Parameter identifiability and optimal control of an SARS-CoV-2 model early in the pandemic

Necibe Tuncer a, Archana Timsina a, Miriam Nuno b, Gerardo Chowell c and Maia Martcheva d

aDepartment of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL, USA; bDepartment of Biostatistics, University of California, Davis, CA, USA; cDepartment of Population Health Sciences, Georgia State University, Atlanta, GA, USA; dDepartment of Mathematics, University of Florida, Gainesville, FL, USA

ABSTRACT
We fit an SARS-CoV-2 model to US data of COVID-19 cases and deaths. We conclude that the model is not structurally identifiable. We make the model identifiable by prefixing some of the parameters from external information. Practical identifiability of the model through Monte Carlo simulations reveals that two of the parameters may not be practically identifiable. With thus identified parameters, we set up an optimal control problem with social distancing and isolation as control variables. We investigate two scenarios: the controls are applied for the entire duration and the controls are applied only for the period of time. Our results show that if the controls are applied early in the epidemic, the reduction in the infected classes is at least an order of magnitude higher compared to when controls are applied with 2-week delay. Further, removing the controls before the pandemic ends leads to rebound of the infected classes.

1. Introduction
Mathematical models have been widely implemented to study the spread and interventions aimed to mitigate the impact of the COVID-19 pandemic. Model complexity has varied significantly, from the simple Susceptible, Exposed, Infected, Recovered (SEIR) model involving a few parameters to more elaborate models with many more parameters that incorporate testing efforts, contact tracing and social distancing protocols [8,12,26,29,31,38,40,41]. For instance, Tang et al. [29] showed that intensive contact tracing followed by quarantine and isolation was effective for reducing the reproduction number and risk of transmission in a model that incorporated clinical progression of disease. In an extended SEIR model, Zou et al. considered testing and under-reporting of COVID-19 cases, which was later trained by machine learning algorithms [41]. Lu et al. illustrated some of the limitations of model fitting based on confirmed cases alone in the absence
of information about testing efforts over time [19]. Researchers have used the cumulative number of cases, quarantined population [29], number of confirmed and fatality cases [41], at early stages of the pandemic. However, the ability to accurately forecast epidemics is highly dependent on (i) amount and type of data, whether there exists sufficient data to identify the model parameters, and (ii) model structure, whether the model is structured to reveal its parameters from the ideal noise free data [33,34].

Insights from mathematical models that incorporate features of disease transmission, incubation period, and other features of a virus can be used to improve our understanding of the spread of diseases and investigate optimal intervention strategies. We are often challenged with the decision of implementing models that are simple enough to provide meaningful inference given the limited data but also models that realistic enough to capture key underlying virus dynamics and epidemiological states. The implementation of more complex features often involves models with large number of unknown parameters and epidemiological states that while important are often not observed or captured. The variability and lack of accuracy in the projections based on some mathematical models have raised concern among public health officials, specially during the early phase of the COVID-10 pandemic. Therefore, it is important that modelling studies of COVID-19 investigate the question of parameter identifiability [21]. A model is said to be (structurally) identifiable, if its parameters can be determined uniquely from the observations such as daily cases and fatalities. If a model is unidentifiable, then infinitely many parameters would give the same observation. In that situation, multiple parameter combinations satisfying the parameter correlations would fit the same data sets, which will result in misleading and inaccurate predictions. Formally studying the parameter identifiability for COVID-19 models will be valuable for several reasons. First, it will provide some guidance about the limitations of the models themselves being proposed in providing valuable predictions and inferences based on data that is currently available. Findings of studies on parameter identifiability are also essential for increasing awareness about the importance of timely access to data that are instrumental for improving estimates and projections.

In this study, we propose a model to study the spread of SARS-CoV-2 and assess the effect of social intervention measures to curtail the spread of the COVID-19 pandemic during the early phase of the outbreak. The findings of this study are most relevant to understanding the spread of the pandemic as it unfolds initially. To our knowledge, this is the first study that has considered the question of both structural and practical identifiability of a COVID-19 model, but the structural identifiability of several COVID-19 models has been recently considered in [21]. Structural and practical identifiability of epidemic models for Ebola, Cholera, Dengue, Rift Valley Fever and Zika have been studied before [5,9,10,32–35]. Even though, it is not possible to identify individual parameters of the epidemic model from the observations, two independent studies confirmed that it is possible to accurately estimate the basic reproduction number from the epidemic data [9,33]. In this study, we investigate the structural and practical identifiability of a COVID-19 model that considers six epidemic states. Susceptible (S) individuals can be transitioned into a latency period (E), become infectious with no symptoms and not being tested (I_n), or can move into symptomatic infectious and tested state (I_s), symptomatic isolated and treated (I_h), and recovered individuals (R).
In Section 2, we consider a COVID-19 model with no infection control measures (e.g. social distancing), derive the basic reproductive number, and investigate parameter identifiability using daily COVID-19 cases and deaths data. In Section 3, we discuss structural and practical identifiability of the model. We determine structural identifiability using the Differential Algebra approach given its ability to eliminate the unobserved state variables. Using data on daily incidence of cases and deaths, we establish that the proposed model is not structurally identifiable but we improve its structural identifiability by fixing some of the parameters from external information. We further investigate parameters practical identifiability by performing Monte Carlo simulations on a constrained minimization problem numerically from noisy observations. This approach is local and is similar in nature to approaches with derived 95% confidence intervals, bootstrapping or sensitivity analysis. Although Monte Carlo Simulations (MCS) are not a novel approach, the novelty in our approach is that we are applying MCS only after we have determined that the parameter estimation problem is well-posed. This framework is expected to lead to much more reliable solutions. A subset of two parameters were deemed practically unidentifiable. Given that shelter-in-place order were initiated in March 19, 2020, Section 4 incorporates an optimal control with control variables of reduction in transmission rates for exposed and asymptomatic infected individuals, as well as self-isolation for symptomatic infected individuals. The forward–backward sweep method is used to approximate the solution of the model with controls. Section 5 provides a discussion of the findings described in this study.

Structural non-identifiability arises from excessive parametrization of the proposed model or by choice of a completely wrong model structure [22,32]. This problem can often be resolved by reformulating the model to reduce the number of states and parameters or fix parameter values that are less relevant to model predictions of interest. Lack of practical identifiability can be remedied by employing an alternative model of lower complexity, collecting more data about model states to better characterize the system dynamic features, increasing spatial–temporal resolution of the data to better constrain the model parameters, and/or reducing the number of parameters that are jointly estimated.

2. A basic compartmental model for COVID-19 infection

We introduce a mathematical model of COVID-19 by splitting the total population, \( N(t) \), into non-intersecting classes based on the disease status. In US individuals with symptoms of COVID-19 were asked to self isolate at home. We denote this class by \( I_s(t) \). If clinical symptoms were eventually developed, individuals were admitted to a hospital. We denote the hospitalized individuals who are isolated and treated by \( I_h(t) \). Individuals who were not tested but may have had COVID-19 without symptoms are denoted by \( I_n(t) \). So, the total population \( N(t) \) is divided into sub-populations that consists of susceptible individuals \( S(t) \), exposed individuals \( E(t) \), symptomatic and infectious individuals that are tested \( I_s(t) \), asymptomatic and infectious individuals that are not tested and remain in the community \( I_n(t) \), severe symptomatic individuals who are hospitalized and isolated \( I_h(t) \), and recovered individuals \( R(t) \). The total population size is \( N(t) = S(t) + E(t) + I_s(t) + I_n(t) + I_h(t) + R(t) \). We use standard incidence but since the isolated individuals do not mix in the population, we divide by \( N - I_h \). This implicitly assumes that the infection rate of hospital workers from the hospitalized COVID-19 individuals is
relatively low and can be neglected. The model takes the form:

\[
\begin{align*}
S' &= -\frac{\beta_E S I_n + \beta_s I_s S}{N - I_h}, \\
E' &= \frac{\beta_E S I_n + \beta_s I_s S}{N - I_h} - kE, \\
I_n' &= (1 - \rho)kE - \gamma_n I_n, \\
I_s' &= \rho kE - (\alpha + \gamma_s) I_s, \\
I_h' &= \alpha I_s - (\gamma_h + \nu) I_h, \\
R' &= \gamma_n I_n + \gamma_s I_s + \gamma_h I_h.
\end{align*}
\]

Susceptible individuals acquire COVID-19 infection upon effective contact with an exposed individual or an infectious individual who may or may not show COVID-19 symptoms. The COVID-19 infection rate is assumed to be different for each infection class to account for the heterogeneity of the contact rates of infectious individuals with or without the clinical COVID-19 symptoms. Namely, the parameters \(\beta_E, \beta_n\) and \(\beta_s\) represent the transmission rates for individuals in \(E, I_n\) and \(I_s\) compartments respectively. It is estimated that infectiousness starts 5, 8 and 11 days before the incubation period ends [14]. Recent papers have modelled the latent period and the pre-symptomatic period separately [13]. In this study, to keep the model simple we lumped these two stages in a class called \(E\), which denotes the incubation period. The average transmission rate of two stages is not zero. Hence \(\beta_E \neq 0\). Therefore, newly infected individuals that move to the incubation class \(E\) become infectious while they are still in class \(E\). After the incubation period, which lasts for \(1/k\) days, a fraction \((0 < \rho < 1)\) of exposed individuals start showing clinical symptoms of COVID-19 and get tested. The remaining \((1 - \rho)\) fraction of the exposed individuals do not show any clinical symptoms of COVID-19. Infectious individuals might develop severe infection and get hospitalized at rate \(\alpha\). Infectious individuals in \(I_n, I_s\) and \(I_h\) classes recover at rates \(\gamma_n, \gamma_s\) and \(\gamma_h\) respectively. A flow diagram of the model (1) is shown in Figure 1. The state variables and parameters of the model are described in Tables 1 and 2, respectively.

