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

Zander Memon, Charlotte Moser, Zoe Plzak

American University, UW-La Crosse, Boston University

January 21st, 2023

Acknowledgements

This research would not have been possible without the support of graduate student Anu Agarwal and our mentor Dr.Tingting Tang.

We also wish to thank Chris O'neill and Vadim Ponomarenko or their support throughout the REU program.

Finally, we are grateful for NSF grant that supported all mathematics REU students at San Diego State University in 2022.

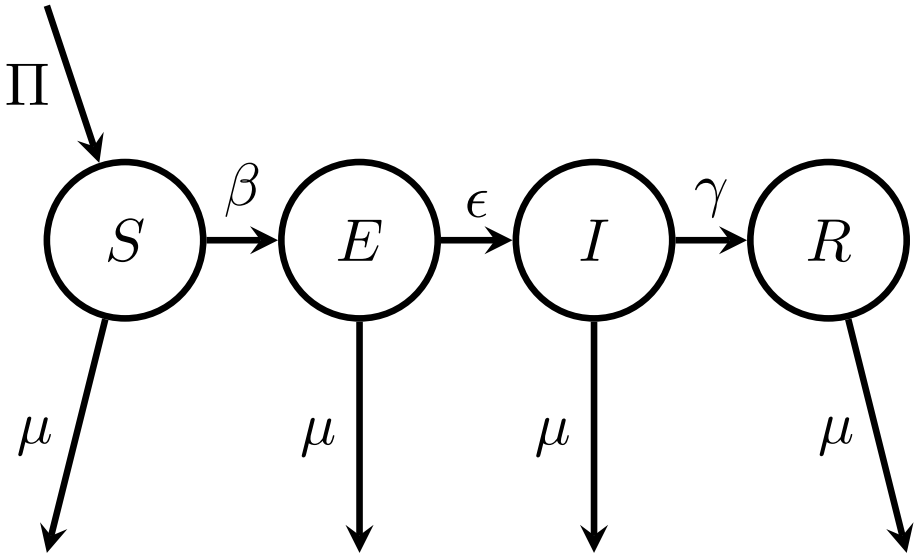
Epidemic Modeling and the SIR/SEIR Models

2 Main Ideas for Modeling:

- Given data about number of cases, can we retroactively understand information about the disease?
- Given biological information about a disease, can we predict the severity and future outcome of an epidemic?

SEIR Model

$$\frac{dS}{dt} = -\beta SI + \Pi N - \mu S$$
$$\frac{dE}{dt} = \beta SI - \epsilon E - \mu E$$
$$\frac{dI}{dt} = \epsilon E - \gamma I - \mu I$$
$$\frac{dR}{dt} = \gamma I - \mu R$$

