



Data-Driven Modeling for Different Stages of Pandemic Response

Aniruddha Adiga¹, Jiangzhuo Chen¹, Madhav Marathe^{1,3}, Henning Mortveit^{1,2},
Srinivasan Venkatramanan^{1*} and Anil Vullikanti^{1,3}

Abstract | Some of the key questions of interest during the COVID-19 pandemic (and all outbreaks) include: where did the disease start, how is it spreading, who are at risk, and how to control the spread. There are a large number of complex factors driving the spread of pandemics, and, as a result, multiple modeling techniques play an increasingly important role in shaping public policy and decision-making. As different countries and regions go through phases of the pandemic, the questions and data availability also change. Especially of interest is aligning model development and data collection to support response efforts at each stage of the pandemic. The COVID-19 pandemic has been unprecedented in terms of real-time collection and dissemination of a number of diverse datasets, ranging from disease outcomes, to mobility, behaviors, and socio-economic factors. The data sets have been critical from the perspective of disease modeling and analytics to support policymakers in real time. In this overview article, we survey the data landscape around COVID-19, with a focus on how such datasets have aided modeling and response through different stages so far in the pandemic. We also discuss some of the current challenges and the needs that will arise as we plan our way out of the pandemic.

1 Introduction

As the SARS-CoV-2 pandemic has demonstrated, the spread of a highly infectious disease is a complex dynamical process. A large number of factors are at play as infectious diseases spread, including variable individual susceptibility to the pathogen (e.g., by age and health conditions), variable individual behaviors (e.g., compliance with social distancing and the use of masks), differing response strategies implemented by governments (e.g., school and workplace closure policies and criteria for testing), and potential availability of pharmaceutical interventions. Governments have been forced to respond to the rapidly changing dynamics of the pandemic, and are becoming increasingly reliant on different modeling and analytical techniques to understand, forecast, plan, and respond; this includes

statistical methods and decision support methods using multi-agent models, such as: (i) forecasting epidemic outcomes (e.g., case counts, mortality, and hospital demands), using a diverse set of data-driven methods, e.g., ARIMA type time-series forecasting, Bayesian techniques and deep learning, e.g.,^{1–5}, (ii) disease surveillance^{6,7}, and (iii) counter-factual analysis of epidemics using multi-agent models, e.g.,^{8–13}; indeed, the results of Refs.^{11,14} were very influential in the early decisions for lockdowns in a number of countries.

The specific questions of interest change with the stage of the pandemic. In the pre-pandemic stage, the focus was on understanding how the outbreak started, epidemic parameters, and the risk of importation to different regions. Once outbreaks started the acceleration stage, the focus is on determining the growth rates, the

¹ Biocomplexity Institute and Initiative, Charlottesville, USA.

² Department of Systems Engineering and Environment, University of Virginia, Charlottesville, USA.

³ Department of Computer Science, University of Virginia, Charlottesville, USA.

*sriini@virginia.edu

differences in spatio-temporal characteristics, and testing bias. In the mitigation stage, the questions are focused on non-prophylactic interventions, such as school and work place closures and other social-distancing strategies, determining the demand for healthcare resources. In the suppression stage, the focus shifts to using prophylactic interventions, combined with better testing and tracing. These phases are not linear and overlap with each other. For instance, the acceleration and mitigation stages of the pandemic might overlap spatially, temporally as well as within certain social groups.

Different kinds of models are appropriate at different stages, and for addressing different kinds of questions. For instance, statistical and machine learning models are very useful in forecasting and short-term projections. However, they are not very effective for longer term projections, understanding the effects of different kinds of interventions, and counter-factual analysis. Mechanistic models are very useful for such questions. Simple compartmental-type models, and their extensions, namely, structured metapopulation models, are useful for several population-level questions. However, once the outbreak has spread, and complex individual- and community-level behaviors are at play, multi-agent models are most effective, since they allow for a more systematic representation of complex social interactions, individual and collective behavioral adaptation, and public policies.

As with any mathematical modeling effort, data play a big role in the utility of such models. Till recently, data on infectious diseases were very hard to obtain due to various issues, such as privacy and sensitivity of the data (since it is information about individual health), and logistics of collecting such data. The data landscape during the SARS-CoV-2 pandemic has been very different: a large number of datasets are becoming available, ranging from disease outcomes (e.g., time-series of the number of confirmed cases, deaths, and hospitalizations), some characteristics of their locations and demographics, healthcare infrastructure capacity (e.g., number of ICU beds, number of healthcare personnel, and ventilators), and various kinds of behaviors (e.g., level of social distancing, usage of PPEs); see Refs.^{15–17} for comprehensive surveys on available datasets.

However, using these datasets for developing good models, and addressing important public health questions remain challenging. The goal of this article is to use the widely accepted stages of a pandemic as a guiding framework

to highlight a few important problems that require attention in each of these stages. We will aim to provide a succinct model-agnostic formulation while identifying the key datasets needed, how they can be used, and the challenges arising in that process. We will also use SARS-CoV-2 as a case study unfolding in real time, and highlight some interesting peer-reviewed and preprint literature that pertains to each of these problems. An important point to note is the necessity of randomly sampled data, e.g., data needed to assess the number of active cases and various demographics of individuals that were affected. Census provides an excellent rationale. It is the only way one can develop rigorous estimates of various epidemiologically relevant quantities.

There have been numerous surveys on the different types of datasets available for SARS-CoV-2, e.g.,^{15–18}, as well as different kinds of modeling approaches. However, they do not describe how these models become relevant through the phases of pandemic response. An earlier similar attempt to summarize such response-driven modeling efforts can be found in Ref.¹⁹; based on the 2009–H1N1 experience, this paper builds on their work and discusses these phases in the present context and the SARS-CoV-2 pandemic. Although the paper touches upon different aspects of model-based decision-making, we refer the readers to a companion article in the same special issue²⁰ for a focused review of models used for projection and forecasting.

2 Background

Multiple organizations including CDC and WHO have their frameworks for preparing and planning response to a pandemic. For instance, the Pandemic Intervals Framework from CDC¹ describes the stages in the context of an influenza pandemic; these are illustrated in Fig. 1. These six stages span investigation, recognition, and initiation in the early phase, followed by most of the disease spread occurring during the acceleration and deceleration stages. They also provide indicators for identifying when the pandemic has progressed from one stage to the next²¹. As envisioned, risk evaluation [i.e., using tools like Influenza Risk Assessment Tool (IRAT) and Pandemic Severity Assessment Framework (PSAF)] and early case identification characterize the first three stages, while non-pharmaceutical

¹ <https://www.cdc.gov/flu/pandemic-resources/national-strategy/intervals-framework.html>.

