Projecting COVID-19 cases and hospital burden in Ohio

Wasiur R. KhudaBukhsh [a,*], Caleb Deen Bastian [b], Matthew Wascher [c], Colin Klaus [d,e,f], Saumya Yashmohini Sahai [g], Mark H. Weir [e,f,h], Eben Kenah [e,f,**], Elisabeth Root [i], Joseph H. Tien [d,e,f,j], Grzegorz A. Rempała [d,e,f,j,**]

[a] School of Mathematical Sciences, University of Nottingham, University Park Nottingham NG7 2RD, United Kingdom
[b] Program in Applied and Computational Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544, USA
[c] Department of Mathematics, University of Dayton, 300 College Park, Dayton, OH 45469, USA
[d] Mathematical Biosciences Institute, The Ohio State University, 1735 Neil Avenue, Columbus, OH 43210, USA
[e] College of Public Health, The Ohio State University, Cunz Hall, 1841 Neil Avenue, Columbus, OH 43210, USA
[f] Infectious Diseases Institute, 208 Bricker Hall, 190 North Oval Mall, Columbus, OH 43210-1358, USA
[g] Department of Computer Science and Engineering, The Ohio State University, 395 Dreese Laboratories, 2015 Neil Avenue, Columbus, OH 43210, USA
[h] The Sustainability Institute, The Ohio State University, 74 W. 18th Avenue, Columbus, OH 43210, USA
[i] Institute for Disease Modeling, The Bill & Melinda Gates Foundation, Seattle, WA, USA
[j] Department of Mathematics, The Ohio State University, 231 West 18th Avenue, Columbus, OH 43210-1174, USA

ARTICLE INFO

ABSTRACT

As the Coronavirus 2019 disease (COVID-19) started to spread rapidly in the state of Ohio, the Ecology, Epidemiology and Population Health (EEPH) program within the Infectious Diseases Institute (IDI) at The Ohio State University (OSU) took the initiative to offer epidemic modeling and decision analytics support to the Ohio Department of Health (ODH). This paper describes the methodology used by the OSU/IDI response modeling team to predict statewide cases of new infections as well as potential hospital burden in the state. The methodology has two components: (1) A Dynamical Survival Analysis (DSA)-based statistical method to perform parameter inference, statewide prediction and uncertainty quantification. (2) A geographic component that down-projects statewide predicted counts to potential hospital burden across the state. We demonstrate the overall methodology with publicly available data. A Python implementation of the methodology is also made publicly available. This manuscript was submitted as part of a theme issue on "Modelling COVID-19 and Preparedness for Future Pandemics".

1. Introduction

The coronavirus 2019 (COVID-19) disease resulted in 580 million confirmed cases and 6.4 million deaths reported globally as of July 2022 (Center for Systems Science and Engineering (CSSE) at the Johns Hopkins University, 2021). As evidenced by epidemics in many countries, such as Italy and Spain, COVID-19 patient case loads have the potential to overwhelm healthcare systems (Grasselli et al., 2020). In the early stages of the COVID-19 pandemic in Ohio, the potential demand for beds in hospitals and intensive care units was a primary concern.

The Infectious Diseases Institute (IDI) at the Ohio State University working in conjunction with the College of Public Health (CPH), the Department of Mathematics, and the Sustainability Institute (SI) established a working relationship with the Ohio Department of Health (ODH) to act as a service to the State, beginning in 2018. Based on this initial collaborative relationship, the Ecology Epidemiology and Population Health (EEPH) program within IDI took the initiative to offer epidemic modeling and decision analytic support for the ODH response to the COVID-19 pandemic within the state of Ohio. The most immediate need was to predict the number of cases and the potential load on the healthcare system.

As in any computational modeling effort, there are many possible approaches and methods to model emerging epidemics. Modeling approaches include compartmental models (Predictive Healthcare at Penn Medicine, 2020; Childs et al., 2020; Weitz, 2020), statistical models (IHME COVID-19 health service utilization forecasting team and Murray, 2020), and agent-based simulations (Ferguson et al., 2020). Different methods are suited to different situations in terms of available data, potential scenarios, local conditions, urgency, and goals. For the state of Ohio, the OSU-IDI group approached the modeling challenge on two fronts by developing: 1) projected statewide estimates of a time series of COVID-19 incidence, and 2) a geographic component that transforms the output of the statewide model to hospital burden by county or "hospital catchment" area.

2. Methods

2.1. Generalization of methods used

The predictive statewide model for Ohio comprises a dynamic network model, where network edges (contacts between nodes in the network) can interact with each other (Newman, 2018), but it has three key distinguishing features:

1. The model considers a dynamic network where the edges can be deactivated over time—supporting social distancing impacts more accurately.
2. The law of large numbers yields a set of differential equations describing the disease process on a large network (Jacobsen et al., 2018) without requiring simulation methods.
3. The solutions of these differential equations can be used to estimate model parameters using a principled statistical approach based on survival analysis. We write an explicit likelihood for the parameters given data on times of illness onset (KhudaBukhsh et al., 2020). Critically, this allows for accurate quantification of the uncertainty in the model predictions.

The DSA approach retains the tractability of an analytic model while incorporating complex human networks to better represent social interactions and distancing.

The projected statewide model takes data on illness onset dates of confirmed cases as input and produces estimates of COVID-19 incidence (i.e., new cases) in Ohio over time. This output is not age-stratified, but the age distribution of the new cases is assumed to match the age distribution of confirmed cases when projecting the number of hospitalizations in the next step of the model.

Estimates of the number of hospitalized COVID-19 cases are derived in the geographic component of the model. Because illness severity and the risk of hospitalization for COVID-19 patients vary according to age and comorbidities (Verity et al., 2020; CDC COVID-19 Response Team, 2020a), we use local age structure and population density to distribute case counts from the statewide epidemic model across smaller geographic areas within the state. Within each geographic unit, we use its own age distribution to project hospitalization rates over time. Consequently, counties or hospital catchment areas that have a different demographic structure (e.g., older or younger populations) will differ in their COVID-19 hospitalizations over time.

2.2. Detailed methods of the predictive model

There are challenges to using traditional compartmental models (Niehus et al., 2020) to estimate future COVID-19 incidence. One of the most fundamental challenges is that these methods require knowledge of the size of the susceptible population. In our predictive model, we use an approach called *Dynamical Survival Analysis* (DSA) (KhudaBukhsh et al., 2021; Somekh et al., 2022; Wascher et al., 2021), which is an extension of *survival dynamical systems* (KhudaBukhsh et al., 2020; Bastian and Rempala, 2020). Three key strengths of the DSA approach for predicting the course of novel virus epidemics such as COVID-19 are:

- It does not require knowledge of the size of the susceptible population,
- It does not require information on overall disease prevalence in the population.
- It does not require prior knowledge of the shape of the epidemic curve.

Details on this approach and the model development can be found in Appendix A.

Since testing was initially focused on the most symptomatic and severe cases, we started with almost no information on the number of asymptomatic infections or those with less severe symptoms who are not tested. Because of the novel nature of this virus and the resulting pandemic, analysis of the epidemic curve cannot be based on previous epidemics caused by other viruses (e.g., SARS-CoV-1, or influenza). Because it relies on illness onset times rather than counts of new cases, the DSA method can incorporate a partial epidemic curve like the one produced by early testing for COVID-19 in Ohio, which was affected by both limited testing capacity and undetected mild or asymptomatic infections. The DSA method has been developed to handle incomplete data in a manner that allows accurate quantification of uncertainty. The method is strengthened by its simplicity as it requires only a single differential equation. Parameters for the model are inferred using empirical temporal data on new illnesses, and the target output is a time series of expected future illnesses.

Initially, we used maximum likelihood estimates (MLEs) to fit the DSA model to Ohio data. However, as a substantial amount of data became available, we adopted a Hamiltonian Monte Carlo-based Bayesian approach to parameter inference and uncertainty quantification. An implementation of our inference method in the Python programming language is available as a GitHub repository (Bastian and KhudaBukhsh, 2020).

Our projected statewide model provides robust estimates of cases over time given partially observed daily counts of new illnesses. This is ideal in a setting where testing capacity is limited or changing due to constraints on lab capacity or detection limits. The counts of new illnesses are often known as an observed *epidemic curve* in the literature (Wu and McGoogan, 2020). Our approach is derived from the general stochastic model of a pathogen spread across a contact network where the nodes represent individuals in a community (Jacobsen et al., 2018). As a working model of a contact network we use a type of random graph called a dynamic configuration model (CM) (Bollobás, 1998). The mean-field approximation for epidemic processes on such networks is sometimes referred to as a pairwise model (Kiss et al., 2017).

Assumptions. All models have specific assumptions used to develop and implement them. First, we assume that each individual in the network (node) has a number of neighbors (their degree). Subsequently, a local Markovian infectious pressure changes their status from *susceptible* (S) to *infective* (I) to *removed* (R). We further assume that the R individuals are no longer able to transmit the infection and cannot be reinfected. For the dynamics of the network model, individuals transition between the S, I, and R compartments based on the following principles that allow the extraction of a mathematical model of the ongoing epidemic based on observable data:

1. Disease spread occurs over a network of contacts. An infectious individual can infect their immediate neighbors at a fixed positive rate. The average number of a person's contacts is positive.
2. Each infected individual recovers from infection at a positive rate or is restricted from contacting their network neighbors through mandatory or voluntary isolation at a positive rate.
3. Each infected individual has an infectious period that is sampled from an exponential distribution.
4. People who are ill remain infectious, and a partial count of new illnesses is observed over time with a negligible chance of false positives.