Figure 1. Flow chart of model (1) showing the transition of individuals between the classes defined by disease status.
Table 1. State variables of the COVID-19 model (1) and their definitions.

| Variables | Definition |
|-----------|------------|
| $S(t)$    | Number of susceptible individuals at time $t$ |
| $E(t)$    | Number of individuals who are exposed to COVID-19 and are in their incubation period at time $t$ |
| $I_n(t)$  | Number of infectious individuals who are not tested and not showing any clinical COVID-19 symptoms at time $t$ |
| $I_s(t)$  | Number of infectious individuals who are tested and showing clinical COVID-19 symptoms at time $t$ |
| $I_h(t)$  | Number of hospitalized individuals due to COVID-19 at time $t$ |
| $R(t)$    | Number of recovered individuals at time $t$ |

Table 2. Definition of the COVID-19 model parameters.

| Parameter | Definition |
|-----------|------------|
| $\beta_E, \beta_n, \beta_s$ | Transmission rate for exposed $E$, asymptomatic $I_n$, and symptomatic $I_s$ individuals |
| $q_E, q_n$ | Relative infectiousness of exposed and non-symptomatic individuals |
| $1/k$      | Incubation period |
| $\rho$     | Proportion of individuals that undergo testing and test positive (Proportion of infected individuals that show COVID-19 symptoms) |
| $\gamma_n, \gamma_s, \gamma_h$ | Recovery rate for asymptomatic $I_n$, symptomatic $I_s$, and hospitalized $I_h$ individuals |
| $\alpha$  | Hospitalization rate of severely infected symptomatic $I_s$ individuals |
| $\nu$     | Disease induced mortality |

We compute the reproduction number of (1) to obtain

$$R_0 = \left( \frac{\beta_E}{k} + (1 - \rho) \frac{\beta_n}{\gamma_n} + \rho \frac{\beta_s}{\gamma_s + \alpha} \right) \frac{S(0)}{N(0)},$$

where $N(0) = S(0) + E(0) + I_n(0) + I_s(0) + I_h(0) + R(0)$ is the initial total population size. The reproduction number consists of three terms, each of which corresponds to secondary cases produced by one type of infectious individuals. The first term is the number of secondary infections produced by exposed individuals in an entirely susceptible population during their lifespan as exposed, the second term is the number of secondary infections produced by asymptomatic and uncounted individuals in an entirely susceptible population during their lifespan as asymptomatic; finally, the third term corresponds to the number of secondary infections produced by symptomatic individuals during their lifespan as symptomatic and counted.

3. **Identifiability analysis: can we uniquely estimate the model parameters?**

In the literature, the term *identifiability analysis* is used to describe whether it is possible to uniquely recover model parameters from a given set of data. It is a crucial step in parameter estimation problem and is usually performed in two steps. The first step is the *structural identifiability* which is a theoretical property of the epidemic model. Structural
identifiability explores whether the epidemic model is structured to reveal its parameters from the observations. Here we use the term observation to refer to continuous smooth output where the data is measured at the discrete time points. If a compartmental epidemic model is structurally unidentifiable, it means that different sets of parameters could produce the same output. Hence a structurally unidentifiable epidemic model should not be used to analyse epidemiological data or design intervention strategies. If an epidemic model is structurally identifiable, then the second step is the practical identifiability. Practical identifiability considers the actual epidemiological data which is contaminated with noise.

3.1. Structural identifiability analysis

To study the structural identifiability of the epidemic model, we represent the model (1) in the following compact form:

$$x'(t) = f(x, p), \quad x(0) = x_0 \quad (2)$$

where $x$ denotes the state variables and $p$ denotes the parameters of the model of (1). The observations, such as daily incidences and deaths, are functions of the state variables. Let $y_1(t)$ denote the daily incidences and $y_2(t)$ denote daily reported deaths, then in terms of model variables the observations take the following forms;

$$y_1(t) = \rho k E \quad \text{and} \quad y_2(t) = \nu I_h \quad (3)$$

In compact form, we denote the observations as $y(t) = (y_1(t), y_2(t))$ such that

$$y(t) = g(x, p) = (\rho k E, \nu I_h). \quad (4)$$

Structural identifiability determines whether the model (2) is structured to reveal its parameters, $p$, from observations (4). Structural identifiability is a theoretical result which considers the ‘perfect’ scenario assuming that observation is given continuously for all time and is noise free. A model is said to be structurally (uniquely) identifiable if all the parameters in $p$ can be recovered from the observations $y$. If even one parameter cannot be identified, the mode is said to be structurally unidentifiable. Several methods for determining structural identifiability have been introduced and studied widely [10,22,33]. Here, we will use the Differential Algebra approach to study the structural identifiability [22]. The goal of Differential Algebra approach is to eliminate the unobserved state variables using the system of differential equations in (2) and to obtain input–output equations. We obtain the input–output equations of the model (1) using the observations (3), which consist of a couple of monic differential polynomials in the observed state variables and their derivatives. The coefficients of these polynomials are rational functions of the parameters of the model. We obtain the following input–output equations using DAISY [1].

$$\frac{dy_1}{dt} - \frac{dy_2}{dt} \frac{\beta_s k \rho (\alpha - \gamma_n + \gamma_s)}{\alpha \beta_E v} - y_1 \frac{-\beta_E \gamma_n + \beta_n k \rho - \beta_s k \rho}{\beta_E} - y_2 \frac{f_1(p)}{\alpha \beta_E v} = 0 \quad (5)$$

$$\frac{d^2 y_2}{dt^2} - \frac{dy_1}{dt} \frac{\alpha \beta_E v (\alpha + v + \gamma_s + \gamma_n)}{\beta_s k \rho (-\alpha + \gamma_n - \gamma_s)} + y_1 \frac{f_2(p)}{\beta_s k \rho (-\alpha + \gamma_n - \gamma_s)} + y_2 \frac{f_3(p)}{-\alpha + \gamma_n - \gamma_s} = 0 \quad (6)$$
where
\[
\begin{align*}
f_1(p) &= \beta_s k \rho (\alpha v + \alpha \gamma_n - v \gamma_n + v \gamma_s - \gamma_n \gamma_h + \gamma_s \gamma_h) \\
f_2(p) &= \alpha v (-\alpha \beta_E \gamma_n + \alpha \beta_s k \rho - \alpha \beta_n k - \beta_E v \gamma_n - \beta_E \gamma_n \gamma_h + \beta_n v k \rho \\
&\quad - \beta_n v k + \beta_n \gamma_s k \rho - \beta_n \gamma_h k - \beta_n k \gamma_n - \beta_n k \rho - \beta_s v k \rho - \beta_s \gamma_n k \rho - \beta_s k \gamma_h \rho) \\
f_3(p) &= \alpha v^2 + 2 \alpha v \gamma_n + \alpha \gamma_n^2 - v \gamma_n + v \gamma_s - 2 \beta_E v \gamma_n + 2 v \gamma_s \gamma_h - \gamma_n \gamma_h^2 + \gamma_s \gamma_h^2.
\end{align*}
\]

Input–output equations (5) and (6) are differential algebraic polynomials of outputs \(y_1\) and \(y_2\) whose coefficients consist of model parameters. We say that the model (2) is structurally identifiable if the mapping from parameter space to the coefficients of input–output equations (5) and (6) are one-to-one \([32,34]\). Within differential algebra approach, we define the structural identifiability as follows \([10,34]\):

**Definition 3.1:** Let \(c(p)\) denote the coefficients of the input–output equations (5) and (6). That is
\[
c(p) = \left( -\frac{\beta_s k \rho (\alpha - \gamma_n + \gamma_s)}{\alpha \beta_E}, -\frac{-\beta_E \gamma_n + \beta_n k \rho - \beta_n k - \beta_s k \rho}{\beta_E}, \frac{f_1(p)}{\alpha \beta_E v}, \frac{f_2(p)}{\beta_s k \rho (\alpha + \o + \gamma_n - \gamma_s)}, \frac{f_3(p)}{-\alpha + \gamma_n - \gamma_s} \right)
\]
with \(p = (\beta_E, \beta_n, \beta_s, k, \rho, \gamma_n, \gamma_s, \alpha, r, v)\) denoting the parameters of the model (1). Then we say that the model (1) is structured to reveal its parameters from the observations (3) if and only if
\[
c(p) = c(\hat{p}) \implies p = \hat{p}
\]

Following the definition of the structural identifiability, we set \(\hat{p} = (\hat{\beta}_E, \hat{\beta}_n, \hat{\beta}_s, \hat{k}, \hat{\rho}, \hat{\gamma}_n, \hat{\gamma}_s, \hat{\alpha}, \hat{\gamma}_h, \hat{\nu})\). Then solving \(c(p) = c(\hat{p})\), we obtain the following parameter correlations:
\[
\begin{align*}
\hat{\beta}_E &= \frac{\beta_s k \rho (\hat{\alpha} - \gamma_n + \hat{\gamma}_s)}{\hat{\beta}_s k \rho (\hat{\alpha} - \hat{\gamma}_n + \hat{\gamma}_s)} \\
\hat{\beta}_n(\rho - 1) &= \frac{\beta_s k \rho}{\hat{\beta}_s k \rho (\hat{\alpha} - \hat{\gamma}_n + \hat{\gamma}_s)} (\hat{\beta}_E \gamma_n - \hat{\beta}_E \gamma_n^2 - \hat{\alpha} \hat{\beta}_E \hat{\gamma}_n + \hat{\beta}_E \gamma_n \hat{\gamma}_n + \hat{\beta}_E \gamma_n \hat{\gamma}_s - \hat{\beta}_E \hat{\gamma}_n \hat{\gamma}_s \\
&\quad - \hat{\alpha} \hat{\beta}_n k + \hat{\beta}_n \gamma_n k + \hat{\alpha} \hat{\beta}_n k \rho - \hat{\beta}_n \gamma_n k \rho + \hat{\beta}_s \gamma_n \hat{k} \rho - \hat{\beta}_s \gamma_n \hat{k} \rho + \hat{\beta}_s \gamma_n \hat{k} \rho)
\end{align*}
\]
\[
\gamma_s + \alpha = \hat{\gamma}_s + \hat{\alpha}, \quad v \alpha = \hat{\nu} \hat{\alpha}, \quad \gamma_n + v = \hat{\gamma}_n + \hat{\nu}
\]

(8)

As it is evident in (8), parameters of the model (1) cannot be obtained from the observations of daily incidences and deaths. Some parameter combinations can be identified such as \(\gamma_s + \alpha\) or \(v \alpha\), but individual parameters \(\gamma_s, \alpha\) or \(v\) cannot be identified. Even if \(\alpha\) is fixed to \(\hat{\alpha}\) in (8), then we can only obtain \(\gamma_s = \hat{\gamma}_s, v = \hat{\nu}\) and \(\gamma_n = \hat{\gamma}_n\), and the parameters \(\beta_E, \beta_n, \beta_s, \gamma_n, k, \rho\) still cannot be identified. Thus we have proved the following Proposition 3.1.
Proposition 3.1: The parameters of the model (1) are not structurally identifiable from the observations of daily incidences and deaths.