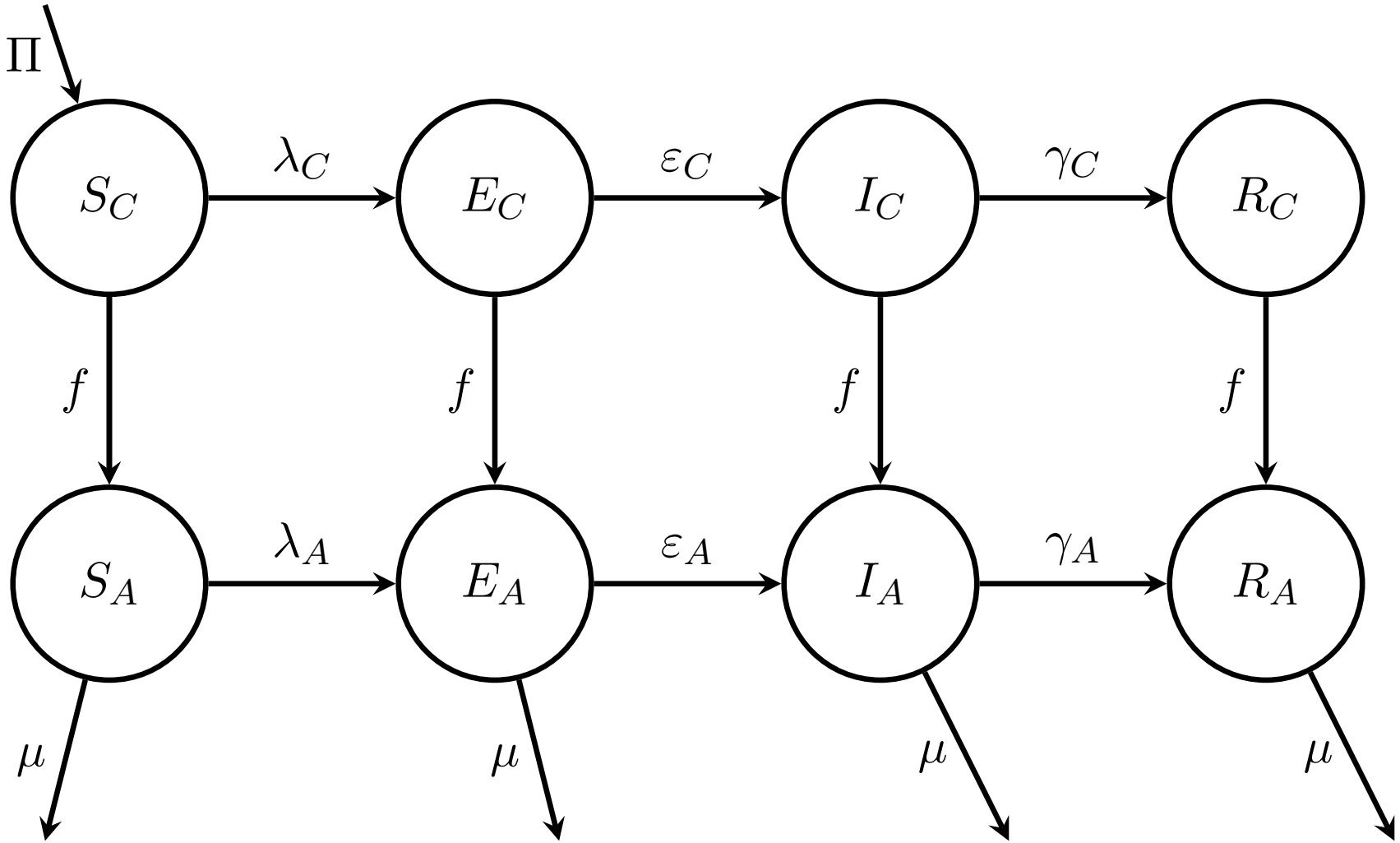


Where $\Pi = \mu$

Age-Dependent Model

- $S_{C/A}$: the number of susceptible people, assuming that everyone is susceptible at t=0.
- $E_{C/A}$: the number of exposed people who made successful contact with an infected person but do not have symptoms.
- $I_{C/A}$: the number of infected people.
- $R_{C/A}$: the number of recovered people, assuming that recovery grants immunity

Age-Dependent Model



Our Known and Unknown Parameters

While most of our parameters are constants, here, λ is a function of I_C, I_A, E_C, E_A.
Children are frequency dependent while adults are density dependent.

- $\lambda_C(I_C, I_A, E_C, E_A) = \beta_{CC}\frac{I_C}{N_C} + \beta_{AC}\frac{I_A}{N_A} + \xi_{CC}\frac{E_C}{N_C} + \xi_{AC}\frac{E_A}{N_A}$
- $\lambda_A(I_C, I_A, E_C, E_A) = \beta_{AA}I_A + \beta_{CA}I_C + \xi_{AA}E_A + \xi_{CA}E_C$

The Model Equations

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

Where $\Pi = \mu$

Parameter Values

Parameter	Range	Reasoning
β_{CC}	$[0.1, 1.0]$	Estimated from reasonable R_0
β_{AC}	$[.05, 1.0]$	Estimated from reasonable R_0
β_{AA}	$[5 \times 10^{-7}, 2 \times 10^{-5}]$	Estimated from reasonable R_0
β_{CA}	$[1 \times 10^{-8}, 5 \times 10^{-7}]$	Estimated from reasonable R_0
γ_C	$[0.02, 1]$	Infectious from 1 to 50 days
γ_A	$[0.01, 1]$	Infectious from 1 day to 100 days
μ	$[4 \times 10^{-5}, 8 \times 10^{-5}]$	Adult age from 35 to 70
f	$[1 \times 10^{-4}, 5 \times 10^{-4}]$	Child age from 5.5 to 27
ϵ_C	$[0.04, 1]$	Incubation from 1 and 25 days
ϵ_A	$[0.02, 1]$	Incubation from 1 to 50 days