Estimating transmission rates. The complete model description and detailed development can be found in Appendix. In brief, the statistical model is used to estimate the number of people and timing of transfer between states *S*, *I*, and *R*. Then through substitution, a term S_t, the number of susceptible people at time *t* is developed. Embedded within this S_t is an improper survival function for the time to the onset of illness in a randomly chosen susceptible person. Using this survival function, the time to infection of a randomly selected susceptible person within a large population follows a temporal pattern determined by probability laws. After we estimate the time series of susceptibles, we can estimate the probability of a randomly selected susceptible individual being infected during the lifetime of an epidemic. From this we develop the conditional probability of remaining susceptible past time *t*.

Estimating dropout and recovery rates. The impact of social distancing is accomplished via a generalized approach to a person being dropped from the network. As has been visualized elsewhere, when a person is removed from the network, their neighbors and contacts within the network are removed, which limits transmission. This is accomplished by estimating the rate of infectious contact within a network (β in the Appendix) and then considering drop outs as a function of recovery rate (details in the Appendix).

Changes in parameters due to interventions. Introducing and then easing restrictions on social distancing in the state has likely resulted in changes in parameters. The effect of such changes is incorporated into the predictive model.

Estimating hospitalization and ICU admission onset. After the statewide model produces time series estimates of cases across the state, these are then translated into estimates of hospitalizations and subsequent ICU admission. The complete derivation of this model is in Appendix B. In short, this is developed in three steps:

1. Estimate case onsets for each age group based on the population and age distribution in each geographic area modeled.
2. Estimate the number of severe cases that will need hospitalization based first on CDC data (CDC COVID-19 Response Team, 2020b) and then updated with Ohio-specific data as those data become available in sufficient quantity to support predictive modeling.
3. Estimate a probability distribution (details in Appendix B) for the time from case onset to hospitalization as a function of age, sex, and race. Using the same variables, we also estimated the time from hospitalization to discharge or death and the probability of being admitted to the ICU. This provides a means to model hospital occupancy based on Ohio-specific data.

These results are then integrated with the geographic modeling methods described next.

2.3. Geographic modeling methods

The statewide model described above produces a time series of estimated incident cases of COVID-19 illness. These outputs are then used to assess hospital burden (e.g., hospitalizations and ICU admissions) using the age structure of the population and the estimated length of time a patient will be in a hospital or ICU bed. This involved a three-step estimation process:

1. The proportion of the total Ohio population residing in each geographic area (e.g., county or hospital catchment area) was estimated from U.S. Census data. Daily case counts were distributed across geographic areas using this proportion, essentially distributing cases by population density. This created a time series of projected incident case counts by geographic area.

2. To account for age differentials in hospitalization and ICU use noted by both the popular press and in the scientific literature, we estimated the proportion of new cases expected to require hospitalization or admission to the ICU using the age distribution in each geographic unit at each time step. Initially, we used estimates from the Morbidity and Mortality Weekly Report (MMWR) published by CDC (CDC COVID-19 Response Team, 2020b). These nationwide estimates of the age distribution of hospitalization were used only while the number of hospitalizations and ICU admissions in Ohio remained too low to make reliable forecasts. Ohio rates were substituted in subsequent runs of the model. In counties with an older age structure, a larger proportion of cases would convert into a hospitalization.

3. We used the estimated number of new hospitalizations and the average length of stay (LOS) to estimate the total hospital burden for each day in the time series. Initially, the average LOS was derived from the literature, but we used observed LOS from Ohio hospitals when enough COVID-19 cases were identified to provide robust estimates. We used a bi-modal distribution, with an average LOS of 5 days for non-ICU patients and 14 days for those requiring ICU care. For each time step, the number of new hospitalized cases estimated from the model was added to the number of cases still in the hospital. The cases that "timed out" of their hospital stay due to the LOS parameter were subtracted from the total. This created an estimated net number of patients in the hospital for each day in the time series which accounted for "patient stacking" over time.

Once daily hospital counts were estimated, we compared these to the reported number of COVID available beds pooled across all hospitals in a geographic area. This allowed us to understand when and where hospital bed need might exceed current capacity.

2.3.1. Description of data

Demographic data. Demographic data on the age distribution for Ohio counties and ZIP Codes were obtained from the U.S. Census Bureau's 5-year American Community Survey 2014–2018 estimates (US Census Bureau, 2022a). The ACS data were used because the sample size was large enough for small geographies for reasonable standard errors and stable estimates. The Census Bureau does not develop data products for the United States Postal Service (USPS) zip codes. Rather, ZIP Code Tabulation Areas (ZCTAs) are generalized areal representations of USPS zip code service areas. We mapped the Census ZCTAs to zip codes and used population estimates for ZCTAs. We used table B01001 that breaks out population counts by sex and 5 year age groups. Fig. 1 shows the distribution of the high risk population (age 55+) by county and hospital catchment area (see Section 2.3.2) in Ohio.

OHA hospital and bed data. The Ohio Hospital Association (OHA) provided data on the location of all hospitals in the state and the number of registered beds for each facility. Beds were broken out into several categories: Airborne Isolation, Critical Care, General Medical/Surgical Beds, and Extracorporeal Membrane Oxygenation (ECMO) beds. These bed types represent two levels of care required for COVID-19 patients, general hospital care for severe disease and ICU beds for patients requiring ventilation. The OHA data contains information that constitutes a "trade secret" under Ohio Revised Code Section 1333.61.

Description of OHA hospital market share data. The OHA provided data on hospital market share derived from administrative hospital claims from the past year (January 2019–December 2019, inclusive). For each hospital, patient encounters were grouped by patient zip code. A hospital's market area included zip codes that represented the top 80% of all encounters at that hospital.
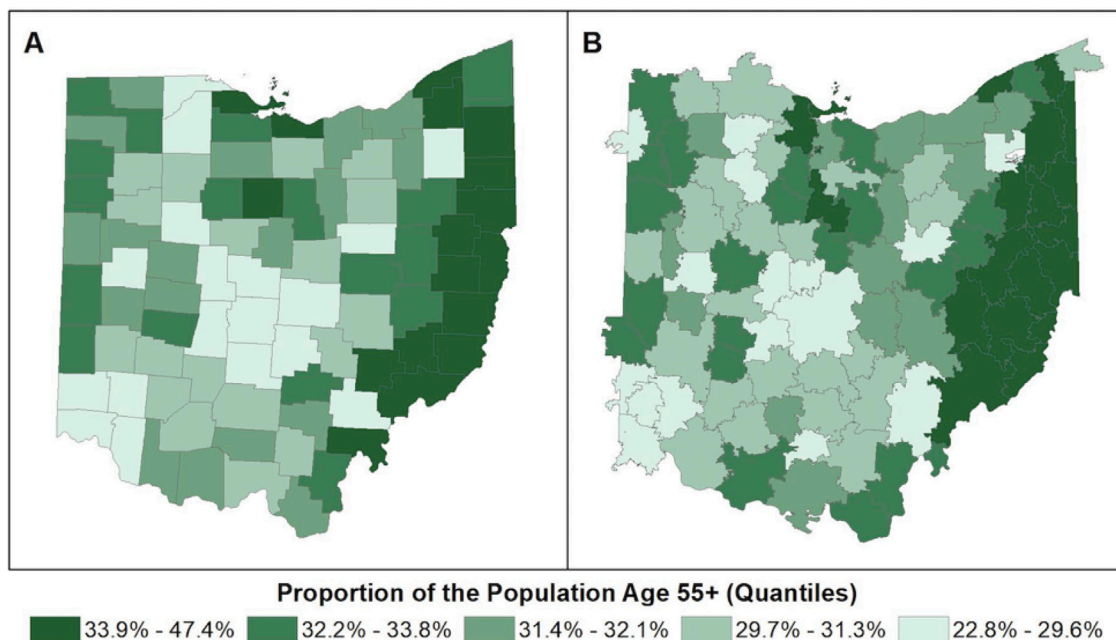
Fig. 1. Proportion of the population over the age of 55 in (A) counties and (B) hospital catchment areas.

2.3.2. Definition of hospital catchment areas

We developed small area estimates from state-level model predictions for counties and hospital catchment area (HCA). Boundary files for the 88 Ohio counties were obtained from the U.S. Census Cartographic Boundary Files (US Census Bureau, 2022b). We developed hospital catchment areas using an approach modified from the Dartmouth Atlas Project (The Dartmouth Institute for Health Policy and Clinical Practice, 2022). We define an HCA as a collection of zip codes whose residents receive most of their hospital care from the hospital(s) in that region. Fig. 1-B shows the HCAs developed and mapped. Note that there are HCAs that cross over into neighboring states. This is explained in more detail below.

