Investigating the Structural and Practical Identifiability of Adult-Child Compartment Epidemic Models

Zander Memon
Charlotte Moser
Zoe Plzak

Project Advisor: Dr. Tingting Tang

Grad Student Advisor: Anuradha Argwal

A Final Report for the 2022 San Diego State University
Mathematics REU

Mathematics NSF REU
San Diego State University
August 2022

Investigating the Structural and Practical Identifiability of Adult-Child Epidemic Models

Tingting Tang[1], Zander Memon[2], Charlotte Moser[3], Zoe Plzak[4], and Anuradha Agarwal[1]

[1]San Diego State University
[2]American University
[3]University of Wisconsin-La Crosse
[4]Boston University

August 5, 2022

Abstract: Compartment modeling has been used extensively in epidemics for both the understanding and prediction of infectious disease. One very important question in epidemic modeling is the balance between the complexity of the model and its generalization accuracy. Complex models may be able to explain existing experiment data better due to their abundance in parameters rather than describing the underlying phenomenon. In this study, we construct an age-based SEIR model that separates the population into Child and Adult compartments. We calculate the basic reproduction number, R_0, and compare our model to the existing simple SEIR model. We determine the structural identifiability of both models using existing algorithms and compute the practical identifiability of both models using Monte Carlo Simulation and Profile Likelihood approaches. We demonstrate the practical identifiability of the simple SEIR model and the practical non-identifiability of the SEIR aged-based model. We then present relevant parameter dependencies that construct a full set of practically identifiable parameters for the age-based SEIR model.

Acknowledgements

This research was funded by the National Science Foundation's Research Experience for Undergraduates program at San Diego State University. The authors would like to thank the NSF and San Diego State University for this funding. Additionally the authors would like to extend sincere thanks to Dr.Tingting Tang for her mentorship, support, and insight throughout the research process. We are also grateful for the guidance of graduate student Anuradha Agarwal whose support has been integral to the success of this project.

Contents

1 Introduction

The SIR model of epidemics has long been a key tool in epidemiology for understanding disease outbreaks and their spread. With the COVID-19 pandemic approaching an endemic state, there is a need for further research to understand, model, and predict disease dynamics. Previous research suggests that SEIR and SIR models can accurately represent and predict disease outbreaks, and that compartments within disease models can be further separated to model multiple interacting populations [1]. Studies such as those completed by [2] use age-structured SIR models to compare social-distancing and vaccination strategies within specific populations. Ram and Schaposnik conclude from their model that age-targeted measures can significantly change the outcome of an epidemic when one assumes that the contact rate between populations is variable, even if the probability of contracting COVID given contact is equal for every age group.

Rather than imposing strict measures on a community as a whole, by identifying the most vulnerable groups, one can implement effective age-specific measures to reduce the impact of an epidemic. For example, one study using COVID data from Wuhan, China, determined that the reduction in incidence data was greatest when employing distancing strategies amongst children and older individuals, but low amongst working adults [3]. Another study suggests while contact rate is insufficient to account for the magnitude of these differences, varied susceptibility between populations may play a role as well, implying that modeling multiple populations may offer insight into the influence of age in an individual's disease response [4].

Complex models can be used to predict outcomes and case numbers based on preexisting knowledge of biological information, such as recovery period and the infectiousness of a disease, but questions remain as to how one can use a model to determine the biological information about a model using data about case numbers. Analysis in structural and practical identifiability has provided insight into what models can be used to find unique parameters given perfect and imperfect data collected about a disease.

In this study, we develop a mathematical model separating the population into two age-specific categories based off an understanding of COVID-19's differing impact on children and adults. Our model incorporates the separate recovery and incubation rates for children, who have been less impacted by COVID-19, and adults, as well as the differences in infectiousness and contact rates between each population. The age-segregated model suggests that more information can be derived about disease modeling by looking at the interactions between overlapping populations than by assuming data fits a single homogeneous population.

1.1 R_0

The **basic reproductive number**, R_0, is an essential number that provides information about the dynamics of a disease. SEIR and SIR models use a system of ODEs to compute R_0. R_0 represents the number of new infections that can come from one infected individual in a totally susceptible population, and is the determinant of whether a disease will die out or become endemic. $R_0 > 1$ implies that a disease is self-sustaining and will become an epidemic, while if $R_0 < 1$, the disease dies out. Computing R_0 can give scientists a deeper understanding of disease dynamics and progress. Additionally, R_0 can serve as a metric on which to compare different epidemic models. Numerous studies in COVID data have reported an average R_0 value of between 2 and 3 across geographic locations [4].

1.2 Identifiability

While predicting R_0 and related disease progression is essential for combating future pandemics, it can be challenging to identify the necessary parameters for calculating the basic reproductive number. When studying a model designed to fit preexisting data, we must begin by investigating whether or not our model not only fits the data, but can accurately identify information about the disease and make predictions about future disease-preventing measures. We turn therefore to structural and practical identifiability, which determine whether a model's parameters can be uniquely identified from data.

A model is structurally identifiable if the model parameters can be uniquely determined based on the model formulation with continuous data and no noise. Structural identifiability has two sub-categories: global and local. A parameter is said to be **globally structurally identifiable** if the parameter can be uniquely recovered from the given output equation. Similarly, a parameter is said to be **locally structurally identifiable** if there are a finite number of solutions for the parameter that can be recovered from the output equation. On the contrary, if there are an infinite amount of values for one or more parameters, our model is considered to be non-identifiable. Often models that are structurally identifiable are not practically identifiable, but structural identifiability is a necessary condition for practical identifiability.

Because structural identifiability requires impossibly perfect conditions, it does not provide information on the applicability of the model to real-world data. Practical identifiability can fill in these gaps in knowledge. Establishing practical identifiability of a model can avoid inaccurate results in parameter estimation. However, practical identifiability is an incredibly local behavior and can be difficult to determine beyond specific parameter values and a specific timescale.

2 The SEIR Model

2.1 Simple SEIR Model

We consider the standard SEIR model defined by the set of equations

$$\frac{dS}{dt} = -\lambda S + \Pi N - \mu S \tag{1}$$

$$\frac{dE}{dt} = \lambda S - \epsilon E - \mu E \tag{2}$$

$$\frac{dI}{dt} = \epsilon E - \gamma I - \mu I \tag{3}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{4}$$

where S, E, I, and R represent the number of Susceptible, Exposed, Infected, and Recovered individuals in a system. Π represents the birth rate of the population and μ is the death rate. New infections are created at rate λ, which is a function of both Exposed and Infected individuals. We let $\lambda = \beta I + \xi E$ where β is the infectiousness of Infected individuals and ξ is the infectiousness of Exposed individuals. New infections have an incubation rate of ϵ and recovery rate of γ. We assume in the Simple model that Exposed individuals are less infectious than Infected

individuals, and that the population is stable, with $\Pi = \mu$. A diagram of the model is shown in Figure 1.

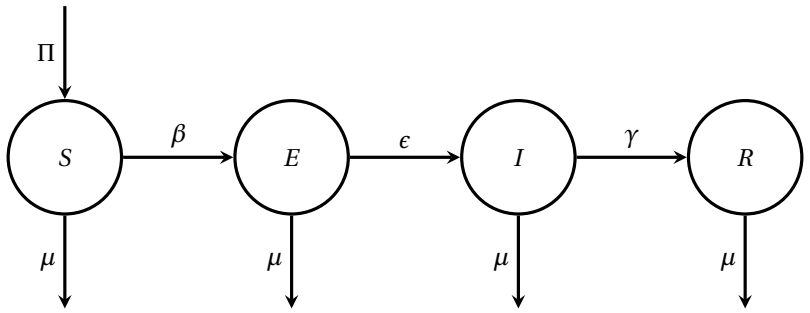

Figure 1: Diagram for Simple SEIR Model

2.2 Adult-Child Compartment Model

Our study expands on the Simple model from 2.1 by subdividing the compartments into age-specific compartments. Here, we let S_C, E_C, I_C, and R_C represent the number of children in each stage of disease progression at a given time t. Similarly, we let S_A, E_A, I_A, and R_A represent the number of adults in each stage at time t.

The SEIR model takes the following form with the subsequent parameter definitions:

$$\begin{aligned}
\frac{dS_C}{dt} &= \Pi N_A - fS_C - \lambda_C S_C \\
\frac{dS_A}{dt} &= fS_C - \lambda_A S_A - \mu S_A \\
\frac{dE_C}{dt} &= \lambda_C S_C - \epsilon_C E_C - fE_C \\
\frac{dE_A}{dt} &= \lambda_A S_A + fE_C - \epsilon_A E_A - \mu E_A \\
\frac{dI_C}{dt} &= \epsilon_C E_C - fI_C - \gamma_C I_C \\
\frac{dI_A}{dt} &= fI_C + \epsilon_A E_A - \gamma_A I_A - \mu I_A \\
\frac{dR_C}{dt} &= \gamma_C I_C - fR_C \\
\frac{dR_A}{dt} &= \gamma_A I_A + fR_C - \mu R_A
\end{aligned} \tag{5}$$

Table 1: Parameter Definitions for SEIR Model

Parameter	Definition
Π	The natural birth rate for the population.

μ	The natural death rate for the population. We assume that the birth rate is equal to the death rate. We assume the average age of an adult is 40 and set μ to be $\frac{1}{365(40)}$
f	The rate at which children become adults. We consider individuals aged 10 and under to be children, with the assumption that the average age of children 10 and under is 5. We then calculate f to be $\frac{1}{365(5)}$.
ϵ	The rate at which exposed individuals become infected, or the incubation rate. We assume that individuals have a reduced or nonexistent infectiousness at this stage.
γ	The rate at which infected individuals recover. We differentiate between γ_C and γ_A on the assumption that children recover at a faster rate than adults.
λ	The force of infection, or the rate at which susceptible individuals become exposed.

As in the Simple SEIR model, because our model assumes that Exposed individuals can be both non-infectious and slightly infectious, λ is expressed as a function of Infected and Exposed individuals. λ_c is a function of the number of Infected and Exposed individuals in the Child and Adult compartments, as well as the number of contacts made between groups and the likelihood an individual in one group will successfully contract a disease upon contact with an Infected or Exposed individual from another group. We assume that children's contacts are frequency dependent, while adults' are density dependent. This is because we assume children come into contact with others at a fixed rate while adults come into contact with others at a rate that is proportional to changes in the population N. λ_c is the summation of the 4 groups Susceptible Children may come into contact with, each of which is multiplied by the transmission rates β for Infected individuals and ξ for Exposed, and scaled by the population size N to account for frequency dependency. This suggests that the rate of exposure to disease is a function of both population dynamics (i.e. the number of Infected and Exposed individuals) and of the particular disease transmission rates. We allow β and ξ to vary according to 4 different rates of transmission: Adult-Adult, Adult-Child, Child-Adult, and Child-Child. This is based on the assumption that transmission rate is not constant across age groups [4].

Table 2: Parameter Definitions for SEIR Model: Child Compartment

Parameter	Definition
λ_C	The rate at which Susceptible Children become exposed. $$\lambda_C(I_C, I_A, E_C, E_A) = \beta_{CC}\frac{I_C}{N_C} + \beta_{AC}\frac{I_A}{N_A} + \xi_{CC}\frac{E_C}{N_C} + \xi_{AC}\frac{E_A}{N_A} \quad (6)$$
β_{CC}	The rate at which Infected Children transmit disease to Susceptible Children.

Parameter	Definition
β_{AC}	The rate at which Infected Adults transmit disease to Susceptible Children.
ξ_{CC}	The rate at which Exposed Children transmit disease to Susceptible Children.
ξ_{AC}	The rate at which Exposed Adults transmit disease to Susceptible Children.

Similarly, we define λ_A to be a function of the number of Exposed and Infected individuals in the Children and Adult groups respectively. Since we define adults' contact rate to be density dependent, the number of Infected and Exposed individuals is impacted by changes in the total population, and we do not scale λ_A by N.

Table 3: Parameter Definitions for SEIR Model: Adult Compartment

Parameter	Definition
λ_A	The rate at which Susceptible Adults become exposed. $$\lambda_A(I_C, I_A, E_C, E_A) = \beta_{AA} I_A + \beta_{CA} I_C + \xi_{AA} E_A + \xi_{CA} E_C \quad (7)$$
β_{AA}	The rate at which Infected Adults transmit disease to Susceptible Adults.
β_{CA}	The rate at which Infected Children transmit disease to Susceptible Adults.
ξ_{AA}	The rate at which Exposed Adults transmit disease to Susceptible Adults.
ξ_{CA}	The rate at which Exposed Children transmit disease to Susceptible Adults.

3 R_0 Calculation

3.1 R_0 Simple SEIR Model

The value for R_0 in the simple model is obtained by using the Next Generation Matrix method. We begin by identifying $\mathscr{F}(\mathbf{x})$, the terms in each compartment that create new infections, and $\mathscr{V}(\mathbf{x})$, the movement between compartments.

$$\mathscr{F} = \begin{pmatrix} \beta SI + \xi SE \\ \\ 0 \end{pmatrix}$$

$$\mathscr{V} = \begin{pmatrix} \epsilon E + \mu E \\ -\epsilon E + \gamma I + \mu I \end{pmatrix}$$

We then linearize and evaluate $\mathscr{F}$ and $\mathscr{V}$ at the disease-free equilibrium, where $S_0 \approx N$. We write denote S_0 as our initial population condition N.

$$F = \begin{bmatrix} \xi S & \beta S \\ 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} \epsilon + \mu & 0 \\ -\epsilon & \gamma + \mu \end{bmatrix}$$

R_0 is the largest eigenvector of the matrix FV^{-1}.

$$FV^{-1} = \begin{bmatrix} \dfrac{\beta\epsilon S}{(\gamma+\mu)(\epsilon+\mu)} + \dfrac{\xi S}{\epsilon+\mu} & \dfrac{\beta S}{\gamma+\mu} \\ 0 & 0 \end{bmatrix}$$

$$R_0 = \frac{\beta\epsilon S}{(\gamma+\mu)(\epsilon+\mu)} + \frac{\xi S}{\epsilon+\mu}$$

3.2 R_0 Adult Child SEIR Model

The value for R_0 in this model is again obtained by using the Next Generation Matrix method. We find the respective input and output vectors for the infected compartments E_C, E_A, I_C, and I_A.

$$\mathscr{F} = \begin{pmatrix} S_C(\beta_{cc}\dfrac{I_c}{N_c} + \beta_{AC}\dfrac{I_A}{N_A} + \xi_{CC}\dfrac{E_c}{N_c} + \xi_{AC}\dfrac{E_A}{N_A}) \\ S_A(\beta_{AA}I_A + \beta_{CA}I_C + \xi_{AA}E_A + \xi_{CA}E_C) \\ 0 \\ 0 \end{pmatrix}$$

$$\mathscr{V} = \begin{pmatrix} \epsilon_C E_C + f E_C \\ \epsilon_A E_A - f E_C + \mu E_A \\ -\epsilon_C E_C f I_C + \gamma_C I_C \\ -f I_C - \epsilon_A E_A + \gamma_A I_A + \mu I_A \end{pmatrix}$$

We linearize both matrices with respect to the disease free equilibrium.

$$F = \frac{\partial \mathscr{F}}{\partial E_C, E_A, I_C, I_A} = \begin{bmatrix} \dfrac{\xi_{CC}S_{C0}}{N_C} & \dfrac{\xi_{AC}S_{C0}}{N_A} & \dfrac{\beta_{CC}S_{C0}}{N_C} & \dfrac{\beta_{AC}S_{C0}}{N_A} \\ \xi_{CA}S_{A0} & \xi_{AA}S_{A0} & \beta_{CA}S_{A0} & \beta_{AA}S_{A0} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V = \frac{\partial \mathscr{V}}{\partial E_C, E_A, I_C, I_A} = \begin{bmatrix} \epsilon_C + f & 0 & 0 & 0 \\ -f & \epsilon_A + \mu & 0 & 0 \\ -\epsilon_C & 0 & f + \gamma_C & 0 \\ 0 & -\epsilon_A & -f & \gamma_A + \mu \end{bmatrix}$$

Here F can be interpreted as the rate at which new secondary infectious are introduced by Exposed Children, Exposed Adults, Infected Children, and Infected Adults in a given period of time. $F(1, j)$ is the rate at which new infections occur in children. Similarly, $F(2, j)$ is the rate

at which new infections occur in adults. For example, $F(1,1)$ is the rate of new infections in children by Exposed Children per unit time.