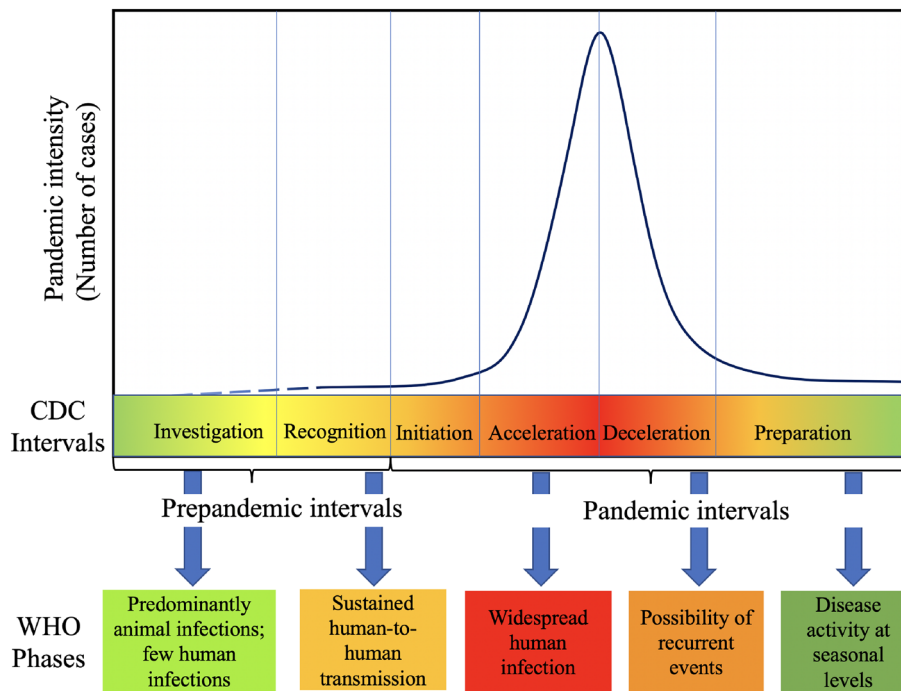


Figure 1: CDC Pandemic Intervals Framework and WHO phases for influenza pandemic.

interventions (NPIs) and available therapeutics become central to the acceleration stage. The deceleration is facilitated by mass vaccination programs, exhaustion of susceptible population, or unsuitability of environmental conditions (such as weather). A similar framework is laid out in WHO's pandemic continuum² and phases of pandemic alert³. While such frameworks aid in streamlining the response efforts of these organizations, they also enable effective messaging. To the best of our knowledge, there has not been a similar characterization of mathematical modeling efforts that go hand in hand with supporting the response.

3 Modeling for Stages of a Pandemic

For summarizing the key models, we consider four of the stages of pandemic response: pre-pandemic, acceleration, mitigation, and suppression. Here, we provide the key problems in each stage, the datasets needed, the main tools and techniques used, and pertinent challenges. We structure our discussion based on our experience with modeling the spread of COVID-19 in the

US, done in collaboration with local and federal agencies.

- Pre-pandemic (Sect. 4): in the initial time period, there are few human infections, and the key questions involve understanding the epidemiological parameters, and the risks of importation to different countries. The primary sources of data used in this stage include line lists, clinical investigations and prior literature on similar diseases, and mobility data such as airline flows, and information on travel restrictions.
- Acceleration (Sect. 5): this stage is relevant once the epidemic takes root within a country. There is usually a big lag in surveillance and response efforts, and the key questions are to model spread patterns at different spatio-temporal scales, and to derive short-term forecasts and projections. A broad class of datasets is used for developing models, including mobility, populations, land use, and activities. These are combined with various kinds of time-series data and covariates such as weather for forecasting.
- Mitigation (Sect. 6): in this stage, different interventions, which are mostly non-pharmaceutical in the case of a novel pathogen, are implemented by government agencies, once

² <https://www.cdc.gov/flu/pandemic-resources/planning-preparedness/global-planning-508.html>.

³ <https://www.who.int/csr/disease/swineflu/phase/en/>.

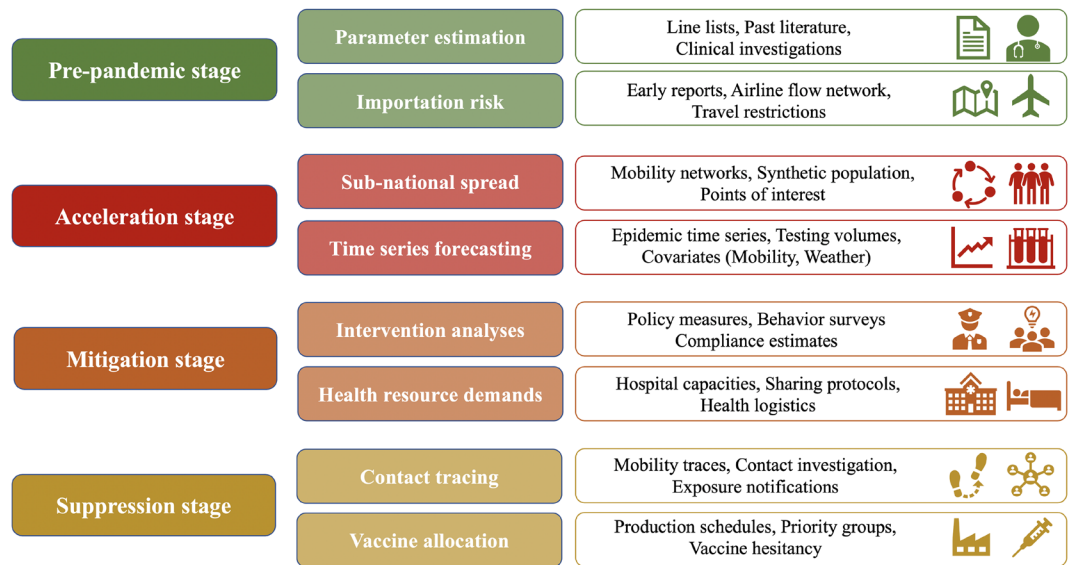


Figure 2: Summary of the data needs in different stages described in Sect. 3.

the outbreak has taken hold within the population. This stage involves understanding the impact of interventions on case counts and health infrastructure demands, taking individual behaviors into account. The additional datasets needed in this stage include those on behavioral changes and hospital capacities.

- **Suppression (Sect. 7):** this stage involves designing methods to control the outbreak by contact tracing and isolation and vaccination. Data on contact tracing, associated biases, vaccine production schedules, and compliance and hesitancy are needed in this stage.

Figure 2 gives an overview of this framework and summarizes the data needs in these stages. These stages also align well with the focus of the various modeling working groups organized by CDC which include epidemic parameter estimation, international spread risk, sub-national spread forecasting, impact of interventions, healthcare systems, and university modeling. In reality, one should note that these stages may overlap, and may vary based on geographical factors and response efforts. Moreover, specific problems can be approached prospectively in earlier stages, or retrospectively during later stages. This framework is thus meant to be more conceptual than interpreted along a linear timeline. Results from such stages are very useful for policymakers to guide real-time response.

4 Pre-pandemic Stage

Consider a novel pathogen emerging in human populations that is detected through early cases involving unusual symptoms or unknown etiology. Such outbreaks are characterized by some kind of spillover event, mostly through zoonotic means, like in the case of COVID-19 or past influenza pandemics (e.g., swine flu and avian flu). A similar scenario can occur when an incidence of a well-documented disease with no known vaccine or therapeutics emerges in some part of the world, causing severe outcomes or fatalities (e.g., Ebola and Zika.) Regardless of the development status of the country where the pathogen emerged, such outbreaks now the risk of causing a worldwide pandemic due to the global connectivity induced by human travel.

