

## Highlights

### **Projecting COVID-19 Cases and Subsequent Hospital Burden in Ohio**

Wasiur R. KhudaBukhsh, Caleb Deen Bastian, Matthew Wascher, Colin Klaus, Saumya Yashmohini Sahai, Mark Weir, Eben Kenah, Elisabeth Root, Joseph H. Tien, Grzegorz Rempala

- We present a novel statistical approach called Dynamic Survival Analysis (DSA) to model an epidemic curve with incomplete data. The DSA approach is advantageous over standard statistical methods primarily because it does not require prior knowledge of the size of the susceptible population, the overall prevalence of the disease, and also the shape of the epidemic curve.
- The principal motivation behind the study was to obtain predictions of case counts of COVID-19 and the resulting hospital burden in the state of Ohio during the early phase of the pandemic.
- The proposed methodology was applied to the COVID-19 incidence data in the state of Ohio to support the Ohio Department of Health (ODH) and the Ohio Hospital Association (OHA) with predictions of hospital burden in each of the Hospital Catchment Areas (HCAs) of the state.

# Projecting COVID-19 Cases and Subsequent Hospital Burden in Ohio

Wasiur R. KhudaBukhsh<sup>a,\*</sup>, Caleb Deen Bastian<sup>b</sup>, Matthew Wascher<sup>c</sup>, Colin Klaus<sup>d,ef</sup>, Saumya Yashmohini Sahai<sup>g</sup>, Mark Weir<sup>e,f,h</sup>, Eben Kenah<sup>e,f</sup>, Elisabeth Root<sup>i</sup>, Joseph H. Tien<sup>d,ef,j</sup> and Grzegorz Rempala<sup>d,ef,j,\*\*</sup>

<sup>a</sup>*School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*

<sup>b</sup>*Applied Mathematics, Princeton University and Massive Dynamics, Princeton, NJ, USA*

<sup>c</sup>*Department of Mathematics, University of Dayton, 300 College Park, Dayton, Ohio 45469, USA*

<sup>d</sup>*Mathematical Biosciences Institute, The Ohio State University, 1735 Neil Avenue, Columbus OH 43210, USA*

<sup>e</sup>*College of Public Health, The Ohio State University, Cunz Hall, 1841 Neil Avenue, Columbus, OH 43210, USA*

<sup>f</sup>*Infectious Diseases Institute, 208 Bricker Hall, 190 North Oval Mall, Columbus, OH 43210-1358, USA*

<sup>g</sup>*Department of Computer Science and Engineering, The Ohio State University, 395 Drees Laboratories, 2015 Neil Avenue, Columbus, OH 43210, USA*

<sup>h</sup>*The Sustainability Institute, The Ohio State University, 74 W. 18th Avenue, Columbus, OH 43210, USA*

<sup>i</sup>*Institute for Disease Modeling, The Bill & Melinda Gates Foundation, Seattle, Washington, USA*

<sup>j</sup>*Department of Mathematics, The Ohio State University, 231 West 18th Avenue Columbus, OH 43210-1174, USA*

---

## ARTICLE INFO

**Keywords:**

COVID-19

Dynamic Survival Analysis

SIR model

Prediction

---

## ABSTRACT

As the Coronavirus 2019 (COVID-19) disease started to spread rapidly in the state of Ohio, the Ecology, Epidemiology and Population Health (EPPH) 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 Dynamic 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 has been made available publicly.

---

## 1. Introduction

The coronavirus 2019 (COVID-19) disease has resulted in 267 million confirmed cases and 5 million deaths reported globally as of 8 December 2021 [10]. As evidenced by epidemics in many countries, particularly in Italy and Spain, COVID-19 patient case loads have the potential to overwhelm healthcare systems [14].

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 (EPPH) program within IDI took the initiative to offer epidemic modeling and decision analytics support to ODH preparatory planning and response to the COVID-19 pandemic, specifically the Ohio epidemic.

---

\*Corresponding author

\*\*Principal corresponding author

✉ [wasiur.khudaBukhsh@nottingham.ac.uk](mailto:wasiur.khudaBukhsh@nottingham.ac.uk) (W.R. KhudaBukhsh); [cbastian@princeton.edu](mailto:cbastian@princeton.edu) (C.D. Bastian); [mwascher1@udayton.edu](mailto:mwascher1@udayton.edu) (M. Wascher); [klaus.68@mbi.osu.edu](mailto:klaus.68@mbi.osu.edu) (C. Klaus); [sahai.17@osu.edu](mailto:sahai.17@osu.edu) (S.Y. Sahai); [weir.95@osu.edu](mailto:weir.95@osu.edu) (M. Weir); [kenah.1@osu.edu](mailto:kenah.1@osu.edu) (E. Kenah); [root.145@osu.edu](mailto:root.145@osu.edu) (E. Root); [tien.20@osu.edu](mailto:tien.20@osu.edu) (J.H. Tien); [rempala.3@osu.edu](mailto:rempala.3@osu.edu) (G. Rempala)

🌐 <https://wasiur.xyz/> (W.R. KhudaBukhsh); <https://cph.osu.edu/people/mweir> (M. Weir);

<https://cph.osu.edu/people/ekenah> (E. Kenah); <https://u.osu.edu/root.145/> (E. Root);

