. With more accurate data related to mixing and care continuum proportions specific to each jurisdiction, the model can be further improvised to project incidence and prevalence estimates for each jurisdiction and make informed real-time decisions about care metrics. We assumed no change in behavioral, demographic, and disease transmission factors over time. Thus, the incidence estimates are sensitive to changes in these factors.

In summary, our results show that the aggregated national incidence estimates in 2030 with the implementation of the EHE plan, were not sensitive to jurisdictional mixing when assuming jurisdictional heterogeneity in care ,however, within each jurisdiction, incidence were significantly sensitive to mixing. The consequences of this were that the decisions inferred from the model to reach the EHE goal were sensitive to jurisdictional mixing and jurisdictional heterogeneity in care. The estimates for overall reductions in incidence between continuing status-quo versus implementing EHE plan were also sensitive to jurisdictional mixing assumptions but were not sensitive to assumptions on heterogeneity in care.

Table 6.1 Scenarios simulated using the jurisdictional model

<table>
<tr><th>Scenario no.</th><th>Scenario Name (mixing data; care data)</th><th>Mixing data‡</th><th colspan="2">Care data (rates of HIV-diagnosis, care-drop-out, and level of PrEP coverage)</th></tr>
<tr><td></td><td></td><td></td><td>EHE jurisdictions</td><td>non-EHE jurisdictions</td></tr>
<tr><td colspan="5">TO EVALUATE THE IMAPCT OF MIXING</td></tr>
<tr><td>[1]</td><td>No-mixing; Baseline national</td><td>No</td><td>National estimates kept constant, at 2018 national values, for all jurisdictions and for all years</td><td>National estimates kept constant, at 2018 national values, for all jurisdictions and for all years</td></tr>
<tr><td>[2]</td><td>Level-1-Mixing; Baseline national</td><td><table><tr><th>Risk group</th><th>HM</th><th>HF</th><th>MSM</th></tr><tr><td>Same juri</td><td>90 %</td><td>90%</td><td>85 %</td></tr><tr><td>Other juri same state</td><td>10 %</td><td>10%</td><td>15 %</td></tr><tr><td>Other states</td><td>0</td><td>0</td><td>0</td></tr></table></td><td>National estimates kept constant, at 2018 national values, for all jurisdictions and for all years</td><td>National estimates kept constant, at 2018 national values, for all jurisdictions and for all years</td></tr>
<tr><td>[3]</td><td>Level-2-mixing; Baseline national</td><td><table><tr><th>Risk group</th><th>HM</th><th>HF</th><th>MSM</th></tr><tr><td>Same juri</td><td>57 %</td><td>65%</td><td>47 %</td></tr><tr><td>Other juri same state</td><td>43 %</td><td>35%</td><td>53 %</td></tr><tr><td>Other states</td><td>0</td><td>0</td><td>0</td></tr></table></td><td>National estimates kept constant, at 2018 national values, for all jurisdictions and for all years</td><td>National estimates kept constant, at 2018 national values, for all jurisdictions and for all years</td></tr>
</table>

Table 6.1 (Continued)

[4]	***Level-3-mixing; Baseline national***	***Risk group***	***HM***	***HF***	***MSM***	National estimates kept constant, at 2018 national values, for all jurisdictions and for all years	National estimates kept constant, at 2018 national values, for all jurisdictions and for all years
		Same juri	57 %	65%	47 %		
		Other juri same state	28 %	23%	31 %		
		Other states	14 %	12%	22 %		
[5]	***No-mixing; EHE national***	No				National estimates calibrated to nationally achieve EHE targets (95-95-95) by 2025, and kept constant at 2025 value thereafter	National estimates kept constant at 2018 national values until 2025, and thereafter, calibrated to nationally achieve EHE targets (95-95-95) by 2030
[6]	***Level-1-mixing; EHE national***	***Risk group***	***HM***	***HF***	***MSM***	National estimates calibrated to nationally achieve EHE targets (95-95-95) by 2025, and kept constant at 2025 value thereafter	National estimates kept constant at 2018 national values until 2025, and thereafter, calibrated to nationally achieve EHE targets (95-95-95) by 2030
		Same juri	90 %	90%	85 %		
		Other juri same state	10 %	10%	15 %		
		Other states	0	0	0		
[7]	***Level-2-mixing; EHE national***	***Risk group***	***HM***	***HF***	***MSM***	National estimates calibrated to nationally achieve EHE targets (95-95-95) by 2025, and kept constant at 2025 value thereafter	National estimates kept constant at 2018 national values until 2025, and thereafter, calibrated to nationally achieve EHE targets (95-95-95) by 2030
		Same juri	57 %	65%	47 %		
		Other juri same state	43 %	35%	53 %		
		Other states	0	0	0		

Table 6.1 (Continued)

[8]	***Level-3-mixing; EHE national***	***Risk group***	***HM***	***HF***	***MSM***	National estimates calibrated to nationally achieve EHE targets (95-95-95) by 2025, and kept constant at 2025 value thereafter	National estimates kept constant at 2018 national values until 2025, and thereafter, calibrated to nationally achieve EHE targets (95-95-95) by 2030
		Same juri	57 %	65%	47 %		
		Other juri same state	28 %	23%	31 %		
		Other states	14 %	12%	22 %		
TO EVALUATE THE IMAPCT OF CARE CONTINUUM ACCESS							
[9]	***No-mixing; Baseline jurisdiction-specific***	No				Jurisdiction-specific estimates kept constant for all years at 2018 baseline values.	Jurisdiction-specific estimates kept constant for all years at 2018 baseline values.
[10]	***Level-1-mixing; Baseline jurisdiction-specific***	***Risk group***	***HM***	***HF***	***MSM***	Jurisdiction-specific estimates kept constant for all years at 2018 baseline values.	Jurisdiction-specific estimates kept constant for all years at 2018 baseline values.
		Same juri	90 %	90%	85 %		
		Other juri same state	10 %	10%	15 %		
		Other states	0	0	0		
[11]	***Level-2-mixing; Baseline jurisdiction-specific***	***Risk group***	***HM***	***HF***	***MSM***	Jurisdiction-specific estimates kept constant for all years at 2018 baseline values.	Jurisdiction-specific estimates kept constant for all years at 2018 baseline values.
		Same juri	57 %	65%	47 %		
		Other juri same state	43 %	35%	53 %		
		Other states	0	0	0		

Table 6.1 (Continued)

[12]	***Level-3-mixing; Baseline jurisdiction-specific***	***Risk group***	***HM***	***HF***	***MSM***	Jurisdiction-specific estimates kept constant for all years at 2018 baseline values.	Jurisdiction-specific estimates kept constant for all years at 2018 baseline values.
		Same juri	57 %	65%	47 %		
		Other juri same state	28 %	23%	31 %		
		Other states	14 %	12%	22 %		
[13]	***No-mixing; EHE jurisdiction-specific***	No				Jurisdiction-specific estimates calibrated to achieve EHE targets (95-95-95) within each jurisdiction by 2025, and kept constant at 2025 value thereafter	Jurisdiction-specific estimates kept constant at 2018 values until 2025, and thereafter, calibrated to achieve EHE targets (95-95-95) by 2030 within each jurisdiction
[14]	***Level-1-mixing; EHE jurisdiction-specific***	***Risk group***	***HM***	***HF***	***MSM***	Jurisdiction-specific estimates calibrated to achieve EHE targets (95-95-95) within each jurisdiction by 2025, and kept constant at 2025 value thereafter	Jurisdiction-specific estimates kept constant at 2018 values until 2025, and thereafter, calibrated to achieve EHE targets (95-95-95) by 2030 within each jurisdiction
		Same juri	90 %	90%	85 %		
		Other juri same state	10 %	10%	15 %		
		Other states	0	0	0		
[15]	***Level-2-mixing; EHE jurisdiction-specific***	***Risk group***	***HM***	***HF***	***MSM***	Jurisdiction-specific estimates calibrated to achieve EHE targets (95-95-95) within each jurisdiction by 2025, and kept constant at 2025 value thereafter	Jurisdiction-specific estimates kept constant at 2018 values until 2025, and thereafter, calibrated to achieve EHE targets (95-95-95) by 2030 within each jurisdiction
		Same juri	57 %	65%	47 %		
		Other juri same state	43 %	35%	53 %		
		Other states	0	0	0		

Table 6.1 (Continued)

[16]	***Level-3-mixing; EHE jurisdiction-specific***	***Risk group***	***HM***	***HF***	***MSM***	Jurisdiction-specific estimates calibrated to achieve EHE targets (95-95-95) within each jurisdiction by 2025, and kept constant at 2025 value thereafter	Jurisdiction-specific estimates kept constant at 2018 values until 2025, and thereafter, calibrated to achieve EHE targets (95-95-95) by 2030 within each jurisdiction
		Same juri	57 %	65%	47 %		
		Other juri same state	28 %	23%	31 %		
		Other states	14 %	12%	22 %		

Table 6.2 Percentage difference between incidence estimates for no mixing when compared to mixing for year 2018 and 2030. For scenarios that keep care metrics (HIV-diagnosis rate, care-drop-out rate, and PrEP coverage) fixed at the 2018 baseline values (Scenarios 1-4 and 9-12).

Scenario # (Baseline)	1	2	3	4	9	10	11	12
	Homogeneity in care across jurisdictions				Heterogeneity in care across jurisdictions			
	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3
Overall	Ref	7% and 11%	24% and 22%	24% and 26%	Ref	2% and 2%	7% and 5%	8% and 9%
EHE	Ref	-3% and -4%	-10% and -8%	-5% and -4%	Ref	-1% and -1%	-2% and -2%	0% and 0%
Non-EHE	Ref	19% and 28%	67% and 55%	60% and 60%	Ref	5% and 7%	18% and 15%	19% and 21%

Values represent the % change in incidence estimate in mixing scenario compared to no mixing scenario and calculated as 100(mixing scenario - no mixing scenario)/mixing scenario).
The range refers to the minimum and maximum value.

Table 6.3 Percentage difference between incidence estimates for no mixing when compared to mixing for year 2018 and 2030. For scenarios that scale up care metrics (HIV-diagnosis rate, care-drop-out rate, and PrEP coverage) to reach EHE targets by 2025 for EHE jurisdictions and by 2030 for non-EHE jurisdictions (Scenarios 5-8 and 13-16).

Scenario # (EHE targets)	5	6	7	8	13	14	15	16
	Homogeneity in care across jurisdictions				Heterogeneity in care across jurisdictions			
	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3
Overall	Ref	7% and 4%	24% and 6%	24% and 10%	Ref	2% and -1%	7% and -3%	8% and 1%
EHE	Ref	-3% and 0%	-10% and 1%	-5% and 7%	Ref	-1% and 4%	-2% and 8%	0% and 15%
Non-EHE	Ref	19% and 7%	67% and 9%	60% and 13%	Ref	5% and -4%	18% and -11%	19% and -9%

Values represent the % change in incidence estimate in mixing scenario compared to no mixing scenario and calculated as 100(mixing scenario - no mixing scenario)/mixing scenario).
The range refers to the minimum and maximum value

Table 6.4 Percentage difference between incidence estimates across jurisdictions for no mixing when compared to mixing for baseline year 2018. For scenarios that assume jurisdictional heterogeneity and scale up care metrics (HIV-diagnosis rate, care-drop-out rate, and PrEP coverage) to reach EHE targets by 2025 for EHE jurisdictions and by 2030 for non-EHE jurisdictions (Scenarios 13-16).

Scenario #	13	14	15	16
	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3
Overall	Ref	-9% to 30%	-31% to 109%	-27% to 94%
EHE	Ref	-9% to 11%	-31% to 39%	-27% to 46%
Non-EHE	Ref	-5% to 30%	-18% to 109%	-11% to 94%

Footnotes:
Values represent the % change in incidence estimate in mixing scenario compared to no mixing scenario and calculated as 100(mixing scenario - no mixing scenario)/mixing scenario).
The range refers to the minimum and maximum value across jurisdictions.

Table 6.5 Percentage difference between HIV testing interval estimates for risk group MSM. Difference calculated across jurisdictions for no mixing when compared to mixing for mid-intervention years (2022 for EHE and 2028 for non-EHE) and target year (2025 for EHE and 2030 for non-EHE). For scenarios that assume jurisdictional heterogeneity and scale up care metrics (HIV-diagnosis rate, care-drop-out rate, and PrEP coverage) to reach EHE targets by 2025 for EHE jurisdictions and by 2030 for non-EHE jurisdictions (Scenarios 13-16).

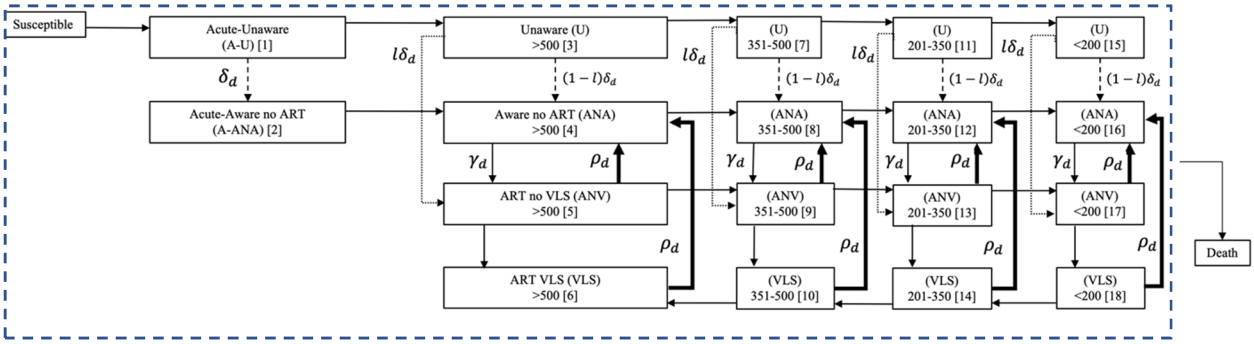
		Mid intervention year (2022 for EHE and 2028 for non-EHE)				Target year (2025 for EHE and 2030 for non-EHE)			
Scenario #		13	14	15	16	13	14	15	16
Jur. type	Interval Cohort	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3
EHE	< 2 years	Ref	-14% to 9%	-32% to 25%	-35% to 20%	Ref	-23% to 5%	-44% to 13%	-45% to 12%
	2-4 years	Ref	-9% to 10%	-23% to 36%	-20% to 34%	Ref	2% to 10 %	4% to 34%	6% to 30%
Non-EHE	< 2 years	Ref	-2% to 15%	0% to 53%	-16% to 48%	Ref	-4% to 13%	-7% to 50%	-14% to 37%
	2-4 years	Ref	10% to 15%	34% to 60%	36% to 37%	Ref	-	-	-

Values represent the % change in HIV-testing interval estimate in mixing scenario compared to no mixing scenario and calculated as 100(mixing scenario - no mixing scenario)/mixing scenario).
The range refers to the minimum and maximum value across jurisdictions for the specific time interval cohort (i.e., < 2 years or 2-4 years).

Table 6.6 Percentage reduction in cumulative incidence over time (2018-2030) from baseline scenario compared to EHE target scenario.

Scenario # (EHE targets - baseline)	1 and 5	2 and 6	3 and 7	4 and 8	9 and 13	10 and 14	11 and 15	12 and 16
	Homogeneity in care across jurisdictions				Heterogeneity in care across jurisdictions			
	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3	No-mixing	Mixing level 1	Mixing level 2	Mixing level 3
Overall	-24%	-27%	-30%	-30%	-30%	-31%	-33%	-33%
EHE	-41%	-40%	-38%	-40%	-45%	-44%	-41%	-41%
Non-EHE	-5%	-14%	-24%	-23%	-9%	-15%	-24%	-25%

Values represent the % reduction in cumulative incidence estimate for baseline scenario and EHE target scenario (i.e., (EHE target scenario-baseline scenario)/ baseline scenario).



Dotted arrow is disease stage specific HIV-diagnosis rate when linked to care at diagnosis, represented by $l\delta_d$

Dashed arrow is disease stage specific HIV-diagnosis rate when not linked to care at diagnosis, represented by $(1-l)\delta_d$

Thick arrow is disease stage specific care-drop-out rate, represented by ρ_d

γ_d represents the disease stage specific rate of re-entry-to-care

Figure 6.1 Compartmental simulation model. Transition diagram for care continuum stages and disease stages.

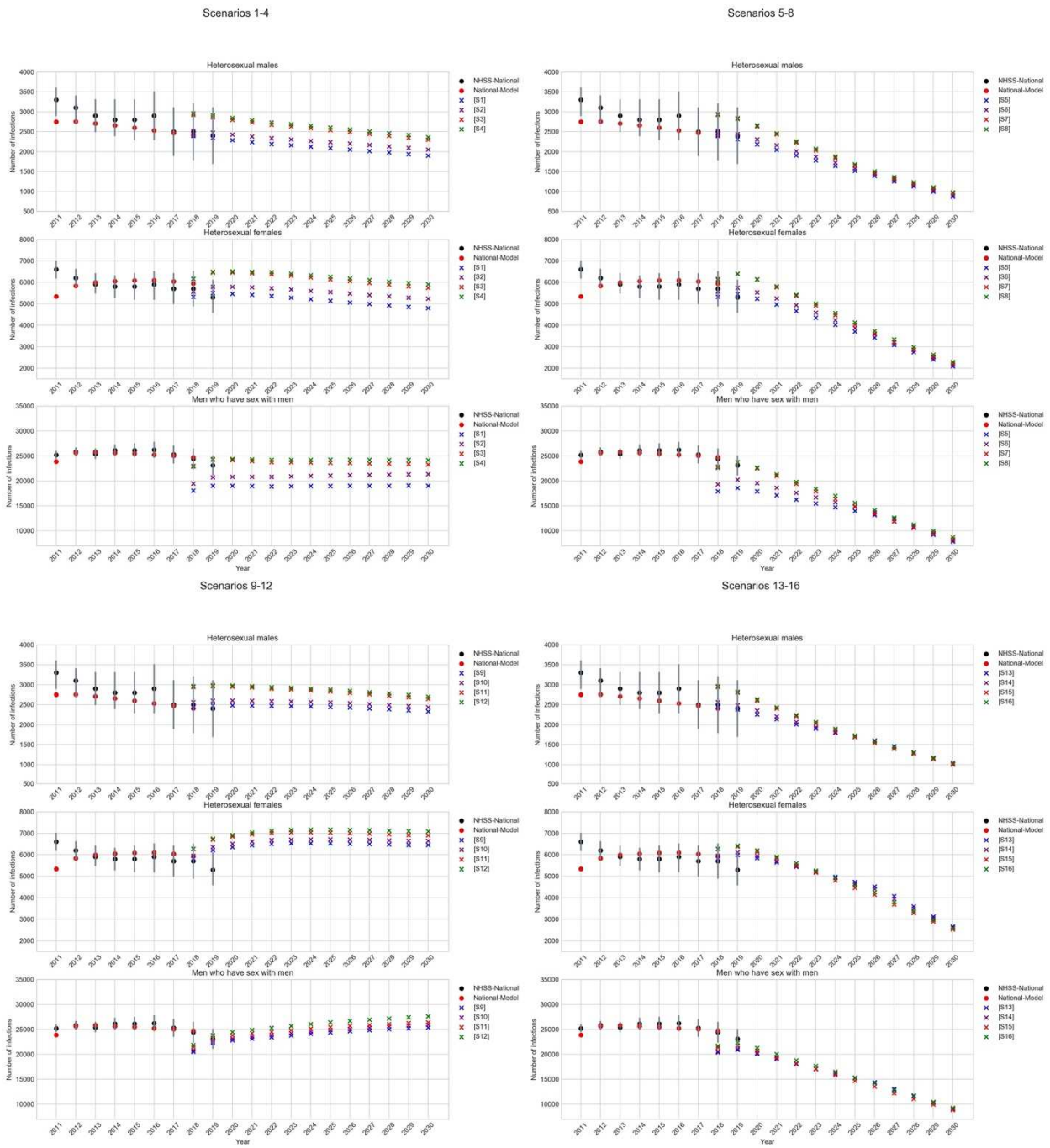


Figure 6.2 Comparing annual risk-group specific incidence simulated from the National-Model and 16 Jurisdictional-Model scenarios (S1 to S16) with NHSS estimates. Comparing national estimates for the period 2011 to 2019. Comparing model projections between the 16 Jurisdiction-Model scenarios for the period 2020 to 2030. NHSS: National HIV Surveillance System; NHSS-National: national level estimates from NHSS.

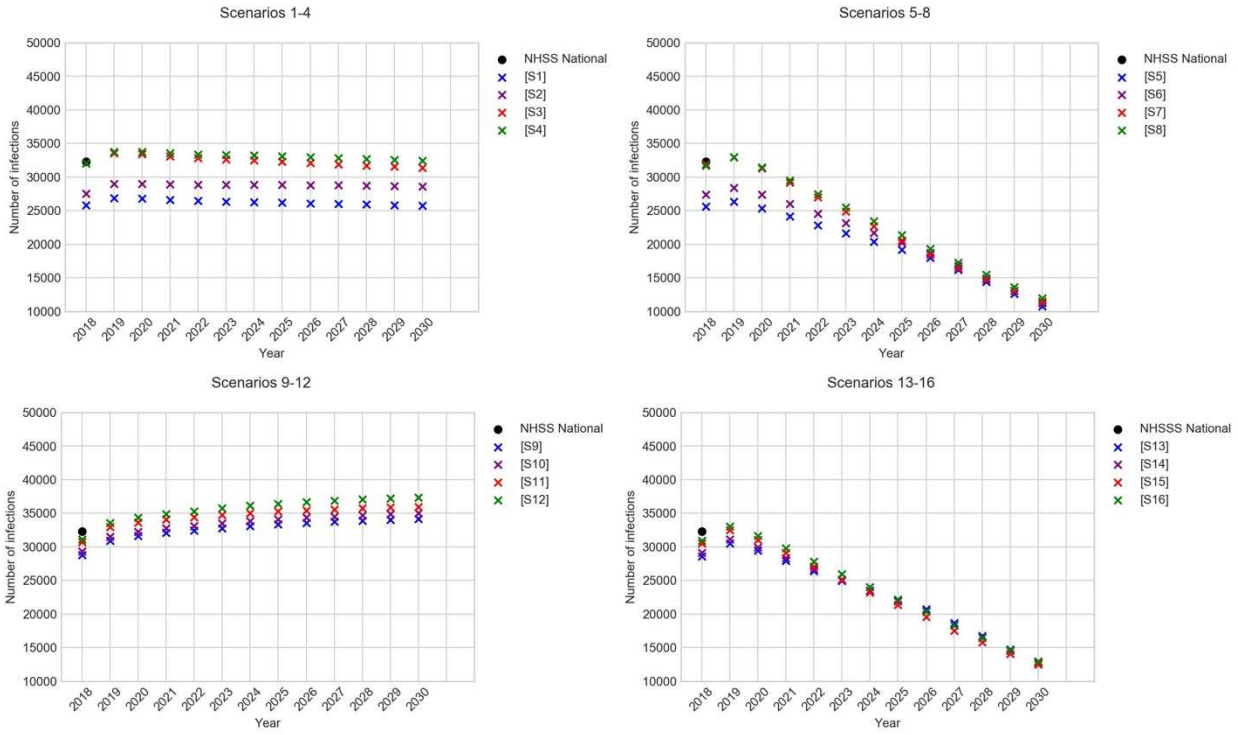


Figure 6.3 Comparing annual incidence simulated from the 16 Jurisdictional-Model scenarios (S1 to S16) with NHSS estimate for the year 2018. Comparing model projections between the 16 Jurisdiction-Model scenarios for the period 2019 to 2030. NHSS: National HIV Surveillance System; NHSS-National: national level estimates from NHSS.

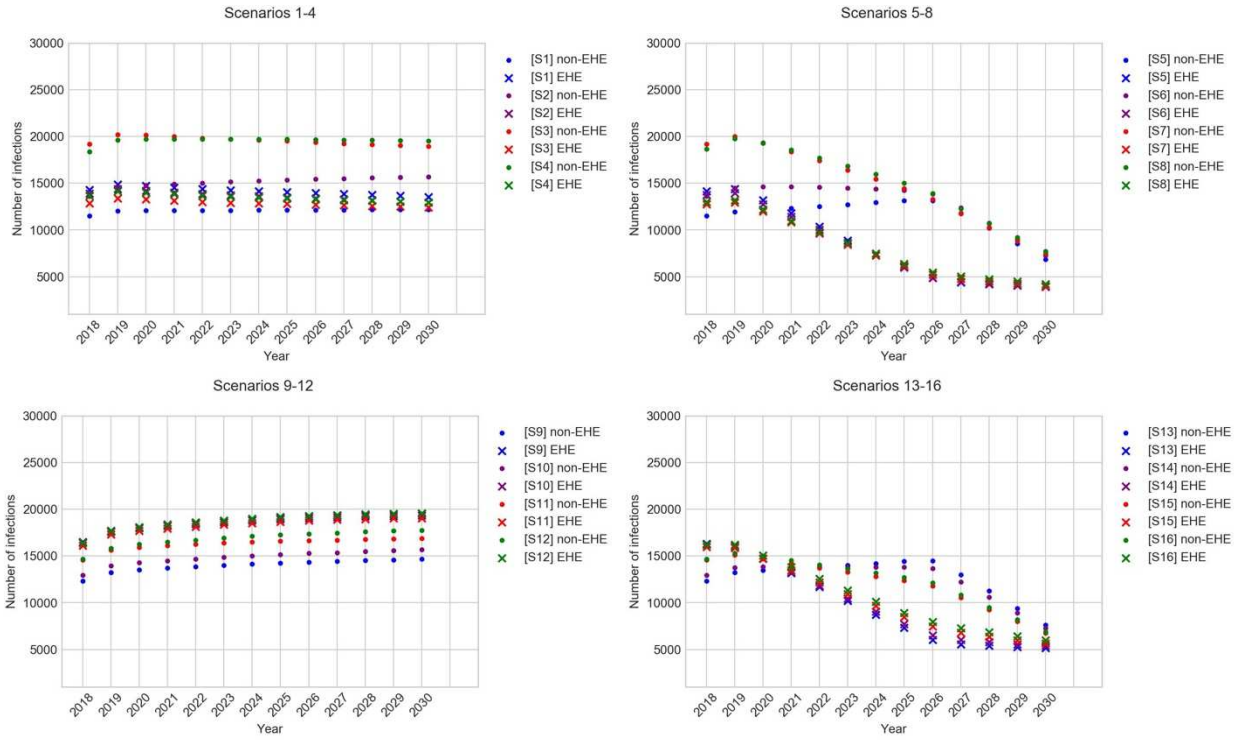


Figure 6.4 Comparing annual incidence projections between the simulated 16 Jurisdictional-Model scenarios (S1 to S16) for aggregate EHE and non-EHE jurisdictions for the period 2018 to 2030.

Figure 6.5.a EHE jurisdiction; Comparing percentage difference of overall incidence between mixing scenarios. Comparing no mixing (Scenario 13) with Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16) for each EHE jurisdiction for year 2018.

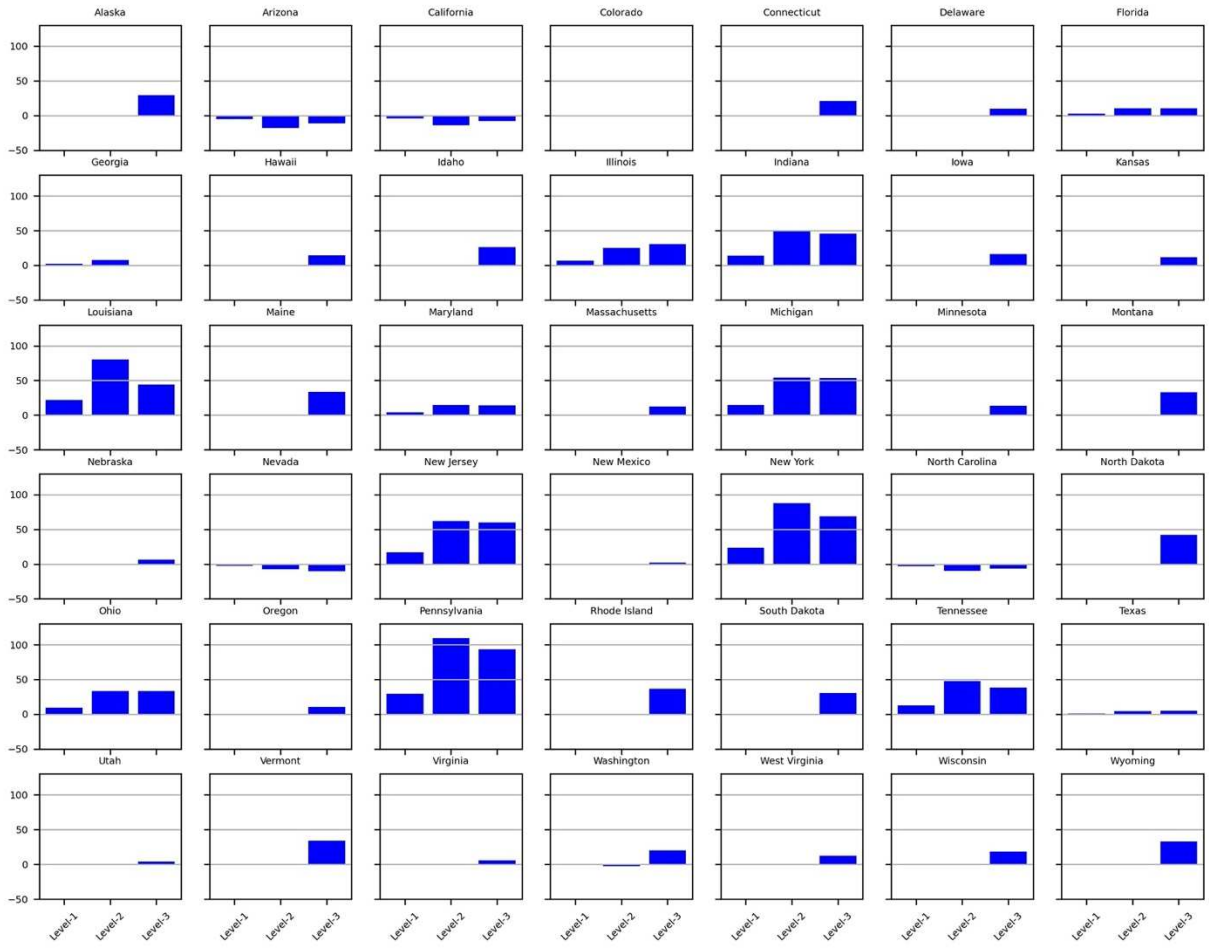


Figure 6.5.b Non-EHE jurisdiction; Comparing percentage difference of overall incidence between mixing scenarios. Comparing no mixing (Scenario 13) with Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16) for each non-EHE jurisdiction for year 2018.

Figure 6.6.a EHE jurisdictions; Comparing HIV interval between testing in months across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group MSM. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for EHE jurisdictions applied in years 2019-2025. For each year the order of HIV interval between testing is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively. Testing interval limit on y-axis was capped at 4 years (48 months) as intervals during intervention years did not exceed 4 years.

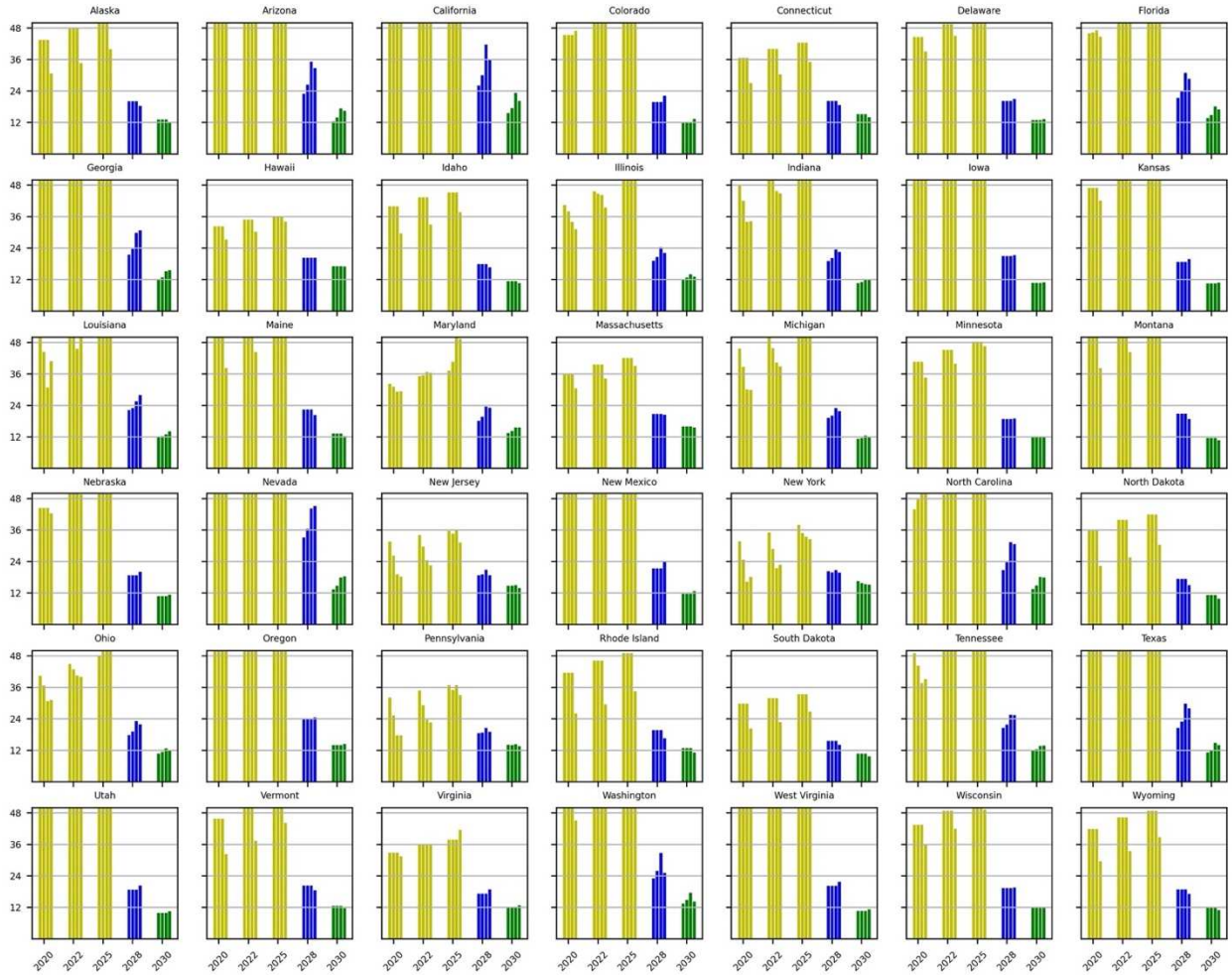


Figure 6.6.b Non-EHE jurisdictions; Comparing HIV interval between testing in months across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group MSM. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for non-EHE jurisdictions applied in years 2026-2030. For each year the order of HIV interval between testing is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively. Testing interval limit on y-axis was capped at 4 years (48 months) as intervals during intervention years did not exceed 4 years.

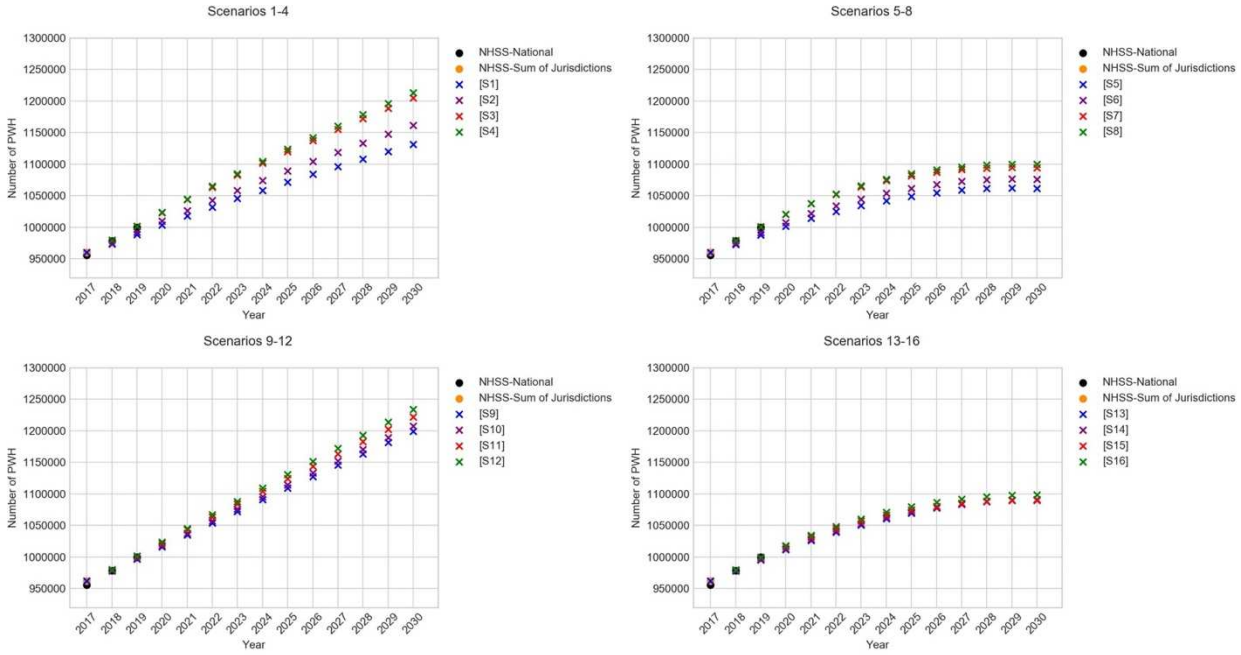


Figure 6.7 Comparing PWH simulated from the 16 Jurisdiction -Model scenarios (S1 to S16) for the period with NHSS national and aggregate jurisdiction estimates for the period 2018 to 2030. NHSS: National HIV Surveillance System; NHSS-National: national level estimates from NHSS; NHSS-Sum of Jurisdictions: sum of jurisdiction level estimates from NHSS.

Chapter 7: Conclusions

Modeling tools are an asset in understanding existing epidemics and emerging epidemics. The studies presented in this dissertation provide valuable contributions to existing literature on the frameworks required to tackle disease outbreaks on different levels and scales of population (school/university level, county level, state level, or nation level). The methodologies presented serve as a tool for analyzing trends in epidemic outcomes (infections, reported cases, hospitalizations, and deaths) and evaluate mitigation measures (pharmaceutical and non-pharmaceutical interventions) that successfully reduce the said outcomes. The intellectual merit of this research is the frameworks that can analyze varied interventions for different scales of population sizes. Further, the research also evaluates two key parameters that have impact on transmission of infection: mixing of populations across different jurisdictions, and changes in population demographics and access to care across jurisdictions.

The outcomes of this research hope to help make decisions not just for the epidemics modeled (i.e., COVID-19 and HIV/AIDS) but also for disease epidemics that share similar transmission dynamics and disease natural history. Also, the case study regions only serve as an example, but the models can be re-particularized for other regions. The models and outcomes were developed to aid public health policy makers in evaluating decisions either for novel disease outbreaks or for existing epidemics.

References

1. Brauer F. Mathematical epidemiology: Past, present, and future. Infectious Disease Modelling. 2017;2: 113–127. doi:10.1016/j.idm.2017.02.001

2. Lang T. Plug COVID-19 research gaps in detection, prevention and care. In: WORLD VIEW [Internet]. 15 Jul 2020 [cited 2 Nov 2021]. Available: https://www.nature.com/articles/d41586-020-02004-1

3. Tatapudi H, Das R, Das T. Impact assessment of full and partial stay-at-home orders, face mask usage, and contact tracing: An agent-based simulation study of COVID-19 for an urban region. Global Epidemiology. 2020;2. doi:https://doi.org/10.1016/j.gloepi.2020.100036

4. Tatapudi H, Das TK. Impact of school reopening on pandemic spread: A case study using an agent-based model for COVID-19. Infectious Disease Modelling. 2021;6: 839–847. doi:10.1016/j.idm.2021.06.007

5. Tatapudi H, Das R, Das T. Impact of Vaccine Prioritization Strategies on Mitigating COVID-19: An Agent-Based Simulation Study using an Urban Region in the United States. MedRxiv. 2021.

6. Zhao X, Tatapudi H, Corey G, Gopalappa C. Threshold analyses on combinations of testing, population size, and vaccine coverage for COVID-19 control in a university setting. Eksin C, editor. PLoS ONE. 2021;16: e0255864. doi:10.1371/journal.pone.0255864

7. HIV.gov. U.S. Statistics. 2 Jun 2021 [cited 12 Aug 2021]. Available: https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics

8. NIH. HIV Treatment. [cited 12 Aug 2021]. Available: https://hivinfo.nih.gov/understanding-hiv/fact-sheets/hiv-treatment-basics

9. CDC. PrEP Effectiveness. 13 May 2021 [cited 20 Aug 2021]. Available: https://www.cdc.gov/hiv/basics/prep/prep-effectiveness.html

10. AHEAD. What is AHEAD? [cited 12 Aug 2021]. Available: https://ahead.hiv.gov/about

11. HIV.gov. What Is Ending the HIV Epidemic in the U.S.? 2 Jun 2021 [cited 12 Aug 2021]. Available: https://www.hiv.gov/federal-response/ending-the-hiv-epidemic/overview

12. AHEAD. FAQs. In: EHE Goals: What are the 2025 goals and 2030 goals for each of the indicators? [Internet]. [cited 20 Aug 2021]. Available: https://ahead.hiv.gov/faqs

13. Krebs E, Enns B, Wang L, Zang X, Panagiotoglou D, Del Rio C, et al. Developing a dynamic HIV transmission model for 6 U.S. cities: An evidence synthesis. Cipriano L, editor. PLoS ONE. 2019;14: e0217559. doi:10.1371/journal.pone.0217559

14. Nosyk B, Zang X, Krebs E, Min JE, Behrends CN, Del Rio C, et al. Ending the Epidemic in America Will Not Happen if the Status Quo Continues: Modeled Projections for Human Immunodeficiency Virus Incidence in 6 US Cities. Clinical Infectious Diseases. 2019;69: 2195–2198. doi:10.1093/cid/ciz1015

15. Zang X, Krebs E, Min JE, Pandya A, Marshall BDL, Schackman BR, et al. Development and Calibration of a Dynamic HIV Transmission Model for 6 US Cities. Med Decis Making. 2020;40: 3–16. doi:10.1177/0272989X19889356

16. Nosyk B, Zang X, Krebs E, Enns B, Min JE, Behrends CN, et al. Ending the HIV epidemic in the USA: an economic modelling study in six cities. The Lancet HIV. 2020;7: e491–e503. doi:10.1016/S2352-3018(20)30033-3

17. Krebs E, Zang X, Enns B, Min JE, Behrends CN, Del Rio C, et al. Ending the HIV Epidemic Among Persons Who Inject Drugs: A Cost-Effectiveness Analysis in Six US Cities. The Journal of Infectious Diseases. 2020;222: S301–S311. doi:10.1093/infdis/jiaa130

18. Zang X, Krebs E, Mah C, Min JE, Marshall BDL, Feaster DJ, et al. Can the 'Ending the HIV Epidemic' initiative transition the USA towards HIV/AIDS epidemic control? AIDS. 2020;34: 2325–2328. doi:10.1097/QAD.0000000000002668

19. Nosyk B, Krebs E, Zang X, Piske M, Enns B, Min JE, et al. "Ending the Epidemic" Will Not Happen Without Addressing Racial/Ethnic Disparities in the United States Human Immunodeficiency Virus Epidemic. Clinical Infectious Diseases. 2020;71: 2968–2971. doi:10.1093/cid/ciaa566

20. Quan AML, Mah C, Krebs E, Zang X, Chen S, Althoff K, et al. Improving health equity and ending the HIV epidemic in the USA: a distributional cost-effectiveness analysis in six cities. The Lancet HIV. 2021; S2352301821001478. doi:10.1016/S2352-3018(21)00147-8

21. Khatami SN, Gopalappa C, Mechanical and Industrial Engineering Department, University of Massachusetts Amherst, Amherst, MA 01003, USA. A reinforcement learning model to inform optimal decision paths for HIV elimination. MBE. 2021;18: 7666–7684. doi:10.3934/mbe.2021380

22. Khurana N, Yaylali E, Farnham PG, Hicks KA, Allaire BT, Jacobson E, et al. Impact of Improved HIV Care and Treatment on PrEP Effectiveness in the United States, 2016–2020. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2018;78: 399–405. doi:10.1097/QAI.0000000000001707

23. Grey JA, Bernstein KT, Sullivan PS, Purcell DW, Chesson HW, Gift TL, et al. Estimating the Population Sizes of Men Who Have Sex With Men in US States and Counties Using

Data From the American Community Survey. JMIR Public Health Surveill. 2016;2: e14. doi:10.2196/publichealth.5365

24. the SPREAD Programme, Paraskevis D, Pybus O, Magiorkinis G, Hatzakis A, Wensing AM, et al. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. Retrovirology. 2009;6: 49. doi:10.1186/1742-4690-6-49

25. Rothenberg R, Muth SQ, Malone S, Potterat JJ, Woodhouse DE. Social and Geographic Distance in HIV Risk. Sexually Transmitted Diseases. 2005;32: 506–512. doi:10.1097/01.olq.0000161191.12026.ca

26. Kerani RP, Handcock MS, Handsfield HH, Holmes KK. Comparative Geographic Concentrations of 4 Sexually Transmitted Infections. Am J Public Health. 2005;95: 324–330. doi:10.2105/AJPH.2003.029413

27. Board AR, Oster AM, Song R, Gant Z, Linley L, Watson M, et al. Geographic Distribution of HIV Transmission Networks in the United States. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2020;85: e32–e40. doi:10.1097/QAI.0000000000002448

28. United States Census Bureau. Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2019. Jun 2020 [cited 4 Nov 2021]. Available: https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/counties/asrh/cc-est2019-alldata.csv

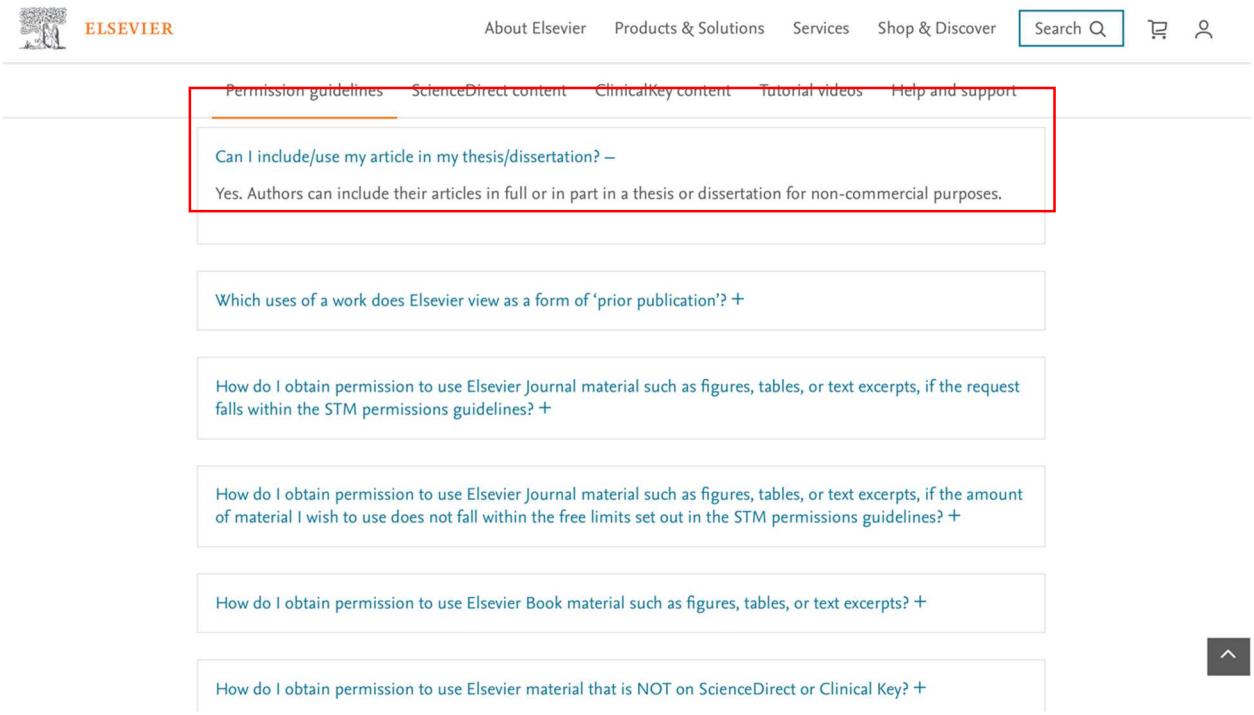
29. NCHHSTP AtlasPlus. [cited 4 Nov 2021]. Available: https://www.cdc.gov/nchhstp/atlas/index.htm

30. Clinical Info HIV.gov. Recommendations for the Use of Antiretroviral Drugs in Pregnant Women with HIV Infection and Interventions to Reduce Perinatal HIV Transmission in the United States. 29 Dec 2020 [cited 8 Nov 2021]. Available: https://clinicalinfo.hiv.gov/en/guidelines/perinatal/prep

Appendix A: General Information About Appendices

Appendices include copyright authorizations for published material and complete published material in the journals of Global Epidemiology, Infectious Disease Modeling, and PLOS ONE, materials submitted to the journal of BMC Medical Research Methodology, and other supporting information for Chapter 6.