In this approach, we assume that the initial conditions are given and fixed. An important advantage of using Differential Algebra approach in structural identifiability analysis is that we obtain parameter correlations as in (8). These parameter correlations are then used to obtain a structurally identifiable model [10,32]. However, the correlations in (8) are highly complex. In an attempt to get a structurally identifiable model, we rewrite the parameter correlations into the following simpler forms;

\[
\frac{\beta_E}{\beta_s k \rho (\hat{\alpha} - \gamma_n + \hat{\gamma}_s)} = \frac{\hat{\beta}_n (\hat{\rho} - 1)}{\hat{\beta}_s \hat{\rho}} \left( \hat{\alpha} - \hat{\gamma}_n + \hat{\gamma}_s \right) + \frac{\hat{\beta}_E}{\hat{\beta}_s \hat{\rho}} \left( \hat{\alpha} - \gamma_n + \hat{\gamma}_s \right) (\gamma_n - \hat{\gamma}_n)
\]

\[
+ \frac{1}{\hat{\alpha} - \hat{\gamma}_n + \hat{\gamma}_s} (\gamma_n - \hat{\gamma}_n)
\]

\[
\gamma_s + \alpha = \hat{\gamma}_s + \hat{\alpha}, \quad \nu \alpha = \hat{\nu} \hat{\alpha}, \quad \gamma_n + \nu = \hat{\gamma}_n + \hat{\nu}
\]

Next, we factor out \( \beta_s \) from the incidence term in model (1) as follows:

\[
\beta_s \frac{q_E S + q_n I_n + I_s}{N - I_h}
\]

where \( q_E = \frac{\beta_E}{\beta_s} \) and \( q_n = \frac{\beta_n}{\beta_s} \). With this setting, if we fix parameter \( \alpha \) at \( \hat{\alpha} \), and \( \gamma_n \) at \( \hat{\gamma}_n \) then the parameter correlations in (9) become

\[
\frac{q_E}{k \rho} = \frac{\hat{q}_E}{\hat{k} \hat{\rho}}, \quad \frac{q_n (\rho - 1)}{\rho} = \frac{\hat{q}_n (\hat{\rho} - 1)}{\hat{\rho}}, \quad \alpha = \hat{\alpha}, \quad \gamma_n = \hat{\gamma}_n, \quad \gamma_s = \hat{\gamma}_s,
\]

\[
\gamma_h = \hat{\gamma}_h, \quad \nu = \hat{\nu}.
\]

We can see from (10) that the parameters \( \alpha, \gamma_n, \gamma_s, \gamma_h \) and \( \nu \) can be identified; parameter combinations \( \frac{q_E}{k \rho} \) and \( \frac{q_n (\rho - 1)}{\rho} \) can be identified; but parameter \( \beta_s \) cannot be identified. For instance, the parameter correlation \( \frac{q_E}{k \rho} \) implies that infinitely many values of the parameters \( q_E, k \) and \( \rho \) would generate the same output as long as the parameter relationship \( \frac{q_E}{k \rho} \) remains constant at \( \frac{\hat{q}_E}{\hat{k} \hat{\rho}} \). Same is true for parameters \( \rho \) and \( q_n \) satisfying the relationship \( \frac{q_n (\rho - 1)}{\rho} \). In the view of these results, we obtain the following Proposition 3.2.

Proposition 3.2: If the parameters, \( \alpha, \gamma_n, k \) and \( \rho \) are fixed, then solving \( c(p) = c(\hat{p}) \), we obtain

\[
q_E = \hat{q}_E, \quad q_n = \hat{q}_n, \quad k = \hat{k}, \quad \rho = \hat{\rho}, \quad \alpha = \hat{\alpha}, \quad \gamma_n = \hat{\gamma}_n, \quad \gamma_s = \hat{\gamma}_s,
\]

\[
\gamma_h = \hat{\gamma}_h, \quad \nu = \hat{\nu}.
\]

Since, \( \beta_s \) still cannot be obtained we say that the model (1) is structurally unidentifiable from the observations of daily incidences and deaths.
Remark 3.1: The model (1) is not structurally identifiable, because the parameter $\beta_s$ does not appear in the parameter relations in (10). But all the other parameters are identifiable. We use the following strategy in an attempt to make $\beta_s$ an identifiable parameter; we use the basic reproduction number $R_0$ to determine the infection rate $\beta_s$ as

$$\beta_s = R_0 / \left( \frac{q_E}{k} + (1 - \rho) \frac{q_n}{\gamma_n} + \rho \frac{1}{\gamma_s + \alpha} \right) S(0) / N(0).$$

If we use estimate from the literature of $R_0$, and fix $R_0$ at that value, then $\beta_s$ will be identifiable as well.

### 3.1.1. Structural identifiability of COVID-19 epidemic models

We take this opportunity and study the structural identifiability of some COVID-19 models. Several models are proposed within the last year for COVID-19 infection and most of these models, including our model (1), are extensions of the following Kermack–McKendrick type SEIRepidemic model without human demography.

$$\begin{align*}
S' &= -\beta E S + \beta IS, \\
E' &= \beta E S + \beta IS - kE, \\
I' &= kE - (\nu + \gamma)I, \\
R' &= \gamma I.
\end{align*}$$

Model 1 (M1):

As before we assume that the observations are daily incidences and COVID-19 deaths. In terms of the state variables of the model (M1), observations becomes

$$y_1(t) = kE, \quad y_2(t) = \nu I.$$

We have determined the input–output equations of the model M1 and show that it is structurally (globally) identifiable and obtain the following result (see Appendix for details).

**Proposition 3.3:** The model (M1) is structurally identifiable from the observations of daily incidences $y_1(t) = kE$, and deaths $y_2(t) = \nu I$.

In Proposition 3.3, it is assumed that all new incidences, $kE$ are observed. However, this is not the case in reality. A key feature of COVID-19 cases is that majority of new infections are generated by individuals who do not show any clinical symptoms of COVID-19. A big portion of these asymptomatic individuals will remain undetected and only some might be detected through random testing or contact tracing. This leads to the fact that only the fraction of new incidences are reported. We continue with the identifiability analysis assuming that only fractions of daily incidences are observed, that is if we set $y_1(t) = \rho kE$ for $0 < \rho < 1$ then the parameter correlations for model (M1) become

$$\frac{\hat{\beta}_E}{\hat{\beta}} = \frac{\hat{\beta}_E}{\hat{\beta}}, \quad \nu \hat{\beta} = \hat{\nu} \hat{\beta}, \quad \rho \hat{\beta} = \hat{\rho} \hat{\beta}, \quad \gamma \hat{\beta} + \hat{\nu} \hat{\beta} = \hat{\gamma} \hat{\beta} + \hat{\nu} \hat{\beta}, \quad k = \hat{k}$$

Assuming a much realistic scenario that only a fraction of actual incidences can be observed, makes the simple SEIR model unidentifiable. Thus we have obtained the following result for model (M1).
**Proposition 3.4:** The model \((M_1)\) with a fraction of observed cases is not structurally identifiable from the observations of daily incidences, \(y_1(t) = \rho kE\), and deaths \(y_2(t) = \nu I_s\).

**Remark 3.2:** From parameter correlations (11), we see that if we set \(\rho = \hat{\rho}\), that is if we assume that the fraction of the incidences being reported daily is known, then the model \((M_1)\) becomes structurally identifiable from daily incidences and deaths.

The next model, we will study is the one where infected class \(I\) is further divided into two subclasses, as asymptomatic individuals \(I_n\) and symptomatic individuals \(I_s\). Then we obtain the following model:

\[
\begin{align*}
S' &= -\beta_EES + \beta_nI_nS + \beta_sI_sS, \\
E' &= \frac{\beta_EES + \beta_nI_nS + \beta_sI_sS}{N} - kE, \\
I'_n &= (1 - \rho)kE - \gamma_nI_n, \\
I'_s &= \rho kE - (\nu + \gamma_s)I_s, \\
R' &= \gamma_nI_n + \gamma_sI_s.
\end{align*}
\]

**Model 2 \((M_2)\):**

In this model, we assume that all the infected individuals who are showing symptoms are getting tested. That is for this model \((M_2)\), observations of daily incidences and deaths take the form \(y_1(t) = \rho kE\), \(y_2(t) = \nu I_s\). Determining the input–output equations and then solving \(c(p) = c(\hat{p})\), gives the following parameter correlations;

\[
\frac{\beta_E}{\beta_s \rho} = \frac{\hat{\beta}_E}{\hat{\beta}_s \hat{\rho}}, \quad \frac{\beta_n(1 - \rho)}{\beta_s \rho} = \frac{\hat{\beta}_n(1 - \hat{\rho})}{\hat{\beta}_s \hat{\rho}}, \quad \frac{\gamma_n}{\nu} = \frac{\hat{\gamma}_n}{\hat{\nu}}, \quad \frac{\gamma_s}{\nu} = \frac{\hat{\gamma}_s}{\hat{\nu}}, \quad \frac{k}{\nu} = \frac{\hat{k}}{\hat{\nu}}.
\]

**Proposition 3.5:** The model \((M_2)\) is not structurally identifiable from the observations of daily incidences, \(y_1(t) = \rho kE\), and deaths \(y_2(t) = \nu I_s\).

If we assume both the incubation period and the fraction of infected individuals who are showing symptoms are known, and set \(k = \hat{k}\), \(\rho = \hat{\rho}\). Furthermore, if we set as before, \(q_E = \frac{\hat{\beta}_E}{\hat{\beta}_s}\) and \(q_n = \frac{\hat{\beta}_n}{\hat{\beta}_s}\), then the following parameters of the model \((M_2)\) can be identified,

\[q_E = \hat{q}_E, \quad q_n = \hat{q}_n, \quad \gamma_n = \hat{\gamma}_n, \quad \gamma_s = \hat{\gamma}_s, \quad \nu = \hat{\nu}, \quad \rho = \hat{\rho}, \quad k = \hat{k}\]

Since all the parameter but \(\beta_s\) can be identified. We use the same strategy as before and determine \(\beta_s\) from the basic reproduction number of the model \((M_2)\). We summarize the results in the following proposition.