Table 4: Biological Interpretation of F:

Value	Interpretation
$F(1,1)$	Number of new infections created in children by Exposed Children per time spent while infectious.
$F(1,2)$	Number of new infections created in children by Exposed Adults per time spent while infectious.
$F(1,3)$	Number of new infections created in children by Infected Children per time spent while infectious.
$F(1,4)$	Number of new infections created in children by Infected Adults per time spent while infectious.
$F(2,1)$	Number of new infections created in adults by Exposed Children per time spent while infectious.
$F(2,2)$	Number of new infections created in adults by Exposed Adults per time spent while infectious.
$F(2,3)$	Number of new infections created in adults by Infected Children per time spent while infectious.
$F(2,4)$	Number of new infections created in adults by Infected Adults per time spent while infectious.

Because V is the average rate at which individuals transition between compartments, $W = V^{-1}$ is the expected time spent in each compartment.

Table 5: Biological Interpretation of W:

Value	Interpretation
$W(1,1)$	Expected time an individual spends in E_C given that they are introduced by E_C.
$W(2,1)$	Expected time an individual spends in E_A given that they are introduced by E_C.
$W(3,1)$	Expected time an individual spends in I_C given that they are introduced by E_C.
$W(4,1)$	Expected time an individual spends in I_A given that they are introduced by E_C.
$W(2,2)$	Expected time an individual spends in E_A given that they are introduced by E_A.
$W(4,2)$	Expected time an individual spends in I_A given that they are introduced by E_A.
$W(3,3)$	Expected time an individual spends in I_C given that they are introduced by I_C.
$W(4,3)$	Expected time an individual spends in I_A given that they are introduced by I_C.
$W(4,4)$	Expected time an individual spends in I_A given that they are introduced by I_A.

We use Matlab to calculate the inverse matrix of V, V^{-1}, and identify its eigenvectors. Then

$$W = \begin{bmatrix} \frac{1}{\epsilon_C + f} & 0 & 0 & 0 \\ \frac{f}{(\epsilon_C + f)(\epsilon_A + \mu)} & \frac{1}{\epsilon_A + \mu} & 0 & 0 \\ \frac{\epsilon_C}{(\epsilon_C + f)(f + \gamma_C)} & 0 & \frac{1}{f + \gamma_C} & 0 \\ \frac{\epsilon_A f^2 + \epsilon_A \epsilon_C f + \epsilon_A f \gamma_C + \epsilon_C f \mu}{(\epsilon_C + f)(f + \gamma_C)(\epsilon_A + \mu)(\gamma_A + \mu)} & \frac{\epsilon_A}{(\epsilon_A + \mu)(\gamma_A + \mu)} & \frac{f}{(f + \gamma_C)(\gamma_A + \mu)} & \frac{1}{\gamma_A + \mu} \end{bmatrix}$$

We then compute FV^{-1}.

$$FV^{-1} = \begin{bmatrix} A & B & C & D \\ E & F & G & H \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where the coefficients are as follows:

$$A = \frac{S_{C0}\xi_{CC}}{N_C(\epsilon_C + f)} + \frac{S_{C0}\beta_{CC}\epsilon_C}{N_C(\epsilon_C + f)(f + \gamma_C)} + \frac{S_{C0}f\xi_{AC}}{N_A(\epsilon_C + f)(\epsilon_A + \mu)} + \frac{S_{C0}\beta_{AC}(\epsilon_A f^2 + \epsilon_A \epsilon_C f + \epsilon_A f \gamma_C + \epsilon_C f \mu)}{N_A(\epsilon_C + f)(f + \gamma_C)(\epsilon_A + \mu)(\gamma_A + \mu)}$$

$$B = \frac{S_{C0}\xi_{AC}}{N_A(\epsilon_A + \mu)} + \frac{S_{C0}\beta_{AC}\epsilon_A}{N_A(\epsilon_A + \mu)(\gamma_A + \mu)}$$

$$C = \frac{S_{C0}\beta_{CC}}{N_C(f + \gamma_C)} + \frac{S_{C0}\beta_{AC}f}{N_A(f + \gamma_C)(\gamma_A + \mu)}$$

$$D = \frac{S_{C0}\beta_{AC}}{N_A(\gamma_A + \mu)}$$

$$E = \frac{S_{A0}\xi_{CA}}{\epsilon_C + f} + \frac{S_{A0}\beta_{CA}\epsilon_C}{(\epsilon_C + f)(f + \gamma_C)} + \frac{S_{A0}f\xi_{AA}}{(\epsilon_C + f)(\epsilon_A + \mu)} + \frac{S_{A0}\beta_{AA}(\epsilon_A f^2 + \epsilon_A \epsilon_C f + \epsilon_A f \gamma_C + \epsilon_C f \mu)}{(\epsilon_C + f)(f + \gamma_C)(\epsilon_A + \mu)(\gamma_A + \mu)}$$

$$F = \frac{S_{A0}\xi_{AA}}{\epsilon_A + \mu} + \frac{S_{A0}\beta_{AA}\epsilon_A}{(\epsilon_A + \mu)(\gamma_A + \mu)}$$

$$G = \frac{S_{A0}\beta_{CA}}{f + \gamma_C} + \frac{S_{A0}\beta_{AA}f}{(f + \gamma_C)(\gamma_A + \mu)}$$

$$H = \frac{S_{A0}\beta_{AA}}{\gamma_A + \mu}$$

Note here that $F(1,1)V^{-1}(1,1)$ is the number of new Exposed Children caused by one individual introduced in E_C during their time in E_C. Additionally, $F(1,2)V^{-1}(2,1)$ is the number of new Exposed Children caused by one individual introduced in E_C during their time in E_A. It follows that $FV^{-1}(1,1)$ is the number of new Exposed Children by an Exposed Child over their infectious lifetime.

Table 6: Biological Interpretation of the coefficients of FV^{-1}:

Coefficient	Interpretation
A	The number of new infections in children created by an Exposed Child over their infectious lifetime.
B	The number of new infections in children created by an Exposed Adult over their infectious lifetime.
C	The number of new infections in children created by an Infected Child over their infectious lifetime.
D	The number of new infections in children created by an Infected Adult over their infectious lifetime.
E	The number of new infections in adults created by an Exposed Child over their infectious lifetime.
F	The number of new infections in adults created by an Exposed Adult over their infectious lifetime.
G	The number of new infections in adults created by an Infected Child over their infectious lifetime.
H	The number of new infections in adults created by an Infected Adult over their infectious lifetime.

The eigenvalues of this matrix are then:

$$\frac{A}{2}+\frac{F}{2}-\frac{(A^2-2AF+F^2+4BE)^{1/2}}{2} \text{ and } \frac{A}{2}+\frac{F}{2}+\frac{(A^2-2AF+F^2+4BE)^{1/2}}{2}$$

Since we know that all of the parameters are > 0, the maximum eigenvalue is

$$\frac{A}{2}-\frac{F}{2}+\frac{(A^2-2AF+F^2+4BE)^{1/2}}{2}$$

Then

$$R_0=\frac{A}{2}+\frac{F}{2}+\frac{(A^2-2AF+F^2+4BE)^{1/2}}{2} \tag{8}$$

Given our coefficient interpretations above, R_0 can be interpreted as the summation of the number of new infections created in children by individuals who start in the Exposed Children compartment, and the number of new infections created in adults by individuals who start in the Exposed Adult compartment. Since Exposed Children and Exposed Adults move through the Infected compartments over their infectious lifetimes, R_0 accounts for infections created by both Exposed and Infected individuals.

4 Structural Identifiability

In this section we introduce structural identifiability and compute the structural identifiability of our SEIR Child-Adult epidemic model. We introduce various methods for computing structural identifiability and outline the algorithms behind each method. We conclude with a brief dicussion about relevant output vectors that produce structurally identifiable results.

The motivation behind structural identifiability lies in recovering unique parameter values from perfect data or continuous, noise-free data. This is in contrast with practical identifiability, which tests whether the model parameters can be reasonably recovered from imperfect or incomplete data. A parameter is said to be **globally structurally identifiable** if the parameter can be uniquely recovered from the given output equation. Similarly, a parameter is said to be **locally structurally identifiable** if there are a finite number of solutions for the parameter that can be recovered from the output equation. On the contrary, if there are an infinite amount of values for one or more parameters our model is considered to be structurally non-identifiable. Often models that are structurally identifiable are not practically identifiable, but structural identifiability is a necessary condition for practical identifiability. As such, it is often necessary to evaluate the structural identifiability of our model parameters before computing practical identifiability.

We begin by defining structural identifiability in terms of output input functions. We can represent our SEIR model in the form of [5] below,

$$\frac{dx}{dt} = f(x(t), p), x(0) = x_0.$$

Here $f(x(t), p)$ is the input function where $x(t)$ represents the state variables and p represents a specific parameter from our parameter set. We then define our output function as $g(x(t), p)$. Additionally, note that our output observations are given in discrete time intervals: $\{y_{i=1}^n\}$. Then we can write $\{y_{i=1}^n\}$ as the following equation:

$$y_i = g(x(t_i), \hat{p}) + E_i$$

where $\hat{p}$ represents the true value of each parameter in the parameter set, $g(x(t_i))$ is the output of our model at each time step, and E_i is the observation measurement error that represents the difference between the observed and expected values. In our calculations, we assume that E_i follows a normal distribution with mean 0 and variance σ^2. We define global and local structurally identifiability below.

Definition 4.1. A parameter P is **globally structurally identifiable** if for every q in the set of estimated parameters, the equation $g(x(t), p) = g(x(t), q)$ if and only if $p = q$.

Definition 4.2. Let $N(P)$ denote the neighborhood of the parameter p. The parameter p is **locally structurally identifiable** if for every p there exists an open neighborhood $N(P)$ such that for every $q \in N(P)$ where q is in the set of estimated parameters, the equation $g(x(t), p) = g(x(t), q)$ if and only if $p = q$.

Next we explore 3 methods of computing structural identifiability and the results garnered from each method.

4.1 DAISY

We first computed the structural identifiability of our model in DAISY (Differential Algebra for Identifiability of Systems). DAISY provides a certificate of correctness while other methods such as **SIAN** in Maple and **IdentifiabilityAnalysis** in Mathematica are only correct with high probability of accuracy. DAISY works by taking in a vector of our variables, a vector of our unknown parameters, the number of state variables, the number inputs, the number of outputs, and our model equations.

DAISY uses a differential algebra approach to determine whether or not a model is structurally identifiable. The program first ranks the parameters, inputs, outputs, variables and their derivatives. A typical ranking places the input and output variables at the lower rank, while the highest rank is reserved for the state variables of the system. From there, DAISY calculates the characteristic set of the model equations, or a minimal set of differentiable polynomials. The characteristic polynomials, one through the number of outputs, are not in terms of the state variables and therefore show the relationship between the input and output variables.

Next, Ritts algorithm uses repeated Lie derivatives to eliminate the non-observed state variables from the system of equations and then uses this to find the input-output relation of the system. Daisy then normalizes the polynomials to be monotonic in order to fix the values of the coefficients of the polynomials. The coefficients of these polynomials form a function of our unknown parameters.

Finally, DAISY computes an exhaustive summary, or a map from the parameter space $\mathbb{P}$ to $\mathbb{R}^{\nu}$ where $j = 1, ..., \nu_i$ and j is an index running over the monomial indices of the set of differential polynomials. This serves to linearly parameterize the input-output relation. We check identifiability by checking that this map $p : \mathbb{P} \to \mathbb{R}^{\nu}$ is injective. This computation uses the Buchberger algorithm that calculates the Groebner basis. The structure of the Groebner basis allows us to determine the structural identifiability of a system of differential equations.

While DAISY provides a certificate of accuracy, it suffers from expression swell as models get more computationally complex. Because of this, attempts to compute the structural identifiability of our SEIR Child-Adult failed to produce any results.

4.2 SIAN

Next we used **SIAN** (Structural Identifiability ANalyzer), a package in Maple to compute the structural identifiability of our SEIR Child-Adult model. SIAN combines a differential algebra approach and a Taylor series approach. Structural identifiability in SIAN can be analyzed as a map between the parameter values and initial conditions to the output data. By reducing the output functions to their truncated Taylor series, we can effectively reduce the dimension of this map to a finite dimension.

SIAN first constructs the maximal polynomial system: a system of algebraic equations that define the input-output equations, the initial conditions, and the parameter set. Next SIAN truncates the polynomial system of equations based on the Jacobian condition, finding the unique solution of the triangular system. Then, random specializations of integer values are computed for the parameters and the initial conditions of the truncated system. Finally, SIAN outputs the set of locally and globally structurally identifiable parameters.

It is important to note that SIAN does not take in initial conditions, and instead takes in Sc(0), Sa(0), Ec(0), Ea(0), Ic(0), Ia(0), Rc(0), Ra(0) as additional parameters.