Appendix B: Copyrights for Published Materials in Global Epidemiology Journal and Infectious Disease Modelling Journal



Appendix C: Published Materials in Global Epidemiology Journal

Global Epidemiology 2 (2020) 100036

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: https://www.journals.elsevier.com/global-epidemiology

Research Paper

Impact assessment of full and partial stay-at-home orders, face mask usage, and contact tracing: An agent-based simulation study of COVID-19 for an urban region

Hanisha Tatapudi [a,*], Rachita Das [b], Tapas K. Das [a]

[a] Department of Industrial and Management System Engineering, University of South Florida, Tampa, FL, USA
[b] Miller School of Medicine, University of Miami, Miami, FL, USA

ARTICLE INFO

Article history:
Received 2 August 2020
Received in revised form 10 October 2020
Accepted 15 October 2020
Available online 19 October 2020

Keywords:
COVID-19
SARS-CoV-2
Agent-based simulation model
Intervention strategies
Face mask
Contact tracing

ABSTRACT

Purpose: Social intervention strategies to mitigate COVID-19 are examined using an agent-based simulation model. Outbreak in a large urban region, Miami-Dade County, Florida, USA is used as a case study. Results are intended to serve as a planning guide for decision makers.
Methods: The simulation model mimics daily social mixing behavior of the susceptible and infected generating the spread. Data representing demographics of the region, virus epidemiology, and social interventions shapes model behavior. Results include daily values of infected, reported, hospitalized, and dead.
Results: Results show that early implementation of complete stay-at-home order is effective in flattening and reversing the infection growth curve in a short period of time. Whereas, using Florida's Phase II plan alone could result in 75% infected and end of pandemic via herd immunity. Universal use of face masks reduced infected by 20%. A further reduction of 66% was achieved by adding contact tracing with a target of identifying 50% of the asymptomatic and pre-symptomatic.
Conclusions: In the absence of a vaccine, the strict stay-at-home order, though effective in curbing a pandemic outbreak, leaves a large proportion of the population susceptible. Hence, there should be a strong follow up plan of social distancing, use of face mask, contact tracing, testing, and isolation of infected to minimize the chances of large-scale resurgence of the disease. However, as the economic cost of the complete stay-at-home-order is very high, it can perhaps be used only as an emergency first response, and the authorities should be prepared to activate a strong follow up plan as soon as possible. The target level for contact tracing was shown to have a nonlinear impact on the reduction of the percentage of population infected. Increase in contact tracing target from 20% to 30% appeared to provide the largest incremental benefit.

List of abbreviations and acronyms

WHO	World Health Organization
AB	Agent-based
CI	Confidence interval
aOR	Adjusted odds ratio
SEIR	Susceptible exposed infected recovered/removed
GB	Giga bytes
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
COVID-19	Coronavirus Disease 2019.

Introduction

Emergence of the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) was first reported on December 31, 2019 in Wuhan, China and subsequently declared a global pandemic on March 11 by the World Health Organization (WHO) [3,52]. As of Sept. 9, 2020, the number of reported cases worldwide has reached over 27.5 million causing 897,789 deaths. The number of infected cases continues to rise quite significantly [2]. The U.S. has been among the hardest hit by the coronavirus pandemic with 6.3+ million reported infections and 189,538 reported deaths (>21% of the total reported deaths worldwide) so far. However, as the daily new cases, hospital admissions, and deaths began to decline in mid-May, most States in the U.S. began phased lifting of their social intervention measures. For example, Florida adopted a three phased approach: Phase I (which began in May 18, 2020) allowed most businesses and workplaces to reopen with up to 50% of their building capacities and with large events constrained to

* Corresponding author.
E-mail addresses: tatapudi@mail.usf.edu (H. Tatapudi), das@usf.edu (T.K. Das).

https://doi.org/10.1016/j.gloepi.2020.100036

25%; Phase II (began in June 5, 2020) allowed all businesses to reopen for up to 50–75% of their capacities and also permitted events in large venues with no more than 50% of their capacities; Phase III will be akin to a complete reopening for which neither a date nor the criteria have been declared. A summary of Florida's phased intervention plan can be seen in Figure A1 (in the Appendix). As the reopening entered Phase II, Florida, along with many other states, began to see sharp increases in daily new infections (e.g., Florida reported over 15,000 new cases on July 11, 2020 along with a test positivity rate reaching over 15%).

In this paper, we investigate a few 'what-if' scenarios for social intervention policies including if the stay-at-home order were not lifted, if the Phase II order continues unaltered, what impact will the universal face mask usage have on the infections and deaths, and finally, how do the benefits of contact tracing vary with various target levels for identifying asymptomatic and pre-symptomatic. We conduct our investigation by first developing a comprehensive agent-based simulation model for COVID-19, and then using a major urban outbreak region (Miami-Dade County of Florida, USA with 2.8 million population) as the case study for the model.

Methodology

Agent-based (AB) simulation models have been widely used to mimic complex social contact processes and transmission dynamics of influenza and respiratory type viruses. One of the early applications of AB simulation model for influenza can be found in [16]. The authors examined flexible immunization routines and variable vaccine response patterns on 1957 Asian and 1968 Hong Kong pandemic strains of influenza A. In [18,19,24], authors presented AB simulation models with a detailed approach to generate new infections using calculation of force of infection received by susceptible from infected for potential pandemic outbreaks of A(H5N1) influenza virus in Southeast Asia and U.S./U.K. Assessment of disease burden from a potential A(H7N9) outbreak in the U.S. was studied using an AB simulation model in [43]. AB models have been used for examining both pharmaceutical and non-pharmaceutical intervention strategies, for example, mass vaccination for smallpox outbreak [26], effectiveness of targeted antiviral prophylaxis to contain influenza before widespread availability of a vaccine [33], development of mitigation strategies for a potential A(H5N1) outbreak using a design of experiment approach [35], optimal resource allocation among multiple regions of a country during an influenza pandemic [51], and simulation-based reinforcement learning framework for dynamic mitigation of influenza [14].

A number of papers have appeared in recent literature using AB models to examine COVID-19 outbreak. The papers include: an AB model integrated with mobility data to evaluate intervention measures such as testing, social distancing, contact tracing and quarantine for a potential second wave of the COVID-19 pandemic in the Boston metropolitan area [1]; a general framework of an AB simulator for COVID-19 using geo-spatial data to evaluate intervention measures such as closure of community locations, shops, social distancing, face mask, isolation, among others [34]; health and economic impacts of social interventions for COVID-19 [42]; a probabilistic approach using an AB model to simulate COVID-19 transmission and evaluate mitigation strategies in a closed built environment [15]; and a granular AB model of COVID-19 in Australia to compare school closures and varying levels of intervention strategies such as air travel, case isolation, quarantine, and social distancing [9].

COVID-19 in the U.S. has been handled in a very decentralized manner, where mitigation policies were developed and adopted at municipality/county levels. For example, several counties within the State of Florida, U.S.A. adopted the universal facemask policy while others didn't. As a result, spread of SARS-CoV-2 has not been uniform. Hence, we focused on developing a granular model that can be region-specific and can yield outcomes to guide regional decision makers. While we have presented the model development in the main body of the paper using Miami-Dade County (a county of Florida with 2.8 million population) as a case study region, we have also presented a step-by-step approach (in the Appendix) on how the model can be implemented for other regions. The virus transmission model that is embedded within our AB model follows closely the approach used in [18] that modeled a potential A(H5N1) pandemic. Instead of using real-time mobility data, access to which is limited in the U.S. for privacy reasons, we used detailed census data for generating both population and establishments in the region. This information together with people's daily schedules, guided by the prevailing social-behavioral norms, were used to generate social mixing process for the COVID-19 pandemic. The model is first calibrated and validated with evolving daily age-specific reported data on infection, hospitalization, and deaths. We then extended the model into future months to forecast pandemic impact under various intervention conditions.

Models for COVID-19 in literature are either data-driven (e.g., [4,5,11,17,53]) or compartmental like SEIR (susceptible, exposed, infected, and recovered) or their variants (e.g. [1,28,41,54,55]). Observation data driven models are well suited for understanding the past progression of a pandemic and also for estimating parameters characterizing virus epidemiology. However, they offer limited prediction ability for the future, especially in situations where conditions are changing (e.g., testing and treatment ability, social interventions, people's behavior, and response). SEIR type compartmental models guided by differential equations have been most widely used for communicable diseases, some early examples are [13,29,30]. Such models are aggregate in nature and assume uniform behavior of the population over time. Hence, these models also do not adapt well to changing pace of disease transmission. As stated in [10], an agent-based modeling approach is more suitable for a detailed consideration of individual attributes, specific disease natural history, and complex societal interventions that change over time.

The AB simulation model replicates the dynamics of the pandemic outbreak by incorporating: 1) population demography of the outbreak region for all age groups and their employment categories, 2) establishment information concerning numbers, sizes, and compositions of households, schools, workplaces, and community places, 3) daily schedules for people of all age groups before and during the intervention orders (e.g., stay-at-home), 4) isolation of infected and quarantine of household members, 5) closure of schools, workplaces, and community places, 6) compliance to isolation and quarantine requirements, and 7) epidemiological parameters of the virus. The key epidemiological parameters include: disease natural history with average lengths of latent, incubation, symptomatic, and recovery periods; distribution of infectiousness; percent asymptomatic; and fatality rate.

Each day, the AB model tracks the following for each person: 1) hourly movements and locations based on their daily schedules that depend on age, employment status, prevailing intervention orders, and quarantine/isolation status; 2) hourly contacts with other susceptible and infected; 3) force of infection accumulation; 4) start of infection; 5) visit/consult with a doctor (if symptomatic and insured); 6) testing (if infected and visited/consulted a doctor or asymptomatic chosen for testing either randomly or via contact tracing); 7) test reporting delay; 8) disease progression of infected; 9) hospitalization (if infected with acute illness); and 10) recovery or death. The AB model reports daily and cumulative values of actual infected, doctor visits, tested, reported cases, hospitalized, recovered, and deaths, for each age category. A schematic diagram depicting the algorithmic sequence and parameter inputs for the AB simulation model is presented in Fig. 1.

Our AB simulation model works as follows. It begins by generating the individual people according to the U.S. census data that gives population attributes including age (see Table A1) and occupational distribution (see Table A4). Thereafter, it generates the households based on their composition characterized by the number of adults and children (see Table A2). The model also generates, per census data, schools (see Table A3) and the workplaces and other community locations

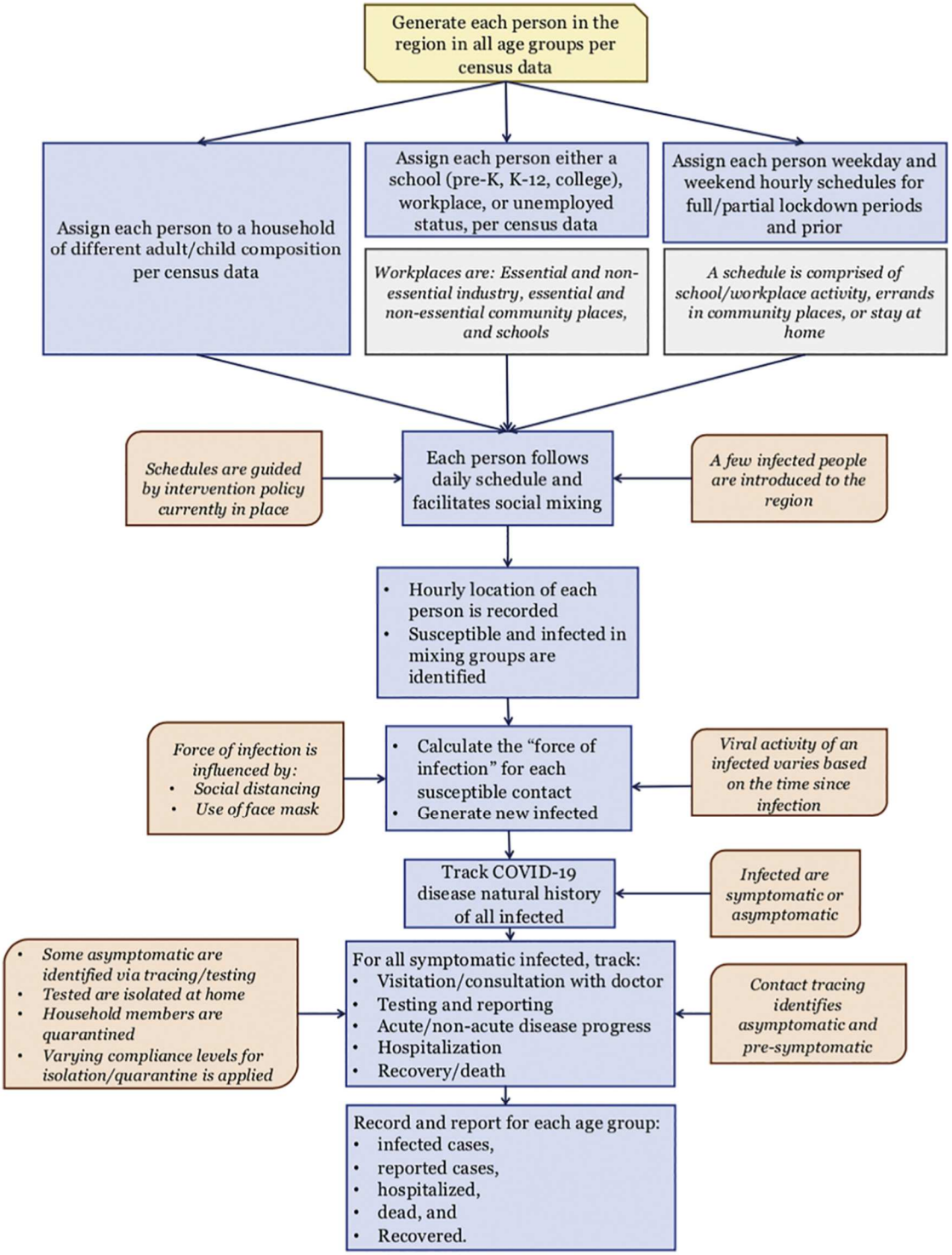


Fig. 1. A schematic of agent-based simulation model for COVID-19.

```
Main
for simulation_day = 0,…,max_day
    if simulation_day = 0
    generate_entities()
        generate_businesses()
        initalize_oubtreak()        // Introduces a small number of infected individuals who are contagious
    begin_outbreak()            // Begin interaction among infected and susceptible and start community spread
    report (infected, tested, reported, hospitalized, death) // Keeps track of daily age specific numbers

Function: Generate entities
// Generates population in the region based on census and assigns them to household and
   workplace/school (including those unemployed who stay at home)
for household = 1,…,total_household_types
    n1 = number of adults
    n2 = number of children
    for i = 1,…,n1
        assign_age_household_workplace()
    for j = 1,…,n2
        assign_age_household_school()

Function: Generate businesses
// Generates businesses in the region based on census data and
// classifies them as essential and non-essential industries and community places
for business = 0,…,total_business_types
    initialize_mixing_groups()          // Creates smaller mixing groups (departments) within each
                                           business and assigns employees

Function: Begin outbreak
// Assigns daily schedules based on prevailing social intervention, tracks hourly social mixing, accounts for
   viral transmission, creates new infections, tests and reports infected, tracks disease progress for infected, monitors
   stay at home/hospitalization, and records recovery and death
if simulation_day < lockdown_day
    schedule = schedule_1       // Pre-pandemic schedule
if (simulation_day >= lockdown_day) and (simulation_day < Phase_I)
    schedule = schedule_2       // Schedule after lockdown begins (closure of non-essential places and schools and
                                   partial opening of essential places with limited capacity)
if (simulation_day >= Phase_I) and (simulation_day < Phase_II)
    schedule = schedule_3       // Schedule during Phase I of reopening (schools remain closed, partial opening of
                                   essential and non-essential with limited capacity)
if (simulation_day >= Phase_II) and (simulation_day < Full_reopening)
    schedule = schedule_4       // Schedule during Phase II of reopening (schools remain closed, increased partial
                                   opening of essential and non-essential places)
for hour = 1,…,24
    disease_progress()          // Monitors disease condition along disease natural history on an hourly basis for
                                   each infected individual
    tracking_individuals()      // Tracks infected and susceptible in the same location (mixing group) by the hour and
                                   calculates added force of infection for susceptible and tracks viral accumulation
    infection_process()         // At the end of each day, infection status of each susceptible with
                                   any viral accumulation is determined
    testing()                   // Tests symptomatic and some selected asymptomatic based on test availability and
                                   reports outcome considering test sensitivity and reporting delay
    hospitalization()           // Tracks hospitalization of symptomatic developing acute condition
    recovery() or death()       // Tracks recovery or death for individuals in home/hospitals
```

Fig. 2. Pseudo-code for agent-based simulation model of COVID-19.

(see Table A4). Each individual is assigned a household, while maintaining the average household composition, and, depending on the age, either a school or a workplace (considering employment levels). A daily (hour by hour) schedule is assigned to every individual, chosen from a set of alternative schedules, based on their attributes. The schedules vary between weekdays and weekends and also depend on the prevailing social intervention orders (see Table A5).

Simulation begins on the day when one or more infected people are introduced to the region (referred to as simulation day 1). Simulation model tracks hourly movements of each individual (susceptible and infected) every day, and records for each susceptible the number of infected contacts and their identification at each location. Based on the level of infectiousness of each infected contact (which depends on the day of his/her infectiousness period), the model calculates the daily force of infection received by each susceptible from all infected contacts at all hours of the day [18]. Daily force of infection is considered to accumulate. However, it is assumed that if a susceptible does not gather any additional force of infection (i.e., does not come in contact with any infected) for two consecutive days, the cumulative force of infection for the susceptible reduces to zero. At the end of each day, the model uses cumulative force of infection to calculate the probability of infection for each susceptible. The model updates the infection status of all individuals to account for new infections as well as disease progression of infected individuals. A pseudo-code in Fig. 2 depicts the major elements and structure of the agent-based simulation program.

Epidemiological models and other parameters that guide the AB model are described next. Fig. 3 presents a schematic of the disease natural history of COVID-19, parameters of which are given in Table A6. Once infected, an individual simultaneously begins the latency and the incubation periods. The individual becomes infectious after the latent period is complete but displays symptoms (unless asymptomatic) at the end of the incubation period. The period between end of latency and end of incubation is referred to as pre-symptomatic, a time when the infectiousness grows rapidly and almost reaches its peak. Symptomatic cases either follow a non-acute progression (majority of cases, not requiring hospitalization) or acute progression (requiring hospitalization). Cases for whom disease does not become acute, enter a recovery period after infectiousness ends. Those with acute disease progression (generally toward the end of the infectious period) are hospitalized. After the hospital stay period, cases either recover or die. For average lengths of recovery and hospitalized periods that are used in the AB model, see Table A6. There is some evidence based on animal experimentation that recovered individuals may become immune to reinfection [31,40]. But other studies remain inconclusive [32]. Hence, due to lack of established data on this matter, our model considers the recovered cases to be immune to further COVID-19 infections.

The duration and intensity of infectiousness is considered to be guided by a lognormal density function (see Fig. 4). The function is truncated at the average length of the infectiousness period (which is considered to be 9.5 days). Asymptomatic cases are assumed to follow a similar infectiousness intensity profile but scaled by a factor (C_k in the force of infection calculation (1), see Table A7).

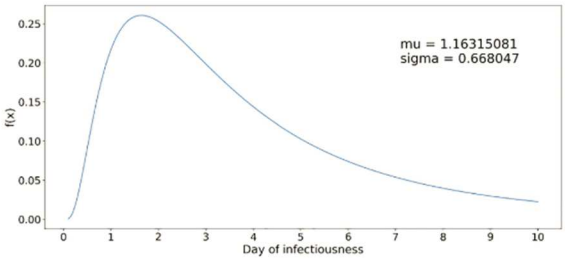


Fig. 4. Lognormal distribution function for infectiousness profile of a COVID-19 case.

The AB model estimates the probability of infection for a susceptible i using the accumulated value of *daily force of infection* (λ_i), which is calculated as follows.

$$\lambda_i = \sum_{k|h_k=h_i} \frac{I_k \beta_h \kappa(t-\tau_k)\rho_k[1+C_k(\omega-1)]}{n_i^{\alpha}} + \sum_{j,k|l_k^j=l_i^j} \frac{I_k \beta_p^j \kappa(t-\tau_k)\rho_k[1+C_k(\omega-1)]}{m_i^j}. \quad (1)$$

The first component in (1) accounts for the force experienced by susceptible individual i at home from other infected household members k. The second component captures the force experienced at schools/workplaces/community places for work and also at community places visited for daily errands; this happens when a susceptible i is in the same location type j where infected k is at hour t. The definition and values of the parameters of (1) are given in Table A7. Eq. (1) is a modified version of the force of infection equation given in [18], which has three components that separately calculate force of infection received at home, at indoor workplaces, and at outdoor community. For the sake of simplicity, we have considered only the first two components, home and indoor workplaces, where most of the COVID-19 transmission is assumed to be taking place. We have assumed that the mode of virus transmission at indoor community places that are routinely visited by people as part of their daily errands (like grocery stores, home goods stores, dine-in/take-out restaurants, etc.) is similar to that of indoor workplace transmission.

The force of infection is gathered by a susceptible individual each day from all infected contacts in his/her mixing groups (home, school/workplace, and community places). The cumulative value of λ_i is used

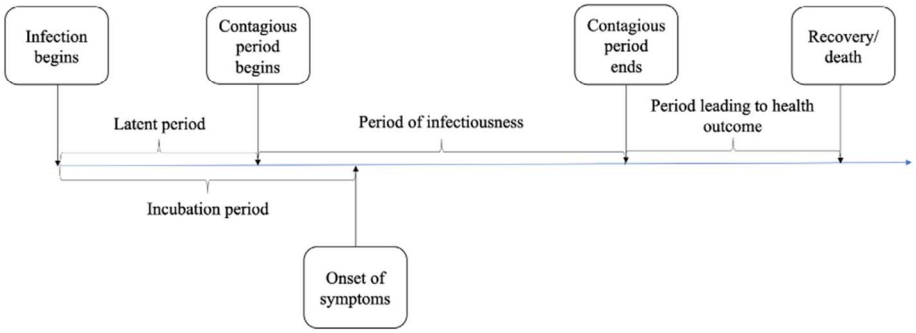


Fig. 3. Disease natural history of COVID-19 (see Table A6 for average lengths of the periods).

at the end of each day to calculate the probability of infection as $1 - exp^{-\lambda_i}$.

The AB model incorporates all applicable intervention orders like stay-at-home, school and workplace closure, isolation of symptomatic cases at home, and quarantine of household members of those who are infected. The model also considers: varying levels of compliances for isolation and quarantine, lower on-site staffing levels of essential work and community places during stay-at-home order, restricted daily schedule of people during various social intervention periods, phased lifting of interventions, use of face masks in workplaces, schools and community places with varying compliance levels, and contact tracing with different target levels to identify asymptomatic and pre-symptomatic cases. The timeline for social interventions implemented in the model are summarized in Table A8.

Other salient considerations in the implementation of our AB model are as follows. Across all age groups, 35% of the infected cases were considered asymptomatic [7]. Approximately, 20% (20%) of Florida residents are reported as uninsured and do not have access to a primary care physician [49]. Uninsured people thus considered not to have the doctor referral required for most of the testing facilities in Miami-Dade County, and hence not tested. All symptomatic cases with health insurance were assumed to visit/consult with a doctor. Depending on their symptoms, travel history, and contact history, some of them were given referrals for testing. We considered that only a small percentage of cases visiting/consulting a doctor were given referrals in the early months of the pandemic (until the middle of April 2020), due to the shortage of testing and restrictive CDC guidelines on who could be tested [6]. However, as CDC relaxed its test eligibility guidelines [39] and the capacity to test increased in Florida, we gradually increased the probability of getting a test referral from a doctor closer to 100% by early June 2020 for symptomatic cases (see Table A9). We also considered in our model that a small fraction (reaching only up to 10% over time) of the asymptomatic cases are randomly tested through various community testing protocols, e.g., at elderly care facilities, healthcare facilities, workplaces, etc. Note that we did not consider co-infection, and therefore all cases that were tested in our simulation model had COVID-19. Hence, each test yielded a positive outcome with a probability equal to the test sensitivity (see Table A9). Based on the data reported on Florida COVID-19 dashboard, a test result reporting delay of up to 10 days on average was considered at the start of the pandemic, which was progressively reduced (see Table A9). All symptomatic cases with or without testing were considered to isolate at home with a given probability of compliance. The probability of compliance was considered to vary during the length of the symptomatic period of infection. For this purpose, we divided the symptomatic period into three parts: I, II, and III, and assumed a lower isolation compliance in parts I and III and higher in part II, when the illness is more apparent. See Table A10 for the isolation compliance probabilities. Susceptible members of the households with one or more infected cases are considered to quarantine themselves. We also assumed a level of compliance for quarantine (see Table A10). We used hospitalization and death data reported for Miami-Dade County [20] for each age group to obtain probabilities of hospitalization of the reported cases, and probabilities of death for those who are hospitalized (see Table A11).

Though we have implemented our AB simulation model for a specific region, it is quite general in its usability for other urban regions with similar demography, societal characteristics, and intervention measures. In our model, Tables A1–A4 summarize the demographic inputs (age and household distribution, number of schools for various age groups, and number of workplaces of various types and sizes). These data will need to be curated from both national and local census records. Social interventions vary from region to region and hence the data in Table A8 will need to be updated. Similarly, testing availability, test sensitivity, and test outcome reporting delay may also vary significantly from region to region, and thus Table A9 will also need to be updated. The rest of the data (in Tables A5, A6, A7, A10, and A11) are related to epidemiology of COVID-19. These are unlikely to be significantly different, though some adjustments of these based on population demographics may be needed.

Model calibration

The AB model utilizes a large number of parameters, which are *demographic parameters*, *epidemiological parameters*, and *social intervention parameters*. We kept almost all of the above parameters fixed at their respective chosen values and calibrated the model by changing values for only a few. The calibrated parameters include the transmission coefficients used in calculating force of infection at home, work, school, and community places (β_h and β_p^j). The choice of the values of transmission coefficients was initially guided by [18] and the prevailing estimates of R0, and thereafter adjusted at different points in time during the calibration period (first 127 days of the simulation starting on

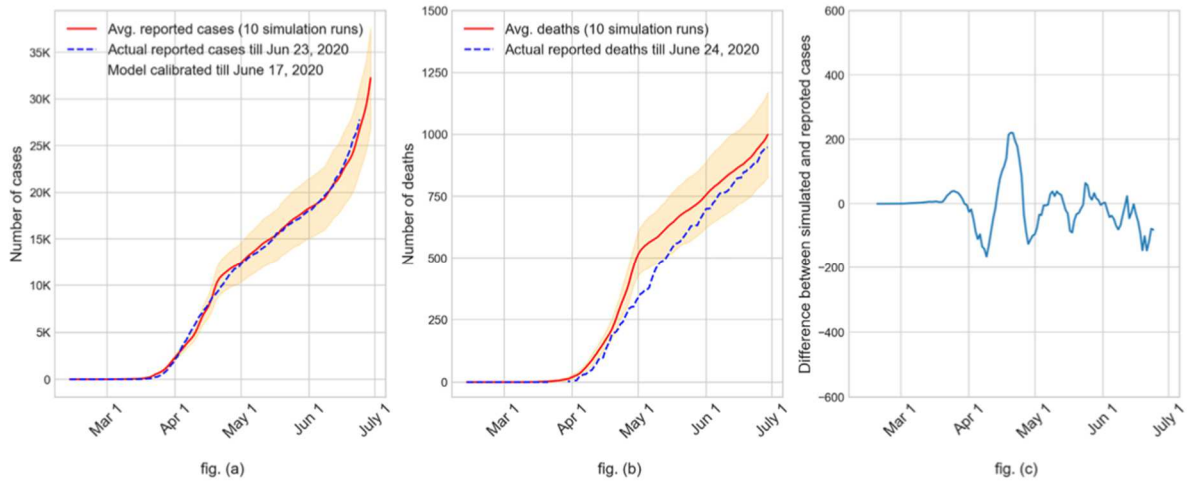


Fig. 5. Validation of AB simulation model results using the State reported numbers of 'cumulative infected cases' in fig. (a) and 'cumulative deaths' in fig. (b) for Miami-Dade County, Florida, USA. Fig. (c) difference of the 7-day moving average between cumulative reported cases from simulation and surveillance (mean difference — 8 and 95% CI (−145 to 130)).

Table 1
Summary of the key results from the AB simulation model implemented on a sample urban outbreak region (Miami-Dade County of Florida, USA) with population of 2.8 million.

Interventions ➡ / Outcomes ⬇	If Stay-At-Home order were not lifted (started March 17, 2020)	If Phase I reopening continued (started May 18, 2020)	If Phase II reopening continues without alterations (started June 5, 2020)	If Phase II reopening continues with mandatory use of face masks (started June 25, 2020)	If Phase II reopening continues with use of face masks and contact tracing with 50% target (starting June 30, 2020)
Time when pandemic subsides below a threshold	Early Aug. 2020	July 2021	End-Oct. 2020	End-Nov. 2020	End-Sept. 2020
Total number of infections (95% C.I.)	162K (136K – 188K)	600K (530K – 670K)	2.17M (2.16M – 2.18M)	1.74M (1.73M – 1.75M)	581K (447K – 716K)
Total number of reported cases (95% C.I.)	23K (19K – 27K)	220K (186K – 254K)	866K (854K – 877K)	714K (702K – 726K)	247K (178K – 316K)
Total number of hospitalizations (95% C.I.)	4.1K* (3.5K – 4.8K)	37.5K* (31.7K – 43.4K)	149K* (147K – 151K)	120K* (119K – 122K)	35.2K* (25.6K – 44.8K)
Total number of deaths (95% C.I.)	1K** (0.9K – 1.2K)	9.4K** (7.9K – 10.8K)	36.4K** (35.8K – 36.9K)	29.7K** (29.3K – 30.2K)	8.8K** (6.5K – 11.1K)

*The numbers presented in the table were computed in mid-late June. Per State reported data, in the months of July and August, the probability of hospitalization reduced significantly across all age groups by over 80%. Hence, our estimates for total number of hospitalizations are much higher than the expected outcome.
**Since the deaths are simulated by applying a probability on those hospitalized, estimates of the number of deaths from our model are much higher than the expected outcome.

February 12 until the start of Phase II reopening order on June 5, 2020). The transmission coefficient for school was only applicable for the first 35 days of the simulation period until the beginning of the stay-at-home order on March 17, 2020. The only other parameters that were calibrated are the number of errands in the daily schedules under various intervention conditions and the percentage of workers in essential (e.g., utility services and grocery stores) and non-essential (e.g., offices and restaurants) workplaces who physically reported to work during different intervention periods. Calibration of the above parameters was done so that the daily cumulative numbers of reported infected cases and deaths from the AB simulation model closely match the values published in the Florida COVID-19 dashboard until June 17, 2020. Fig. 5 (a) and (b) show the daily cumulative mean values (with 95% confidence intervals) for the reported infected cases and deaths as obtained from the simulation model. The dotted lines represent the actual numbers reported in the Florida COVID-19 dashboard for Miami-Dade County [22]. Fig. 5 (c) depicts the difference of the 7-day moving average values of simulated and actual reported number of cases.

Once the model is calibrated and validated with available reported data on infected and dead, we extended the model into the future to predict outcomes. The only parameters that were altered after the calibration period are to reflect the expected changes in social interventions, e.g., order mandating use of face mask, re-closing some community places, expected increase in contact tracing, and changes in community response via daily schedule restrictions. Hence, the parameters that were changed after the calibration period included those for daily schedules, transmission coefficients, testing and contact tracing rates, and compliance to isolation and quarantine. Most of the parameter values used in the AB model were obtained from government archives and research literature, for which references are provided

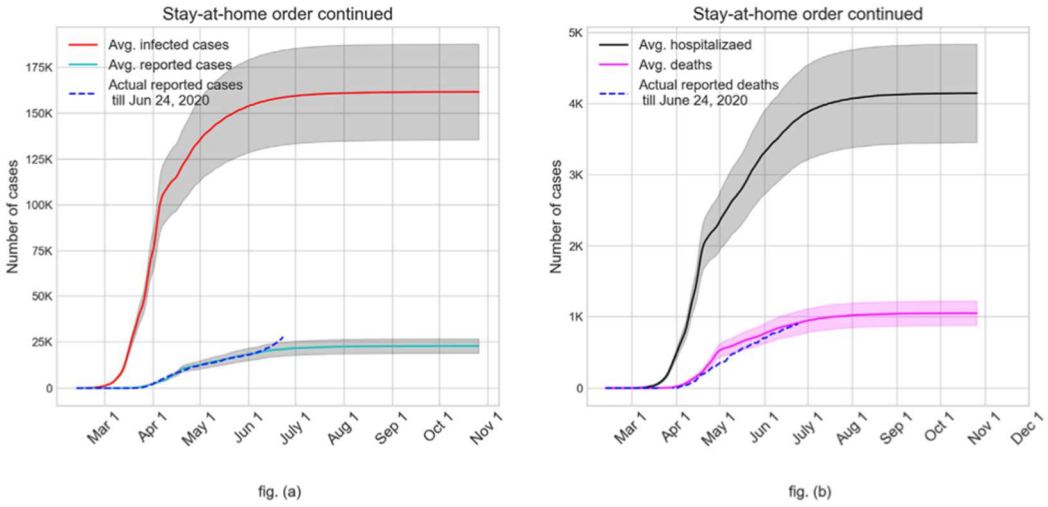


Fig. 6. Growth of actual and reported infected cases (fig. (a)) and hospitalizations and deaths (fig. (b)) if stay-at-home order were not lifted.

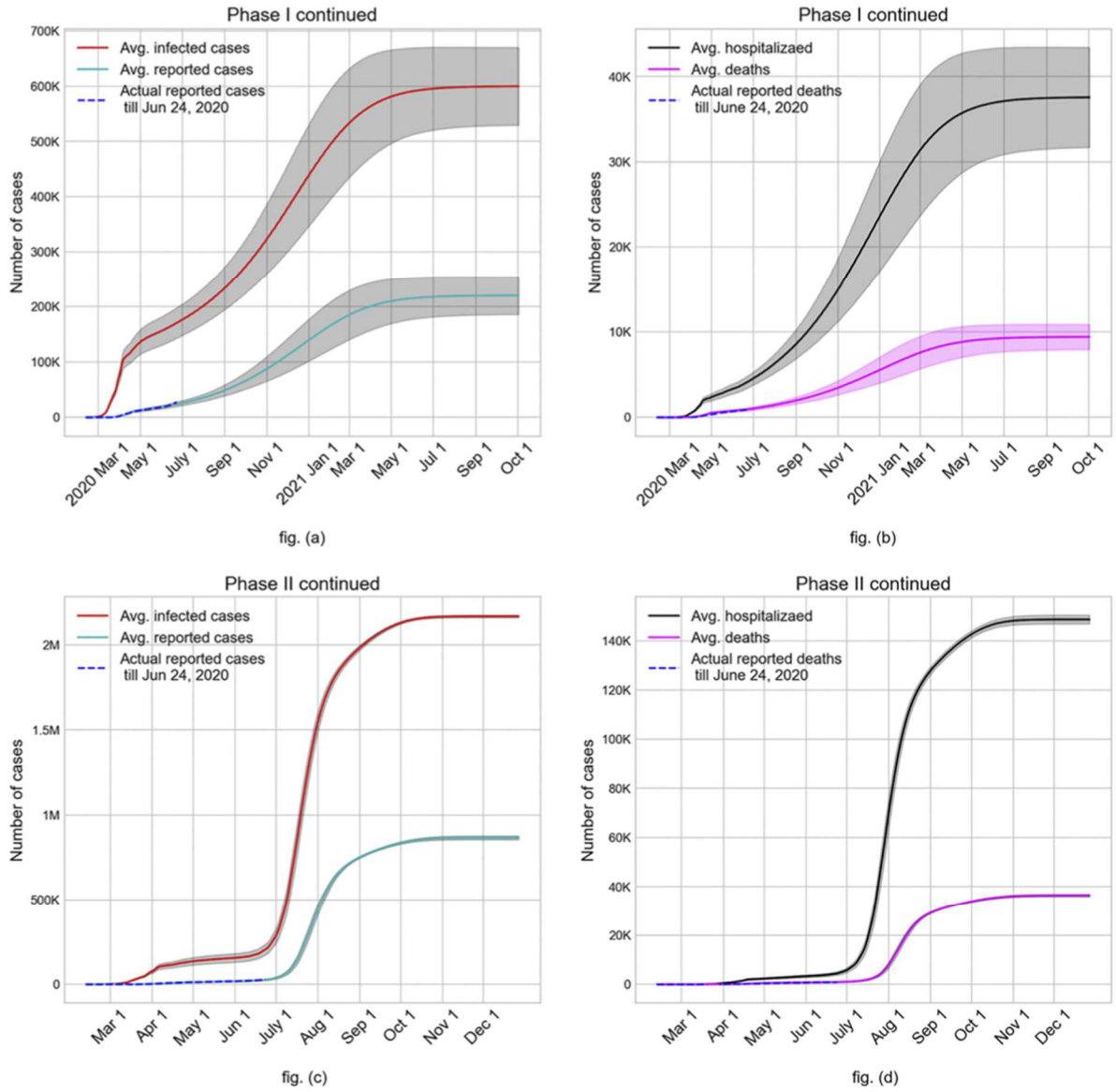


Fig. 7. Case study outcomes (average values with 95% C.I.) of continuing with Phase I reopening (fig. (a) and fig. (b)) and Phase II reopening without face mask and contact tracing (fig. (c) and fig. (d)).

(see Tables A1-A11). For some of the parameters for which we could not find an archived data source, we used expert opinion and current media reports.

Results

We used our model to predict the rate of growth of infected cases, reported cases, hospitalizations, and deaths for the case study region (Miami-Dade County, Florida, USA) under various social intervention scenarios. First, we allowed the model to mimic retrospectively the progress of the pandemic under three separate intervention scenarios for a large number of days. The scenarios are: stay-at-home order continued without reopening until pandemic subsides, Phase I of reopening continued without moving into Phase II, and Phase II of reopening continued without the use of face mask or any other changes. Thereafter, we conducted a prospective examination of the impact we are likely to see in coming days from the use of face masks and contact tracing.

We first present a summary of the key results of our study (see Table 1), from which a number of key insights can be derived that may apply to other similar urban regions experiencing respiratory/influenza type virus outbreaks. Early imposition of stay-at-home order appears to have been quite effective in first flattening and then reversing the growth curve. Per our model, if the stay-at-home order was allowed to remain enforced, the pandemic would have subsided with a relatively low percentage of the population (5.8%) infected and approximately 0.04% dead within six months of inception; 50 or below daily new

infections was used as the criterion to consider that pandemic has subsided in Miami-Dade County. If the extent of social mixing akin to Phase II reopening of Florida is in place for an urban region (without the use of face mask and contact tracing), the pandemic would likely have raged for 8–9 months and subside only after reaching herd immunity with over 75% of the population infected and 1.3% of the population dead. Universal use of face masks of surgical variety was shown by the model to reduce average total infected, hospitalized, and dead by 20%, 19%, and 15%, respectively. Aggressive contact tracing with a goal to identify 50% of the asymptomatic and pre-symptomatic was also projected to have a very significant positive impact with an average reduction of 66% of total infected. The average reductions in total infected with 40%, 30%, and 20% contact tracing targets were found to be 58%, 41%, and 14%, respectively.

In what follows, we expound the results from our study. Figs. 6 and 7 show the simulation results for the retrospective examination scenarios with average values (with 95% CIs in shade) of daily cumulative cases of actual infected, reported, hospitalized, and dead. The blue dotted lines represent the actual numbers of infected and dead as reported in the Florida COVID-19 dashboard till June 24 (our calibration period was till June 17).

Fig. 6 shows a strong influence of continuing with the stay-at-home order in curbing the COVID-19 growth within approximately 6 months from its inception with on average less than 5.8% of the population infected, 0.15% hospitalized, and 0.037% dead; 50 or below daily new infections was used as the criterion to consider that pandemic has subsided in Miami-Dade County. Such a quick suppression of a virus outbreak always leaves the possibility of resurgence, for which an effective plan of contact tracing, testing, isolation, and support for those isolated (when needed) should be in place.

Fig. 7 shows the expected outcomes of continuing with the Phase I order and the Phase II order. Fig. 7(a) demonstrates a clear upward swing of the number of infected by the end of May as a result of Phase I reopening, in contrast to stay-at-home scenario where the numbers actually begin to drop at the end of May. The upward trajectory continues for nearly 12 months after reopening before curving down and subsiding the pandemic in July 2021. This scenario would have resulted in on average approximately 21% of the population infected (see fig. (a)), 1.3% hospitalized, and 0.34% dead (see fig. (b)). Figs. 7(c) and 7 (d) depict the rather grim outcome of continuing with Phase II order without face mask where over 75% of the population gets infected, 5.5% of the population hospitalized, and 1.3% dead. The steep multifold increase in the number of infected in late June after the Phase II reopening in June 5 results in an expected end of the pandemic via herd immunity by late October 2020. We note that at the time of implementing the simulation model (late June 2020) to obtain the above results, the percentage of hospitalized people that died was quite high, especially at the higher age groups. However, by the time of revising the manuscript per reviewer comments and resubmission (early September 2020), the death rate of hospitalized COVID-19 cases has dropped significantly. Hence our model's predictions for the number of deaths are much higher than what is expected.

Hereafter, we used our model in a prospective examination of the pandemic progression under Phase II with the use of face mask and contact tracing. Universal use of face mask in workplaces, schools and community places, where maintaining social distancing is not always feasible, was added to the Phase II guidelines starting June 25, 2020 in Miami-Dade County. In an article that analyzed data from the literature for SARS, MERS, and COVID-19 outbreaks, it is shown that the average value of the adjusted odds ratio (aOR) of getting an infection after wearing surgical variety face masks versus without wearing face mask is 0.33 [12]. This can be interpreted as the likelihood of getting infected if wearing a surgical variety face mask is one third of what it would be for not wearing a mask. Hence, we considered a 67% reduction in the transmission coefficient (β_p^j) used in calculating the force of infection (see eq. (1)), assuming a 100% compliance in the use of surgical variety masks at workplaces, schools and community places. We also tested the impact of 30% and 45% reductions in the transmission coefficient value (β_p^j), which translate to approximately 50% and 70% compliance for face mask usage, respectively. The anticipated impact of face mask usage together with Phase II order on the average cumulative numbers of infected are shown in Fig. 8(a). It also depicts the risk difference between the average values of cumulative infected without and with the universal use of face mask. It may be noted that since the infections grow slower with the use of face mask, the cumulative risk difference rises to almost 875 K in the middle of August and then settles down close to 430 K, when pandemic is predicted to subside by the end of November 2020. Fig. 8(b) depicts the daily values of the average infected for Phase II without and with the universal face mask policy. As expected, the peak of daily infections with face mask usage shifted to a slightly later date and the downward trend began after a smaller percentage (31%) of the total population are infected compared to 36% without the use of face mask.

Though the universal use of face mask together with the Phase II order is likely to reduce a large number of infections (an estimated 430 K), this strategy still leaves a high percentage (63%) of the total population infected before the pandemic is predicted to subside, likely after reaching herd immunity. While a vaccine is still unavailable, perhaps

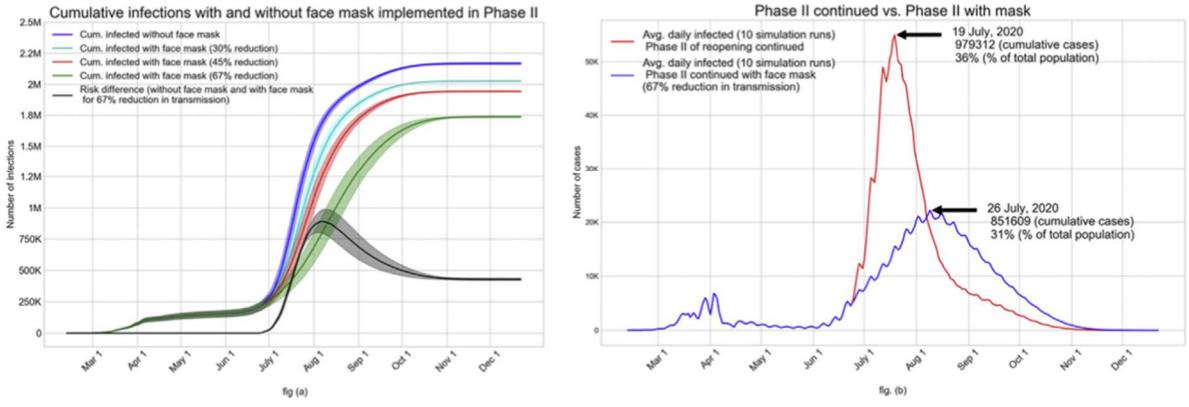


Fig. 8. (a): Impact of face mask usage starting June 25 (together with Phase II order) on the average cumulative infected for all compliance levels; fig. (b): Impact of universal use of face mask on the average daily infected.

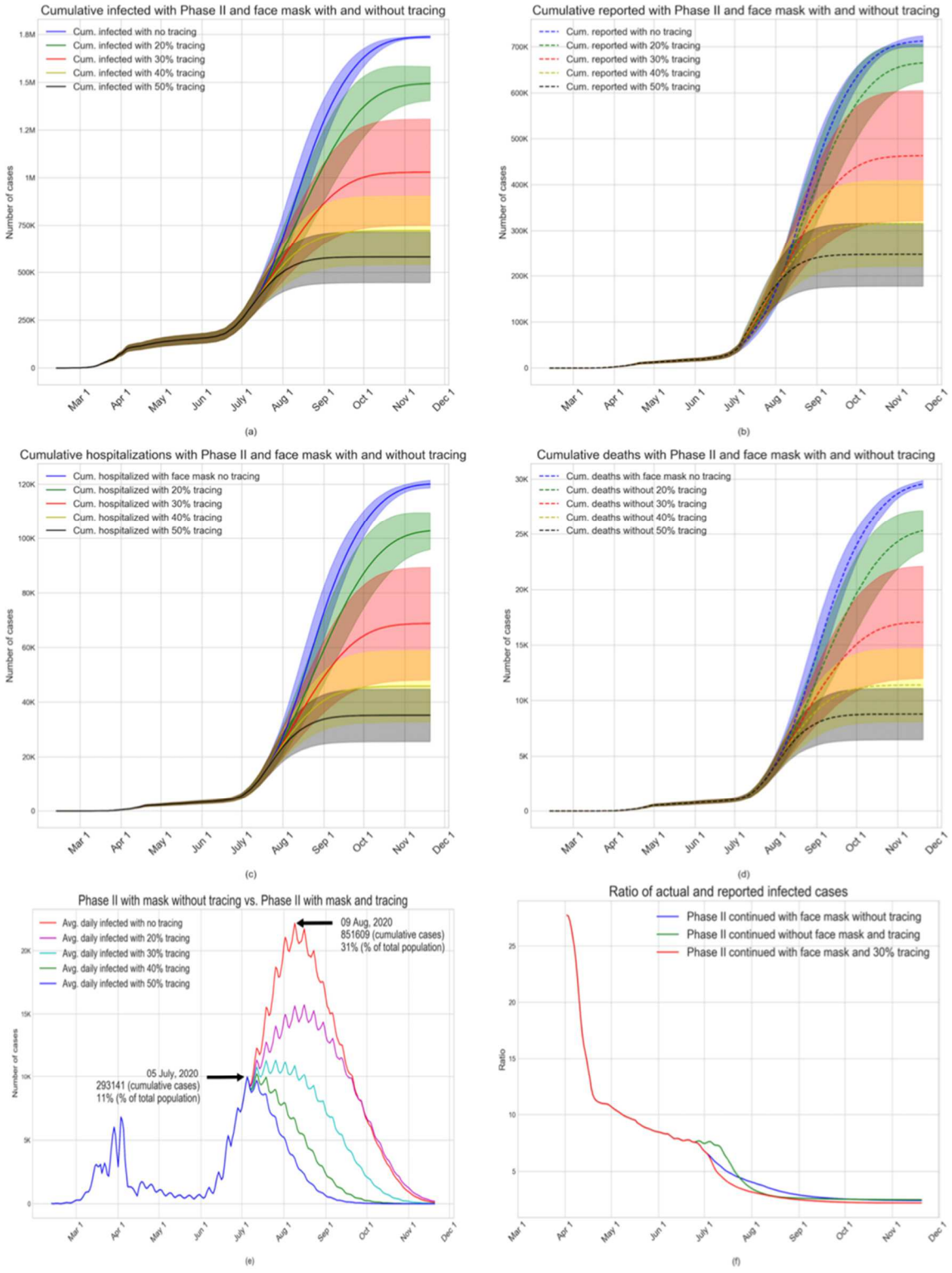


Fig. 9. Impact of contact tracing, starting on June 30, during Phase II with universal face mask usage.

the only other way to reduce the size of this impacted population is through contact tracing. We used our model to examine a number of different contact tracing strategies by adding them to the scenario of Phase II with universal face mask usage. We implemented contact tracing starting June 30 with a number of different targets (20%, 30%, 40%, and 50%) of identifying asymptomatic and pre-symptomatic cases. The impact on the average cumulative values of actual infected and the reported cases are shown in Figs. 9(a) and 9(b). It can be observed that contact tracing can significantly reduce the number of people infected. With the 50% target for contact tracing (which is an aggressive goal), the average cumulative number of infected by the time the daily number of new infections falls below a threshold (possibly by the end of September 2020) would reduce to 581 K from over 1.73 million with Phase II and face mask alone (a 66% reduction). The corresponding reductions in cumulative infections and the associated times for pandemic to subside that can be expected from contact tracing targets of 40%, 30%, and 20% are 58% (mid-October), 41% (mid-November), and 14% (mid-December). It may also be noted that the impact of contact tracing target on the reduction of cumulative infected is nonlinear. Figs. 9(c) and 9(d) show the average cumulative numbers of hospitalizations and deaths. Expected reductions in hospitalization achieved from contact tracing targets of 50%, 40%, 30%, and 20% compared to the use of face mask alone (during Phase II) are 71%, 62%, 43%, and 14%, respectively. The corresponding expected reductions in the number of deaths are 70%, 62%, 43%, and 14%, respectively. We note again that since we completed our simulations runs in late June, the death rates of those hospitalized for all ages have dropped multi-fold during the months of July and August. Thus, our predicted numbers of deaths as depicted in Fig. 9(d) are much higher than what is expected.