<https://people.math.osu.edu/tien.20/> (J.H. Tien); <https://cph.osu.edu/people/grempala> (G. Rempala)

ORCID(s): 0000-0003-1803-0470 (W.R. KhudaBukhsh); 0000-0002-7117-7773 (E. Kenah); 0000-0002-6307-4555 (G. Rempala)

Just as in any computational modeling effort, there are many approaches and methods to choose from to model emerging epidemics. Different methods are suited to different situations in terms of data, scenarios, conditions, and urgency. This is no different for modeling the COVID-19 pandemic, either nationally or for the state of Ohio. Modeling approaches include: compartmental models [11, 28, 35], statistical models [16], and agent-based simulations [13]. For the state of Ohio, the OSU-IDI group has 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 [24], 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 [17] 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 [19]. Critically, this allows for a more accurate quantification of the uncertainty in the model predictions.

Because of these features, 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, which occurs in the next step of the model.

This is done separately for illness severity and hospitalization since these rates vary for COVID-19 patients based on age and comorbidity stratification [31, 33]. Rather, the projected statewide model provides an estimate of the number of new and cumulative cases over time across the entirety of Ohio.

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 [31, 33], we use 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 [25] 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 *Dynamic Survival Analysis* (DSA) [20, 30, 34], which is an extension of *survival dynamical systems* [6, 19]. Details on this approach and the model development can be found in Appendix A.

Three key strengths of the DSA approach for predicting 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.

Since testing has focused on the most symptomatic and severe cases, we have 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 influenzas. 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 testing for COVID-19 in Ohio, which is affected by both limited testing capacity and undetected asymptomatic infections. The DSA method has been developed to handle such limited 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 made available as a GitHub repository [4].

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 [36]. 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 [17]. As a working model of a contact network we use a type of random graph called a dynamic configuration model (CM) [7]. The mean-field approximation for epidemic processes on such networks is sometimes referred to as a pairwise model [21].

*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—an unknown, but readily assumed component of contemporary models. 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, that is, an infectious individual can infect their immediate neighbors at a fixed positive rate. It is implicitly assumed that the average number of a person's contacts is also 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. Once infected, an infected individual has an infectious period that is assumed to follow 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 misdiagnosis (*i.e.*, false positives).

*Estimating Transmission Rates:* The complete model description and detailed development can be found in the 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 approach, 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 ( $\beta$  in the appendix), and then considering drop outs as a function of recovery rate (details in 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.

### **2.2.1. 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 can be seen in Appendix B. In short, this is developed in three steps and is integrated with the geographic modeling methods described next.

1. Estimate case onset for each age group based on age stratification in each geographic area modeled.
2. Estimate the number of severe cases that will need hospitalization based first on CDC data [32] and then updated with Ohio-specific data as those data become more conducive to 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.

### **2.3. Geographic Modeling Methods**

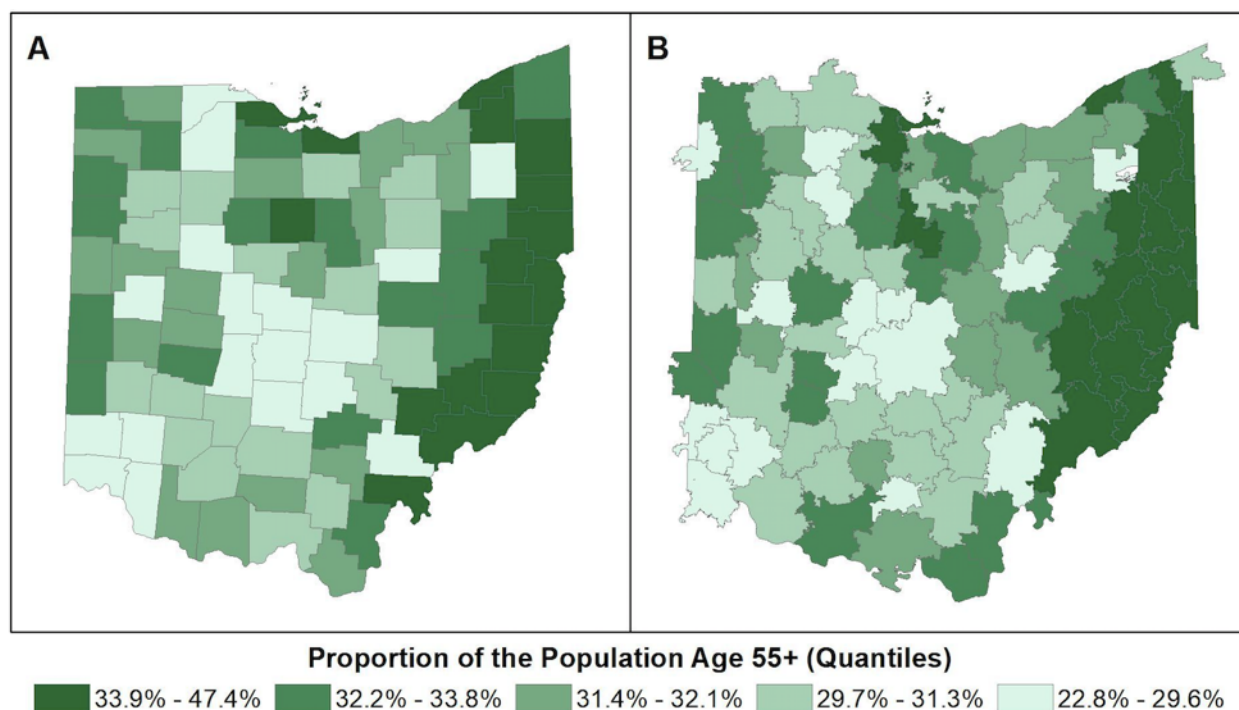
The time series of estimated cases of COVID-19 illness are output from the statewide model described above. These outputs are then used to assess hospital burden (*eg.* hospitalizations and ICU admissions) using age structure of the population and the estimated length of time a patient will be in a hospital 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 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 [32]. 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. We estimated the number of hospitalized cases by assuming higher rates of hospitalization for older segments of the population. Thus, 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 [8]. 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