Basic Reproduction Number

Definition

R_0 is the number of new infections that can come from one infected individual in a totally susceptible population

- $R_0 < 1$ will not create an epidemic
- $R_0 > 1$ might create an epidemic

We use the next generation matrix method to compute our R_0

Next Generation Matrix Method

Approach: calculate a matrix whose ij-th term represents the average number of infections in the ith compartment created by individuals in the jth compartment

1. Identify diseased compartments
2. Identify $\mathcal{F}$ (new secondary infections) and $\mathcal{V}$ (movement **out** of compartments)

Linearize at the Disease Free Equilibrium and find F (the Jacobian of $\mathcal{F}$) and V (the Jacobian of $\mathcal{V}$)

Calculating R_0 for the Adult-Child Model

The R_0 of our Adult-Child compartment depends on:

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

$$+ \frac{N_C \beta_{AC}(\epsilon_A f^2 + \epsilon_A \epsilon_C f + \epsilon_A f \gamma_C + \epsilon_C f \mu)}{N_A(\epsilon_C + f)(f + \gamma_C)(\epsilon_A + \mu)(\gamma_A + \mu)}$$

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

$$\text{E} = \frac{N_A \xi_{CA}}{\epsilon_C + f} + \frac{N_A \beta_{CA}\epsilon_C}{(\epsilon_C + f)(f + \gamma_C)} + \frac{N_A f \xi_{AA}}{(\epsilon_C + f)(\epsilon_A + mu)} +$$

$$\frac{N_A \beta_{AA}(\epsilon_A f^2 + \epsilon_A \epsilon_C f + \epsilon_A f \gamma_C + \epsilon_C f \mu)}{(\epsilon_C + f)(f + \gamma_C)(\epsilon_A + \mu)(\gamma_A + \mu)}$$

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

Relating SEIR Model to Adult-Child Model

- Goal: compare simple SEIR to SEIR Adult-Child
- How do we relate β to β_{AA},β_{CC},β_{AC}, and β_{CA}?
- How do ϵ and γ relate to ϵ_C, ϵ_A, γ_C, and γ_A?