**Proposition 3.6:** If the parameters \(k\) and \(\rho\) are fixed, then \(q_E, q_n, \gamma_n, \gamma_s\) and \(\nu\) can be identified. But the model \((M_2)\) remains structurally unidentifiable from the observations of daily incidences and deaths, since \(\beta_s\) cannot be identified.

However, the infection rate \(\beta_s\) can be determined from the identifiable model parameters through basic reproduction number \(R_0\),

\[
R_0 = \beta_s \left( \frac{q_E}{k} + (1 - \rho) \frac{q_n}{\gamma_n} + \rho \frac{1}{\nu + \gamma_s} \right) \frac{S(0)}{N(0)}.
\]
Our intention for considering models \((M_1)\) and \((M_2)\) is to find the most complex COVID-19 model which is intrinsically identifiable, that is identifiable without any restrictions on the parameters. It is worth mentioning that the models \((M_1), (M_2)\) and \((1)\) are nested models. Model \((M_1)\) is nested in model \((M_2)\), and model \((M_2)\) is nested in \((1)\). The simplest model \((M_1)\) is the only structurally identifiable model from the observations of daily incidences and casualties. As the model gets more complex as we increase the number of infectious classes in an attempt to capture the transmission dynamics of the COVID-19 infection, the epidemic model becomes unidentifiable.

### 3.2. Parameter estimation

We obtain the parameter estimates of the model \((1)\) by fitting the model to the observed daily COVID-19 incidences and deaths in the USA. The time series data from March 3 to April 26, 2020, is collected [36]. We are interested in estimating the epidemiologically important parameters. In particular, the parameters to be estimated from data are infection rate of \(S\) by exposed individuals \(\beta_{SQ}\), infection rate by individuals who are not showing any clinical COVID-19 symptoms \(\beta_{SN}\), infection rate by individuals with clinical COVID-19 symptoms \(\beta_s\), recovery rate of individuals with clinical symptoms \(\gamma_s\), recovery rate of hospitalized individuals \(\gamma_h\) and death rate of hospitalized individuals \(\nu\). Based on the structural identifiability analysis of the epidemic model \((1)\), we fix some model parameters such as the incubation period to estimate the model parameters accurately. The literature search of peer-reviewed publications reporting the incubation period of SARS-COV-2 listed the incubation period as ranging from 2 to 18 days [11]. We set the incubation period \(1/k\) as 14 days. Mild or asymptomatic cases usually recover within 14 days [2]. Since we have already set the incubation period to 14 days, we set the recovery period of the asymptomatic individuals \(1/\gamma_n\) to 1 day. The mean estimated basic reproduction number of SARS-COV-2 early in the pandemic is found to vary from 1.95 to 6.47 [16]. We set \(R_0 = 4.3\) for our parameter estimation problem. For the proportion of asymptomatic individuals, several studies give a wide range, namely 30–70\% of individuals who have tested positive have mild or no symptoms ([39] and the references therein). Due to this large uncertainty, we take the numbers in the Diamond Princess cruise ship; 712 individuals in the cruise ship Diamond Princess tested positive for COVID-19, 410 of these infected individuals in the cruise ship were asymptomatic. Therefore, we set the fraction of asymptomatic individuals to \(410/712 = 58\%\) [27]. That is we take \(\rho = 0.42\). Hospitalization rate of infected individuals with clinical symptoms of COVID-19 per day \(\alpha\) is not known. However, CDC reports that at an average 15\% of infected individuals are admitted to hospitals [4]. Thus we set

\[
\frac{\alpha}{\alpha + \gamma_s} = 0.15 \quad \text{which implies } \alpha = 0.1765\gamma_s
\]

We use the daily US COVID-19 cases and deaths from March 3 to April 26, 2020, as the observations of the parameter estimation problem. Let \(\{t_1, t_2, \ldots, t_n\}\) denote the discrete time when the number of US COVID-19 cases are observed, and let \(\{t_1, t_2, \ldots, t_m\}\) denote the discrete time when the number of US COVID-19 deaths are observed. Both observations are contaminated with noise and thus

\[
Y_1^i = y_1(t_i) + \epsilon_i \quad i = 1, 2, \ldots, n \quad \text{and} \quad Y_2^j = y_2(t_j) + \epsilon_j \quad j = 1, 2, \ldots, m \quad (15)
\]
Table 3. Parameter estimation results: parameter estimates of the model (1) obtained by solving the optimization problem (16).

| Parameter | \( \hat{\beta}_s \) | \( q_E \) | \( q_n \) | \( \alpha \) | \( \gamma_s \) | \( \gamma_h \) | \( \nu \) |
|-----------|----------------|--------|--------|--------|--------|--------|--------|
| Estimated value | \( 6.6 \times 10^{-5} \) | 589.8 | 97394.6 | 0.006 | 0.03 | 0.14 | 0.86 |

We estimated the parameters of the model (1) \( \mathbf{p} = [\beta_s, q_E, q_n, \gamma_s, \gamma_h, \nu] \) by solving the following constrained optimization problem:

\[
\hat{\mathbf{p}} = \min_{\mathbf{p}} \left( \frac{1}{n} \sum_{i=1}^{n} |y_1(t_i) - \hat{Y}_1|^2 + \frac{1}{m} \sum_{j=1}^{m} |y_2(t_j) - \hat{Y}_2|^2 \right) \quad \text{w.r.t} \quad \mathbf{p} > 0
\]

and \( \beta_s = \frac{R_0}{k(1 - \rho) \gamma_n + \rho \frac{1}{\gamma_s + \alpha}} \frac{S(0)}{N(0)}, \quad \nu < 1, \quad 1/28 < \gamma_h < 1/7, \quad \gamma_s < 1, \quad \alpha = 0.1765\gamma_s \) (16)

where \( \hat{Y}_1 \) and \( \hat{Y}_2 \) are average daily US COVID-19 cases and deaths respectively from March 3 to April 26, 2020:

\[
\hat{Y}_1 = \frac{1}{n} \sum_{i=1}^{n} Y_1^i \quad \text{and} \quad \hat{Y}_2 = \frac{1}{m} \sum_{j=1}^{m} Y_2^j. \quad (17)
\]

We use the confirmed COVID-19 cases for the parameter estimation problem. However, it is worth mentioning that the due to limited (testing only individuals with moderate to severe symptoms especially early in the pandemic) and imperfect (false negative results) testing, high number of cases in the USA were undetected or under-reported [39]. It is estimated that the actual COVID-19 cases is 3–20 times higher than the reported cases [39]. In our modeling approach, we take this into consideration with \( \rho \)-proportion of individuals who are getting tested. Another consequence of limited testing capacities is that true initial population size \( N(0) \) is not captured in the data. We pre-estimate the initial effective population size, \( N(0) \), and then fix the initial conditions, since the goal of this study is whether we can uniquely determine the epidemiologically important COVID-19 parameters such as the transmission rates of the different infectious classes. We present the estimated parameters in Table 3, and model (1) predictions with daily COVID-19 cases and deaths in Figure 2.

### 3.3. Practical identifiability analysis

We continue the identifiability analysis by performing Monte Carlo Simulations (MCS) and study whether the parameters of the model can be estimated by solving the constraint minimization problem numerically (16) from the noisy observations \( \mathbf{Y} = (Y_1^i, Y_2^j) \) (15). This type of identifiability is called practical identifiability [22]. To determine the practical identifiability, first we estimate the model parameters by (16) using the reported daily incidences and deaths. Then using the estimated parameters \( \hat{\mathbf{p}} \), we generate synthetic data sets \( \mathbf{Y} = (Y_1^i, Y_2^j) \) which are contaminated with error. We perform the MCS with two different error structures: (i) normal distribution and (ii) Poisson distribution.
Figure 2. Number of COVID-19 cases and deaths in the USA from March 3 to April 26, 2020 (red dots) plotted together with the model predictions (blue curve) with the estimated values presented in Table 3. Pre-estimated initial conditions are $S(0) = 3,000,000, E(0) = 2500, I_n(0) = 1000, I_s(0) = 63, l_b(0) = 20, R(0) = 0$.

(i) Error structure with normal distribution: We assume that the measurement error is proportional to data. Namely we assume that the measurement error has the following form:

$$\epsilon_i = y_1(t_i)\varepsilon_i$$ and similarly $$\epsilon_j = y_2(t_j)\varepsilon_j$$

where $\varepsilon_i$ and $\varepsilon_j$ are independent and identically distributed with mean zero and standard deviation $\sigma_0$. Thus the random variables $Y_i^1$ and $Y_j^1$ have mean $y_1(t_i)$ and $y_2(t_j)$ respectively, and standard deviation $y_1(t_i)\sigma_0$ and $y_2(t_j)\sigma_0$ respectively. In particular, we generate $M = 1000$ data sets using normal distribution whose mean is the model predictions and standard deviations are $y_1(t_i)\sigma_0$ and $y_2(t_j)\sigma_0$ respectively. That is

$$Y_i^1 = \mathcal{N}(y_1(t_i, \hat{p}), y_1(t_i, \hat{p})\sigma_0) \quad \text{and} \quad Y_j^1 = \mathcal{N}(y_2(t_j, \hat{p}), y_2(t_j, \hat{p})\sigma_0)$$

(18)

where $\mathcal{N}(\mu, \sigma)$ denotes the normal distribution with mean $\mu$ and standard deviation $\sigma$.

We generate $M = 1000$ synthetic data sets for each measurement errors, by increasing the errors gradually from $\sigma_0 = 1\%$, to $\sigma_0 = 5\%, 10\%$ and $\sigma_0 = 20\%$. In Figure 3, we present a single synthetic incidence ($Y_i^1$) and deaths ($Y_j^1$) for each measurement error $\sigma_0$. Comparing Figure 3 with the actual data in Figure 2, it appears the actual data has about 20% error.