**Figure 1:** Proportion of the population over the age of 55 in A) counties and B) hospital catchment areas

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. Figure 1 shows the distribution of the high risk population (age 55+) by county and Hospital Catchment Area 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* 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.

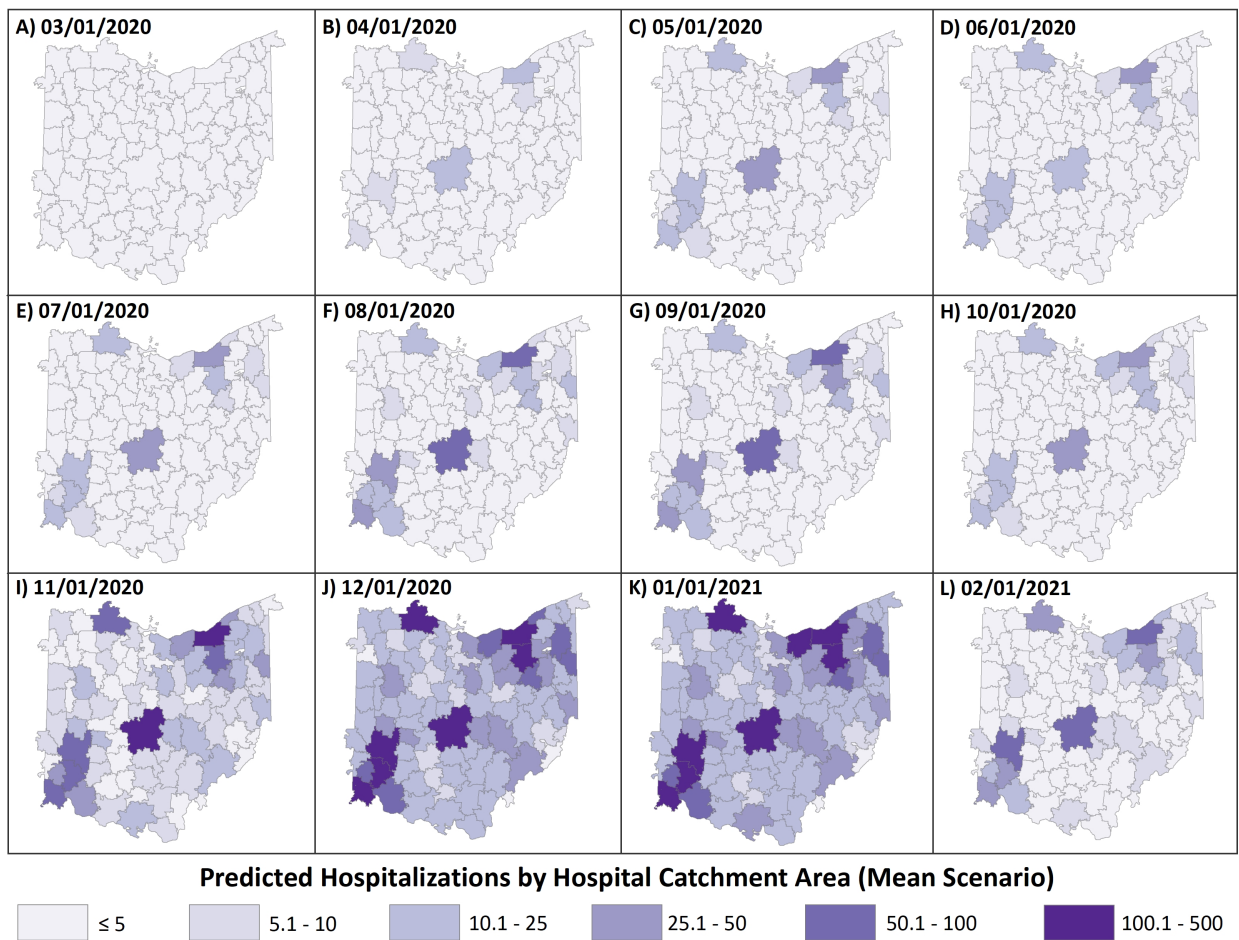
### 2.3.2. Definition of Hospital Catchment Areas

We developed small area estimates from state-level model predictions for counties and hospital catchment areas (HCA). Boundary files for the 88 Ohio counties were obtained from the U.S. Census Cartographic Boundary Files [9]. We developed hospital catchment areas using an approach modified from the Dartmouth Atlas Project [1]. We define a hospital catchment area as a collection of ZIP codes whose residents receive most of their hospital care from the hospital(s) in that region. Figure 1-B shows the HCAs developed and mapped. Note that in figure 1-B there are HCAs that cross over into neighboring states. This is explained in more detail in the subsection 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 [26]. 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 OH.
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.