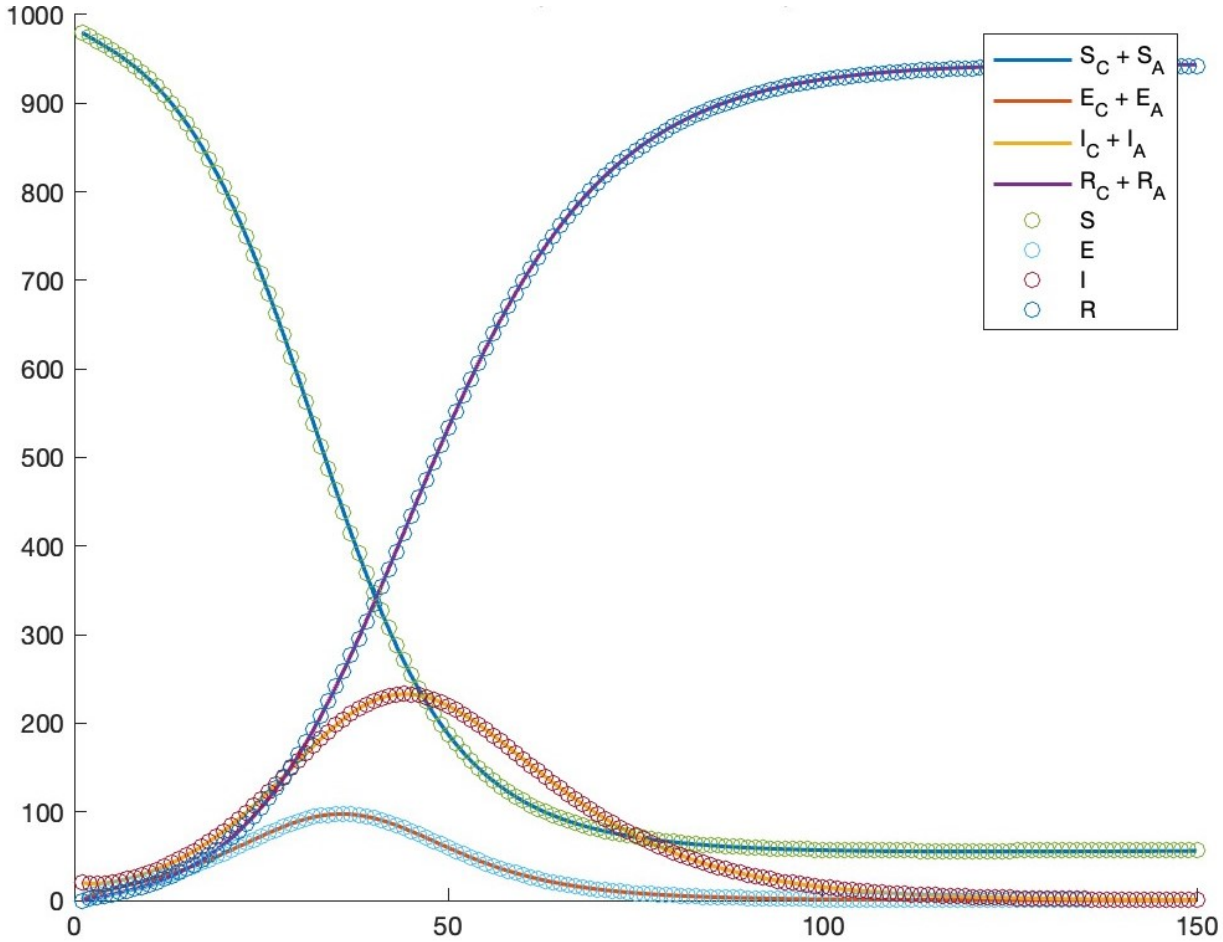
Base Case

- Begin with a base case where Child values = Adult values
- Account for frequency and density dependent β values

Adult-Child	**Relation**	**Adult-Child**	**Relation**
β_{AA}	β	β_{CC}	βN_C
β_{CA}	β	β_{AC}	βN_A
ϵ_C	ϵ	ϵ_A	ϵ
γ_C	γ	γ_A	γ
$N_C = 250$	$N - N_A$	$N_A = 750$	$N - N_C$

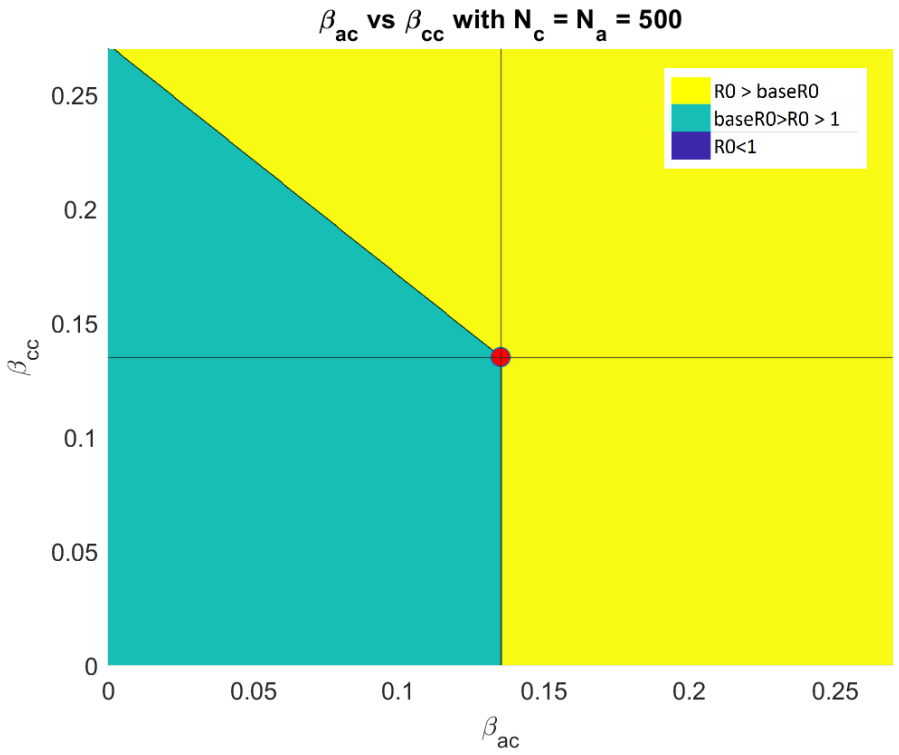
$\beta \approx 0.00027$, $\epsilon = 0.25$, $\gamma = 0.087$