(ii) Error structure with poisson distribution: For the Poisson distribution, the mean and the standard deviation are equal.

$$Y_i^1 = \mathcal{P}(y_1(t_i, \hat{p})) \quad \text{and} \quad Y_j^1 = \mathcal{P}(y_2(t_j, \hat{p}))$$

(19)

where $\mathcal{P}(\mu)$ denotes the Poisson distribution with mean $\mu$. We present the synthetic data generated by Poisson distribution in Figure 4. It is apparent from Figure 2, Figure 3 and Figure 4 that the error structure in actual data most likely has normal distribution with $\sigma_0 = 20\%$. 
Figure 3. Top row: A single set of incidences data generated for the Monte Carlo Simulations is presented for each measurement error. The incidence data is generated by taking the estimated parameters given in Table 3 as true parameters $\hat{\mathbf{p}}$ and generating random incidences with normal distribution whose mean is $y_1(t, \hat{\mathbf{p}})$ and the standard deviation is $y_1(t, \hat{\mathbf{p}})\sigma_0$. Bottom row: The simulated deaths data is generated by taking the estimated parameters given in Table 3 and as true parameters $\hat{\mathbf{p}}$ and generating random deaths with normal distribution whose mean is $y_2(t, \hat{\mathbf{p}})$ and the standard deviation is $y_2(t, \hat{\mathbf{p}})\sigma_0$ as in (18). First column is for $\sigma_0 = 1\%$, the second column is for $\sigma_0 = 5\%$, the third column is for $\sigma_0 = 10\%$ and the fourth column is for $\sigma_0 = 20\%$. We stopped at $\sigma_0 = 20\%$, because comparison with the actual data suggests that the measurement error is about 20%.

Figure 4. Presents a single set of incidences and deaths data generated by Poisson error distribution. Data is generated by taking the estimated parameters given in Table 3 as true parameters $\hat{\mathbf{p}}$ and generating random data with Poisson distribution.

We then fit the COVID-19 epidemic model (1) to each $M = 1000$ data sets by solving the optimization problem (16) and then calculate the Average Relative Errors (AREs) of each parameter for each noise level. The AREs are computed as

$$\text{ARE}(p^k) = 100\% \frac{1}{M} \sum_{j=1}^{M} \frac{|\hat{p}^k - p_j^k|}{\hat{p}^k}$$

where $p^k$ is the $k^{th}$ parameter in $\mathbf{p}$ and $\hat{\mathbf{p}}$ are the parameters that generate the random variables $Y_1^i$ and $Y_2^j$. AREs of the parameters for each measurement error are presented
Table 4. Monte Carlo simulation results: Average relative errors (AREs) of parameters of the COVID-19 model (1) are presented after performing parameter estimation problem (16) to \( M = 1000 \) synthetic data sets for each measurement error \( \sigma_0 \).

| Parameter | \( \sigma_0 = 1\% \) ARE | \( \sigma_0 = 5\% \) ARE | \( \sigma_0 = 10\% \) ARE | \( \sigma_0 = 20\% \) ARE |
|-----------|-------------------------|-------------------------|-------------------------|-------------------------|
| \( \beta_s \) | 78 | 238 | 294 | 426 |
| \( q_E \) | 13.7 | 41.2 | 77.7 | 628.6 |
| \( q_n \) | 9.1 | 11.9 | 14.3 | 18.1 |
| \( \nu \) | 2.6 | 4.9 | 6.1 | 7.9 |
| \( \gamma_s \) | 1.3 | 3.4 | 4.6 | 7.6 |
| \( \gamma_h \) | 3.7 | 11.8 | 14.4 | 17.5 |

Table 5. Monte Carlo simulation results: average relative errors (AREs) and 95% confidence intervals of parameters of the COVID-19 model (1) are presented after performing parameter estimation problem (16) to \( M = 1000 \) synthetic data sets with Poisson error structure.

| Parameter | ARE | Confidence Interval |
|-----------|-----|---------------------|
| \( \beta_s \) | 49 | [6.4 \times 10^{-4}, 3.0 \times 10^{-4}] |
| \( q_E \) | 10.7 | [142, 674] |
| \( q_n \) | 8.2 | [21673, 99999] |
| \( \nu \) | 3.3 | [0.825, 0.953] |
| \( \gamma_s \) | 2.2 | [0.0298, 0.0334] |
| \( \gamma_h \) | 7.4 | [0.0845, 0.1429] |

in Table 4 and Table 5. AREs give insight about the practical identifiability of the model parameters. In Table 4, we present the AREs of each parameter when the error structure in the data is assumed to be normal. AREs of each parameter increases as the measurement error level (\( \sigma_0 \)) increases. We see that two of the parameters, namely \( \beta_s \) and \( q_E \), have high AREs compared to other parameters. The AREs of parameters \( \beta_s \) and \( q_E \) are more than the 20 times the measurement error when \( \sigma_0 = 20\% \). On the other hand, the AREs of the parameters \( q_n, \nu, \gamma_s \) and \( \gamma_h \), are bounded by the measurement error at \( \sigma_0 = 20\% \). Based on AREs in Table 4, we conclude that the parameters \( \beta_s \) and \( q_E \) are not practically identifiable. Parameters \( q_n, \nu, \gamma_s \) and \( \gamma_h \) are practically identifiable.

Next we run the MCS by generating \( M = 1000 \) data sets using Poisson error structure with mean \( y_1(t, \hat{p}) \) and \( y_2(t, \hat{p}) \). We then compute the AREs and the results are presented in Table 5. Checking the ARE of each parameter in Table 5, we see that the parameters \( \beta_s \) and \( q_E \) have higher ARE compared to other parameters and declare these parameters to be practically unidentifiable. Comparing the ARE values in Table 4 with ARE values in Table 5, regardless of the error structure assumption, Normal vs. Poisson, we see that in both cases the parameters \( \beta_s \) and \( q_E \) have significantly higher AREs compared to other parameters. In conclusion, error structure does not influence the parameter identifiability.

We continue with the identifiability analysis by assessing the evolution of the parameter identifiability. Thus we investigate how the parameter identifiability changes as function of time. We performed the MCS simulations at three different stages of the pandemic; (i) before it reached its peak, (ii) at the peak and (iii) after the peak. Since normal error
Table 6. Monte Carlo simulation results: Average relative errors (AREs) of parameters of the COVID-19 model (1) are presented after performing parameter estimation problem (16) to \( M = 1000 \) synthetic data sets for each measurement error \( \sigma_0 \). Error structure is assumed to be normal.

| Parameter | \( \sigma_0 = 1\% \) | \( \sigma_0 = 5\% \) | \( \sigma_0 = 10\% \) | \( \sigma_0 = 20\% \) |
|-----------|---------------------|---------------------|---------------------|---------------------|
| Before the peak | ARE | ARE | ARE | ARE |
| \( \beta_s \) | 106 | 512 | 1023 | 2520 |
| \( q_E \) | 24.5 | 52.2 | 74.1 | 126.2 |
| \( q_n \) | 19.5 | 24.6 | 23.4 | 25.7 |
| \( \nu \) | 2.2 | 4.9 | 6.1 | 7.8 |
| \( \gamma_s \) | 1.7 | 3.7 | 5.6 | 13.1 |
| \( \gamma_h \) | 8.4 | 14.3 | 16.7 | 17.3 |

| Parameter | \( \sigma_0 = 1\% \) | \( \sigma_0 = 5\% \) | \( \sigma_0 = 10\% \) | \( \sigma_0 = 20\% \) |
|-----------|---------------------|---------------------|---------------------|---------------------|
| At the peak | ARE | ARE | ARE | ARE |
| \( \beta_s \) | 21 | 179 | 270 | 1359 |
| \( q_E \) | 10.5 | 33.9 | 120.1 | 267 |
| \( q_n \) | 8.5 | 14.1 | 17.4 | 24.9 |
| \( \nu \) | 2 | 4 | 6 | 9.6 |
| \( \gamma_s \) | 2.4 | 4.9 | 7.7 | 15.6 |
| \( \gamma_h \) | 51.2 | 98.9 | 115.7 | 123.3 |

Before the peak At the peak

structure is a better representation of the error structure in the actual data, we performed the MC simulations with normal error structure. We present the AREs of each parameter before the peak and at the peak in Table 6. AREs presented in Table 4 are the AREs of the parameters after the peak. The AREs of the infection rate of symptomatic individuals, \( (\beta_s) \), decrease as the pandemic progresses, same is true for \( q_n \). We do not observe the decline of the AREs for the parameter \( q_E \). On the other hand, AREs of the parameters \( \nu \) and \( \gamma_s \) remain almost the same for all three phases of the pandemic. And AREs of the parameter \( \gamma_h \) increase from before the peak to at the peak phase and then decrease again from the peak to the after the peak phase. For this practical identifiability analysis, we conclude that as the pandemic progresses, as we add more data points to the parameter estimation problem, it is not possible to say that the identifiability of the parameters improves.