4.3 IdentifiabilityAnalysis

In this section we present structural identifiability results from Wolfram Mathematica, using a package called **IdentifiabilityAnalysis**. IdentifiabilityAnalysis takes in initial conditions either as fixed values or as parameters to be estimated. By eliminating the need to take repeated Lie derivatives in computing the Jacobian, IdentifiabilityAnalysis allows for the computation of structural identifiability for much larger systems. Additionally, IdentifiabilityAnalysis further reduces the computational complexity by computing all calculations modulo a large prime number.

IdentifiabilityAnalysis also uses a differential algebra approach. In summary, the package works by computing a non-linear algebraic system, whose rank of the Jacobian determines the identifiability of the system. IdentifiabilityAnalysis reduces computational complexity by directly computing the entries of the Jacobian matrix via computing the power series expansion of the partial derivatives of the input variables with respect to the parameters and the initial conditions. Additionally, the package further simplifies the structural identifiability computation by performing all calculations modulo a large prime number. There is a degree of inaccuracy that can stem from these simplifications, but the probability of a rank error in the computation of the rank of the Jacobian can be bounded from above and decreases as value of the prime number modulo increases. The steps of the algorithm are outlined below:

1. Assign random integers as the values for the parameters and initial conditions.
2. The input of the system is specialized to a truncated random integer coefficient power series.
3. Compute the truncated power series solutions of the state variables, the partial derivative of the input variables with respect to the initial conditions, and the partial derivatives of the input variables with respect to the parameter set.
4. Utilize the truncated power series solutions and substitute them into the partial derivatives in step 3. This produces the power series representations of the output sensitivity derivatives.
5. Identify the coefficients of the truncated power series of the output derivatives with the coefficients of a general Taylor expansion of the output sensitivity derivatives. These give the entries of the Jacobian matrix.
6. Calculate the rank of the Jacobian. If the Jacobian has a full rank the system is said to be structurally identifiable. If the matrix is rank deficient, then the non-identifiable parameters are found by determining which columns of the Jacobian whose removal does not change the rank.

We began by computing structural identifiability for a basic SEIR model defined below. Then we can compare the results of the SEIR Child-Adult model with the basic SEIR model to see whether the increased complexity of our adapted model leads to a structurally non-identifiable model. We define our model as follows:

$$\frac{dS}{dt} = -\beta\frac{SI}{N},$$
$$\frac{dE}{dt} = \beta\frac{SI}{N} - \eta E,$$

$$\frac{dI}{dt} = \eta E - \alpha I,$$
$$\frac{dR}{dt} = \alpha I$$

where β is the transmission rate, $\frac{1}{\eta}$ is the latent period where an individual is infected but not yet infectious, and α is the recovery rate. Additionally in this model $S(0) = S_0, E(0) = E_0, I(0) = 0$, and $R(0) = 0$. The parameters to be estimated are: $\{S(0), I(0), E(0), R(0), \alpha, \beta, \eta, N\}$ The structural identifiability results were computed in SIAN and are shown in the table below.

output vector	fixed parameters	locally identifiable parameters	globally identifiable parameters
$I(t), N$	f, μ	N, β, I(0), α, η, S(0), E(0)	N, β, I(0)
$I(t)$	f, μ	I(0), α, η, S(0), E(0)	I(0)
R	f, μ	α, η, E(0), In(0), S(0), R(0)	S(0), R(0)

Table 7: Structural Identifiability of SEIR Parameters: Base SEIR Case

4.4 Non-Infectious E SEIR Model

In this section we present the results of structural identifiability analysis for the non-infectious E case where $\xi = 0$.

output vector	fixed parameters	locally identifiable parameters	globally identifiable parameters
I_C, I_A	f, μ	All	All
I_C	f, μ	All	NA
I_A	f, μ	All	NA
$I_C + I_A$	f, μ	All	NA
R_C, R_A	f, μ	All	All
$R_C + R_A$	f, μ	All	NA
C_C, C_A	f, μ	All	NA
$C_C + C_A$	f, μ	All but C_{A0} and C_{C0}	NA
C_C, C_A, R_C, R_A	f, μ	All	All
$C_C + R_C, C_A + R_A$	f, μ	All	NA
$C_C + C_A, R_C + R_A$	f, μ	All but C_{A0} and C_{C0}	NA
I_C, I_A	f, μ, β_{CA}	All	All
I_C, I_A	$f, \mu, \beta_{CC}, \beta_{CA}$	All	All
I_C, I_A	$f, \mu, \gamma_C, \gamma_A, \epsilon_C, \epsilon_A$	All	All
I_C, I_A	f, μ, β-dependencies	All	All

Table 8: Structural Identifiability of SEIR Parameters: Non-Infectious E: MAPLE Results

output vector	fixed parameters	identifiable parameters	non-identifiable
I_C, I_A	f, μ, β_{CC}	All	None
I_C, I_A	$f, \mu, \beta_{CC}, \beta_{CA}$	All	None
I_C, I_A	$f, \mu, \gamma_C, \gamma_A, \epsilon_C, \epsilon_A$	All	None

Table 9: Structural Identifiability of SEIR Parameters: Mathematica

In the structural results above, both SIAN and Mathematica take in initial conditions as estimated parameters. "NA" indicates a kernel error in SIAN where computations of global structural identifiability failed. Additionally, note that Mathematica only provides information on

local structural identifiability, while SIAN outputs both local and global structural identifiability. We observe that R_C, R_A output vectors produce a full globally structurally identifiable set of parameters.

4.5 Infectious E SEIR Model

The table below contains the structural identifiability results for the original parameter set: $\{\beta_{AC}, \beta_{CA}, \beta_{CC}, \beta_{AA}, \gamma_C, \gamma_A, \epsilon_C, \epsilon_A\}, \xi_{CC}, \xi_{CA}, \xi_{AC}, \xi_{AA}\}$ in addition to the initial condition set described above.

output vector	fixed parameters	locally identifiable parameters	globally identifiable parameters
I_C, I_A	f, μ	All	NA
I_C	f, μ	All	NA
I_A	f, μ	All	NA
$I_C + I_A$	f, μ	All	NA
R_C, R_A	f, μ	All	All
$R_C + R_A$	f, μ	All	NA
C_C, C_A	f, μ	All	NA
$C_C + C_A$	f, μ	All	NA
C_C, C_A, R_C, R_A	f, μ	All	All
$C_C + R_C, C_A + R_A$	f, μ	All	NA
$C_C + C_A, R_C + R_A$	f, μ	All but $C_c(0)$ and $C_a(0)$	NA
I_C, I_A	$f, \mu, \gamma_C, \gamma_A, \epsilon_C, \epsilon_A$	All	All
I_C, I_A	f, μ, β-dependencies	All	All
I_C, I_A	f, μ, β and ξ-dependencies	All	All

Table 10: Structural Identifiability of SEIR Parameters: Infectious E: MAPLE Results

output vector	fixed parameters	identifiable parameters	non-identifiable
I_C, I_A	f, μ	All	None
$I_C + I_A$	f, μ	All	None
R_C, R_A	f, μ	All	None
$R_C + R_A$	f, μ	All	None
C_C, C_A	f, μ	None	All
$C_C + R_C, C_A + R_A$	f, μ	All	None
$C_C + C_A, R_C + R_A$	f, μ	All	None

Table 11: Structural Identifiability of SEIR Parameters: Infectious E: Mathematica

Observe in the tables above that the output vectors R_C, R_A produce a full set of globally identifiable parameters and C_C, C_A produce no identifiable parameters.

5 Practical Identifiability

5.1 Monte Carlo Simulation

In this paper we use **Monte Carlo Simulations** as shown in Algorithm 1, to determine the practical identifiability of our SEIR Child-Adult model. We generate M=100 prevalence data sets for both Infected Children and Infected Adult compartments across a time span of T days. This

is repeated for six different noise levels: $\sigma = 0, 1, 5, 10, 20, 30$. Next we fit our model to each of the M data sets, and estimate our parameters using the optimization packages Fmincon and Fminsearchbnd in Matlab. We then compute the average relative error (ARE) across the 100 estimated parameter values. A parameter set p is practically identifiable if $\text{ARE}(p_\sigma^{(k)}) \leq \sigma$ for all σ.

Algorithm 1 Monte Carlo Simulation

$t_j = 1, 2, \ldots, T$
Let p denote true parameter value.

for $\sigma = 0\%, 1\%, 5\%, 10\%, 20\%, 30\%$ **do**
  for $i = 1, 2, \ldots, 100$ **do**
    $\hat{I}_{iC}(t_j, p) = I_C(t_j, p) + \alpha_C$
    $\hat{I}_{iA}(t_j, p) = I_A(t_j, p) + \alpha_A$
    α_C follows $N(I_C(t_j, p), I_C(t_j, p)\sigma)$
    α_A follows $N(I_A(t_j, p), I_A(t_j, p)\sigma)$

    Optimize error function:
    error $= \sum(\frac{1}{I_{iC}(t_j)^2}(I_{iC}(t_j) - \hat{I}_{iC}(t_j))^2) + \sum(\frac{1}{I_{iA}(t_j)^2}(I_{iA}(t_j)$
    $-\hat{I}_{iA}(t_j))^2)$
    $\hat{p}_i(\sigma)$ is the optimal set of parameter values for the ith generated
    data set.
  end for

  Compute ARE
  Let k be the number of estimated parameters
  for $n = 1, 2, \ldots, k$ **do**
    $\text{ARE}(p_\sigma^{(k)}) = 100\% \times \frac{1}{100} \sum_{i=1}^{100} \frac{p^{(k)} - \hat{p}_{\sigma i}^{(k)}}{p^{(k)}}$
  end for
end for

Note that a different estimated parameter may impact the time of peak infection. Therefore, the outputs, our Infected compartments, go to zero at a different pace for the infectious and non-infectious E cases as shown in Figure 2. To fix this, we run our MATLAB code for a shorter length of time in order to eliminate the days at which the output returns values of 0 since it makes the weight in our least squares error function undefined.

$$\text{weight}_C = \frac{1}{I_C^2}, \text{weight}_A = \frac{1}{I_A^2}$$

Practical identifiability changes when only given data before the peak infection, so it is important that we always test practical identifiability on data that goes beyond the peak time of infection [6]. When adjusting the time span, we verify that the data includes the time of peak infection.

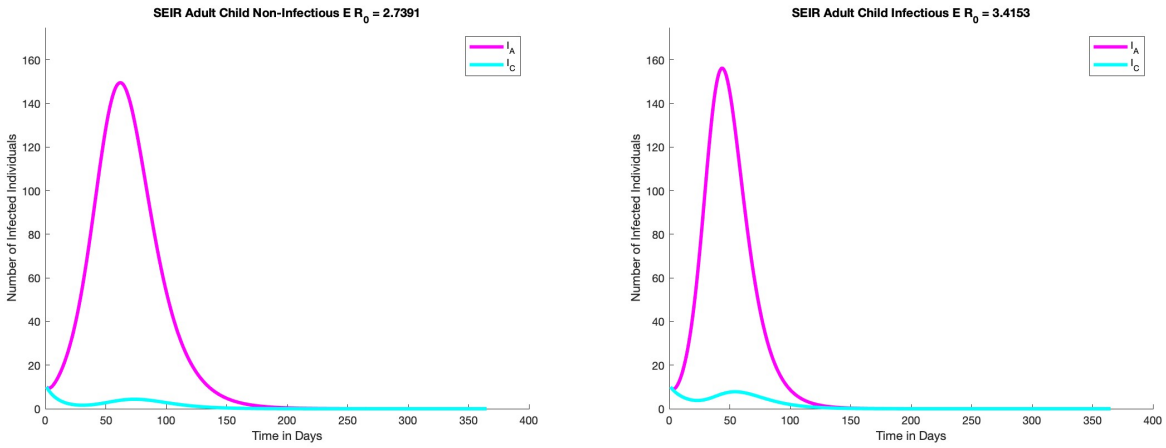

Figure 2: Differences in dynamics between infectious and non-infectious E cases. For the non-infectious case, time of peak infection for children is at day 71 and time of peak infection for adults is at day 62. For the infectious E case, time of peak infection for children is at day 53 and time of peak infection for adults is at day 43

For some cases, Fmincon computes a singular matrix due to the initial guess being equivalent to our true parameter values. We instead use Fminsearchbnd to avoid inaccurate results in these instances.

We set our initial conditions for the adult child model as shown in Table 12. Our initial conditions are based off literature reviews for international COVID-19 data. The percentage of individuals under 15 is 25% of the world population [7]. Lowering this percentage to account for the lower proportion of individuals under 10 years old than that of 15 years would not bring R_0 under 1 as shown in Figure 3 nor would it cause a large change in the value of R_0. Therefore, we define 25% of the population as children under 10 for our simulations.

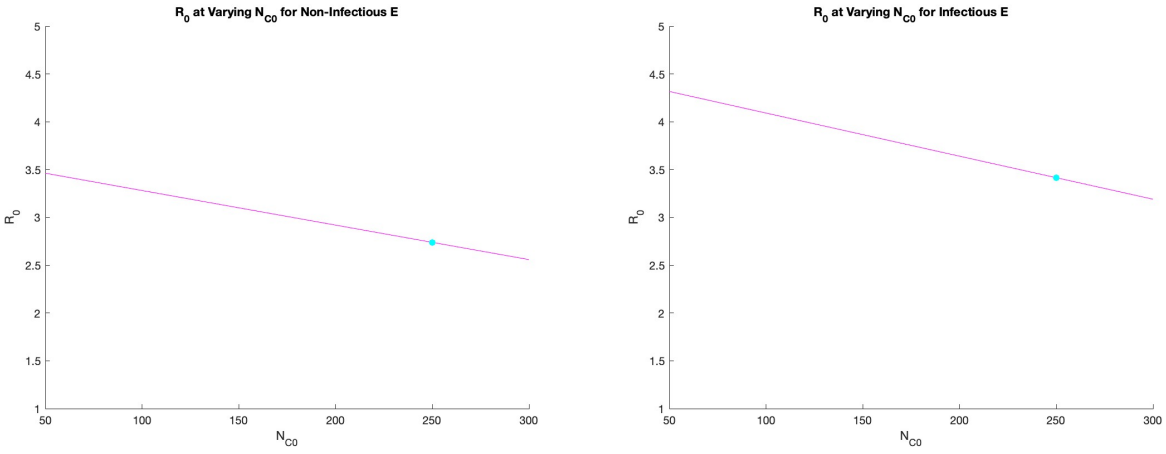

Figure 3: R_0 at proportion of children varying from 5% to 30%

Table 12: Initial Conditions

Variable	Initial Value	Variable	Initial Value

N_C	250	N_A	750
S_C	240	S_A	740
E_C	0	E_A	0
I_C	10	I_A	10
R_C	0	R_A	0

5.1.1 Non-Infectious E SEIR Model

Since practical identifiability is a local behavior, we must fix our true parameter values as given in Table 13. We estimate our parameter values based on COVID-19 data and a reasonable R_0 to approximate a model of the pandemic.