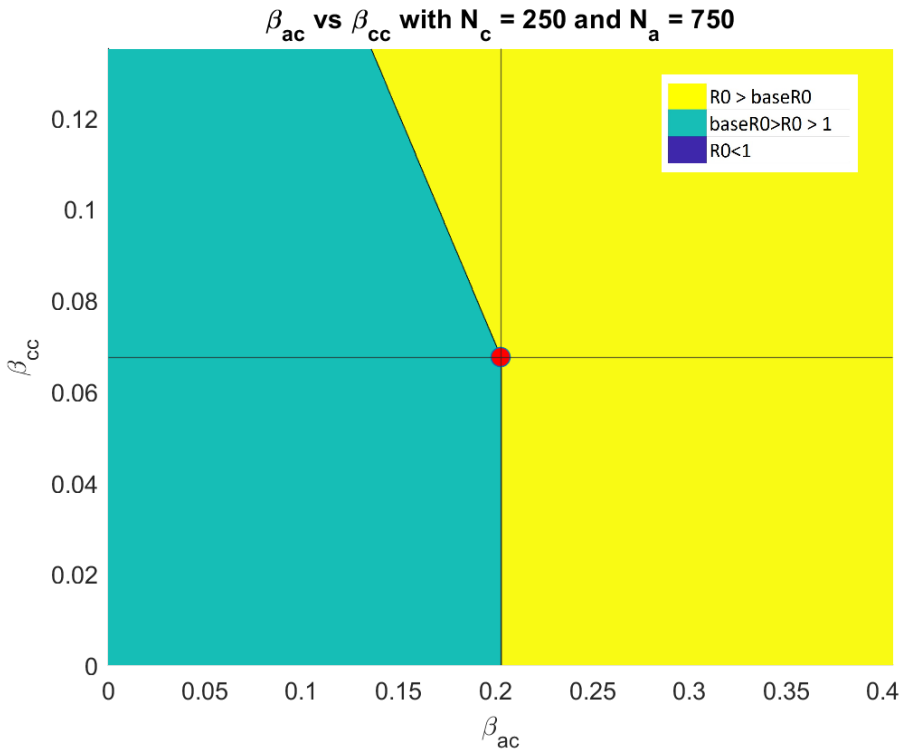
Base Case



Adult Child SEIR $R_0 = 3.1$ and base case SEIR $R_0 = 3.1$

Comparing R_0 : Adult to Child vs Child to Child





Isolated Child Populations

When our β values are not identical, what happens if we just look at one isolated population (just adults or just children)?

Standard SEIR	**Adult-Child Model**	**Value**
β	$\frac{\beta_{CC}}{N_C}$	0.00000108
ϵ	ϵ_C	0.3
γ	γ_C	0.1

How do we account for population size and the number of people interacting?