HCA definitions depend on the integrated use of geospatial methods that grouped each zip code in the state with the most geographically proximate hospital and modified these groupings using data on hospital market share by zip code. Only hospitals with acute care beds that could be used for COVID-19 patients were included in the analysis. We excluded facilities such as long-term acute care (LTAC), hospice, orthopedic, rehabilitation or psychiatric/behavioral health hospitals and freestanding ERs. We defined HCAs using three steps:

1. The location of all hospitals in Ohio, and those in Michigan, Indiana, West Virginia, Kentucky and Pennsylvania located on the border with Ohio, was used to generate a Voronoi diagram with hospitals as generating points (Okabe et al., 2000). We included hospital in neighboring states in the Voronoi analysis to avoid edge effects, which would attribute patients to Ohio hospitals that typically use hospitals in other states. This yielded 233 distinct areas, one for each hospital generator. Twenty-eight areas were subsequently deleted because they included no area within Ohio.
2. Voronoi polygons were overlaid with zip codes. ZIP codes were assigned to the Voronoi polygon if their centroid fell within the polygon. This ensured that each ZIP code was associated with the most geographically proximate hospital and divided the state into groupings of zip codes assigned to each hospital.
3. Using the OHA hospital market share file, we examined the level of agreement between each Voronoi polygon and market share zip codes for each hospital. In cases where adjacent Voronoi polygons were generated by hospitals that also shared 60% or

more of their market share zip codes, we aggregated these polygons to create one hospital catchment area. In large metropolitan areas with many hospitals, we aggregated groupings of Voronoi polygons to create regional catchment areas. In rural areas, this typically resulted in aggregating two adjacent polygons.

Using this procedure, we generated 96 HCAs for the state of Ohio. Some HCAs included zip codes from neighboring states, and some Ohio ZIP codes were included in HCAs for non-Ohio hospitals. HCAs are shown in Fig. 2.

3. Results

Here, we present a brief summary of our model fits and predictions. One of the hallmarks of the COVID-19 pandemic has been the change in human behavior due to various interventions throughout the course of the pandemic. As a consequence, the COVID-19 epidemic curve in Ohio deviated much from a typical epidemic curve. The Ohio curve initially followed an exponential growth phase with high R_0, then a phase of steady linear growth and a decline with R_0 close to one, and then finally, due to reopening of the state, another phase of exponential growth with R_0 greater than one. As such, it is natural to believe the parameters are potentially different in these different phases. Therefore, we fit the DSA model with multiple change points (see posterior distributions in Figs. 6 and 7 in Appendix A).

In Fig. 3, we compare actual counts of daily new infections against model predictions. The fitted trajectories follow the observed epidemic curve well. The inner confidence bound is the true posterior confidence bound obtained point-wise. However, it is worth noting that predicted trajectories underestimate the variance or over-dispersion in the observed counts of daily new infections. Therefore, we adopted an empirical variance adjustment method to account for this possible underestimation. The broader confidence bound corresponds to the variance-adjusted trajectories. More details on the variance adjustment method and other diagnostic plots are in Appendix B.

The daily counts thus predicted are then down-projected into predictions of hospitalization surge across Ohio in the geographic component of the model. During the early phase of the epidemic, only short-term predictions (a few weeks) of hospitalizations were provided to the ODH and the OHA. As more data became available, and the
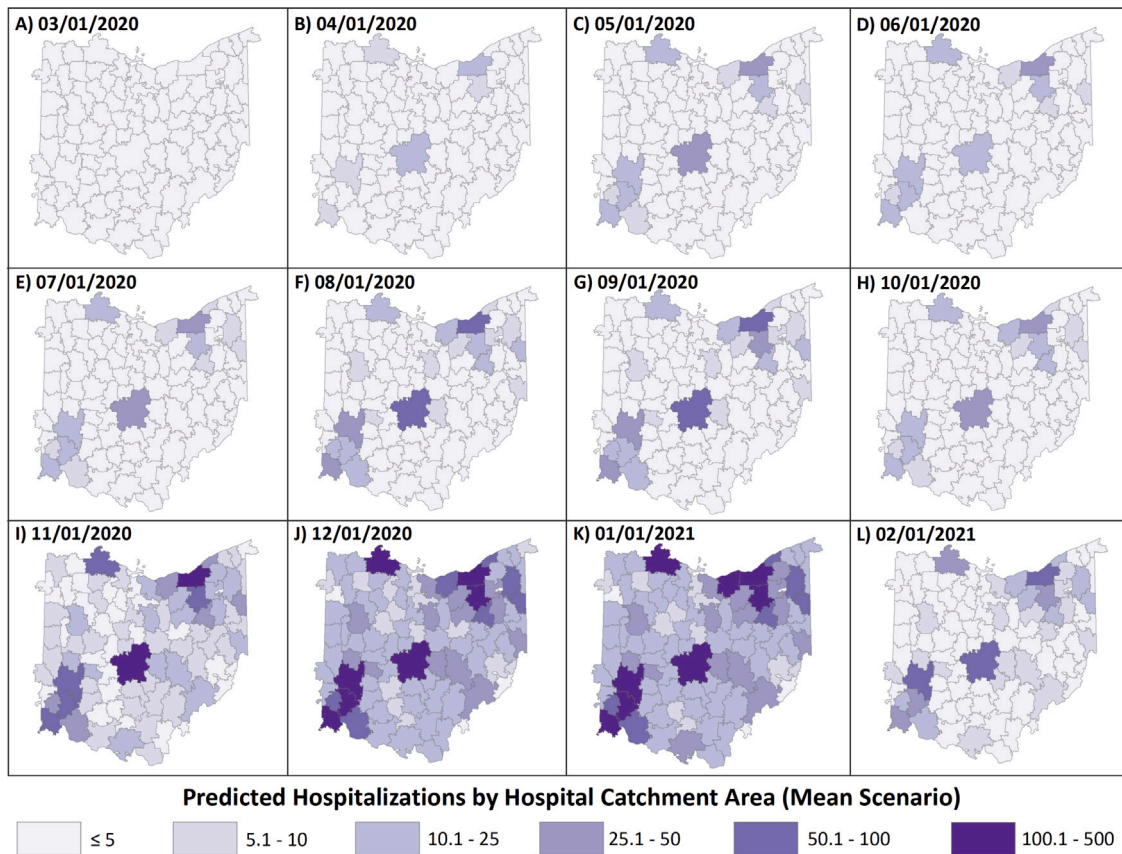
Predicted Hospitalizations by Hospital Catchment Area (Mean Scenario)

≤ 5   5.1 - 10   10.1 - 25   25.1 - 50   50.1 - 100   100.1 - 500

Fig. 2. Estimates of hospital burden mapped to hospital catchment areas for 12 days during the pandemic.
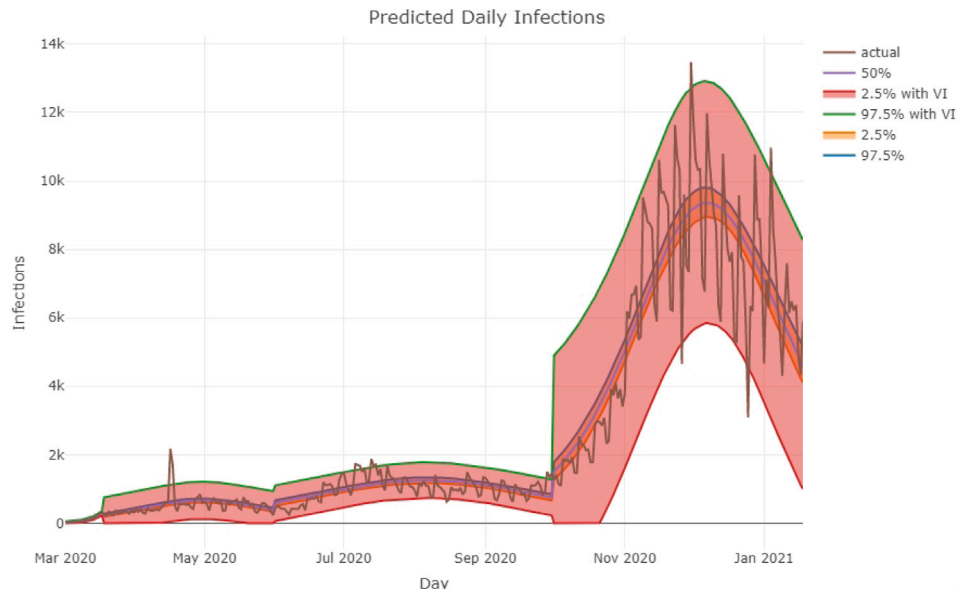

Fig. 3. Predictions based on the fitted model versus actual daily incidence in Ohio as reported by the Ohio Department of Health. We assumed three change points here corresponding to four density segments. Day zero corresponds to March 1, 2020. The first change point is on March 17, 2020. The second change point is on June 1, 2020, after which the state observed another phase of exponential growth. The third change point is on October 1, 2021, which marks the beginning of second wave characterized by a severe exponential growth phase. The broader (outer) confidence bound corresponds to the variance-adjusted trajectories.

learnt model parameters became more robust and stable, slightly longer term predictions were provided, and changes in the model parameters were monitored regularly. When significant changes in the model parameter were observed, the prediction model was updated accordingly and fresh predictions of hospitalizations were made available to the ODH and the OHA. In Fig. 2, we show snapshots of predicted hospitalizations across the 96 HCAs for the state of Ohio on 12 different days from March 1, 2020 to February 1, 2021.