4. Forecasting multiple outbreak waves and determining control strategies

4.1. Predicting the future epidemic waves

Kermack–McKendrick type SEIR epidemic model (M1) or its variations such as (M2) and (1) can only forecast a single outbreak wave. Due to their nature, these types of outbreak models cannot predict multiple epidemic waves. In this section, we give a methodology on how to use an epidemic model to predict multiple waves. We use the methodology used in sub-epidemic curves introduced in [6]. Authors in [6] used a single generalized-logistic growth model which describes the incidences to forecast multiple epidemic waves. Here we generalize this methodology to outbreak models. We model the epidemic wave which consist of three overlapping sub-epidemics using the COVID-19 model (1):

\[
\begin{align*}
S_i' &= -A_i(t) \left( \frac{\beta_E_i E_i S_i + \beta_n_i I_n_i S_i + \beta_s_i I_s_i S_i}{N_i - I_{hi}} \right), \\
E_i' &= A_i(t) \left( \frac{\beta_E_i E_i S_i + \beta_n_i I_n_i S_i + \beta_s_i I_s_i S_i}{N_i - I_{hi}} - k_i E_i \right), \\
I_{ni}' &= A_i(t) \left( (1 - \rho_i) k_i E_i - \gamma_n_i I_{ni} \right), \\
I_{si}' &= A_i(t) \left( \rho_i k_i E_i - (\alpha_i + \gamma_s_i) I_{si} \right), \\
I_{hi}' &= A_i(t) \left( \alpha_i I_{si} - (\gamma_h_i + \nu_i) I_{hi} \right), \\
R_{i}' &= A_i(t) \left( \gamma_n_i I_{ni} + \gamma_s_i I_{si} + \gamma_h_i I_{hi} \right),
\end{align*}
\]
Figure 5. Blue curve is the prediction of the sub-epidemic model (21) to US COVID-19 cases and deaths (bars).

where

\[ A_i(t) = \begin{cases} 
0 & \text{for } t < t_i \\
1 & \text{for } t \geq t_i 
\end{cases} \]

for \( i = 1, 2, 3 \) with \( t_1 < t_2 < t_3 \) and \( t_1 = 0 \). For the sub-epidemic model (21), the observation of incidences and deaths becomes respectively as follows:

\[ y_1(t) = \rho_1 k_1 E_1 + \rho_2 k_2 E_2 + \rho_3 k_3 E_3 \quad \text{and} \quad y_2(t) = \nu_1 I_{h1} + \nu_2 I_{h2} + \nu_3 I_{h3}. \]

To estimate the model parameters, we modified the optimization problem (16) for the sub-epidemic model (21). The estimated parameters are presented in Table 7. The fit is shown in Figure 5.
Table 7. Parameter estimation results: parameter estimates of the sub-epidemic model (21).

| Parameter | $\beta_s$ | $q_E$ | $q_n$ | $\gamma_s$ | $\gamma_n$ | $\nu$ |
|-----------|----------------|--------|--------|-------------|-------------|--------|
| Estimated value | $4.2 \times 10^{-5}$ | 5428 | 47368 | 0.05 | 0.14 | 0.24 |

4.2. An optimal control scenario

US states initiated various community mitigation policies in March 2020. Some of the widely implemented strategies included the issuance of orders requiring persons to stay home, schools were closed and universities transitioned to online instruction. Stay-at-home orders started from California on March 19, 2020, and ultimately were issued in all 50 states, covering 316 million people between March 19 and May 20, 2020 [30]. To model the social distancing, we introduce control $u_1(t)$ which effectively reduces the transmission rates for the exposed $E$ and non-symptomatic individuals $I_n$. To model the self-isolation required from symptomatic individuals $I_s$, we introduce a control $u_2(t)$. Although, direct correspondence between the real life controls, social distancing and self-isolation to our controls $u_1$ and $u_2$ is not clear, such an investigation still yields to useful observations, particularly in the extreme case, when we want to achieve maximum effect. Both controls belong to the admissible set

$$A = \{ u \in L^1(0, T) | 0 \leq u(t) \leq 1 \}$$

where $T$ is the final time considered. The model with the controls takes the following form:

$$
\begin{align*}
S' &= -\beta_s q_E (1 - u_1) ES + q_n (1 - u_1) I_n S + (1 - u_2) I_h S - \frac{N - I_h}{N - I_n} \\
E' &= \beta_s q_E (1 - u_1) ES + q_n (1 - u_1) I_n S + (1 - u_2) I_h S - kE \\
I'_n &= (1 - \rho) kE - \gamma_n I_n \\
I' &= \rho I_s - (\alpha + \gamma_s) I_s \\
I'_h &= \alpha I_s - (\gamma_h + \nu) I_h \\
R' &= \gamma I_n + \gamma_s I_s + \gamma_h I_h
\end{align*}
$$

(22)

The main goal of the quarantine was to reduce the total number of infected individuals over the time $[0, T]$. To express that we use the following cost functional:

$$
\mathcal{J}(u_1, u_2) = \int_0^T \left( E + I_n + I_s + I_h + B_1 u_1^2 + B_2 u_2^2 \right) dt = \int_0^T L(x, u) dt
$$

(23)

Our goal is to find an optimal control functions $(u_1^*, u_2^*)$ such that

$$J(u_1^*, u_2^*) = \min_{u \in A} J(u_1, u_2) \quad \text{subjected to the system (20)}$$

We do not consider isolation to be a control variable, as isolation was applied only to severe cases. Since 80–85% are mild to moderate [23], hospitalizations are determined by the progression to severe symptoms and cannot be controlled.
Given an optimal pair 

Theorem 4.1: Given an optimal pair \((x^*,u^*)\), there exists adjoint variables \(\lambda_i\), \(i = 1, 2, \ldots, 6\) such that

\[
\lambda_1' = (\lambda_1 - \lambda_2) \frac{\beta_E(1-u_1)E + \beta_n(1-u_1)I_n + \beta_s(1-u_2)I_s}{S + E + I_n + I_s + R} \\
+ (\lambda_2 - \lambda_1) \frac{\beta_E(1-u_1)ES + \beta_n(1-u_1)I_nS + \beta_s(1-u_2)I_sS}{(S + E + I_n + I_s + R)^2} \\
\]

\[
\lambda_2' = -1 + (\lambda_1 - \lambda_2) \frac{\beta_E(1-u_1)S}{S + E + I_n + I_s + R} + (\lambda_2 - \lambda_1) \\
\times \frac{\beta_E(1-u_1)ES + \beta_n(1-u_1)I_nS + \beta_s(1-u_2)I_sS}{(S + E + I_n + I_s + R)^2} \\
+ (\lambda_2 - \lambda_3)k + (\lambda_3 - \lambda_4)\rho k \\
\]

\[
\lambda_3' = -1 + (\lambda_1 - \lambda_2) \frac{\beta_n(1-u_1)S}{S + E + I_n + I_s + R} + (\lambda_2 - \lambda_1) \\
\times \frac{\beta_E(1-u_1)ES + \beta_n(1-u_1)I_nS + \beta_s(1-u_2)I_sS}{(S + E + I_n + I_s + R)^2} \\
+ (\lambda_3 - \lambda_6)\gamma_n \]

\[
\lambda_4' = -1 + (\lambda_1 - \lambda_2) \frac{\beta_s(1-u_2)S}{S + E + I_n + I_s + R} + (\lambda_2 - \lambda_1) \\
\times \frac{\beta_E(1-u_1)ES + \beta_n(1-u_1)I_nS + \beta_s(1-u_2)I_sS}{(S + E + I_n + I_s + R)^2} \\
\]
Setting

\[
\begin{align*}
\lambda'_5 &= -1 + (\lambda_5 - \lambda_6)r + \lambda_5v \\
\lambda'_6 &= (\lambda_2 - \lambda_1)\frac{\beta_E(1 - u_1)E + \beta_n(1 - u_1)I_n + \beta_2(1 - u_2)I_sS}{(S + E + I_n + I_s + R)^2}
\end{align*}
\]

with \(\lambda_i(T) = 0\) for \(i = 1, 2, \ldots, 6\). The optimal controls \(u^* = (u_1^*, u_2^*)\) are represented by

\[
\begin{align*}
u_1^* &= \min\left(1, \max\left(0, \frac{(\lambda_2 - \lambda_1)S(\beta_EE + \beta_nI_n)}{2B_1(S + E + I_n + I_s + R)}\right)\right) \\
u_2^* &= \min\left(1, \max\left(0, \frac{(\lambda_2 - \lambda_1)\beta_2I_sS}{2B_2(S + E + I_n + I_s + R)}\right)\right)
\end{align*}
\]

**Proof:** Pontryagin minimum principle states that for the optimality of control \(u^*(t)\) and corresponding trajectory \(x^*\) where \(\forall t \in [0, T]\), it is necessary that there exist a nonzero adjoint vector function \(\lambda^*(t)\) which is a solution to the adjoint system

\[
\begin{cases}
\lambda'(t) = -\frac{\partial H(x, u, \lambda)}{\partial x} \\
\lambda(T) = 0
\end{cases}
\]

Setting

\[
\lambda'_1 = -\frac{\partial H}{\partial S}, \quad \lambda'_2 = -\frac{\partial H}{\partial E}, \quad \lambda'_3 = -\frac{\partial H}{\partial I_n}, \quad \lambda'_4 = -\frac{\partial H}{\partial I_s}, \quad \lambda'_5 = -\frac{\partial H}{\partial I_h}, \quad \lambda'_6 = -\frac{\partial H}{\partial R}
\]

we obtain the adjoint system with \(\lambda(T) = 0\). By the optimality condition,

\[
\frac{\partial H}{\partial u_1} = 2B_1u_1 + \frac{\lambda_1\beta_EE + \lambda_1\beta_nI_nS}{(N - I_h)} - \frac{\lambda_2\beta_EE + \lambda_2\beta_nI_nS}{(N - I_h)} = 0
\]

we obtain

\[
u_1^* = \frac{(\lambda_2 - \lambda_1)S(\beta_EE + \beta_nI_n)}{2B_1(S + E + I_n + I_s + R)}
\]

Similarly,

\[
\frac{\partial H}{\partial u_2} = 2B_2u_2 + \frac{\lambda_1\beta_2I_sS - \lambda_2\beta_2I_sS}{N - I_h} = 0 \implies u_2^* = \frac{(\lambda_2 - \lambda_1)\beta_2I_sS}{2B_2(S + E + I_n + I_s + R)}
\]

Notice that \(\frac{\partial^2 H}{\partial u_1^2} = 2B_1 \geq 0\), and \(\frac{\partial^2 H}{\partial u_2^2} = 2B_2 \geq 0\). Since, \(u^* \in \mathcal{A}\) we set

\[
\begin{align*}
u_1^* &= \min\left(1, \max\left(0, \frac{(\lambda_2 - \lambda_1)S(\beta_EE + \beta_nI_n)}{2B_1(S + E + I_n + I_s + R)}\right)\right) \quad \text{and} \\
u_2^* &= \min\left(1, \max\left(0, \frac{(\lambda_2 - \lambda_1)\beta_2I_sS}{2B_2(S + E + I_n + I_s + R)}\right)\right)
\end{align*}
\]
4.2.1. Numerical simulations with optimal control

We use forward–backward sweep method [18] to approximate the solution of model (22) with controls $u_1, u_2$. We use MATLAB built-in function `ode15s` to approximate the solutions of the model (22) with controls by taking initial conditions $x(0) = x_0$ and also use the `ode15s` to solve the adjoint differential differential equations backward in time by taking the value at the final time as $\lambda(T) = 0$. We use the following forward–backward sweep algorithm to get the optimal control functions [18].