Two questions become relevant at this stage: what are the epidemiological attributes of this disease, and what are the risks of importation to a different country? While the first question involves biological and clinical investigations, the latter is more related to societal and environmental factors.

4.1 Epidemiological Parameter Estimation

One of the crucial tasks during early disease investigation is to ascertain the transmission and severity of the disease. These are important dimensions along which the pandemic potential is characterized, because together they determine the overall disease burden, as demonstrated

within the Pandemic Severity Assessment Framework²². In addition to risk assessment for right-sizing response, they are integral to developing meaningful disease models.

Formulation: Let $\Theta = \{\theta_T, \theta_S\}$ represent the transmission and severity parameters of interest. They can be further subdivided into sojourn time parameters θ^δ and transition probability parameters θ^p . Here, Θ corresponds to a continuous time Markov chain (CTMC) on the disease states. The problem formulation can be represented as follows:

Given $\Pi(\Theta)$, the prior distribution on the disease parameters and a dataset \mathcal{D} , estimate the posterior distribution $\mathbf{P}(\Theta|\mathcal{D})$ over all possible values of Θ . In a model-specific form, this can be expressed as $\mathbf{P}(\Theta|\mathcal{D}, \mathcal{M})$ where \mathcal{M} is a statistical, compartmental, or agent-based disease model.

Data needs: To estimate the disease parameters sufficiently, line lists for individual confirmed cases are ideal. Such datasets contain, for each record, the date of confirmation, possible date of onset, severity (hospitalization/ICU) status, and date of recovery/discharge/death. Furthermore, age and demographic/co-morbidity information allow development of models that are age- and risk-group stratified. One such crowd-sourced line list was compiled during the early stages of COVID-19²⁴ and later released by CDC for US cases²⁵. Data from detailed clinical investigations from other countries such as China, South Korea, and Singapore were also used to parameterize these models²⁶. In the absence of such datasets, past parameter estimates of similar diseases (e.g., SARS and MERS) were used for early analyses.

Modeling approaches: For a model-agnostic approach, the delays and probabilities are obtained by various techniques, including Bayesian and Ordinary Least Squares fitting to various delay distributions. For a particular disease model, these are estimated through model calibration techniques such as MCMC and particle filtering approaches. A summary of community estimates of various disease parameters is provided at <https://github.com/midas-network/COVID-19>. Further such estimates allow the design of pandemic planning scenarios varying in levels of impact, as seen in the CDC scenarios page⁴. See Refs.^{27–29} for methods and results related to estimating COVID-19 disease

parameters from real data. Current models use a large set of disease parameters for modeling COVID-19 dynamics; they can be broadly classified as transmission parameters and hospital resource parameters. For instance, in our work, we currently use parameters (with explanations), as shown in Table 1.

Challenges: Often, these parameters are model-specific, and hence, one needs to be careful when reusing parameter estimates from literature. They are related but not identifiable with respect to population-level measures such as basic reproductive number R_0 (or effective reproductive number R_{eff}) and doubling time which allow tracking the rate of epidemic growth. Also the estimation is hindered by inherent biases in case ascertainment rate, reporting delays and other gaps in the surveillance system. Aligning different data streams (e.g., outpatient surveillance, hospitalization rates, and mortality records) is in itself challenging.

4.2 International Importation Risk

When a disease outbreak occurs in some part of the world, it is imperative for most countries to estimate their risk of importation through spatial proximity or international travel. Such measures are incredibly valuable in setting a timeline for preparation efforts, and initiating health checks at the borders. Over centuries, pandemics have spread faster and faster across the globe, making it all the more important to characterize this risk as early as possible.

Formulation: Let \mathcal{C} be the set of countries, and $\mathcal{G} = \{\mathcal{C}, \mathcal{E}\}$ an international network, where edges (often weighted and directed) in \mathcal{E} represent some notion of connectivity. The importation risk problem can be formulated as below:

Given $C_o \in \mathcal{C}$ the country of origin with an initial case at time 0, and C_i the country of interest, using \mathcal{G} , estimate the expected time taken T_i for the first cases to arrive in country C_i .

In its probabilistic form, the same can be expressed as estimating the probability $P_i(t)$ of seeing the first case in country C_i by time t .

Data needs: Assuming we have initial case reports from the origin country, the first data needed is a network that connects the countries of the world to represent human travel. The most common source of such information is the airline network datasets, from sources such as IATA, OAG, and OpenFlights; Marie Isabelle et al.³⁰ provides

⁴ <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html>.

Table 1: COVID-19-specific parameters that we currently use in our modeling and studies.

Parameter	Values	Description
Transmissibility (R_0)	2.5 [2.0, 3.0]	Basic reproduction number
Incubation period	5 days	Time from infection to onset
Latent period	3–5 days	Time from infection to infectious
Percent symptomatic	65%	Infected people that exhibit symptoms
Infectious period	5–6 days	Duration of infectiousness
Infection detection rate	15%	One confirmed case corresponds to 7 cases
Serial interval	7 days	Time from infection to next generation infection
Onset to hospitalization	6.2 days	Time from symptoms to hospitalization
Hospitalization to ventilation	1–2 days	Time from hospitalization to ventilation
Duration hospitalized	3–8 days	Time spent in the hospital
Duration ventilated	2–7 days	Time spent on a ventilator
Percent hospitalized	5.5%	Symptomatic individuals becoming hospitalized
Percent ventilated	13%	Hospitalized patients that require ventilation

Please note that the estimated values evolve in time; the values in the table are based on the best guess 2020-04-14 version of “COVID-19 Pandemic Planning Scenarios” document prepared by the Centers for Disease Control and Prevention (CDC) SARS-CoV-2 Modeling Team²³

a systematic review of how airline passenger data have been used for infectious disease modeling. These datasets could either capture static measures such as number of seats available or flight schedules, or a dynamic count of passengers per month along each itinerary. Since the latter has intrinsic delays in collection and reporting, for an ongoing pandemic, they may not be representative. During such times, data on ongoing travel restrictions³¹ become important to incorporate. Multi-modal traffic will also be important to incorporate for countries that share land borders or have heavy maritime traffic. For diseases such as Zika, where establishment risk is more relevant, data on vector abundance or prevailing weather conditions are appropriate.

Modeling approaches: Simple structural measures on networks (such as degree, PageRank) could provide static indicators of vulnerability of countries. By transforming the weighted, directed edges into probabilities, one can use simple contagion models (e.g., Independent Cascades) to simulate disease spread and empirically estimate expected time of arrival. Global metapopulation models (GLEaM) that combine SEIR type dynamics with an airline network have also been used in the past for estimating importation risk. Brockmann and Helbing³² used a similar framework to quantify effective distance on the network which seemed to be well correlated

with time of arrival for multiple pandemics in the past; this has been extended to COVID-19^{8, 33}. In Ref.³⁴, the authors employ air travel volume obtained through IATA from ten major cities across China to rank various countries along with the IDVI to convey their vulnerability. Wu et al.³⁵ consider the task of forecasting international and domestic spread of COVID-19 and employ Official Airline Group (OAG) data for determining air traffic to various countries, and De Salazae et al.³⁶ fit a generalized linear model for observed number of cases in various countries as a function of air traffic volume obtained from OAG data to determine countries with potential risk of under-detection. Also, Gilbert et al.³⁷ provide Africa-specific case study of vulnerability and preparedness using data from Civil Aviation Administration of China.