In Fig. 4, we compare the predicted total hospital admissions in Ohio from October 2 with the true reported cases (Ohio Hospitals
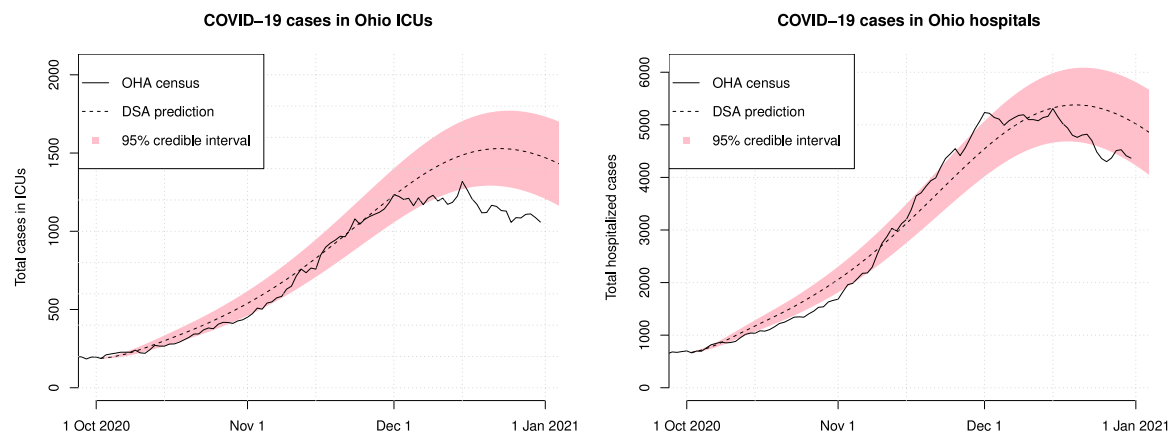
Fig. 4. (Left) Predictions of ICU admissions in Ohio after the third change point (October 1, 2021). **(Right)** Predictions of total hospital admissions in Ohio after the third change point (October 1, 2021).

Association, 2022) during the second wave of the epidemic, *i.e.*, after the third change point October 1, 2021. The figure shows good agreement between our predictions and the reported cases. Additional comparative figures on hospital admissions are provided in Appendix B.

4. Discussion

The susceptible–infectious–recovered (SIR) framework is the basis for many COVID-19 epidemic models (Predictive Healthcare at Penn Medicine, 2020; Childs et al., 2020; Li et al., 2020; Weitz, 2020). Several websites provide implementations of the SIR framework and present scenarios with different interventions such as social distancing that affect the transmission parameter β (Predictive Healthcare at Penn Medicine, 2020; Childs et al., 2020). Recent work (Childs et al., 2020) has attempted to extend the basic SIR model to include additional compartments corresponding to incubation, presymptomatic infectious, stages of symptomatic infectious, and hospitalization. An important advantage of the DSA framework is that it can be extended to include network-based models and non-Markov models by altering the form of the likelihood, and choices between the models can be guided by standard statistical methods.

Current SIR tools (Predictive Healthcare at Penn Medicine, 2020; Childs et al., 2020) are suited for scenario exploration, like illustrating the *flattening of the curve* under different levels of social distancing compliance. Unfortunately, they are less suited for forecasts because model parameters are not calibrated based upon case or outcome data. A recent attempt at taking a meta-population approach with SEIR dynamics in each patch, to estimate model parameters (including proportion of asymptomatic infection) based upon case counts from China was given in Li et al. (2020). This resulted in a spatially explicit model without age-structure. Conversely, Weitz (2020) is an example of a nonspatial model with age structure: Weitz extends the SEIR framework to include age groups, and estimates model parameters based upon hospitalization and death counts from Georgia. Further work from the University of Washington uses mortality data as model inputs, but takes a purely statistical approach in fitting a sigmoidal curve to cumulative COVID-19 deaths using a mixed-effects model (IHME COVID-19 health service utilization forecasting team and Murray, 2020). Another alternative is an agent-based model such as that used in Ferguson et al. (2020). Similar to Li et al. (2020), Ferguson et al. (2020) forecast a very large number of infections with COVID-19.

Considering the level of uncertainty in the data being input into all of these models, we believe that a model output of a time series of illnesses, hospitalizations, and ICU admissions is a more reasonable approach. Similarly, the simplified framework and structure of our approach allow for greater flexibility under uncertain data inputs into the model and a lack of information on key parameters such as the total

susceptible population and the proportion of cases that are subclinical or asymptomatic. It turned out that the resulting predictions were remarkably accurate.

A significant advantage of the DSA method is that it requires fewer parameters and is usually less computationally expensive than the agent-based models, such as Agrawal et al. (2020), Ferguson et al. (2020), which are used quite successfully to run large-scale simulations for the purpose of forward predictions and the analysis of what-if scenarios. Barring dropping of edges, the DSA method does not provide the flexibility to test arbitrary what-if scenarios involving individual human behaviors because the method is based on population-level equations.

Our approach is quite general and flexible. We have recently used variations of the DSA method with encouraging results to model the foot-and-mouth disease (FMD) outbreak in the United Kingdom in 2001 (Di Lauro et al., 2022), the spread of influenza A(H1N1) on the Washington State University campus in 2009 (KhudaBukhsh et al., 2020), and the Ebola epidemic in the Democratic Republic of Congo (DRC) in 2018–2020 (Vossler et al., 2022).

CRediT authorship contribution statement

Wasiur R. KhudaBukhsh: Methodology, Software, Writing – original draft. **Caleb Deen Bastian:** Methodology, Software, Writing – original draft. **Matthew Wascher:** Methodology, Software, Writing – original draft. **Colin Klaus:** Methodology, Software, Writing – original draft. **Saumya Yashmohini Sahai:** Software, Writing – original draft. **Mark H. Weir:** Conceptualization, Methodology, Writing – original draft. **Eben Kenah:** Conceptualization, Methodology, Formal analysis, Writing – original draft. **Elisabeth Root:** Conceptualization, Methodology, Writing – original draft. **Joseph H. Tien:** Conceptualization, Methodology, Writing – original draft. **Grzegorz A. Rempała:** Conceptualization, Methodology, Writing – original draft.

Data availability & software

The software to perform model fit along with a data example is available from Bastian and KhudaBukhsh (2020). The onset data used in this paper can be downloaded from the Ohio Department of Health (ODH) COVID-19 dashboard. Probability distributions for the time from disease onset to hospitalization, length of stay, and probability of admission to the ICU were based on analysis of non-public data from ODH.

Acknowledgment

Funding

Additional materials

A video presentation describing the methodology used for epidemic size predictions is available from Rempała (2020).

Appendix A. Predicting statewide cases of COVID-19

This derivation is taken from Bastian et al. (2020). We will outline the derivation of a simple but powerful general modeling framework that provides robust estimates of the quantities relevant to monitoring local outbreaks where only limited amount of information is available through partially observed daily counts of new (symptomatic) infections. Our approach is derived from the general stochastic model of a contagion spread across a contact network of nodes representing individuals in a community (Jacobsen et al., 2018). As a working model of a contact network we use a dynamic configuration model (CM)-type random graph (see, e.g., Bollobás (1998)). Such a model is often also referred to as a *pairwise model* (Kiss et al., 2017).

Briefly, we assume that each node has its degree and that the nodes may change their status from the initial "Susceptible" (S) to "Infective" (I) (or "Infectious") and, finally, to "Removed" (R), according to their local Markovian infectious pressure (hazard of infection). We assume that the "Removed" individuals are no longer able to pass infection and cannot be reinfected.

A.1. Network dynamics assumptions

The dynamic model of individuals transition between the states S, I and R is based on several simple principles that allow to extract a mathematical model of the ongoing epidemic and relate it to observable data:

1. The spread occurs over a network of contacts, that is, an infectious individual may only infect his/her immediate neighbors at fixed rate $\beta > 0$; it is assumed that the average number of node's contacts is $\mu > 0$.
2. The infected individual may recover at rate $\gamma > 0$ or be restricted from contacting his/her network neighbors either through mandatory or voluntary quarantine at rate $\delta > 0$.
3. The infected individuals have an infectious period that is a random sample from an exponential distribution.
4. The symptomatic infectives are infectious and the partial count of new infectives is observed over time with a negligible chance of false positives.

We note that the model as described here extends previous work studying epidemics on CM-random networks (e.g., Newman, 2003) and that this formulation above can also account for extensions of the basic SIR compartments to include a latent period (up to 14 days for COVID-19, see Niehus et al., 2020), as well as more general staged progression models (Hethcote, 2000).

Instead of directly analyzing the stochastic CM model described above, which is challenging due to heterogeneity in the number of contacts and the evolution of the connectivity structure (e.g., Bartlett, 1960; Kermark and Mckendrick, 1927), we make use of the general results on the mean field approximation (Durrett, 2007; Ball and Neal, 2008) and the convergence of the random infection hazard in large networks. Specifically, as shown in Jacobsen et al. (2018) under the assumption of the Poisson-type (binomial, Poisson, or negative binomial) degree distribution, the mean field approximation of the dynamics is given by the following set of differential equations (where dots denote time derivatives):

$$
\begin{aligned}
\dot{x}_S &= -\beta x_D x_S \\
\dot{x}_I &= \beta x_D x_S - \gamma x_I \\
\dot{x}_D &= \beta(1-\kappa)x_D^2 + \left(\kappa\mu\beta x_S^{2\kappa-1} - \tilde{\gamma}\right) x_D
\end{aligned}
\tag{1}
$$

where the pair (x_S, x_I) describes the relative number of susceptibles and infecteds, $x_D = x_{SI}/x_S$ is the relative density of infectious connections, κ is the average contact network density,[1] and $\tilde{\gamma} = \beta + \gamma + \delta$. The usual initial conditions are $x_S(0) = 1$, $x_I(0) = \rho > 0$, and $x_D(0) = \mu\rho$.