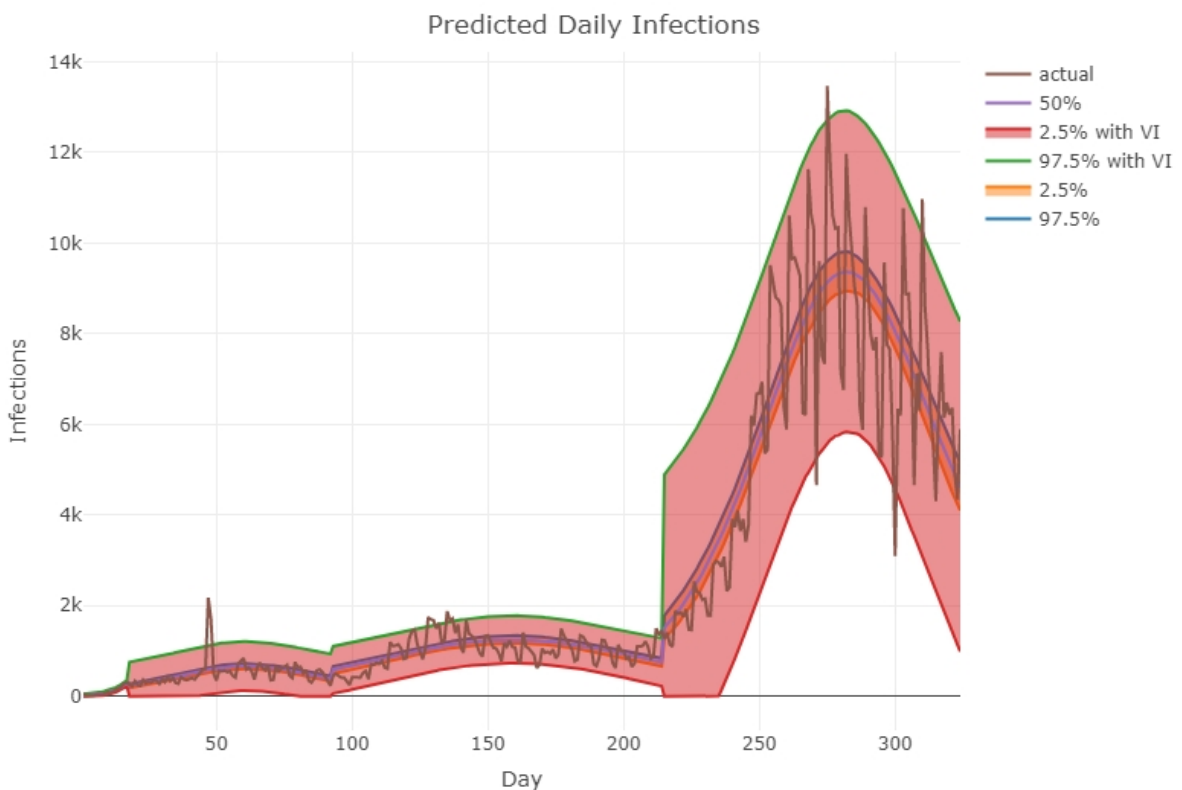


**Figure 2:** Estimates of hospital burden mapped to hospital catchment areas for 12 days during the pandemic

Using this procedure, we generated 96 Hospital Catchment Areas 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 Figure 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, then a phase of steady linear growth and a decline, and then finally, due to reopening of the state, another phase of exponential growth. 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.



**Figure 3:** Predictions based on the fitted model vs actual daily incidence in Ohio as reported by the Ohio Department of Health. We have 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.

In Figure 3, we compare actual counts of daily new infections against the model predictions. As one can see, the fitted trajectories follow the observed epidemic curve reasonably 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 provided in Appendix B. The daily counts thus predicted are then down-projected into predictions of hospitalization surge across the state in the geographic component of the model.

## 4. Discussion

### 4.1. Comparison with Other Approaches

Although, due to the significant differences in approach, it is difficult to directly compare our projective model to other contemporary methods, some indirect comparisons may be offered. The susceptible-infectious-recovered (SIR) framework is the basis for many COVID-19 epidemic models [11, 22, 28, 35]. Several websites provide implementations of the SIR framework and present scenarios with different interventions such as social distancing affecting the transmission parameter  $\beta$  [11, 28], thus allowing for different hypothetical scenarios to be explored. A key difference, however, is that phase-space variables from the SIR model are infections, rather than symptomatic cases. Consequently, knowledge of asymptomatic to symptomatic infections is needed to translate output from these models to cases and hospitalizations. Recent work [11] has attempted to extend the basic SIR model to include additional compartments corresponding to incubation, presymptomatic infectious, stages of symptomatic infectious, and hospitalization. Considering the lack of clear information on these types of outcomes, it is clearly difficult to implement this model as such detailed data are not currently available.