Challenges: Note that arrival of an infected traveler will precede a local transmission event in a country. Hence, the former is more appropriate to quantify in early stages. Also, the formulation is agnostic to whether it is the first infected arrival or first detected case. However, in real world, the former is difficult to observe, while the latter is influenced by security measures at ports of entry (land, sea, and air) and the ease of identification for the pathogen. For instance, in the case of COVID-19, the long incubation period and the high likelihood of asymptomaticity could have

resulted in many infected travelers being missed by health checks at PoEs. We also noticed potential administrative delays in reporting by multiple countries fearing travel restrictions.

5 Acceleration Stage

As the epidemic takes root within a country, it may enter the acceleration phase. Depending on the testing infrastructure and agility of surveillance system, response efforts might lag or lead the rapid growth in case rate. Under such a scenario, two crucial questions emerge that pertain to how the disease may spread spatially/socially and how the case rate may grow over time.

5.1 Sub-national Spread Across Scales

Within the country, there is need to model the spatial spread of the disease at different scales: state, county, and community levels. Similar to the importation risk, such models may provide an estimate of when cases may emerge in different parts of the country. When coupled with vulnerability indicators (socio-economic, demographic, and co-morbidities) they provide a framework for assessing the heterogeneous impact the disease may have across the country. Detailed agent-based models for urban centers may help identify hotspots and potential case clusters that may emerge (e.g., correctional facilities, nursing homes, food processing plants, etc. in the case of COVID-19).

Formulation: Given a population representation \mathcal{P} at appropriate scale and a disease model \mathcal{M} per entity (individual or sub-region), model the disease spread under different assumptions of underlying connectivity \mathcal{C} and disease parameters Θ . The result will be a spatio-temporal spread model that results in $Z_{s,t}$, the time-series of disease states over time for region s .

Data needs: Some of the common datasets needed by most modeling approaches include: (1) social and spatial representation, which includes Census, and population data, which are available from Census departments (see, e.g.,³⁸), and Landscan³⁹, (2) connectivity between regions (commuter, airline, road/rail/river), e.g.,^{30,31}, (3) data on locations, including points of interest, e.g., OpenStreetMap⁴⁰, and (4) activity data, e.g., the American Time Use Survey⁴¹. These datasets help capture where people reside and how they move around, and come in contact with each other. While some of these are static, more

dynamic measures, such as GPS traces, become relevant as individuals change their behavior during a pandemic.

Modeling approaches: Different kinds of structured metapopulation models^{8, 42–45} and agent-based models^{46–48} have been used in the past to model the sub-national spread; we refer to Refs.^{13, 49, 50} for surveys on different modeling approaches. These models incorporate typical mixing patterns, which result from detailed activities and co-location (in the case of agent-based models), and different modes of travel and commuting (in the case of metapopulation models).

Challenges: While metapopulation models can be built relatively rapidly, agent-based models are much harder—the datasets need to be assembled at a large scale, with detailed construction pipelines, see, e.g.,^{46–48}. Since detailed individual activities drive the dynamics in agent-based models, schools and workplaces have to be modeled, to make predictions meaningful. Such models will get reused at different stages of the outbreak, so they need to be generic enough to incorporate dynamically evolving disease information. Finally, a common challenge across modeling paradigms is the ability to calibrate the model to the dynamically evolving spatio-temporal data from the outbreak—this is especially challenging in the presence of reporting biases and data insufficiency issues.

5.2 Growth Rate and Time-Series Forecasting

Given the early growth of cases within the country (or sub-region), there is need for quantifying the rate of increase in comparable terms across the duration of the outbreak (accounting for the exponential nature of such processes). These estimates also serve as references when evaluating the impact of various interventions. As an extension, such methods and more sophisticated time-series methods can be used to produce short-term forecasts for disease evolution.

Formulation: Given the disease time-series data within the country $Z_{s,t}$ until data horizon T , provide scale-independent growth rate measures $G_s(T)$, and forecasts $\hat{Z}_{s,u}$ for $u \in [T, T + \Delta T]$, where ΔT is the forecast horizon.

Data needs: Models at this stage require datasets such as (1) time-series data on different kinds of disease outcomes, including case counts,

mortality, hospitalizations, along with attributes, such as age, gender, and location, e.g.,^{51–55}, (2) any associated data for reporting bias (total tests and test positivity rate)⁵⁶, which need to be incorporated into the models, as these biases can have a significant impact on the dynamics, and (3) exogenous covariates (mobility and weather), which have been shown to have a significant impact on other diseases, such as Influenza, e.g.,⁵⁷.

Modeling approaches: Even before building statistical or mechanistic time-series forecasting methods, one can derive insights through analytical measures of the time-series data. For instance, the effective reproductive number, estimated from the time-series⁵⁸, can serve as a scale-independent metric to compare the outbreaks across space and time. Additionally multiple statistical methods ranging from autoregressive models to deep learning techniques can be applied to the time-series data, with additional exogenous variables as input. While such methods perform reasonably for short-term targets, mechanistic approaches as described earlier can provide better long-term projections. Various ensembling techniques have also been developed in the recent past to combine such multi-model forecasts to provide a single robust forecast with better uncertainty quantification. One such effort that combines more than 30 methods for COVID-19 can be found at the COVID Forecasting Hub⁵. We also point to the companion paper for more details on projection and forecasting models.

Challenges: Data on epidemic outcomes usually have a lot of uncertainties and errors, including missing data, collection bias, and backfill. For forecasting tasks, these time-series data need to be near real time, else one needs to do both nowcasting, as well as forecasting. Other exogenous regressors can provide valuable lead time, due to inherent delays in disease dynamics from exposure to case identification. Such frameworks need to be generalized to accommodate qualitative inputs on future policies (shutdowns, mask mandates, etc.), as well as behaviors, as we discuss in the next section.

6 Mitigation Stage

Once the outbreak has taken hold within the population, local, state and national governments attempt to mitigate and control its spread by considering different kinds of interventions. Unfortunately, as the COVID-19 pandemic has

shown, there is a significant delay in the time taken by governments to respond. As a result, this has caused a large number of cases, a fraction of which lead to hospitalizations. Two key questions in this stage are: (1) how to evaluate different kinds of interventions, and choose the most effective ones, and (2) how to estimate the healthcare infrastructure demand, and how to mitigate it. The effectiveness of an intervention (e.g., social distancing) depends on how individuals respond to them, and the level of compliance. The healthcare resource demand depends on the specific interventions which are implemented. As a result, both these questions are connected, and require models which incorporate appropriate behavioral responses.

6.1 Intervention Analyses

In the initial stages, only non-prophylactic interventions are available, such as: social distancing, school and workplace closures, and use of PPEs, since no vaccines and anti-virals are available. As mentioned above, such analyses are almost entirely model-based, and the specific model depends on the nature of the intervention and the population being studied.