Table 13: Fixed and Estimated Parameter Values

Parameter	Value	Source	Parameter	Value	Source
μ	0.00008	[8] [9]	f	0.0005	[8] [10]
β_{CC}	0.01	[8] [11]	β_{AA}	0.00027	Estimated
β_{CA}	0.00005	[11]	β_{AC}	0.01	[8]
γ_C	0.1	[8]	γ_A	0.074	[12]
ϵ_C	0.3	[8]	ϵ_A	0.2	[13]

We begin by running the Monte Carlo Simulation where we estimate all disease specific parameters to get Table 14.

Table 14: Practical Identifiability for β_{AA}, β_{CC}, β_{AC}, β_{CA},γ_A,γ_C, ϵ_C,and ϵ_A where $T = 365$ using fmincon

σ	$\boldsymbol{\beta_{AA}}$	β_{CC}	$\boldsymbol{\beta_{AC}}$	β_{CA}
0	0.0000	0.0000	0.0000	0.0000
1	0.0141	0.0771	0.0127	0.0860
5	0.1368	7.4768	0.7954	2.2646
10	0.4139	39.1975	1.9703	9.3628
20	1.8227	97.3919	3.8395	39.5859
30	5.0465	168.6005	6.2050	80.4438
σ	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000
1	0.0159	0.0169	0.0130	0.0401
5	0.2317	0.9040	0.4418	0.4982
10	0.8576	4.1867	1.7974	1.6478
20	2.9308	10.0851	5.1290	6.5954
30	6.0575	15.9486	9.2823	15.4116

In Table 14, the transmission rates from children, β_{CC}, and β_{CA}, appear to not be practically identifiable while the remainder of the parameters are. We next fix the least practically identifiable of the two, β_{CC}, to see if the other, β_{CA}, becomes practically identifiable in Table 15.

Table 15: Practical Identifiability for β_{AA}, β_{AC}, β_{CA},γ_C,γ_A, ϵ_C,and ϵ_A where $T = 365$ using fmincon

σ	$\boldsymbol{\beta_{AA}}$	$\boldsymbol{\beta_{AC}}$	β_{CA}	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0096	0.0107	0.0130	0.0124	0.0114	0.0116	0.0225
5	0.1145	0.7071	2.0670	0.2121	0.3166	0.3324	0.4549
10	0.4106	1.6597	9.0988	0.8735	0.7145	0.7901	1.6547
20	2.1995	3.8992	46.8843	2.7272	1.8460	3.0890	8.0240
30	6.0667	6.7388	90.5261	5.8050	3.4994	9.8910	17.3581

In Table 15, β_{CA} is still not practically identifiable but the other six parameters are. This indicates that fixing β_{CA} will allow us to obtain a full set of practically identifiable parameters. We confirm this in Table 16.

Table 16: Practical Identifiability for β_{AA}, β_{AC}, γ_A, γ_C, ϵ_C, and ϵ_A where $T = 365$ using fmincon

σ	$\boldsymbol{\beta_{AA}}$	$\boldsymbol{\beta_{AC}}$	$\boldsymbol{\gamma_A}$
0	0.0000	0.0000	0.0000
1	0.0117	0.0165	0.0222
5	0.1344	0.5610	0.1861
10	0.4622	1.8485	0.7366
20	1.5650	3.9218	2.9964
30	4.2548	6.1123	6.0440

σ	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000
1	0.0138	0.0156	0.0266
5	0.2609	0.2526	0.4822
10	0.7965	0.7622	1.6507
20	1.8750	2.2417	5.1656
30	3.6714	10.6538	11.9416

Since the infection rates, β values, are the most difficult parameter values to pinpoint, we seek additional conclusions about the identifiability of these parameters for our model. To further investigate the practical identifiablity of the β parameters, we fix the other four parameters to obtain the ARE values in Table 17.

Table 17: Practical Identifiability for β_{AA},β_{CC},β_{AC}, and β_{CA} where $T = 365$ using fmincon

σ	$\boldsymbol{\beta_{AA}}$	$\boldsymbol{\beta_{CC}}$	$\boldsymbol{\beta_{AC}}$	$\boldsymbol{\beta_{CA}}$
0	0.0000	0.0003	0.0001	0.0002

1	0.0769	0.3680	0.1086	0.1528
5	0.3878	1.9307	0.5698	0.7323
10	0.9186	4.4049	1.2515	1.7253
20	3.5974	14.1137	2.1176	7.9444
30	7.4914	23.0994	4.6589	17.1020

Fixing the rate at which exposed individuals become infected and the rates at which infected individuals recover reveals a full set of practically identifiable parameters at our fitted parameter values. The fact that we can estimate infection rates accurately with noisy data may be very helpful in future disease models.

To further investigate practically identifiable models for the non-infectious E case, we sought a way to estimate a singular β that is scaled appropriately for each adult child infection rate. Since the infection rate is the product of the probability of infection and the contact rate, we can compute these scalars given some knowledge on the disease-specific relationship between adults and children in these categories. We use COVID specific findings to generate our scalars. We then compute the ARE for a single β that determines all 4 unknown βs.

We set β equal to the transmission rate, β_{AA}, from infectious adults to susceptible adults. We assume that the contact rate from children to children is less than the contact rate from adults to adults is thus we multiply β by $\frac{1}{2}$ for β_{CC}. We also estimate the contact rate from children to adults to be a third of the contact rate from adults to adults, so we multiply β by $\frac{1}{3}$ for β_{CA}. Additionally, we assume the contact rate from adults to children to be about a third of the contact rate from adults to adults, so we multiply β by $\frac{1}{3}$ for β_{AC}. Since children are at most half as likely to contract COVID compared to adults, the probability of transmission to children will be half of the probability of transmission to adults thus we multiply β by $\frac{1}{2}$ for β_{CC} and β_{AC} [8]. The probability of transmission from children is $\frac{24515}{42739}$ of the probability of transmission from adults, so we multiply β by $\frac{24515}{42739}$ for β_{CC} and β_{CA} [11]. These relationships result in the dependencies shown in Table 18.

Table 18: β Dependencies

Parameter	Dependency	Parameter	Dependency
β_{AA}	β	$\frac{\beta_{CC}}{N_{C0}}$	$\frac{1}{4}\frac{24515}{42739}\,\beta$
β_{CA}	$\frac{1}{3} * \frac{24515}{42739}\,\beta$	$\frac{\beta_{AC}}{N_{A0}}$	$\frac{1}{6}\,\beta$

Table 19: Practical Identifiability for β γ_A,γ_C,ϵ_C, and ϵ_A where $T = 365$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0203	0.0242	0.0188	0.0142	0.0676
5	0.1681	0.2574	0.1733	0.5270	0.6985
10	0.3988	0.7500	0.4012	1.1210	1.6238
20	1.0902	2.7449	0.8450	3.3212	4.1651

30	2.6538	5.8913	1.7347	10.1056	9.1182

When we compute practical identifiability for a collective β, along with the ϵ and γ parameters, we obtain a full set of practically identifiable parameters as shown in 19. This provides an alternative model given different information about a disease.

5.1.2 Infectious E SEIR Model

The non-infectious E SEIR model is a special case of the infectious E model where all ξ values are fixed at 0. Here we present the practical identifiability of the infectious E SEIR adult child model. The fitted COVID specific parameters for the ξs are given by Table 20. The remaining parameters are set at the same values presented in Table 13.

Table 20: ξ Fixed and Estimated Parameter Values

Parameter	Value	Source	Parameter	Value	Source
ξ_{AA}	0.0002	[14]	ξ_{CC}	0.0067	[14] [8] [11]
ξ_{CA}	0.000033	[14] [11]	ξ_{AC}	0.0067	[14] [8]

We begin by running the Monte Carlo Simulation where we estimate all diseases specific parameter values which produces Table 21.

Table 21: Practical Identifiability for β_{AA}, β_{CC}, β_{AC}, β_{CA}, γ_C, γ_A, ϵ_C, ϵ_A, ξ_{CC}, ξ_{AA}, ξ_{CA}, and ξ_{AC} where $T = 250$ using fmincon

σ	$\boldsymbol{\beta_{AA}}$	β_{CC}	β_{AC}	β_{CA}	γ_C	$\boldsymbol{\gamma_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9412	7.9757	5.7989	19.7460	1.8837	0.1015
5	3.2467	26.4452	11.8255	41.7726	8.8579	0.3991
10	6.2684	44.1012	20.9476	44.0497	16.4534	0.8823
20	14.3341	72.9835	38.9088	51.7101	23.7573	2.5946
30	25.8113	92.9567	39.7756	76.2023	29.0818	6.2513
σ	ϵ_C	ϵ_A	ξ_{CC}	ξ_{AA}	ξ_{CA}	ξ_{AC}
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	8.2914	3.0401	329.1744	2.9627	1389.5	9.2826
5	30.7098	8.3567	428.4256	10.4277	4032.4	35.9406
10	79.6158	11.9891	929.4349	17.8005	7539.1	73.0132
20	117.4588	21.9284	2780.7	36.9233	11731	109.9530
30	142.8740	24.0662	3164.7	60.6729	15816	89.8377

Since the majority of the parameters are not practically identifiable, we move on to investigating the practical identifiability of the infection rates in Table 22. For the infectious E case this is all of the β and ξ values.

Table 22: Practical Identifiability for β_{AA}, β_{CC}, β_{AC}, β_{CA}, ξ_{CC}, ξ_{AA}, ξ_{CA}, and ξ_{AC} where $T = 250$ using fmincon

σ	β_{AA}	β_{CC}	$\boldsymbol{\beta_{AC}}$	β_{CA}
0	0.0000	0.0000	0.0000	0.0000
1	2.4440	2.7943	2.8675	13.6825
5	6.0076	7.1509	5.5580	30.8122
10	10.4381	16.9261	13.4978	44.4221
20	20.4924	24.5778	18.4858	78.2155
30	39.6615	32.2456	18.1156	156.6028
σ	ξ_{CC}	ξ_{AA}	ξ_{CA}	ξ_{AC}
0	0.0000	0.0000	0.0000	0.0000
1	114.7923	2.1610	1950.3	0.7762
5	217.7204	7.8126	5102.3	3.8035
10	490.2204	16.4853	8966.2	7.4770
20	629.6905	38.8814	16905	19.5418
30	553.0994	65.7712	28387	34.7203

This set only has one practically identifiable parameter β_{AC}, so we try to make further simplifications by only estimating a singular β and ξ. We assume the relationships between the child and adult contact rates and probability of infection are the same for exposed individuals. Therefore, we use the same scalars in Table 18 for the β values as we do for the corresponding ξ values as shown in Table 23.

Table 23: ξ Dependencies

Parameter	Dependency	Parameter	Dependency
ξ_{AA}	ξ	$\frac{\xi_{CC}}{N_{C0}}$	$\frac{1}{4}\frac{24515}{42739}\,\xi$
ξ_{CA}	$\frac{1}{3}\frac{24515}{42739}\,\xi$	$\frac{\xi_{AC}}{N_{A0}}$	$\frac{1}{6}\,\xi$

We run the Monte Carlo Simulation for the infectious E case with these dependencies to obtain Table 24.

Table 24: Practical Identifiability for β,γ_C,γ_A, ϵ_C, ϵ_A, ξ where $T = 250$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\gamma_A}$
0	0	0	0
1	0.1484	0.0543	0.0309
5	1.5585	0.3015	0.2195
10	3.8483	0.7723	0.5322
20	11.2767	2.4018	1.8939

30	23.9931	4.8543	3.7280

σ	ϵ_C	ϵ_A	ξ
0	0	0	0
1	0.1467	0.1954	0.2635
5	2.6161	1.6134	3.1677
10	5.8175	3.4819	7.6676
20	10.2561	7.2099	19.3825
30	16.7237	10.2737	37.5742

Since we still do not yield a full set of practically identifiable parameters, we make ξ dependant on beta. The scalars in Table 18 will still hold for the β parameters, but the ξ scalars in Table 23 will be adjusted to include the relationship between β and ξ. The rate at which exposed individuals infect susceptible individuals is $\frac{2}{3}$ the rate at which infected individuals infect susceptible individuals [14]. For that reason, we multiply the estimated β value by $\frac{2}{3}$ to compute ξ and obtain Table 25.

Table 25: ξ Dependencies for ξ Dependant on β

Parameter	Dependency	Parameter	Dependency
ξ_{AA}	$\frac{2}{3}\beta$	$\frac{\xi_{CC}}{N_{C0}}$	$\frac{2}{3}\frac{1}{4}\frac{24515}{42739}\beta$
ξ_{CA}	$\frac{2}{3}\frac{1}{3}\frac{24515}{42739}\beta$	$\frac{\xi_{AC}}{N_{A0}}$	$\frac{2}{3}\frac{1}{6}\beta$

With these dependencies, we can estimate a singular β value instead of all eight infection rates. We run the Monte Carlo Simulation for this case and obtain Table 26.

Table 26: Practical Identifiability for $\beta, \gamma_C, \gamma_A, \epsilon_C$, and ϵ_A when ξ is dependant on β where $T = 250$ using fminsearchbnd

σ	β	γ_C	γ_A	ϵ_C	ϵ_A
0	0	0	0	0	0
1	0.0190	0.0464	0.0612	0.6337	0.1567
5	0.0886	0.2060	0.2923	3.0605	0.7388
10	0.2075	0.4240	0.8978	5.7178	1.9162
20	0.6626	0.9380	3.0319	13.6892	4.4418
30	1.6097	1.4879	6.4126	19.1212	7.6412

By making the ξ values dependent on a singular β value, we obtain a full set of practically identifiable parameters for the adult child infectious E case.

5.2 Identifiability Map

In the above section, we analyze the practical identifiability of parameters given different unknowns using the same estimated values for each table. We now evaluate the practical identi-

fiability of a given set of unknown parameters using a range of estimated values given in Table 27. We approximate a range of acceptable values for each infection rate based on its definition and the average range of R_0.