The current SIR tools [11, 28] 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 as 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 [22]. This resulted in a spatially explicit model without age-structure. Conversely, [35] is an example of a model with age structure but not spatial: 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 [16]. Another alternative is an agent-based model such as that used in [13]. Similar to [22], [13] 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 rationalize that a model output of a time series of illnesses is the more reasonable approach. Similarly, using the simplified framework and structure of our approach allow for greater flexibility under uncertain data inputs into the model, and a lack of typically required information such as total susceptible population, or asymptomatic rates.

## Acknowledgment

We would like to acknowledge Dominick Winecki for developing useful shell scripts during the early hectic days of the pandemic. We also acknowledge Ian Dunn for his help with the GIS software and graphics.

## Funding

WKB acknowledges the President's Postdoctoral Scholars Program (PPSP) of the Ohio State University (OSU). The work of GAR and EK was partially funded by the National Science Foundation (NSF) Rapid Grant DMS-2027001. EK and WKB were partially funded by National Institute of Allergy and Infectious Disease (NIAID) grant R01 AI116770. The content is solely the responsibility of the authors and does not represent the official views of the NSF or NIAID.

## Data Availability & Software

The software to perform model fit along with a data example is available from [4]. The onset data used in this paper can be downloaded from the Ohio Department of Health (ODH) COVID-19 dashboard at <https://coronavirus.ohio.gov/dashboards/overview>. 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.

## Additional Materials

A video presentation describing the methodology used for epidemic size predictions is available from [29].

### A. Predicting Statewide Cases of COVID-19

#### A.1. Detailed Description of Project Mode Development

This derivation is taken in whole from the manuscript in preparation [5]. 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 - *e.g. illnesses*. Our approach is derived from the general stochastic model of a contagion spread across a contact network of nodes representing individuals in a community [17]. As a working model of a contact network we use a dynamic configuration model (CM)-type random graph (see, *e.g.*, [7]). Such a model is often also referred to as a *pairwise model* [21].

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). We assume that the “Removed” individuals are no longer able to pass infection and cannot be reinfected.

#### A.2. 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 stay infected for a random amount of time distributed according to 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 misdiagnosis (false positives).

We note that the model as described here extends previous work studying epidemics on CM-random networks (*e.g.*, [23]) 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 [25]), as well as more general staged progression models [15].

Instead of directly analyzing the stochastic CM model described above, which is challenging due to the heterogeneity in the number of contacts and the evolution of the connectivity structure (*e.g.*, [3, 18]), we will make use of the general results on the mean field approximation [2, 12], and the convergence of the random infection hazard in large networks. Specifically, as shown in [17] under the assumption of the Poisson-type degree distribution (for instance the node degree is Poisson or negative-binomial distributed), the mean field approximation of the dynamics is given by the following set of differential equations (dots denote time derivatives)

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

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,  $\kappa$  is the average contact network density<sup>1</sup>, and  $\tilde{\gamma} = \beta + \gamma + \delta$ . The usual initial conditions are  $x_S(0) = 1$ ;  $x_I(0) = \rho > 0$ ;  $x_D(0) = \mu\rho$ .

<sup>1</sup>The network parameter  $\kappa > 0$  is defined as the ratio of network mean excess degree and mean degree, see [17] for details. It is known that  $\kappa = 1$  corresponds to the Poisson degree network whereas  $\kappa > 1$  corresponds to the negative binomial one.

### A.3. Estimating Reproduction Rate

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 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 the reduced system with only one equation describing the decay of susceptibles. To simplify notation, denote  $S_t := x_S(t)$  to obtain

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

where  $S_0 = 1$  and  $\tilde{\rho} = \beta\mu\rho$ ,  $\tilde{\gamma} = \beta + \gamma + \delta$ , and  $\tilde{\beta} = \mu\beta$ .

Note that the equation (2) is defined for  $\kappa = 1$  by taking the limit  $\kappa \rightarrow 1$ . The value of the basic reproduction number  $R_0$  for both reduced (2) and full (1) system is

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

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

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

Instead of thinking about  $S_t$  as a proportion of susceptibles it is convenient to think about  $S_t$  as an improper survival function of the *time to infection* of a single randomly chosen susceptible. Then  $S_t$  has an improper density  $-\dot{S}_t$  (see [19]). It is improper since  $\int_0^\infty -\dot{S}_t dt = 1 - S_\infty = \tau < 1$  where  $\tau$  is defined below (see also [6, 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  $T \rightarrow \infty$  then  $\tau_T \rightarrow \tau_\infty = \tau$ , the probability of a randomly selected susceptible individual *being infected during the lifetime of an epidemic*. (One may think of  $\tau$  also 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 [19]) is given by

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

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

### A.4. Estimating Dropout and Recovery Rates

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

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

To estimate  $\gamma$  we consider now the recovery density. As shown in [19], 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

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

The analogous likelihood as in (4) can be now produced using the conditional recovery density

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

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

This likelihood is used to estimate  $\gamma$  directly conditional on estimated  $S_t$  (and thus  $f_\infty$ ).