Formulation: Given an abstract model, \mathcal{M} , the general goals are (1) to evaluate the impact of an intervention (e.g., school and workplace closure, and other social distancing strategies) on different epidemic outcomes (e.g., average outbreak size, peak size, and time to peak), and (2) find the most effective intervention from a suite of interventions, with given resource constraints. The specific formulation depends crucially on the model and type of intervention. Even for a single intervention, evaluating its impact is quite challenging, since there are a number of sources of uncertainty, and a number of parameters associated with the intervention (e.g., when to start school closure, how long, and how to restart). Therefore, finding uncertainty bounds is a key part of the problem.

Data needs: While all the data needs from the previous stages for developing a model are still there, representation of different kinds of behaviors is a crucial component of the models in this stage; this includes: use of PPEs, compliance to social distancing measures, and level of mobility. Statistics on such behaviors are available at

⁵ <https://covid19forecasthub.org/>.

a fairly detailed level (e.g., counties and daily) from multiple sources, such as (1) the COVID-19 Impact Analysis Platform from the University of Maryland⁵⁴, which gives metrics related to social distancing activities, including level of staying home, out of county trips, and out of state trips, (2) changes in mobility associated with different kinds of activities from Google⁵⁹, and other sources, (3) survey data on different kinds of behaviors, such as usage of masks⁶⁰.

Modeling approaches: As mentioned above, such analyses are almost entirely model-based, including structured metapopulation models^{8, 42–45}, and agent-based models^{46–48}. Different kinds of behaviors relevant to such interventions, including compliance with using PPEs and compliance to social distancing guidelines, need to be incorporated into these models. Since there is a great deal of heterogeneity in such behaviors, it is conceptually easiest to incorporate them into agent-based models, since individual agents are represented. However, calibration, simulation, and analysis of such models pose significant computational challenges. On the other hand, the simulation of metapopulation models is much easier, but such behaviors cannot be directly represented—instead, modelers have to estimate the effect of different behaviors on the disease model parameters, which can pose modeling challenges.

Challenges: There are a number of challenges in using data on behaviors, which depends on the specific datasets. Much of the data available for COVID-19 are estimated through indirect sources, e.g., through cell phone and online activities, and crowd-sourced platforms. This can provide large spatio-temporal datasets, but has unknown biases and uncertainties. On the other hand, survey data are often more reliable, and provides several covariates, but is typically very sparse. Handling such uncertainties, rigorous sensitivity analysis, and incorporating the uncertainties into the analysis of the simulation outputs are important steps for modelers.

6.2 Healthcare Resource Demands

The COVID-19 pandemic has led to a significant increase in hospitalizations. Hospitals are typically optimized to run near capacity, so there have been fears that the hospital capacities would not be adequate, especially in several countries in Asia, but also in some regions in the US. Nosocomial transmission could further increase this burden.

Formulation: The overall problem is to estimate the demand for hospital resources within a population—this includes the number of hospitalizations, and more refined types of resources, such as ICUs, CCUs, medical personnel, and equipment, such as ventilators. An important issue is whether the capacity of hospitals within the region would be overrun by the demand, when this is expected to happen, and how to design strategies to meet the demand—this could be through augmenting the capacities at existing hospitals, or building new facilities. Timing is of essence, and projections of when the demands exceed capacity are important for governments to plan.

Data needs: The demands for hospitalization and other healthcare resources can be estimated from the epidemic models mentioned earlier, by incorporating suitable health states, e.g.,^{43, 61}; in addition to the inputs needed for setting up the models for case counts, datasets are needed for hospitalization rates and durations of hospital stay, ICU care, and ventilation. The other important inputs for this component are hospital capacity, and the referral regions (which represent where patients travel for hospitalization). Different public and commercial datasets provide such information, e.g.,^{62, 63}.

Modeling approaches: Demand for healthcare resources is typically incorporated into both metapopulation and agent-based models, by having a fraction of the infectious individuals transition into a hospitalization state. An important issue to consider is what happens if there is a shortage of hospital capacity. Studying this requires modeling the hospital infrastructure, i.e., different kinds of hospitals within the region, and which hospital a patient goes to. There is typically limited data on this, and data on hospital referral regions, or Voronoi tessellation can be used. Understanding the regimes in which hospital demand exceeds capacity is an important question to study. Nosocomial transmission is typically much harder to study, since it requires more detailed modeling of processes within hospitals.

Challenges: There is a lot of uncertainty and variability in all the datasets involved in this process, making its modeling difficult. For instance, forecasts of the number of cases and hospitalizations have huge uncertainty bounds for medium- or long-term horizon, which is the kind of input necessary for understanding hospital demands, and whether there would be any deficits.

7 Suppression Stage

The suppression stage involves methods to control the outbreak, including reducing the incidence rate and potentially leading to the eradication of the disease in the end. Eradication in case of COVID-19 appears unlikely as of now; what is more likely is that this will become part of seasonal human coronaviruses that will mutate continuously much like the influenza virus.

7.1 Contact Tracing

Contact tracing problem refers to the ability to trace the neighbors of an infected individual. Ideally, if one is successful, each neighbor of an infected neighbor would be identified and isolated from the larger population to reduce the growth of a pandemic. In some cases, each such neighbor could be tested to see if the individual has contracted the disease. Contact tracing is the workhorse in epidemiology and has been immensely successful in controlling slow moving diseases. When combined with vaccination and other pharmaceutical interventions, it provides the best way to control and suppress an epidemic.

Formulation: The basic contact tracing problem is stated as follows: Given a social contact network $G(V, E)$ and subset of nodes $S \subset V$ that are infected and a subset $S_1 \subset S$ of nodes identified as infected, find all neighbors of S . Here, a neighbor means an individual who is likely to have a substantial contact with the infected person. One then tests them (if tests are available), and following that, isolates these neighbors, or vaccinates them or administers anti-viral. The measures of effectiveness for the problem include: (i) maximizing the size of S_1 , (ii) maximizing the size of set $N(S_1) \subseteq N(S)$, i.e., the potential number of neighbors of set S_1 , (iii) doing this within a short period of time, so that these neighbors either do not become infectious, or they minimize the number of days that they are infectious, while they are still interacting in the community in a normal manner, (iv) the eventual goal is to try and reduce the incidence rate in the community—thus if all the neighbors of S_1 cannot be identified, one aims to identify those individuals who when isolated/treated lead to a large impact; (v) and finally verifying that these individuals indeed came in contact with the infected individuals and thus can be asked to isolate or be treated.

Data needs: Data needed for the contact tracing problem include: (i) a line list of individuals who are currently known to be infected (this is needed

in case of human-based contact tracing). In the real world, when carrying out human contact tracer-based deployment, one interviews all the individuals who are known to be infectious and reaches out to their contacts.

Modeling approaches: Human contact tracing is routinely done in epidemiology. Most states in the US have hired such contact tracers. They obtain the daily incidence report from the state health departments and then proceed to contact the individuals who are confirmed to be infected. Earlier, human contact tracers used to go from house to house and identify the potential neighbors through a well-defined interview process. Although very effective it is very time-consuming and labor intensive. Phones were used extensively in the last 10–20 years as they allow the contact tracers to reach individuals. They are helpful- but have the downside that it might be hard to reach all individuals. During COVID-19 outbreak, for the first time, societies and governments have considered and deployed digital contact tracing tools^{64–68}. These can be quite effective but also have certain weaknesses, including, privacy, accuracy, and limited market penetration of the digital apps.