Fig. 9(e) shows the impact of contact tracing starting on June 30, 2020 on the average daily infected values. It is interesting to note from the figure that an aggressive contact tracing/testing and isolation of those found infected appears to be capable of quickly turning the tide on new infections. Various COVID-19 dashboards maintained by government and private agencies have been reporting data including numbers of infected (tested positive), hospitalized, and dead. But the actual numbers of infected people in the outbreak regions remain a subject of expert opinion. Speculations abound place the ratio of actual to reported numbers of infected to as high as 10. As our simulation model yields estimates of the actual number of infected, we calculated the daily values of the ratio of average actual infected to average reported for a few scenarios: Phase II continued, Phase II with universal face mask usage, and Phase II with universal face mask usage and contact tracing with a 30% target. Values of these ratios are shown in Fig. 9(f). It can be seen that in the initial days of the pandemic, the ratios are very high (close to 30), which we believe is due to under testing together with long reporting delay. However, as the testing of the symptomatic increased and reporting delay decreased over time, the ratios came down sharply to 10 and continued to fall to near 7. The ratios are expected to further decrease gradually to about 2.5 as the predicted values of daily new infections begin to fall starting late July and early August.

Discussion

We have presented in this paper an agent-based simulation model for COVID-19 pandemic to serve as a policy evaluation tool for public health decision makers. We have implemented the model on one of the epicenters of COVID-19 outbreak in the U.S. (Miami-Dade County of Florida, an urban metropolitan region with 2.8 million population). The model implementation demonstrates the efficacy of our model in both retrospective and prospective assessment of a number of social intervention strategies via their impact on the numbers of infected, hospitalized, and deaths. We have analyzed sensitivity of a number of intervention parameters (partial closures, compliance of face mask usage, and contact tracing targets) in order to support the task of decision making by our public health policy makers.

Our simulation model is written in C/C++ and implemented using GNU General Public License [25].

Our model offers the flexibility to implement a variety of societal conditions including test availability, test reporting delay, stay-at-home order, partial reopening, selective closures of schools and workplaces when infections reappear, use of face mask with various levels of compliance, contact tracing, vaccinations, and use of antivirals. Only a subset of these conditions has been examined and reported in this paper. At the time of revising the manuscript for resubmission for publication, in order to retool the model to reflect changing conditions, we further calibrated our model until July 15, 2020. The recalibrated model was validated by comparing age-specific values of average number of cases reported, hospitalized, and dead for months March to September with the data published in the Florida COVID-19 dashboard for Miami-

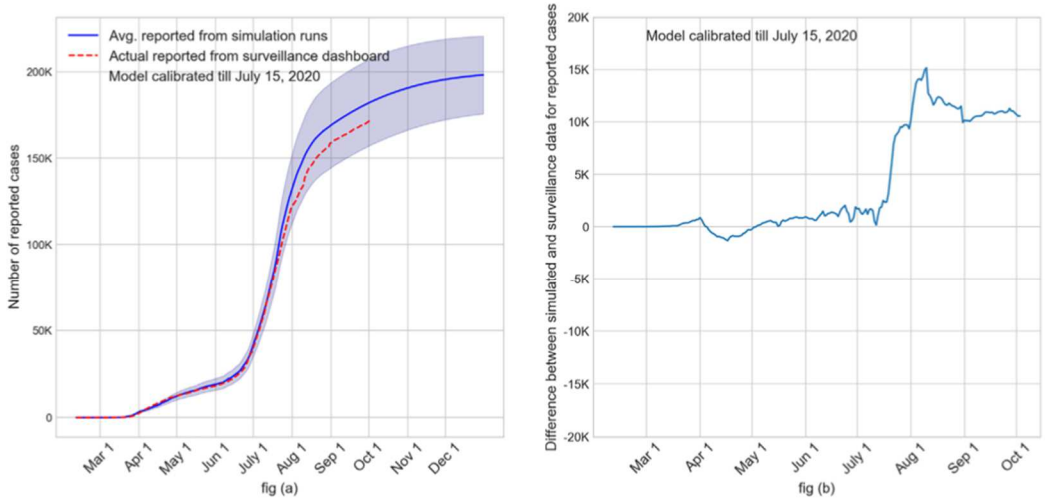


Fig. 10. Validation of calibrated simulation model using state reported surveillance numbers for cumulative reported cases. Fig. (a): Cumulative plot of the average reported cases from simulation (with 95% CI) along with surveillance data. Fig (b): Plot of the difference between cumulative values of average reported cases from simulation and surveillance.

Dade County until September 30, 2020 [22]. See Fig. A2 in Appendix for monthly comparisons. Fig. 10 (a) shows the cumulative values (with 95% confidence interval) of the average reported cases as obtained from the recalibrated simulation model. The dotted line represents the actual numbers of reported cases from the Florida COVID-19 dashboard. Fig. 10 (b) shows the error between simulated and surveillance values for the cumulative number of reported cases.

Our agent-based model has several limitations. First and foremost, the simulation model is an abstraction of how a pandemic impacts a large and complex society. Though our model deliberately introduces some variabilities, somewhat pre-defined daily schedules are used to approximate a highly dynamic contact process of an urban region. Also, the contact process does not account for significant variabilities in the types and lengths of interactions even within each mixing groups. We did not assign geographic locations (latitude and longitude) for households, businesses, schools, and community places, and assumed them to be uniformly distributed over the region. It is common for urban population centers and associated establishments to grow in clusters, for which the contact patterns are expected to be different from those in uniformly dispersed regions. We did not consider special events like parties, games, and street protests, some of which are known to have caused superspreading of the virus and case increases. Finally, and perhaps most importantly, the model uses a large number of parameters (listed in Tables A1 through A11 in the Appendix), and hence the model predictions are influenced by the choice of those values. We have used published data from the government archives and research literature for most of the parameters. In absence of established data source, we have used expert opinion and media reports. Hence, the model results, as presented in this paper, are only expected outcomes based on currently available information.

Each scenario of our case study with 10 replicates (with different seeds) takes approximately 8–12 h to run in a standard desktop computer with Intel Core i7 with 16GB memory. In the interest of presenting our observations quickly to the public health decision makers, while COVID-19 is still rampant in the region, we chose to use a limited number (10) of replicates. As the main purpose of this paper is to conduct a broad what-if analysis, we do not believe that use of a small number of replicates has negatively influenced our observations. The trends and observations derived from our results are only intended to be used for planning and guidance of public health decision makers.

As part of our continuing (future) work, we plan to use our model to examine the impact of reopening of K-12 schools and colleges/universities for the new academic year, which began at the end of August and early September. We also plan to use our AB model to assess efficacy of various prioritization strategies (based on age, risk, and work groups) for the vaccines that are anticipated to be available in limited quantities by the beginning of 2021.

Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Step-by-step approach for implementing the AB simulation model for a pandemic outbreak region.

In what follows, we present a detailed outline of a step-by-step approach for implementing our AB simulation model for a pandemic outbreak in a region caused by influenza or respiratory viruses, as in COVID-19 that is affecting the world. The main body of the paper uses Miami Dade County of Florida, USA as the case study region. Here we use another region, Boston city/Suffolk County, as an example.

Step 1: Model input data gathering.

The following data must be gathered as the first step for applying our AB simulation model to another region like Boston city/Suffolk County.

1) *Population distribution by age*: U.S. census bureau provides county specific age distribution data. For the chosen region, e.g., Boston city/Suffolk county, this information can be found in [A1].
2) *Household composition of adults and children*: Family compsition with sizes of adults and children may not be directly available for all regions, it can be calculated using other sources, like [A2], which provides Boston's households by type and size. Composition of adult and children in these households can be assumed to be similar to another comparable city/county region in the U.S for which data is available.
3) *Distribution of categories of workplaces*: U.S. Census bureau provides, for various industry types, the number of establishments and the corresponding size of employment[A3]. This national data can be proportioned for the specific region if county data is not avaialble.
4) *Characterization of workplaces (essential or non-essential)*: Essential and non-essential workplace characterization can be gathered from state level data, if available. In most regions within the U.S., this characterization can be assumed to be similar. A list of essential and non-essential characterization for the state of Massachusetts can be found in [A4].
5) *Employment levels at various workplace categories, and unemployment*: This data is provided by U.S. Census for some counties and states. Refer to [A3] for data for Boston. When this data is unavailable, national estimates can be used in proportion to the population of the region.
6) *Percentage uninsured*: U.S. Census data's quick facts provides this information [A5].
7) *Distribution of schools (K-12, colleges and universities)*: Total number of schools and staffing levels are categorized as public and private. For information on public schools, see [A7], and for private schools, see [A8].
8) *Distribution of school attendance (K-12, colleges and universities)*: Enrollment levels in public and private schools for K-12 and universities can be found in [A6].
9) *Social intervention policies*: Multi-phased policies are commonly used in most regions in the U.S. However, the dates and the nature of these intervention policies vary widely from region to region. For example, Alaska, Arizona and Georgia, among others, didn't have a mandatory face mask policy, while other states did. Some states did not have a statewide policy, but each municipality adopted their own.
   a) *Lockdown policy*: Time-varying policy applicable for Boston city/Suffolk County can be found in [A11].
   b) *Policies for phased lifting of social interventions*: Time-varying policy applicable for Boston city/Suffolk County can be found in [A11].
   c) *Use of face mask*: Boston implemented a facemask policy in early May in an executive order by the State [A9].
   d) *Implementation of contact tracing*: Massachusetts state government provides a dashboard on the community health outcomes for COVID-19. Details on the success of contact tracing in the communities along with the outcome measures varying over time can be found in [A10].
   e) *Policy for returning to school*: School reopening policies also widely vary from state to state and also among counties within a state. Information on Boston's public school reopening policy can be found in [A12]. It is important to frequently check sources on school policy as they are transient. For example, Boston planned to reopen on October 15, but shifted to October 22 after seeing an increase in the number of cases.

10) *Time varying testing of symptomatic and asymptomatic*: Limited testing availability has been a serious concern in many U.S. regions that suffered from a high level of disease spread. Time varying data on test availability and test outcome reporting delay are difficult to find in indexed literature during a pandemic. Hence, these can be assumed from regional news reports, test reporting data, and/or other grey literature.
11) *Number of reported, hospitalized and deaths*: Daily data and archived data on the number of people reported positive, hospitalized and dead can be found from dashboard in [A10].
12) *Probability of hospitalization for reported cases and probability of death for hospitalized*: This can be calculated from [A10] based on age specific reporting.

The information contained in the following tables for Miami Dade County are likely to be same for other regions like Boston city/Suffolk County.

13) Daily schedules for people can be assumed to be same as in Table A5.
14) Disease natural history parameters for COVID-19 can be assumed to be same across regions within a country, see Table A6.
15) Some of the parameters in Table A7 for calculating the force of infection need to be calibrated (see Step 3). However, the remaining parameters in Table A7 can be assumed to be same.
16) Self-isolation compliance for symptomatic cases, and quarantine compliance for household members can be assumed to be same for different urban regions within a country, as in Table A10.

Step 2: Updating the simulation model.

Once the input data collection is complete, the next step is to update the model parameters as follows.

1) *Update the simulation model with all gathered input data from step 1*: After gathering data, it needs to be curated and transformed into .txt files to be read by the simulation model. Some of the data are directly coded in the model, where applicable.
2) *Decide simulation begin date*: Simulation begin date depends on the outbreak region and is based on the date of the first reported case. Up to 14 days before the first reported case can be used as a potential date for simulation model to begin.
3) *Decide simulation end date*: Simulation end date is chosen as desired by the modeler.
4) Number of initial infected cases: Most Departments of Health (DOH) provide a count and characterization of the number of initial infected cases with travel histories. One can identify these initial infected cases during the first month of the outbreak and use those cases to initiate social mixing and community spread.

Step 3: Calibrate and validate simulation model.

Once the simulation model is updated with the input data for the region, the model is calibrated using a small applicable subset of input data and the model output is validated with actual surveillance data from the region, as follows.

1) Generate multiple seeds for the uniform random variables that are used to calculate the probabilities of infection, hospitalization, death, testing, symptomatic, disease severity, test sensitivity, compliance for isolation and quarantine, among others. Simulation output from each seed is considered a replicate. Using output data from all replicates, an average value and a corresponding confidence interval for each output measure are calculated.
2) A set of initial values of the transmission coefficients for home, school, work, and community places are assumed (based on current literature and published models for outbreaks of similar diseases). These transmission coefficients (along with other parameters, see Table A7) are used for calculating force of infection, which is then used to calculate the probability of infection. Different sets of transmission coefficient values are selected for different reference points in time in the simulation, depending on changes in social intervention status and significant current events. For example, the transmission coefficients are appropriately calibrated (reduced) on the day universal use of facemask is announced. Also, percentage testing of asymptomatic and pre-symptomatic are increased when contact tracing begins. Street protests combined with Independence Day holiday in early July 2020 are examples of current events that may require adjustment of transmission coefficient values.
3) Other parameters that are considered suitable for the AB model calibration are probability of running errand that guides daily schedule and probability of employees reporting to work for essential and non-essential businesses. These values can also be assumed to change over time during a pandemic depending on the phased intervention policies implemented by the government in the outbreak region.
4) The simulation is calibrated for a chosen period. In this study, the model was calibrated up to July 15th, as reported data was available until that date for validation purposes at the time of model calibration.
5) Results for reported cases, hospitalized, and dead for all age groups are gathered from the simulation model for each seed.
6) Average values (with confidence interval) are computed for the numbers of reported, hospitalized and dead.
7) For model validation, the simulated average values for the reported, hospitalized, and dead and compared with actual surveillance data.
8) Alter calibration parameters as needed to obtain desired level of validation accuracy. Measure validation accuracy is calculated as the difference in the seven-day moving average between simulated and surveillance data.

Step 4: Implement calibrated model for prediction

1) Run calibrated simulation model for all seeds for a desired prediction period beyond the calibration/validation time.
2) Extract age specific data for total infected, reported cases, hospitalized, and dead from simulation for each seed.
3) Report mean and confidence interval.

References for step by step approach

[A1] United Census Bureau. ACS Demographic And Housing Estimates: 2019, https://data.census.gov/cedsci/table?q=Boston%20city,%20Suffolk%20County,%20Massachusetts&tid = ACSDP1Y2019.DP05&hidePreview = false; [accessed 10.06.2020].
[A2] The Boston Planning and Development Agency. Boston by the numbers 2018, http://www.bostonplans.org/getattachment/3e8bfacf-27c1-4b55-adee-29c5d79f4a38; [accessed 10.06.2020].
[A3] United Census Bureau. Annual Business Survey: Statistics for Employer Firms by Industry, Sex, Ethnicity, Race, and Veteran Status for the U.S., States, Metro Areas, Counties, and Places: 2017, https://data.census.gov/cedsci/table?q=employment%20size%20boston%20&t=Employment&tid=ABSCS2017.AB1700CSA01&hidePreview=false; [accessed 10.06.2020].
[A4] Mass.gov. COVID-19 Essential Services, https://www.mass.gov/info-details/covid-19-essential-services; [accessed 10.06.2020].
[A5] United Census Bureau. QuickFacts Boston City, Massachusetts, https://www.census.gov/quickfacts/bostoncitymassachusetts#qf-flag-X; [accessed 10.06.2020].
[A6] United Census Bureau. School Enrollment, https://data.census.gov/cedsci/table?q=Boston%20city,%20Suffolk%20County,%20Massachusetts&t = Education&tid = ACSST1Y2019.S1401&hidePreview = false; [accessed 10.06.2020].
[A7] BPS Communications Office. Boston Public Schools At A Glance 2019–2020, https://www.bostonpublicschools.org/cms/lib/MA01906464/Centricity/Domain/187/BPS%20at%20a%20Glance%202019-20_FINAL.pdf; [accessed 10.06.2020].

[A8] Private School Review. Largest Massachusetts Private Schools, https://www.privateschoolreview.com/school-size-stats/massachusetts; [accessed 10.06.2020].
[A9] AARP.org. State-by-State Guide to Face Mask Requirements, https://www.aarp.org/health/healthy-living/info-2020/states-mask-mandates-coronavirus.html; [accessed 10.06.2020].
[A10] Mass.gov. COVID-19 Response Reporting, https://www.mass.gov/info-details/covid-19-response-reporting; [accessed 10.06.2020].
[A11] Mass.gov. COVID-19 State of Emergency, https://www.mass.gov/info-details/covid-19-state-of-emergency; [accessed 10.06.2020].
[A12] Boston Public Schools. Important Information About School Reopening, https://www.bostonpublicschools.org/reopening2020; [accessed 10.06.2020].

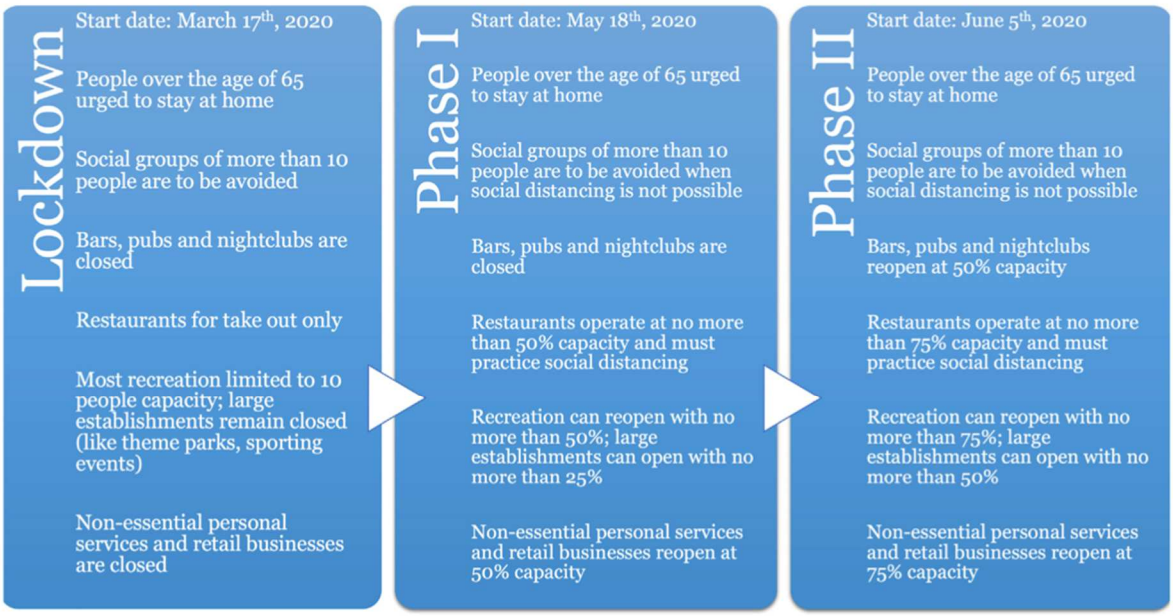


Fig. A1. Florida's phased social intervention plan for COVID-19 pandemic [44].

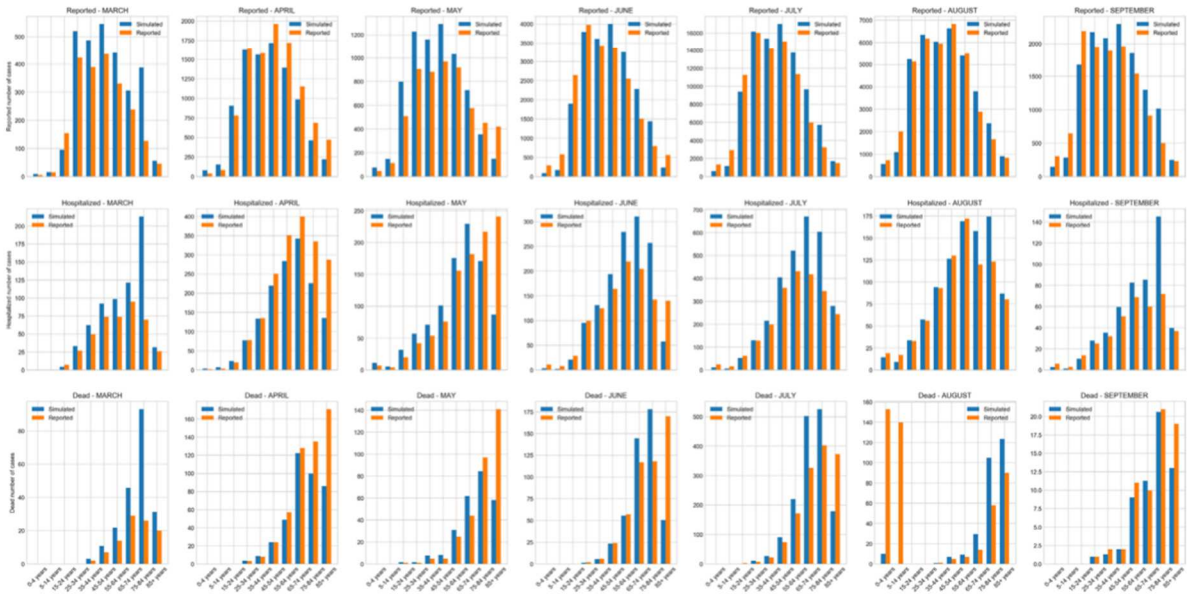


Fig. A2. Validation of simulation model by comparing monthly average values of reported cases, hospitalized, and deaths (in orange) with the corresponding reported values from Florida COVID-19 Dashboard (in blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table A1
Age distribution of Miami-Dade population [47].

Age range	Percentage
0–5	6.9
6–9	4.2
10–14	5.8
15–17	3.5
18–22	6.1
23–29	9.5
30–64	47.8
65–99	16.2

Table A2
Household composition of adults and children in Miami-Dade County of Florida, USA [37].

Household type	Percentage
1 adult 0 children	23
1 adult 1 children	1
2 adults 0 children	26
1 adult 2 children	3
2 adults 1 children	6
3 adults 0 children	10
1 adult 3 children	1
2 adults 2 children	6
4 adults 0 children	8
1 adult 4 children	1
2 adults 3 children	8
3 adults 3 children	4
4 adults 3 children	3

Table A3
Distribution of educational institutions in Miami-Dade County; all children ages 6–22 are assumed to attend school; among all children in ages 0–5, only 50,540 are assumed to attend school; the remaining children stay at home [36,50].

Age range	School type	Percentage	Number of schools	Number of classes in each school	Average number of students in each class	Average number of students in attending each school
0–5	*Pre-K*	26	190	14	19	266
6–9	*Elementary*	16	323	20	19	380
10–14	*Junior*	22	147	44	25	1100
15–17	*Senior*	13	232	22	19	418
18–22	*College/University*	23	9	375	50	18,750

Table A4
Distribution of different types of workplaces in Miami-Dade County. All industries and community places are classified into essential or non-essential. Essential industries remain functional with a percentage of their workforce reporting during interventions like stay-at-home or phased reopening. Non-essential industries are considered to operate remotely. Essential industries include wholesale trade, waste management and remediation services, agriculture, forestry, fishing and hunting, mining, quarrying, oil and gas extraction, utilities, construction, manufacturing, transportation and warehousing. Non-essential industry includes finance and insurance, real estate and rental and leasing, professional, scientific and technical services, management of companies and enterprises, administrative and support for waste management and remediation services, educational services, other services except public administration. Essential community includes grocery stores, convenience stores, pharmacies and drug stores, home centers, health care and social assistance. Non-essential community includes retail, arts, entertainment and recreation and accommodation and food services. For details on education institutions, see Table A3. The total employed population in Miami-Dade County is 1,171,404 [45,46,48].

Category of workplaces	Number of establishments	Average number of people working in each	Number of mixing groups in each establishment	Average number of people in each mixing group
Essential industry	19,378	6	1	6
Essential industry	356	100	2	50
Essential industry	205	200	4	50
Essential industry	81	400	8	50
Essential industry	12	600	12	50
Essential industry	6	1000	20	50
Non-essential industry	37,002	6	1	6
Non-essential industry	546	100	2	50
Non-essential industry	346	200	4	50
Non-essential industry	100	400	8	50
Non-essential industry	26	600	12	50
Non-essential industry	8	1000	20	50
Essential community	10,709	6	1	6
Essential community	237	100	2	50

(continued on next page)

Table A4 (*continued*)

Category of workplaces	Number of establishments	Average number of people working in each	Number of mixing groups in each establishment	Average number of people in each mixing group
Essential community	236	200	4	50
Essential community	41	400	8	50
Essential community	17	600	12	50
Essential community	19	1000	20	50
Non-essential community	14,571	6	1	6
Non-essential community	601	100	2	50
Non-essential community	296	200	4	50
Non-essential community	87	400	8	50
Non-essential community	32	600	12	50
Educational institution	190	42	14	3
Educational institution	323	60	20	3
Educational institution	147	176	44	4
Educational institution	232	66	22	3
Educational institution	9	1500	375	4

Table A5
Daily schedules for adults and children.

	Employed Adults	Unemployed Adults	Children
Weekday (regular schedule before social intervention)	• At work from 8 AM – 4 PM • One errand (for one hour) between 5 and 7 PM • Home from 7 PM – 8 AM	• Runs 3 errands during three randomly selected hours between 8 AM and 7 PM • Home from 7 PM – 8 AM	• At school from 8 AM – 3 PM • Home 4 PM – 8 AM
Weekend (regular schedule before social intervention)	• Runs 3 errands during three randomly selected hours between 8 AM and 7 PM	• Runs 3 errands during three randomly selected hours between 8 AM and 7 PM	• Engages in 3 activities from 8 AM - 7 PM
If isolated or Quarantined or under stay-at-home order	• Runs one errands during a randomly selected hour between 8 AM and 7 PM with probability 0.2 (and applicable compliance probability)	• Runs one errands during a randomly selected hour between 8 AM and 7 PM with probability 0.2 (and applicable compliance probability)	Stay at home
If isolated or Quarantined or otherwise under Phase I or Phase II order	• Runs 2 errands during two randomly selected hours between 8 AM and 7 PM with probability 0.4 in Phase I and 0.8 in Phase II (and applicable compliance probabilities for isolated and quarantined)	• Runs 2 errands during two randomly selected hours between 8 AM and 7 PM with probability 0.4 in Phase I and 0.8 in Phase II (and applicable compliance probabilities for isolated and quarantined)	• Stay at home

Table A6
Disease natural history parameters of COVID-19.

Disease natural history parameters	Average value
Latent period	3 days [27]
Incubation period	5.5 days [38]
Contagious period	9.5 days [27]
Asymptomatic percentage	35% [7]
Health outcome parameters	**Average value**
Percent of Florida residents without health insurance	20% [49]
Time to visit doctor for symptomatic	1–2 days after onset of symptoms
Symptomatic (who visits/consults doctor) not hospitalized assumed to be recovered for normal activity	14 days from after doctor visit
Symptomatic who do not visit/consult doctor assumed to resume normal activity	8.5 days after onset of symptoms
Time to hospitalization	5–9 days after onset of symptoms [23]
Hospitalized period / time to recovery/death for hospitalized	9–12 days after hospitalization [8]

Table A7
Parameters of the force of infection (eq. (1)).

Parameter	Description	Value			
I_k	Infected status of an individual k	1 if infected and 0 otherwise			
β_h	Transmission coefficient at home[a]		**Period**		
		0.7	Before stay-at-home orders		
		0.015	During stay-at-home orders		
		0.5	During Phase I		
		0.5	During Phase II		
β_p^l	Transmission coefficient at school, work, and community places[a]	**School**	**Work**	**Community**	**Period**

Table A7 (*continued*)

Parameter	Description	Value			
		0.7	0.8	0.18	Before stay-at-home order
		0.0	0.025	0.0065	During stay-at-home order
		0.0	0.5	0.057	During Phase I
		0.0	0.5	0.22	During Phase II
$\kappa(t)$	Infectiousness at time t (t denotes the elapsed time after completion of latency)[b]	Lognormal function value at t with $\mu = 1.16315081$ and $\sigma = 0.668047$			
ρ_k	Relative infectiousness of individual k[c]	1			
C_k	Scaling factor for mild/asymptomatic vs severe infection[c]	1 if severe/symptomatic, 0 for mild/asymptomatic			
ω	Scaling factor for infectiousness for a mild vs severe infection[c]	2 for severe infection relative to a mild one			
α	Scaling factor for household size[c]	0.8			
n_i	Number of people in the household of individual i	Calculated from simulation			
m_i^j	Number of people in the place type j where individual i is	Calculated from simulation			

[a] Choice of these parameters were guided by Ferguson [18], literature estimates of R0 for SARS-CoV-2, and prevailing interventions. Transmission coefficients were subsequently calibrated to arrive at the values given here.
[b] Parameters of the lognormal distribution function were selected to have the mean length of infection as 4 days and a standard deviation of 3 days [27].
[c] Selected from Ferguson [18].

Table A8
Social intervention order timeline for Miami-Dade County [44].

Intervention policy implemented at Miami-Dade County, Florida, U.S.	Date of implementation	Day of Simulation
Stay at home policy	March 172,020	35
Phase I reopening	May 182,020	97
Phase II reopening	June 52,020	115
Mandatory usage of Face mask	June 252,020	135
Contact tracing (assumed to begin)	June 302,020	140

Table A9
Time varying data of testing rate, test sensitivity, and test result reporting delay. (values are based on expert opinion, news reports, and Florida COVID-19 Dashboard data [22])

Date range	Proportion of asymptomatic cases randomly tested (before contact tracing)	Proportion of symptomatic cases who are tested*	Test sensitivity	Test result reporting delay
Feb 12, 2020 – March 12, 2020	0.05	0.2	0.8	10
March 13, 2020 – April 11, 2020	0.05	0.2	0.8	7
April 12, 2020 – May 11, 2020	0.06	0.3	0.8	7
May 12, 2020 – June 5, 2020	0.07	0.6	0.8	6
June 6, 2020 – July 10, 2020	0.1	0.85	0.8	2
July 11, 2020 onwards	0.1	0.9	0.9	2*

* Test result reporting delay appeared to have increased again to up to a week or more in Florida, but not implemented in the model.

Table A10
Self-isolation compliance for symptomatic cases and quarantine compliance for household members.

Parameter	Value*
Isolation compliance of adults in Part I of symptomatic period	75%
Isolation compliance of adults in Part II of symptomatic period	95%
Isolation compliance of adults in Part III of symptomatic period	90%
Isolation compliance of children in Part I of symptomatic period	80%
Isolation compliance of children in Part II of symptomatic period	99%
Isolation compliance of children in Part III of symptomatic period	95%
Duration of Part I of symptomatic period	1.5 days
Duration of Part II of symptomatic period	3.5 days
Duration of Part III of symptomatic period	2 days
Quarantine compliance of susceptible in households with infected	75%

* Values are assumed based on expert opinion.

Table A11
Probability of hospitalization for reported cases and probability of death for hospitalized [21].

Age range	Probability of hospitalization among those reported	Probability of death among those hospitalized
0–4	0.08	0.00
5–14	0.03	0.00
15–24	0.04	0.01
25–34	0.06	0.03
35–44	0.10	0.05
45–54	0.14	0.07
55–64	0.20	0.13
65–74	0.34	0.25
75–84	0.46	0.37
85–100	0.49	0.57

References

[1] Aleta A, Martin-Corral D, Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. medRxiv. 2020. https://doi.org/10.1101/2020.05.06.20092841.

[2] Almukhtar S, Aufrichtig A, Bloch M, Calderone J, Collins K, Conlen M, et al. Nytimes.com. Coronavirus Map: Tracking the Global Outbreak; 2020 Nytimes.com. https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html?auth=link-dismiss-google1tap; accessed 07.07.2020.

[3] Andersen K, Rambaut A, Lipkin W, Holmes E, Garry R. The proximal origin of SARS-CoV-2. Nat Med. 2020;26. https://doi.org/10.1038/s41591-020-0820-9.

[4] Barmparis GD, Tsironis GP. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. Chaos Solitons Fractals. 2020;135:109842. https://doi.org/10.1016/j.[s.2020.109842.

[5] Candido D, Claro I, Jesus J, Marciel de Souza W, Moreira F, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. medRxiv. 2020. https://doi.org/10.1101/2020.06.11.20128249.

[6] Centers for Disease Control and Prevention. CDC releases consolidated COVID-19 testing recommendations. https://www.cdc.gov/media/releases/2020/s0613-covid19-testing-recommendations.html;; 2020. accessed 07.08.2020.

[7] Centers for Disease Control and Prevention. COVID-19 Pandemic Planning Scenarios. https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-archive/planning-scenarios-2020-05-20.pdf;; 2020. accessed 06.22.2020.

[8] Centers for Disease Control and Prevention. Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19). https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html;; 2020. accessed 07.08.2020.

[9] Chang SL, Harding N, Zacherson C, Cliff OM, Prokopenko M. Modeliling tranmission and control of the COVID-19 pandemic in Australia; 2020 arXiv.2003.10218.

[10] Chao D, Halloran E, Obenchain V, Longini I. FluTE, a publicly available stochastic influenza epidemic simulation model. PLoS Comput Biol. 2010;6(1):e1000656. https://doi.org/10.1371/journal.pcbi.1000656.

[11] Chintalapudi N, Battineni G, Amentaa F. COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: a data driven model approach. J Microbiol Immunol Infect. 2020;53(3):396–403. https://doi.org/10.1016/j.jmii.2020.04.004.

[12] Chu D, Akl E, Duda S, Solo K, Yaacoub S, Schünemann H, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet. 2020;395(10242):1973–87. https://doi.org/10.1016/S0140-6736(20)31142-9.

[13] Colizza V, Barrat A, Barthelemy M, Valleron A, Vespignani A, et al. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med. 2007;4(1):e13. https://doi.org/10.1371/journal.pmed.0040013.

[14] Das TK, Savachkin A, Zhu Y. A large-scale simulation model of pandemic influenza outbreaks for development of dynamic mitigation strategies. IIE Transactions. 2008;40:893–905. https://doi.org/10.1080/07408170802165856.

[15] D'Orazio M, Bernardini G, Quagliarini E. How to restart? An agent-based simulation model towards the definition of strategies of COVID-19 "second phase" in public buildings. arXiv. 2020arXiv:2004.12927.

[16] Elveback LR, Fox JP, Ackerman E, Langworthy A, Boyd M, Gatewood L. An influenza simulation model for immunization studies. Am J Epidemiol. 1976;103(2):152–65. https://doi.org/10.1093/oxfordjournals.aje.a112213.

[17] Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis. J Med Virol. 2020;92(6):645–59. https://doi.org/10.1002/jmv.25750.

[18] Ferguson NM, Cummings DAT, Fraser C, Cajka J, Cooley P, Burke D. Strategies for mitigating an influenza pandemic. Nature. 2006;442(7101):448–52. https://doi.org/10.1186/s12889-017-4884-5.

[19] Ferguson NM, Cumming DAT, Cauchemez S, Fraser C, Riley S, Meeyai A, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia, 437; 2005; 209–14. https://doi.org/10.1038/nature04017 7056.

[20] Florida Department of Health. Coronavirus: summary of persons being monitored, persons under investigation, and cases. http://ww11.doh.state.fl.us/comm/_partners/covid19_report_archive/state_reports_latest.pdf;. accessed 07.08.2020.

[21] Florida COVID-19 Response. COVID-19: summary of persons being monitored, persons tested, and cases. http://ww11.doh.state.fl.us/comm/_partners/covid19_report_archive/state_reports_latest.pdf;; 2020. accessed 07.15.2020.

[22] Florida Department of Health, Division of Disease Control and Health Protection. Florida's COVID-19 Data and Surveillance Dashboard. https://experience.arcgis.com/experience/96dd742462124fa0b38ddedb9b25e429;. accessed 07.07.2020.

[23] Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 — COVID-NET, 14 states, march 1–30, 2020. MMWR. 2020;69(15):458–64http://dx.doi.org/10.15585/mmwr.mm6915e3.

[24] Germann T, Kadau K, Longini I, Macken C. Mitigation strategies for pandemic influenza in the United States. Proc Natl Acad Sci U S A. 2006;103(15):5935–40. https://doi.org/10.1073/pnas.0601266103.

[25] GNU Operating System. GNU General Public License. https://www.gnu.org/licenses/gpl-3.0.en.html;. accessed 07.07.2020.

[26] Halloran ME, Longini IM, Nizam A, Yang Y. Containing bioterrorist smallpox. Science. 2002;298(5597):1428–32. https://doi.org/10.1126/science.1074674.

[27] Health Information and Quality Authority. Evidence summary for COVID-19 viral load over course of infection. https://www.hiqa.ie/sites/default/files/2020-04/Evidence-Summary_COVID-19_duration-of-infectivity-viral-load_0.pdf;; 2020. accessed 05.08.2020.

[28] Hou C, Chen J, Zhou Y, Hua L, Yuan J, He S, et al. The effectiveness of quarantine of Wuhan City against the Corona virus disease 2019 (COVID-19): a well-mixed SEIR model analysis. J Med Virol. 2020;92(7):841–8. https://doi.org/10.1002/jmv.25827.

[29] Kaufman HJ, Edlund S, Douglas VJ. Infectious Disease Modeling: Creating a Community to Respond to Biological Threats. Statistical Communications in Infectious Diseases. 2009;1(1). https://doi.org/10.2202/1948-4690.1001.

[30] Kermack WO, McKendrick GA. A contribution to the mathematical theory of epidemics. Bull Math Biol. 1991;53:1–2. https://doi.org/10.1098/rspa.1933.0106.

[31] Kirkcaldy R, King B, Brooks J. COVID-19 and Postinfection immunity. JAMA. 2020;323(22):2245. https://doi.org/10.1001/jama.2020.7869.

[32] Linlin B, Deng W, Gao H, Xiao C, Liu J, Xue J, et al. Lack of Reinfection in Rhesus Macaques Infected with SARS-CoV-2 bioRxiv ; 2020. https://doi.org/10.1101/2020.03.13.990226.

[33] Longini IM, Halloran ME, Nizam A, Yang Y. Containing Pandemic Influenza with Antiviral Agents. Am J Epidemiol. 2004;159(7):623–33.

[34] Mahmood I, Arnabnejad H, Suleimenova D, Sassoon I, Marshan A, Serrano A, et al. FACS: A geospatial agent-based simulator for analyzing COVID-19 spread and public health measures on local regions. http://bura.brunel.ac.uk/handle/2438/20914; 2020.

[35] Martinez D, Das T. Design of non-pharmaceutical intervention strategies for pandemic influenza outbreaks. BMC Public Health. 2014;14(1). https://doi.org/10.1186/1471-2458-14-1328.

[36] Miami Dade County Public Schools. STATISTICAL HIGHLIGHTS 2018–2019. http://drs.dadeschools.net/StatisticalHighlights/M970%20-%20ATTACHMENT%20%20Statistical%20Highlights%2018-19.pdf;; 2019.

[37] Miami Matters. Households/Income Data for County: Miami-Dade. http://www.miamidadematters.org/demographicdata?id=414&sectionId=936;; 2020. accessed 05.08.2020.

[38] Midas-network, 2020. COVID-19/parameter_estimates/2019_novel_coronavirus/, https://github.com/midas-network/COVID-19/tree/master/parameter_estimates/2019_novel_coronavirus; 2020 [accessed 07.07.2020]

[39] Nitkin K, Johns Hopkins Medicine. Coronavirus Screening Test Developed at Johns Hopkins. https://www.hopkinsmedicine.org/coronavirus/articles/screening-test.html;; 2020.

[40] Ota M. Will we see protection or reinfection in COVID-19? Nat Rev Immunol. 2020;20(6):351. https://doi.org/10.1038/s41577-020-0316-3.

[41] Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling medrXiv ; 2020. https://doi.org/10.1101/2020.02.16.20023465.

[42] Silva PCL, Batista PVC, Lima HS, Alves MA, Guimaraes FG, Silva RCP. COVID-ABS: an agent-based model of COVID-19 epidmeic to simulate health and economic effects of social distancing interventions. Chaos Solitions Fractals. 2020;139:110088. https://doi.org/10.1016/j.chaos.2020.110088.

[43] Silva W, Das T, Izurieta R. Estimating disease burden of a potential A(H7N9) pandemic influenza outbreak in the United States. BMC Public Health. 2017;17(1). https://doi.org/10.1186/s12889-017-4884-5.

[44] The City of Miami - Florida. COVID-19 Updates. https://www.miamigov.com/Notices/News-Media/COVID-19-Updates;; 2020. accessed 05.08.2020.

[45] United Census Bureau. All Sectors: County Business Patterns by Legal Form of Organization and Employment Size Class for U.S., States, and Selected Geographies. https://data.census.gov/cedsci/table?q=miami%20dade&g=0500000US12086&tid=CBP2017.CB1700CBP&n=4231%3A4232%3A4233%3A4234%3A4235%3A4236%3A4237%3A4238%3A4239%3A4241%3A4242%3A4243%3A4244%3A4245%3A4246%3A4247%3A4248%3A4249&vintage=2017;; 2017. accessed 05.08.2020.

[46] United Census Bureau. All Sectors: County Business Patterns by Legal Form of Organization and Employment Size Class for U.S., States, and Selected Geographies: 2017, https://data.census.gov/cedsci/table?q=miami%20dade&g=0500000US12086&tid=CBP2017.CB1700CBP&t=Employment%20History&n=4231%3A4233%3A4234%3A4235%3A4236%3A4237%3A4238%3A4239%3A4241%3A4242%3A4243%3A4244%3A4245%3A4246%3A4247%3A4248%3A4249&hidePreview=true; [accessed 05.08.2020].

[47] United Census Bureau. ACS DEMOGRAPHIC AND HOUSING ESTIMATES. https://data.census.gov/cedsci/table?q=miami%20dade&g=0500000US12086&tid=ACSDP1Y2018.DP05;. accessed 05.08.2020.

[48] United Census Bureau. Annual Business Survey: Statistics for Employer Firms by Industry, Sex, Ethnicity, Race, and Veteran Status for the U.S., States, Metro Areas, Counties, and Places: 2017, https://data.census.gov/cedsci/table?q=miami%20dade&g=0500000US12086&hidePreview=false&tid=ABSCS2017.AB1700CSA01&vintage=2017&layer=VT_2018_050_00_PY_D1&cid=DP05_0001E&t=Employment%20Size; [accessed 05.08.2020].

[49] United States Census Bureau. QuickFacts Miami-Dade County, Florida. https://www.census.gov/quickfacts/fact/table/miamidadecountyflorida/POP060210. accessed 05.19.2020.

[50] United States Census Bureau. SEX BY SCHOOL ENROLLMENT BY LEVEL OF SCHOOL BY TYPE OF SCHOOL FOR THE POPULATION 3 YEARS AND OVER. https://data.census.gov/cedsci/table?q=miami%20dade&g=0500000US12086&tid=ACSDT1Y2018.B14002&vintage=2018&t=School%20Enrollment%3AType%20of%20School. accessed 05.08.2020.

[51] Uribe-Sánchez A, Savachkin A, Santana A, Prieto-Santa D, Tapas KD. A predictive decision-aid methodology for dynamic mitigation of influenza pandemic, 33; 2011; 721–49. https://doi.org/10.1007/s00291-011-0249-0 3.

[52] Wu F, Zhao S, Yu B, Chen Y, Wang W, Song Z, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265–9. https://doi.org/10.1038/s41586-020-2008-3.

[53] Yang S, Cao P, Du P, Wu Z, Zhang Z, Yang L, et al. Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis. Ann Transl Med. 2020;8(4):128. https://doi.org/10.21037/atm.2020.02.66.

[54] Yang Z, Zeng Z, Wang K, Wong S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020;12(3):165–74. https://doi.org/10.21037/jtd.2020.02.64.

[55] Zhang Y, Jiang B, Yuan J, Tao Y. The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study medRxiv ; 2020. https://doi.org/10.1101/2020.03.04.20031187.

Appendix D: Published Materials in Infectious Disease Modelling Journal

Infectious Disease Modelling 6 (2021) 839–847

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Impact of school reopening on pandemic spread: A case study using an agent-based model for COVID-19

Hanisha Tatapudi*, Tapas K. Das

Department of Industrial and Management System Engineering, University of South Florida, Tampa, FL, USA

ARTICLE INFO

Article history:
Received 25 March 2021
Received in revised form 9 June 2021
Accepted 10 June 2021
Available online 8 July 2021
Handling Editor: Dr HE DAIHAI HE

Keywords:
COVID-19
SARS-CoV-2
Agent-based simulation model
School reopening
School transmission rate
Age-specific impact

ABSTRACT

This article examines the impact of partial/full reopening of school/college campuses on the spread of a pandemic using COVID-19 as a case study. The study uses an agent-based simulation model that replicates community spread in an urban region of U.S.A. via daily social mixing of susceptible and infected individuals. Data representing population demographics, SARS-CoV-2 epidemiology, and social interventions guides the model's behavior, which is calibrated and validated using data reported by the government. The model indicates a modest but significant increase (8.15%) in the total number of reported cases in the region for a complete (100%) reopening compared to keeping schools and colleges fully virtual. For partial returns of 75% and 50%, the percent increases in the number of reported cases are shown to be small (2.87% and 1.26%, respectively) and statistically insignificant. The AB model also predicts that relaxing the stringency of the school safety protocol for sanitizing, use of mask, social distancing, testing, and quarantining and thus allowing the school transmission coefficient to double may result in a small increase in the number of reported infected cases (2.14%). Hence for pandemic outbreaks from viruses with similar characteristics as for SARS-CoV-2, keeping the schools and colleges open with a modest campus safety protocol and in-person attendance below a certain threshold may be advisable.

1. Introduction

When the COVID-19 pandemic hit the U.S. in late February of 2020, policy makers across the nation implemented a lockdown by closing non-essential businesses and switching to virtual operation for schools/colleges and some of the essential workplaces. By August 2020, as several regions in the U.S. saw a decline in the number of new cases, the school districts decided to either partially or fully reopen school/college campuses for students to return. We developed a plan for a model-based study of the impact school/college reopening had on the community by comparing the increase in the number of infected cases between continued virtual operation and various levels of reopening (50%, 75%, and 100%). Since many adverse events took place soon after reopening of schools/colleges in September of 2020, such as setting in of the winter weather prompting people to be indoors, U.S. presidential election rallies in September and October, Thanksgiving holidays in

* Corresponding author.
E-mail addresses: tatapudi@usf.edu (H. Tatapudi), das@usf.edu (T.K. Das).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2021.06.007

Abbreviations

List of abbreviations and acronyms

AB	Agent-based;
CI	Confidence interval
SEIR	Susceptible exposed infected recovered/removed
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
COVID-19	Coronavirus Disease 2019
CDC	Centers for Disease Control and Prevention
SARS	Severe Acute Respiratory Syndrome
pre-K	Pre-kindergarten

the last week of November, and Christmas holidays, the total number of infections increased immensely in the last months of 2020. Hence, a key question that we examined is what portion of the increased cases was contributed by the reopening of schools/colleges. This paper presents our findings and conclusions, which we believe will be useful for decision makers in potential future pandemic outbreaks of similar (SARS) virus types.

To aid our investigation, we developed a comprehensive AB simulation model that mimics pandemic spread using COVID-19 outbreak in an urban outbreak region as a case study; the region we focused on is Miami-Dade County of Florida, USA with 2.8 million population. The AB model yields estimate for age-stratified numbers of actual infections, reported cases, hospitalized, and dead. The model was calibrated and validated using publicly available data from the region until end of September 2020 as schools reopened on September 30. We froze the model parameters to their calibrated values till end September and ran the simulation till the end of December 2020 with school reopening as the only major new event. We examined various reopening scenarios (e.g., 50%, 75%, and 100% return to campus) and different estimated values of transmission coefficients at schools and colleges. Transmission coefficients were assumed to be 1.5x, 2x, 2.5x, and 3x the estimated transmission coefficient for essential workplaces, where control measures are easier to implement and maintain. Results from different scenarios were compared with those for the respective base cases (0% student return and 1.5x school transmission coefficient).

For pandemics before COVID-19, school closure was considered an important and impactful social intervention (Sypsa & Hatzakis, 2009). Several studies examined the impact of school closure strategies (Cauchemez et al., 2009; Earn et al., 2012; Germann et al., 2006; Halder et al., 2010; Lee et al., 2010; Sypsa & Hatzakis, 2009). However, the strategies evaluated in many of these studies were implemented for a limited time window, after which a complete reopening occurred with minor variations. In contrast, COVID-19 has persisted for most of 2020 and is continuing into 2021, and thus, in addition to school closure policies, there is also a need to understand school reopening policies. Another distinctive difference between COVID-19 and the past pandemics is that the latter have mostly been caused by influenza viruses, which affects children and older adults more than the other age groups, whereas COVID-19 affects older adults more.

Several studies have evaluated the impact of school closure and reopening for COVID-19. Although most studies have concluded that school reopening may not pose a significant burden on the epidemic (Courtemanche et al., 2020; Iwata et al., 2020), some others have contradicted this statement (Domenico et al., 2021; Head et al., 2020; Keeling et al., 2020). Study presented in (Keeling et al., 2020) used an SEIR transmission model to evaluate eight different strategies under varying degrees of mixing by reopening schools for different class groups. They concluded that reopening increases mixing, however this can be constrained to keep the reproductive number under 1 using interventions. A similar conclusion was drawn from a different study (Domenico et al., 2021) that used a stochastic discrete age-structured transmission model that implemented a progressive and prompt reopening strategy with varying percentage of attendance. Study presented in (Head et al., 2020) evaluated reopening under two susceptibility assumptions (one where ages below 20 are half as susceptible as adults, and the other where young population is equally as susceptible as adults) and two transmission contexts (high and moderate community transmission). They recommended a hybrid-learning approach within smaller cohorts of 20 students for elementary schools and 10 students for high schools.

Our choice of the use of an agent based (AB) simulation model was deliberate to accommodate population and regional demographics and fine-grained dynamics of the pandemic, as well as the range of time-varying social interventions that were in place. Most available SEIR type continuum models are often only qualitative approximations that compromise epidemiological realism (Siettos & Russo, 2013), which is especially true for a complex and stochastically evolving pandemic like COVID-19. However, for AB models, uncertainty in parameters including structure of the contact network, distribution of individual attributes, and presence of feedbacks could expand the confidence ranges of the outputs (Rahmandad & Sterman, 2008). As age played a critical role in the transmission, disease progression, hospitalizations, and deaths in COVID-19, our ability to calibrate and validate an age-stratified model and obtain age-stratified predictions was also a significant benefit of choosing the AB modeling approach. However, in the interest of conciseness of the paper, we have presented only cumulative results and have not shown their age-stratified breakdowns. Readers are referred to (Tatapudi et al., 2020) for age-stratified results.