A.2. Likelihood and reproduction numbers

For the purpose of statistical analysis of the system (1) we make an assumption that only the empirical counts of the new infected are available in practice. Dividing the last equation in (1) by the first one, solving for x_D in terms of x_S, and substituting back into the first equation, we obtain a reduced system with only one equation describing the decay of susceptibles. To simplify notation, denote $S_t := x_S(t)$ to obtain

$$
-\dot{S}_t = \tilde{\beta}(1-S_t^K)S_t^K + \frac{\tilde{\gamma}}{1-K}S_t(1-S_t^{K-1}) + \tilde{\rho}S_t^K
\tag{2}
$$

where $S_0 = 1$ and $\tilde{\rho} = \beta\mu\rho$, $\tilde{\gamma} = \beta + \gamma + \delta$, and $\tilde{\beta} = \mu\beta$. Note that Eq. (2) is defined for $K = 1$ by taking the limit $K \to 1$. The value of the basic reproduction number is

$$
R_0 = \kappa\beta/\gamma.
$$

for both the full (1) and reduced (2) systems.

The condition $\kappa = 1$ implies the Poisson degree assumption for the pairwise model and reduces equation (2) to

$$
-\dot{S}_t = \tilde{\beta}\left(S_t - S_t^2\right) + \tilde{\gamma}\,S_t \log\left(S_t\right) + \tilde{\rho}S_t \quad \text{and} \quad S_0 = 1.
\tag{3}
$$

Instead of thinking about S_t as a proportion of susceptibles it is convenient to think about S_t as an improper survival function for *the time to infection* of a single randomly chosen susceptible. Then S_t has an improper density $-\dot{S}_t$ (see KhudaBukhsh et al., 2020). It is improper since $\int_0^\infty -\dot{S}_t \, dt = 1 - S_\infty = \tau < 1$ where τ is defined below (see also (Bastian and Rempala, 2020, Example 2)). Under this survival function interpretation, the infection time for a randomly selected initially susceptible individual (in an infinite population) follows a temporal pattern according to the probability law S_t given by (2) or (3).

When we observe only a partial epidemic trajectory, say until time T, then the observed infection time is conditional on the infection occurring by time T, that is, on an event that has probability $\tau_T = 1 - S_T$. It is easy to show that as $T \to \infty$ then $\tau_T \to \tau_\infty = \tau$, the

---

[1] The network parameter $\kappa > 0$ is defined as the ratio of network mean excess degree and mean degree, see Jacobsen et al. (2018) for details. It is known that $K = 1$ corresponds to the Poisson degree network whereas $K > 1$ corresponds to the negative binomial one.
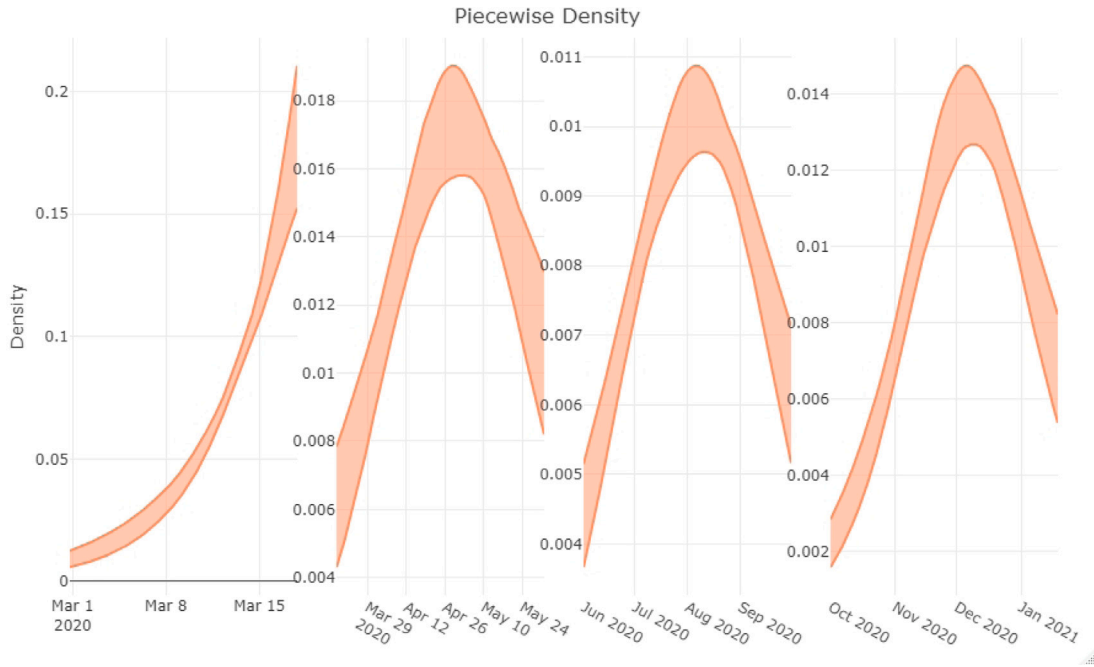
Fig. 5. Estimated density of infection times in the four segments of the epidemic.

probability of a randomly selected susceptible individual *being infected during the lifetime of an epidemic*. One may also think of τ as the final proportion of infected in the epidemic in infinite population. Then the conditional density of symptom onset is

$$f_T(t) = -\dot{S}_t / \tau_T,$$

which is simply the scaled derivative of the probability of staying susceptible past time t (denoted S_t). Accordingly, setting $\theta = (\kappa, \beta, \gamma, \rho)$, the approximate likelihood of the joint symptom times (epidemic curve) of n observed new cases by current time T in an infinite population (see KhudaBukhsh et al., 2020) is given by

$$\mathcal{L}(\theta | t_1, \ldots, t_n, T) = \prod_{i=1}^{n} f_T(t_i). \tag{4}$$

Although the expression above looks simple, note that the function $f_T(t)$ depends upon the vector of parameters θ only implicitly through the differential Eqs. (2) or (3).

A.3. Estimating dropout and recovery rates

Recall that $\tilde{\gamma} = \beta + \gamma + \delta$. Given $\tilde{\gamma}$, we may estimate the recovery rate γ from the recovery density. Then we have the approximate expression for drop-out

$$\delta \simeq \tilde{\gamma} - \gamma$$

assuming β is negligible.

To estimate γ, we consider now the recovery density. As shown in KhudaBukhsh et al. (2020), the density of daily recovery times is given by

$$g(t) = \int_0^t f_\infty(u) \gamma e^{-\gamma(t-u)} \, du.$$

For practical model fitting, a shift parameter $\varepsilon \in \mathcal{R}$ may be needed as

$$\bar{g}_\varepsilon(t) = \frac{g(t + \varepsilon \wedge 0)}{\int_0^\infty g(u + \varepsilon \wedge 0) \, du}.$$

The overall density of recovery is then the following mixture

$$\tilde{g}(t) = \frac{1}{1+\rho} \bar{g}_\varepsilon(t) + \frac{\rho}{1+\rho} \gamma e^{-\gamma t}.$$

Using the conditional recovery density

$$\check{g}_T(t) = \frac{\tilde{g}(t)}{\int_0^T \tilde{g}(t) \, dt} \quad \text{for} \quad t \in [0, T].$$

a likelihood analogous to Eq. (4) can now be produced:

$$\mathcal{L}(\gamma | t_1, \ldots, t_k, T) = \prod_{i=1}^{n} \check{g}_T(t_i). \tag{5}$$

This likelihood is used to estimate γ directly conditional on estimated S_t (and thus f_∞).

A.4. Incorporation of a change point

As outlined in an IDI blog post by the COVID-19 response modeling team (OSU / IDI COVID-19 Modeling Response Team, 2020), the initial exponential growth phase was followed by a plateau, a slow ascent, a decline, and then, a second exponential growth phase. This pattern can be attributed to changes in human behavior. As a consequence of the unusual epidemic curve, we need to accommodate change points in the model. Specifically, we assume the parameter vector θ takes different values in different segments of the epidemic separated by the change points. For the sake of simplicity, we illustrate the revised likelihood function with a single change point.

Let $A = (0, T^\star]$, and $B = (T^\star, T]$ denote the two segments of the epidemic with a change point at $T^\star < T$. We assume the parameter θ takes value θ_A in segment A and θ_B in segment B. Then, the conditional densities in each of the segments are

$$f_A(t) = \frac{-\dot{S}_t(\theta_A) \mathbb{1}_A(t)}{1 - S_{T^\star}(\theta_A)},$$
$$f_B(t) = \frac{-\dot{S}_t(\theta_B) \mathbb{1}_B(t)}{S_{T^\star}(\theta_B) - S_T(\theta_B)}, \tag{6}$$

where $S_t(\theta_A)$ (or $S_t(\theta_B)$) is the solution to either (2) or (3) with $\theta = \theta_A$ (or $\theta = \theta_B$). Write $\tau_T(\cdot) = 1 - S_T(\cdot)$. Then, the conditional density over the entire time interval $[0, T]$ is