Challenges: These include: (i) inability to identify everyone who is infectious (the set S)—this is virtually impossible for COVID-19 like disease unless the incidence rate has come down drastically and for the reason that many individuals are infected but asymptomatic; (ii) identifying all contacts of S (or S_1)—this is hard, since individuals cannot recall everyone they met, certain folks that they were in close proximity might have been in stores or social events and thus not known to individuals in the set S . Furthermore, even if a person is able to identify the contacts, it is often hard to reach all the individuals due to resource constraints (each human tracer can only contact a small number of individuals).

7.2 Vaccine Allocation

The overall goal of the vaccine allocation problem is to allocate vaccine efficiently and in a timely manner to reduce the overall burden of the pandemic.

Formulation: The basic version of the problem can be cast in a very simple manner (for networked models): Given a graph $G(V, E)$ and a budget B on the number of vaccines available,

find a set S of size B to vaccinate so as to optimize certain measure of effectiveness. The measure of effectiveness can be (i) minimizing the total number of individuals infected (or maximizing the total number of uninfected individuals); (ii) minimizing the total number of deaths (or maximizing the total number of deaths averted); (iii) optimizing the above quantities but keeping in mind certain equity and fairness criteria (across socio-demographic groups, e.g., age, race, income); (iv) taking into account vaccine hesitancy of individuals; (v) taking into account the fact that all vaccines are not available at the start of the pandemic, and when they become available, one gets limited number of doses each month; (vi) deciding how to share the stockpile between countries, state, and other organizations; (vii) taking into account efficacy of the vaccine.

Data needs: As in other problems, vaccine allocation problems need as input a good representation of the system; network-based, metapopulation-based, and compartmental mass action models can be used. One other key input is the vaccine budget, i.e., the production schedule and timeline, which serves as the constraint for the allocation problem. Additional data on prevailing vaccine sentiment⁶⁹ and past compliance to seasonal/neonatal vaccinations are useful to estimate coverage.

Modeling approaches: The problem has been studied actively in the literature⁷⁰; network science community has focused on optimal allocation schemes, while public health community has focused on using metapopulation models⁷¹ and assessing certain fixed allocation schemes based on socio-economic and demographic considerations⁷². Game theoretic approaches⁷³ that try and understand strategic behavior of individuals and organization has also been studied.

Challenges: The problem is computationally challenging and, thus, most of the time simulation based optimization techniques⁷⁴ are used. Challenge to the optimization approach comes from the fact that the optimal allocation scheme might be hard to compute or hard to implement. Other challenges include fairness criteria⁷⁵ (e.g., the optimal set might be a specific group) and also multiple objectives that one needs to balance.

8 Discussion

While the above sections provide an overview of salient modeling questions that arise during the key stages of a pandemic, mathematical and computational model development is equally, if not more important, as we approach the post-pandemic (or more appropriately "inter-pandemic") phase. Often referred to as "peace time" efforts, this phase allows modelers to retrospectively assess individual and collective models on how they performed during the pandemic. To encourage continued development and identifying data gaps, synthetic forecasting challenge exercises⁷⁶ may be conducted where multiple modeling groups are invited to forecast synthetic scenarios with varying levels of data availability. Another set of models that are quite relevant for policymakers during the winding down stages are those that help assess overall health burden and economic costs of the pandemic.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements

The authors would like to thank members of the Biocomplexity COVID-19 Response Team and Network Systems Science and Advanced Computing (NSSAC) Division for their thoughtful comments and suggestions related to epidemic modeling and response support. We thank members of the Biocomplexity Institute and Initiative, University of Virginia for useful discussion and suggestions. This work was partially supported by National Institutes of Health (NIH) Grant R01GM109718, NSF BIG DATA Grant IIS-1633028, NSF DIBBS Grant OAC-1443054, NSF Grant No.: OAC-1916805, NSF Expeditions in Computing Grant CCF-1918656, CCF-1917819, NSF RAPID CNS-2028004, NSF RAPID OAC-2027541, US Centers for Disease Control and Prevention 75D30119C05935, DTRA sub-contract/ARA S-D00189-15-TO-01-UVA. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

Received: 27 August 2020 Accepted: 15 September 2020
Published online: 16 November 2020

References

- Adhikari B, Xu X, Ramakrishnan N, Prakash BA (2019) EpiDeep: exploiting embeddings for epidemic forecasting. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, KDD'19, pp 577–586, New York, NY, USA, 2019. Association for Computing Machinery. <https://doi.org/10.1145/3292500.3330917>.
- Perone G (2020) An arima model to forecast the spread and the final size of covid-2019 epidemic in Italy (first version on SSRN 31 March). SSRN Electron J
- Desai A, Kraemer M, Bhatia S, Cori A, Nouvellet P, Herring M, Cohn E, Carrion M, Brownstein J, Madoff L, Lassmann B (2019) Real-time epidemic forecasting: challenges and opportunities. *Health Secur* 17:268–275
- Reich NG, McGowan CJ, Yamana TK, Tushar A, Ray EL, Osthus D, Kandula S, Brooks LC, Crawford-Crudell W, Gibson GC, Moore E, Silva R, Biggerstaff M, Johansson MA, Rosenfeld R, Shaman JL (2019) Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. *PLoS Comput Biol* 15
- Funk S, Camacho A, Kucharski AJ, Eggo RM, John EW (2018) Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics* 22:56–61 (The RAPIDD Ebola Forecasting Challenge)
- Healthmap. <https://healthmap.org/en/>. Accessed 28 Oct 2020
- Fung I, Tse Z, Fu K-W (2015) The use of social media in public health surveillance. *West Pac Surv Response J WPSAR* 6:3–6
- Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Piontti AP, Mu K, Rossi L, Sun K, Viboud C, Xiong X, Yu H, Halloran ME, Longini IM, Vespignani A (2020) The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 368(6489):395–400. <https://science.sciencemag.org/content/368/6489/395>
- Britton T (2020) Basic prediction methodology for COVID-19: estimation and sensitivity considerations. *medRxiv*. <https://www.medrxiv.org/content/10.1101/2020.03.27.20045575v2>
- Rocklöv J, Sjödin H, Wilder-Smith A (2020) COVID-19 outbreak on the diamond princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. *J Travel Med* 27(3):taaa030
- Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S, Boonyasiri A, Cucunubá Perez Z, Cuomo-Dannenburg G et al (2020) Report 9: impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. Imperial College Technical Report, 2020. <https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf>
- Eubank S, Guclu H, Anil Kumar VS, Marathe M, Srinivasan A, Toroczkai A, Wang N (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429:180–184
- Marathe M, Vullikanti A (2013) Computational epidemiology. *Commun ACM* 56(7):88–96
- IHME COVID, Murray CJL et al (2020) Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *MedRxiv*. <https://www.medrxiv.org/content/10.1101/2020.03.27.20043752v1>
- Alamo T, Reina DG, Mammarella M, Abella A (2020) Open data resources for fighting COVID-19. *arXiv preprint*. <https://arxiv.org/pdf/2004.06111.pdf>
- Alamo T, Reina DG, Millán P (2020) Data-driven methods to monitor, model, forecast and control COVID-19 pandemic: leveraging data science, epidemiology and control theory. *arXiv preprint*. <https://arxiv.org/pdf/2006.01731.pdf>
- Shuja J, Alanazi E, Alasmay W, Alashaikh A (2020) COVID-19 datasets: a survey and future challenges. *medRxiv*. <https://www.medrxiv.org/content/10.1101/2020.05.19.20107532v2.full.pdf+html>
- Sameni R (2020) Mathematical modeling of epidemic diseases; a case study of the COVID-19 coronavirus. *arXiv preprint*. <https://arxiv.org/pdf/2003.11371.pdf>
- Wu Joseph T, Cowling Benjamin J (2011) The use of mathematical models to inform influenza pandemic preparedness and response. *Exp Biol Med* 236(8):955–961
- Adiga A, Dubhashi D, Lewis B, Marathe M, Venkatraman S, Vullikanti A (2020) Mathematical models for COVID-19 pandemic: a comparative analysis. *J IISc*. <https://doi.org/10.1007/s41745-020-00200-6>
- Holloway R, Rasmussen SA, Zaza S, Cox NJ, Jernigan DB, Influenza Pandemic Framework Workgroup (2014) Updated preparedness and response framework for influenza pandemics. *Morb Mortal Wkly Rep* 63(6):1–18
- Reed C, Biggerstaff M, Finelli L, Koonin LM, Beauvais D, Uzicanin A, Plummer A, Bresee J, Redd SC, Jernigan DB (2013) Novel framework for assessing epidemiologic effects of influenza epidemics and pandemics. *Emerg Infect Dis* 19(1):85
- Centers for Disease Control and Prevention (2020) COVID-19 pandemic planning scenarios. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-h.pdf>. Accessed 14 Sept 2020
- Xu B, Gutierrez B, Mekaru S, Sewalk K, Goodwin L, Loskill A, Cohn EL, Hsuen Y, Hill SC, Cobo MM (2020) Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci Data* 7(1):1–6
- CDC. COVID-19 case surveillance public use data | data | centers for disease control and prevention. <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf>. Accessed 24 Aug 2020
- Li L-Q, Huang T, Wang Y-Q, Wang Z-P, Liang Y, Huang T-B, Zhang H-Y, Sun W, Wang Y (2020) COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. *J Med Virol* 92(6):577–583

27. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, Hens N (2020) Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance* 25(17):2000257
28. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 172(9):577–582
29. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, Cowling BJ, Lipsitch M, Leung GM (2020) Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 26(4):506–510
30. Marie Isabelle MM, Hall IM, Christley RM, Leach S, Read JM (2019) The use and reporting of airline passenger data for infectious disease modelling: a systematic review. *Eurosurveillance* 24(31):1800216
31. Venkatramanan S (2020) Flight cancellations related to 2019-nCoV (COVID-19). University of Virginia Dataverse
32. Brockmann D, Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science* 342:1337–1342
33. Adiga A, Venkatramanan S, Schlitt J, Peddireddy A, Dickerman A, Bura A, Warren A, Klahn BD, Mao C, Xie D, Machi D. Evaluating the impact of international airline suspensions on the early global spread of COVID-19. *medRxiv*. 2020 Jan 1.
34. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer Moritz UG, Khan K (2020) Potential for global spread of a novel coronavirus from China. *J Travel Med*
35. Wu JT, Leung K, Leung GM (2020) Forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*
36. De Salazar PM, Niehus R, Taylor A, Buckee CO, Lipsitch M (2020) Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. *medRxiv*. <https://www.medrxiv.org/content/10.1101/2020.02.04.20020495v2>
37. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boelle PY, D'Ortenzio E, Yazdanpanah Y, Eholie SP, Altmann M, Gutierrez B (2020) Preparedness and vulnerability of African countries against introductions of 2019-nCoV. *medRxiv*
38. Beckman R, Baggerly J, Keith A, McKay M (1996) Creating synthetic baseline populations. *Transp Res A* 30:415–429
39. Landscan. <https://landscan.ornl.gov/landscan-datasets>. Accessed 28 Oct 2020
40. Openstreetmap. https://wiki.openstreetmap.org/wiki/Downloading_data. Accessed 28 Oct 2020
41. American time use survey. <https://www.bls.gov/tus/>. Accessed 28 Oct 2020
42. Balcan D, Colizza V, Gonçalves B, Hao H, Ramasco JJ, Vespignani A (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci* 106:21484–21489
43. Venkatramanan S, Chen J, Fadikar A, Gupta S, Higdon D, Lewis B, Marathe M, Mortveit H, Vullikanti A (2019) Optimizing spatial allocation of seasonal influenza vaccine under temporal constraints. *PLoS Comput Biol* 15(9):e1007111
44. Gomes Marcelo F C, Pastore Ana, y Piontti AP, Rossi L, Chao DL, Longini IM, Halloran ME, Vespignani A (2014) Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLoS Curr* 6:2014
45. Zhang Q, Sun K, Chinazzi M, y Piontti AP, Dean NE, Rojas DP, Merler S, Mistry D, Poletti P, Rossi L, Bray M, Elizabeth Halloran M, Longini IM, Vespignani A (2017) Spread of zika virus in the Americas. *PNAS* 114(22):E4334–E4343
46. Eubank S, Anil Kumar VS, Marathe MV, Srinivasan A, Wang N (2006) Structure of social contact networks and their impact on epidemics. *DIMACS Ser Discrete Math Theor Comput Sci* 70:181
47. Barrett CL, Beckman RJ, Khan M, Anil Kumar VS, Marathe MV, Stretz PE, Dutta T, Lewis B (2009) Generation and analysis of large synthetic social contact networks. In: *Winter simulation conference*, pp 1003–1014
48. Longini IM, Nizam A, Shufu X, Ungchusak K, Hanshaworakul W, Cummings DA, Halloran EM (2005) Containing pandemic influenza at the source. *Science* 309(5737):1083–1087
49. Allen LJS, Brauer F, Van den Driessche P, Wu J (2008) *Mathematical epidemiology*, vol 1945. Springer, Berlin
50. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256
51. Amazon Web Services (2020) A public data lake for analysis of COVID-19 data. <https://aws.amazon.com/blogs/big-data/a-public-data-lake-for-analysis-of-covid-19-data/>. Accessed 28 Oct 2020
52. MIDAS network (2020) MIDAS 2019 novel coronavirus repository. <https://github.com/midas-network/COVID-19>. Accessed 28 Oct 2020
53. The New York Times (2020) Coronavirus (COVID-19) data in the United States. <https://github.com/nytimes/covid-19-data>. Accessed 28 Oct 2020
54. COVID-19 impact analysis platform. <https://data.covid.umd.edu/>. Accessed 28 Oct 2020
55. Biocomplexity Institute (2020) COVID-19 surveillance dashboard. <http://ncov.bii.virginia.edu/dashboard/>. Accessed 28 Oct 2020
56. The covid tracking project. <https://covidtracking.com/>. Accessed 28 Oct 2020
57. Shaman J, Pitzer V, Viboud C, Grenfell B, Lipsitch M (2010) Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol* 8:e1000316