2. Methodology

Original version of our AB simulation model for COVID-19 was presented in (Tatapudi et al., 2020). The model is particularized for an urban metropolitan region in the U.S. (Miami-Dade County of Florida with 2.8 million population). The AB model generates individual people according to the U.S. census data (by age and occupational distribution), households (by adult and children distribution), schools, workplaces, and community locations. A daily (hour by hour) schedule is assigned to every individual, chosen from a set of alternative schedules, based on their attributes. The model also incorporates temporal changes in the social interventions that were in place during most of 2020. Interventions include complete and partial lockdowns, school closure and reopening, face mask mandate, and limited contact tracing. The model also considers varying levels of compliances for isolation and quarantine, lower on-site staffing levels of essential work and community places

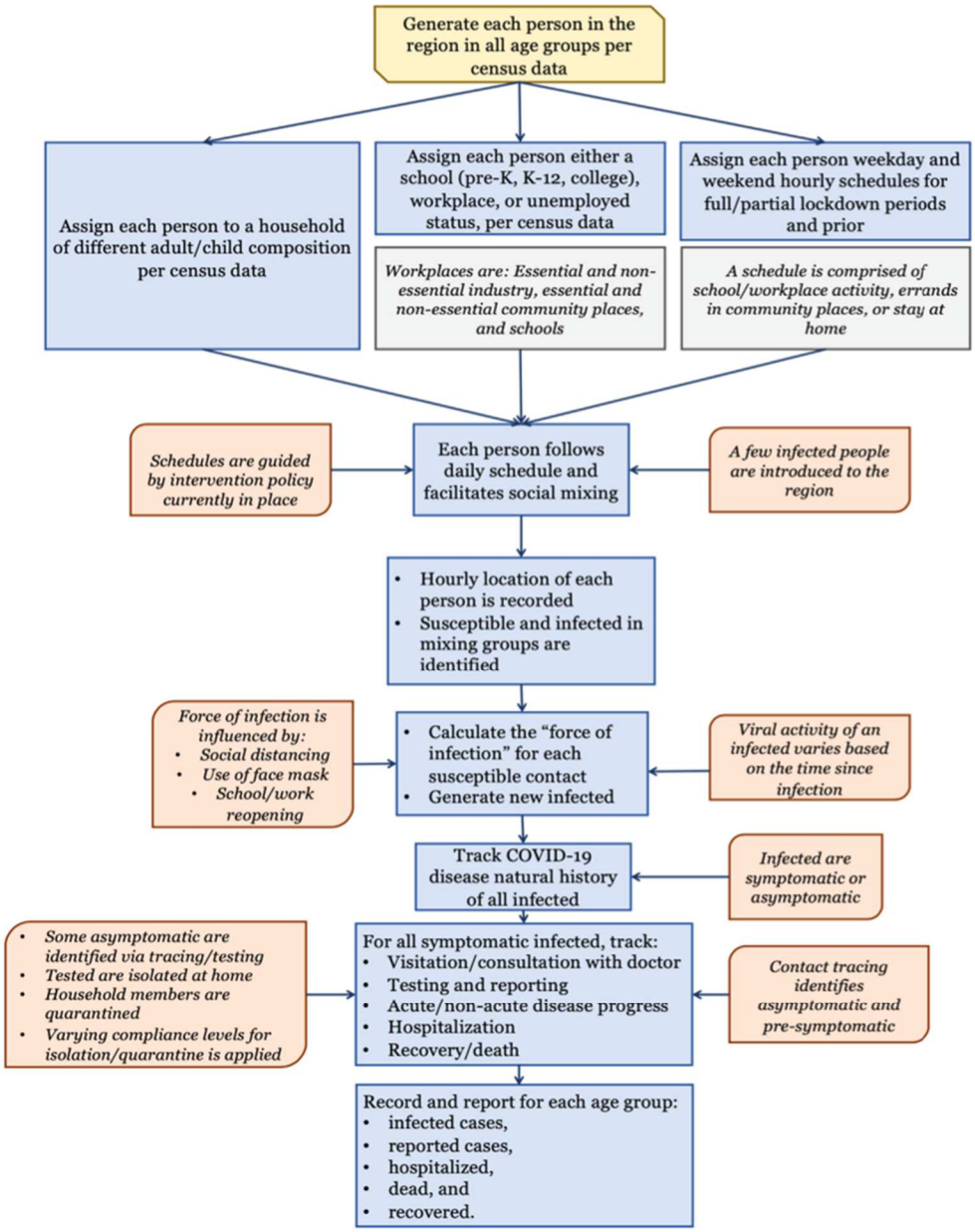


Fig. 1. A flow diagram for the agent-based simulation model for COVID-19 (Tatapudi et al., 2020) {Color}.

during stay-at-home order, restricted daily schedule of people during various social intervention periods. The AB model reports daily and cumulative values of actual infected, doctor visits, tested, reported cases, hospitalized, recovered, and deaths, for each age category. For more specific details on the input data used for building the model, the algorithmic sequence, and the functional structure of the simulation model, reader is referred to (Tatapudi et al., 2020). Fig. 1 shows the schematic diagram depicting the algorithmic sequence for the AB simulation model. The model was calibrated using parameters for transmission coefficients at home, work, school, and community places (see (Tatapudi et al., 2020)). Calibration was done so that the daily cumulative numbers of reported cases from the AB simulation model closely matched the values published in the Florida COVID-19 dashboard until September 30, 2020. Fig. 2 shows the daily cumulative reported cases of infection (average from 13 runs with 95% confidence intervals) from AB simulation model. The dotted line in Fig. 2 shows the cumulative case growth from the surveillance data from the Florida COVID-19 dashboard for Miami-Dade County (Florida'sD-19 Data a).

3. Results

We used our model to predict the incremental growth of infected cases, reported cases, hospitalizations, and deaths for the region for various levels of student return and an estimated value of the school transmission coefficient (2.0x), assuming that the social mixing conditions other than school reopening are the same throughout the remaining months of 2020. Results are compared with the baseline scenario of 0% return, in which operation of all schools and colleges remains fully virtual until the end of the year 2020. In Miami-Dade County, 21.5% of the population (approximately 600,000 out of 2.8 million people) attend schools (pre-K through 12th grade) and colleges (community colleges, four-year colleges, and universities). Fig. 3 summarizes the model outcomes for all four levels of return to school, i.e., 0%, 50%, 75% and 100%. In the model implementation, for scenarios with partial return (50% and 75%), students were rotated. The graphs show average values with 95% C.I. (from thirteen simulation runs with different seeds) of the cumulative numbers of (actual) infected cases, reported cases, hospitalized, and dead. Since the CIs for different level of returns mostly overlap with each other, the average and the CI values for the last day of simulation are noted on figures. As expected, shortly after the schools and colleges reopen (September 30), the curves begin to diverge. Table 1 summarizes the numerical values of the outcomes from the graphs in Fig. 3 for December 31, 2020. The average numbers do increase for reopening with 50% and 75% returns, but the increases are small. For example, the cumulative reported cases by the year end are higher than the base case with 0% return by only 1.26% (p-value 0.59) and 2.87% (p-value 0.22), for 50% and 75% returns, respectively). Notably, the percentage cumulative increase in reported cases for 100% return is almost threefold higher compared to 75% return. When compared with the base case with 0% return, 100% return resulted in a significantly higher cumulative reported cases (by 8.15% with p-value 0.00118). It can also be seen from Fig. 3 that for up to 75% return, the daily reported cases reach a decreasing pattern by December 31 and fall near or below a threshold of 100 new daily cases. Whereas for 100% return, the trend for cumulative reported cases remains increasing and the daily reported number on December 31 is still relatively high at 390. It appears that the social mixing process caused by student return of higher than 75% crosses a threshold leading to a sustained (non-decreasing) pattern of new infections until the end

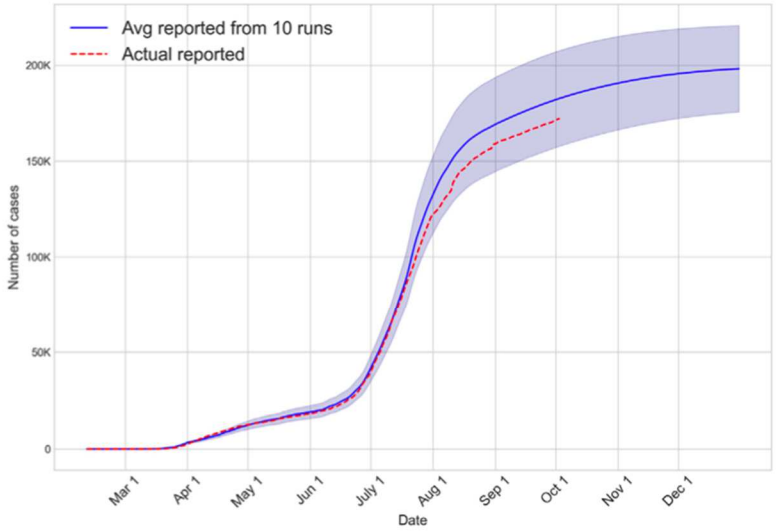


Fig. 2. Cumulative plot of the reported cases of infection from AB simulation (average with 95% CI in shade) from Feb 2020–Dec 2020 along with surveillance data in dotted line. {Color}.

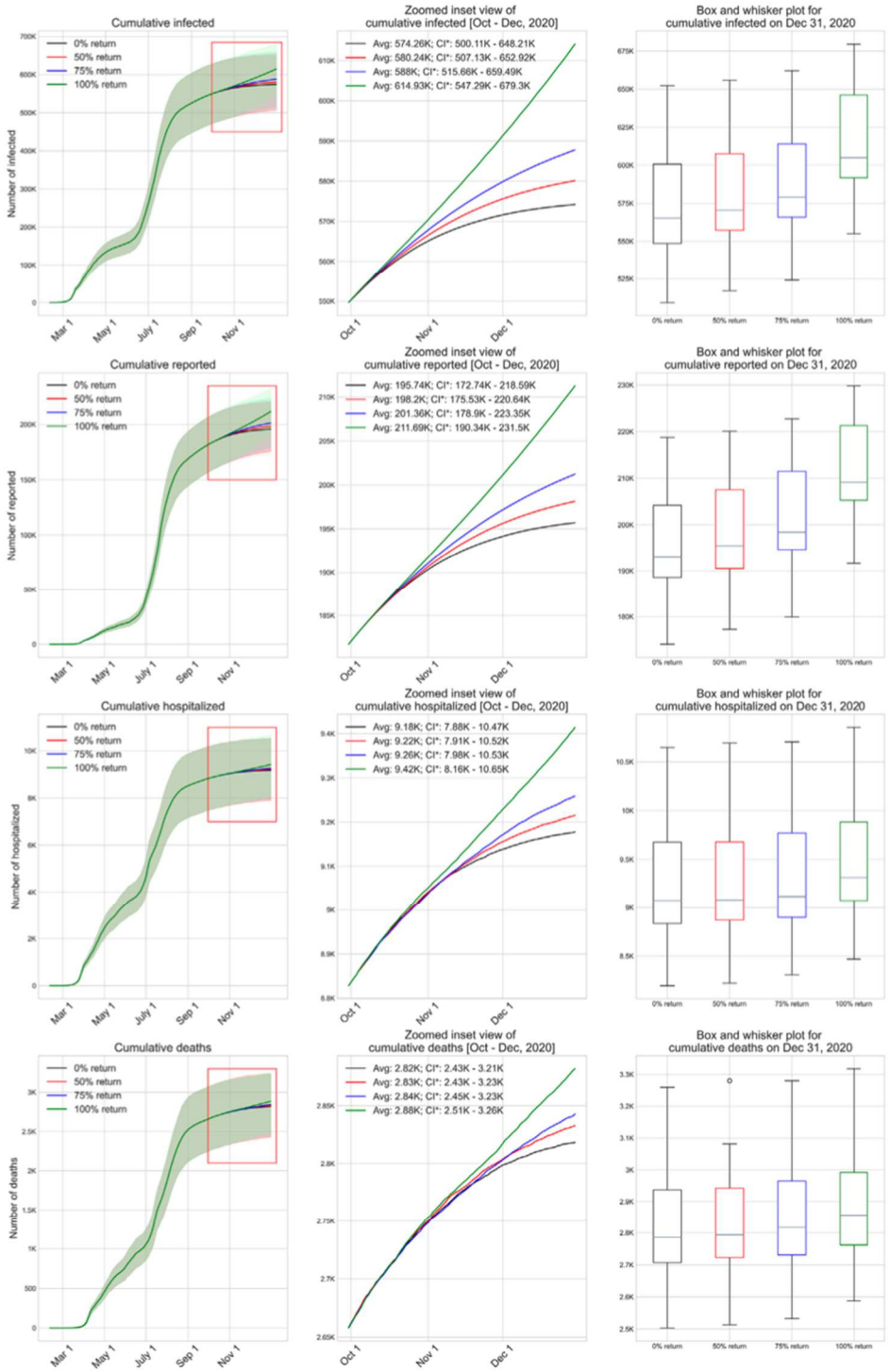


Fig. 3. Sensitivity analysis with various levels of school reopening on September 30, 2020 (between no return to 100% return) on the spread of COVID-19 assuming school transmission coefficient is 2.0x times the workplace transmission coefficient {Color}.

Table 1
Summary of COVID-19 outcomes on Dec 31, 2020 for various levels of student return to school and colleges, as reported by the AB simulation model; the numbers correspond to school transmission coefficient being 2.0 x transmission coefficient at workplaces.

Percentage of student return				
Outcomes on Dec 31, 2020	0% (base case)	50%	75%	100%
Daily reported cases	34	59	113	390
Cumulative infections (%increase over base case)	574261	580242 (1.04%)	587999 (2.39%)	614932 (7.08%)
Cumulative reported cases (%increase over base case)	195736	198204 (1.26%)	201356 (2.87%)	211689 (8.15%)
Cumulative hospitalizations (%increase over base case)	9178	9217 (0.42%)	9262 (0.92%)	9421 (2.65%)
Cumulative deaths (%increase over base case)	2818	2833 (0.53%)	2843 (0.89%)	2884 (2.34%)

of the year. Since the numbers of hospitalization and deaths are derived from the reported cases, graphs for those display similar patterns.

Hereafter, we conducted a sensitivity analysis for the transmission coefficient at schools and colleges. Ensuring effective safety protocol at schools and colleges is resource intensive and many school districts don't have the human and financial wherewithal to adopt a safety protocol (cleaning, sanitizing, social distancing, testing, and quarantining) of the highest standard as recommended by CDC (Centers for Disease Contr, 2020). Consequently, the transmission coefficient may vary significantly among schools and colleges. In our sensitivity analysis, we have examined a number of scenarios where transmission coefficient at schools and colleges are 1.5x, 2.0x, 2.5x, and 3x that of the workplaces (offices and businesses); we have assumed 1.5x as the base case. For this analysis, we maintained the level of student return at 50%, since a survey of Miami-Dade County school district showed that nearly 50% of students intended to return to campus (Strauss, 2020).

Fig. 4 depicts the outcomes from scenarios with four different transmission coefficient values at schools and colleges. Table 2 summarizes the numerical values of the outcomes from Fig. 4 for December 31, 2020. The impact of increase in transmission coefficient for schools and colleges follow an expected increasing trend. However, the increases in cumulative reported cases for scenarios with 2.0x, 2.5x, and 3.0x transmission coefficients compared to the base case of 1.5x were not statistically significant (p-values are 0.80, 0.59, and 0.36, respectively).

4. Discussion

This study presents an AB simulation model aided investigation of how the reopening of schools and colleges likely has impacted the numbers of actual infected, reported, hospitalized, and dead from COVID-19 pandemic in a densely populated urban region of U.S.A. The AB model is highly flexible and is able to mimic scenarios with a number of chosen levels (%) of students returning to schools and colleges for in-person instruction. The model also accommodates varying levels safety precautions adopted by schools and colleges represented by the values of the transmission coefficient. Results of the model-based investigation can be summarized as follows: 1) for up to 75% return, new additional cases from reopening of schools and colleges is likely to be small, 2) 100% return is expected to cause a steady but modest increase in the number of additional cases till the end of the year, and 3) even when the transmission coefficient at schools and colleges is three times that of the workplaces, a scenario that is indicative of inadequate safety practice, additional number of reported cases is not expected to rise drastically (less than 2.5%). In summary, in order to reduce incremental infections from school reopening, keeping the level of student return below a threshold appears to be important. Also, not having the resources to implement a very high level of campus safety protocol may not be a critical barrier to school reopening. Our findings appear to concur with those reported in the recent studies and media reports (Auger et al., 2020; Gewertz, 2020; Insights for education. C, 2020; Lopez, 2020; Mandavilli, 2020; Taxin & Beam, 2020). It is worth mentioning, as noted in the introductory paragraph, that the actual number of infections, hospitalizations, and deaths reported between September and December of 2020 were much higher than predicted by our model. For example, the number of reported infections by end of December 2020 in Miami-Dade County was almost 100 K higher than predicted by the model. It is common knowledge that this sharp increase in the months following school reopening was caused by multitudes of other events leading to higher social mixing that took place in the U.S. Our model allowed us to maintain the societal conditions at the pre-school reopening status and thus examine the true impact of school reopening.

The AB model has limitations as it is an abstraction of how a pandemic impacts a large and complex society. A limited number of pre-set daily schedules are used to approximate a highly dynamic contact process, and it does not account for variabilities in types and lengths of interactions. All the schools, work, and community places are assumed to be uniformly distributed in the simulated region, though they usually tend to grow in clusters in urban settings. We also did not implement the quarantining and on-campus testing in schools and colleges. At the time of completing this short communication, we have also completed a study examining two issues for COVID-19: the impact of vaccine (Pfizer/BioNTech and Moderna) administration in the early months of 2021, and an efficacy comparison among different prioritization strategies being considered by various countries and localities. Our findings are presented in (Tatapudi et al., 2021).

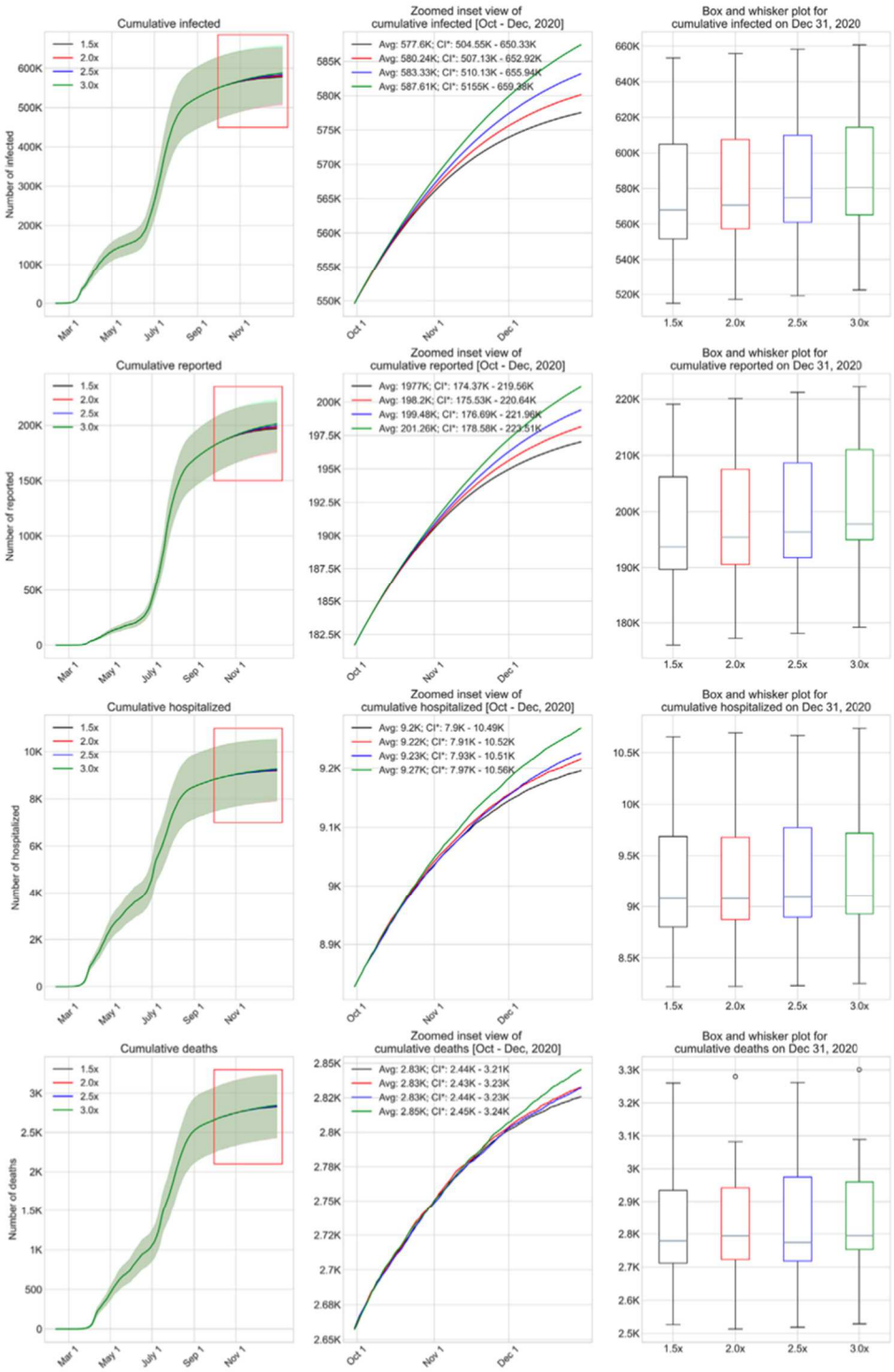


Fig. 4. Sensitivity analysis with various values of school transmission coefficient (considered to be 1.5x to 3.0x of the workplace transmission coefficient) on the spread of COVID-19 with 50% of the students returning to campus for in-person instruction beginning on September 30, 2020 {Color}.

Table 2
Summary of COVID-19 outcomes on Dec 31, 2020, obtained from the AB simulation model for 50% student return and various transmission coefficient values.

	Transmission coefficient at school			
Outcomes on Dec 31, 2020	1.5 x transmission coefficient of workplaces (base case)	2.0 x transmission coefficient of workplaces	2.5 x transmission coefficient of workplaces	3.0 x transmission coefficient of workplaces
Daily reported cases	46	59	80	101
Cumulative infections (%increase over base case)	577604	580242 (0.46%)	583327 (0.99%)	587611 (1.73%)
Cumulative reported cases (% increase over base case)	197065	198204 (0.578%)	199484 (1.23%)	201259 (2.13%)
Cumulative hospitalizations (% increase over base case)	9197	9217 (0.22%)	9227 (0.33%)	9269 (0.78%)
Cumulative deaths (%increase over base case)	2826	2833 (0.248%)	2832 (0.21%)	2845 (0.67%)

Author statement

Hanisha Tatapudi: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – review & editing, Data curation. Tapas Das: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Writing – original draft, Visualization, Supervision.

Funding

The authors have not been funded by any grant or organization for this work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Auger, K., Shah, S., Richardson, T., Hartley, D., Hall, M., Warniment, A., et al. (2020). Association between statewide school closure and COVID-19 incidence and mortality in the US. *Journal of the American Medical Association, 324*(9), 859–870.

Cauchemez, S., Ferguson, N. M., Wachtel, C., Tegnell, A., Saour, G., Duncan, B., et al. (2009 Aug). Closure of schools during an influenza pandemic. *The Lancet Infectious Diseases, 9*(8), 473–481.

Centers for Disease Control and Prevention. (2020). Operating schools during COVID-19: CDC's considerations [internet]. [cited 2020 Oct 19]. Available from https://www.cdc.gov/coronavirus/2019-ncov/community/schools-childcare/schools.html.

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). Strong social distancing measures in the United States reduced the COVID-19 growth rate. *Health Affairs, 39*(7), 1237–1246.

Domenico, L., Pullano, G., Sabbatini, C., Boëlle, P., & Colizza, V. (2021). Modelling safe protocols for reopening schools during the COVID-19 pandemic in France. *Nature Communications, 12*(1), 1073.

Earn, D., He, D., Loeb, M., Fonseca, K., Lee, B., & Dushoff, J. (2012). Effects of school closure on incidence of pandemic influenza in Alberta, Canada. *Annals of Internal Medicine, 156*(3), 173–181.

Florida's COVID-19 data and surveillance dashboard,. Florida Department of Health, Division of Disease Control and Health Protection.

Germann, T., Kadau, K., Longini, I., & Macken, C. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences, 103*(15), 5935–5940.

Gewertz C. Schools need to Be bolder' about reopening, Public Health Expert Says, [Internet]. Education Week. [cited 2020 Oct 19]. Available from: https://www.edweek.org/ew/articles/2020/11/03/schools-need-to-be-bolder-about-reopening.html.

Halder, N., Kelso, J., & Milne, G. (2010). Developing guidelines for school closure interventions to be used during a future influenza pandemic. *BMC Infectious Diseases, 10*(221).

Head, J., Andrejko, K., Cheng, Q., Collender, P., Phillips, S., Boser, A., et al. (2020). The effect of school closures and reopening strategies on COVID- 19 infection dynamics in the san Francisco Bay Area: A cross- sectional survey and modeling analysis. *MedRxiv [Internet].* https://doi.org/10.1101/2020.08.06.20169797. Available from.

Insights for education. COVID-19 and Schools: What we can learn from six months of closures and reopening. [Internet]. [cited 2020 Oct 19]. Available from https://education.org/facts-and-insights#f09a6e46-8c5f-4d01-8297-d2a3f6c8f873.

Iwata, K., Doi, A., & Miyakoshi, C. (2020). Was school closure effective in mitigating coronavirus disease 2019 (COVID-19)? Time series analysis using Bayesian inference. *International Journal of Infectious Diseases, 99*, 57–61.

Keeling, M., Tildesley, M., Atkins, B., Penman, B., Southall, E., Guyver-Fletcher, G., et al. (2020). The impact of school reopening on the spread of COVID-19 in England. *MedRxiv*.

Lee, B., Brown, S., Cooley, P., Potter, M., Wheaton, W., Voorhees, R., et al. (2010). Simulating school closure strategies to mitigate an influenza epidemic. *Journal of Public Health Management and Practice, 16*(3), 252–261.

Lopez G. What we've learned so far from school reopenings in the US [Internet]. Vox. [cited 2020 Oct 19]. Available from. https://www.vox.com/future-perfect/21494352/school-reopenings-covid-coronavirus-pandemic-in-person-teaching.

Mandavilli A. Schoolchildren seem unlikely to Fuel coronavirus surges, scientists say [internet]. New York Times. [cited 2020 Oct 19]. Available from. https://www.nytimes.com/2020/10/22/health/coronavirus-schools-children.html.

Rahmandad, H., & Sterman, J. (2008 May). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science, 54*(5), 998–1014.

Siettos, C. I., & Russo, L. (2013 May 15). Mathematical modeling of infectious disease dynamics. *Virulence, 4*(4), 295–306.

Strauss, V. (2020). Florida education commissioner orders Miami to open schools earlier than planned [Internet]. *Washington Post*. [cited 2020 Oct 19]. Available from https://www.washingtonpost.com/education/2020/09/27/florida-education-commissioner-orders-miami-open-schools-earlier-than-planned/.

Sypsa, V., & Hatzakis, A. (2009 Jun 18). School closure is currently the main strategy to mitigate influenza A(H1N1)v: A modeling study. *Euro Surveillance, 14*(24), 19240.

Tatapudi, H., Das, R., & Das, T. (2020 Nov 1). Impact assessment of full and partial stay-at-home orders, face mask usage, and contact tracing: An agent-based simulation study of COVID-19 for an urban region. *Global Epidemiology, 2*, 100036.

Tatapudi, H., Das, R., & Das, T. (2021). Impact of vaccine prioritization strategies on mitigating COVID-19: An agent-based simulation study using an urban region in the United States. *MedRxiv*.

Taxin A, Beam A. California sees No link from school openings to virus spread [internet]. NBC Bay Area. [cited 2020 Oct 19]. Available from: https://www.nbcbayarea.com/news/california/no-link-seen-between-california-school-openings-virus-cases/2376111/.

Appendix E: Article Submitted to BMC Medical Research Methodology Journal

Impact of Vaccine Prioritization Strategies on Mitigating COVID-19: An Agent-Based Simulation Study using an Urban Region in the United States

Hanisha Tatapudi[1], Rachita Das[2], and Tapas K. Das [1]

[1] Department of Industrial and Management System Engineering, University of South Florida, Tampa, Florida, USA.

[2] Miller School of Medicine, University of Miami, Miami, Florida, USA.

Corresponding author: Hanisha Tatapudi

Email: tatapudi@usf.edu

as quickly as possible, after providing for the most vulnerable. As much of the population worldwide is yet to be vaccinated, results from this study should aid public health decision makers in effectively allocating their limited vaccine supplies.

Keywords: Vaccination strategies, COVID-19, Agent-based simulation model, Vaccination policies, Vaccination prioritization

ABSTRACT

Background

Approval of novel vaccines for COVID-19 had brought hope and expectations, but not without additional challenges. One central challenge was understanding how to appropriately prioritize the use of limited supply of vaccines. This study examined the efficacy of the various vaccine prioritization strategies using the vaccination campaign underway in the U.S.

Methods

The study developed a granular agent-based simulation model for mimicking community spread of COVID-19 under various social interventions including full and partial closures, isolation and quarantine, use of face mask and contact tracing, and vaccination. The model was populated with parameters of disease natural history, as well as demographic and societal data for an urban community in the U.S. with 2.8 million residents. The model tracks daily numbers of infected, hospitalized, and deaths for all census age-groups. The model was calibrated using parameters for viral transmission and level of community circulation of individuals. Published data from the Florida COVID-19 dashboard was used to validate the model. Vaccination strategies were compared using a hypothesis test for pairwise comparisons.

Results

Three prioritization strategies were examined: a minor variant of CDC's recommendation, an age-stratified strategy, and a random strategy. The impact of vaccination was also contrasted with a no vaccination scenario. The study showed that the campaign against COVID-19 in the U.S. using vaccines developed by Pfizer/BioNTech and Moderna 1) reduced the cumulative number of infections by 10% and 2) helped the pandemic to subside below a small threshold of 100 daily new reported cases sooner by approximately a month when compared to no vaccination. A comparison of the prioritization strategies showed no significant difference in their impacts on pandemic mitigation.

Conclusions

Even though vaccines for COVID-19 were developed and approved much quicker than ever before, their impact on pandemic mitigation was small as the explosive spread of the virus had already infected a significant portion of the population, thus reducing the susceptible pool. A notable observation from the study is that instead of adhering strictly to a sequential prioritizing strategy, focus should perhaps be on distributing the vaccines among all eligible

as quickly as possible, after providing for the most vulnerable. As much of the population worldwide is yet to be vaccinated, results from this study should aid public health decision makers in effectively allocating their limited vaccine supplies.

Keywords: Vaccination strategies, COVID-19, Agent-based simulation model, Vaccination policies, Vaccination prioritization

INTRODUCTION

SARS-CoV-2 and resulting COVID-19 disease has been raging world-wide since early 2020, killing over 2.0 million globally and nearly 450,000 in the United States by the end of January 2021 [1]. A significant winter swell in cases was underway in the U.S. between November 2020 and January 2021 despite protective measures in place such as face mask usage, limited contact tracing, travel restrictions, social distancing practices, and partial community closures. To combat this, many promising novel vaccines were developed, of which two (Pfizer/BioNTech and Moderna) were authorized for emergency use in mid-December 2020 by the U.S. Food and Drug Administration (USFDA) [2]. Data from initial trials of cohorts greater than 30,000 people showed that these vaccines, given in two doses, were safe and have ~95% effectiveness in preventing COVID-19 [3]. Vaccine deployment in the U.S. began soon after USFDA approval.

Implementing an effective vaccination campaign was essential to dramatically reduce the infection, hospitalization, and death rates, but posed many unique challenges. Vaccine prioritization and allocation strategy was at the forefront of the challenges to effectively vaccinate communities. Strategy was influenced by a number of key factors: 1) limited initial vaccine supply in the months following release, 2) transmission and severity of COVID-19 varying by segment of the population, 3) vaccine approvals only for adults, and 4) acceptability and compliance in the community for two dose vaccination [4].

U.S. Centers for Disease Control (CDC) released an outline prioritizing healthcare personnel, first responders, persons with high risk medical conditions for COVID-19, and older adults >65 years. These groups were given priority for vaccination in phase 1, when supply was limited. In phase 2 (supply increased to begin to meet demand) and phase 3 (supply was greater than demand), other population groups were vaccinated based on age and availability [5]. Vaccine allocation structures with basic similarities and some key differences were used by countries around the world. For example, after healthcare workers, France's vaccine allocation scheduled other general workers regardless of age who they had determined to be at high risk of contracting and spreading the virus due to contact with the general public. This includes retail, school, transportation, and hospitality staff [6]. Such differences in vaccine prioritization strategies were untested at the time of the study and warranted modeling and examination.

The goal of this paper was to investigate the impact of vaccination on the pandemic via outcome measures such as numbers of infected, hospitalized, and deaths in the months following December 15, 2020 when vaccination would begin in the U.S. Two specific objectives of our investigation were: 1) to assess the expected impact of the vaccination program on mitigating COVID-19, and 2) to inform public health officials on the comparative benefits, if any, of the different vaccine prioritization strategies. We conducted our investigation by using our agent-based (AB) simulation model for COVID-19 that was presented during the early phase of the pandemic [7]. For this study, we first extended calibration of our model till December 30, 2020 to ensure that our model appropriately tracked the explosive increase in cases that started with the onset of winter and the year-end holiday period in 2020. We then enhanced the AB simulation model by adding a framework for vaccination. This included: vaccination priorities for people based on attributes including profession and age, use of two different vaccines by Pfizer/BioNTech and Moderna with their contracted quantities and approximate delivery timelines, vaccine acceptance, transition period between each priority group, vaccination rate, and immunity growth for vaccinated starting with the first dose.

As in [7], we implemented our calibrated AB model, augmented with vaccination, for Miami-Dade County of the U.S. with 2.8 million population, which had been an epicenter of COVID-19 in the U.S. We conducted our investigation by implementing a number of prioritization strategies and obtaining the corresponding numbers of total infections, reported infections, hospitalizations, and deaths. The strategies implemented were 1) minor variant of the CDC recommended strategy, 2) age stratified strategy, and 3) random strategy. We also implemented a no vaccination case. These strategies are explained in a later section. We compared and contrasted the results to assess vaccination efficacy and relative performances of the priority strategies. We made a number of key observations from the results, which we believe will help public health officials around the world to choose effective vaccine prioritization strategies to mitigate the negative impacts of COVID-19.

LITERATURE REVIEW

On a global scale, equitable and ethical distribution of vaccines for all (low, medium, and high-income) countries is an important question. As the world leader in promoting global health, World Health Organization (WHO) released an evidence-based framework for vaccine-specific recommendations [8]. WHO proposed vaccine prioritization for

three potential scenarios of transmission: community transmission, sporadic cases or cluster of cases, and no cases. Each scenario has three stages and focuses on different risk groups. COVID-19 pandemic resembles "community transmission." For this, the first stage focused on healthcare workers and older adults with highest risk; second stage continued the focus on older adults and people with comorbidities, sociodemographic groups, and educational staff; and the third stage focused on essential workers and social/employment groups unable to physically distance themselves.

National Academy of Sciences, Engineering, and Medicine (NAESM) developed a more comprehensive phased framework for equitable allocation of COVID-19 vaccine [9]. The first phase prioritized healthcare workers and first responders, people with high risk comorbidities, and older adults in crowded living conditions; second phase focused on K-12 school staff and child care workers, essential workers, people with moderate risk comorbidities, people living in shelters, physically and mentally disabled people and staff that provide care, employment settings where social distancing was not possible, and remaining older adults; third phase prioritized young adults, children, and workers; and fourth phase included everyone else. No specific studies had been presented to the literature at the time of this research that had evaluated the efficacy of the proposed vaccination priorities for mitigating COVID-19.

A number of studies can be found in the literature on vaccination strategies for controlling outbreaks of other viruses. The work presented in [10] analyzes the effect of both CDC guided targeted vaccination strategy as well as a mass vaccination strategy for seasonal Influenza outbreaks in the U.S. The study found that a mass vaccination policy reaped the most benefits both in terms of cost and quality-of-life years (QALYs) lost. Authors in [11] use a genetic algorithm to find optimal vaccine distribution strategies that minimize illness and death for Influenza pandemics with age specific attack rates similar to the 1957–1958 A(H2N2) Asian Influenza pandemic and the 1968–1969 A(H3N2) Hong Kong Influenza pandemic. They consider coverage percentage under varying vaccine availability and developed an optimal vaccination approach that was 84% more effective than random vaccination. A study reported in [12] examined vaccination to prevent interpandemic Influenza for high-risk groups and children, and recommended concentrating on schoolchildren, who were most responsible for transmission, and then extended to high-risk groups. A compartmental model in [13] was used to develop optimal strategies to reduce the morbidity and mortality of the H1N1 pandemic. The study found that age specific vaccination schedules had the most beneficial impact on mortality.

It can be concluded from the above review of relevant literature that there is no 'one size fits all' strategy for vaccination to either prevent a pandemic outbreak or mitigate one. Virus epidemiology and corresponding disease characteristics, as well as the efficacy and supply of the vaccine must be considered in developing an effective vaccination prioritization strategy. Our paper aims to address this need by presenting a detailed AB simulation modeling approach and using it to assess efficacy of vaccine prioritization strategies for COVID-19.

METHODOLOGY

Published COVID-19 modeling approaches were either data-driven models, as in [14]–[18], or variants of SEIR type compartmental models as in [19]–[23]. Data driven models are very well suited for understanding the past progression of a pandemic and also for estimating parameters characterizing virus epidemiology. However, these models offer limited ability to predict the future progression of a pandemic that could be dynamically evolving with regards to virus epidemiology, disease manifestations, and sociological conditions. Compartmental models, on the other hand, are aggregate in nature and do not adapt well to changing dynamics of disease transmission. An AB modeling approach is considered to be more suitable for a detailed accounting of individual attributes, specific disease natural history, and complex social interventions [24]. Hence we adopted the AB model approach to conduct our study.

The AB simulation-based methodology was particularized using data for Miami Dade County of Florida, with 2.8 million population, an epicenter for COVID-19 spread in the South-Eastern United States. A step by step approach for building such a model for another region can be found in [7]. The methodology began by generating the individual people according to the U.S. census data that gives population attributes including age (see Table A1 in [7]) and occupational distribution (see Table A4 in [7]). Thereafter, it generated the households based on their composition characterized by the number of adults and children (see Table A2 in [7]). The model also generated, per census data, schools (see Table A3 in [7]) and the workplaces and other community locations (see Table A4 in [7]). Each person was assigned a workplace and household based on the numbers, sizes, and compositions of households, schools and workplaces derived from census data (see Tables A2, A3, and A4 in [7] for distribution of household, schools, and workplaces, respectively). A daily (hour by hour) schedule was assigned to every individual, chosen from a set of alternative schedules based on their attributes. The schedules vary between weekdays and weekends

and also depend on the prevailing social intervention orders (see Table A1 in Additional file 1). The methodology incorporated means to implement a number of intervention orders including full and partial closures/reopening of schools and workplaces [7], [25], isolation and quarantine of infected individuals and household members, use of face mask, contact tracing of asymptomatic and pre-symptomatic individuals, and vaccination priorities. The timeline for interventions implemented in the model are summarized in Table 1.

Table 1: Social intervention order timeline for Miami-Dade County

Intervention policy implemented at Miami-Dade County, Florida, U.S.	**Date of implementation**	**Day of Simulation**
Stay at home policy	March 17 2020	35
Phase I reopening	May 18 2020	97
Phase II reopening	June 5 2020	115
Mandatory usage of face mask	June 25 2020	135
Contact tracing (assumed to begin)	June 30 2020	140
Phase III reopening	September 25 2020	227
School reopening	September 30 2020	232
Vaccination begin day	December 15 2020	308

The AB model also included a number of uncertainties such as 1) time varying values of testing rates for symptomatic and asymptomatic, test sensitivity, and test result reporting delay (see Table A9 in [7]), 2) self-isolation compliance for symptomatic cases and quarantine compliance for susceptible household members (see Table A10 in [7]), 3) time varying and age specific probabilities of 'hospitalization among reported cases' and 'death among hospitalized' (see Tables A2 and A3 in Additional file 1), 4) mask usage compliance (100% compliance was assumed to reduce transmission coefficient by 33% [26]), 5) contact tracing level (assumed to be at 15% based on [27], [28]), 6) percentage return to school (considred to be 50% based on [29]), 7) willingness to vaccinate (considered to be 60% based on survey results in [30], [31]), 8) variations in daily schedules during various phases of social interventions (see Table A5 in [7]), 9) percentage of asymptomatic among infected (assumed to be 35% [32]), and 10) vaccine efficacy (95% [3]).

On the first day of the simulation, the model introduced a few initial infected cases in the community and began the social mixing process. Each day the model tracked the following for each person: 1) hourly movements and locations based on their daily schedules that depend on age, employment status, prevailing social intervention

orders, and quarantine/isolation status; 2) hourly contacts with other susceptible and infected; 3) vaccination status and immunity, 4) force of infection accumulation; 5) start of infection; 6) visits/consultation with a doctor (if symptomatic and insured); 7) testing (if infected and visited/consulted a doctor or asymptomatic chosen for testing either randomly or via contact tracing); 8) test reporting delay; 9) disease progression (if infected); 10) hospitalization (if infected and acutely ill); and 11) recovery or death (if infected). The AB model reports daily and cumulative values of actual infected, reported infections, hospitalized, and deaths, for each age category. A schematic diagram depicting the algorithmic sequence and parameter inputs for the AB simulation model is presented in Figure 1.

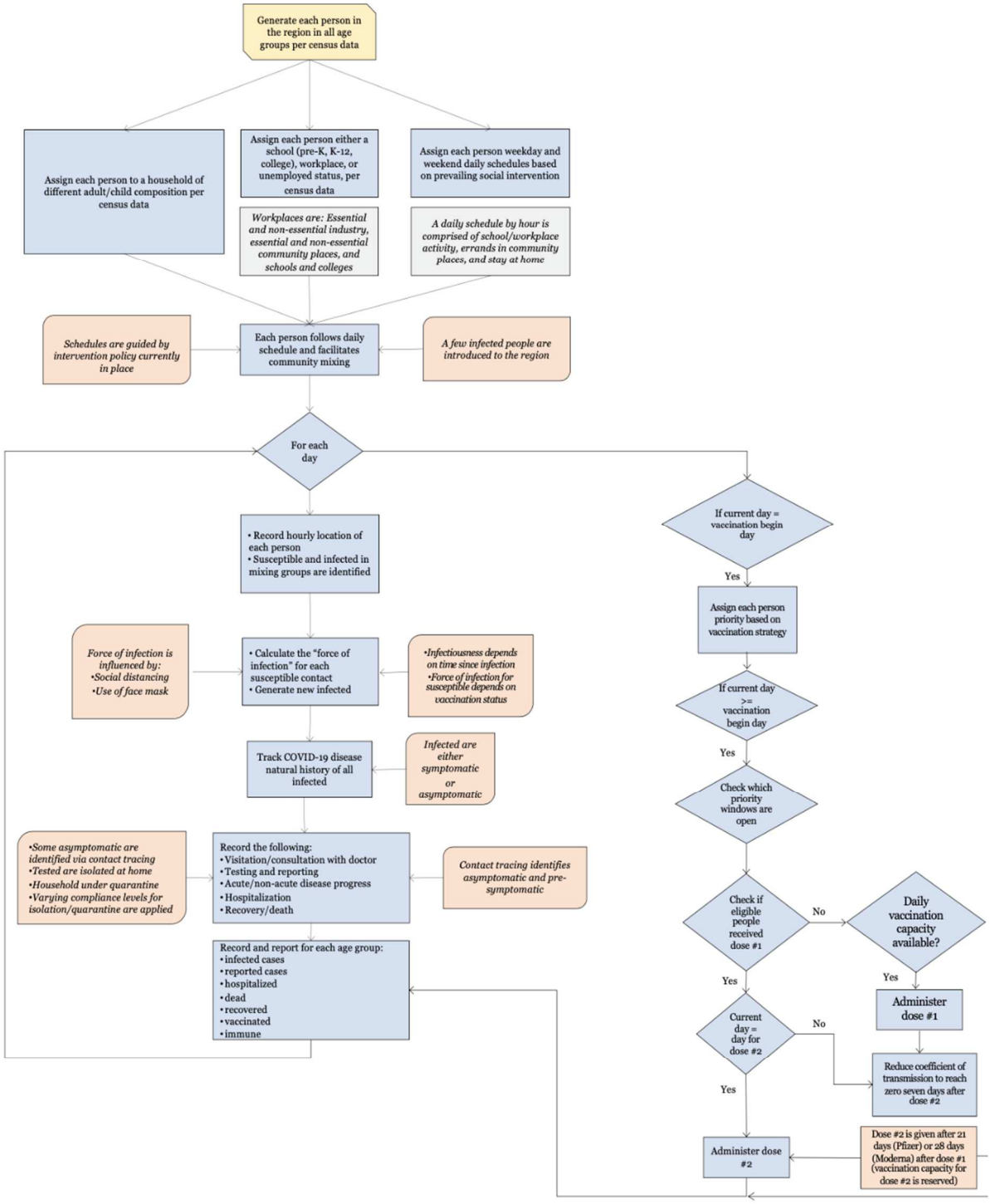


Figure 1: Schematic of the AB model for mimicking COVID-19 spread under social interventions and vaccination in the U.S.

For the susceptible, if vaccinated, since the exact immune response from vaccine was not known at the time of the study, we assumed a linearly increasing partial immunity for susceptible after they received the first dose, attaining full immunity after seven days after the second dose; we only considered vaccines made by Pfizer/BioNTech and Moderna, for which the second dose was administered 21 and 28 days after the first dose, respectively. The model updated the infection status of all individuals to account for new infections as well as disease progressions of infected individuals. A pseudo-code in Figure A1 in Additional file 1 depicts the major elements and structure of the AB simulation program.

For the infected, we considered the following epidemiological parameters: disease natural history with average lengths of latent, incubation, symptomatic, and recovery periods; distribution of infectiousness; percent asymptomatic; and fatality rate. The infected people were considered to follow a disease natural history as shown in Figure 2, parameters of which can be found in Table A6 of [7]. The model assumed that the recovered cases became fully immune to further COVID-19 infections. However, since this assumption was not fully supported by data at the time of the study, people recovered from COVID-19 were also considered candidates for vaccination. The duration and intensity of infectiousness was guided by a lognormal density function (see Figure 3). The function was truncated at the average length of the infectiousness period (which was considered to be 9.5 days). Asymptomatic cases were assumed to follow a similar infectiousness intensity profile but scaled by a factor , as used in the force of infection calculation (1) (see Table A4 in Additional file 1).

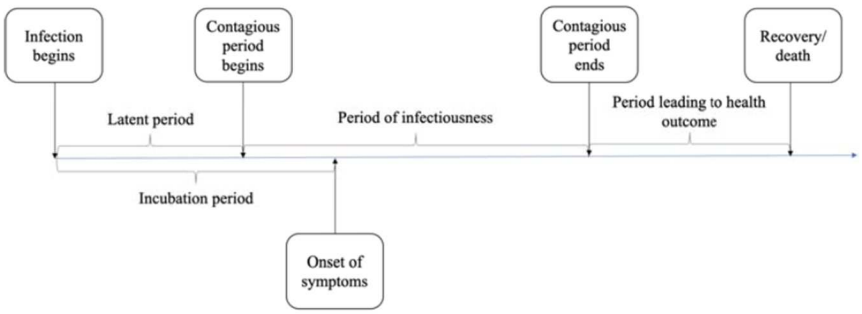


Figure 2: Disease natural history of COVID-19 [7]

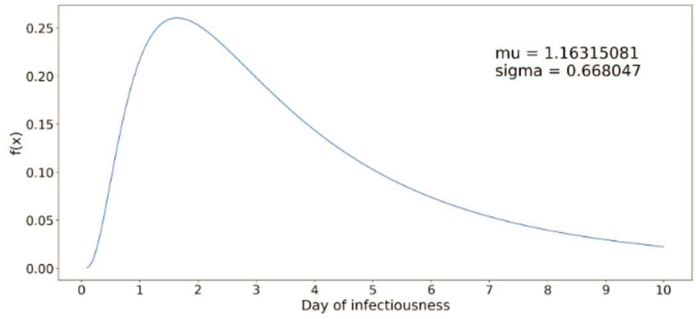


Figure 3: Lognormal distribution function for infectiousness profile of a COVID-19 case [7]

In what follows we describe how we compute the force of infection and used in determining the probability of infection of individuals. Each susceptible was assumed to ingest viral particles from each infected they come in contact with during the day. The total amount of ingestion of viral particles for a susceptible i was measured as the force of infection (λ_i) using a modified version of the equation presented in [33]. The equation for force of infection (as presented in [33]) has three components to account for ingestions experienced from infected contacts at home, workplace/school/indoor community locations, and outdoor locations. As we had not found evidence in the literature of any significant COVID-19 transmissions at outdoor locations, we eliminated the third component and used only the first two elements of the equation as shown in (1). The first component of (1) accounts for the daily force of infection experienced by a susceptible i from those infected at home, and the second component accounts for workplace/school/indoor community locations.

$$\lambda_i = \sum_{k|h_k=h_i} \frac{I_k \beta_h \kappa(t-\tau_k) \rho_k [1+C_k(\omega-1)]}{n_i^{\alpha}} + \sum_{j,k|l_k^j=l_i^j} \frac{I_k \beta_p^j \kappa(t-\tau_k) \rho_k [1+C_k(\omega-1)]}{m_i^j}. \quad (1)$$

The definitions and values of parameters of (1) can be found in Table A4 in Additional file 1. The daily force of infection was considered to accumulate. However, it was assumed that if a susceptible does not gather any additional force of infection (i.e., does not come in contact with any infected) for two consecutive days, the cumulative force of infection for the susceptible reduces to zero. At the end of each day, the cumulative value of λ_i was used to calculate the probability of infection for susceptible i as $1 - exp^{-\lambda_i}$. This probability was used to classify a susceptible individual as infected in the simulation model.