$$f_T(t) = \frac{\tau_{T^\star}(\theta_A)}{\tau_T(\theta_B) - \tau_{T^\star}(\theta_B) + \tau_{T^\star}(\theta_A)} f_A(t)$$
$$+ \frac{\tau_T(\theta_B) - \tau_{T^\star}(\theta_B)}{\tau_T(\theta_B) - \tau_{T^\star}(\theta_B) + \tau_{T^\star}(\theta_A)} f_B(t). \tag{7}$$
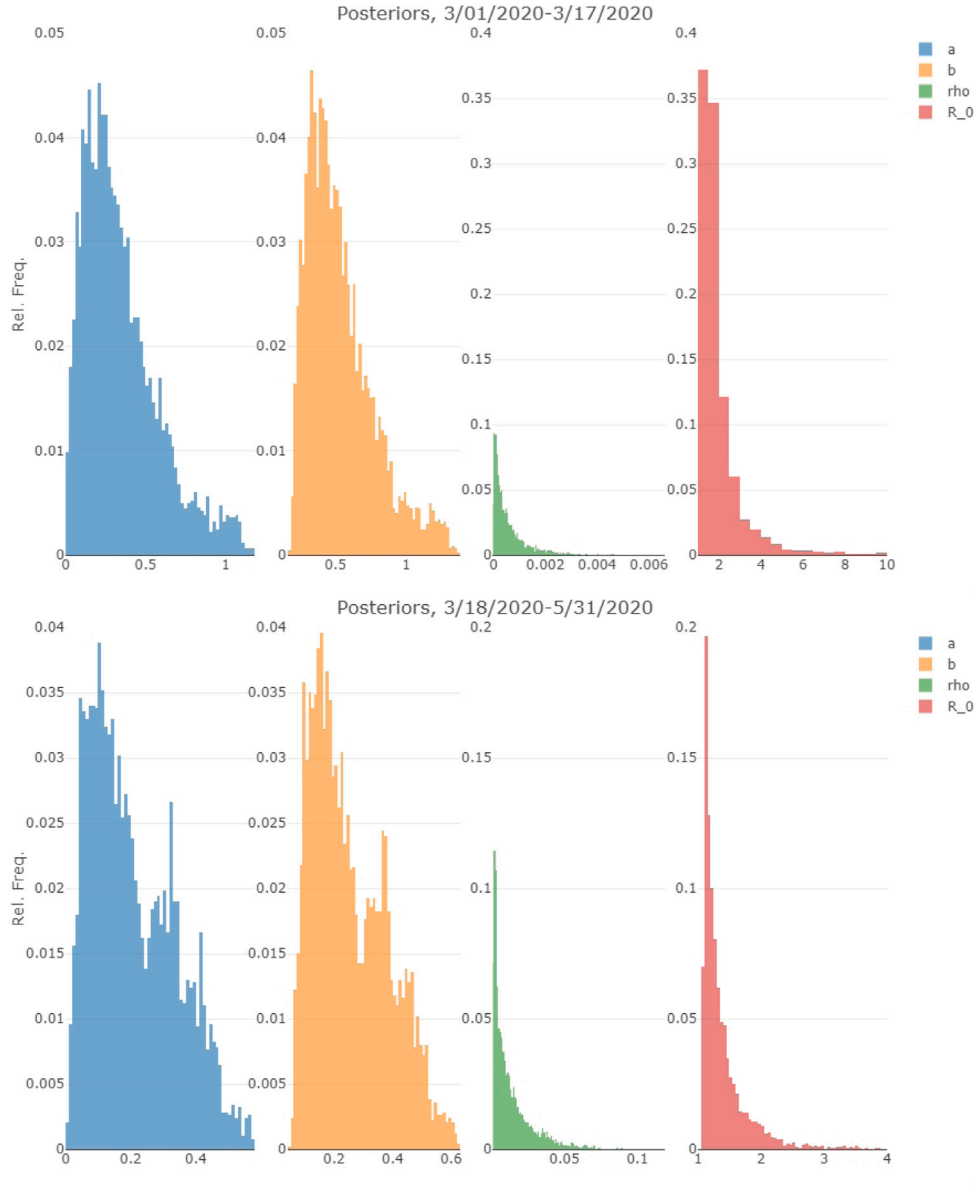
Fig. 6. Posterior densities of the fitted parameters in the first two segments of the pandemic. Here, $a = \tilde{\gamma}$, and $b = \tilde{\beta}$.

The final likelihood function looks exactly like (4) with the conditional density f_T given above in (7).

A.5. Estimating the final size of an outbreak

We assume that the size of an outbreak k_∞ is a fixed integer representing the likely number of total infections in the contact network of the confirmed cases only. Hence k_∞ is not the prevalence of the disease in the population but rather an estimate of the total outbreak size in the community of n individuals where we see infections. We estimate n at any given time T by the discount estimator $\hat{n}_T = k_T/(1 - S_T)$ where k_T is the number of cases observed by time T. Then we estimate the total number of cases by the end of an epidemic as

$$\hat{k}_\infty = \frac{\tau k_T}{1 - S_T}$$

where $\tau = 1 - S_\infty$ is the final probability of infection defined in Section 2.2.

A.6. Prediction and uncertainty quantification

We follow a Hamiltonian Monte Carlo-based Bayesian approach for prediction and uncertainty quantification. We assume independent non-informative priors for the parameters $\tilde{\beta}, \tilde{\gamma}$ and $\tilde{\rho}$. We approximate the posterior density of θ by drawing posterior samples using the *No U-Turn Sampling* (NUTS) method. As a point estimate of θ, we take the mean of the posterior density. The cumulative epidemic curve obtained as a solution to either (2) or (3) corresponding to the point estimate defined above gives us the most likely trajectory. The posterior samples of $\tilde{\beta}, \tilde{\gamma}$ and $\tilde{\rho}$ are used to generate a pointwise Monte Carlo confidence interval around the most likely trajectory. In other words, we essentially generate predicted trajectories corresponding to the posterior samples and then compute appropriate quantiles at desired time points to get the confidence interval. However, as shown in Fig. 3, the confidence bound obtained this way is narrow and suggests that the procedure might be underestimating the variance or over-dispersion in the observed daily new case counts. This over-dispersion could be a result of testing delays or other systematic issues with data collection. Therefore, we adopt an empirical variance adjustment method.
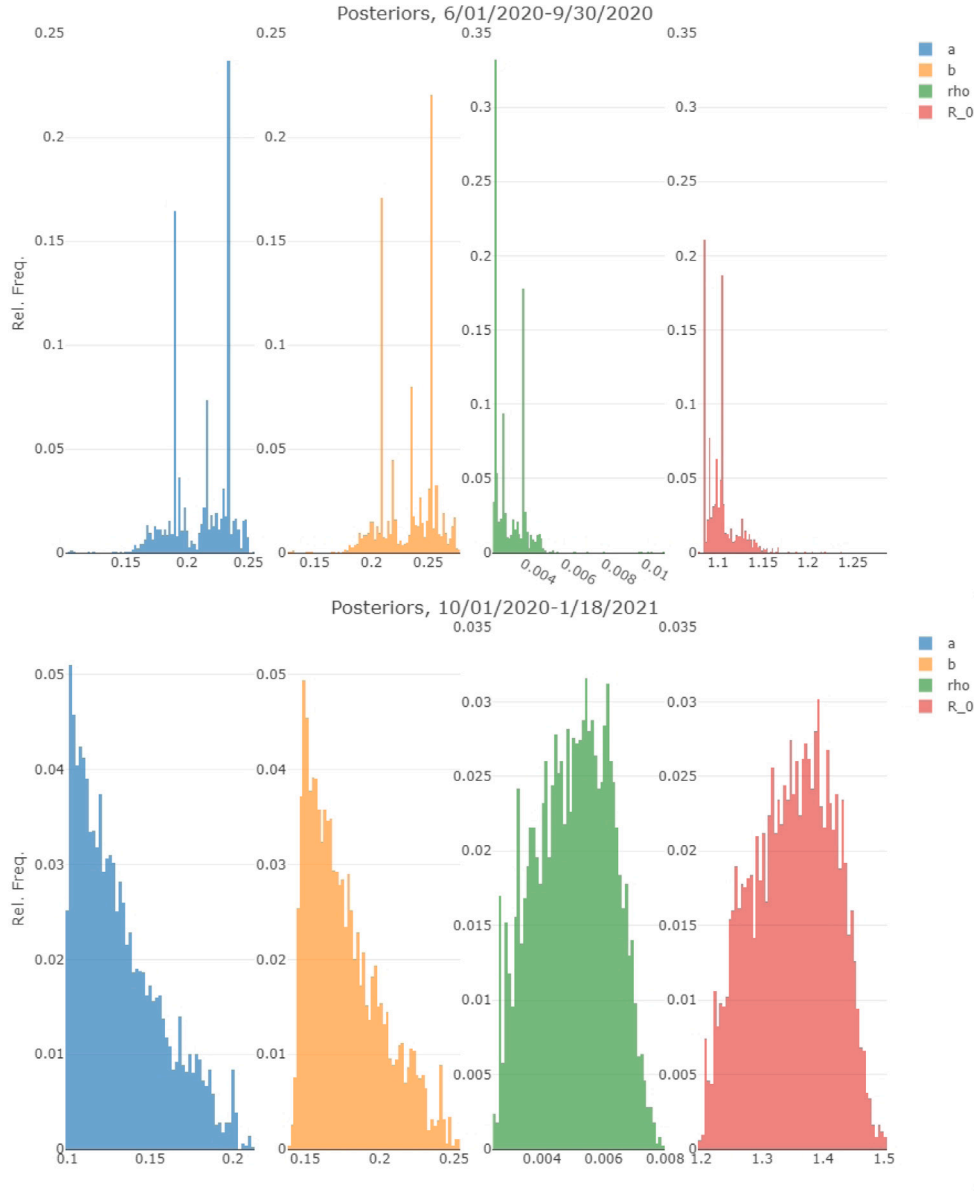
Fig. 7. Posterior densities of the fitted parameters in the last two segments of the pandemic. Here, $a = \tilde{\gamma}$, and $b = \tilde{\beta}$.