Table 27: Fixed Parameter Ranges

Parameter	Value Range	Source
μ	0.00008	[8] [9]
f	0.0005	[8] [10]
β_{CC}	[0.001,1.0]	Estimated using R_0
β_{AA} and β	$[5.0*10^{-7}, 4.0*10^{-4}]$	Estimated using R_0
β_{CA}	$[1.0*10^{-8}, 0.005]$	Estimated using R_0
β_{AC}	[0.01,1.0]	Estimated using R_0
γ_C	[0.02,1.0]	Child recovery period from 1 to 50 days
γ_A	[0.01,1.0]	Adult recovery period from 1 to 100 days
ϵ_C	[0.04,1.0]	Child incubation period from 1 to 25 days
ϵ_A	[0.02,1.0]	Adult incubation period from 1 to 50 days

Similar to when we change our model from infectious to non-infectious, the dynamics change when we change the value of different model parameters. For this reason, we must continue to adjust the time span to account for these differences. Again we verify the time of peak infection precedes T. We denote time of peak infection for children as $TPIC$ and time of peak infection for adults as $TPIA$.

Note that for the approximate parameter ranges proposed below, more investigation is necessary to confirm that all values between the bounds yield a fully practically identifiable set. Additionally, we need to take a further look outside of those bounds to verify that no other values yield a practically identifiable set.

5.2.1 Dependant β Model

Because our model is practically identifiable for prevalence data when we use a single β value on which the β values are dependent, we return to this model. We define our β values using the dependencies in Table 18 and use the fixed values we defined in Table 13. For each parameter, then, we calculate the ARE for our $\beta, \gamma_C, \gamma_A, \epsilon_C$, and ϵ_A when the parameter is at different values within its value range. For β we use the parameter range given for β_{AA}.

We begin by running the Monte Carlo Simulation on the two ends of the parameter range for β while keeping all other parameters at their values given in Table 13 to obtain tables 28 and 29.

Table 28: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\beta = 0.0000005$, $T = 195$, $TPIC = 1$, and $TPIA = 1$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1526	0.0072	0.0046	0.0284	0.0162

5	0.5268	0.0367	0.0325	0.1300	0.1388
10	0.9886	0.1128	0.0908	0.3342	0.3771
20	5.9613	0.4510	0.3062	0.8449	3.4152
30	23.4016	1.0635	0.7872	4.9230	18.0515

Table 29: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\beta = 0.0004$, $T = 200$, $TPIC = 53$, and $TPIA = 45$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0944	0.0400	0.0557	0.3870	0.2187
5	0.5990	0.2459	0.3777	3.3116	1.3005
10	1.5917	0.6304	0.8402	7.3785	3.5212
20	4.6029	1.9077	1.6981	14.2028	8.8168
30	12.0869	4.2887	2.6892	21.8770	18.9709

Since both ends of the range yield a full set of practically identifiable parameters, we next investigate if the model stays practically identifiable between these two β values in tables 30 and **??**.

Table 30: Practical Identifiability for β, γ_A, γ_C, ϵ_C, and ϵ_A when $\beta = 0.000004$, $T = 125$, $TPIC = 1$, and $TPIA = 1$ using fminsearchbnd

σ	β	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	ϵ_A
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	3.0249	0.1610	0.0256	10.1617	3.8420
5	10.6874	0.7040	0.1449	20.0109	21.5699
10	24.5039	1.6141	0.4012	37.8625	57.3678
20	46.2204	3.1104	0.8756	68.8195	124.5263
30	80.5635	5.9262	1.9789	60.3148	108.3582

Table 31: Practical Identifiability for β, γ_A, γ_C, ϵ_C, and ϵ_A when $\beta = 0.00016$, $T = 365$, $TPIC = 1$, and $TPIA = 103$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0616	0.0528	0.0518	0.8005	0.2497
5	0.3738	0.3134	0.2964	4.2328	1.5077
10	0.8908	0.9731	0.7320	9.2118	3.0265

20	2.5577	3.2361	2.3114	18.9912	7.7423
30	5.4925	6.9734	5.4811	39.6907	16.5372

We must continue to investigate the practical identifiability of our model for different values of β. However, to get an idea of the rest of the parameters' ranges, we move on to map the identifiability of γ_A. Similar to β, we begin mapping identifiability for this parameter by computing the ARE for the ends of the parameter range to obtain tables 32 and 33.

Table 32: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.01$, $T = 365$, $TPIC = 61$, and $TPIA = 60$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0872	0.0257	0.0579	0.7055	0.2529
5	0.4095	0.1703	0.3222	3.7036	1.2640
10	0.6244	0.4891	0.8397	7.8757	2.5991
20	1.5976	2.0153	3.1139	13.4555	5.0651
30	2.3225	4.5024	6.8350	24.7370	9.4149

Table 33: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 1.0$, $T = 30$, $TPIC = 1$, and $TPIA = 1$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0002	0.0000	0.0001	0.0001	0.0001
1	0.5935	0.3870	0.1139	6.8586	0.5597
5	2.6122	1.6227	0.6681	64.2359	3.0674
10	4.5095	3.0188	1.4551	105.0389	4.8214
20	10.8615	7.1470	3.2963	167.7244	12.5190
30	15.2703	9.6019	5.5453	187.4323	20.4713

The lower end of the range yields a fully practically identifiable set of parameters, but the upper end of the range does not. We now search for the boundary where this difference occurs in tables 34, 35, 36, 37, 38, 39, 40, and 41.

Table 34: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.5$, $T = 50$, $TPIC = 1$, and $TPIA = 1$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0	0	0	0	0
1	0.7698	0.5149	0.1335	2.3705	0.4567

5	3.8951	2.5776	0.7108	17.5978	2.4769
10	7.1725	4.6693	1.6443	58.6901	4.7893
20	15.8487	10.1910	3.0086	131.7590	10.2119
30	23.0252	15.8001	5.5299	166.8361	14.8819

Table 35: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.3$, $T = 50$, $TPIC = 1$, and $TPIA = 1$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9063	0.6958	0.2208	1.9476	0.7221
5	4.5156	3.5064	1.1067	16.9069	3.7440
10	9.7009	7.6026	2.1805	67.3387	6.2244
20	17.7700	14.1401	5.3830	115.2633	12.1130
30	23.2036	18.7352	7.8462	154.2190	19.3232

Table 36: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.2$, $T = 200$, $TPIC = 1$, and $TPIA = 1$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1161	0.1002	0.0757	0.2536	0.0563
5	1.0969	0.9728	0.7196	3.9257	0.8851
10	2.4272	2.1770	1.5926	11.8892	2.1884
20	5.9350	5.5649	4.1856	44.7258	7.2584
30	10.7033	10.2882	8.1632	81.9746	15.3438

Table 37: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.15$, $T = 250$, $TPIC = 1$, and $TPIA = 74$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1173	0.0913	0.0741	1.0623	0.2773
5	0.5879	0.4738	0.4586	5.0617	1.3024
10	1.2467	1.1197	0.9629	11.6072	2.7184
20	3.9498	3.8812	3.3953	29.5546	7.4124
30	7.7456	7.8386	6.8562	61.8900	15.5141

Table 38: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.12$, $T = 250$, $TPIC = 1$, and $TPIA = 71$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1050	0.0795	0.0717	0.6580	0.2970
5	0.5715	0.4553	0.3861	3.5190	1.5241
10	1.0884	1.0103	0.8070	8.2884	2.6952
20	3.3318	3.5152	2.7407	16.7910	7.7009
30	6.5547	7.4148	5.9020	42.8971	15.0314

Table 39: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.1$, $T = 250$, $TPIC = 1$, and $TPIA = 66$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0770	0.0628	0.0553	0.6864	0.2476
5	0.4081	0.3632	0.3144	2.9394	1.2332
10	0.8235	0.9421	0.8326	7.2856	2.1684
20	2.4699	3.3543	2.3183	16.1494	6.8075
30	4.2397	6.9850	4.7081	31.9194	8.6453

Table 40: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.09$, $T = 250$, $TPIC = 1$, and $TPIA = 65$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0604	0.0597	0.0551	0.6126	0.2131
5	0.3325	0.3285	0.2833	3.3429	1.1299
10	0.7540	0.9451	0.6931	6.8818	2.4445
20	1.6257	3.0971	1.7805	13.8924	4.6627
30	2.4306	6.2701	3.7298	31.9039	5.5677

Table 41: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_A = 0.08$, $T = 250$, $TPIC = 72$, and $TPIA = 63$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0680	0.0586	0.0739	0.6610	0.2533
5	0.2859	0.3065	0.3618	3.4984	1.0518
10	0.5687	0.8722	0.6122	5.6828	2.0875
20	1.0476	2.8438	1.5677	13.0673	3.8018
30	1.4684	6.2476	2.7375	19.3249	5.9203

The values for which γ_A yields a practically identifiable set of parameters is approximately between 0.01 and 0.08 meaning this model can accurately estimate the parameters for an adult infectious period between 100 days and 12.5 days.

Now we explore the identifiability map of γ_C by computing the ARE at either end of its value range to obtain tables 42 and 43.

Table 42: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_C = 0.02$, $T = 365$, $TPIC = 87$, and $TPIA = 60$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0542	0.0555	0.0240	0.5173	0.1911
5	0.3248	0.3720	0.1723	4.5781	1.2319
10	0.7707	1.0024	0.5528	11.6625	2.4693
20	2.4551	3.5967	2.1855	45.5253	6.8717
30	4.6247	7.4720	4.4239	107.0400	13.0349

Table 43: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_C = 1.0$, $T = 125$, $TPIC = 1$, and $TPIA = 64$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0	0	0	0	0
1	0.1761	0.0886	0.1250	0.5875	0.5145
5	0.9121	0.4555	0.6306	3.5189	2.6420
10	1.6102	1.0020	1.2339	6.5729	4.6476
20	3.3351	3.3952	2.9149	13.1267	9.9573
30	6.0022	7.6331	5.4516	16.6192	18.8283

At the lower end of the range for γ_C we do not obtain a full set of practically identifiable parameters, but at the upper end we do. We again search for the boundary at which this difference

occurs in tables 44, 45, 46, and 47.

Table 44: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_C = 0.5$, $T = 175$, $TPIC = 1$, and $TPIA = 64$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0949	0.0590	0.0913	0.5674	0.3001
5	0.4937	0.3543	0.3842	2.5545	1.6503
10	0.9733	0.8024	0.8979	4.9262	2.9986
20	1.9435	3.0143	2.0185	9.8859	6.3943
30	2.7831	6.4262	3.3870	13.5106	8.6469

Table 45: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_C = 0.06$, $T = 250$, $TPIC = 74$, and $TPIA = 61$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0629	0.0489	0.0497	0.7528	0.2414
5	0.2483	0.3193	0.2383	3.8534	0.9425
10	0.5110	0.8971	0.4691	7.0392	2.0844
20	1.1235	2.9026	1.0902	15.4620	4.0582
30	1.9380	6.4332	2.0229	30.2558	7.1187

Table 46: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_C = 0.07$, $T = 250$, $TPIC = 73$, and $TPIA = 61$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0571	0.0563	0.0489	0.6220	0.2180
5	0.2560	0.3310	0.2274	3.5929	1.0088
10	0.5858	0.8259	0.5591	7.1709	2.2544
20	1.3345	2.9469	1.3355	15.1935	5.0824
30	1.9820	6.0588	2.2404	31.3140	7.3856

Table 47: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\gamma_C = 0.08$, $T = 250$, $TPIC = 72$, and $TPIA = 61$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0601	0.0539	0.0549	0.6302	0.2286
5	0.3493	0.2942	0.3003	3.3335	1.2996
10	0.5389	0.7892	0.6388	6.9523	1.9320
20	1.2480	2.8015	1.4044	13.6964	4.5506
30	2.1026	6.0151	2.4467	26.5302	7.5133

The approximate range at which γ_C yields a full set of practically identifiable parameters is between 0.08 and 1.0. Meaning, diseases with a child infectious period between 1 and 12.5 days can be accurately approximated with this model.

We now begin mapping the practical identifiability of ϵ_C by testing the practical identifiability of the parameter for the upper and lower bounds of the parameter range. 48 and 49

Table 48: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_C = 0.04$, $T = 365$, $TPIC = 1$, and $TPIA = 62$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0423	0.0547	0.0604	0.0323	0.1476
5	0.2433	0.3321	0.3732	0.1719	0.9253
10	0.4974	0.9131	0.6405	0.3146	1.8875
20	1.2767	2.8125	2.4442	0.7729	4.1982
30	1.6866	6.1637	4.2758	1.1305	5.2616

Table 49: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_C = 1.0$, $T = 365$, $TPIC = 68$, and $TPIA = 62$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1050	0.0106	0.0431	0.6645	0.1777
5	0.6528	0.0636	0.3087	5.5361	1.0142
10	1.0013	0.1027	0.6832	18.6103	1.5313
20	1.3953	0.1790	1.0230	32.6951	2.1079
30	2.0743	0.4397	1.2414	32.0216	2.7311

For the lower bound of ϵ_C we obtain a full set of practically identifiable parameters, but at

the upper bound we do not. We now search for the boundary where this difference arises in tables 50, 51, 52, 53, 54, and 55.

Table 50: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_C = 0.5$, $T = 200$, $TPIC = 69$, and $TPIA = 62$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0684	0.0506	0.0561	0.3727	0.2256
5	0.3973	0.3052	0.3494	5.1916	1.3385
10	0.8619	0.8072	0.7935	11.2485	2.9712
20	1.3392	2.8840	1.7248	21.4163	4.7539
30	2.2909	6.2832	3.0552	38.5169	7.9084

Table 51: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_C = 0.4$, $T = 200$, $TPIC = 71$, and $TPIA = 62$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0555	0.0538	0.0567	0.2839	0.1770
5	0.4361	0.3276	0.3166	4.2993	1.5107
10	0.8546	0.8591	0.7896	8.7362	2.9626
20	1.4645	2.9444	1.6213	16.9206	5.0296
30	2.3966	6.1685	3.1914	33.0964	7.8241

Table 52: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_C = 0.32$, $T = 200$, $TPIC = 71$, $TPIA = 62$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0617	0.0561	0.0652	0.6329	0.2183
5	0.4045	0.3636	0.3612	3.0451	1.3623
10	0.7589	0.8121	0.7352	7.7220	2.6152
20	1.5601	2.9351	1.6034	15.3607	5.3677
30	2.3926	6.0988	3.0268	22.4301	7.8491

Table 53: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_C = 0.33$, $T = 200$, $TPIC = 70$, and $TPIA = 62$ using fminsearchbnd

σ	β	γ_A	γ_C	ϵ_C	ϵ_A
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0790	0.0558	0.0665	0.6381	0.2718
5	0.3735	0.2999	0.3431	3.2312	1.2658
10	0.7621	0.8286	0.7933	8.6262	2.5789
20	1.5376	2.9175	1.7058	16.0619	5.2035
30	2.5923	6.1457	3.0176	27.7041	8.6299

Table 54: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_C = 0.34$, $T = 200$, $TPIC = 70$, and $TPIA = 62$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0825	0.0541	0.0754	0.7351	0.2946
5	0.4708	0.3237	0.3158	3.4873	1.6645
10	0.7392	0.7920	0.6243	6.3699	2.5597
20	1.7422	2.7708	1.5111	16.9196	5.9840
30	2.2649	6.0224	2.8441	22.5027	7.5418

Table 55: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_C = 0.35$, $T = 200$, $TPIC = 70$, and $TPIA = 62$ days using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0851	0.0579	0.0688	0.7339	0.2909
5	0.4388	0.3026	0.3772	3.8199	1.5526
10	0.8358	0.8117	0.8194	7.9414	2.8325
20	1.4788	2.9844	1.8227	17.7312	5.0258
30	2.1887	5.9902	3.1726	31.0935	7.8666

The approximate range for ϵ_C at which we obtain a full set of practically identifiable parameters is between 0.04 and 0.34. Therefore, a disease that has an child incubation period between 2.9 and 25 days can be accurately represented by this model.