As observed in [7], though it was implemented for a specific region, our model is quite general in its usability for other urban regions with similar demography, societal characteristics, and intervention measures. In our model, demographic inputs (age and household distribution, number of schools for various age groups, and number of workplaces of various types and sizes) were curated from both national and local census records. Social interventions vary from region to region and the related parameters can be easily updated. Similarly, the data related to epidemiology of COVID-19 were unlikely to significantly vary from one region to another, though some adjustments of these based on population demographics may be needed.

Model Calibration

The AB model utilized a large number of parameters, which were *demographic*, *epidemiological*, and *social intervention parameters*. We kept almost all of the above parameters fixed at their respective chosen values and calibrated the model by changing values for only a few. The calibrated parameters include the transmission coefficients used in calculating force of infection at home, work, school, and community places (β_h and β_p^j) (see Table A4 in Additional file 1). The choice of the values of transmission coefficients was initially guided by [34] and thereafter adjusted at different points in time during the calibration period (until December 30, 2020). The only other parameters that were calibrated are the number of errands in the daily schedules under various intervention conditions and the percentage of workers in essential (e.g., healthcare, utility services, and grocery stores) and non-essential (e.g., offices and restaurants) workplaces who physically reported to work during different intervention periods (see Table A1 in Additional file 1). Calibration of the above parameters was done so that the daily cumulative numbers of reported infected cases from the AB simulation model closely matched the values published in the Florida COVID-19 dashboard until December 30, 2020. Figure 4 shows the cumulative values (with 95% confidence intervals) as well as the daily numbers of the reported infected cases, hospitalizations, and deaths as obtained from the simulation model. The dotted lines represent the actual numbers reported in the Florida COVID-19 dashboard for Miami-Dade County [35], the trend of which was closely followed by the numbers obtained from our AB simulation model. The actual numbers reported by the Florida COVID-19 dashboard showed large variations, which was due to reporting delays.

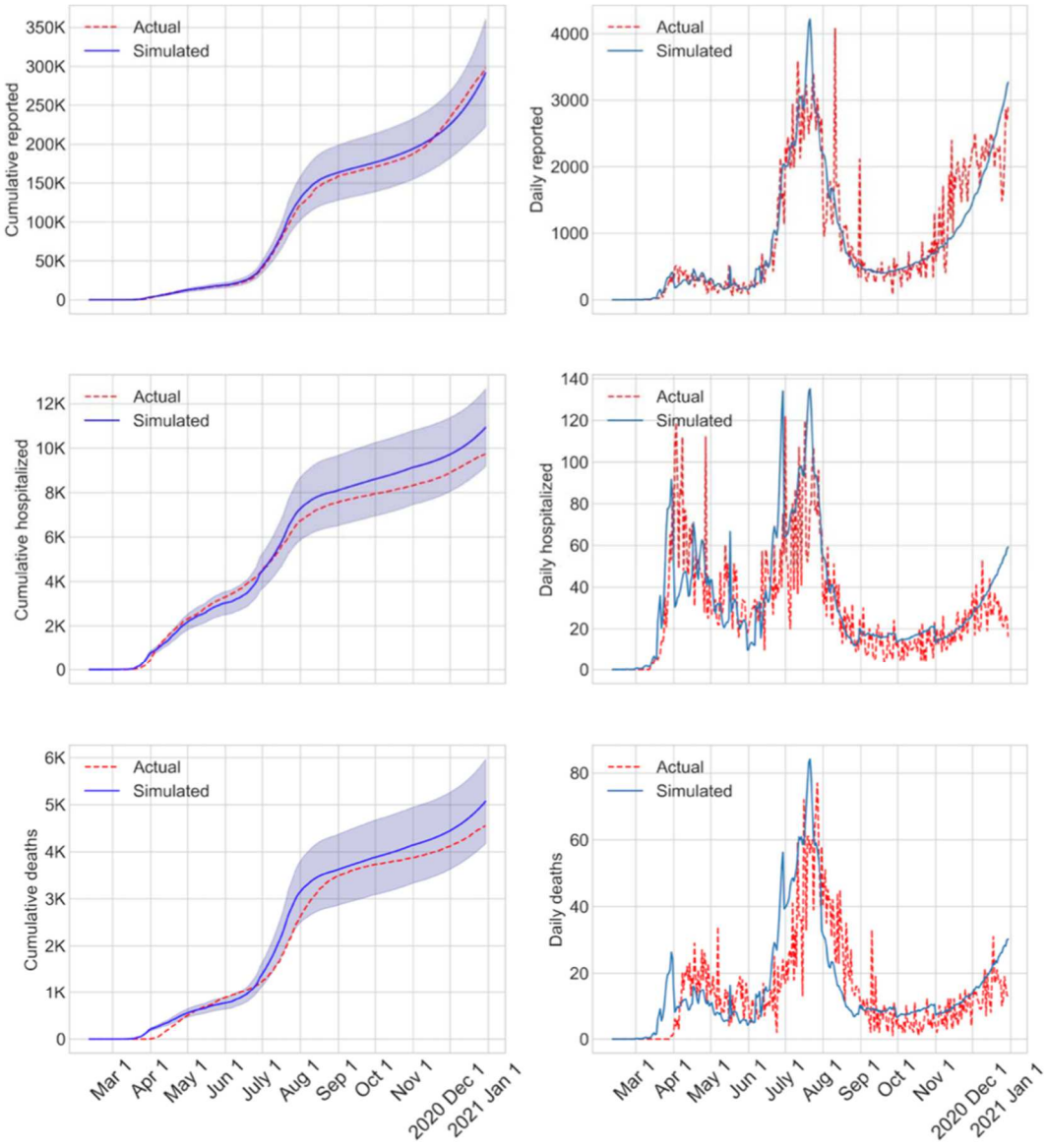


Figure 4: Validation graphs with cumulative and daily values of infected, reported, and deaths calibrated until Dec 30, 2020

VACCINE PRIORITIZATION STRATEGIES

We used our AB model to examine the expected benefits of the ongoing vaccination in the U.S. using the limited supply of two types of vaccines developed by Pfizer/BioNTech and Moderna, which had the emergency approvals for distribution from the USFDA. We considered the number of vaccine doses that the two companies were contracted by the U.S. government to supply, which include the initial contracts for 100 million doses from each

company and the more recent contract for an additional 100 million doses from Pfizer/BioNTech. This amounted to a total of 300 million doses, which could inoculate 150 million people with two required doses. To our knowledge, the total supply was being apportioned among the states and the counties depending on the population. Florida had approximately 6.5% of the U.S. population and the Miami Dade County had 13% of Florida's population. Hence, we assumed that Miami Dade County would receive approximately 2.54 million doses and would be able to vaccinate 1.27 million people out of the total 2.8 million population. We also assumed that the vaccine deliveries will occur in batches starting in late December 2020 till late June 2021. Our study goal was to first determine the extent of reduction in the number of infections, hospitalizations, and deaths that we can expect to realize from the vaccination process in comparison with if no vaccines were available. Thereafter, we conducted a comparative study between three different vaccination priority schemes to determine if the outcomes (number of reported cases, hospitalized, and deaths) from those were statistically significant.

	Priority 1	Priority 2	Priority 3	Priority 4	Priority 5
Age stratified strategy	• Healthcare providers • Nursing home residents • *Dec 15, 2020 – Jan 13, 2021*	• People of ages 65 and over • *Jan 14 – Feb 12 2021*	• People of ages 55 to 64 • *Feb 13 – Mar 14 2021*	• People of ages 45 to 54 • *Mar 15 – Apr 13, 2021*	• People of ages 16 to 44 • *Apr 14 – Jul 15, 2021*
Minor CDC variant strategy	• Healthcare providers • Nursing home residents • *Dec 15, 2020 – Jan 13, 2021*	• First responders • Educators • People of ages 75 and over • *Jan 14 – Feb 12, 2021*	• People of ages 65 to 74 • *Feb 13 – Mar 30, 2021*	• People of ages 16 to 64 • *Apr 1 – Jul 15, 2021*	• NA
Random strategy	• Healthcare providers and nursing home residents • *Dec 15 2020 – Jan 13, 2020*	• All people of ages 16 and over • *Jan 14 – Jul 15, 2021*	• NA	• NA	• NA

Figure 5: Vaccine prioritization strategies examined using AB simulation model for COVID-19 in the U.S.

The priority strategies that were examined are broadly described here; a more complete description is presented in Figure 5. In the absence of a declared timeline for transition of eligibility from one priority group to the next, we assumed 30 days between transition. This period was extended to allow all eligible and willing to be vaccinated when the phased vaccine supply fell short of the number of people in the eligible priority group. The first strategy that we implemented was a minor ***variant of the CDC recommended strategy***: Priority 1: healthcare providers and

nursing home residents; Priority 2: first responders, educators, and people of ages 75 and over; Priority 3: people of ages 65 to 74; Priority 4: people of ages 16 to 64. The CDC recommended strategy also included in priority 3 people of ages 16 to 64 with specific health conditions. Since we did not track health conditions in our AB model, we limited our priority 3 to people of ages 65 and above only. The second strategy that we implemented was an ***age-stratified strategy***: Priority 1: healthcare providers and nursing home residents; Priority 2: people of ages 65 and over; Priority 3: people of ages 55 to 64; Priority 4: people of ages 45 to 54; Priority 5: people of ages 16 to 44. The third strategy that we implemented was a ***random strategy***: Priority 1: healthcare providers and nursing home residents; Priority 2: all people of ages 16 and over. People with prior COVID-19 history were not excluded and 60% of the people were considered willing to vaccinate [30], [31].

RESULTS

Figure 6 shows the plots indicating the growth over time in the number of actual reported vaccinations together with those from the three vaccination prioritization strategies that we had implemented. The vaccination began on December 15, 2020. As evident from the figure, the vaccination growth of the CDC variant strategy aligned closely with the acutal reported numbers. In the random strategy, the growth in vaccination occured faster as this strategy opened eligibility to all ages 16 and above sooner than other strategies. The age dependent strategy further staggers the eligibility and hence the vaccination grew slower than the CDC variant strategy. The flattening of the vaccination growth curves was representative of the limited vaccine supply that was considered in our simulation. The continued growth of the acutal reported vaccinations (red dotted line) was indicative of the increase in vaccine supplies since the time of our model implementation in December 2020.

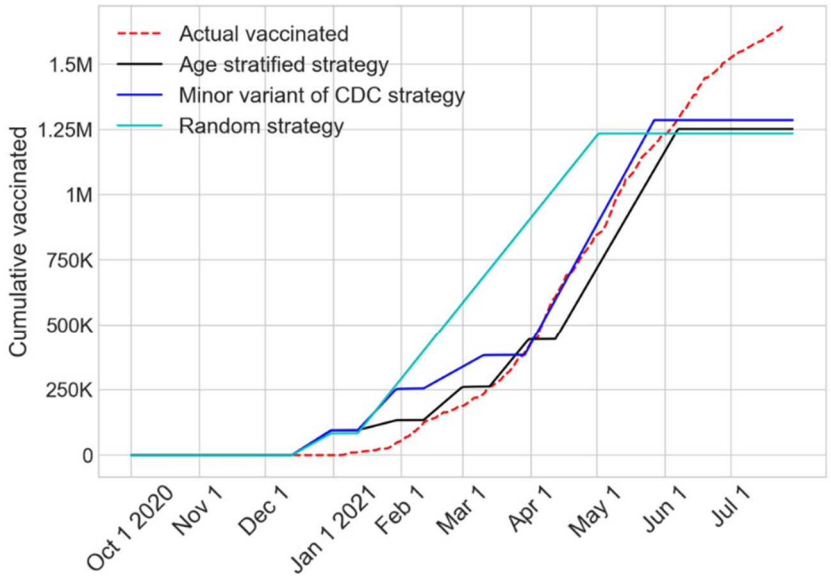


Figure 6: Cumulative values of actual reported and simulated vaccinations until July 31, 2021

Figure 7 shows the trends of the cumulative numbers of infected cases, reported cases, hospitalized, and deaths from the three vaccine prioritization strategies together with the no vaccination scenario until July 31, 2021 (last day of simulation). Since the confidence intervals for the cumulative values of the strategies overlap significantly, for the purpose of comparison, we chose to focus on the cumulative values on the last day of simulation. These are presented in Table 2, which summarizes the values and their 95% confidence intervals. The table also provides the time frame when the pandemic, according to the model, subsides with the new reported cases falling below the threshold of 100. We note that, the reported cases presented in Figure 7 and Table 2 were obtained directly from the simulation model. However, the hospitalization and death numbers were calculated by applying the time-varying and age-specific probabilities (as presented in Tables A2 and A3 in Additional file 1) on the reported cases.

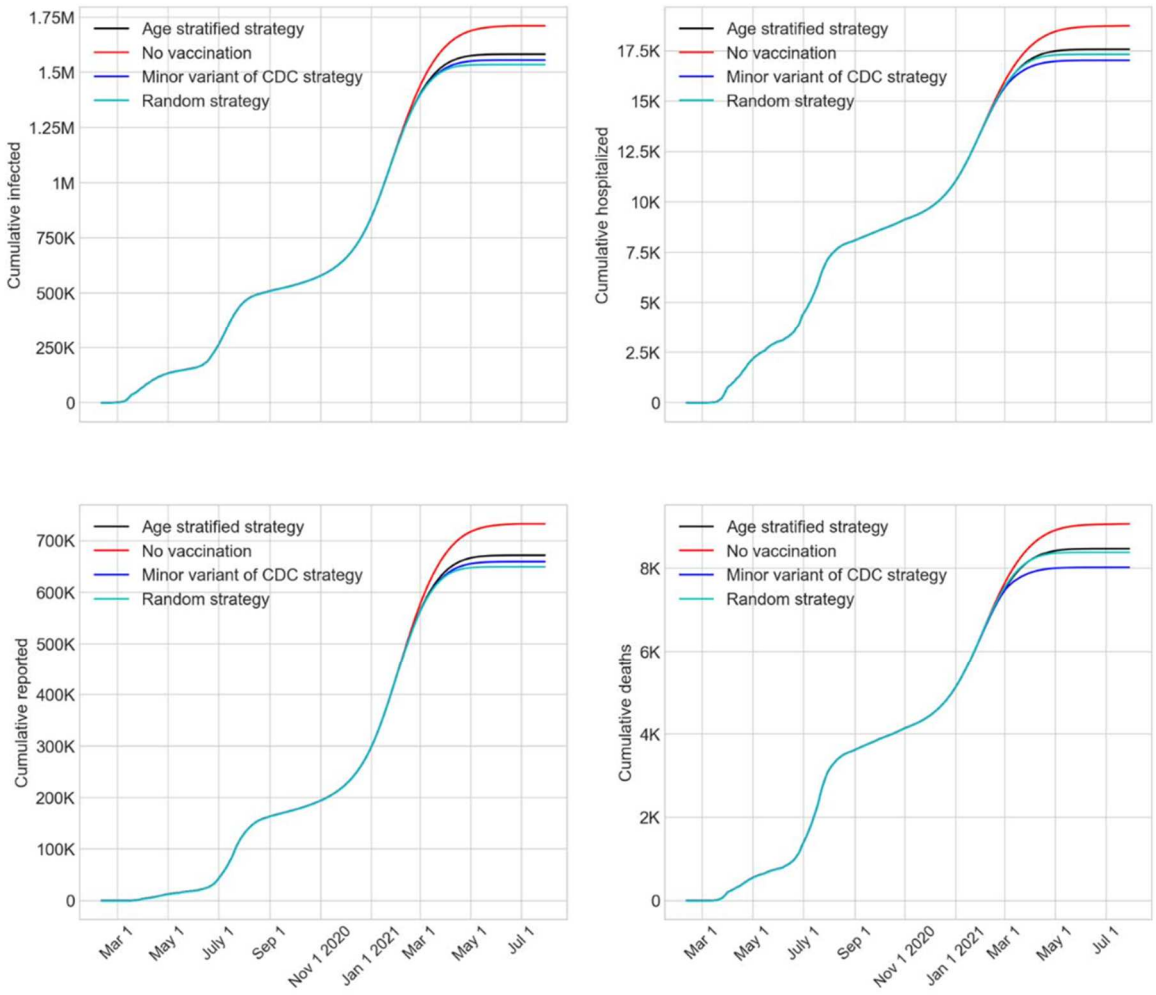


Figure 7: Impact of vaccination strategies on cumulative numbers of infected, reported, hospitalized, and deaths from COVID-19 until July 31, 2021

Table 2: Summary of expected cumulative values (with 95% confidence intervals) on July 31, 2021 obtained by the AB model for the vaccine prioritization strategies

Outcome / Prioritization Strategy	Infected Cases	Reported Cases	Hospitalized	Deaths	Date when new reported cases fell below 100
No vaccination	1.71M (1.66M – 1.76M)	732K (695K – 770K)	18.7K (18.3K – 19.1K)	9K (8.8K – 9.3K)	June 12, 2021
Minor variant of CDC	1.55M (1.38M – 1.73M)	659K (567K – 752K)	17K (15.7k – 18.3k)	8K (7.2K – 8.8K)	May 17, 2021

Age stratified	1.58M (1.43M – 1.73M)	672K (589K – 754K)	17.6K (16.6k – 18.5k)	8.5K (7.9k – 8.9k)	May 22, 2021
Random	1.54M (1.32M – 1.75M)	649K (538K – 761K)	17.3K (15.9K – 18.7K)	8.4K (7.6K – 9.1K)	May 7, 2021

Since the performance of the three vaccine prioritization strategies, as shown in Figure 7, appeared to be similar, we conducted simple pairwise statistical comparisons of the numbers of reported cases (from column 3 of Table 2) using a test of hypothesis. According to the test results, the variant of the CDC strategy produced a statistically significantly lower number of reported cases than no vaccination (p-value 0.0204). Similar results were found for age-stratified and random strategies when compared to no vaccination. However, a comparison of the reported cases among the three vaccine prioritization strategies showed no significant statistical difference (p-values near 0.4). Pairwise comparison of the hospitalized and deaths (in columns 4 and 5 of Table 2) showed that the CDC variant strategy produced statistically significantly lower numbers compared to no vaccination (p-values 0.0014 and 0.0015, respectively). However, similar to the reported cases, the three vaccination strategies did not statistically differ among themselves in terms of hospitalizations and deaths.

The numbers from Table 2 also indicate that the CDC variant strategy achieved a reduction of 9%, 10%, 9%, and 11% for total infected, reported cases, hospitalized, and deaths, respectively, compared to the outcomes with no vaccination. Moreover, the CDC variant resulted in the pandemic subsiding below a small threshold of 100 new reported cases about a month sooner when compared with no vaccination. The CDC variant strategy also spared 5.6% of the population form being infected. The above seemingly low impact of vaccination may be attributed to the explosive growth of new reported cases that occurred in the winter months (Nov. 2020-Mar. 2021), which likely had significantly reduced the pool of susceptible people.

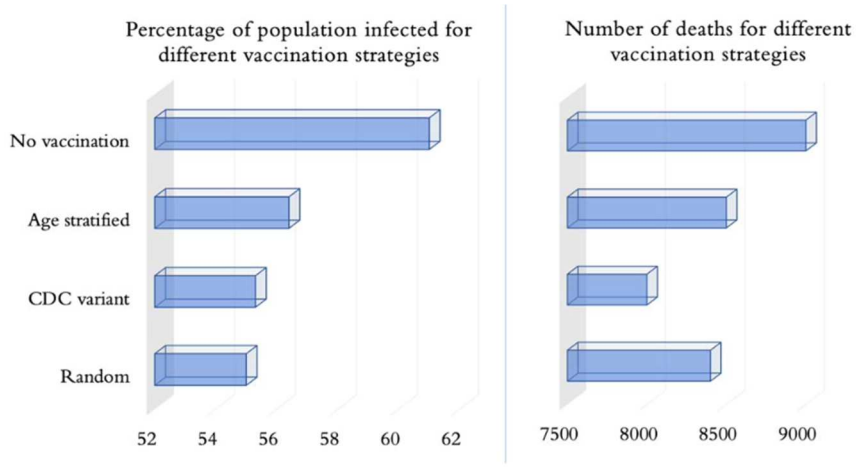


Figure 8: Visualization of the impacts of vaccination strategies on the percentage of total population infected and the total numbers of deaths on July 31, 2021

Figure 8 offers a further visual comparison of the impact of vaccination strategies on the percentage of population infected and the total number of deaths. A few interesting observations can be made from the figure as follows. The random strategy yielded the lowest percentage of population infected (even lower than CDC variant) as it made the age groups most active in social mixing eligible for vaccination sooner than all other strategies. For the same reason however, the random strategy yielded more deaths than CDC variant as it caused a delay in vaccination of the most vulernable elderly population who were not exclusively prioritized. Interstingly, the random strategy also performed better than the age stratified strategy in both measures of infected and deaths.

CONCLUDING REMARKS

We have developed a detailed agent-based simulation model for mimicking the spread of COVID-19 in an urban region (Miami-Dade County, Florida) of the U.S. The model was calibrated using transmission coefficients and parameters guiding the daily schedules of people, and was validated using the actual reported cases from the Florida COVID-19 dashboard till December 30, 2020 (see Figure 4). On this validated model, we incorporated the vaccination process that started in the U.S. on December 15, 2020 using two different vaccines developed by Pfizer/BioNTech and Moderna. Based on the government contracts at the time of our study, we assumed availability of an estimated 2.54 million doses for Miami-Dade County to inoculate 1.27 million people (among the total population of 2.8 million) on a 2 dose regimen.

Model results indicated that the use of the available vaccines reduced the spread of the virus and helped the pandemic to subside below a small threshold of 100 daily new cases by mid-May 2021, approximately a month sooner than if no vaccines were available. Also, the vaccination was shown to reduce the number of infections by approximately 10% compared to no vaccination, which translates to sparing 5.6% of the total population from being infected. We note that, even though the vaccines were developed and approved for human use at a much faster rate than ever accomplished before, the accelerated growth of the infections, especially with the onset of the winter in the northern hemisphere, reduced the expected benefits of vaccination.

Another noteworthy finding of this study was that there were no statistical differences between the numbers of reported cases resulting from different vaccination prioritization strategies that were tested. This information should give more latitude to decision makers in urban regions across the world for distribution of the limited supply of COVID-19 vaccine. Our results suggested that instead of adhering strictly to a sequential prioritizing strategy, focus should be on distributing the vaccines among all eligible as quickly as possible.

Though our AB model is well suited to study future progression of COVID-19 and other similar respiratory type viruses, it has some limitations. The simulation model is highly granular, which while being a strength presents a challenge of appropriately estimating the wide array of its input parameters. Though we have attempted to address this challenge by conducting a sensitivity analysis on some of the critical parameters, such as levels of face mask usage, contact tracing, societal closure, and school reopening, this analysis could be extended to many other parameters. As mentioned under vaccination strategies, our model did not include health conditions that were relevant to COVID-19 (like pulmonary disease, obesity, heart problems) as attributes for people. Hence, we were not able to implement one element of the CDC recommended prioritization strategy that recommends people aged 16-64 years with underlying medical conditions to be considered in priority 3 (see Figure 5). Also, we did not consider any vaccine wastage due to complexities associated with refrigeration, distribution, and human error. We also assumed that vaccination of all priority groups occured uniformly over the eligibility periods considered, which may not reflect the reality. Also, at the time of the study, there was little available literature on the rate of immunity growth each day from the two dose vaccines, therefore, we assumed a linear growth starting with the first dose and

culminating (full immunity) seven days after the second dose. Moreover, the model did not consider, the emergence of new strains of virus as the pandemic progresses, and the lower level of immunity offered by the vaccine for breakthrough infection. Finally, changes in the virus behavior might impact the comparative outcomes of different vaccine prioritization strategies, as presented in this paper.

At the time of the publication of our study, vaccine availability is still limited in many parts of the world. Our results will provide useful information for healthcare policy makers in judiciously allocating their COVID-19 vaccine supply among their population. We also believe that our findings on vaccine prioritization strategies will serve as a resource for the decision makers for future outbreaks of similar respiratory viruses. Finally, as only a limited number of studies examining vaccine prioritization strategies have been presented to the open literature, our research makes a significant addition.

LIST OF ABBREVIATIONS

AB – Agent-based

SEIR – Susceptible – exposed – infected – recovered/removed

SARS-CoV-2 – Severe acute respiratory syndrome coronavirus 2

COVID-19 – Coronavirus Disease 2019

USFDA – United States Food and Drug Administration

CDC – Centers for Disease Control

NAESM – NationalAcademy of Sciences, Engineering, and Medicine

QALY – Quality of Life Years

WHO – World Health Organization

DECLARATIONS

Ethics approval and consent to participate: Individual human data was not used in our study. Only aggregate data made available in Florida COVID-19 Dashboard was used.

Consent for publication: Not applicable

Availability of data and materials: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests: The authors declare that they have no competing interests

Funding: Not applicable

Authors' contributions:

Hanisha Tatapudi: Conceived and designed the model, Selection of model input parameters and data gathering, Coding and testing of the model, Design and perform the experiments, Output analysis and review, Manuscript preparation and review

Rachita Das: Selection of model input parameters and data gathering, Output analysis and review, Manuscript preparation and review

Tapas K Das: Conceived and designed the model, Selection of model input parameters and data gathering, Coding and testing of the model, Design and perform the experiments, Output analysis and review, Manuscript preparation and review

Acknowledgements: Not applicable

Authors' information:

Corresponding author:

Hanisha Tatapudi – corresponding author

e-mail: tatapudi@usf.edu.

Telephone: +1 (813) 453-3577

Rachita Das

e-mail: rachi95@gmail.com

Telephone: +1 (813) 527-1133

Tapas K. Das

e-mail: das@usf.edu

Telephone: +1 (813) 843-0285

REFERENCES

[1] A. Jordan *et al*., "Coronavirus in the U.S.: Latest Map and Case Count," *NYtimes.com*, Jan. 25, 2021. Accessed: Jan. 25, 2021. [Online]. Available: https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html

[2] "COVID-19 Vaccines," *U.S. Food & Drug Administration*, Jan. 12, 2021. Accessed: Jan. 25, 2021. [Online]. Available: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine

[3] "Pfizer and BioNtech conclude Phase 3 study of COVID-19 vaccine candidate, meeting all primary efficacy endpoints," *Pfizer*, Nov. 18, 2020. Accessed: Jan. 25, 2021. [Online]. Available: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-conclude-phase-3-study-covid-19-vaccine

[4] J. Buckner, G. Chowell, and M. Springborn, "Optimal Dynamic Prioritization of Scarce COVID-19 Vaccines," *Medrxiv*, Sep. 2020, doi: 10.1101/2020.09.22.20199174.

[5] "How CDC Is Making COVID-19 Vaccine Recommendations," *CDC*, Dec. 30, 2020. Accessed: Jan. 25, 2021. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations-process.html

[6] L. Roope, P. Clarke, and R. Duch, "Who should get the coronavirus vaccine first? France and the UK have different answers," *The Conservation*, Nov. 16, 2020. Accessed: Jan. 25, 2021. [Online]. Available: https://theconversation.com/who-should-get-the-coronavirus-vaccine-first-france-and-the-uk-have-different-answers-149875

[7] H. Tatapudi, R. Das, and T. Das, "Impact assessment of full and partial stay-at-home orders, face mask usage, and contact tracing: An agent-based simulation study of COVID-19 for an urban region," *Glob. Epidemiol.*, vol. 2, no. 100036, Nov. 2020, doi: https://doi.org/10.1016/j.gloepi.2020.100036.

[8] "Evidence to recommendations for COVID-19 vaccines: Evidence framework," *WHO*, Dec. 10, 2020. Accessed: Jan. 25, 2020. [Online]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-SAGE-Framework-Evidence-2020-1

[9] "Framework for equitable allocation of COVID-19 vaccine." The National Academy Press. Accessed: Jan. 25, 2021. [Online]. Available: https://www.nap.edu/read/25917/chapter/1

[10] K. Clements, J. Chancellor, N. Kristin, K. DeLong, and D. Thompson, "Cost-Effectiveness of a Recommendation of Universal Mass Vaccination for Seasonal Influenza in the United States," *Value Health*, vol. 14, no. 6, pp. 800–811, Sep. 2011, doi: https://doi.org/10.1016/j.jval.2011.03.005.

[11] R. Patel, I. Longini, and E. Halloran, "Finding optimal vaccination strategies for pandemic influenza using genetic algorithms," *J. Theor. Biol.*, vol. 234, no. 2, pp. 201–212, May 2005, doi: https://doi.org/10.1016/j.jtbi.2004.11.032.

[12] I. Longini and E. Halloran, "Strategy for Distribution of Influenza Vaccine to High-Risk Groups and Children," *Am. J. Epidemiol.*, vol. 161, no. 4, pp. 303–306, Feb. 2005, doi: https://doi.org/10.1093/aje/kwi053.

[13] D. Knipl and G. Röst, "Modelling the strategies for age specific vaccination scheduling during influenza pandemic outbreaks," *Math Biosci Eng*, vol. 8, no. 11, pp. 123–139, Jan. 2011, doi: 10.3934/mbe.2011.8.123.

[14] G. Barmparis and G. Tsironis, "Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach," *135*, vol. 109842, May 2020, doi: 10.1016/j.[s.2020.109842.

[15] D. Candido, I. Claro, J. Jesus, W. Marciel de Souza, F. Moreira, and S. Dellicour, "Evolution and epidemic spread of SARS-CoV-2 in Brazil," *Science*, vol. 369, no. 6508, pp. 1255–1260, Sep. 2020, doi: 10.1126/science.abd2161.

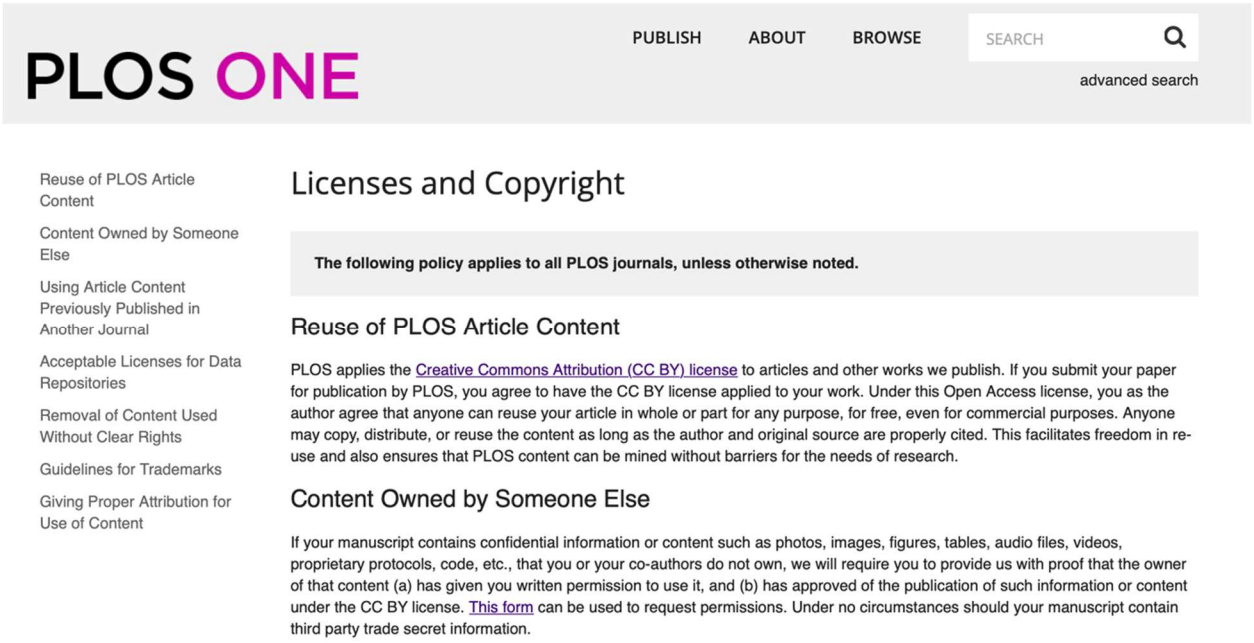
[16] N. Chintalapudi, G. Battineni, and F. Amentaa, "COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach," *J Microbiol Immunol Infect*, vol. 53, no. 3, pp. 396–403, Jun. 2020, doi: 10.1016/j.jmii.2020.04.004.

[17] Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the COVID□19 outbreak and effectiveness of government interventions: A data□driven analysis," *J Med Virol*, vol. 92, no. 6, pp. 645–659, doi: 10.1002/jmv.25750.

[18] S. Yang, P. Cao, P. Du, Z. Wu, Z. Zhang, and L. Yang, "Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis," *Ann Transl Med*, vol. 8, no. 4, p. 128, Feb. 2020, doi: 10.21037/atm.2020.02.66.

[19] A. Aleta, D. Martin-Corral, A. Piontti, M. AJelli, M. Litvinova, and M. Chinazzi, "Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic," *Medrxiv*, May 2020, doi: https://doi.org/10.1101/2020.05.06.20092841.
[20] C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, and S. He, "The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID□19): A well□mixed SEIR model analysis," *J Med Virol*, vol. 92, no. 7, pp. 841–848, Jul. 2020, doi: https://doi.org/10.1002/jmv.25827.
[21] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling," *MedRxiv*, Feb. 2020, doi: https://doi.org/10.1101/2020.02.16.20023465.
[22] Z. Yang, Z. Zeng, K. Wang, S. Wong, W. Liang, and M. Zanin, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *J Thorac Dis*, vol. 12, no. 3, pp. 165–174, Mar. 2020, doi: 10.21037/jtd.2020.02.64.
[23] Y. Zhang, B. Jiang, J. Yuan, and Y. Tao, "The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study," *MedRxiv*, Mar. 2020, doi: https://doi.org/10.1101/2020.03.04.20031187.
[24] D. Chao, E. Halloran, V. Obenchain, and I. Longini, "FluTE, a publicly available stochastic influenza epidemic simulation model," *PLoS Comput Biol*, vol. 6, no. 1, p. e1000656, Jan. 2010, doi: 10.1371/journal.pcbi.1000656.
[25] H. Tatapudi and T. K. Das, "Impact of school reopening on pandemic spread: A case study using an agent-based model for COVID-19," *Infect. Dis. Model.*, vol. 6, pp. 839–847, 2021, doi: 10.1016/j.idm.2021.06.007.
[26] D. K. Chu *et al.*, "Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis," *The Lancet*, vol. 395, no. 10242, pp. 1973–1987, Jun. 2020, doi: 10.1016/S0140-6736(20)31142-9.
[27] C. Vazquez, "Coronavirus in Miami-Dade: Contact tracing failures and talk of how to spend federal money," *Local10*, Jul. 27, 2020. https://www.local10.com/news/local/2020/07/27/coronavirus-in-miami-dade-contact-tracing-failures-and-talk-of-how-to-spend-federal-money/ (accessed Jul. 30, 2021).
[28] "Contact Tracing Is Failing in Many States. Here's Why," *NY Times*, Jul. 31, 2020. https://www.nytimes.com/2020/07/31/health/covid-contact-tracing-tests.html (accessed Jul. 30, 2021).
[29] Miami-Dade County Public Schools, "A guide to the reopening of Miami-Dade County public schools," *Reopen smart return safe*, Aug. 2020. http://pdfs.dadeschools.net/reopening/Reopen%20SMART%20Return%20SAFE%20-%20A%20Guide%20to%20the%20Reopening%20of%20Miami-Dade%20County%20Public%20Schools.pdf (accessed Aug. 03, 2021).
[30] M. Beusokem, "Survey: COVID vaccine willingness waned since April," *Center for Infectious Disease Research and Policy*, Dec. 30, 2020. Accessed: Jan. 28, 2021. [Online]. Available: https://www.cidrap.umn.edu/news-perspective/2020/12/survey-covid-vaccine-willingness-waned-april
[31] L. Saad, "U.S. Readiness to Get COVID-19 Vaccine Steadies at 65%," *Gallup News*, Jan. 12, 2021. Accessed: Jan. 28, 2021. [Online]. Available: https://news.gallup.com/poll/328415/readiness-covid-vaccine-steadies.aspx
[32] CDC, "COVID-19 Pandemic Planning Scenarios," *Centers for Disease Control and Prevention*. https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-archive/planning-scenarios-2020-05-20.pdf (accessed May 05, 2021).
[33] N. M. Ferguson *et al.*, "Strategies for containing an emerging influenza pandemic in Southeast Asia," *Nature*, vol. 437, no. 7056, pp. 209–214, Sep. 2005, doi: 10.1038/nature04017.
[34] N. Ferguson, D. Cummings, C. Fraser, J. Cajka, P. Cooley, and D. Burke, "Estimating disease burden of a potential A(H7N9) pandemic influenza outbreak in the United States," *Nature*, vol. 442, no. 7101, pp. 448–452, Apr. 2006, doi: https://doi.org/10.1038/nature04795.
[35] "Florida's COVID-19 Data and Surveillance Dashboard," *Florida Department of Health, Division of Disease Control and Health Protection*.
[36] "Evidence summary for COVID-19 viral load over course of infection," *Health Information and Quality Authority*, 2020. https://www.hiqa.ie/sites/default/files/2020-04/Evidence-Summary_COVID-19_duration-of-infectivity-viral-load_0.pdf (accessed Jul. 30, 2021).

Appendix F: Copyrights for Published Materials in PLOS ONE



PLOS ONE

PUBLISH ABOUT BROWSE SEARCH

advanced search

Reuse of PLOS Article Content

Content Owned by Someone Else

Using Article Content Previously Published in Another Journal

Acceptable Licenses for Data Repositories

Removal of Content Used Without Clear Rights

Guidelines for Trademarks

Giving Proper Attribution for Use of Content

Licenses and Copyright

The following policy applies to all PLOS journals, unless otherwise noted.

Reuse of PLOS Article Content

PLOS applies the Creative Commons Attribution (CC BY) license to articles and other works we publish. If you submit your paper for publication by PLOS, you agree to have the CC BY license applied to your work. Under this Open Access license, you as the author agree that anyone can reuse your article in whole or part for any purpose, for free, even for commercial purposes. Anyone may copy, distribute, or reuse the content as long as the author and original source are properly cited. This facilitates freedom in re-use and also ensures that PLOS content can be mined without barriers for the needs of research.

Content Owned by Someone Else

If your manuscript contains confidential information or content such as photos, images, figures, tables, audio files, videos, proprietary protocols, code, etc., that you or your co-authors do not own, we will require you to provide us with proof that the owner of that content (a) has given you written permission to use it, and (b) has approved of the publication of such information or content under the CC BY license. This form can be used to request permissions. Under no circumstances should your manuscript contain third party trade secret information.

Appendix G: Published Materials in PLOS ONE Journal

PLOS ONE

RESEARCH ARTICLE

Threshold analyses on combinations of testing, population size, and vaccine coverage for COVID-19 control in a university setting

Xinmeng Zhao[1☯], Hanisha Tatapudi[1☯], George Corey[2], Chaitra Gopalappa[1]*

1 Department of Mechanical and Industrial Engineering, University of Massachusetts Amherst, Amherst, MA, United States of America, **2** University Health Services, University of Massachusetts Amherst, Amherst, MA, United States of America

☯ These authors contributed equally to this work.
* chaitrag@umass.edu

OPEN ACCESS

Citation: Zhao X, Tatapudi H, Corey G, Gopalappa C (2021) Threshold analyses on combinations of testing, population size, and vaccine coverage for COVID-19 control in a university setting. PLoS ONE 16(8): e0255864. https://doi.org/10.1371/journal.pone.0255864

Editor: Ceyhun Eksin, Texas A&M University College Station, UNITED STATES

Received: July 22, 2020

Accepted: July 26, 2021

Published: August 9, 2021

Data Availability Statement: All relevant data are within the paper and its S1 Text and S1–S6 Tables.

Funding: Gopalappa and Zhao were partly funded from NSF 1915481.

Competing interests: The authors have declared that no competing interests exist.

Abstract

We simulated epidemic projections of a potential COVID-19 outbreak in a residential university population in the United States under varying combinations of asymptomatic tests (5% to 33% per day), transmission rates (2.5% to 14%), and contact rates (1 to 25), to identify the contact rate threshold that, if exceeded, would lead to exponential growth in infections. Using this, we extracted contact rate thresholds among non-essential workers, population size thresholds in the absence of vaccines, and vaccine coverage thresholds. We further stream-lined our analyses to transmission rates of 5 to 8%, to correspond to the reported levels of face-mask-use/physical-distancing during the 2020 pandemic. Our results suggest that, in the absence of vaccines, testing alone without reducing population size would not be sufficient to control an outbreak. If the population size is lowered to 34% (or 44%) of the actual population size to maintain contact rates at 4 (or 7) among non-essential workers, mass tests at 25% (or 33%) per day would help control an outbreak. With the availability of vaccines, the campus can be kept at full population provided at least 95% are vaccinated. If vaccines are partially available such that the coverage is lower than 95%, keeping at full population would require asymptomatic testing, either mass tests at 25% per day if vaccine coverage is at 63–79%, or mass tests at 33% per day if vaccine coverage is at 53–68%. If vaccine coverage is below 53%, to control an outbreak, in addition to mass tests at 33% per day, it would also require lowering the population size to 90%, 75%, and 60%, if vaccine coverage is at 38–53%, 23–38%, and below 23%, respectively. Threshold estimates from this study, interpolated over the range of transmission rates, can collectively help inform campus level preparedness plans for adoption of face mask/physical-distancing, testing, remote instructions, and personnel scheduling, during non-availability or partial-availability of vaccines, in the event of SARS-Cov2-type disease outbreaks.

Introduction

The COVID-19 pandemic caused by the SARS-CoV-2 virus has caused significant disease and economic burdens since its first outbreak in December 2019. Because of the absence of an effective vaccine, as of June 2020 at the time of this study and since March 2020, the main intervention for the prevention of COVID-19 transmissions had been to reduce contacts between people through lockdowns of non-essential organizations and services [1]. However, lockdowns are a threat to the economic stability of a nation as seen by the unprecedented rise in unemployment rates [2, 3]. Therefore, while lockdowns are a good short-term strategy, for a long-term strategy, or until a vaccine becomes widely available, it has become necessary to identify alternate strategies and lifestyles that control the disease burden while minimizing the economic burden. Interventions that are effective include the use of face masks, physical distancing between persons at a recommended 6ft, and contact tracing and testing or mass testing to enable early diagnosis in the asymptomatic stage of infection [4]. However, removal of lockdowns should be strictly accompanied by a reopening plan that rapidly and efficiently enables the adoption of the above interventions to avoid an epidemic rebound. In addition to public health agencies, all members of a community, in both public and private sectors, play a key role in the development and implementation of a reopening plan that is most suited for their organization [5]. Among these sectors, universities and colleges bear a special burden to develop a reopening plan that include changes to a range of activities related to teaching, research, dining, housing, and extra-curricular activities.

We developed a compartmental differential equations model to simulate epidemic projections of a potential COVID-19 outbreak in a population of 38,000 individuals, which is representative of undergraduate and graduate students, faculty, and staff in a typical residential university in the United States. We simulated epidemic projections of potential outbreaks under varying combinations of contact tracing and testing, and mass testing, to identify combinations that would reduce the effective reproduction numberR_e to a value below the epidemic threshold of 1. R_e is directly proportional to the duration of infectiousness, transmission rate (the probability of transmission per contact per day, representing the infectiousness of the virus), and contact rate (the number of contacts per person per day) [6]. Asymptomatic testing through trace and test or mass tests lead to diagnosis in the asymptomatic phase of the infection, and thus, if persons diagnosed with infection are successfully quarantined, it reduces the duration of exposure [7–9] and thus reduce R_e. Physical distancing by the recommended 3 or 6ft and use of face masks can reduce transmission rate, and thus reduce R_e [10, 11]. Reducing contact rate such as through transitioning to remote work to reduce population density on campus directly reduces R_e. Thus, different types of interventions help reduce each of these components of R_e. Here, we evaluated different combinations of test rates, transmission rates, and contact rates that help reduce R_e to below 1 to identify minimum levels of testing, physical distancing and face mask use, and population density necessary for effective control of an outbreak.

While it is generally known that increasing contact tracing and testing is necessary, studies evaluating testing at an organizational level, such as university, were only recently emerging at the time of this study in June 2020. One study that analyzed contact tracing in the general populations estimated that reducing R_0 of 1.5 to an R_e of 1 requires more than 20% of contacts traced, reducing R_0 of 2.5 to an R_e of 1 requires more than 80% of contacts traced, and reducing R_0 of 3.5 to an R_e of 1 requires more than 100% of contacts traced [12]. A modeling study applied to the Boston area [13] estimated that the best way-out scenario is a Lift and Enhanced Testing (LET) with 50% detection and 40% of contacts traced. According to this, the number of individuals that need to be traced per 1000 persons is below 0.1 under partial reopening and

below 0.15 under total reopening. Models for a university were only recently emerging at the time of this study in June 2020 [14–17], but typically, most studies combine transmission rate and contact rate as one metric in the evaluation of testing.

In this study, instead of using a product of transmission rate and contact rate as one metric as typically done, we evaluated these separately, due to the following reasons. First, it helps systematically evaluate different interventions considering that different types of interventions help reduce each of the three components of R_e, testing reduces duration of exposed infectious stage, transitioning to remote classes reduce contact rate, vaccinations reduce the number of contacts who are potential disease carriers, and face mask use and 6ft distancing reduces transmission rate. Second, while adoption of each of these decisions are made at an organizational level, adherence and feasibility of face mask and 6ft distancing are highly influenced by individual behaviors and thus have a larger range of uncertainty. Third, while physical distancing and use of face masks can reduce transmission rate, the baseline transmission rate and expected reductions could vary based on multiple factors such as indoor vs. outdoor settings and ventilation, proper use and type of face mask [10, 11, 18, 19], mode of transmission [20–23], and viral load in the index person [8, 24]. Fourth, though we specifically focus this study on COVID-19 caused by the original SARS-CoV-2 virus, studying varying levels of transmission rates could help extrapolate findings to new variants or future outbreaks of viral respiratory infections with similar disease progressions [24], especially in the early stages when specific data is lacking but when the same non-pharmaceutical interventions, such as face masks, physical distancing, remote instructions, and testing, are suitable options.

To systematically inform these analyses, we first evaluated different combinations of trace and test rate, mass test rate, and transmission rate for a range of contact rates, to identify the threshold contact rates that maintain infection cases below certain set levels of tolerance. We then used the contact rate thresholds to identify the population size thresholds, i.e., the maximum population size on campus, which could help inform decisions related to campus activities such as the fraction of classes to transition to remote. We also used the contact rate thresholds to identify the vaccine coverage thresholds for a post-vaccine era, i.e., the vaccine coverage necessary for a campus to return to a normal population size. We also identify, under each intervention combination, the number of trace and tests and quarantines. These metrics could collectively help inform development of a preparedness plan for reopening a university during the COVID-19 pandemic or to set protocols in the event of future outbreaks.

Methodology

Simulation methodology

We developed a compartmental model for simulating epidemic projections over time. The epidemiological flow diagram for the compartmental model is depicted in Fig 1A. Each box is an epidemiological state, and each arrow represents a transition from one state to another. Note, each compartment is further split by age and gender, but for clarity of notations, we do not include it in the equations below.

Let π_t = [S, L, E, I, Q_L, Q_E, Q_I, H, R, D] be a vector, with each element representing the number of people in a compartment at time t, specifically,

S = the number of susceptible individuals at time t,

L = the number of exposed, but asymptomatic and not infectious individuals (latent stage; also, the non-infectious phase of incubation stage) at time t,

E = the number of asymptomatic or pre-symptomatically infectious individuals (infectious phase of the incubation stage) at time t,

I = the number of infectious individuals (symptomatic and infectious stage) at time t,

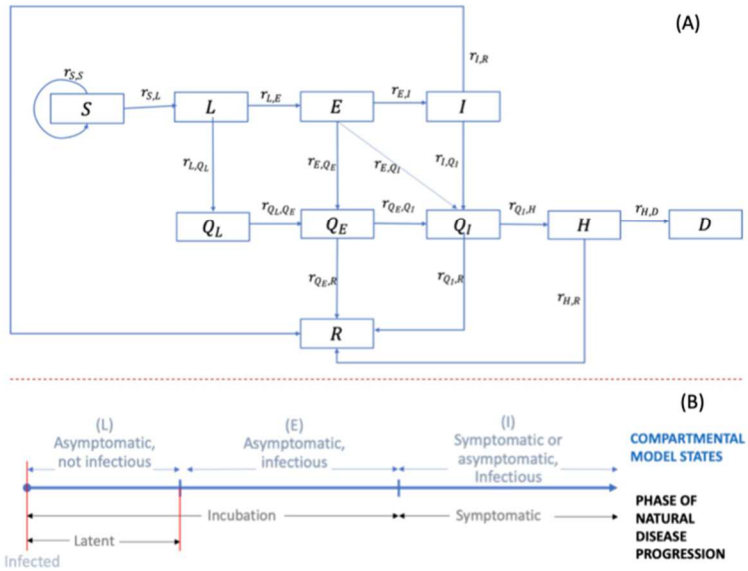


Fig 1. Overview of the extended SEIR compartmental model. (A) Compartmental model flow diagram. (B) Natural disease progression of SARS-COV-2 virus in infected patients. S = susceptible, L = exposed and not infectious (Latent stage) (asymptomatic), E = asymptomatic and infectious, I = symptomatic and infectious, Q_L = exposed and not infectious (Latent) and Quarantined (diagnosed), Q_E = asymptomatic and infectious and Quarantined (diagnosed), Q_I = Infectious and Quarantined (diagnosed), H = Hospitalized, R = Recovered, and D = Deaths.

https://doi.org/10.1371/journal.pone.0255864.g001

Q_L = the number of exposed, asymptomatic and not infectious (latent) and quarantined individuals (diagnosed) at time *t*,

Q_E = the number of asymptomatic or pre-symptomatically infectious and quarantined individuals (diagnosed) at time *t*,

Q_I = the number of infectious and quarantined individuals (diagnosed) at time *t*,

H = the number of hospitalized individuals at time *t*,

R = the number of recovered individuals at time *t*, and

D = the number of deaths at time *t*.