(1) Take initial guesses for $u = (u_1(t), u_2(t))$
(2) Solve for the state function differential Equation (22) using `ode15s` forward in time.
(3) Solve for the adjoint function differential Equation $\lambda'_i(t) = \frac{\partial J}{\partial x_i}$, $\lambda(T) = 0$ using `ode15s` backward in time.
(4) We update the new values of $u = (u_1(t), u_2(t))$ using (26) and (27).
(5) Iterate the same process until the following convergence criteria is met for every state variable, adjoint and control functions

$$\frac{||x^k - x^{k-1}||}{||x^k||} < \delta \quad \text{and} \quad \frac{||u^k - u^{k-1}||}{||u^k||} < \delta \quad \text{and} \quad \frac{||\lambda^k - \lambda^{k-1}||}{||\lambda^k||} < \delta \quad \text{for } \delta > 0.$$

For the optimal control simulations, we use the parameter values obtained in the fitting as presented in Table 3. For the first simulation, we solve the optimal control problem with controls social distancing and self-isolation setting off at time $t = 0$ and effective for 75 days. Figure 6 presents the state of infectious populations with and without social distancing and self-isolation. We see that the controls are most effective when it is applied at 100%, that is at their upper bound of 1. We recognize that achieving 100% social distancing or isolation is not practically possible. However our results show that the higher the social distancing or isolation the lower the infection. Therefore, depending on the level of the cost parameters and our cost structure, one should aim at the maximum possible social distancing and isolation. Even if we set a realistic upper bound, $u_{\text{max}}$ for the controls which is strictly less than 1, the optimal control will still reside at the $u_{\text{max}}$ most of the time. For the next simulation, we apply the social distancing and self-isolation from March 19 till May 20, for about 2 months as this was the case for the USA. We perform the iterative process to find the optimal pair which would be only effective from March 19 till May 20 that would minimize the objective function (23) subject to the COVID-19 model (22) starting on March 3 and ending on June 23, 2020. In Figure 7, we only present the infectious states with controls, since the infectious populations without controls are same as in Figure 6 second row. Comparing Figure 6 and Figure 7, we see that, one of the main differences in starting the social distancing and self-isolation 2 weeks after the launch of the outbreak is that the number of infectious classes are at least 1 magnitude order higher if not more. Further, we see, comparing Figure 6 with Figure 7 that if we had continue to apply the controls there would not have been an uptick in cases.

5. Discussion

Mathematical compartmental models can provide valuable insight about the spread of SARS-CoV-2 in a population. In particular, they can provide guidance about the role of
specific interventions in reducing the impact of an ongoing disease outbreak. All models are limited by factors that include the levels of complexity in the epidemiological states being considered, the parameters to be estimated, and the limited data that often involves delay and uncertainty. These models have the potential to provide valuable guidance for pandemic preparedness and planning to reduce the burden of an outbreak, particularly during the early phase of an outbreak. In this study, we proposed a COVID-19 epidemic model that considers six epidemic states and incorporates important phases of the coronavirus life cycle, such as incubation and latency period. Scenarios that incorporate social distancing efforts are implemented to limit the spread of SARS-CoV-2 were considered in
our study. We computed the reproduction number, estimated model parameters based on daily cases and deaths, and conducted structural and practical identifiability analysis for the proposed model.

We show that parameters of the proposed model are not structurally identifiable from observations of daily incidence and deaths. We further establish that for a sequence of models, nested in the proposed model, only the SEIR model where all incidences are observed, is structurally identifiable. For our proposed model, we show that if the parameters $\alpha, \gamma_n, k, \rho$ are fixed from external information, then all parameters are structurally identifiable, except $\beta_s$. To identify $\beta_s$ we use estimates of $R_0$ from the literature.

We further investigated the practical identifiability of these parameters by performing Monte Carlo Simulations and found that transmission rate for exposed $\beta_E = q_E \beta_s$ and symptomatic infectious $\beta_s$ were not practically identifiable. This result largely persists even if we use outside information for the proportion of individuals who progress to hospitalization. A surprising outcome of the practical identifiability analysis was that fixing the proportion of individuals who become hospitalized in fact makes the parameters of the model practically identifiable. The practical identifiability conclusions remain the same regardless of the error structure (Poisson or Normal) chosen. We also investigated whether taking a shorter time series compared to taking a longer time series affected identifiability results. For the most part identifiability improved with the longer data set but we think that the results were not strong enough and uniform enough to be conclusive.

The investigation of the optimal control problem shows that social distancing should be practiced at 100% to reduce the disease quickly, while isolation can be practiced at more relaxed conditions, particularly as time progresses. Social distancing should be practiced at 100% for most of the time interval even if the cost of that is very high. High levels of social distancing lead to exceptionally high reduction in the number infected individuals without symptoms and hospitalized individuals which leads to quick curtailing of the pandemic and maintenance of susceptible individuals at higher level.

An important finding of our study is our investigation of social distancing and self-isolation as control measures. We investigated two scenarios: control is active during the entire studied interval of time and the control is active only when the lock down was in effect, which is a subset of time. With the first scenario both optimal controls gradually go to the maximum and stay there for the remainder of the studied duration. In this case, the epidemic states are well controlled and reach peaks at least an order of magnitude smaller than with the second scenario when the control is applied with delay. In the second scenario, the optimal controls jump to the maximum and stay there for the entire allowed duration. However, the numbers in the infected classes are much larger, and as soon as the control is stopped, these classes start growing in magnitude. The observation we make here support again the rule of thumb that control measures have to be applied as early as possible in an epidemic as strongly as possible. Our numerical simulations for each of the epidemic states of the proposed model show the critical importance of the timely implementation of control measures, particularly in reducing the number of individuals at risk of exposure and infection from the coronavirus. The implications of these findings are most relevant during the very early phase of the pandemic as the simulations are based on data captured very early-on. To further investigate the trajectories of these epidemic states and parameters as the pandemic progresses beyond its early phase, we fit overlapping epidemic waves to the data of the USA as described by [6]. This allows us to investigate how the
virus and epidemic characteristics of the COVID-19 pandemic have evolved over the year since it first became established in the USA. To accurately model the spread of the coronavirus and the role of specific interventions, it is important that we update modelling efforts to account for new virus information, as well as additional relevant interventions such as testing and contact tracing [3,12,17]. It is also important to consider that morbidity and mortality associated with COVID-19 differ among individuals in the population and such trends in risk have also evolved over time [7,24,37]. During the early phase of the COVID-19 pandemic in the USA older age individuals seemed to have experienced a higher burden while in the current months of this pandemic, younger and mid-age individuals seem to experience higher risk of infection. A consistent observation with this pandemic is that individuals of older age, and higher pre-existing comorbidities such as heart disease, diabetes and obesity have a higher risk of infection and severe COVID-19 experience [15,25,28].

In summary, this study presents a COVID-19 model that investigates the spread of SARS-CoV-2 and the role control interventions play to reduce the burden associated with the COVID-19 pandemic. In particular, our study of the structural and practical identifiability of the proposed model highlights the importance of parameter identifiability and the limitations inherent to uncertainty in data available, particularly during the early phase of a novel pandemic. We have presented some unique methodology for turning unidentifiable into identifiable parameters. Using a mixed approach to the problem of parameter identifiability, we are able to provide specific guidance about the identifiability of a subset of parameters that can be useful in guiding some of the model predictions. We further investigate the relevance of control measures through social distancing efforts and self-isolation. We show that the timely implementation of control measures can be extremely valuable to reduce the burden of this novel pandemic.

We present all Matlab code for parameter estimation and Monte Carlo Simulations at https://github.com/NecibeTuncer/Identifiability_COVID-19_Early_in_Pandemic together with the DAISY and Mathematica results for the identifiability analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

G.C. is partially supported from NSF grants 2125246 and 2026797 and R01 GM 130900.

ORCID

Necibe Tuncer  http://orcid.org/0000-0002-6388-2499

References

[1]  G. Bellu, M.P. Saccomani, S. Audoly, and L. D’Angio, Daisy – differential algebra for identifiability of systems, Available at https://daisy.dei.unipd.it.
[2]  H.R. Bhapkar, P.N. Mahalle, N. Dey, and K.C. Santosh, Revisited covid-19 mortality and recovery rates: Are we missing recovery time period?, J. Med. Syst. 44 (2020), pp. 1–5.
[3] A. Bilinski, F. Mostashari, and J.A. Salomon, Modeling contact tracing strategies for covid-19 in the context of relaxed physical distancing measures, JAMA network open 3(8) (2020), pp. e2019217.

[4] CDC, Severe outcomes among patients with coronavirus disease 2019 (covid-19) – united states, February 12–March 16, 2020, MMWR 69 (2020), pp. 343–346.

[5] K. Roosa and G. Chowell, Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models, Theoretical Biol. Med. Model. 16 (2019), pp. 1–15.

[6] G. Chowell, A. Tariq, and J.M. Hyman, A novel sub-epidemic modeling framework for short-term forecasting epidemic waves, BMC. Med. 17 (2019), pp. 1–18.

[7] R.M. Colombo, M. Garavello, F. Marcellini, and E. Rossi, An age and space structured sir model describing the covid-19 pandemic, medRxiv 10 (2020), pp. 1–20.

[8] R. Dandekar and G. Barbastathis, Quantifying the effect of quarantine control in covid-19 infectious spread using machine learning, medRxiv2020

[9] M.C. Eisenberg and Y-H. Kao, Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment, Epidemics 25 (2018), pp. 89–100.

[10] M.C. Eisenberg, S.L. Robertson, and J.H. Tien, Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease, J. Theor. Biol. 324 (2013), pp. 84–102.

[11] C. Elias, A. Sekri, P. Leblanc, M. Cucherat, and P. Vanhems, The incubation period of covid-19: A meta-analysis, Int. J. Infect. Dis. 104 (2021), pp. 708–710.

[12] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri, Modelling the covid-19 epidemic and implementation of population-wide interventions in italy, Nat. Med. 26 (2020), pp. 855–860.

[13] G. González-Parra and A.J. Arenas, Qualitative analysis of a mathematical model with presymptomatic individuals and two sars-cov-2 variants, Computational and Applied Mathematics 40 (2021), pp. 1–25.