We now examine the final parameter value for this model by running the two bounds of the parameter range for ϵ_A in tables 56 and 57.

Table 56: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_A = 0.02$, $T = 365$, $TPIC = 1$, and $TPIA = 182$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1201	0.0517	0.0858	0.2669	0.1372
5	0.8768	0.3864	0.5908	7.4865	0.9522
10	1.8407	1.1002	1.3510	18.0360	1.9276
20	3.5683	3.6569	3.1224	37.6624	3.2308
30	6.7231	7.7429	6.1685	71.4130	6.1083

Table 57: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_A = 1.0$, $T = 200$, $TPIC = 53$, and $TPIA = 43$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0313	0.0372	0.0533	0.3123	0.2008
5	0.2516	0.3118	0.3607	3.2858	2.5752
10	0.4861	0.6359	0.7862	7.3625	5.3795
20	1.2389	2.3549	1.2953	12.3389	11.1462
30	2.8553	5.4215	2.4808	23.2932	17.6404

Since the lower bound is not practically identifiable and the upper bound is, we look for the point where this difference originates in tables 58, 59, 60, and 61

Table 58: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_A = 0.5$, $T = 200$, $TPIC = 58$, and $TPIA = 48$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0459	0.0541	0.0622	0.4518	0.2849
5	0.2615	0.3034	0.3124	3.1812	1.8595
10	0.5296	0.7561	0.6809	5.7708	3.5288
20	1.2794	2.4571	1.5607	12.5906	7.3047
30	2.6126	5.3281	2.4416	26.4995	12.4513

Table 59: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_A = 0.1$, $T = 200$, $TPIC = 88$, and $TPIA = 80$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1495	0.0660	0.0982	0.1596	0.3029
5	0.9038	0.4067	0.5825	3.3273	1.8324
10	1.9797	1.1830	1.4066	7.6359	3.9461
20	3.9303	3.8493	3.2782	13.5797	7.8095
30	7.3111	8.0145	6.8909	42.1158	14.7275

Table 60: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_A = 0.15$, $T = 200$, $TPIC = 77$, and $TPIA = 68$ using fmincon

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	ϵ_C	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1071	0.0611	0.0706	0.1406	0.2863
5	0.5572	0.3415	0.4028	3.4576	1.5000
10	1.1798	1.0831	0.8718	6.7083	3.1112
20	2.3730	3.1926	2.2335	15.9705	6.3385
30	3.0958	6.6831	3.9475	30.8652	8.0729

Table 61: Practical Identifiability for $\beta, \gamma_A, \gamma_C, \epsilon_C$, and ϵ_A when $\epsilon_A = 0.16$, $T = 200$, $TPIC = 75$, and $TPIA = 66$ using fminsearchbnd

σ	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_A}$	$\boldsymbol{\gamma_C}$	$\boldsymbol{\epsilon_C}$	$\boldsymbol{\epsilon_A}$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.1032	0.0565	0.0722	0.6472	0.2967
5	0.4469	0.3491	0.3602	2.9528	1.3133
10	1.0597	0.8090	0.9824	7.1585	2.9838
20	2.0936	3.0961	2.0394	14.0198	5.9350
30	2.4546	6.3243	3.4628	26.5309	6.5759

The values for which ϵ_A yields a practically identifiable set of parameters appears to be between 0.16 to 1.0 meaning this model can accurately estimate the parameters for an incubation period between 1 and 6.25 days.

5.2.2 β Values Estimated Model

Since when we only estimate the beta values, we obtain a full set of practically identifiable parameters, we begin mapping the identifiability for this case as well.

We begin by testing the identifiability for the bounds of β_{AA}'s parameter range in tables 62 and 63.

Table 62: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{AA} = 0.0000005$, $T = 195$, $TPIC = 1$, and $TPIA = 1$ using fminsearchbnd

σ	β_{AA}	β_{CC}	$\boldsymbol{\beta_{AC}}$	$\boldsymbol{\beta_{CA}}$
0	0.0000	0.0000	0.0000	0.0000
1	7.8800	0.4829	0.3249	0.3149
5	36.2588	2.8119	1.8740	1.4790
10	70.4620	5.7901	3.8235	2.8848
20	92.0124	15.2376	9.4562	4.9852
30	103.6934	34.4476	20.4628	7.4153

Table 63: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{AA} = 0.0004$, $T = 200$, $TPIC = 1$, and $TPIA = 46$ using fminsearchbnd

σ	$\boldsymbol{\beta_{AA}}$	$\boldsymbol{\beta_{CC}}$	$\boldsymbol{\beta_{AC}}$	β_{CA}
0	0.0000	0.0000	0.0000	0.0000
1	0.0413	0.6948	0.1626	2.1758
5	0.2108	3.3178	0.8656	10.6242
10	0.5817	7.5968	2.0336	24.8079
20	1.6262	16.3251	5.7896	57.4667
30	3.4505	27.6155	12.7657	109.3969

Since neither bound for β_{AA} yields a full set of practically identifiable parameters, we examine the middle of the range in tables 64, 65, 66, and 67.

Table 64: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{AA} = 0.000004$, $T = 125$, $TPIC = 1$, and $TPIA = 1$ using fminsearchbnd

σ	β_{AA}	β_{CC}	$\boldsymbol{\beta_{AC}}$	$\boldsymbol{\beta_{CA}}$
0	0.0000	0.0000	0.0000	0.0000
1	1.6800	0.6353	0.4954	0.4311
5	7.2588	3.7168	2.7131	2.0269
10	18.5222	7.5874	5.5541	5.0226
20	38.4874	13.7385	9.7193	12.1064
30	58.5029	32.3559	20.3211	19.4429

Table 65: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{AA} = 0.00035$, $T = 200$, $TPIC = 16$, and $TPIA = 42$ using fminsearchbnd

σ	$\boldsymbol{\beta_{AA}}$	$\boldsymbol{\beta_{CC}}$	$\boldsymbol{\beta_{AC}}$	β_{CA}
0	0.0000	0.0000	0.0000	0.0000
1	0.0401	0.8289	0.1726	1.8685
5	0.1914	4.6137	0.9936	8.8989
10	0.4593	7.7794	1.8332	18.6860
20	1.4194	16.9159	5.9301	47.6317
30	2.8489	29.8869	12.2719	89.0031

Table 66: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{AA} = 0.000375$, $T = 200$, $TPIC = 1$, and $TPIA = 37$ using fminsearchbnd

σ	$\boldsymbol{\beta_{AA}}$	$\boldsymbol{\beta_{CC}}$	$\boldsymbol{\beta_{AC}}$	β_{CA}
0	0.0000	0.0000	0.0000	0.0000
1	0.0360	0.6522	0.1344	1.7855
5	0.2258	3.8992	0.9117	10.3123
10	0.5335	7.8651	2.0186	24.3396
20	1.5626	15.8272	5.3868	54.3245
30	3.4221	28.4894	12.5397	118.7473

Table 67: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{CC} = 0.005$, $T = 200$, $TPIC = 1$, and $TPIA = 44$ using fminsearchbnd

σ	β_{AA}	β_{CC}	β_{AC}	β_{CA}
0	0.0000	0.0000	0.0000	0.0000
1	0.0382	1.9334	0.1547	1.4198
5	0.1769	9.7688	0.7611	6.5110
10	0.4018	19.4541	1.8349	14.8081
20	1.0687	41.4784	4.4315	31.1735
30	2.1889	51.6148	10.1552	60.2314

β_{AA} seems to have a shorter range at which we obtain a full set of practically identifiable parameters. More investigation is necessary to make any further conclusions.

We now move on to examine the identifiability map of β_{CC}. We run the Monte Carlo Simulation at the bounds of β_{CC} to obtain tables 68 and 69.

Table 68: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{CC} = 0.001$, $T = 200$, $TPIC = 1$, and $TPIA = 62$ using fminsearchbnd

σ	$\boldsymbol{\beta_{AA}}$	β_{CC}	$\boldsymbol{\beta_{AC}}$	β_{CA}
0	0.0000	0.0000	0.0000	0.0000
1	0.0320	9.4766	0.1503	1.2673
5	0.1828	54.9126	0.8896	6.6359
10	0.4019	82.0071	1.6332	14.9826
20	0.9255	167.6687	3.6063	28.3282
30	2.3132	146.1000	8.4608	66.5138

Table 69: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{CC} = 1.0$, $T = 175$, $TPIC = 16$, and $TPIA = 47$ using fminsearchbnd

σ	$\boldsymbol{\beta_{AA}}$	$\boldsymbol{\beta_{CC}}$	$\boldsymbol{\beta_{AC}}$	$\boldsymbol{\beta_{CA}}$
0	0.0000	0.0000	0.0000	0.0000
1	0.0871	0.1169	0.1945	0.4361
5	0.3865	0.5953	0.8713	1.8842
10	0.7130	1.3087	2.2586	3.2050
20	1.6375	3.1030	5.5667	6.7165
30	2.8441	6.4822	12.0314	11.0614

The upper bound for β_{CC} yields a practically identifiable set, but the lower bound does not. We search for the origin of this difference in Table 70.

Table 70: Practical Identifiability for $\beta_{AA}, \beta_{CC}, \beta_{AC}$, and β_{CA} when $\beta_{CC} = 0.5$, $T = 175$, $TPIC = 23$, and $TPIA = 51$ days using fminsearchbnd

σ	$\boldsymbol{\beta_{AA}}$	$\boldsymbol{\beta_{CC}}$	$\boldsymbol{\beta_{AC}}$	$\boldsymbol{\beta_{CA}}$
0	0.0000	0.0000	0.0000	0.0000
1	0.0965	0.0455	0.2578	0.5413
5	0.5354	0.2178	1.0641	3.0548
10	0.8841	0.5513	2.1287	5.0861
20	2.1386	1.6187	4.4922	11.0511
30	2.7035	3.0057	6.3879	11.8189

β_{CC} seems to have a larger range than β_{AA}. More investigation is necessary for any further details.

5.3 Profile Likelihood

Profile Likelihood is a another approach to identifiability analysis. It is a one-dimensional representation of the likelihood indicating which values of a single parameter component are in statistical agreement with the available measurements. In this section we present a profile likelihood approach to determine confidence intervals some parameters in our parameter set. We outline the methodology behind Profile Likelihood and present example results from a selection of our parameters.

Profile likelihood uses a Maximum Likelihood Estimation (MLE) approach. Here let $y = (y_1, ..., y_m)$ be a random sample from an unknown population. Then let $f(y|\theta)$ represent the probability density function (PDF) that defines the probability of observing y given the parameter θ. Given observed data, the goal of MLE is to find the corresponding PDF that is most likely to have produced the observed data. The likelihood function used to find this PDF is defined below.

Definition 5.1. Given a random sample $y = (y_1, ..., y_m)$ and a probability density function $f(y|\theta)$, the likelihood function $L(\theta|y)$ is the product of the PDF evaluated at y_i's:

$$L(\theta|y) = \prod_{i=1}^{m} f(y_i|\theta).$$

The maximum likelihood function can be found by maximizing the log-likelihood function $ln(L(\theta|y)$. This is possible because $L(\theta|y)$ and $ln(L(\theta|y)$ are monotonically related, and maximizing either returns the same MLE estimate. Then, there exist two cases: when $ln(L(\theta|y)$ and when $ln(L(\theta|y)$ is not differentiable. Begin first with the case where $ln(L(\theta|y)$ is differentiable. Then, $ln(L(\theta|y)$ must satisfy the differentiable equation:

$$\frac{\partial(L(\theta|y)}{\partial\theta_i = 0}$$

for all $\theta_i = \theta_{i,MLE}$ for all $i = 1, .., m$. This partial derivative is also called the likelihood equation. Observe here that the above condition is not sufficient to guarantee a maximum. Thus, the second partial derivative of the likelihood equation must also satisfy:

$$\frac{\partial^2(L(\theta|y)}{\partial^2\theta_i < 0}$$

for all $\theta_i = \theta_{i,MLE}$ for all $i = 1, .., m$. In other words, this requires $ln(L(\theta|y)$ to be convex in the neighborhood of θ_{MLE}.

Now in the case where $ln(L(\theta|y)$ is not differentiable, an analytic form of the solution to MLE cannot be found. Instead, the MLE estimate must be computed numerically using a nonlinear optimization algorithm. This process works by reducing the parameter space to smaller subsets instead of performing an exhaustive search. Over each iteration, the parameter set is slightly modified to get closer to optimized performance. The process continues until the parameter set has converged to the optimal set of parameters.

The profile likelihood of a parameter θ_i is computed as follows:

$$\chi^2_{PL}(\theta_i) = \min_{\theta_{j\neq i}}[\chi^2(\theta)].$$

The profile likelihood approach numerically optimizes $\chi^2(\theta)$ to produce an estimated parameter set $\widehat{\theta}$. Then the optimization approach beings with $\widehat{\theta_i}$ and samples along the profile likelihood of θ_i by taking a step θ_{STEP} in an increasing/decreasing direction of θ_i and subsequently re-optimizing all $\theta_{j\neq i}$. This is repeated until the desired threshold ∇_α is reached or the maximum number of steps is reached.

The value of θ_{STEP} changes depending on the steepness of the likelihood at each point. Consequently, if the likelihood is flat, the value of θ_{STEP} should be large; if the likelihood is steep, the value of θ_{STEP} should be small. We can write this mathematically such that θ_{STEP} should fulfill the following condition:

$$\chi^2(\theta_{last} + \theta_{step}) - \chi^2(\theta_{last}) \approx q\nabla_\alpha$$

where $q \in [0,1]$ and θ_{last} is the parameter values from the previous step.