Epidemic states *L*, *E*, and *I* were formulated such that each state represented a distinct phase along the natural disease progression (see Fig 1B), and they collectively included all phases. Over time, persons from *S* can transition to *L*, *E*, and *I*, and upon diagnoses, transition to Q_L, Q_E, or Q_I, and further to *H*, *R*, or *D*, (transitions represented by arrows in Fig 1A) as discussed below.

Let,

p = transmission rate (probability of transmission per contact per day),

c = contact rate (number of contacts per person per day),

N = total population who are alive,

a_B = symptom-based testing rate,

$a_{C,t}$ = rate of testing through contact tracing at time t,

$a_{U,t}$ = rate of testing through mass testing at time t,

ρ = test sensitivity for asymptomatic testing (through mass tests or trace and test),

$days_L$ = duration in latent period,

$days_{incub}$ = duration in incubation period,

$days_{IR}$ = time from onset of symptoms to recovery,
$days_{Q_IR}$ = time from diagnosis to recovery,
$days_{Q_IH}$ = time from diagnosis to hospitalization,
$days_{HR}$ = time from hospitalization to recovery,
$days_{HD}$ = time from hospitalization to death,
$prop_{hosp}$ = proportion hospitalized, and
$prop_{severe}$ = proportion of cases that are severe.

Then, we can write the equations for transition rates (arrows in Fig 1A) as follows:

$r_{S,L} = \frac{pc(E+I)}{N}$, which assumes that only infected persons in E and I can transmit, persons in Q_E and Q_I self-quarantine, and persons in L and Q_L are not infectious.

$$r_{L,E} = \frac{1}{days_L}$$

$r_{E,Q_I} = \frac{(prop_{severe})}{days_{incub} - days_L}$, which assumes that only a proportion of cases that are severe ($prop_{severe}$) get diagnosed immediately because of exhibition of symptoms, we use the proportion hospitalized as a proxy for severe cases; the denominator is based on the assumption that the duration in state E is equal to the difference between the duration of the incubation period and the latent period.

$r_{E,I} = \frac{(1-prop_{severe})}{days_{incub} - days_L}$, which follows from above.

$r_{I,Q_I} = a_B$, which assumes that under symptom-based testing, only persons who show moderate to severe symptoms get diagnosed and those who show mild symptoms do not.

$r_{Q_I,H} = \frac{prop_{hosp}}{days_{Q_IH}}$, for $prop_{hosp}$ we use the proportion of persons hospitalized among those diagnosed through symptom-based testing.

$r_{L,Q_L} = \rho a_{U,t} + (1 - \rho a_{U,t})\rho a_{C,t}$, which assumes that under the implementation of both mass testing and contact tracing and testing, persons diagnosed through mass test will not be tested again on the same day through contact tracing (as our time unit is daily).

$r_{E,Q_E} = \rho a_{U,t} + (1 - \rho a_{U,t})\rho a_{C,t}$, which is similar to above.

$r_{I,R} = (1 - a_B)\left[\rho a_{U,t} + \left(1 - \rho a_{U,t}\right)\rho a_{C,t}\right] + \frac{1}{days_{IR}}$, which assumes that persons with mild cases that did not get diagnosed through symptom-based testing have a chance of getting tested through additional testing options, and self-quarantine upon diagnosis. Note that we did not separately model asymptomatic cases but incorporated that into the symptom-based testing rate (a_B) by considering that 35% of cases are mild to no symptoms and thus do not have a chance of being diagnosed through symptom-based testing.

$$r_{Q_L,Q_E} = \frac{1}{days_L}$$

$r_{Q_E,Q_I} = \frac{[a_B(1-prop_{severe})+(prop_{severe})]}{days_{incub} - days_L}$, theoretically, r_{Q_E,Q_I} should be the same as $r_{E,I}$, however, as the rate of transitioning from Q_I to H is fixed to fit to the proportion hospitalized under symptom-based tests, if extensive testing is conducted, the number of persons in Q_I would increase, thus, incorrectly inflating the number of persons who are hospitalized; To avoid this, we modified the equation to consider that the number of persons flowing into Q_I would be equal to the proportion flowing from I to Q_I under symptom-based testing.

$r_{Q_E,R} = \frac{(1-[a_B(1-prop_{severe})+(prop_{hosp})])}{days_{incub} - days_L}$, which follows from the above equation.

$$r_{Q_I,R} = \frac{1 - prop_{hosp}}{days_{Q_IR}}$$

$$r_{H,R} = \frac{prop_{recover}}{days_{HR}}$$

$$r_{H,D} = \frac{(1 - prop_{recover})}{days_{HD}}$$

Note: $r_{s,s}$ is the testing rate (either through mass test or trace and test). We assumed that susceptible persons go back to the susceptible state after testing, i.e., we did not explicitly track false positives.

The values and ranges for the above epidemic parameters used in the compartmental simulation model are presented in Table A in S1 Text.

We simulate the epidemic over time using the following system of differential equations

$$\pi_{t+1} = \pi_t + \pi_t Q_t.dt$$

where, Q_t = a matrix of transition rates between states (arrows in Fig 1A), and dt = time-step. We use a time-unit of per day for the transition rates in Q_t and set $dt = \frac{1}{10}$, and thus, the model simulates every 10^{th} of a day.

The expansion of the system of differential equations are as follows:

$$S_{t+1} = S_t + (-r_{S,L}S_t)\ dt$$

$$L_{t+1} = L_t + (r_{S,L}S_t - (r_{L,E} + r_{L,Q_L})L_t)\ dt$$

$$E_{t+1} = E_t + (r_{L,E}\ L_t - (r_{E,I} + r_{E,Q_E} + r_{E,Q_I} + r_{E,R})E_t)\ dt$$

$$I_{t+1} = I_t + (r_{E,I}E_t - (r_{I,Q_I} + r_{I,R})I_t)\ dt$$

$$Q_{L,t+1} = Q_{L,t} + (r_{L,Q_L}\ L_t - (r_{Q_L,Q_E})Q_{L,t})\ dt$$

$$Q_{E,t+1} = Q_{E,t} + (r_{E,Q_E}\ E_t - (r_{Q_E,Q_I} + r_{Q_E,R})Q_{E,t})\ dt$$

$$Q_{I,t+1} = Q_{I,t} + ((r_{E,Q_I}\ E_t + r_{Q_E,Q_I}\ Q_{E,t} + r_{I,Q_I}I_t) - (r_{Q_I,H} + r_{Q_I,R})Q_{I,t})\ dt$$

$$H_{t+1} = H_t + (r_{Q_I,H}Q_{I,t} - (r_{H,R} + r_{H,D})H_t)\ dt$$

$$\mathrm{R}_{t+1} = R_t + (r_{Q_E,R}Q_{E,t} + r_{Q_I,R}Q_{I,t} + r_{I,R}I_t + r_{H,R}H_t)\ dt$$

$$D_{t+1} = D_t + r_{H,D}H_t dt$$

We can further expand by substitution of the rate terms with their equations as follows:

$$S_{t+1} = S_t + \left(-\frac{pc(E+I)}{N}\right)S_t dt$$

$$L_{t+1} = L_t + \left(\frac{pc(E+I)}{N}S_t - \left(\frac{1}{days_L} + \rho a_{U,t} + (1 - \rho a_{U,t})\rho a_{C,t}\right)L_t\right)\ dt$$

$$E_{t+1} = E_t + \left(\frac{1}{days_L}L_t - \left(\frac{(1-prop_{severe})}{days_{incub}-days_L} + (\rho a_{U,t} + (1-\rho a_{U,t})\rho a_{C,t}) + \frac{(prop_{severe})}{days_{incub}-days_L}\right)E_t\right)dt$$

$$I_{t+1} = I_t + \left(\frac{(1-prop_{severe})}{days_{incub}-days_L}E_t - \left(a_B + (\rho a_{U,t} + (1-\rho a_{U,t})\rho a_{C,t} + \frac{1}{days_{IR}})\right)I_t\right)dt$$

$$Q_{L,t+1} = Q_{L,t} + \left((\rho a_{U,t} + (1-\rho a_{U,t})\rho a_{C,t})L_t - \left(\frac{1}{days_L}\right)Q_{L,t}\right)\,dt$$

$$Q_{E,t+1} = Q_{E,t} + \left((\rho a_{U,t} + (1-\rho a_{U,t})\rho a_{C,t})E_t - \left(\frac{[a_B(1-prop_{severe}) + (prop_{severe})]}{days_{incub}-days_L} + \frac{1-prop_{hosp}}{days_{Q_IR}}\right)Q_{E,t}\right)\,dt$$

$$Q_{I,t+1} = Q_{I,t} + \left(\frac{(prop_{severe})}{days_{incub}-days_L}E_t + \frac{[a_B(1-prop_{severe}) + (prop_{severe})]}{days_{incub}-days_L}Q_{E,t} + a_B I_t - \left(\frac{prop_{hosp}}{days_{Q_IH}} + \frac{1-prop_{hosp}}{days_{Q_IR}}\right)Q_{I,t}\right)dt$$

$$H_{t+1} = H_t + \left(\frac{prop_{hosp}}{days_{Q_IH}}Q_{I,t} - \left(\frac{prop_{recover}}{days_{HR}} + \frac{(1-prop_{recover})}{days_{HD}}\right)H_t\right)\,dt$$

$$R_{t+1} = R_t + \left(\frac{(1-[a_B(1-prop_{severe}) + (prop_{hosp})])}{days_{incub}-days_L}Q_{E,t} + \frac{1-prop_{hosp}}{days_{Q_IR}}Q_{I,t} + (\rho a_{U,t} + (1-\rho a_{U,t})\rho a_{C,t} + \frac{1}{days_{IR}}\right)I_t + \frac{prop_{recover}}{days_{HR}}H_t)\,dt$$

$$D_{t+1} = D_t + \frac{(1-prop_{recover})}{days_{HD}}H_t dt$$

Input data assumptions and sources for simulation model

For the rates of natural disease progression, we used estimates from other studies in the literature. The description of the data, sources, and values (with ranges and medians where applicable) for all parameters are available in the Table A in S1 Text. Briefly, we assumed an incubation period duration of 5.4 days [25], the first 2.5 days in stage L (not infectious and asymptomatic) [26], and the remaining 2.9 days in stage E (infectious and asymptomatic). We assumed about 65% of cases develop medium to severe symptoms [27] and, in the absence of test and trace or mass test, can be diagnosed through symptom-based testing. We assumed the remaining 35% of cases show mild to no symptoms and can be diagnosed only through trace and test, or universal mass test. We assumed an average duration of 3.5 days from the time of onset of symptoms to hospitalization [28], with the proportion hospitalized varying as a function of age. For mild cases, we assumed an average duration of 7 days from the time of onset of

symptoms to recovery [28]. We assumed case fatality rates vary as a function of age and gender.

Interventions

Mass test and trace and test. We evaluated the following scenarios: mass test only, trace and test only, delayed trace and test only, combination mass test and trace and test, and combination mass test and delayed trace and test, each at different rates of testing, as follows. We evaluated mass testing at rates of 5% 10%, 20%, 25%, and 33% of the population per day, which is equivalent to testing once every 20 days (5% per day over every 20-day period), 10 days, 5 days, 4 days, and 3 days (33% per day every 3-day period), respectively. We modeled the rate of trace and test as the inverse of the time from infection to effective isolation of a contact, i.e., the sum of the number of days passed since contact with an individual (as reported by the index diagnosed person) and the number of days into the future to find, test, and isolate the infected contact. We chose this definition as each component in this duration can vary significantly for every diagnosed person and for each of their contacts. In the case the contact is never found, the duration would be the full duration of infection. Thus, this definition of trace and test can be comparable to data that is typically collected. Specifically, the trace and test rate here should be compared to the average of the inverse of the time from reported contact to either effective isolation of that contact or maximum infection duration (whichever is the least value), averaged over all contacts. We evaluated trace and test rates at levels of 10%, 17%, 20%, 25%, 33%, and 50%, equivalent of 10 days, 6 days, 5 days, 4 days, 3 days, and 2 days, respectively, from the time of transmission of infection to effective isolation of that contact. We evaluated combinations of mass test and trace and test, by varying mass tests between 5% and 33% per day and keeping trace and test at 50% as this higher rate of trace and test maybe more feasible with mass test than symptom-based test only. We assumed trace and test would initiate within the first 5 cases of diagnosis. Considering there may be delays in setting up a trace and test system (such as in events of new outbreaks in the future or failure to respond quickly), to tests its sensitivity, we evaluated scenarios by delaying the initiating of trace and test to after diagnosis of 20 cases. Thus, the scenarios referred to as 'trace and test only' and 'combination tests' refers to initiation of trace and test after first 5 cases of diagnosis. And the scenarios referred to as 'delayed trace and test only' and 'delayed combination tests' refers to delaying initiation of trace and test to after diagnoses of 20 cases. In all scenarios, we applied baseline symptom-based testing, assumed test results are available within 24 hours, and persons testing positive self-quarantine for 14-days. For diagnosis in asymptomatic stages, i.e., diagnosis through trace and test or mass test, we assumed a test sensitivity of 0.9 [29].

Non-pharmaceutical interventions. We evaluated transmission rates (p) of 14% (baseline), 8% (mid), 5.4% (lower-mid), and 2.5% (lowest). The baseline value of p corresponds to an average estimate under no interventions (no physical distancing and no face masks) [11, 30]. A transmission rate of 8% corresponds to the expected rates with the use of face masks in a non-health care setting [11]. Transmission rates of 5.4% and 2.5% correspond to expected rates under 3ft and 6ft physical distancing, respectively [11] (see Table B in S1 Text). We evaluated contact rates between 1 and 25 (c), we did not differentiate between on-campus and off-campus contact rates.

Application to a university setting

Demographic data. We used the Fall 2018 student enrollment data from the University of Massachusetts—Amherst, Amherst, MA, to determine the population size of undergraduate and graduate students and their age and gender distributions [31]. For faculty and staff, we

used the age distribution of persons 25 years and older from the Town of Amherst, MA, where the university is located [32]. To initiate an outbreak, we assumed 4 to 5 infected cases on Day 1, estimated as follows. We assumed that the proportion of incoming students who are infected would be equal to the prevalence of COVID-19 in Massachusetts (MA) in June 2020. We also assumed that all incoming students would be tested, and about 10% of infected cases would be false negatives. Prevalence is unknown, as not all cases are diagnosed and diagnosed cases are not specifically tracked. Therefore, to estimate prevalence of COVID-19 in MA, we used the simulation model to determine the ratio of new diagnosis to persons with infection and applied that ratio to the number of new diagnoses on June 26th in MA. This resulted in about 4 infected cases on day 1 remaining undetected, thus initiating an outbreak. We also assumed that at the beginning of every week, there would be about 3 to 4 infections from outside, calculated by assuming that about 10% of the population are likely to mix with the population outside the university or travel out of Amherst during weekends and are not tested upon return. Based on the above, we initialized the model on Day 1 with 4 infected persons in the Latent stage and added 3 to 4 outside cases to the Latent stage at the beginning of every week. We simulated the model for a 90-day period to represent the duration of the expected Fall 2020 semester.

Tolerance on the number of infected cases for identifying contact rate thresholds

We evaluated contact rate thresholds under three levels of epidemic tolerance: relaxed tolerance, medium tolerance, and tight tolerance. Relaxed tolerance marked the point beyond which there was an exponential growth in infections, the maximum number of infections under this tolerance level was about 170. For medium tolerance, we set the number of infections to less than 77, and for tight tolerance, we set the number of infections to less than 50. The latter two cases correspond to maximum infections for a case fatality rate (CFR) of 2%, which is the reported CFR in the general population for the United States [33]. That is, 1/0.02 gives the 50 cases threshold and 77 is obtained by further dividing that by 65%, which is the proportion of cases with medium to severe symptoms [27], to account for the remaining 35% of cases with mild to no symptoms that were likely unreported and thus not included in the CFR calculation. As the CFR for COVID-19 is much lower in university student aged populations, the use of the alternative tolerance on the number of infections helps avoid spill-over effect of a breakout into the community. Also note that, because of our assumptions for the number of initial cases and cases per week entering the population from outside, the minimum number of cases over the 90-day period would be 45. Therefore, the tolerance of 50 cases is a very tight tolerance. For context, one of the indicators CDC uses to categorize community transmission risk is the number of cases per 100,000 persons during the last 7 days, categorizing as low, moderate, and substantial to high if there were less than 10, 10–49, and greater than 49 cases, respectively [34]. Converting our tolerance levels to the CDC indicator would translate to 35, 15, and 10 cases for relaxed, medium, and tight tolerances, respectively. If we exclude the 45 cases from outside, it translates to 25, 6, and 1 cases for relaxed, medium, and tight tolerances, respectively.

Population behavioral data

While there was limited data on contact rates specific to university students at the time of this original study in June 2020, studies conducted since then have generated some (though limited) data on population behaviors. These data include contact rates and behaviors related to use of non-pharmaceutical interventions such as face mask and 6 ft physical distancing, mostly

either self-reported in surveys or estimations made in other modeling studies informed through university settings. We briefly summarize the data from each study in the Table C in S1 Text. Some of the surveys were specific to university students in the United States while others were either university students in other countries or general populations. Studies on surveys of university students, when partial shutdowns were enforced and universities resorted to varying levels of remote classes, reported 6 to 8 contacts per person per day [35, 36]. However, students who self-reported as providing essential services or caring for non-household members (~23%) reported an average contact rate of about 14 per person per day [37]. Modeling studies that estimated contacts among university students for a scenario prior to the pandemic assume contact rates of 16 to 24 per person per day [38–40]. Using data on face mask use and physical distancing, specifically originating from three surveys of student and general population in the United States [38, 42, 43] and the transmission rates corresponding to these interventions (summarized in Table C in S1 Text), we calculate the expected transmission rate to be between 5% and 8%. We use these estimates to further streamline our analyses.

Interpretation of contact rate thresholds: Size of social circle, population size, and vaccine coverage

We utilize the contact rate thresholds, under the different levels of testing and transmission rate (face mask use and physical distancing), to identify four additional metrics that would help inform campus decisions: first, the contact rate threshold among non-essential workers after accounting for the higher contact rate among essential workers, which would help inform the size of social circles at the individual level and schedule campus activities; second, the threshold values for population size on campus as a proportion of the actual population size, which would help decisions related to the fraction of remote vs. face-to-face classes, on-campus housing, and other campus activities for the era of pre-vaccine availability; third, for the era of post-vaccine availability, the threshold values for vaccination coverage for the university to return to normal (i.e., 100% population size); fourth, the threshold values for population sizes under varying levels of vaccine coverage, which would help decisions related to campus activities in the event that vaccines are only partially available that coverage is not at levels sufficient to fully return to normal. All four metrics would be used alongside decisions related to the level of testing.

The metrics were estimated as follows. Suppose $\hat{C}$ is the contact rate threshold, we estimated the first metric as $\hat{C}_n = (\hat{C} - C_e p_e)/(1 - p_e)$, where $\hat{C}_n$ is the contact rate threshold among non-essential workers (we limit $0 \leq \hat{C}_n \leq \hat{C}$), C_e is the contact rate among essential workers (we assume $C_e = 14$ [37]), and p_e is the proportion of the population who are essential workers (we assume $p_e = 23\%$ [37]).

The interpretation of the second, third, and fourth metrics arise from our simplifying assumption that contact rates are directly proportional to the population density [41], $C = c_0\rho$; $\rho = N/A$; where C is the actual contact rate (under regular face-to-face instructions), ρ is the density, N is the population size, A is the campus area, and c_0 is a constant. Further, we assume that university campuses maintain similar levels of population density under regular work conditions, i.e., though the population sizes may vary across universities, the campus area also changes proportionally so that the population density is similar, and thus, the contact rates under regular work conditions are also similar. Multiple studies reported similar contact rates of 16 to 24 under regular working conditions supporting this assumption [42, 43]. Thus, if our estimated contact rate thresholds (say $\hat{C}$) are lower than the actual contact rates of 16 to 24 (C), given fixed area (A), achieving $\hat{C}$ would require reducing the population size (N) proportional

to the reduction in contact rate, i.e., $\hat{C}/C = \hat{N}/N (= say\ \hat{p})$, implying that the population size on campus should be at a maximum of $\hat{p}$% of its original population size ($\hat{p} = \hat{C}/C$).

The third metric on interpretation of threshold for vaccination coverage (say $\hat{V}$) follows from the above assumptions. Achieving a contact rate threshold of $\hat{C}$ when universities are back to regular face-to-face classes, i.e., $\hat{N} = N$ or $\hat{p} = 1$, would require that $(1 - \hat{C}/C)$ proportion of the population be effectively vaccinated. More precisely, vaccine coverage should be at least $\hat{V} = (1 - (\hat{C}/C))/v_e$, where v_e is vaccine efficacy and corresponds to the chance that a vaccinated individual is fully protected from being infected, and thus, is not a potential disease carrying contact. Intuitively, this is saying that though the actual contact rate is C, because $\hat{V}\%$ are vaccinated and protected from infection or transmitting, the effective contact rate is $\hat{C}$. This implies that a threshold contact rate of $\hat{C}$ can be achieved, while maintaining $\hat{p} = 1$, if $\hat{V}\%$ are vaccinated. The vaccine coverage results presented here were estimated by assuming a vaccine efficacy of 95%, and thus, in the event that this changes, the vaccine coverage results should be adjusted by multiplying with 95% and dividing by the new value.

Following from above, the fourth metric considers the fact that if the actual vaccine coverage (say V) is less than $\hat{V}$, achieving the contact rate threshold ($\hat{C}$) would also require some reduction in N. Specifically, the population size on campus should be at a maximum of $\hat{p}$% of its original population size, with $\hat{p} = 1 - ((1 - V)C - \hat{C})/C$, derived as follows. We can write $(1 - V)C - \hat{C}$ as the number of excess contacts, i.e., the number to reduce after accounting for the proportion vaccinated ($(1–V)C$), the proportion of contacts to reduce would then be $((1 - V)C - \hat{C})/C$, and finally, applying the same assumptions as in the second metric would give the equation for $\hat{p}$. If the vaccination coverage is zero, i.e., $V = 0$, we would get back $\hat{p} = \hat{C}/C$. If $V = 1$, then $\hat{p} = 1 + \hat{C}/C$, which implies that even if $\hat{C} = 0$, the campus can fully open. We bound $0 \leq \hat{p} \leq 1$, such that, even if $\hat{C} > 0$ we interpret this as fully back to normal population size (though it would mathematically imply that the campus can handle a higher density from an epidemic perspective, e.g., influx from outside).

Thus, to keep within the infection tolerance levels, $\hat{C}$ would mark the maximum average contact rate over the full population, $\hat{C}_n$ the maximum average contact rate for non-essential workers after accounting for the higher contact rate among essential workers, $\hat{p}$ the maximum proportion of the population who should return back to campus (either when $V = 0$ or $0 < V \leq \hat{V}$), and $\hat{V}$ the minimum vaccine coverage to fully return back to normal ($\hat{p} = 1$). As the above method of estimation of thresholds incorporate the effectiveness of vaccinations, we can interpret that the interventions, such a testing and use of facemask and social distancing, would be applied to only the unvaccinated persons.

Identifying feasible intervention combinations

We identify three sets of feasible combination results. For the event that vaccines are unavailable, we identify the feasible combinations of testing, contact rate for non-essential workers ($\hat{C}_n$), and population size on campus ($\hat{p}$) that can effectively control an outbreak to below the tolerance levels. We define feasible combinations as those with $\hat{C}_n > 2$ in the transmission rate range of 5% to 8%, which would correspond to the reported use of face mask and physical distancing among the university population (see 'Population behavioral data' above). For the event that vaccines are partially or fully available, we identify the minimum vaccine coverage threshold ($\hat{V}$) for the campus to fully return back to normal ($\hat{p} = 1$), and if the vaccine coverage is below this threshold, the reductions in population size ($\hat{p}$) necessary to control the

epidemic to within the tolerance levels. We also identify suitable testing scenarios for reported levels of face mask use and physical distancing (transmission rate of 5% to 8%), and reported levels of contact rates under regular face-to-face classes (16 to 24 per day) and remote classes (6 to 8 per day). We define suitable as those that avoid exponential growth in cases over the duration of a semester. For the above three sets of combination scenarios, we also present results under the full range of transmission rates in the S1–S3 Tables, which could be useful in the event of change in transmission rates such as emergence of new virus variants.

Results

When vaccines are unavailable ($V = 0\%$), there is no single intervention that can effectively control an outbreak. However, there are multiple feasible combinations of testing, contact rate for non-essential workers ($\hat{C}_n$), and population size on campus ($\hat{p}$) that can be implemented to effectively control an outbreak to keep cases below the relaxed to medium tolerance levels, though none to keep cases below the tight tolerance level (Table 1). Examples of feasible combinations under the relaxed tolerance level include: mass tests only at 25% per day, contact rate for non-essential workers at 2 to 6 per day, and campus population size at 26% to 42%; or trace and test only at 33%, contact rate for non-essential workers at 4 to 8 per day, and campus population size at 31% to 47% (see full list in Table 1). Under the medium tolerance level, only scenarios with combination tests were feasible, examples include: 5% mass test, 50% trace and test, contact rate for non-essential workers at 2 to 5 per day, and campus population size at 26% to 36%; or 33% mass test, 50% trace and test, contact rate for non-essential workers at 8 to 14 per day, and campus population size at 47% to 73% (see full list in Table 1). Note: the range in population size results correspond to mid-points of the range for contact rate (C) of 16 and 24 in Table 1.

The corresponding peak numbers of trace and tests per day (per 10,000 persons) in the above feasible scenarios were at a reasonably manageable level. The relaxed tolerance level had a higher value (14 to 55 per day) than the medium tolerance level (3 to 11 per day) considering the population size on campus were lower in the latter case because of the tighter tolerance (Table 1). The peak number of quarantines per day (per 10,000 persons) for the above feasible scenarios also seem manageable. As with above, the relaxed tolerance level had a higher value (6 to 25 per day) than the medium tolerance level (5 to 6 per day). Combinations of testing, contact rate, and population size for the full range of transmission rates evaluated are presented in S1 Table.

When vaccines become partially or fully available, to keep the population size on campus at 100% ($\hat{p} = 1$), the level of testing necessary to effectively control an outbreak would depend on the vaccine coverage in the population (Table 2). To keep infection cases within the relaxed tolerance level, implementing symptomatic-testing-only will be sufficient if at least 95% ($\hat{V}$) of the population are vaccinated (Table 2). With the addition of mass tests only, 5%, 10%, 20%, 25%, and 33% mass tests per day would be sufficient if at least 89% to 95%, 84% to 89%, 74% to 84%, 63% to 79%, and 53% to 68% ($\hat{V}$) of the population are vaccinated (Table 2), respectively, the range corresponding to transmission rate of 5% to 8%, i.e., the unvaccinated continue to use face masks and maintain physical distancing at current compliance levels.

If vaccine coverage (V) is below 53% (the threshold noted above), it would be necessary to also reduce the population size (Table 2). If vaccine coverage (V) is between 38% and 53%, 23% and 38%, or 8% and 23%, in addition to mass tests at 33% per day, it would be necessary to maintain a population size threshold ($\hat{p}$) of at most 90%, 75%, or 60% on average, respectively, (Table 2) and the unvaccinated continue to use face masks and maintain physical

Table 1. Feasible combinations ⁋ of testing, contact rate, and population size on campus for effective control of a disease outbreak in the absence of a vaccine.

Tolerance	Testing	Contact rate threshold (per day)† for non-essential workers‡	Population size† (if regular contact rate is 16)	Population size† (if regular contact rate is 24)	Peak trace and tests per day (per 10,000 persons)	Peak quarantine per day (per 10,000 persons)
Relaxed tolerance	S+25%U	2–6	31% - 50%	21% - 33%	0–0	6–7
	S+33%U	5–9	44% - 63%	29% - 42%	0–0	7–7
	S+33%T	4–8	38% - 56%	25% - 38%	14–21	18–18
	S+50%T	6–11	50% - 75%	33% - 50%	36–55	22–23
	S+5%U+50%T	10–16	69% - 100%	46% - 67%	33–48	20–20
	S+10%U+50%T	11–18	75% - 100%	50% - 75%	24–35	15–15
	S+20%U+50%T	15–22	94% - 100%	63% - 92%	16–25	13–13
	S+25%U+50%T	16–24	100% - 100%	67% - 100%	14–20	13–12
	S+33%U+50%T	18–25	100% - 100%	75% - 100%	12–15	10–9
	S+5%U+50%dT	4–9	38% - 63%	25% - 42%	36–55	21–25
	S+10%U+50% dT	6–13	50% - 81%	33% - 54%	32–63	21–25
	S+20%U+50% dT	11–18	75% - 100%	50% - 75%	36–54	21–21
	S+25%U+50% dT	13–19	81% - 100%	54% - 79%	28–41	19–18
	S+33%U+50% dT	15–23	94% - 100%	63% - 96%	21–33	16–18
Medium tolerance	S+5%U+50%T	2–5	31% - 44%	21% - 29%	8–9	6–6
	S+10%U+50%T	4–8	38% - 56%	25% - 38%	7–11	6–6
	S+20%U+50%T	5–10	44% - 69%	29% - 46%	4–7	5–5
	S+25%U+50%T	6–11	50% - 75%	33% - 50%	4–6	4–5
	S+33%U+50%T	8–14	56% - 88%	38% - 58%	3–6	5–5
	S+20%U+50% dT	2–5	31% - 44%	21% - 29%	4–4	6–6
	S+25%U+50% dT	4–6	38% - 50%	25% - 33%	4–4	6–5
	S+33%U+50% dT	5–9	44% - 63%	29% - 42%	3–4	6–5
Tight tolerance	No scenarios were feasible ⁋					

Relaxed tolerance: Less than 1 death or 170 cases of infection. This point also marks the point beyond which there was an exponential growth in infections in the simulated runs.

Medium tolerance: Less than 77 cases of infections. Estimated as 1/CFR /%reported cases. We assumed a case fatality rate (CFR) of 2% in the general population in the US [33]; We assumed that 65% of infected cases are reported, which is the proportion showing medium to severe symptoms [27].

Tight tolerance: Less than 50 cases of infection. Estimated as 1/CFR. We assumed a case fatality rate (CFR) of 2% in the general population in the US [33].

S: symptomatic testing, U: Mass test, T: trace and test, dT: delayed trace and test.

⁋ We defined a testing scenario as feasible if estimated contact rate thresholds among non-essential workers were at least 2 when transmission rates were 8% and 5% (corresponding to reported use of face mask and physical distancing [11, 30]). The range of values in the table thus correspond to transmission rate of 8% - 5%.

‡ We assume 23% are essential workers and have a contact rate of 14 per day [37].

† Contact rate threshold (per person per day): the average value for contacts per person per day to keep infections below the tolerance level. These reduced contact rates, from the original rates of 16 to 24 [42, 43], can be achieved through reduction in population size at the noted thresholds.

https://doi.org/10.1371/journal.pone.0255864.t001

distancing at current compliance levels. Note: the population size threshold noted here is the average of the values reported for contact rate (C) of 16 and 24 in Table 2.

Instead of adding mass tests only, addition of trace and tests only to symptom-based testing at the lowest rate of 10% (or highest rate of 50%) will also be sufficient to keep the population

Table 2. Combinations of testing, vaccine coverage, and population size for effective control of a disease outbreak.

Testing	Vaccination coverage	Population size (if regular contact rate is 16)	Population size (if regular contact rate is 24)
S	95% - 100%	100% - 100%	100% - 100%
Mass tests only (% tested per day)			
S+5%U	89% - 95%	100% - 100%	100% - 100%
S+10% U	84% - 89%	100% - 100%	100% - 100%
S+20%U	74% - 84%	100% - 100%	100% - 100%
S+25%U	63% - 79%	100% - 100%	100% - 100%
S+33%U	53% - 68%	100% - 100%	100% - 100%
S+33%U	38% - 53%	100% - 97%	79% - 83%
S+33%U	23% - 38%	85% - 82%	64% - 68%
S+33%U	8% - 23%	70% - 67%	49% - 53%
Trace and tests only			
S+10%T	84% - 95%	100% - 100%	100% - 100%
S+17%T	79% - 89%	100% - 100%	100% - 100%
S+20%T	74% - 84%	100% - 100%	100% - 100%
S+25%T	68% - 84%	100% - 100%	100% - 100%
S+33%T	58% - 74%	100% - 100%	100% - 100%
S+50%T	42% - 63%	100% - 100%	100% - 100%
S+50%T	27% - 42%	100% - 92%	77% - 75%
S+50%T	12% - 27%	87% - 77%	62% - 60%
Trace and tests only (capped at 20%)			
S+10%T	84% - 95%	100% - 100%	100% - 100%
S+17%T	79% - 89%	100% - 100%	100% - 100%
S+20%T	74% - 84%	100% - 100%	100% - 100%
S+20%T	59% - 69%	96% - 94%	84% - 86%
S+20%T	44% - 54%	81% - 79%	69% - 71%
S+20%T	29% - 39%	66% - 64%	54% - 56%
S+20%T	14% - 24%	51% - 49%	39% - 41%

The range of value presented correspond to transmission rate range of 5% - 8%, thus fixing face mask and physical distancing at reported levels.
S: symptomatic testing, U: Mass test, T: trace and test, dT: delayed trace and test.

https://doi.org/10.1371/journal.pone.0255864.t002

size on campus at 100% ($\hat{p} = 1$) if at least 84% to 95% are vaccinated (or 42% to 63% are vaccinated) (Table 2). If vaccine coverage (V) is below 42% it would be necessary to also reduce the population size, keeping it to at most 89% on average if vaccine coverage is between 27% and 42%, and to at most 75% on average if vaccine coverage is between 12% and 27%. Considering that 50% trace and test, equivalent to 2 days from infection to isolation is a very tight timeline, which may be more feasible only with digital tracing, we also evaluated at a maximum of 20% trace and test, equivalent to 5 days from infection to isolation. This level of 20% trace and test only will be sufficient to keep the population size on campus at 100% if vaccine coverage is at least 74% to 84%. If vaccine coverage is below that, it will also require a reduction in population size, e.g., to 75% on average if only 44% to 54% of the population are vaccinated (Table 2). All the above scenarios for trace and tests also correspond to the continued use of face masks and physical distancing at least at current compliance levels (transmission rate of 5% to 8%). The combinations of testing and vaccination coverage under the full range of transmission rates are presented in S2 Table.

The total cases of infections and deaths over a 90-day semester if a fully unvaccinated population is on campus (contact rates of 16 to 24 per person per day as reported for regular face-

Table 3. Suitable testing options ⁋ for effective control of a disease outbreak keeping contact rates at reported levels † ‡.

Testing	Number infected (per 10,000 persons)		Peak trace and tests per day (per 10,000 persons)		Peak quarantine per day (per 10,000 persons)	
Transmission rate (p) —>	5%	8%	5%	8%	5%	8%
S+33%U	23 (20, 27)	44 (32, 67)	0 (0, 0)	0 (0, 0)	4 (3, 5)	7 (5, 13)
S+33%T	30 (25, 36)	60 (42, 89)	14 (9, 17)	22 (14, 28)	12 (9, 14)	24 (18, 33)
S+50%T	23 (21, 26)	34 (29, 41)	20 (12, 24)	29 (19, 36)	10 (9, 13)	18 (14, 22)
S+5%U+50%T	19 (17, 21)	25 (22, 29)	8 (7, 13)	16 (10, 19)	6 (5, 7)	9 (8, 12)
S+10%U+50%T	17 (16, 19)	22 (20, 24)	6 (5, 9)	8 (7, 10)	5 (4, 5)	7 (6, 8)
S+20%U+50%T	16 (15, 16)	19 (18, 21)	4 (3, 4)	4 (4, 5)	4 (3, 4)	5 (4, 5)
S+25%U+50%T	15 (15, 16)	18 (17, 20)	3 (2, 4)	3 (3, 3)	3 (3, 4)	4 (4, 4)
S+33%U+50%T	15 (14, 16)	17 (16, 18)	2 (2, 2)	2 (2, 3)	3 (3, 3)	4 (3, 4)
S+5%U+50%dT	30 (26, 34)	45 (38, 55)	26 (16, 32)	46 (36, 64)	15 (12, 18)	26 (21, 34)
S+10%U+50%dT	24 (22, 27)	34 (29, 40)	11 (9, 18)	24 (15, 32)	10 (8, 12)	17 (14, 21)
S+20%U+50%dT	19 (18, 21)	25 (22, 28)	4 (3, 8)	10 (6, 10)	6 (5, 7)	9 (8, 11)
S+25%U+50%dT	18 (17, 19)	22 (20, 25)	3 (3, 3)	6 (4, 8)	5 (5, 5)	7 (6, 9)
S+33%U+50%dT	17 (16, 17)	19 (18, 22)	2 (2, 2)	3 (2, 6)	4 (4, 4)	6 (5, 7)

S: symptomatic testing, U: Mass test, T: trace and test, dT: delayed trace and test.

†Reported value of contact rate is on average between 6 and 8 per person per day under remote instructions [35, 36]; Using our estimations, this would correspond to population size of about 31% and 42%. Results in the table correspond to this contact rate, presented as average (minimum, maximum).

‡ Reported contact rate is on average between 16 and 24 per person per day under face-to-face instructions [42, 43]. None of the scenarios for this contact rate were suitable, and thus, are not presented in the table.

⁋ We defined a testing scenario as suitable if there were no exponential growth in infections when transmission rates were 5% and 8% (corresponding to reported use of face mask and physical distancing [11, 30]).

https://doi.org/10.1371/journal.pone.0255864.t003

to-face instructions [42, 43]) suggest an exponential growth in infections in most testing scenarios, even if face mask and physical distancing are used at levels reported during the pandemic (transmission rates of 5% to 8%) (S3 Table). With contact rate of 6 to 8 per person per day (corresponding to reported numbers when several universities moved to partial or full remote instructions [35, 36]) and use of facemask and physical distancing at levels reported during the pandemic, an exponential growth in infections was prevented with the following testing scenarios: 33% per day mass test only, at least 33% trace and test only, any of the combination tests, and any of the delayed combination tests (Table 3). In these suitable scenarios, the peak number of trace and tests, per 10,000 persons, varied from 2 to 64 per day, and the peak number of quarantines, per 10,000 persons, varied from 3 to 26 per day (Table 3).

Discussions

This work estimates, under varying combinations of mass test, trace and test, and transmission rate, the contact rate thresholds that would help efficiently control an infectious disease outbreak on residential university campuses in the United States. The metric typically used in the COVID-19 literature for evaluating testing strategies is the reproduction number R_0, which combines the contact rate and transmission rate. As interventions that influence transmission rates are different than those that influence contact rates, separating these parameters help systematically evaluate metrics to inform epidemic control protocols on university campuses. In this study, we extracted four main metrics. First, the contact rate threshold among non-essential workers after accounting for the higher contact rate among essential workers, which could help inform the size of social circles at the individual-level and schedule group activities such as in labs and offices. Second, population size threshold, i.e., the maximum proportion of the

actual population size, which could help university-level activity decisions such as the fraction of classes that should be moved to remote instruction. Third, the threshold values for vaccination coverage for the campus to return to normal, i.e., the minimum vaccination coverage for having 100% of the population back on campus, which would help plan for the period post introduction of vaccines. Fourth, the threshold values for population size if vaccine coverage is below required thresholds, which could help decisions in the event that vaccines are not widely available that coverage (proportion vaccinated) is not at levels sufficient to fully resume normal activities. The fourth metric could especially be useful in the transitionary phase to normality (until vaccines become fully available) and where the results suggest lowering the population size by just a small number, which could be achieved by moving only a few classes online, such that, the overall population density on campus on any given day is lower but most students have most (if not all) of their classes as face-to-face.

While the implementation of the decisions related to the above metrics are driven at the university-level, adherence and feasibility to use of interventions such as face mask and physical distancing could vary by individual behaviors [37, 39, 40]. By separately modeling contact rates and transmission rates in this study, we extracted results corresponding to transmission rates (of 5% to 8%) that match reported behaviors for face mask use and physical distancing [11, 30], and thus evaluated the university-level decisions under these adherence or feasibility ranges.

Our analyses suggest that implementing only testing, only face mask use and physical distancing, or only population size reductions will not be sufficient, but require combinations of these interventions to successfully control an outbreak on university campuses. Further, in the absence of vaccinations, at reported levels of face mask and physical distancing, testing alone without reducing population size would also not be sufficient to control an outbreak. This suggests that university campuses have high population densities that, for effective control of highly virulent infections such as SARS-CoV-2, it would require reducing the population size such as through remote learning.

Although individual interventions are not sufficient, there are multiple choices for combinations of interventions to choose from if vaccines are unavailable. If, along with continuing face mask and physical distancing at current levels, the population size is kept to at most 34% (or 44%) of the actual population size, mass tests only of 25% (or 33%) per day would help control an outbreak (Table 1). The choice between mass tests of 25% per day vs. 33% per day should consider the costs of a greater proportion remote learning (quantitative and qualitative costs) vs. costs of both testing more often and testing a larger population.

An alternative to mass tests only would be trace and test only, along with continuing face mask and physical distancing at current levels and reducing population size. Trace and test only would also be sufficient at rates of 33% (or 50%) if population size is kept to at most 39% (or 52%) (Table 1). These population size range are close to the 34% (or 44%) reported above for 25% (or 33%) per day mass tests only. Trace and test of 33% and 50% correspond to 3 days and 2 days, respectively, from the time an infected person makes contact with an individual to effective isolation of that individual. Feasibility of this short turnaround times would determine the choice between use of mass test vs. trace and test. Turnaround times are expected to be shorter with digital contact tracing, such as smart phone apps, compared to manual tracing, and feasibility and adoption of apps could be higher among university students than general population. However, studies related to its feasibility and adoption followed by adherence to isolation, among other issues such as privacy and alternative digital technologies are only recently emerging [44–47]. Our results also suggest that, if these turnaround times are not achievable and further if there are any delays in trace and test initiation, then trace and test alone is not favorable (none of the delayed trace and test were feasible (Table 2)) and should

instead adopt either mass tests only or mass test with trace and test. Use of mass test with trace and test could improve trace and test due to potential early diagnosis of index persons. Our results suggest that, if mass tests can increase trace and test to 50% (within two days from contact to isolation), there is more flexibility in trade-offs between mass test rates and contact rate thresholds, and thus more flexibility in population size (Table 2).

In the event that vaccines are available, the full population can be back on campus and resume normal activities provided at least 95% of the population is vaccinated (Table 2). If vaccine coverages are lower than 95%, resuming normal activities with the full population size on campus would require additional asymptomatic testing, with the level of testing depending on vaccine coverage. Mass tests of at least 25% per day would be sufficient if vaccine coverage is at least 70%, or mass tests of at least 33% per day would be sufficient if vaccine coverage is at least 59%. If vaccine coverage is below 59%, to control an outbreak, in addition to mass tests at 33% per day, it would also require lowering the population size to 90%, 75%, and 60%, if vaccine coverage is at 46%, 31%, and 16%, respectively (Table 2).

Corresponding to the reported compliance to face mask and physical distancing and reported contact rates of 6 to 8 per person per day (a population size of 31% to 42% as per our estimations), from surveys [35, 37, 39] conducted over the year 2020 when universities transitioned a large proportion of classes to remote instructions and vaccines were unavailable, our results suggest the need for at least 33% mass test only or 33% trace and test only (Table 3). Scenarios that did not meet these criteria led to exponential growths in infections. These results generally match observed cases over the Fall 2020 semester, where several campuses saw cases into the thousands within the first two weeks of opening and were able to quickly control the spread within two to three weeks by temporarily transitioning to remote instructions [48]. While the universities were able to effectively control the outbreak quickly, it was also observed by this study [48] that the infections rapidly spread into the neighboring community, which were less successful in controlling the spread. Therefore, we believe, results obtained from our study, which set tight tolerance levels on infection cases, would be beneficial for developing epidemic response plans that consider the interests of the broader community. Our results also suggest that, with asymptomatic testing only, it would be necessary to have a vaccination coverage threshold of >95% for a university to fully return back to normal. This threshold is much higher than the typical 70% to 80% range used for herd-immunity in the literature for the general population [49], to a small extent because of setting a tighter tolerance but to a large extent because of the higher population density characteristic of university campuses. The latter can also be observed in R_0 values estimated for universities, which in some instances went above 10 even with online instructions [48], while the herd-immunity in the general population is approximately calculated as $1-1/R_0$ using a R_0 of 3.5.

Our work is subject to limitations. Our model is deterministic. We used an average contact rate for all persons in order to estimate threshold values that could help inform university-level decisions. We did not model contact rates to be representative of actual expected networks between individuals. We did not explicitly model other interventions that could reduce transmission rate such as controlled ventilation, filtering air and controlling air flow, which are likely to impact transmissions [50]. The transmission rates also have a large range of uncertainty due to varying individual behaviors, the data used for streamlining the analyses in this study are based on limited data availabilities, however, the extrapolations over the wide range of transmission rates could be utilized. We did not model false positives for any of the testing scenarios and thus susceptible persons immediately return back to susceptible compartment after testing. We did not model other flu like illnesses and thus we did not assess the additional healthcare resource needs such as testing and quarantining because of similarity in symptoms with COVID-19. In estimation of vaccination thresholds, we did not consider the natural

immunity developed among persons who may have been infected previously. The estimation of vaccination thresholds assume that the virus is still prevalent in the larger community and thus there is a chance of the infection entering the population, such as through local or global travel. We assume that the population density is similar across university campuses with contact rates between 16 and 24, and thus, this assumption should be considered when generalizing to campuses.

In conclusion, the results from this study could be used to collectively inform decisions related to testing, population size reductions through remote instructions, size of social circles, personnel scheduling in labs and offices, under scenarios of both unavailability or partial availability of vaccines, and within the observed levels of compliance to face mask use and physical distancing. The analyses conducted here specifically streamlined the results to the COVID-19 disease caused by the SARS-CoV-2 virus. However, given the wide range of transmission rates evaluated here, which were based on results from a meta-analysis study that evaluated SARS-CoV-2 and other viruses of similarly high virulence [11], broader observations from this study could be extrapolated for use in early stages of new outbreaks of similar viral respiratory infections with similar incubation periods [24], where non-pharmaceutical intervention options such as face masks, physical distancing, remote instructions, and testing are the suitable options. As was the case at the time of conducting this study, in the early stages of an outbreak, there is uncertainty in the baseline transmission rate, efficacy of face mask use and physical distancing [11]. Thus, the results over the range of transmission rates might only serve as a preliminary guide, until more information becomes available for more streamlined analyses.

Supporting information

S1 Table. All combinations of testing, contact rates for non-essential workers, and population size for effective control of a disease outbreak in the absence of vaccines. S: symptomatic testing, U: Mass test, T: trace and test, dT: delayed trace and test.
(XLSX)

S2 Table. Combinations of testing, vaccine coverage, and population size for effective control of a disease outbreak. S: symptomatic testing, U: Mass test, T: trace and test, dT: delayed trace and test.
(XLSX)

S3 Table. Epidemic outcomes under varying levels of testing and corresponding resource needs. S: symptomatic testing, U: Mass test, T: trace and test, dT: delayed trace and test.
(XLSX)

S4 Table. All combinations of testing, contact rates for non-essential workers, and population size for effective control of a disease outbreak in the absence of a vaccine.
(XLSX)

S5 Table. Combinations of testing, vaccine coverage, and population size for effective control of a disease outbreak.
(XLSX)

S6 Table. Epidemic outcomes under varying levels of testing and corresponding resource needs.
(XLSX)

S1 Text. Appendix.
(DOCX)

Acknowledgments

We would like to acknowledge Sonza Singh, Shifali Bansal, Seyedeh Nazanin Khatami, and Arman Mohseni Kabir for their assistance in data collection in initial stages of the study, and Dr. Laura Balzer, Dr. Michael Ash, and Dr. Hari Balasubramanian for their comments and inputs.

Author Contributions

Conceptualization: Xinmeng Zhao, Hanisha Tatapudi, George Corey, Chaitra Gopalappa.

Data curation: Xinmeng Zhao, Hanisha Tatapudi, Chaitra Gopalappa.

Formal analysis: Xinmeng Zhao, Hanisha Tatapudi, Chaitra Gopalappa.

Investigation: Xinmeng Zhao, Hanisha Tatapudi, Chaitra Gopalappa.

Methodology: Xinmeng Zhao, Hanisha Tatapudi, George Corey, Chaitra Gopalappa.

Project administration: Chaitra Gopalappa.

Resources: Chaitra Gopalappa.

Software: Xinmeng Zhao, Hanisha Tatapudi, Chaitra Gopalappa.

Supervision: Chaitra Gopalappa.

Validation: Xinmeng Zhao, Hanisha Tatapudi, Chaitra Gopalappa.

Visualization: Xinmeng Zhao, Hanisha Tatapudi, Chaitra Gopalappa.

Writing – original draft: Xinmeng Zhao, Hanisha Tatapudi, George Corey, Chaitra Gopalappa.

Writing – review & editing: Xinmeng Zhao, Hanisha Tatapudi, George Corey, Chaitra Gopalappa.