58. Cori A (2013) Epiestim: a package to estimate time varying reproduction numbers from epidemic curves. R package version, pp 1–1
59. Google COVID-19 community mobility reports. <https://www.google.com/covid19/mobility/>. Accessed 28 Oct 2020
60. Mask-wearing survey data. <https://github.com/nytimes/covid-19-data/tree/master/mask-use>. Accessed 28 Oct 2020
61. Wang X, Pasco RF, Du Z, Petty M, Fox SJ, Galvani AP, Pignone M, Johnston SC, Meyers LA (2020) Impact of social distancing measures on coronavirus disease healthcare demand, central Texas, USA. *Emerg Infect Dis* 26(10)
62. Current hospital capacity estimates —snapshot. <https://www.cdc.gov/nhsn/covid19/report-patient-impact.html>. Accessed 28 Oct 2020
63. Total hospital bed occupancy (COVID-19). <https://www.scdhec.gov/infectious-diseases/viruses/coronavirus-disease-2019-covid-19/hospital-bed-capacity-covid-19>. Accessed 28 Oct 2020
64. Lorch L, Trouleau W, Tsirtsis S, Szanto A, Schölkopf B, Gomez-Rodriguez M (2020) Quantifying the effects of contact tracing, testing, and containment. arXiv preprint arXiv:2004.07641
65. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, Zwahlen M, Senti G, Battegay M, Wilder-Smith A et al (2020) COVID-19 epidemic in switzerland: on the importance of testing, contact tracing and isolation. *Swiss Med Wkly* 150(1112)
66. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Parker M, Bonsall D, Fraser C (2020) Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science*
67. Kretzschmar M, Rozhnova G, van Boven M (2020) Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing. SSRN 3562458. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3562458. Accessed 28 Oct 2020
68. Chan J, Gollakota S, Horvitz E, Jaeger J, Kakade S, Kohno T, Langford J, Larson J, Singanamalla S, Sunshine J et al (2020) Pact: privacy sensitive protocols and mechanisms for mobile contact tracing. arXiv preprint arXiv:2004.03544. <https://arxiv.org/pdf/2004.03544.pdf>
69. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, Swarup S (2017) Semantic network analysis of vaccine sentiment in online social media. *Vaccine* 35(29):3621–3638
70. Medlock J, Galvani AP (2009) Optimizing influenza vaccine distribution. *Science* 325(5948):1705–1708
71. Venkatramanan S, Chen J, Fadikar A, Gupta S, Higdon D, Lewis B, Marathe M, Mortveit H, Vullikanti A (2019) Optimizing spatial allocation of seasonal influenza vaccine under temporal constraints. *PLoS Comput Biol* 15(9):e1007111
72. Tuite AR, Fisman DN, Kwong JC, Greer AL (2010) Optimal pandemic influenza vaccine allocation strategies for the Canadian population. *PloS One* 5(5):e10520
73. Bauch CT, Earn DJ (2004) Vaccination and the theory of games. *Proc Nat Acad Sci* 101(36):13391–13394
74. Patel R, Longini Jr IM, Halloran ME (2005) Finding optimal vaccination strategies for pandemic influenza using genetic algorithms. *J Theor Biol* 234(2):201–212
75. Yi M, Marathe A (2015) Fairness versus efficiency of vaccine allocation strategies. *Value in Health* 18(2):278–283
76. Viboud C, Sun K, Gaffey R, Ajelli M, Fumanelli L, Merler S, Zhang Q, Chowell G, Simonsen L, Vespignani A et al (2018) The rapid ebola forecasting challenge: synthesis and lessons learnt. *Epidemics* 22:13–21



Aniruddha Adiga is a Postdoctoral Research Associate at the NSSAC Division of the Biocomplexity Institute and Initiative. He completed his PhD from the Department of Electrical Engineering, Indian Institute of Science (IISc), Bangalore, India, and has held the position of Postdoctoral fellow at IISc and North Carolina State University, Raleigh, USA. His research areas include signal processing, machine learning, data mining, forecasting, and big data analysis. At NSSAC, his primary focus has been the analysis and development of forecasting systems for epidemiological signals such as influenza-like illness and COVID-19 using auxiliary data sources.



Jiangzhuo Chen is a research associate professor in the Biocomplexity Institute, University of Virginia. He received his PhD in computer science from Northeastern University, and held positions at Virginia Tech before joining UVA. He has more than 10 years of experience in designing and implementing

high-performance computational models for large-scale socio-technical systems. He has co-developed three agent-based large-scale simulation tools, which have been used in many studies to support modeling, forecasting, and impact analysis for infectious diseases including Influenza, Ebola, and COVID-19. His current research interests include agent-based modeling, computational epidemiology, and theory-guided machine learning.



Madhav Marathe is a Distinguished Professor in Biocomplexity, the Division Director of the Networks, Simulation Science and Advanced Computing Division at the Biocomplexity Institute and Initiative, and a Professor in the Department of Computer Science at the University of Virginia (UVA). His research interests are in network science, computational epidemiology, AI, foundations of computing, socially coupled system science, and high-performance computing. Before joining UVA, he held positions at Virginia Tech and the Los Alamos National Laboratory. He is a Fellow of the IEEE, ACM, SIAM, and AAAS.



Henning Mortveit is an associate professor with joint appointments at the Networks, Simulation Science and Advanced Computing Division at the Biocomplexity Institute and Initiative, and at the Department of Engineering Systems and Environment at the University of Virginia (UVA). His research is focused on massively interacting systems (MIS), their mathematical theory as captured through graph dynamical system, modeling and simulation science for MIS, as well as supporting system and architecture design for scientifically reproducible, high-performance simulation science of MIS. Before joining UVA, he held positions at Virginia Tech and at the Los Alamos National Laboratory.



Srinivasan Venkatramanan is a Research Scientist at the Biocomplexity Institute and Initiative, University of Virginia, and his research focuses on developing, analyzing, and optimizing computational models in the field of network epidemiology. He received his PhD from the Department of Electrical and

Communication Engineering, Indian Institute of Science (IISc), and did his postdoctoral research at Virginia Tech. His areas of interest include network science, stochastic modeling, and big data analytics. He has used in-silico models of society to study the spread of infectious diseases and invasive species. Recent research includes modeling and forecasting emerging infectious disease outbreaks (e.g., Ebola, COVID-19), impact of human mobility on disease spread, and resource allocation problems in the context of epidemic control (e.g., seasonal influenza vaccines).



Anil Vullikanti is a Professor in the Biocomplexity Institute and the Department of Computer Science at the University of Virginia. He got his PhD at the Indian Institute of Science, and was a postdoctoral associate at the Max Planck Institute for Computer Science, Saarbrücken, Germany, and the Los Alamos National Lab. His research interests are in the broad areas of network science, dynamical systems, combinatorial optimization, and distributed computing, and their applications to computational epidemiology and social networks.