### A.5. Incorporation of a change point

As outlined in an IDI blog post by the COVID-19 response modeling team [27], 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 resultant unusual epidemic curve, we need to enhance the model to accommodate change points. In particular, we assume the parameter  $\theta$  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  $A = (0, T^*]$ , and  $B = (T^*, T]$  denote the two segments of the epidemic with a change point at  $T^* (< T)$ . We assume the parameter  $\theta$  takes value  $\theta_A$  in segment  $A$  and  $\theta_B$  in segment  $B$ . Then, the conditional densities in each of the segments are defined as

$$f_A(t) = \frac{-\dot{S}_t(\theta_A) \mathbb{1}_A(t)}{1 - S_{T^*}(\theta_A)}, \quad (6)$$

$$f_B(t) = \frac{-\dot{S}_t(\theta_B) \mathbb{1}_B(t)}{S_{T^*}(\theta_B) - S_T(\theta_B)},$$

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 given by

$$f_T(t) = \frac{\tau_{T^*}(\theta_A)}{\tau_T(\theta_B) - \tau_{T^*}(\theta_B) + \tau_{T^*}(\theta_A)} f_A(t) + \frac{\tau_T(\theta_B) - \tau_{T^*}(\theta_B)}{\tau_T(\theta_B) - \tau_{T^*}(\theta_B) + \tau_{T^*}(\theta_A)} f_B(t). \quad (7)$$

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

### A.6. Estimating Size of an Outbreak

We assume that the size of an outbreak  $k_\infty$  is a fixed integer representing the likely number of total infections in the contact network of the confirmed cases only. Hence  $k_\infty$  is not the prevalence of the disease in the population but rather just 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.7. 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  $\theta$  by drawing posterior samples using the *No U-Turn Sampling* (NUTS) method. As a point estimate of  $\theta$ , 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 Figure 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.

In simple terms, the empirical variance adjustment could be explained as follows: We first smooth the observed counts using a kernel smoother with  $\text{span} = 0.2$ . We used the function `supersmoothenr` in R programming language. 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.8. Additional numerical results

Here, we present additional diagnostic figures for our model fit. In Figure 4, 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). The posterior densities of the fitted parameters in the three segments are shown in Figure 4. Finally, as a diagnostic measure of the convergence of the Hamiltonian Monte Carlo chains, we show the trace plots of the fitted parameters in Figure 7.

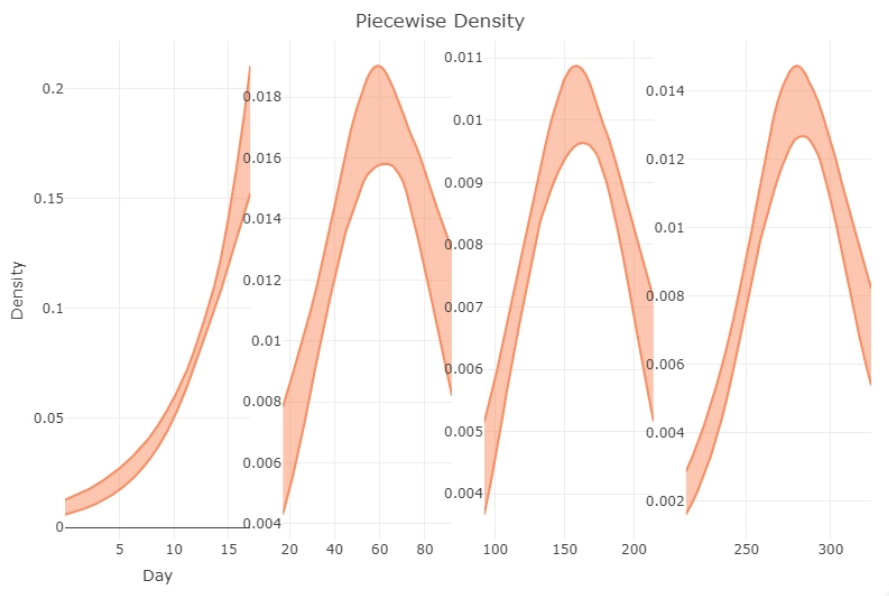


Figure 4: Estimated density of infection times in the four segments of the epidemic.

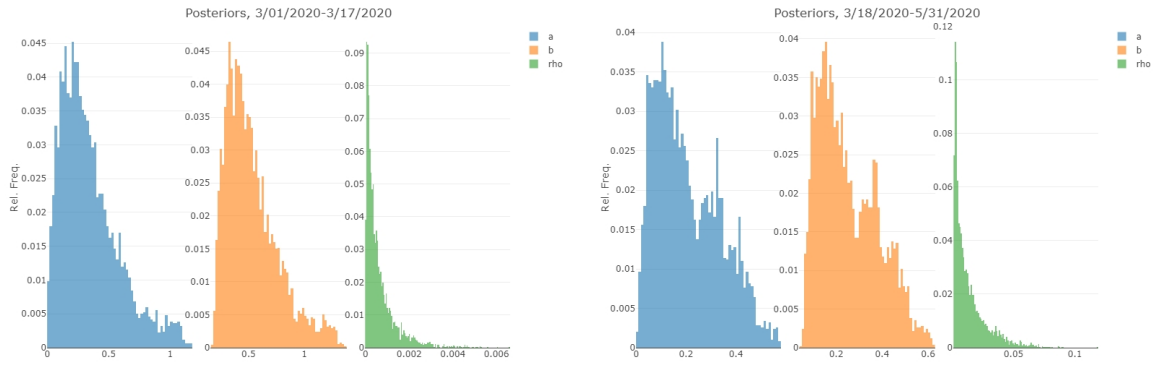


Figure 5: Posterior densities of the fitted parameters in the first two segments of the pandemic.