In simple terms, the empirical variance adjustment could be explained as follows: We first smooth the observed counts using a kernel smoother with span = 0.2. We used the function `supersmoother` in R. Then, we estimate the variance inflation factor by averaging the squared difference of the actual daily new case counts and the smoothed ones. Finally, the variance adjusted confidence bounds are obtained by multiplying the upper confidence bounds from the corresponding a Gaussian approximation with the standard deviation of the variance inflation factor and taking lower confidence bound from the DSA fits.

A.7. Additional numerical results

Here, we present additional diagnostic figures for our model fit. In Fig. 5, we show the estimated density of the infection times in the three segments of the epidemic. This is the density that contributes to the likelihood function (4). Fig. 5 shows the posterior densities of the fitted parameters in the three segments. Fig. 8 shows the trace plots of the fitted parameters as a diagnostic measure of the convergence of the Hamiltonian Monte Carlo chains. In Fig. 9, we show how our

predictions compare against other methods. In particular, we compare our results with the predictions provided by the Institue for Health Metrics and Evaluation.

Appendix B. Estimating hospitalizations from predictions of statewide case numbers

Let $C(t)$ be the trajectory of case onset times as generated by the statewide model. We wish to translate this to hospital census $h_i(t)$ for a geographic region i. Consider an age group $a \in \mathcal{A}$. Let $n_{i,a}$ be the number of individuals in age group a in geographic region i, and let $n_i = \sum_{a \in \mathcal{A}} n_{i,a}$ denote the total population size of i. Let $N = \sum_i n_i$ denote the entire population size of Ohio. Converting $C(t)$ to $h_i(t)$ involves the following steps:

   (i) Estimating the case onsets $c_{i,a}(t)$ for age group a in geographic region i.
  (ii) Deriving from this the onsets of severe cases $s_{i,a}(t)$ that will eventually require hospitalization by age group and geographic region.
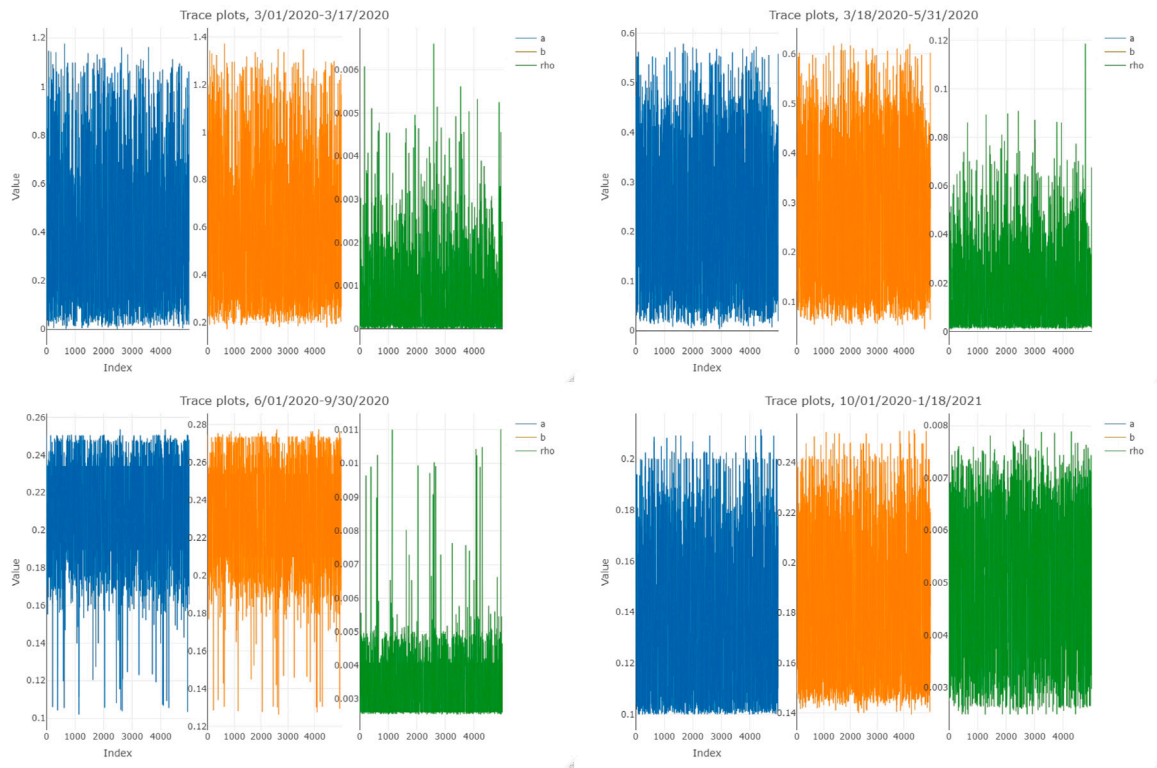
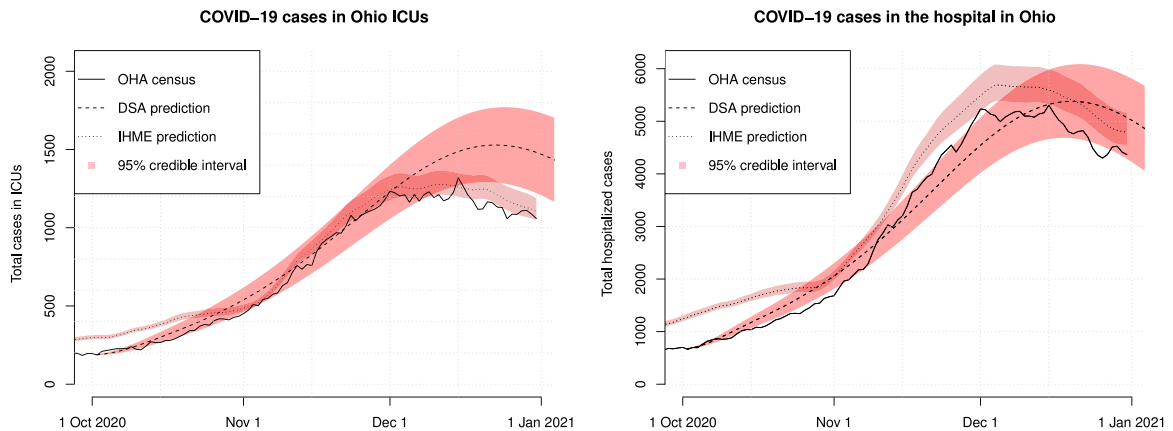**Fig. 8.** Trace plots of the fitted parameters in the four segments.

Fig. 9. In this figure, we compare our predictions from October 2 with predictions downloaded from Institute for Health Metrics and Evaluation (IHME) on November 29, 2022 (Institute for Health Metrics and Evaluation, 2022) and the true reported cases (Ohio Hospitals Association, 2022). It appears that both approaches are able to predict total hospitalizations and ICU admissions reasonably well. The IHME credible intervals are generally narrower than those of our DSA method, but the two sets of predictions are not directly comparable because they are using information from different time periods. **(Left)** Predictions of ICU admissions in Ohio after the third change point (October 1, 2021). **(Right)** Predictions of total hospital admissions in Ohio after the third change point (October 1, 2021).

(iii) Using probability distributions for time from case onset to hospitalization and length of stay to estimate the hospital census $h_{i,a}(t)$.

Estimating case onsets by age group and region. We consider the following factors in estimating $c_{i,a}(t)$ from total case onsets $C(t)$: the population size of i, age structure of i, and the relative likelihood by age of being identified as a COVID-19 case. We first compute the total case onsets $c_i(t) := \sum_{a \in \mathcal{A}} c_{i,a}(t)$ for i in proportion to the population size of i:

$$c_i(t) = \frac{n_i}{N} C(t). \tag{8}$$

To distribute these case onsets by age, define the *relative susceptibility* s_a of age group a as the ratio of the proportion of observed cases in a relative to the proportion of the state's population that is in a.

A relative susceptibility of one corresponds to the proportion of observed cases across the state matching what would be expected if cases were distributed uniformly at random across individuals. Relative susceptibilities larger than one correspond to more cases identified in a than would be expected at random, while relative susceptibilities less than one correspond to fewer identified cases in a than would be expected at random. In the Ohio data, we observe far fewer cases in youth (ages less than 21) than would be expected at random ($s_{a<21} \approx 0.12$), but more cases in the elderly ($s_{a \geq 85} \approx 1.7$). We then compute $c_{i,a}(t)$ by multiplying $c_i(t)$ by a probability incorporating the relative

susceptibility of a and the age structure of i:

$$c_{i,a}(t) = c_i(t) \frac{n_{a,i} s_a}{\sum_{\ell \in \mathcal{A}} n_{i,\ell} s_\ell}. \tag{9}$$

Onset of cases that will eventually require hospitalization. It has been documented that case outcome varies strongly with age (CDC COVID-19 Response Team, 2020b). We use the mean of the probabilities reported in CDC COVID-19 Response Team (2020a) of severe case outcome by age group. Let p_a be the probability of severe case outcome for age group a. Then

$$s_{i,a}(t) = p_a c_{i,a}(t). \tag{10}$$

For simplicity, we do not include other demographic covariates such as gender or comorbidities in the p_a.