[14] X. He, E. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y.C. Lau, J. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. Cowling, F. Li, and G. Leung, Temporal dynamics in viral shedding and transmissibility of covid-19, Nat. Med. 26 (2020), pp. 672–675.

[15] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, and Z. Cheng, Clinical features of patients infected with 2019 novel coronavirus in wuhan, china, The lancet 395(10223) (2020), pp. 497–506.

[16] G.G. Katul, A. Mrad, S. Bonetti, G. Manoli, and A.J. Parolari, Global convergence of covid-19 basic reproduction number and estimation from early-time sir dynamics, PLoS. ONE. 15 (2020), pp. e023980.

[17] A.J. Kucharski, P. Klepac, A. Conlan, S.M. Kissler, M. Tang, H. Fry, J. Gog, and J. Edmunds, CMMID COVID-19 Working Group: et al. Effectiveness of isolation, testing, contact tracing and physical distancing on reducing transmission of sars-cov-2 in different settings, medRxiv 20 (2020), pp. 1151–1160.

[18] S. Lenhart and J.T. Workman, Optimal control applied to biological models, CRC press, 2007.

[19] F.S. Lu, A.T. Nguyen, N.B. Link, J.T. Davis, M. Chinazzi, X. Xiong, A. Vespignani, M. Lipsitch, and M. Santillana, Estimating the cumulative incidence of covid-19 in the united states using four complementary approaches, medRxiv2020

[20] M. Martcheva, An introduction to mathematical epidemiology, Springer, Vol. 61, 2015.

[21] G. Massonis, J.R. Banga, and A.F. Villaverde, Structural identifiability and observability of compartmental models of the covid-19 pandemic, Annu. Rev. Control. 51 (2020), pp. 441–459.

[22] H. Miao, X. Xia, A.S. Perelson, and H. Wu, On identifiability of nonlinear ode models and applications in viral dynamics, SIAM. Rev. Soc. Ind. Appl. Math. 53 (2011), pp. 3–39.

[23] N.J.M. Michelen and C. Stavropoulou, In patients of covid-19, what are the symptoms and clinical features of mild and moderate cases? - cebm, preprint (2020 Apr). Available at http://www.cebm.net/covid-19/in-patients-of-covid-19-what-are-the-symptoms-and-clinical-features-of-mild-and-moderate-case/, [cited 31 Oct 2020].
[24] R. Omori, R. Matsuyama, and Y. Nakata, *The age distribution of mortality from novel coronavirus disease (covid-19) suggests no large difference of susceptibility by age*, Sci. Rep. 10(1) (2020), pp. 1–9.

[25] G. Onder, G. Rezza, and S. Brusaferro, *Case-fatality rate and characteristics of patients dying in relation to covid-19 in italy*, Jama 323(18) (2020), pp. 1775–1776.

[26] J.M. Read, J.R.E. Bridgen, D.A.T. Cummings, A. Ho, and C.P. Jewell, *Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions*, MedRxiv. 2020.

[27] A. Sakurai, T. Sasaki, and S. Kato, *et al Natural history of asymptomatic sars-cov-2 infection*, N. Engl. J. Med. 383 (2020), pp. 885–886.

[28] Y-J. Sun, Y-J. Feng, J. Chen, B. Li, Z.-C. Luo, and P.-X. Wang, *Clinical features of fatalities in patients with covid-19*, Disaster. Med. Public. Health. Prep. 15 (2020), pp. e9–11.

[29] B. Tang, X. Wang, Q. Li, N.L. Bragazzi, S. Tang, Y. Xiao, and J. Wu, *Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions*, J. Clin. Med. 9(2) (2020), pp. 462.

[30] The New York Times, *See which states and cities have told residents to stay at home - the new york times*, Available at https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html. [cited 31 Oct 2020].

[31] C. Tsay, F. Lejarza, M.A. Stadtherr, and M. Baldea, *Modeling, state estimation, and optimal control for the us covid-19 outbreak*, preprint (2020). Available at arXiv:2004.06291.

[32] N. Tuncer, H. Gulbudak, V. Cannataro, and M. Martcheva, *Structural and practical identifiability issues of immuno-epidemiological vector-host models with application to rift valley fever*, Bull. Math. Biol. 78 (2016), pp. 1796–1827.

[33] N. Tuncer, B. Labarre, S. Payoute, and M. Martcheva, *Structural and practical identifiability analysis of zika epidemiological models*, Bull. Math. Biol. 80 (2018), pp. 2209–2241.

[34] N. Tuncer and T. Le, *Structural and practical identifiability analysis of outbreak models*, Math. Biosci. 299 (2018), pp. 1–18.

[35] N. Tuncer, C. Mohanakumar, S. Swanson, and M. Martcheva, *Efficacy of control measures in the control of ebola, liberia 2014–2015*, J. Biol. Dyn. 12(1) (2018), pp. 913–937.

[36] United states coronavirus, *9, 332, 958 cases and 235, 420 deaths - worldometer*, preprint (2020 Oct). Available at https://www.worldometers.info/coronavirus/country/us/, [cited 31 Oct 2020].

[37] B. Wilder, M. Charpignon, J.A. Killian, H.-C. Ou, A. Mate, S. Jabbari, A. Perrault, A.N. Desai, M. Tambe, and M.S. Majumder, *Modeling between-population variation in covid-19 dynamics in hubei, lombardy, and new york city*, Proceedings of the National Academy of Sciences 117(41) (2020), pp. 25904–25910.

[38] J.T. Wu, K. Leung, and G.M. Leung, *Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study*, The Lancet 395(10225) (2020), pp. 689–697.

[39] S.L. Wu, A.N. Mertens, A. Crider, S. Yoshika, N.N.N. Pokpongkiat, S. Djajadi, A. Seth, M.S. Hsiang, J.M. Colford, A. Reingold, B.F. Arnold, A. Hubbard, and J. Benjamin-Chung, *Substantial underestimation of sars-cov-2 infection in the united states*, Nat. Commun. 11 (2020), pp. 1–10.

[40] Z. Yang, Z. Zeng, K. Wang, S-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, and Z. Mai, *et al Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions*, J. Thorac. Dis. 12(3) (2020), pp. 165.

[41] D. Zou, L. Wang, P. Xu, J. Chen, W. Zhang, and Q. Gu, *Epidemic model guided machine learning for covid-19 forecasts in the united states*, medRxiv, 2020.
Appendix

A.1. SEIR model

Input–output equations for the SEIR model (M₁), where \( y₁(t) = kE \), \( y₂(t) = vI \).

\[
0 = \frac{dy₂}{dt} - y₁v + y₂(v + α)
\]

\[
0 = \frac{d³y₁}{dt³} y₁β²v² + \frac{d³y₁}{dt³} y₁y₂βv(2βE - v) + \frac{d³y₁}{dt³} y₁y₂βvk(2βE - v) + \frac{d³y₁}{dt³} y₁y₂βEk²(βE - v) + \frac{d³y₁}{dt³} y₁³β²v²k
\]

\[
+ \frac{d²y₁}{dt²} y₂βvk²(2βE - v) + \frac{d²y₁}{dt²} y₁y₂βEk³(βE - v) - \frac{d²y₁}{dt²} y₁²β²v²k
\]

\[
+ 2\frac{d²y₁}{dt²} y₁y₂βvk(-βE + v)
\]

\[
+ \frac{d²y₁}{dt²} y₂βEk²(-βE + 2v) - 2\frac{d²y₁}{dt²} y₁β²v² + \frac{d²y₁}{dt²} y₂βvk(-2βE - v)
\]

\[
+ \frac{d²y₁}{dt²} y₁²βv²k(-4β - βE + v)
\]

\[
+ \frac{d²y₁}{dt²} y₁y₂vk(ββE + ββEα - 5ββEk - βv² - βvα + 3βvk - β²Ek - βEvk)
\]

\[
+ \frac{d²y₁}{dt²} y₂βEk²(βEα - βEk + v² + να + 3vk) + \frac{d²y₁}{dt²} y₁³β²v²k(-βE + v)
\]

\[
+ \frac{d²y₁}{dt²} y₂βEk²(ββE + ββEα + ββEk - βv² - βvα - β²Ek - βEvk)
\]

\[
+ \frac{d²y₁}{dt²} y₁y₂βEk³(βE + βEα + βEk + v² + να - vk) + 2\frac{dy₁}{dt} β²v²
\]

\[
+ \frac{dy₁}{dt} y₁β²vk(3β + 3βE - v)
\]

\[
+ \frac{dy₁}{dt} y₂vk(-4ββE - 4ββEα - ββEk + βv² + βvα + 2βvk - β²Ek + βEvk)
\]

\[
+ \frac{dy₁}{dt} y₂²d²k(ββE + ββEα + 5ββEk - βvk + 2β²Ek - βEvk)
\]

\[
+ \frac{dy₁}{dt} y₁y₂vk(-ββE² - 2ββEα - 7ββEvk - ββEα² - 7ββEvk - 3ββEk²
\]

\[
+ βv²k + βvαk - 3β²Evk - 3β²Eαk - 3β²Ek² + βE²vk² + βEvαk + βE²vk²)
\]

\[
+ \frac{dy₁}{dt} y₂βEk²(βE² + 2βEα + βEvk + βEα² + βEαk - βEk² + v²k + ναk + 2vk²)
\]

\[
+ 2\frac{dy₁}{dt} y₁³βE²vk²(βv + βα + βk + 2βEk - νk) + \frac{dy₁}{dt} y₁³β²Evk²
\]

\[
× (-2βv² - 4βvα - 3βvk - 2β²α - 3βαk - 6βEvk - 6βEvk - 2β²Ek² + 2v²k + 2vak)
\]

\[
+ \frac{dy₁}{dt} y₁y₂βEk³(2βE² + 4βEα + βEvk + 2βEα² + βEαk + v²k + ναk)
\]

\[
+ y₁³βE²vk³(βv + βα + 2βEvk - νk) + y₁³y₂βEvk³(-βv² - 2βvα
\]

\[
- βα² - 3βEvk - 3βEαk + v²k + ναk) + y₁³y₂β²Evk⁴(v² + 2να + ν²)
\]

(A1)