References

1. Coronavirus Resource Center. Impact of opening and closing decisions by state. In: Johns Hopkins University of Medicine. [Internet]. 8 Jul 2020 [cited 8 Jul 2020]. Available: https://coronavirus.jhu.edu/data/state-timeline
2. Holzer J. The COVID-19 crisis: How do U.S. employment and health outcomes compare to other OECD countries? In: Brookings [Internet]. 2 Jun 2020 [cited 8 Jul 2020]. Available: https://www.brookings.edu/research/the-covid-19-crisis-how-do-u-s-economic-and-health-outcomes-compare-to-other-oecd-countries/
3. Blustein D, Duffy R, Ferreira J, Cohen-Scali V, Cinamon R, Allan B. Unemployment in the time of COVID-19: A research agenda. J Vocat Behav. 2020; 119. https://doi.org/10.1016/j.jvb.2020.103436 PMID: 32390656
4. CDC. How to Protect Yourself & Others. In: Centers for Disease Control and Prevention [Internet]. 24 Apr 2020 [cited 9 Jul 2020]. Available: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html. https://doi.org/10.1097/01.inf.0000437856.09540.11 PMID: 24569199
5. CDC. Interim Guidance for Businesses and Employers Responding to Coronavirus Disease 2019 (COVID-19). In: Centers for Disease Control and Prevention [Internet]. 6 May 2020 [cited 8 Jul 2020]. Available: https://www.cdc.gov/coronavirus/2019-ncov/community/guidance-business-response.html.
6. Barratt H, Kirwan M, Shantikumar S. Epidemic theory (effective & basic reproduction numbers, epidemic thresholds) & techniques for analysis of infectious disease data (construction & use of epidemic curves, generation numbers, exceptional reporting & identification of significant clusters). In: Health Knowledge [Internet]. [cited 8 Jul 2020]. Available: https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/epidemic-theory
7. Kucharski A, Klepac P, Conlan A, Kissler S, Tang M, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a

mathematical modelling study. Lancet Infect Dis. 2020; 20: 1151–1160. https://doi.org/10.1016/S1473-3099(20)30457-6 PMID: 32559451

8. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020; 26: 672–675. https://doi.org/10.1038/s41591-020-0869-5 PMID: 32296168
9. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, et al. COVID-19: towards controlling of a pandemic. The Lancet. 2020; 395: 1015–1018. https://doi.org/10.1016/S0140-6736(20)30673-5 PMID: 32197103
10. MacIntyre CR, Chughtai AA. A rapid systematic review of the efficacy of face masks and respirators against coronaviruses and other respiratory transmissible viruses for the community, healthcare workers and sick patients. International Journal of Nursing Studies. 2020; 108: 103629. https://doi.org/10.1016/j.ijnurstu.2020.103629 PMID: 32512240
11. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet. 2020; 395: 1973–1987. https://doi.org/10.1016/S0140-6736(20)31142-9
12. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health. 2020; 8: e488–e496. https://doi.org/10.1016/S2214-109X(20)30074-7 PMID: 32119825
13. Aleta A, Martín-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat Hum Behav. 2020; 4: 964–971. https://doi.org/10.1038/s41562-020-0931-9 PMID: 32759985
14. Gressman PT, Peck JR. Simulating COVID-19 in a University Environment. Mathematical Biosciences. 2020; 328: 108436. https://doi.org/10.1016/j.mbs.2020.108436 PMID: 32758501
15. Paltiel AD, Zheng A, Walensky RP. COVID-19 screening strategies that permit the safe re-opening of college campuses. Infectious Diseases (except HIV/AIDS); 2020 Jul. https://doi.org/10.1101/2020.07.06.20147702 PMID: 32676614
16. Knowledge @ Wharton. COVID-19 on Campus: How Should Schools Be Redesigned? 22 Jun 2020 [cited 8 Jul 2020]. Available: https://knowledge.wharton.upenn.edu/article/how-should-universities-be-redesigned-in-the-wake-of-covid-19/
17. Vasquez M, Diep F. What Covid-19 Computer Models Are Telling Colleges About the Fall. In: The Chronicle of Higher Education [Internet]. 19 Jun 2020 [cited 8 Jul 2020]. Available: https://www.chronicle.com/article/What-Covid-19-Computer-Models/249027
18. Roos R. Masks plus hand hygiene reduced ILI in college dorm study. In: CIDRAP [Internet]. 22 Jan 2010 [cited 8 Jul 2020]. Available: https://www.cidrap.umn.edu/news-perspective/2010/01/masks-plus-hand-hygiene-reduced-ili-college-dorm-study
19. Aiello AE, Murray GF, Perez V, Coulborn RM, Davis BM, Uddin M, et al. Mask use, hand hygiene, and seasonal influenza-like illness among young adults: A randomized intervention trial. The Journal of Infectious Diseases. 2010; 201: 491–498. https://doi.org/10.1086/650396 PMID: 20088690
20. Meselson M. Droplets and Aerosols in the Transmission of SARS-CoV-2. N Engl J Med. 2020; 382: 2063–2063. https://doi.org/10.1056/NEJMc2009324 PMID: 32294374
21. Jayaweera M, Perera H, Gunawardana B, Manatunge J. Transmission of COVID-19 virus by droplets and aerosols: A critical review on the unresolved dichotomy. Environmental Research. 2020; 188: 109819. https://doi.org/10.1016/j.envres.2020.109819 PMID: 32569870
22. Anderson EL, Turnham P, Griffin JR, Clarke CC. Consideration of the Aerosol Transmission for COVID-19 and Public Health. Risk Analysis. 2020; 40: 902–907. https://doi.org/10.1111/risa.13500 PMID: 32356927
23. Zhang Y, Jiang B, Yuan J, Tao Y. The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study. MedRxiv. 2020. https://doi.org/10.1101/2020.03.04.20031187
24. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DA. Incubation periods of acute respiratory viral infections: a systematic review. The Lancet Infectious Diseases. 2009; 9: 291–300. https://doi.org/10.1016/S1473-3099(09)70069-6 PMID: 19393959
25. Hill A, Levy M, Xie S, Sheen J, Shinnick J, Gheorghe A, et al. Modeling COVID-19 Spread vs Healthcare Capacit. 27 May 2020. Available: https://alhill.shinyapps.io/COVID19seir/
26. Ma S, Zhang J, Zeng M, Yun Q, Guo W, Zheng Y, et al. Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. Infectious Diseases (except HIV/AIDS); 2020 Mar. https://doi.org/10.1101/2020.03.21.20040329

27. CDC. COVID-19 Pandemic Planning Scenarios. In: Centers for Disease Control and Prevention [Internet]. [cited 5 May 2021]. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-archive/planning-scenarios-2020-05-20.pdf

28. MIDAS 2019 Novel Coronavirus Repository. [cited 24 Apr 2020]. Available: https://github.com/midas-network/COVID-19

29. Kobokovich A, West R, Gronvall G. Serology-based tests for COVID-19. In: Johns Hopkins Bloomberg School of Public Health, Center for Healthy Security [Internet]. 7 Jul 2020 [cited 8 Jul 2020]. Available: https://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html

30. Zhao PJ. A Social Network Model of the COVID-19 Pandemic. Epidemiology; 2020 Mar. https://doi.org/10.1111/aphw.12226 PMID: 32945123

31. At a glance, 2019–2020, University of Massachusetts, Amherst. [cited 9 Jun 2020]. Available: https://www.umass.edu/oir/sites/default/files/publications/glance/FS_gla_01.pdf

32. ACS DEMOGRAPHIC AND HOUSING ESTIMATES. In: United States Census Bureau [Internet]. [cited 5 May 2021]. Available: https://data.census.gov/cedsci/table?q=Profile%20of%20General%20Population%20and%20Housing%20Characteristics%20amherst%20massachusetts&tid=ACSDP5Y2019.DP05

33. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B, et al. Mortality Risk of COVID-19. In: Our World in Data [Internet]. [cited 5 May 2020]. Available: https://ourworldindata.org/mortality-risk-covid#the-case-fatality-rate

34. CDC. COVID Data Tracker. [cited 18 Jun 2021]. Available: https://covid.cdc.gov/covid-data-tracker/#county-view

35. Nixon E, Trickey A, Christensen H, Finn A, Thomas A, Relton C, et al. Contacts and behaviours of university students during the COVID-19 pandemic at the start of the 2020/21 academic year. Public and Global Health; 2020 Dec. https://doi.org/10.1101/2020.12.09.20246421

36. Cashore J, Duan N, Janmohamed A, Wan J, Zhang Y, Henderson S. COVID-19 Mathematical Modeling for Cornell's Fall Semester. 2020. Available: https://cpb-us-w2.wpmucdn.com/sites.coecis.cornell.edu/dist/3/341/files/2020/10/COVID_19_Modeling_Jun15-VD.pdf

37. Cohen AK, Hoyt LT, Dull B. A Descriptive Study of COVID-19–Related Experiences and Perspectives of a National Sample of College Students in Spring 2020. Journal of Adolescent Health. 2020; 67: 369–375. https://doi.org/10.1016/j.jadohealth.2020.06.009 PMID: 32593564

38. Brooks-Pollock E, Christensen H, Trickey A, Hemani G, Nixon E, Thomas A, et al. High COVID-19 transmission potential associated with re-opening universities can be mitigated with layered interventions. Infectious Diseases (except HIV/AIDS); 2020 Sep. https://doi.org/10.1101/2020.09.10.20189696

39. COVID-19 Mitigation Behavior Survey Results. In: Kansas State University [Internet]. [cited 26 Apr 2020]. Available: https://www.k-state.edu/covid-19/communications/every-wildcat-a-wellcat/fall-2020-mitigation-behavior-survey-results.html

40. COVID-19 Reopening Survey Data. In: Risk and Social Policy Group [Internet]. [cited 26 Apr 2020]. Available: https://static1.squarespace.com/static/5ec4464f22cd13186530a36f/t/5ef0e413c14aa311f58404fe/1592845331607/FINAL_onepager_wave1.pdf

41. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. Mathematical Biosciences. 2013; 244: 125–134. https://doi.org/10.1016/j.mbs.2013.04.013 PMID: 23665296

42. Danon L, House TA, Read JM, Keeling MJ. Social encounter networks: collective properties and disease transmission. J R Soc Interface. 2012; 9: 2826–2833. https://doi.org/10.1098/rsif.2012.0357 PMID: 22718990

43. Beutels P, Shkedy Z, Aerts M, Van Damme P. Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. Epidemiol Infect. 2006; 134: 1158–1166. https://doi.org/10.1017/S0950268806006418 PMID: 16707031

44. Menges D, Aschmann H, Moser A, Althaus CL, von Wyl V. The role of the SwissCovid digital proximity tracing app during the pandemic response: results for the Canton of Zurich. Infectious Diseases (except HIV/AIDS); 2021 Feb. https://doi.org/10.1101/2021.02.01.21250972

45. Colizza V, Grill E, Mikolajczyk R, Cattuto C, Kucharski A, Riley S, et al. Time to evaluate COVID-19 contact-tracing apps. Nat Med. 2021; 27: 361–362. https://doi.org/10.1038/s41591-021-01236-6 PMID: 33589822

46. Lewis D. Contact-tracing apps help reduce COVID infections, data suggest. In: Nature [Internet]. 21 Feb 2021 [cited 1 May 2021]. Available: https://www.nature.com/articles/d41586-021-00451-y https://doi.org/10.1038/d41586-021-00451-y PMID: 33623147

47. Masel J, Shilen A, Helming B, Rutschman J, Windham G, Judd M, et al. Quantifying meaningful adoption of a SARS-CoV-2 exposure notification app on the campus of the University of Arizona. Epidemiology; 2021 Feb. https://doi.org/10.1101/2021.02.02.21251022

48. Lu H, Weintz C, Pace J, Indana D, Linka K, Kuhl E. Are college campuses superspreaders? A data-driven modeling study. Computer Methods in Biomechanics and Biomedical Engineering. 2021; 1–11. https://doi.org/10.1080/10255842.2020.1869221 PMID: 33439055

49. Aschwanden C. Five reasons why COVID herd immunity is probably impossible. In: Nature [Internet]. 18 Mar 2021 [cited 1 May 2021]. Available: https://www.nature.com/articles/d41586-021-00728-2 https://doi.org/10.1038/d41586-021-00728-2 PMID: 33737753

50. Dai H, Zhao B. Association of the infection probability of COVID-19 with ventilation rates in confined spaces. Build Simul. 2020; 13: 1321–1327. https://doi.org/10.1007/s12273-020-0703-5 PMID: 32837691

Appendix H: Tables for Evaluating the Sensitivity of Jurisdictional Heterogeneity and Mixing in National-level HIV Prevention Analyses: Context of the U.S. Ending the HIV Epidemic Goal

Table H.1 List of jurisdictions modeled

List of counties from CDC Atlas					***List of States and dependent areas from CDC Atlas***			
S. no.	*State*	*County*	***FIPS***	*Modeled/ why not modeled*	*S.no.*	*State*	***FIPS***	*Modeled/ why not modeled*
1.	CA	Alameda County	**6001**	Yes	1.	Alabama	**1**	Yes
2.	MD	Baltimore City	**24510**	Yes	2.	Alaska	**2**	Yes
3.	TX	Bexar County	**48029**	Yes	3.	Arizona	**4**	Yes
4.	NY	Bronx County	**36005**	Yes	4.	Arkansas	**5**	Yes
5.	FL	Broward County	**12011**	Yes	5.	California	**6**	Yes
6.	NV	Clark County	**32003**	Yes	6.	Colorado	**8**	Yes
7.	GA	Cobb County	**13067**	Yes	7.	Connecticut	**9**	Yes
8.	IL	Cook County	**17031**	Yes	8.	Delaware	**10**	Yes
9.	OH	Cuyahoga County	**39035**	Yes	9.	District of Columbia	**11**	Yes
10.	TX	Dallas County	**48113**	Yes	10.	Florida	**12**	Yes
11.	GA	Dekalb County	**13089**	Yes	11.	Georgia	**13**	Yes
12.	DC	District of Columbia	**11001**	Modeled as a state	12.	Hawaii	**15**	Yes

Table H.1 (Continued)

13.	FL	Duval County	**12031**	Yes	13.	Idaho	**16**	Yes
14.	LA	East Baton Rouge Parish	**22033**	Yes	14.	Illinois	**17**	Yes
15.	NJ	Essex County	**34013**	Yes	15.	Indiana	**18**	Yes
16.	OH	Franklin County	**39049**	Yes	16.	Iowa	**19**	Yes
17.	GA	Fulton County	**13121**	Yes	17.	Kansas	**20**	Yes
18.	GA	Gwinnett County	**13135**	Yes	18.	Kentucky	**21**	Yes
19.	OH	Hamilton County	**39061**	Yes	19.	Louisiana	**22**	Yes
20.	TX	Harris County	**48201**	Yes	20.	Maine	**23**	Yes
21.	FL	Hillsborough County	**12057**	Yes	21.	Maryland	**24**	Yes
22.	NJ	Hudson County	**34017**	Yes	22.	Massachusetts	**25**	Yes
23.	WA	King County	**53033**	Yes	23.	Michigan	**26**	Yes
24.	NY	Kings County	**36047**	Yes	24.	Minnesota	**27**	Yes
25.	CA	Los Angeles County	**6037**	Yes	25.	Mississippi	**28**	Yes
26.	AZ	Maricopa County	**4013**	Yes	26.	Missouri	**29**	Yes
27.	IN	Marion County	**18097**	Yes	27.	Montana	**30**	Yes
28.	NC	Mecklenburg County	**37119**	Yes	28.	Nebraska	**31**	Yes
29.	FL	Miami-Dade County	**12086**	Yes	29.	Nevada	**32**	Yes
30.	MD	Montgomery County	**24031**	Yes	30.	New Hampshire	**33**	Data suppressed
31.	NY	New York County	**36061**	Yes	31.	New Jersey	**34**	Yes
32.	FL	Orange County	**12095**	Yes	32.	New Mexico	**35**	Yes
33.	CA	Orange County	**6059**	Yes	33.	New York	**36**	Yes

Table H.1 (Continued)

34.	LA	Orleans Parish	**22071**	Yes	34.	North Carolina	**37**	Yes
35.	FL	Palm Beach County	**12099**	Yes	35.	North Dakota	**38**	Yes
36.	PA	Philadelphia County	**42101**	Yes	36.	Ohio	**39**	Yes
37.	FL	Pinellas County	**12103**	Yes	37.	Oklahoma	**40**	Yes
38.	MD	Prince George's County	**24033**	Yes	38.	Oregon	**41**	Yes
39.	NY	Queens County	**36081**	Yes	39.	Pennsylvania	**42**	Yes
40.	CA	Riverside County	**6065**	Yes	40.	Rhode Island	**44**	Yes
41.	CA	Sacramento County	**6067**	Yes	41.	South Carolina	**45**	Yes
42.	CA	San Bernardino County	**6071**	Yes	42.	South Dakota	**46**	Yes
43.	CA	San Diego County	**6073**	Yes	43.	Tennessee	**47**	Yes
44.	CA	San Francisco County	**6075**	Data suppressed	44.	Texas	**48**	Yes
45.	PR	San Juan Municipio	**72127**	Population demographic data not available for counties and state	45.	Utah	**49**	Yes
46.	TN	Shelby County	**47157**	Yes	46.	Vermont	**50**	Yes
47.	MA	Suffolk County	**25025**	Data suppressed	47.	Virginia	**51**	Yes
48.	TX	Tarrant County	**48439**	Yes	48.	Washington	**53**	Yes
49.	TX	Travis County	**48453**	Yes	49.	West Virginia	**54**	Yes

Table H.1 (Continued)

50.	MI	Wayne County	**26163**	Yes	50.	Wisconsin	**55**	Yes
					51.	Wyoming	**56**	Yes
					52.	American Samoa	**60**	Data not available
					53.	Guam	**66**	Data not available
					54.	Northern Mariana Islands	**69**	Data not available
					55.	Puerto Rico	**72**	Demographic data not available
					56.	U.S. Virgin Islands	**78**	Data not available

EHE jurisdictions are in blue; states that have EHE counties within them are excluded

Table H.2 Rates of care continuum and disease progression used in the matrix G_t

From	***To***	***Progression type***	***Rate****	***Source***
(A-U) [1]	(A-ANA) [2]	Care	Diagnosis rate x (1-linkage to care) x θ_d	Estimated
	(U) >500 [3]	Disease	5.88	[1]
(A-ANA) [2]	(ANA) >500 [4]	Disease	5.88	[1]
(U) >500 [3]	(ANA) >500 [4]	Care	Diagnosis rate x (1-linkage to care) x θ_d	Estimated
	(ANV) >500 [5]	Care	Diagnosis rate x linkage to care x θ_d	Estimated
	(U) 351-500 [7]	Disease	0.286	[1]
(ANA) >500 [4]	(ANV) >500 [5]	Care	0.5	[2]
	(ANA) 351-500 [8]	Disease	0.286	[1]
(ANV) >500 [5]	(ANA) >500 [4]	Care	Dropout rate x φ_d	Estimated
	(VLS) >500 [6]	Care	1.33	[3]
	(ANV) 351-500 [9]	Disease	0.026	[1]
(VLS) >500 [6]	(ANA) >500 [4]	Care	Dropout rate x φ_d	Estimated

* Rates in table represent annual rates input to the simulation model

Table H.2 (Continued)

(U) 351-500 [7]	(ANA) 351-500 [8]	Care	Diagnosis rate x (1-linkage to care) x θ_d	Estimated
	(ANV) 351-500 [9]	Care	Diagnosis rate x linkage to care x θ_d	Estimated
	(U) 201-350 [11]	Disease	0.286	[1]
(ANA) 351-500 [8]	(ANV) 351-500 [9]	Care	0.5	[2]
	(ANA) 201-350 [12]	Disease	0.286	[1]
(ANV) 351-500 [9]	(ANA) 351-500 [8]	Care	Dropout rate x φ_d	Estimated
	(VLS) 351-500 [10]	Care	1.33	[3]
	(ANV) 201-350 [13]	Disease	0.026	[1]
(VLS) 351-500 [10]	(VLS) >500 [6]	Disease	0.385	[1]
	(ANA) 351-500 [8]	Care	Dropout rate x φ_d	Estimated
(U) 201-350 [11]	(ANA) 201-350 [12]	Care	Diagnosis rate x (1-linkage to care) x θ_d	Estimated
	(ANV) 201-350 [13]	Care	Diagnosis rate x linkage to care x θ_d	Estimated
	(U) <200 [15]	Disease	0.33	[1]
(ANA) 201-350 [12]	(ANV) 201-350 [13]	Care	0.5	[2]
	(ANA) <200 [16]	Disease	0.33	[1]
(ANV) 201-350 [13]	(ANA) 201-350 [12]	Care	Dropout rate x φ_d	Estimated
	(VLS) 201-350 [14]	Care	1.33	[3]
	(ANV) <200 [17]	Disease	0.026	[1]
(VLS) 201-350 [14]	(VLS) 351-500 [10]	Disease	0.385	[1]
	(ANA) 201-350 [12]	Care	Dropout rate x φ_d	Estimated

Table H.2 (Continued)

(U) <200 [15]	(ANA) <200 [16]	Care	Diagnosis rate x (1-linkage to care) x θ_d	Estimated
	(ANV) <200 [17]	Care	Diagnosis rate x linkage to care x θ_d	Estimated
(ANA) <200 [16]	(ANV) <200 [17]	Care	1	[4]
(ANV) <200 [17]	(ANA) <200 [16]	Care	Dropout rate x φ_d	Estimated
	(VLS) <200 [18]	Care	1.33	[3]
(VLS) <200 [18]	(VLS) 201-350 [14]	Disease	0.355	[1]
	(ANA) <200 [16]	Care	Dropout rate x φ_d	Estimated

Diagnosis rate and Dropout rate are rates of care metrics estimated monthly for each risk group
Data on linkage to care changes across risk groups, time, and jurisdictions [5]
θ_d= scaling factor for diagnosis rate in disease-stage d, see Table H.3
φ_d= scaling factor for drop-out rate in disease-stage d, see Table H.4

Table H.3 Scaling factor for diagnosis rate in disease-stage d (θ_d)

Disease stage	Risk group: HM	Risk group: HF	Risk group: MSM
Acute	0.090	0.090	0.090
CD_4>500	0.366	0.366	0.366
CD_4 500-350	0.775	0.775	0.776
$CD_4$200-350	1.337	1.337	1.337
CD_4<200	2.353	2.353	2.352

Table H.4 Scaling factor for dropout rate in disease-stage d (φ_d)

Disease stage	Risk group: MSM
Acute	1
CD_4>500	1
CD_4 500-350	1
$CD_4$200-350	1
CD_4<200	0

The rate of incidence or the number of new infections is calculated using a Bernoulli equation (see Section H.3). The rate of deaths from the susceptible population is shown in Table H.5, and for infected compartments is shown in Tables H.6 and H.7.

Table H.5 Death rates for susceptible population by age and risk group [6]

Age	Risk group: HM/MSM	Risk group: HF	Age	Risk group: HM/MSM	Risk group: HF
13	0.000232	0.000138	57	0.009156	0.005647
14	0.000343	0.000172	58	0.009897	0.006043
15	0.000465	0.000211	59	0.010671	0.006441
16	0.000588	0.000251	60	0.011519	0.006886
17	0.00072	0.000293	61	0.012419	0.007391
18	0.000858	0.000336	62	0.013307	0.007931
19	0.000999	0.000379	63	0.014164	0.008508
20	0.001146	0.000425	64	0.015032	0.009142
21	0.001288	0.000472	65	0.016013	0.009874
22	0.001407	0.000515	66	0.017138	0.010717
23	0.001494	0.000551	67	0.018362	0.01166
24	0.001556	0.000582	68	0.019693	0.012711
25	0.00161	0.000612	69	0.021174	0.013894
26	0.001665	0.000646	70	0.022889	0.015285
27	0.001717	0.000684	71	0.024869	0.016878
28	0.001767	0.000729	72	0.027095	0.018607
29	0.001817	0.000779	73	0.029587	0.020466
30	0.001865	0.000833	74	0.032394	0.022522
31	0.001911	0.000887	75	0.035668	0.024929
32	0.00196	0.000939	76	0.039396	0.027729
33	0.002014	0.000988	77	0.043453	0.030855
34	0.002071	0.001034	78	0.047826	0.034321
35	0.002138	0.001085	79	0.052649	0.038211
36	0.002211	0.001143	80	0.058206	0.042771
37	0.002279	0.001205	81	0.064581	0.047992
38	0.002342	0.001271	82	0.071657	0.053678
39	0.002405	0.001345	83	0.079465	0.05981
40	0.002482	0.001429	84	0.088141	0.066584
41	0.002583	0.001524	85	0.097854	0.074258
42	0.00271	0.00163	86	0.108747	0.083053
43	0.00287	0.001748	87	0.120919	0.093123
44	0.003064	0.001881	88	0.134425	0.10454
45	0.003285	0.002029	89	0.149273	0.117305
46	0.003538	0.002195	90	0.165452	0.131392
47	0.003834	0.002386	91	0.182935	0.146753
48	0.004178	0.002605	92	0.201679	0.163331
49	0.004569	0.002851	93	0.221637	0.181064
50	0.004997	0.003118	94	0.242747	0.199886
51	0.005462	0.003403	95	0.263672	0.218908
52	0.005971	0.003714	96	0.284014	0.237815

Table H.5 (Continued)

53	0.006526	0.004052	97	0.303355	0.256265
54	0.007125	0.004415	98	0.321268	0.273894
55	0.007766	0.004813	99	0.337332	0.290328
56	0.008445	0.005233	100	0.354198	0.307747

Table H.6 Death rates for HIV infected without ART [7]

Disease stage	Death rate
CD4 <200	0.116620159
CD4 200-350	0.02371429
CD4 350-500	0.011928287
CD4 >500	0.007968085
Acute	0.007968085

Table H.7 Death rates for HIV infected for different disease stages [8]

Age group	Disease stage CD_4 > 350	Disease stage CD_4 > 200-350	Disease stage CD_4 < 200
13-29	0.003626204	0.005050763	0.014833591
30-39	0.004642914	0.006278342	0.018685359
40-49	0.006073324	0.008337886	0.025242614
50-100	0.046050426	0.015897488	0.011459655

Table H.8 Calibrated values of probability of HIV transmission per sexual act of different types (vaginal and anal) for different risk groups

Risk group	Vaginal acts	Anal acts
HM	0.0003	0.0011
HF	0.0007	0.0082
MSM	0.0022	0.0103

Table H.9 Calibrated values of sexual partnership mixing proportions between different risk groups

Risk group	HM	HF	MSM
HM	0	100.00%	0
HF	98.20%	0	1.80%
MSM	0	40.00%	60.00%

Table H.10 Calibrated values for risk-specific age-related partnership mixing proportions

Risk group	Age group	13-17	18-24	25-29	30-24	35-39	40-44	45-64	65-100
HM	13-17	91.05 %	4.24%	1.12%	1.12%	1.12%	1.12%	0.23%	0%
	18-24	2.30%	92.10 %	1.12%	1.12%	1.12%	1.12%	1.12%	0%
	25-29	6.76%	6.76%	82.00 %	1.12%	1.12%	1.12%	1.12%	0%
	30-24	14.05 %	14.05 %	14.05 %	54.50 %	1.12%	1.12%	1.12%	0%
	35-39	5.39%	5.39%	5.39%	5.39%	76.20 %	1.12%	1.12%	0%
	40-44	4.54%	4.54%	4.54%	4.54%	4.54%	76.20 %	1.12%	0%
	45-64	3.97%	3.97%	3.97%	3.97%	3.97%	3.97%	76.20%	0%
	65-100	0%	0%	0%	0%	0%	0%	0%	0%
HF	13-17	91.05 %	6.95%	0.50%	0.50%	0.50%	0.50%	0.00%	0%
	18-24	6.45%	91.05 %	0.50%	0.50%	0.50%	0.50%	0.50%	0%
	25-29	0.50%	39.80 %	57.70 %	0.50%	0.50%	0.50%	0.50%	0%
	30-24	0.50%	43.00 %	0.50%	54.50 %	0.50%	0.50%	0.50%	0%
	35-39	1.50%	14.70 %	0.50%	0.50%	81.80 %	0.50%	0.50%	0%
	40-44	0.50%	15.70 %	0.50%	0.50%	0.50%	81.80 %	0.50%	0%
	45-64	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00 %	0%
	65-100	0%	0%	0%	0%	0%	0%	0%	0%

Table H.10 (Continued)

MSM	13-17	91.05 %	4.69%	0.48%	0.14%	1.09%	0.79%	1.84%	0%
	18-24	4.82%	48.00 %	4.79%	1.56%	3.80%	32.97 %	4.05%	0%
	25-29	10.36 %	16.44 %	55.90 %	13.47 %	1.94%	1.50%	0.39%	0%
	30-24	0.23%	1.29%	37.58 %	46.30 %	2.68%	8.22%	3.69%	0%
	35-39	6.05%	24.61 %	0.98%	5.31%	55.20 %	1.29%	6.56%	0%
	40-44	3.40%	7.39%	10.49 %	4.34%	9.34%	55.20 %	9.85%	0%
	45-64	11.00 %	10.42 %	9.99%	10.52 %	1.68%	1.18%	55.20%	0%
	65-100	0%	0%	0%	0%	0%	0%	0%	0%

Table H.11 Ranges for age and gender-specific number of sexual acts per year [9]

Age group	Female-upper	Female-lower	Male-upper	Male -lower
13-14	41	20	60	30
15-17	41	20	60	30
18-19	127	73	119	68
20-24	127	73	119	68
25-29	108	62	110	63
30-34	93	51	104	59
35-39	93	51	104	59
40-44	86	48	95	52
45-49	86	48	95	39
50-54	73	40	73	36
55-59	73	32	73	36
60-64	62	35	67	24
65-70	62	35	67	24

*The ranges were transformed to values by taking the average for each age group and gender across upper and lower estimates

H.1 References for Appendix H

1. Khurana N, Yaylali E, Farnham PG, Hicks KA, Allaire BT, Jacobson E, et al. Impact of Improved HIV Care and Treatment on PrEP Effectiveness in the United States, 2016–2020. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2018;78: 399–405. doi:10.1097/QAI.0000000000001707

2. Gardner EM, McLees MP, Steiner JF, del Rio C, Burman WJ. The Spectrum of Engagement in HIV Care and its Relevance to Test-and-Treat Strategies for Prevention of HIV Infection. Clinical Infectious Diseases. 2011;52: 793–800. doi:10.1093/cid/ciq243

3. Clinical Info HIV.gov. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents Living with HIV. 18 Dec 2019 [cited 9 Nov 2021]. Available: https://clinicalinfo.hiv.gov/en/guidelines/adult-and-adolescent-arv/initiation-antiretroviral-therapy

4. Gopalappa C, Farnham PG, Chen Y-H, Sansom SL. Progression and Transmission of HIV/AIDS (PATH 2.0): A New, Agent-Based Model to Estimate HIV Transmissions in the United States. Med Decis Making. 2017;37: 224–233. doi:10.1177/0272989X16668509

5. NCHHSTP AtlasPlus. [cited 4 Nov 2021]. Available: https://www.cdc.gov/nchhstp/atlas/index.htm

6. Social Security. Actuarial Life Table. [cited 9 Nov 2021]. Available: https://www.ssa.gov/oact/STATS/table4c6.html

7. Grover D, Copas A, Green H, Edwards SG, Dunn DT, Sabin C, et al. What is the risk of mortality following diagnosis of multidrug-resistant HIV-1? Journal of Antimicrobial Chemotherapy. 2008;61: 705–713. doi:10.1093/jac/dkm522

8. Prognosis of HIV-1-infected patients up to 5 years after initiation of HAART: collaborative analysis of prospective studies. AIDS. 2007;21: 1185–1197. doi:10.1097/QAD.0b013e328133f285

9. Reece M, Herbenick D, Schick V, Sanders SA, Dodge B, Fortenberry JD. Sexual Behaviors, Relationships, and Perceived Health Among Adult Men in the United States: Results from a National Probability Sample. The Journal of Sexual Medicine. 2010;7: 291–304. doi:10.1111/j.1743-6109.2010.02009.x

Appendix I: Figures for Evaluating the Sensitivity of Jurisdictional Heterogeneity and Mixing in National-level HIV Prevention Analyses: Context of the U.S. Ending the HIV Epidemic Goal

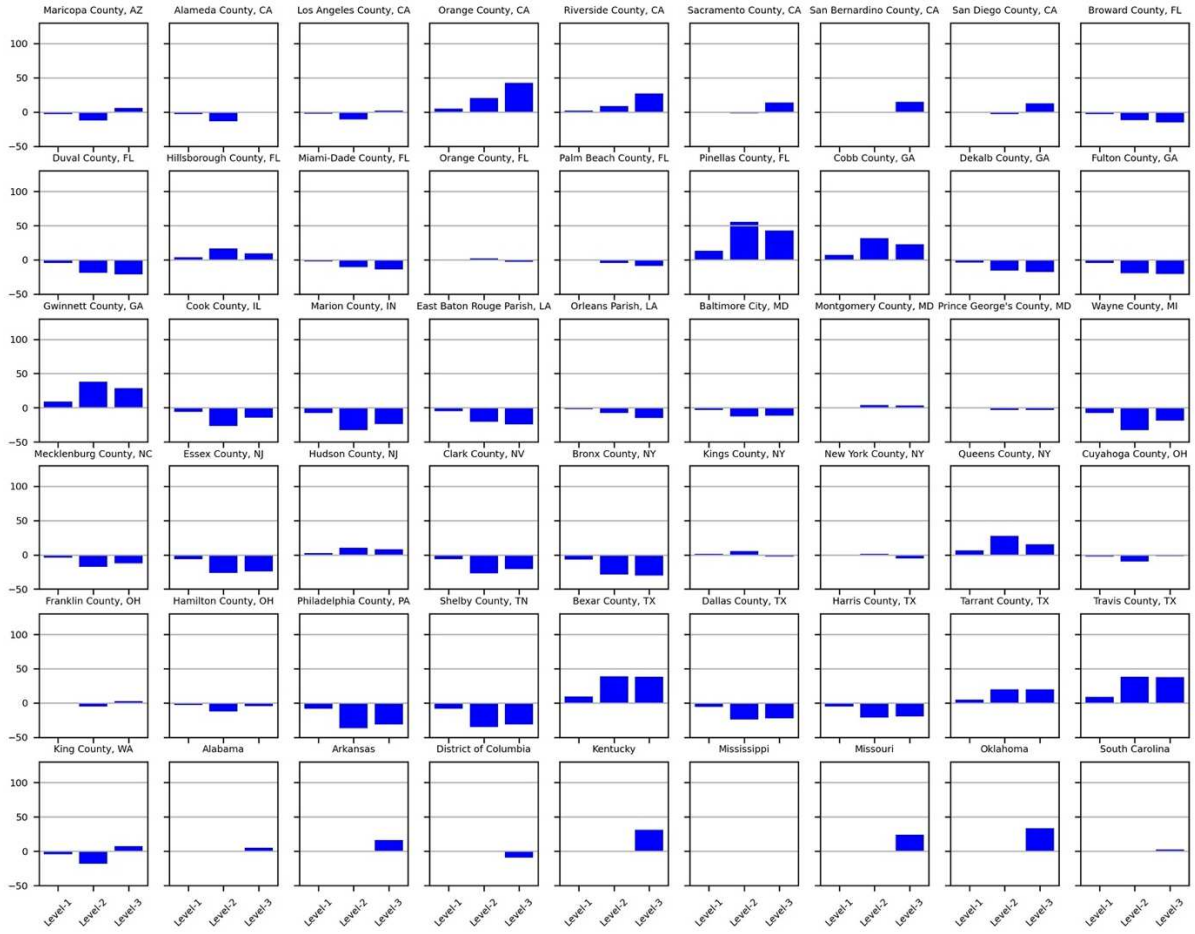


Figure I.1.a EHE jurisdiction; Comparing percentage difference of overall incidence between mixing scenarios for risk group HM. Comparing no mixing (Scenario 13) with Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16) for each EHE jurisdiction for year 2018.

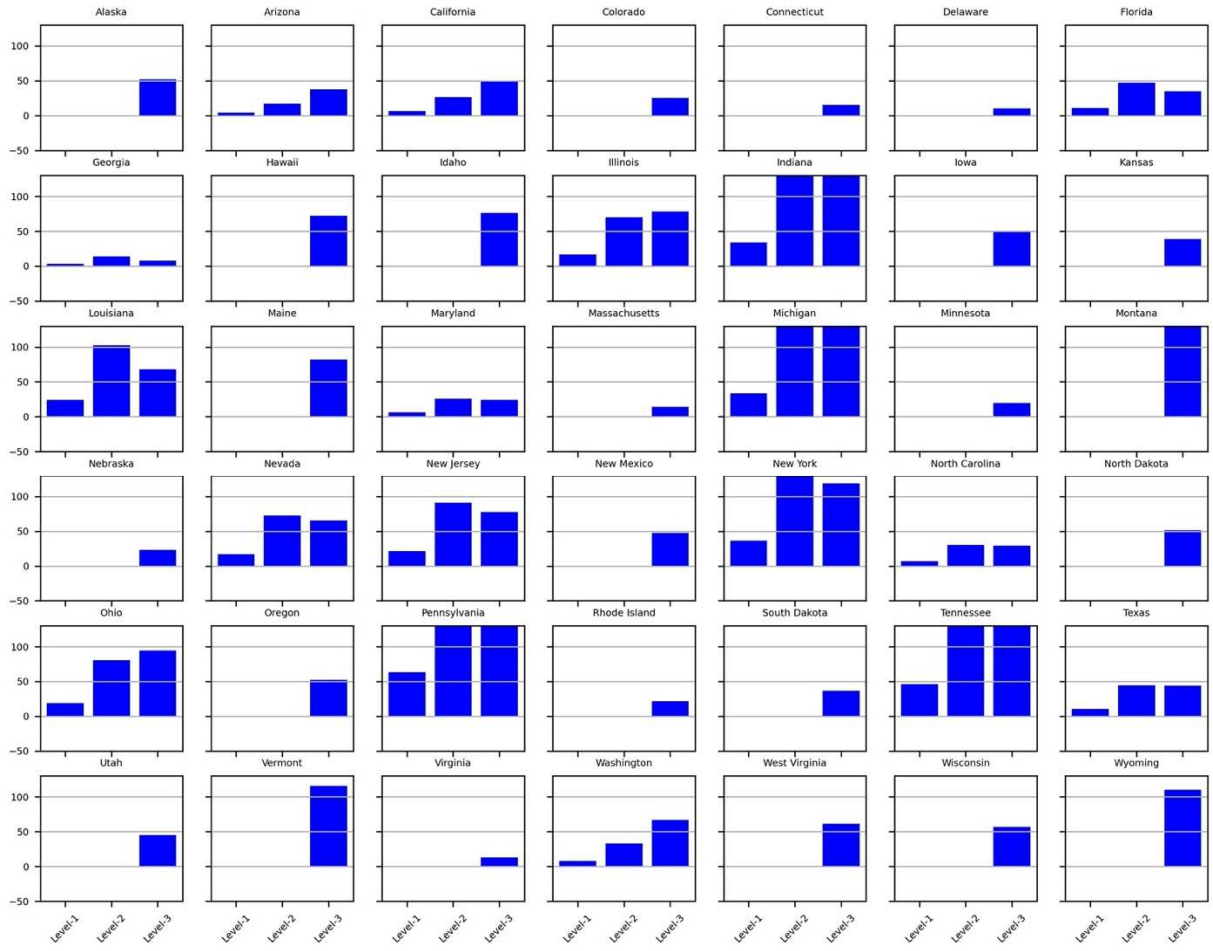


Figure I.1.b Non-EHE jurisdiction; Comparing percentage difference of overall incidence between mixing scenarios for risk group HM. Comparing no mixing (Scenario 13) with Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16) for each non-EHE jurisdiction for year 2018.

Figure I.2.a EHE jurisdiction; Comparing percentage difference of overall incidence between mixing scenarios for risk group HF. Comparing no mixing (Scenario 13) with Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16) for each EHE jurisdiction for year 2018.

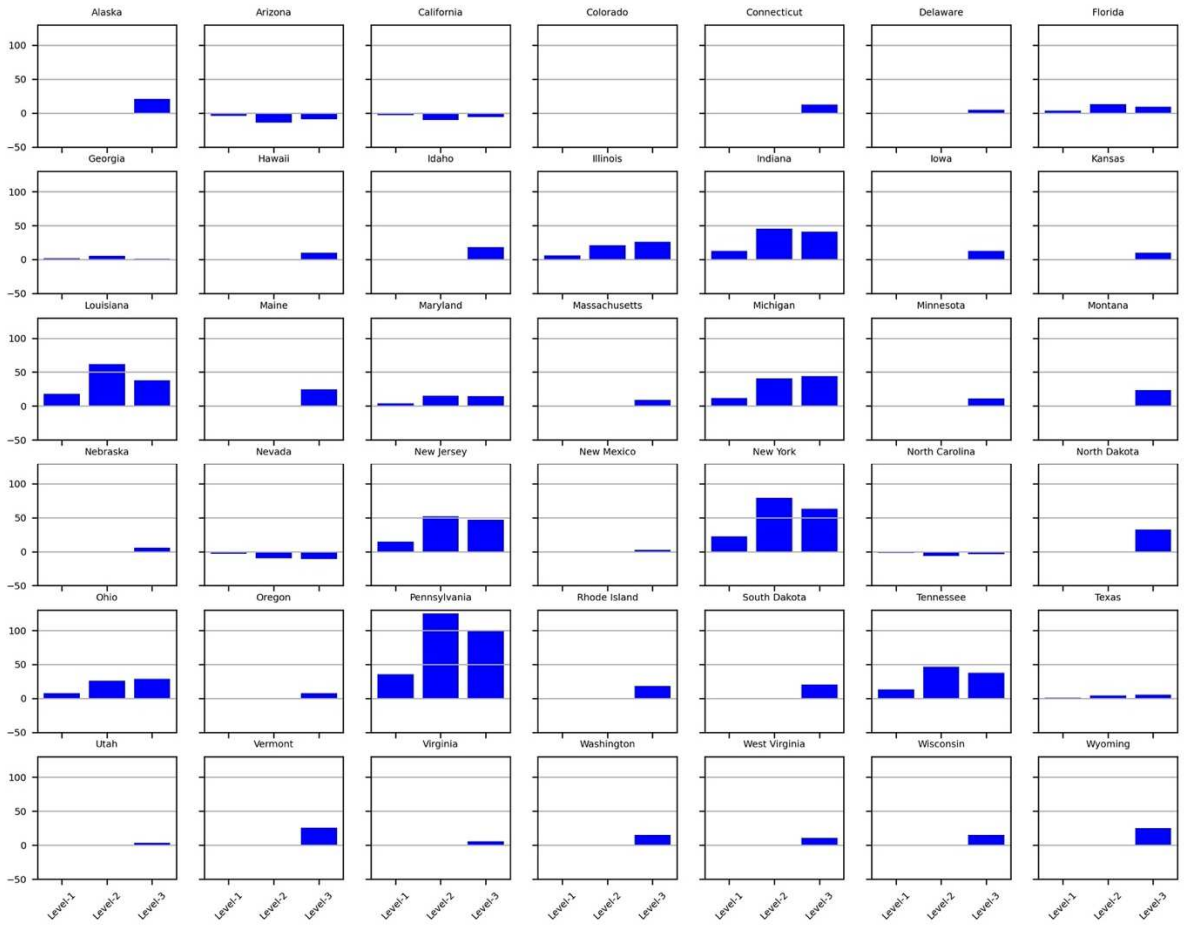


Figure I.2.b Non-EHE jurisdiction; Comparing percentage difference of overall incidence between mixing scenarios for risk group HF. Comparing no mixing (Scenario 13) with Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16) for each non-EHE jurisdiction for year 2018.

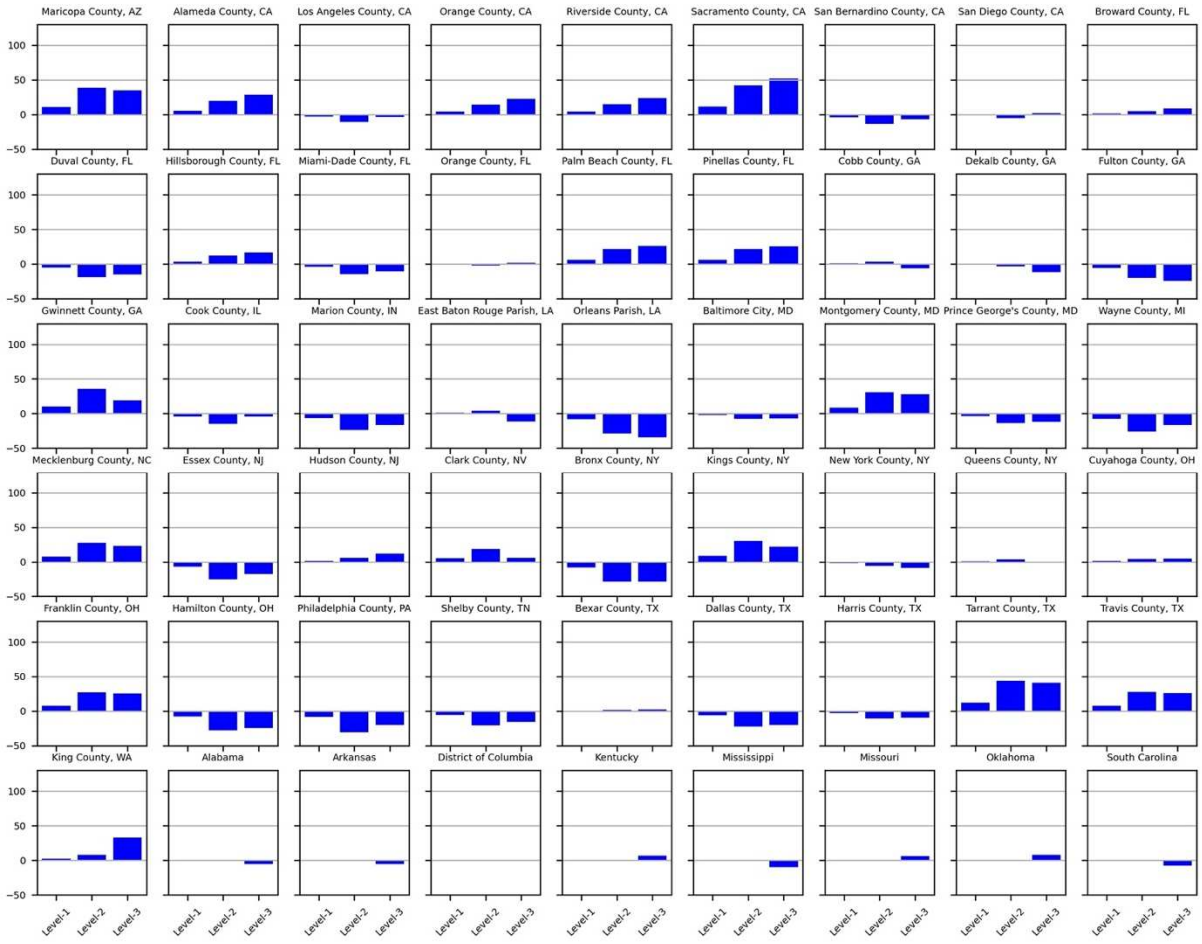


Figure I.3.a EHE jurisdiction; Comparing percentage difference of overall incidence between mixing scenarios for risk group MSM. Comparing no mixing (Scenario 13) with Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16) for each EHE jurisdiction for year 2018.

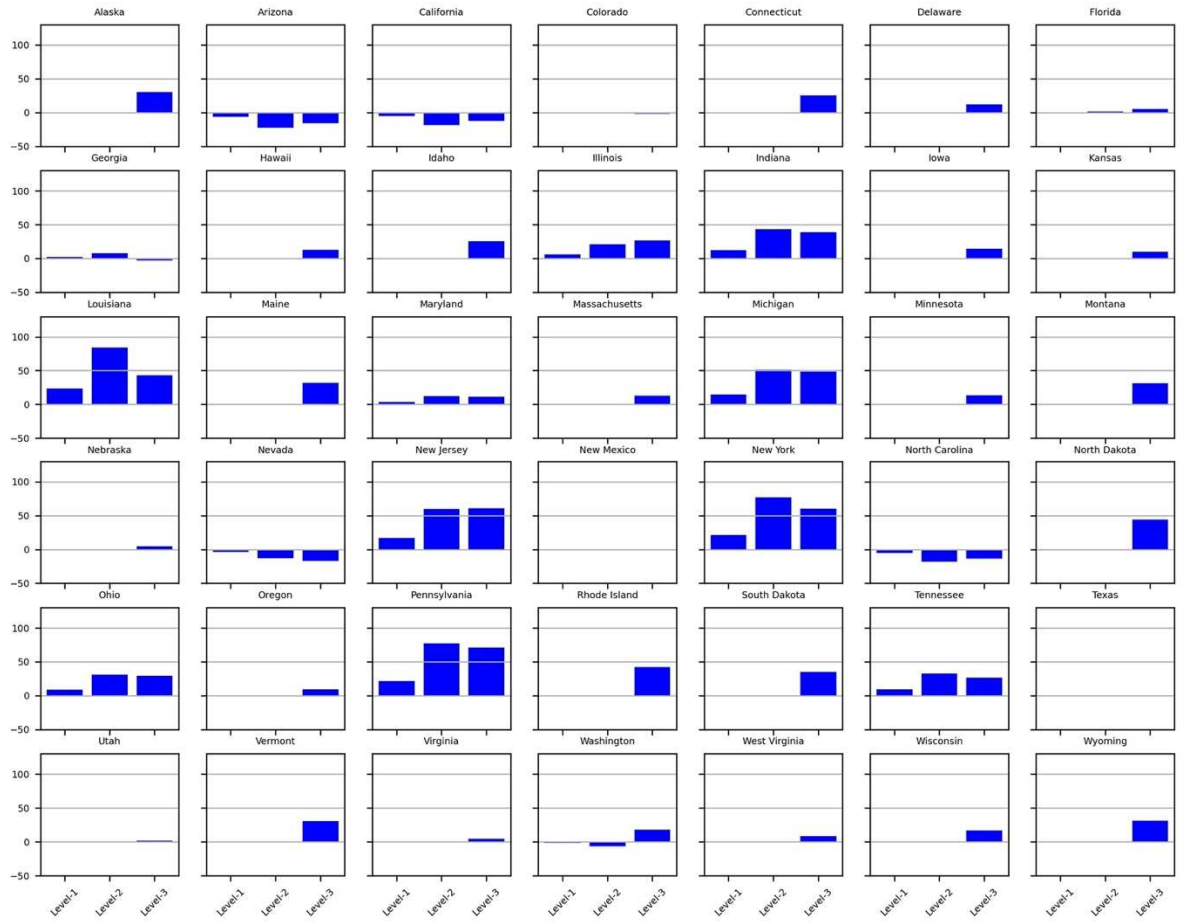


Figure I.3.b Non-EHE jurisdiction; Comparing percentage difference of overall incidence between mixing scenarios for risk group MSM. Comparing no mixing (Scenario 13) with Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16) for each non-EHE jurisdiction for year 2018.