The example results from our profile likelihood approach are below. The fitted parameter set is given by:

$$\beta_{AA} = 0.00027, \beta_{CC} = 0.01, \beta_{AC} = 0.01, \beta_{CA} = 0.00005, \gamma_C = 0.1, \gamma_A = 0.074,$$
$$\epsilon_C = 0.3, \epsilon_A = 0.2, \xi_{CC} = \xi_{AA} = \xi_{CA} = \xi_{AC} = 0$$

and the constants are given by

$$f = .0005, \mu = .00008, N_C = 250, N_A = 750.$$

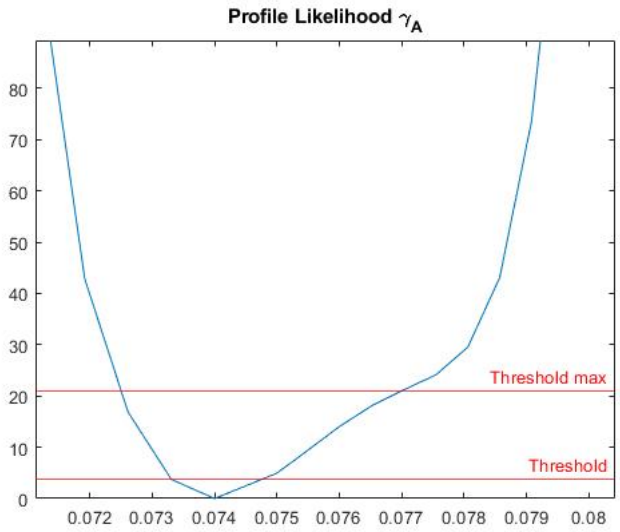

Figure 4: Profile Likelihood results for γ_A: True Value = 0.074

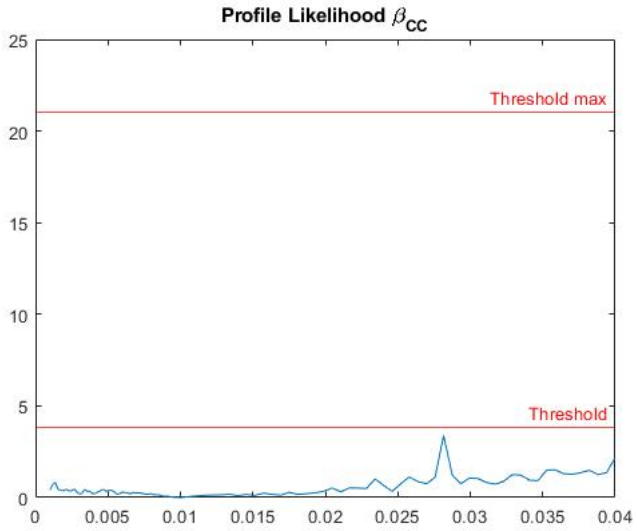

Figure 5: Profile Likelihood results for β_{CC}: True Value = 0.01

6 Comparisons Between the Simple and Adult-Child Model

In order to determine the differences between the Simple and Adult-Child model, we first verify that the two models represent the same epidemic spread. We create a base case set of parameters where the two models have the same outputs by eliminating the differences between the Adult and Child populations. Since our λ_C value is frequency dependent and λ_A is density dependent, we accommodate for this by choosing β and setting our Adult-Child parameters based on their dependencies.

$$\begin{aligned} \beta_{AA} = \beta \quad & \beta_{CA} = \beta \\ \beta_{CC} = \beta N_C \quad & \beta_{AC} = \beta N_A \\ \gamma_A = \gamma_C = \gamma \quad & \epsilon_C = \epsilon_A = \epsilon \end{aligned}$$

We let $\beta = 0.00027$, $\epsilon = 0.25$, and $\gamma = 0.087$, where ϵ and γ are averages of the values used in 5 and β is the value used in our dependent β case. We then plot the solutions of the Simple SEIR model with the sum of the Adult and Child compartment solutions.

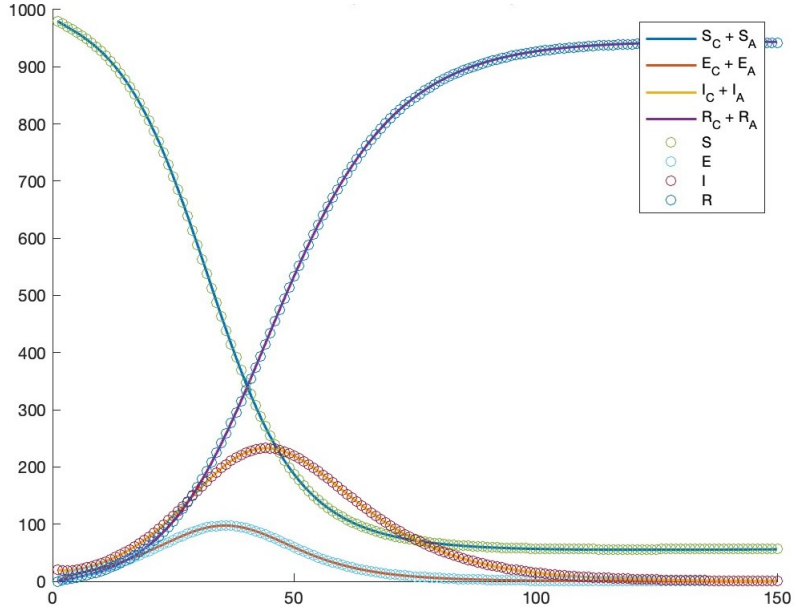

Figure 6: Solutions of Standard and Adult-Child SEIR Models for Base Case

Since the parameters are equal, the solutions of each model in Figure 6 are the same. When we calculate R_0 for each of the models, $R_0 = 3.1009$ for the Adult-Child model and $R_0 = 3.1001$ for the Simple SEIR model. We can attribute the differences in the value to small rounding errors in the calculations. This confirms that our models depict the same epidemic when given the same parameters.

To interpret the effect the β parameters have on R_0 in the Adult-Child model, we demonstrate the differences between the R_0 values by using contour plots to compare how R_0 changes when we vary β. We plot R_0 of the Adult-Child compartment model as β_{CA} and β_{AA} vary and determine its relation to the standard SEIR model: $R_0 < 1$, $1 < R_0 < R_0$ Simple, and R_0 Simple

$< R_0$. We look at these for the cases where the Child and Adult populations are equal, and for the cases where the Child and Adult populations reflect those in 5.

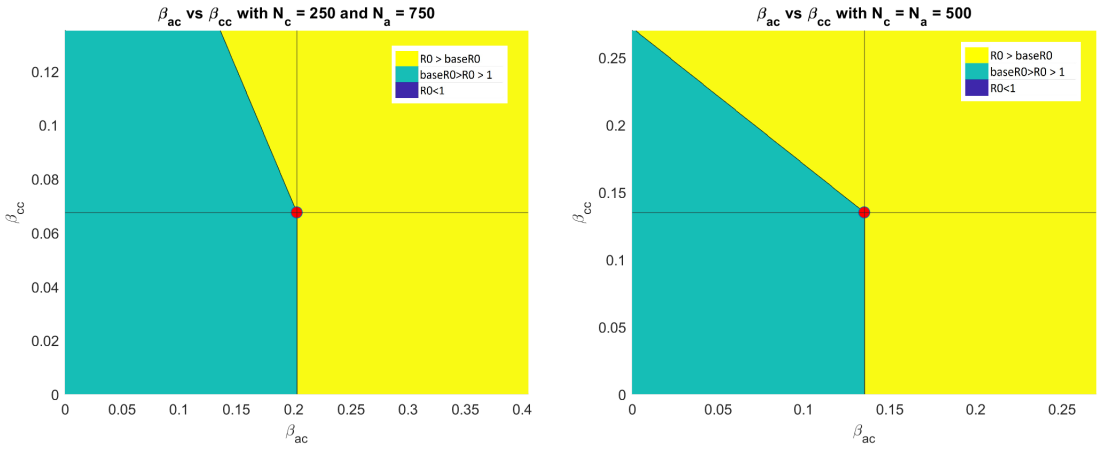

Figure 7: Impact of β_{CA} and β_{AA} on R_0 in Comparison with Simple SEIR Model

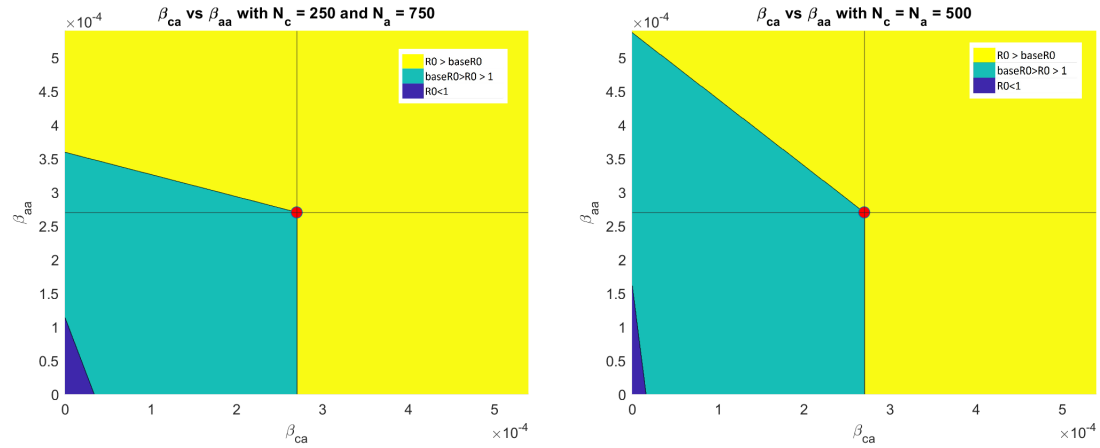

Figure 8: Impact of β_{AC} and β_{CC} on R_0 in Comparison with Simple SEIR Model

In Figure 8, we see that the effect β_{AC} has on R_0 is greater than β_{CC}, and is even more so when the population sizes of Adults and Children are equal. As β_{AC} increases, R_0 grows faster than the initial Simple R_0, while β_{CC} has less impact on increasing R_0. Similarly, βCA has a more significant impact on R_0 than β_{AA} when the population sizes are both equal and unequal. This demonstrates how the interactions between the Adult and Child populations has a significant impact on the epidemic threshold.

We can similarly observe the impact of β_{CA} and β_{AC} by comparing the Adult-Child model to Simple SEIR models. We first use our parameter values from 5 to approximate β, ϵ, and γ by averaging and weighting β values by their respective population sizes as follows:

Table 71: Standard Model Parameters for SEIR Non-Infectious E

Parameter	Equation	Value
β	$(\beta_{AA}+\beta_{CA})\frac{N_A}{N}+(\frac{\beta_{CC}}{N_C}+\frac{\beta_{AC}}{N_A})\frac{N_C}{N}$	0.0002533
γ	$\frac{\gamma_C+\gamma_A}{2}$	0.087
ϵ	$\frac{\epsilon_C+\epsilon_A}{2}$	0.25

We then graph the solutions of both models, summing the Adult and Child compartments to compare to the Simple SEIR model. Figure 9 shows the plot of both solutions in one graph over the course of 150 days.

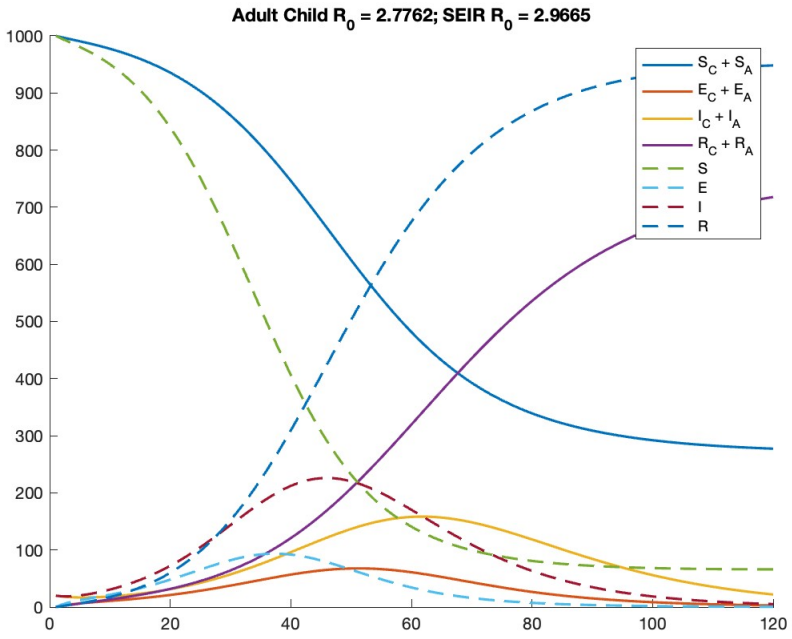

Figure 9: Solutions of Simple and Adult-Child SEIR Models

Note that in this case, the R_0 for the Adult-Child model is less than the R_0 for the Simple SEIR model. Because our parameter values are no longer equal, the Adult-Child model has a later and lower peak infection. Since this model isn't practically identifiable for all parameters, we also look at our dependent-β model, which has a full set of practically identifiable parameters. We calculate a new β using our β dependencies and plot the solutions. Figure 10 shows the solutions of both our dependent-β Adult-Child model and the corresponding Simple SEIR model.

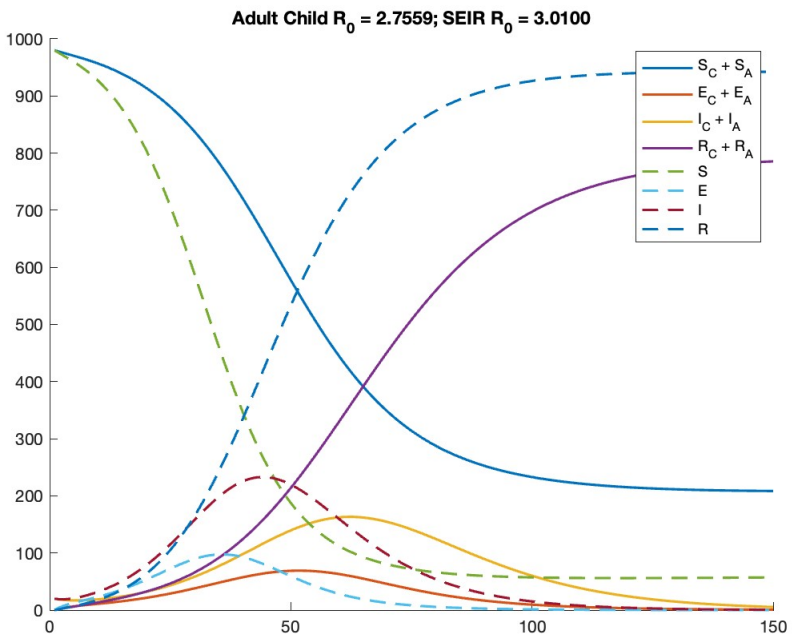

Figure 10: Solutions of Simple and Adult-Child SEIR Models for Dependent Beta

In the Simple SEIR model, the number of Recovered individuals approaches N and the number of Susceptible individuals approaches 0. In the Adult-Child model, however, the solutions appear to approach a herd immunity threshold, where the epidemic dies out without infecting almost the entire population. To understand this key difference, we look at the individual compartments of the Adult-Child model.