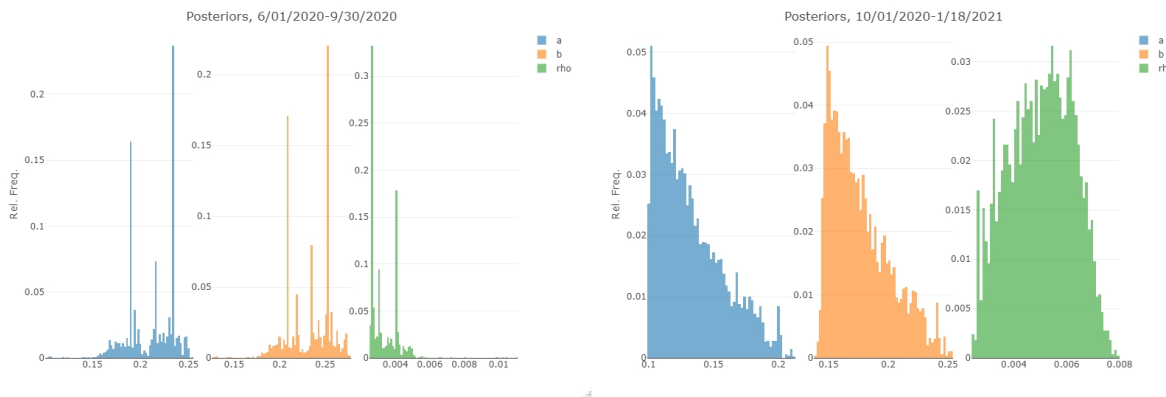


Figure 6: Posterior densities of the fitted parameters in the last two segments of the pandemic.

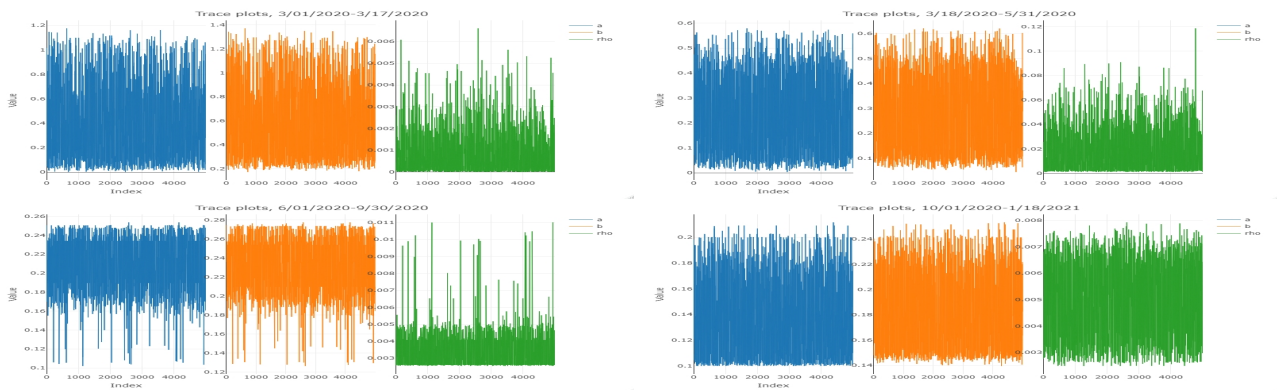


Figure 7: Trace plots of the fitted parameters in the four segments.

## B. Estimating Hospitalizations from Statewide Predictions of 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.
- (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 geographic 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  $i$ :

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

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. For 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  $i$ :

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

*Onset of cases that will eventually require hospitalization.* It has been documented that case outcome varies strongly with age [32]. We use the mean of the probabilities reported in [31] of severe case outcome by age group. Let  $p_a$  be the probability of severe case outcome for age group  $a$ . Then

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

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  $C_{i,a}(t)$  :

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

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

## CRedit authorship contribution statement

**Wasiur R. KhudaBukhsh:** Methodology, software, writing of the manuscript. **Caleb Deen Bastian:** Methodology, software, writing of the manuscript. **Matthew Wascher:** Methodology, software, writing of the manuscript. **Colin Klaus:** Methodology, software, writing of the manuscript. **Saumya Yashmohini Sahai:** Software, writing of the manuscript. **Mark Weir:** Conceptualization of this study, methodology, writing of the manuscript. **Eben Kenah:** Conceptualization of this study, methodology, analysis, and writing of the manuscript. **Elisabeth Root:** Conceptualization of this study, methodology, writing of the manuscript. **Joseph H. Tien:** Conceptualization of this study, methodology, writing of the manuscript. **Grzegorz Rempala:** Conceptualization of this study, methodology, writing of the manuscript.