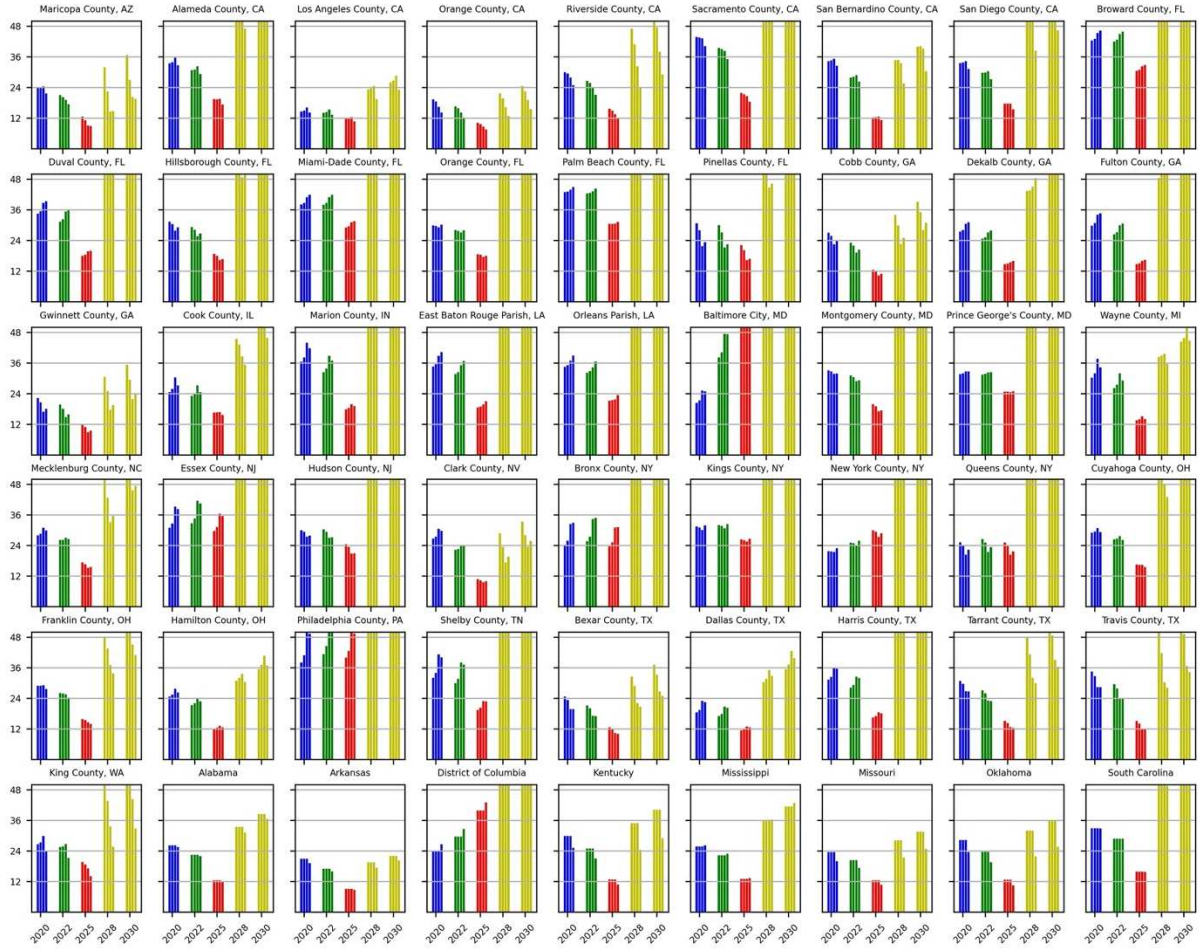


Figure I.4.a EHE jurisdictions; Comparing HIV interval between testing in months across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group HM. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for EHE jurisdictions applied in years 2019-2025. For each year the order of HIV interval between testing is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively. Testing interval limit on y-axis was capped at 4 years (48 months) as intervals during intervention years did not exceed 4 years.

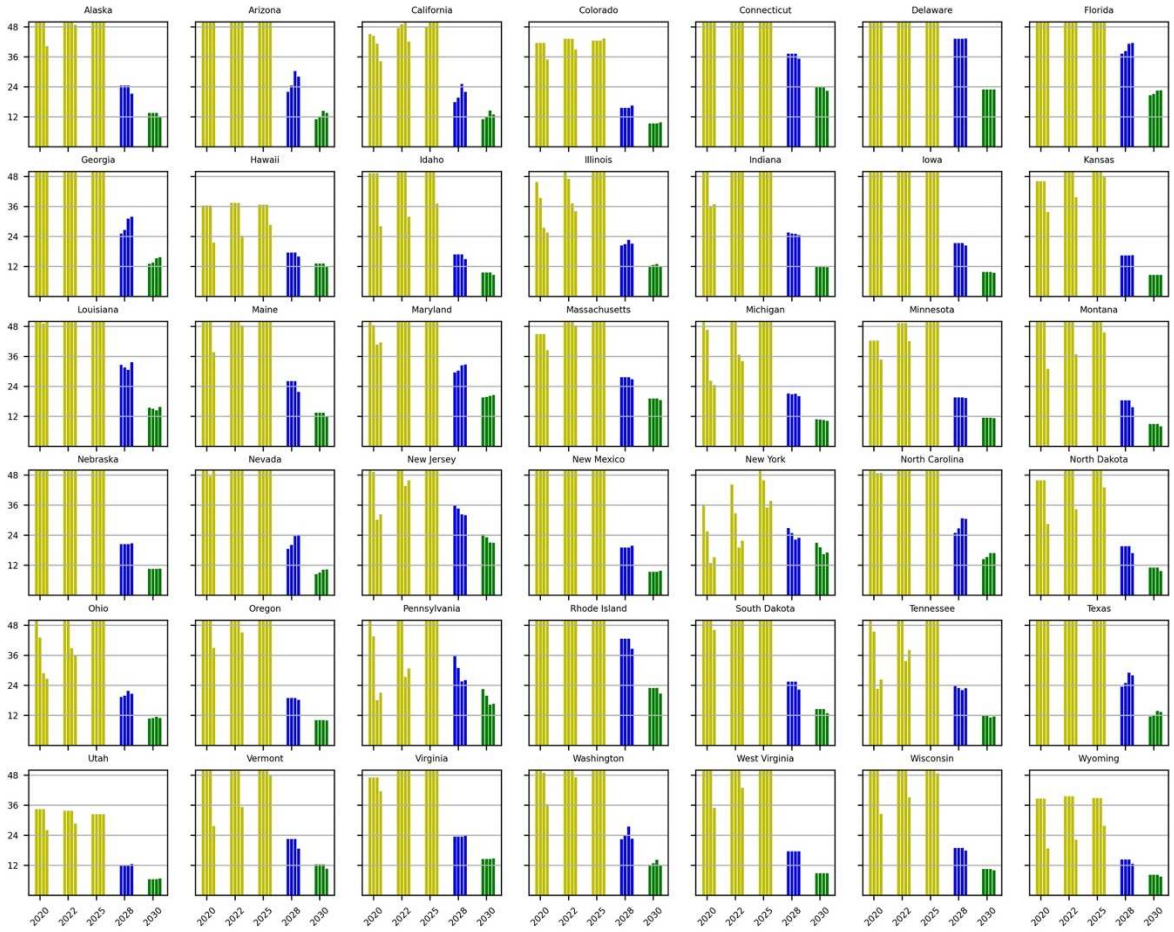


Figure I.4.b Non-EHE jurisdictions; Comparing HIV interval between testing in months across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group HM. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for non-EHE jurisdictions applied in years 2026-2030. For each year the order of HIV interval between testing is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively. Testing interval limit on y-axis was capped at 4 years (48 months) as intervals during intervention years did not exceed 4 years.

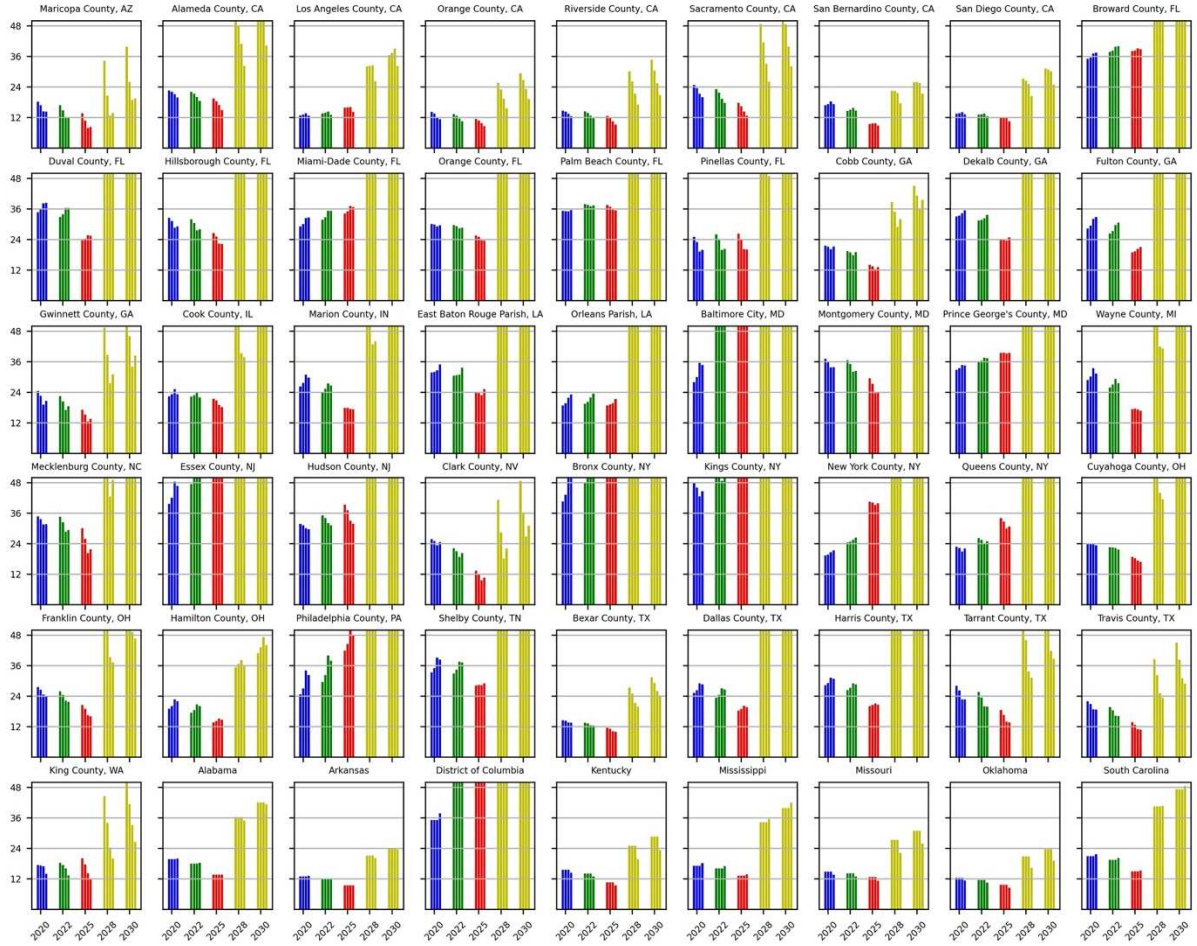


Figure I.5.a EHE jurisdictions; Comparing HIV interval between testing in months across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group HF. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for EHE jurisdictions applied in years 2019-2025. For each year the order of HIV interval between testing is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively. Testing interval limit on y-axis was capped at 4 years (48 months) as intervals during intervention years did not exceed 4 years.

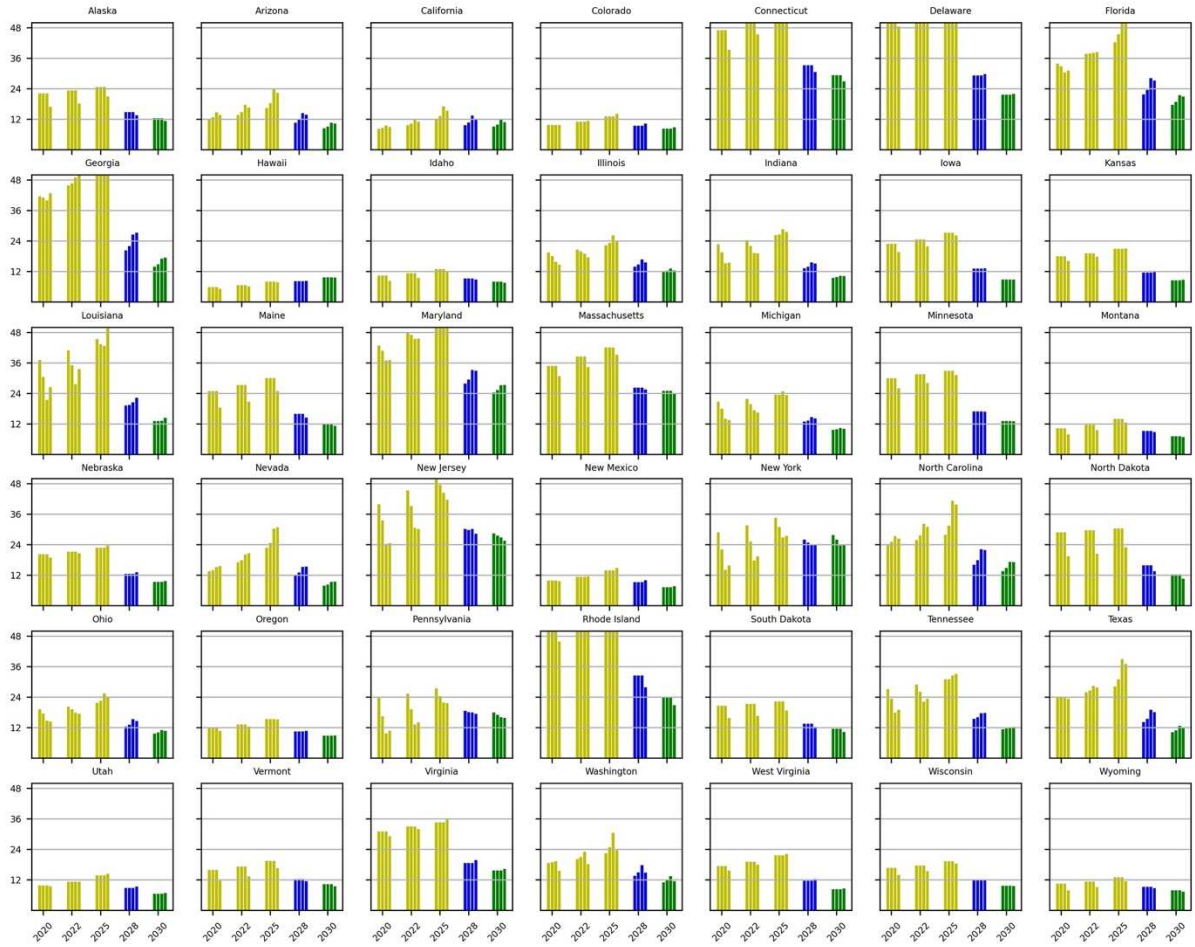


Figure I.5.b Non-EHE jurisdictions; Comparing HIV interval between testing in months across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group HF. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for non-EHE jurisdictions applied in years 2026-2030. For each year the order of HIV interval between testing is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively. Testing interval limit on y-axis was capped at 4 years (48 months) as intervals during intervention years did not exceed 4 years.

Figure I.6.a EHE jurisdictions; Comparing monthly retention in care rate across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group HM. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for EHE jurisdictions applied in years 2019-2025. For each year the order of retention in care is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively.

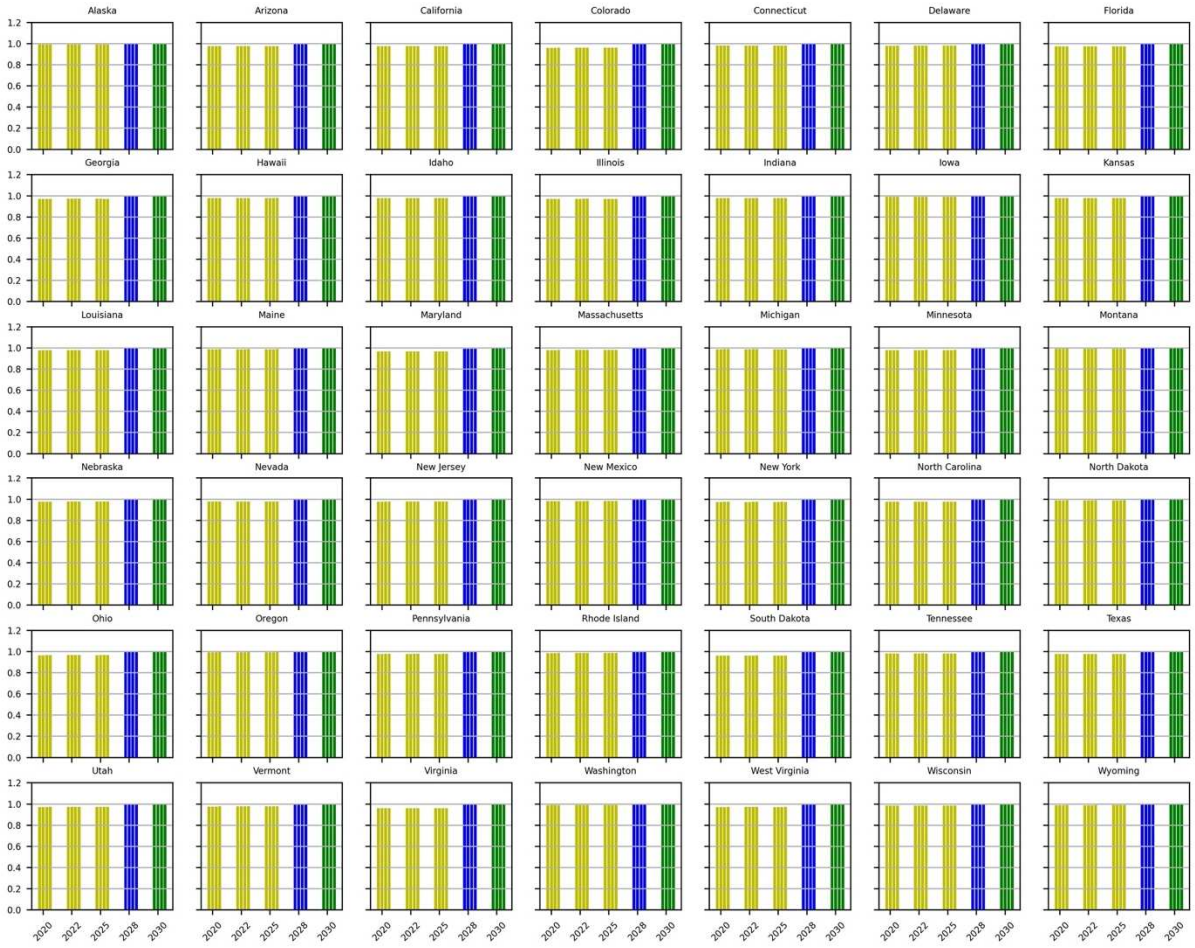


Figure I.6.b Non-EHE jurisdictions; Comparing monthly retention in care rate across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group HM. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for non-EHE jurisdictions applied in years 2026-2030. For each year the order of retention in care is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively.

Figure I.7.a EHE jurisdictions; Comparing monthly retention in care rate across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group HF. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for EHE jurisdictions applied in years 2019-2025. For each year the order of retention in care is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively.

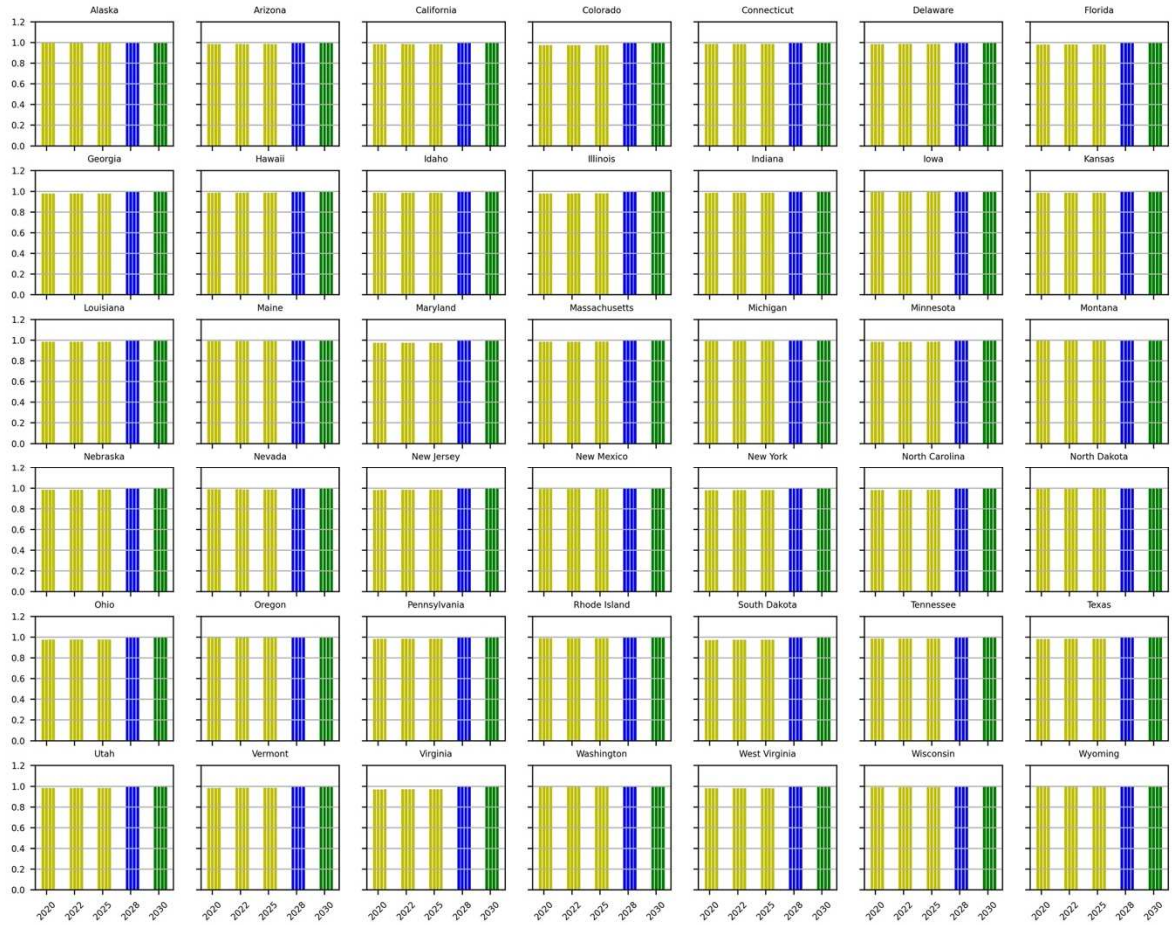


Figure I.7.b Non-EHE jurisdictions; Comparing monthly retention in care rate across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group HF. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for non-EHE jurisdictions applied in years 2026-2030. For each year the order of retention in care is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively.

Figure I.8.a EHE jurisdictions; Comparing monthly retention in care rate across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group MSM. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for EHE jurisdictions applied in years 2019-2025. For each year the order of retention in care is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively.

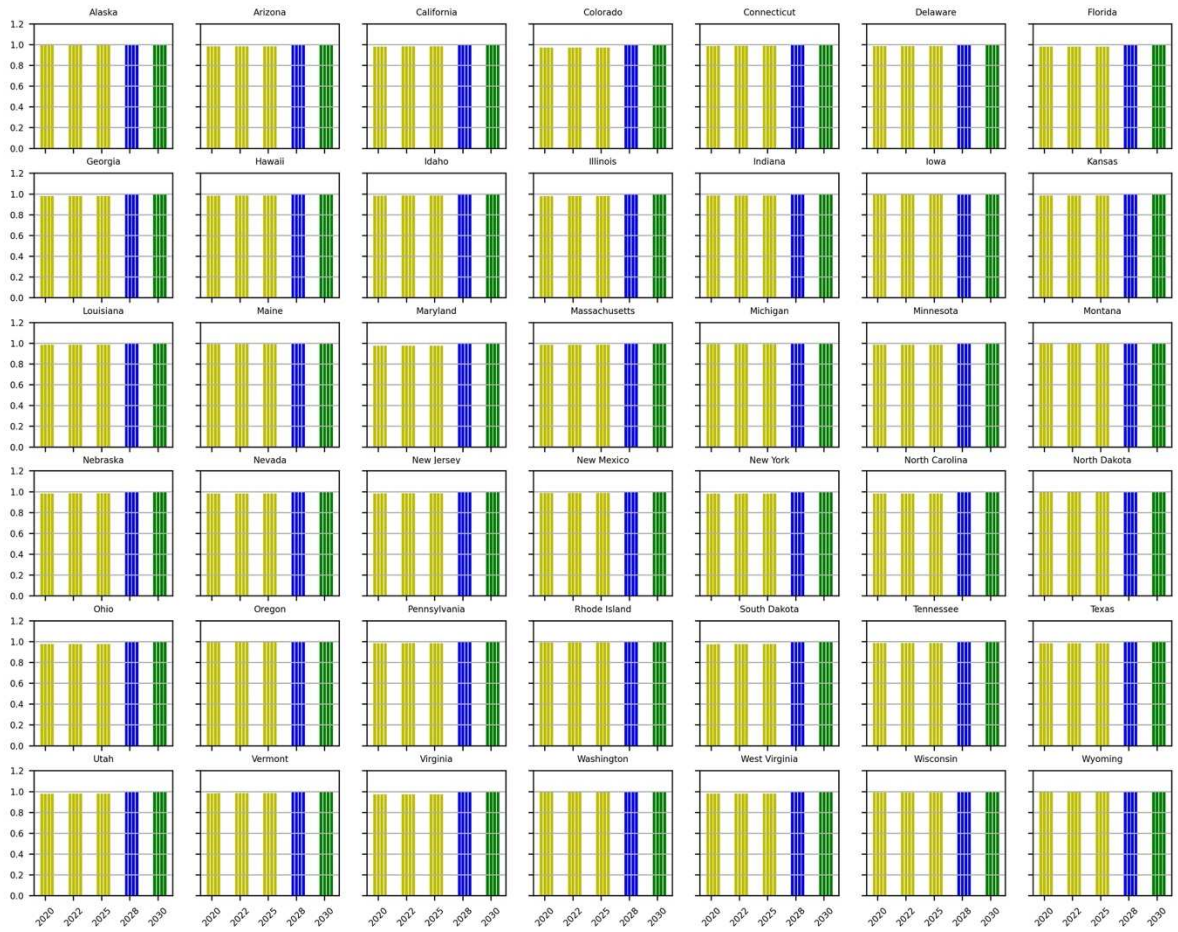


Figure I.8.b Non-EHE jurisdictions; Comparing monthly retention in care rate across years (2020, 2022, 2025, 2028, and 2030) for different mixing scenarios for risk group MSM. Comparing no mixing (Scenario 13), Level-1-mixing (Scenario 14), Level-2-mixing (Scenario 15), and Level-3-mixing (Scenario 16). Interventions for non-EHE jurisdictions applied in years 2026-2030. For each year the order of retention in care is no mixing, Level-1-mixing, Level-2-mixing, and Level-3-mixing, respectively.

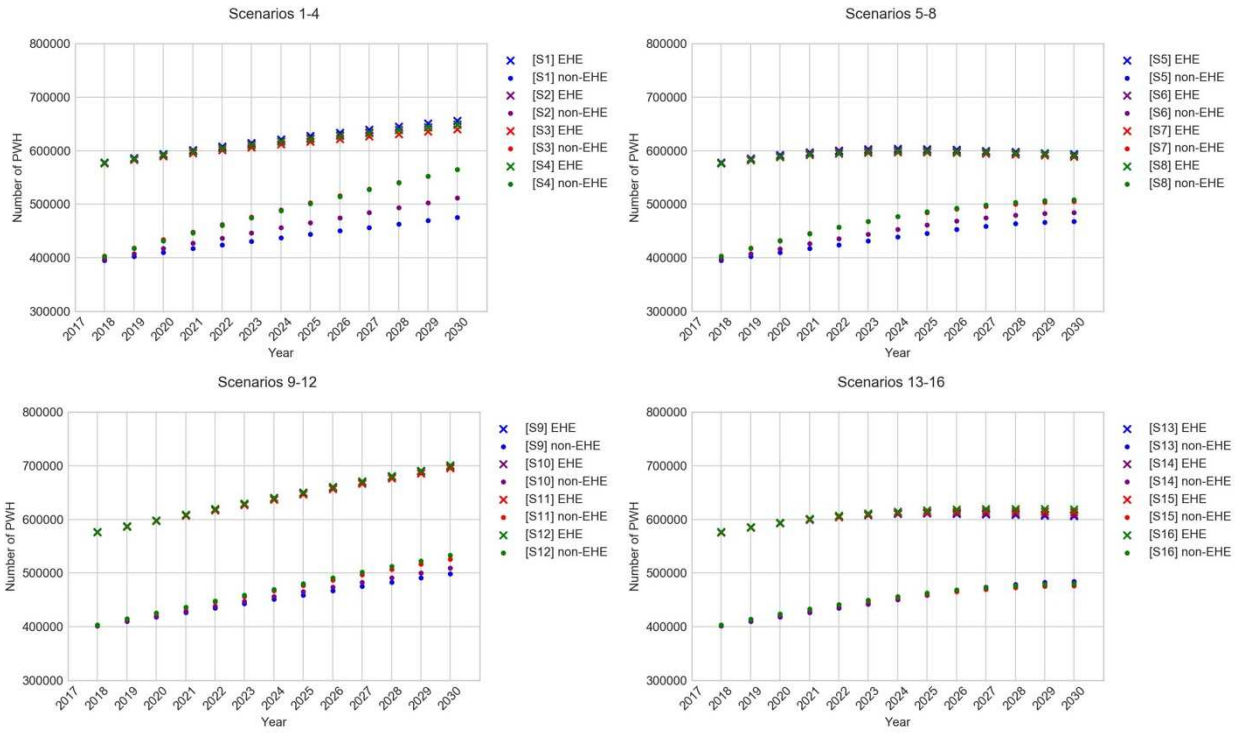


Figure I.9 Comparing annual PWH projections between the simulated 16 Jurisdictional-Model scenarios (S1 to S16) for aggregate EHE and non-EHE jurisdictions for the period 2018 to 2030.

Appendix J: Simulation Compartmental Model

We model HIV epidemic projections through simulation of a continuous-time non-stationary Markov process $\{X_t: \Omega_t, G_t; t = 1:T\}$. For the national model, the state space of the underlying Markov chain is given by $\Omega_{\text{nat}} = \{[S, C, D, De, A, R]\}$, which is a multivariate state. For a jurisdictional model, the state space of the underlying Markov chain is given by $\Omega_{\text{jur}} = \{[S, C, D, De, A, R, J]\}$.

Where, $S \in \{Susceptible\}$, $C \in \{Unaware, \text{Aware no ART}, ART\ no\ VLS, ART\ VLS\}$ representing the care continuum stage of an HIV infected person or susceptible, $D \in \{Acute, CD4 > 500, CD4\ 350 - 500,\ CD4\ 200 - 350, CD4 < 200\}$ representing the HIV disease progression stage of an HIV-infected person or uninfected, $De \in \{Death\}$, $A \in \{13, 14, \dots, 100\}$ representing all ages of a susceptible or HIV-infected person, $R \in \{HM, HF, MSM, \}$, representing risk group of a susceptible or HIV-infected person, and $J \in \{1, 2, \dots, 96\}$, representing the 96 jurisdictions modeled.

Every month people move from one compartment to other. If they move between compartments of care (i.e., C) it is called care continuum progression and if they move between compartments of the disease (i.e., D), it is called disease progression. There are 18 infected compartments, 1 susceptible, and 1 death compartment. Therefore, the simulation compartmental consists of a total of 20 compartments altogether. The matrix G_t represents non-stationary (i.e., varying over time) rate of transition between the states of care continuum and disease progression.

By applying Euler's numerical integration, we can simulate the epidemic over time using $\pi_{t+1} = \pi_t + \pi_t G_t \Delta t$

Where, π_{t+1} is a vector of size $|\Omega_{\text{nat}}|$ or $|\Omega_{\text{jur}}|$ with each element representing the number of people in that stage. We start the simulation with some initial values of π_t (here we used distribution of the U.S. population in 2010 year-end for the national model and 2017 year-end for the jurisdictional model as the initial value of π_t).

We present the transition rates that fill up the matrix G_t, which are obtained through literature studies, surveillance data, or estimated in our model to fit to combinations of π_{t+1} and π_t in Table H.2.

Appendix K: Numerical Estimation of Rates

As numerical estimation of rates is computationally expensive, we derive an analytical approximation of unknown rates by considering a collapsed/simplified state space of the Markov process discussed in Appendix J.

Let $\{Y_t: \bar{\Omega}, \bar{G}_t; t = 1:T\}$, be the Markov process with underlying Markov chain on a collapsed state space for each risk group (HM, HF, MSM). The new state space can be represented as $\bar{\Omega} = \{[\bar{C}, \mathrm{R}, \mathrm{J}]\}$.

Where $\bar{C} \in \{U, A, V\}$, $U = \bigcup_k \bigcup_j \{[U, D_d, A_a]\}$, $A = \bigcup_k \bigcup_j \{[Aware\ no\ ART, D_d, A_a]\}$, $V = \bigcup_k \bigcup_j \{[ART\ no\ VLS, D_d, A_a] \cup [ART\ VLS, D_d, A_a]\}$, $a \in \{13, 14, \ldots, 100\}$, and $d \in \{Acute, CD_4 > 500, CD_4\ 350 - 500,\ CD_4\ 200 - 350, CD_4 < 200\}$.

This method combines all ages and disease progression stages into one state. Further, note that it combines stages ART no VLS and ART VLS as V. π_t considers national data when implementing this method for the national model and π_t considers the jurisdiction specific data for J when implementing this method for the jurisdictional model. To calculate/calibrate the two unknown rates, diagnosis rate and dropout rate, for each risk group, we collapse the compartmental model seen in Figure 6.1 in the Manuscript into the following structure (see Figure K.1). At any time-step t (monthly), values for I_{t-1} and $p_{t-1,s}$ for each $s \in \{U, A, V\}$ are known as they are estimated at the previous time-step of simulation. We iteratively estimate:

- New infections using the Bernoulli equation in Appendix L
- Mortality as discussed in Appendix H using data presented in Tables H.5, H.6, and H.7

- Data for l_t and γ_t are available through literature data as discussed in Appendix H
- Diagnostic rate δ_t is unknown and calibrated along with number of new diagnoses $\delta_{t,risk} N_{t-1,U,d,risk} \theta_d$
- Dropout rate ρ_t is unknown and calibrated along with number of people who dropped out of care $\rho_t I_{t-1} p_{t-1,V}$
- We are only estimating the dropout rate for CD_4 count >200. For CD_4 count <200, we assume dropout is 0 and this is modeled by making φ_d = 0 for CD_4 count < 200.
- Finally, we apply the estimated rates for δ_t and ρ_t into the generator matrix G_t (rates in Table H.2) to estimate π_{t+1}.

Let, $dt = 1/12$ (monthly), $s =$ care continuum stage; $s \in \{U, A, V\}$; U = unaware; A= aware; V = prescribed ART (with VLS + no VLS), $d =$ disease progression stage; $d \in \{ Acute, CD_4 > 500, CD_4$ 350 to 500, CD_4 200 to 350, $CD_4 < 200\}$, I_t = total number of people living with HIV (PWH) (estimated prevalence), $p_{t,s}$= proportion of people in care-stage s at time t, $N_{t,s,d}$ = number of people in care-stage s and disease-stage d at time t, θ_d= scaling factor for diagnosis rate in disease-stage d, φ_d= scaling factor for drop-out rate in disease-stage d, i_t= new infections generated at time t, $m_{t,s,d}$ = number of deaths in the care-stage s and disease-stage d at time t, δ_t = diagnosis rate at time t, ρ_t = dropout rate at time t (dropout rate for CD_4 < 200 = 0, because <200 is AIDS so assume they will stay in care), and γ = re-entry rate (assumed 0.5 for CD_4 >= 200 and 1 for CD_4 < 200).

We estimate rates δ_t and ρ_t by calibrating to surveillance data for $p_{t,s} I_t$; $d \in \{U, A, V\}$. We estimate these rates specific to risk group for the National-model and specific to both risk group and jurisdiction for the Jurisdictional-Model but exclude this in notation for clarity. The

initial data for distribution of population in care stages, i.e., $p_{t,s}$ for both National-model and Jurisdictional-model is taken from NHSS data [5]. Projections for these proportions over time are estimated using the simulation model.

For a sufficiently small incremental time-step, $t + 1$ (we use monthly increments), we can write the equations for the number of people in each stage by formulating it as a system of differential equations.

$$p_{t+1,s}I_{t+1} = p_{t,s}I_t + \frac{dp_{t,s}I_t}{dt} \quad (1)$$

Where, $\frac{dp_{t,s}I}{dt}$ is the rate of the change in $p_{t,s}I$, i.e., the change in the number of infected persons in stage s at time t. We discuss below the derivation of δ_t and ρ_t.

K.1 Estimation of Diagnosis Rates

Expanding (1) for $s = U$ (Unaware stage),

$$I_t p_{t,U,risk} = I_{t-1}p_{t-1,U,risk} + i_{t,risk} - \delta_{t,risk}\sum_d \theta_d I_{t-1,}p_{t-1,U,risk,d} - \sum_d m_{t,U,risk}$$

$$\delta_{t,risk} = \frac{i_{t,risk}+I_{t-1}\,p_{t-1,U,risk}-I_t\,p_{t,U,risk}-\sum_d(m_{t,U,d,risk})}{\sum_d \theta_d I_{t-1}p_{t-1,U,risk,d}}$$

Computationally,

- $I_{t,risk} = I_{t-1,risk} + i_{t,risk} - \sum_d\left(m_{t,U,d,risk} + m_{t,A,d,risk} + m_{t,V,d,risk}\right)$
- $p_{t,U,risk} = p_{t-1,U,risk} + \frac{a_{U,T,risk}-a_{U,T-1,risk}}{1/dt} = \frac{\sum_d N_{t-1,U,d,risk}}{I_{t-1,risk}} + \frac{a_{U,T,risk}-a_{U,T-1,risk}}{1/dt}$
- diagnostic rate $\delta_{t,risk} = \frac{i_{t,risk}+\sum_d N_{t-1,U,d,risk}-I_t\,p_{t,U,risk}-\sum_d(m_{t,U,d,risk})}{\sum_d(N_{t-1,U,d,risk}\theta_d)}$; and
- corresponding number of people that are diagnosed are $\delta_{t,risk}\sum_d N_{t-1,U,d,risk}\theta_d$

K.2 Estimation of Dropout Rates

Expanding (1) for $s = V$ (prescribed ART),

$$I_t p_{t,V,risk} = I_{t-1} p_{t-1,V,risk} + \delta_{t,risk} l_{t,risk} \sum_d I_{t-1} p_{t-1,U,d,risk} \theta_d$$

$$+ \sum_d \gamma_d I_{t-1} p_{t-1,A,d,risk} - \rho_{t,risk} \sum_d I_{t-1} p_{t-1,V,d,risk} \varphi_d - \sum_d \left(m_{t,V,d,risk}\right)$$

$\rho_{t,risk}$

$$= \frac{I_{t-1} p_{t-1,V,risk} + \delta_{t,risk} l_{t,risk} \sum_d I_{t-1} p_{t-1,U,d,risk} \theta_d + \sum_d \gamma_d I_{t-1} p_{t-1,A,d,risk} - \sum_d \left(m_{t,V,d,risk}\right) - I_t p_{t,V,risk}}{\sum_d I_{t-1} p_{t-1,V,d,risk} \varphi_d}$$

Computationally, calculate

- $I_{t,risk} = I_{t-1,risk} + i_{t,risk} - \sum_d \left(m_{t,U,d,risk} + m_{t,A,d,risk} + m_{t,V,d,risk}\right)$
- $p_{t,V,risk} = p_{t-1,V,risk} + \frac{a_{V,T,risk} - a_{V,T-1,risk}}{1/dt} = \frac{\sum_d N_{t-1,V,d,risk}}{I_{t-1,risk}} + \frac{a_{V,T,risk} - a_{V,T-1,risk}}{1/dt}$
- drop-out rate $\rho_{t,risk} = \frac{\sum_d N_{t-1,V,d,risk} + \delta_{t,risk} l_{t,risk} \sum_d \left(N_{t-1,U,d,risk} \theta_d\right) + \sum_d \gamma_{t,d} N_{t-1,A,d,risk} - \sum_d \left(m_{t,V,d,risk}\right) - I_t p_{t,V,risk}}{\sum_d \left(N_{t-1,V,d,risk} \varphi_d\right)}$; and
- corresponding number of people who drop-out are $\rho_{t,risk} \sum_d \left(N_{t-1,V,d,risk} \varphi_d\right)$

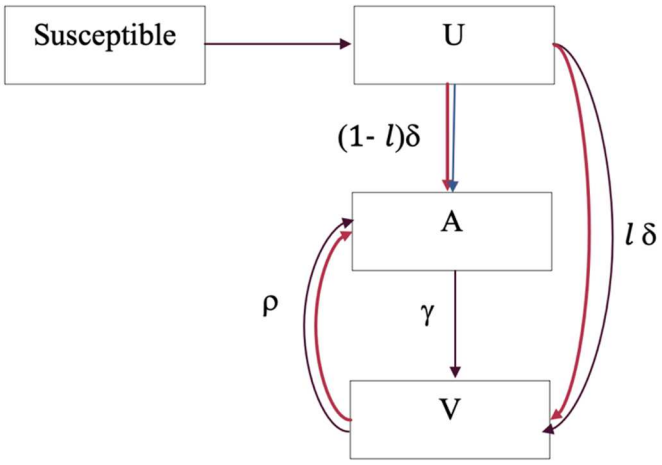


Figure K.1 Flow diagram for disease incidence and transition along the stages of care continuum. S: Population susceptible, U: population Unaware, A: population Aware no ART, V: population with ART no VLS and ART VLS, $\boldsymbol{\delta}$: diagnostic rate, $\boldsymbol{\gamma}$: rate of entering care and treatment among those not in care, and $\boldsymbol{\rho}$: rate of dropping out of care, and $\boldsymbol{l}$: proportion linked to care at diagnosis

Appendix L: Estimation of Incidence using Bernoulli Model

Transmission parameters estimate the rate at which people move from the susceptible compartment to the first infected compartment (i.e., Acute Unaware stage, see Figure K.1 in Manuscript). This is also defined as the number of new infections generated. To determine this, we develop a Bernoulli model, however, the method to estimate the number of new infections is slightly different for the National-model and Jurisdictional-model.

L.1 National Model

We apply the following Bernoulli equation for calculating the number of new infections generated for each susceptible person in age group x_1 and risk group y_1.

Number of new infections in risk group x_1 and age group y_1

$$= S_{x1,y1}[1 - \prod_{i=1}^{30}\{\mathrm{M}_{\mathrm{x}_1,\mathrm{y}_1,\mathrm{i}}\}^{q_{x1,y1,i}}] \quad (2)$$

Where,

$$M_{x_1,y_1,i} = \left\{1 - \left[1 - \left\{\left(1 - [\bar{p}_{v,y_1} p_i]\right)^{m_{v,c}}\left(1 - [p_{v,y_1} p_i]\right)^{m_{v,c'}}\left(1 - [\bar{p}_{a,y_1} p_i]\right)^{m_{a,c}}\left(1 - [p_{a,y_1} p_i]\right)^{m_{a,c'}}\right\}\right]\right\}$$

$$\bar{p}_{v,x1} = p_{v,x1}(1 - \epsilon)$$

$$\bar{p}_{a,x1} = p_{a,x1}(1 - \epsilon)$$

$$m_{v,c} = \frac{n_{v,x1,y1}((1-c_{x1,y1})c_i + c_{x1,y1})}{d_{x1,y1}}$$

$$m_{v,c'} = \frac{n_{v,x1,y1}(1-c_{x1,y1})(1-c_i)}{d_{x1,y1}}$$

$$m_{a,c} = \frac{n_{a,x1,y1}((1-c_{x1,y1})c_i + c_{x1,y1})}{d_{x1,y1}}$$

$$m_{a,c'} = \frac{n_{a,x1,y1}(1-c_{x1,y1})(1-c_i)}{d_{x1,y1}}$$

$$q_{x1,y1,i} = d_{x1,y1} \sum_{x2=1}^{3} \sum_{y2=13}^{100} risk_{x1,x2} age_{y1,y1} \frac{I_{x2,y2,i}}{N_{x2,y2}}$$

$$\left((1-c_{x1,y1})c_i + c_{x1,y1}\right) + (1-c_{x1,y1})(1-c_i) = 1$$

Let, $S_{x1,y1}$ = number of susceptible in risk group x_1, and age group y_1, $p_{v,x1}$ = probability of transmission for vaginal acts for risk group x_1, $p_{a,x1}$ = probability of transmission for anal acts for risk group x_1, ϵ = probability of condom effectiveness, p_i = multiplicative factor for increase in transmission probability based on infected compartment i, $n_{v,x1,y1}$ = number of vaginal acts for risk group x_1, and age group y_1, $n_{a,x1,y1}$ = number of anal acts for risk group x_1, and age group y_1, $c_{x1,y1}$ = proportion of condom use for risk group x_1, and age group y_1, c_i = proportion reduction in number of unprotected acts for risk group x_1, and age group y_1 for people in infected compartment i, $d_{x1,y1}$ = number of partners for risk group x_1, and age group y_1, $risk_{x1,x2}$ = risk specific mixing proportion between risk group x_1 and x_2, $age_{y1,y1}$ = age specific mixing proportion between age group y_1 and y_2, $I_{x2,y2,i}$ = number of infected in risk group x_2, age group y_2, and infected compartment i, and $N_{x2,y2}$ = number of people in risk group x_2, age group y_2.

Where, $x_1 and\ x_2 \in \{HM, HF, MSM\}$, $y_1 and\ y_2 \in \{13,14, \dots, 100\}$, and $i \in \{1, 2, \dots, 18\}$.

To calculate the total number of infections by the model we would sum the Bernoulli equation across all risk groups and age groups. Total number of new infections generated by the national model can be calculated as follows:

$$\sum_{x_1=1}^{3}\sum_{y_1=13}^{100} S_{x_1,y_1}\left(1-\prod_{i=1}^{18}\left\{M_{x_1,y_1,i}\right\}^{q_{x1,y1,i}}\right)$$

We used the national model to calibrate sexual parameters such as transmission probability (see Table H.8), sexual partnership mixing (see Table H.9), and risk-specific age-related mixing (see Table H.10), number of sexual acts (see Table H.11).

L.2 Jurisdictional Model

We apply a modified version of the Bernoulli equation presented for the national model to calculate the number of new infections generated for each susceptible person in age group x_1, risk group y_1 and jurisdiction j =

$$\sum_{x_1=1}^{3}\sum_{y_1=13}^{100} S_{x_1,y_1,j}\left(1-\prod_{i=1}^{18}\left\{M_{x_1,y_1,i}\right\}^{\sum_{\hat{\jmath}=j1}^{jn} mixing_{x_1,j,\hat{\jmath}}\, q_{\hat{\jmath},x_1,y_1,i}}\right)$$

Where, the components different from equation (2) are

$$q_{\hat{\jmath},x_1,y_1,i} = d_{x1,y1}\sum_{x2=1}^{3}\sum_{y2=13}^{100}\left(risk_{x1,x2}age_{y1,y1}\frac{I_{x_2,y_2,i,\hat{\jmath}}}{N_{x_2,y_2,\hat{\jmath}}}\right)$$

Let, $S_{x1,y1,j_1}$ = number of susceptible in risk group x_1, and age group y_1, and geographical jurisdiction j_1, $I_{x2,y2,i,\hat{\jmath}}$ = number of infected in risk group x_2, age group y_2, compartment i, and jurisdiction $\hat{\jmath}$, $N_{x2,y2,\hat{\jmath}}$ = number of people in risk group x_2, age group y_2, and jurisdiction $\hat{\jmath}$, $mixing_{x_1,j,\hat{\jmath}}$ = proportion of mixing of risk group x_1 located in jurisdiction j with PWH located in jurisdiction $\hat{\jmath}$, and $j\ and\ \hat{\jmath} \in \{j_1, j_2, \ldots, j_{96}\}$. The total number of new infections generated by the jurisdictional model can be calculated as follows:

$$= \sum_{j=j_1}^{j_{96}}\sum_{x_1=1}^{3}\sum_{y_1=13}^{100} S_{x_1,y_1,j_1}\left(1-\prod_{i=1}^{18}\left\{M_{x_1,y_1,i}\right\}^{\sum_{\hat{\jmath}=j_1}^{j_{96}} mixing_{x_1,j,\hat{\jmath}}\, q_{\hat{\jmath},x_1,y_1,i}}\right)$$