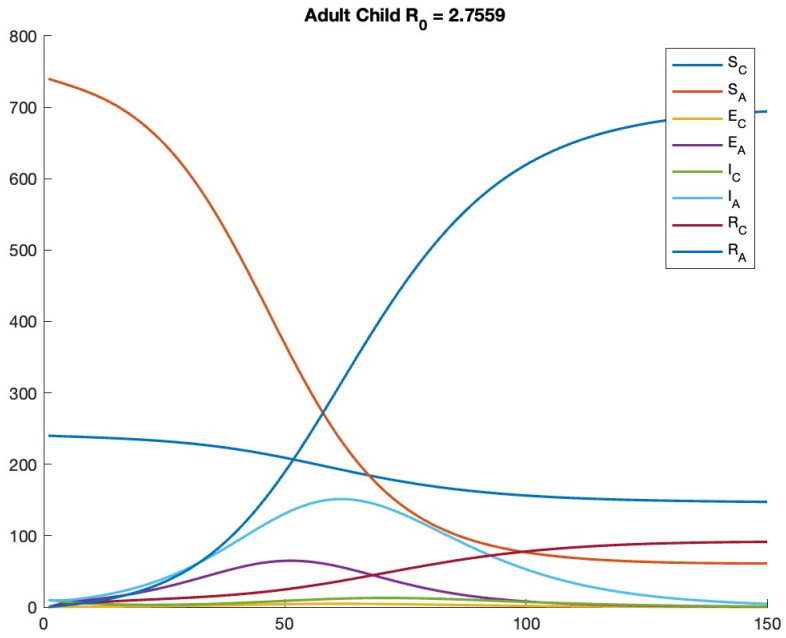

Figure 11: Solutions of Dependent Beta Adult-Child SEIR Model

In Figure 11, we see that while the number of Recovered Adults approaches the Adult population size, the number of Recovered Children does not approach the Child population size and plateaus at a much lower value. This indicates that while the Simple SEIR model seen in 10 models an epidemic, only one of the populations, the Adult population, experiences an epidemic in the Adult-Child model. We look at the impact of the Adult-Child model on the separate populations by comparing them to isolated populations that do not interact with one another. We look first at an isolated Child population that does not interact with the Adult population. Because only children are being modeled in our new Simple SEIR model, we let $\beta = \beta_{CC}$, $\epsilon = \epsilon_C$, and $\gamma = \gamma_C$ to represent the Child to Child infectiousness, Child incubation rate, and Child recovery rate. We plot the Infected Child solution of the Adult-Child model in comparison to the Infected solution of the Simple SEIR model using our Child parameters. Since the Adult-Child model has an initial Susceptible Child population of $N_C = 250$ but a total population size of $N = 1000$, there are 1000 Susceptible individuals that can become Infected and pass on the disease, but only 250 Children that are being observed. We account for this by modeling a Simple SEIR model for both a population size of $N = 250$ and $N = 1000$.

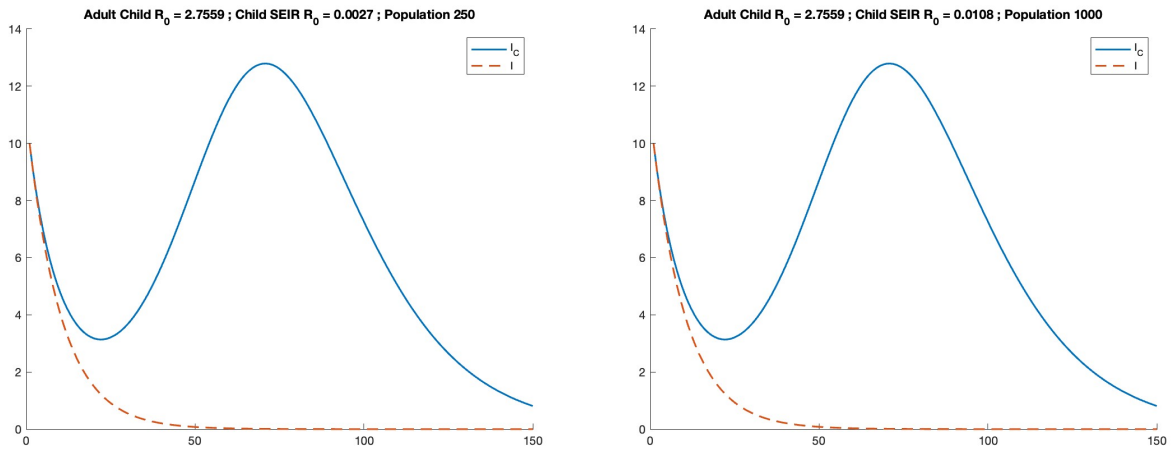

Figure 12: Infected Children for Child-Adult and Simple SEIR Model for Varying Population Sizes

In both cases in Figure 12, the models are non-epidemic, with small R_0 values. We infer that the peak in Infected Child cases in the Adult-Child model results due to the interaction with a more infectious Adult population, rather than with other Infected Children.

We next look at an isolated Adult population that has no contact with the Child population. We let $\beta = \beta_{CC}$, $\epsilon = \epsilon_A$, and $\gamma = \gamma_A$ to represent the Adult to Adult infectiousness, Adult incubation rate, and Adult recovery rate. We plot the Infected Adult solution of the Adult-Child model in comparison to the Infected solution of the Simple SEIR model using our Adult parameters. We use population sizes of $N = 750$ and $N = 1000$ to account for both the initial Susceptible Adult population and total population of the Adult-Child model.

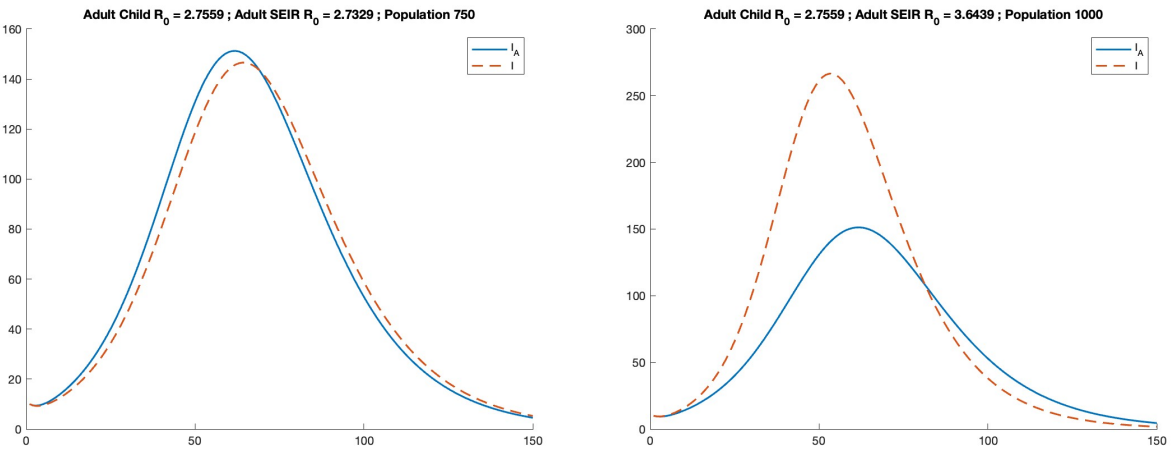

Figure 13: Infected Adults for Child-Adult and Simple SEIR Model for Varying Population Sizes

Both cases in Figure 71 depict epidemics with $R_0 > 1$. When the population size of the Simple SEIR model is $N = 750$, the curve much more closely resembles the Infected Adult solution of the Adult-Child model, but has a smaller peak as a result of the few cases caused by the 250 Children unaccounted for in the Simple model. When the population size of the Simple model is $N = 1000$, the peak is much larger than the Infected Adult solution. The infectiousness of

Adults is significant enough that the 250 additional Adults create more infections than the 250 Children in the Adult-Child model.

7 Conclusion

7.1 Discussion

Mathematical modeling has helped guide decision- and policy-making regarding COVID-19 and other infectious diseases. Incorporating previous knowledge about a disease may help predict the outcome of an epidemic after disease-preventing measures have been taken. As a model becomes more complicated, however, applicability and accuracy can be lost. In this report, we successfully developed a dual-population model with a full set of structurally identifiable parameters. While our SEIR Child-Adult was not practically identifiable for all twelve unknown parameters, we were able to create relevant β and ξ dependencies that produced a full set of practically identifiable parameters.

Additionally, because not all parameters in the simple SEIR model are practically identifiable, our model offers a higher accuracy when using imperfect, real-world data. Our age-based SEIR model can provide additional insight about disease spread and dynamics without compromising identifiability properties. Our comparisons of R_0 and peak infection values indicate the importance of segregating COVID-19 data by age in order to develop accurate forecasting methods.

Our structural identifiability conclusions suggest that recovered data might be a reliable output vector for epidemic modeling. For example, we have shown that all models parameters are globally identifiable when the output vectors are given by R_A, R_C. Moreover we have shown that when the output vector is given in terms of cumulative data, all model parameters are structurally non-identifiable. These conclusions suggest that recovered, prevalence, and incidence data are useful for our SEIR Adult-Child model.

7.2 Future Work

In our model, we make several assumptions that impact our conclusions. We assume that the population of our model remains constant. Since in many cases the birth and death rates do not match, a varying population could change the outcome of the model and requires further study. Our model also does not take into account disease-related deaths, assuming all Infected individuals move into the Recovered compartment.

Our model divides the population into two strict age-based categories focused on children and adults, but recent conclusions about COVID-19 present other age categories that are uniquely impacted by the disease when compared to the general population. Further studies could be made into the elderly population, which experienced higher mortality rates than children; working age adults, who contribute to a significant amount of person-to-person contact; or into immuno-compromised individuals, who may never "age out" of their compartment the way children do but have higher susceptibility to disease and frequently longer recovery periods. Populations can also be separated into more than two compartments, like multiple age groups, which would increase the number of parameters in the model.

The practical identifiability analysis in this paper is restricted to prevalence data for which the Infected Adult and Infected Children data are separately reported. Further questions may

arise when looking at prevalence data that does not differentiate between populations, or single-population data. Investigations into other structurally identifiable output vectors from our structural identifiability analysis could provide different results.

At the time of this report, our investigations into the practical identifiability of parameters within specified ranges is still incomplete. Questions remain as to the exact range for which a parameter within the dependant β and estimated β model may be estimated before becoming practically non-identifiable. Additionally, none of the parameter ranges have been investigated for the ξ dependent on β model. It would also be beneficial to investigate if adjusting the equations for the ξ and β dependencies can make ξ practically identifiable as it is nearly practically identifiable.

Our analysis suggests that Profile Likelihood may be a successful method of computing confidence intervals for identifiability. The success of Profile Likelihood is heavily dependent on the optimization function used and the initial guess chosen. Because of this, future work could include computing Profile Likelihood for a wide range of initial guess values and optimization functions.

References

[1] Julie C Blackwood and Lauren M Childs. An introduction to compartmental modeling for the budding infectious disease modeler. 2018.

[2] Vishaal Ram and Laura P Schaposnik. A modified age-structured sir model for covid-19 type viruses. Scientific reports, 11(1):1–15, 2021.

[3] Kiesha Prem, Yang Liu, Timothy W Russell, Adam J Kucharski, Rosalind M Eggo, Nicholas Davies, Stefan Flasche, Samuel Clifford, Carl AB Pearson, James D Munday, et al. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. The Lancet Public Health, 5(5):e261–e270, 2020.

[4] Michael T Meehan, Diana P Rojas, Adeshina I Adekunle, Oyelola A Adegboye, Jamie M Caldwell, Evelyn Turek, Bridget M Williams, Ben J Marais, James M Trauer, and Emma S McBryde. Modelling insights into the covid-19 pandemic. Paediatric respiratory reviews, 35:64–69, 2020.

[5] Harvey Thomas Banks, Shuhua Hu, and W Clayton Thompson. Modeling and inverse problems in the presence of uncertainty. CRC Press, 2014.

[6] Necibe Tuncer and Trang T Le. Structural and practical identifiability analysis of outbreak models. Mathematical biosciences, 299:1–18, 2018.

[7] Age structure, 2022.

[8] Edward Goldstein, Marc Lipsitch, and Muge Cevik. On the effect of age on the transmission of sars-cov-2 in households, schools, and the community. The Journal of infectious diseases, 223(3):362–369, 2021.

[9] America's age profile told through population pyramids, Oct 2021.

[10] Pop1 child population: Number of children (in millions) ages 0–17 in the united states by age, 1950–2020 and projected 2021–2050, Jul 2020.

[11] Young Joon Park, Young June Choe, Ok Park, Shin Young Park, Young-Man Kim, Jieun Kim, Sanghui Kweon, Yeonhee Woo, Jin Gwack, Seong Sun Kim, et al. Contact tracing during coronavirus disease outbreak, south korea, 2020. Emerging infectious diseases, 26(10):2465, 2020.

[12] Andrew William Byrne, David McEvoy, Aine B Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth A Lane, Conor McAloon, et al. Inferred duration of infectious period of sars-cov-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic covid-19 cases. BMJ open, 10(8):e039856, 2020.

[13] Xiuquan Nie, Lieyang Fan, Ge Mu, Qiyou Tan, Mengyi Wang, Yujia Xie, Limin Cao, Min Zhou, Zhuang Zhang, and Weihong Chen. Epidemiological characteristics and incubation period of 7015 confirmed cases with coronavirus disease 2019 outside hubei province in china. The Journal of infectious diseases, 222(1):26–33, 2020.

[14] Xiang Ren, Yu Li, Xiaokun Yang, Zhili Li, Jinzhao Cui, Aiqin Zhu, Hongting Zhao, Jianxing Yu, Taoran Nie, Minrui Ren, et al. Evidence for pre-symptomatic transmission of coronavirus disease 2019 (covid-19) in china. Influenza and other respiratory viruses, 15(1):19–26, 2021.