## References

- [1] . . The Dartmouth Atlas of Healthcare. URL: <https://www.dartmouthatlas.org/faq/#research-methods-faq>. <https://www.dartmouthatlas.org/faq/#research-methods-faq>.
- [2] Ball, F., Neal, P., 2008. Network epidemic models with two levels of mixing. *Mathematical Biosciences* 212, 69–87.
- [3] Bartlett, M.S., 1960. Stochastic population models in ecology and epidemiology. Methuen.
- [4] Bastian, C.D., KhudaBukhsh, W.R., 2020. Python Code for Fitting DSA Analysis. [https://github.com/wasiur/dynamic\\_survival\\_analysis](https://github.com/wasiur/dynamic_survival_analysis).
- [5] Bastian, C.D., KhudaBukhsh, W.R., Pan, Y., Kenah, E., Rempala, G.A., . Predicting the size and duration of the outbreaks of covid-19 under minimal assumptions. Technical Report, The Ohio State University College of Public Health, April 2020.
- [6] Bastian, C.D., Rempala, G.A., 2020. Throwing stones and collecting bones: Looking for Poisson-like random measures. *Mathematical Methods in the Applied Sciences* 43, 4658–4668.
- [7] Bollobás, B., 1998. *Random Graphs*. Springer.
- [8] Bureau, U.C., a. American community survey. <https://www.census.gov/programs-surveys/acs>.
- [9] Bureau, U.C., b. Mapping files a. URL: <https://www.census.gov/geographies/mapping-files.html>. <https://www.census.gov/geographies/mapping-files.html>.
- [10] 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.
- [11] 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/>.
- [12] Durrett, R., 2007. *Random graph dynamics*. volume 200. Cambridge university press.
- [13] 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. <https://doi.org/10.25561/77482>.
- [14] Grasselli, G., Pesenti, A., Cecconi, M., 2020. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy. *Lancet* doi: 10.1001/jama.2020/4031.
- [15] Hethcote, H.W., 2000. The mathematics of infectious diseases. *SIAM Review* 42, 599–653.
- [16] 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. medRxiv URL: <https://www.medrxiv.org/content/early/2020/03/30/2020.03.27.20043752>, doi:10.1101/2020.03.27.20043752, arXiv:<https://tinyurl.com/4m9va4ee>.
- [17] 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. *Journal of Biological Dynamics* 12, 746–788.
- [18] Kermack, M., McKendrick, A., 1927. Contributions to the mathematical theory of epidemics. part i. *Proc. R. Soc. A* 115, 700–721.
- [19] KhudaBukhsh, W.R., Choi, B., Kenah, E., Rempala, G.A., 2020. Survival dynamical systems: individual-level survival analysis from population-level epidemic models. *Interface Focus* 10, 20190048.
- [20] KhudaBukhsh, W.R., Khalsa, S.K., Kenah, E.E., Grzegorz A.Rempala, J.H.T., 2021. Covid-19 dynamics in an ohio prison. medRxiv URL: <https://www.medrxiv.org/content/10.1101/2021.01.14.21249782v1>. preprint.
- [21] Kiss, I.Z., Miller, J.C., Simon, P.L., 2017. *Mathematics of Epidemics on Networks*. Springer.
- [22] Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (COVID-19). medRxiv <https://doi.org/10.1101/2020.02.14.20023127>.
- [23] Newman, M.E., 2003. The structure and function of complex networks. *SIAM review* 45, 167–256.
- [24] Newman, M.E.J., 2018. *Networks*. 2nd ed., Oxford University Press.
- [25] 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 .
- [26] Okabe, A.; Boots, B., Sugihara, K., 2000. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. 2nd ed., Wiley.
- [27] 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/>.
- [28] at Penn Medicine, P.H., 2020. COVID-19 Hospital Impact Model for Epidemics (CHIME). URL: <https://penn-chime.phl.io/>. <https://penn-chime.phl.io/>.



- [29] Rempala, G.A., . 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>.
- [30] Somekh, I., KhudaBukhsh, W.R., Root, E.D., Boker, L.K., Rempala, G., Simões, E.A., Somekh, E., 2021. Quantifying the Population-Level Effect of the COVID-19 Mass Vaccination Campaign in Israel: A Modeling Study. SSRN URL: <https://ssrn.com/abstract=3875446>. preprint.
- [31] Team, C.C.R., 2020a. Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019 — united states, february 12–march 28, 2020. *Morbidity and Mortality Weekly Report* 69, 382–386.
- [32] Team, C.C.R., 2020b. Severe outcomes among patients with coronavirus disease 2019 (covid-19) — united states, february 12–march 16, 2020. *Morbidity and Mortality Weekly Report* 69, 343–346.
- [33] 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., Donnelly, C.A., Ghani, A., Ferguson, N.M., 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infectious Diseases* [doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7).
- [34] Wascher, M., Schnell, P.M., KhudaBukhsh, W.R., Quam, M., Tien, J.H., Rempala, G.A., 2021. Monitoring SARS-COV-2 transmission and prevalence in populations under repeated testing. medRxiv URL: <https://www.medrxiv.org/content/10.1101/2021.06.22.21259342v1>. preprint.
- [35] Weitz, J.S., 2020. COVID-19 near-term epidemic risk assessment for Georgia. <https://github.com/jsweitz/covid-19-ga-summer-2020>.
- [36] 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 72 314 cases from the chinese center for disease control and prevention. *Journal of the American Medical Association* .