Hospital census over time. The trajectories $s_{i,a}(t)$ correspond to onset times of cases that will eventually require hospitalization. To translate these into hospital census of COVID-19 patients, let $w_a(t)$ be the probability distribution for time between case onset and hospitalization, and $q_a(t)$ the probability distribution for length of stay in hospital for age group a. Let $A_{i,a}(t)$ and $D_{i,a}(t)$ denote hospital admissions and discharges, respectively, for location i and age group a at time t. Then $A_{i,a}(t)$ corresponds to the convolution of $s_{i,a}$ with w_a, and $D_{i,a}(t)$ corresponds to the convolution of $A_{i,a}$ with $q_a(t)$. Let $C_{i,a}(t)$ denote the hospital census for location i and age group a at time t. The difference between admissions and discharges corresponds to the change in hospital census, so integrating gives $C_{i,a}(t)$:

$$
\begin{aligned}
A_{i,a}(t) &= (s_{i,a} * w_a)(t) \\
D_{i,a}(t) &= (A_{i,a} * q_a)(t) \\
C_{i,a}(t) &= C_{i,a}(0) + \int_0^t A_{i,a}(\xi) - D_{i,a}(\xi)\, d\xi.
\end{aligned}
\tag{11}
$$

For simplicity, we have not distinguished between hospital bed types (e.g. ICU vs. non-ICU) above.

References

Agrawal, S., Bhandari, S., Bhattacharjee, A., Deo, A., Dixit, N.M., Harsha, P., Juneja, S., Kesarwani, P., Swamy, A.K., Patil, P., Rathod, N., Saptharishi, R., Shriram, S., Srivastava, P., Sundaresan, R., Vaidhiyan, N.K., Yasodharan, S., 2020. City-scale agent-based simulators for the study of non-pharmaceutical interventions in the context of the COVID-19 epidemic. J. Indian Inst. Sci. 100 (4), 809–847,

Ball, F., Neal, P., 2008. Network epidemic models with two levels of mixing. Math. Biosci. 212 (1), 69–87.

Bartlett, M.S., 1960. Stochastic Population Models in Ecology and Epidemiology. Methuen.

Bastian, C.D., KhudaBukhsh, W.R., 2020. Python code for fitting DSA analysis. https://github.com/wasiur/dynamic_survival_analysis.

Bastian, C.D., KhudaBukhsh, W.R., Pan, Y., Kenah, E., Rempała, G.A., 2020. Predicting the Size and Duration of the Outbreaks of COVID-19 Under Minimal Assumptions. Technical Report, The Ohio State University College of Public Health.

Bastian, C.D., Rempała, G.A., 2020. Throwing stones and collecting bones: Looking for poisson-like random. Math. Methods Appl. Sci. 43.

Bollobás, B., 1998. Random Graphs. Springer.

CDC COVID-19 Response Team, 2020a. Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019 — United States, February 12–March 28, 2020. Morb. Mortal. Wkly. Rep. 69, 382–386.

CDC COVID-19 Response Team, 2020b. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) — United States, February 12–March 16, 2020. Morb. Mortal. Wkly. Rep. 69, 343–346.

Center for Systems Science and Engineering (CSSE) at the Johns Hopkins University, 2021. COVID-19 dashboard. online. URL https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6. (Accessed 12 July 2021).

Childs, M., Kain, M., Kirk, D., Harris, M., Ritchie, J., Couper, L., Delwel, I., Nova, N., Mordecai, E., 2020. Potential long-term intervention strategies for COVID-19. https://Covid-Measures.Github.Io/.

Di Lauro, F., KhudaBukhsh, W.R., Kiss, I.Z., Kenah, E., Jensen, M., Rempała, G.A., 2022. Dynamic survival analysis for non-Markovian epidemic models. J. R. Soc. Interface 19 (191), 20220124.

Durrett, R., 2007. Random Graph Dynamics. Vol. 200. Cambridge University Press.

Ferguson, N.M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L.C., van Elsland, S., Thompson, H., Verity, R., Volz, E., Wang, H., Wang, Y., Walker, P.G.T., Walters, C., Winskill, P., Whittaker, C., Donnelly, C.A., Riley, S., Ghani, A.C., 2020. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. http://dx.doi.org/10.25561/77482.

Grasselli, G., Pesenti, A., Cecconi, M., 2020. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: Early experience and forecast during an emergency response. JAMA 323 (16), 1545–1546.

Hethcote, H.W., 2000. The mathematics of infectious diseases. SIAM Rev. 42 (4), 599–653.

IHME COVID-19 health service utilization forecasting team, Murray, C.J., 2020. Forecasting COVID-19 Impact on Hospital Bed-Days, ICU-Days, Ventilator-Days and Deaths by US State in the Next 4 Months. Cold Spring Harbor Laboratory Press, http://dx.doi.org/10.1101/2020.03.27.20043752, MedRxiv.

Institute for Health Metrics and Evaluation, 2022. COVID-19 estimate downloads. URL https://www.healthdata.org/covid/data-downloads. Last Accessed 29 November 2022.

Jacobsen, K.A., Burch, M.G., Tien, J.H., Rempala, G.A., 2018. The large graph limit of a stochastic epidemic model on a dynamic multilayer network. J. Biol. Dyn. 12 (1), 746–788.

Kermack, M., Mckendrick, A., 1927. Contributions to the mathematical theory of epidemics. Part I. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 115, 700–721.

KhudaBukhsh, W.R., Choi, B., Kenah, E., Rempała, G.A., 2020. Survival dynamical systems: individual-level survival analysis from population-level epidemic models. Interface Focus 10 (1), 20190048.

KhudaBukhsh, W.R., Khalsa, S.K., Kenah, E.E., Rempała, G.A., Tien, J.H., 2021. COVID-19 dynamics in an Ohio prison. Preprint. MedRxiv. URL https://www.medrxiv.org/content/10.1101/2021.01.14.21249782v1.

Kiss, I.Z., Miller, J.C., Simon, P.L., 2017. Mathematics of Epidemics on Networks. Springer.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020. Substantil undocumented infection facilitates the rapid dissemination of novel coronavirus (COVID-19). MedRxiv. https://doi.org/10.1101/2020.02.14.20023127.

Newman, M.E., 2003. The structure and function of complex networks. SIAM Rev. 45 (2), 167–256.

Newman, M.E.J., 2018. Networks, second ed. Oxford University Press.

Niehus, R., Salazar, P.M.D., Taylor, A., Lipsitch, M., 2020. Quantifying bias of COVID-19 prevalence and severity estimates in wuhan, china that depend on reported cases in international travelers. MedRxiv.

Ohio Hospitals Association, 2022. OHA COVID-19 dashboard. URL https://ohiohospitals.org/covid19data. (Last Accessed from Within US: 29 November 2022.

Okabe, A., Boots, B., Sugihara, K., 2000. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, second ed. Wiley.

OSU / IDI COVID-19 Modeling Response Team, 2020. A note from the OSU / IDI COVID-19 modeling response team. online. URL https://u.osu.edu/eeph/a-note-from-the-osu-idi-covid-19-modeling-response-team/.

Predictive Healthcare at Penn Medicine, 2020. COVID-19 hospital impact model for epidemics (CHIME). https://penn-chime.phl.io/.

Rempała, G.A., 2020. Mathematical Models of Epidemics: Tracking Coronavirus using Dynamic Survival Analysis. https://mbi.osu.edu/events/seminar-grzegorz-rempala-mathematical-models-epidemics-tracking-coronavirus-using-dynamic.

Somekh, I., KhudaBukhsh, W.R., Root, E.D., Boker, L.K., Rempała, G.A., Simões, E.A.F., Somekh, E., 2022. Quantifying the population-level effect of the COVID-19 mass vaccination campaign in Israel: A modeling study. Open Forum Infect. Dis. 9 (5).

The Dartmouth Institute for Health Policy and Clinical Practice, 2022. The Dartmouth Atlas of Healthcare. https://www.dartmouthatlas.org/faq/#research-methods-faq.

US Census Bureau, 2022. American Community Survey. https://www.census.gov/programs-surveys/acs.

US Census Bureau, 2022. Mapping Files A. https://www.census.gov/geographies/mapping-files.html.

Verity, R., Okell, L.C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P.G.T., Fu, H., Dighe, A., Griffin, J.T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., FitzJohn, R., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Laydon, D., Nedjati-Gilani, G., Riley, S., van Elsland, S., Volz, E., Wang, H., Wang, Y., Xi, X., Donnelli, C.A., Ghani, A., Ferguson, N.M., 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect. Dis. 20.

Vossler, H., Akilimali, P., Pan, Y., KhudaBukhsh, W.R., Kenah, E., Rempała, G.A., 2022. Analysis of individual-level data from 2018–2020 Ebola outbreak in Democratic Republic of the Congo. Sci. Rep. 12 (1), 5534.

Wascher, M., Schnell, P.M., KhudaBukhsh, W.R., Quam, M., Tien, J.H., Rempała, G.A., 2021. Monitoring SARS-COV-2 transmission and prevalence in populations under repeated testing. Preprint. MedRxiv. URL https://www.medrxiv.org/content/10.1101/2021.06.22.21259342v1.

Weitz, J.S., 2020. COVID-19 near-term epidemic risk assessment for Georgia. https://github.com/jsweitz/covid-19-ga-summer-2020.

Wu, Z., McGoogan, J.M., 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese center for disease control and prevention. JAMA 323 (13), 1239–1242.