Comparing Infectious Results: Child Infections

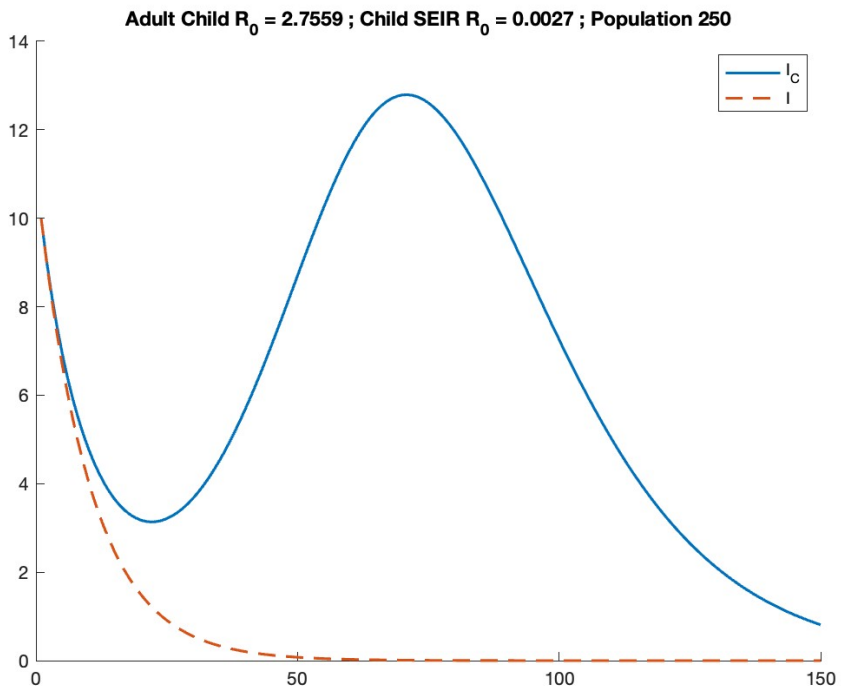


Figure: SEIR: $I(0) = 250$, SEIR-CA: $I_A(0) = 750, I_C(0) = 250$

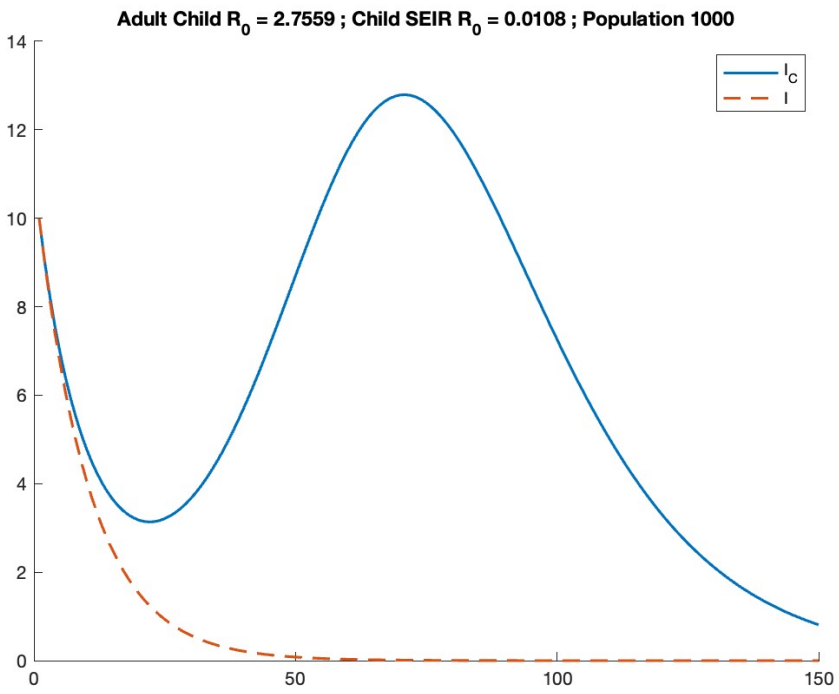


Figure: SEIR: $I(0) = 1000$, SEIR-CA: $I_A(0) = 750, I_C(0) = 250$

What is Identifiability?

Given data, we can create and fit any model. But can we use a model to predict data and understand a disease? If we ask a question, will the model give us multiple different answers or one true value?

Structural and Practical Identifiability

- **Structural Identifiability:** A model is structurally identifiable if the model parameters can be uniquely determined based on the model formulation with continuous data and no noise.
- **Practical Identifiability:** A model is practically identifiable if we can recover *reasonable* model parameters given incomplete or imperfect data.

Computing Structural Identifiability

We compute structural identifiability using **Identifiability Analysis** in Mathematica.

- works by computing a non-linear algebraic system, whose rank of the Jacobian determines the identifiability of the system.
- reduces computational complexity by directly computing the entries of the Jacobian matrix and by performing all calculations modulo a large prime number.
- uses local algebraic identifiabilty to produce a Boolean output where true means all parameters in the parameter set p are locally identifiable, while false means one or more parameters in the parameter set p are not locally identifiable.

Structural Identifiability 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: Structural Identifiability of SEIR Parameters Non-Infectious E

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: Structural Identifiability of SEIR Parameters: Infectious E

Practical Identifiability Motivation

- A model could be structurally identifiable, but it still could be practically unidentifiable due to quality of data, which often needs to be addressed in model parameter estimation.
- We focus on practical identifiability given prevalence data I_C, I_A
- Method: Monte Carlo Simulation
- Practical identifiability is a very local behavior, so we fix our true parameters at COVID specific values in our investigation.

Parameter	Value	Parameter	Value
μ	0.00008	f	0.0005
β_{CC}	0.01	β_{AA}	0.00027
β_{CA}	0.00005	β_{AC}	0.01
γ_C	0.1	γ_A	0.074
ϵ_C	0.3	ϵ_A	0.2

Parameter	Value
ξ_{AA}	0.0002
ξ_{CC}	0.0067
ξ_{CA}	0.000033
ξ_{AC}	0.0067

Practical Identifiability Motivation

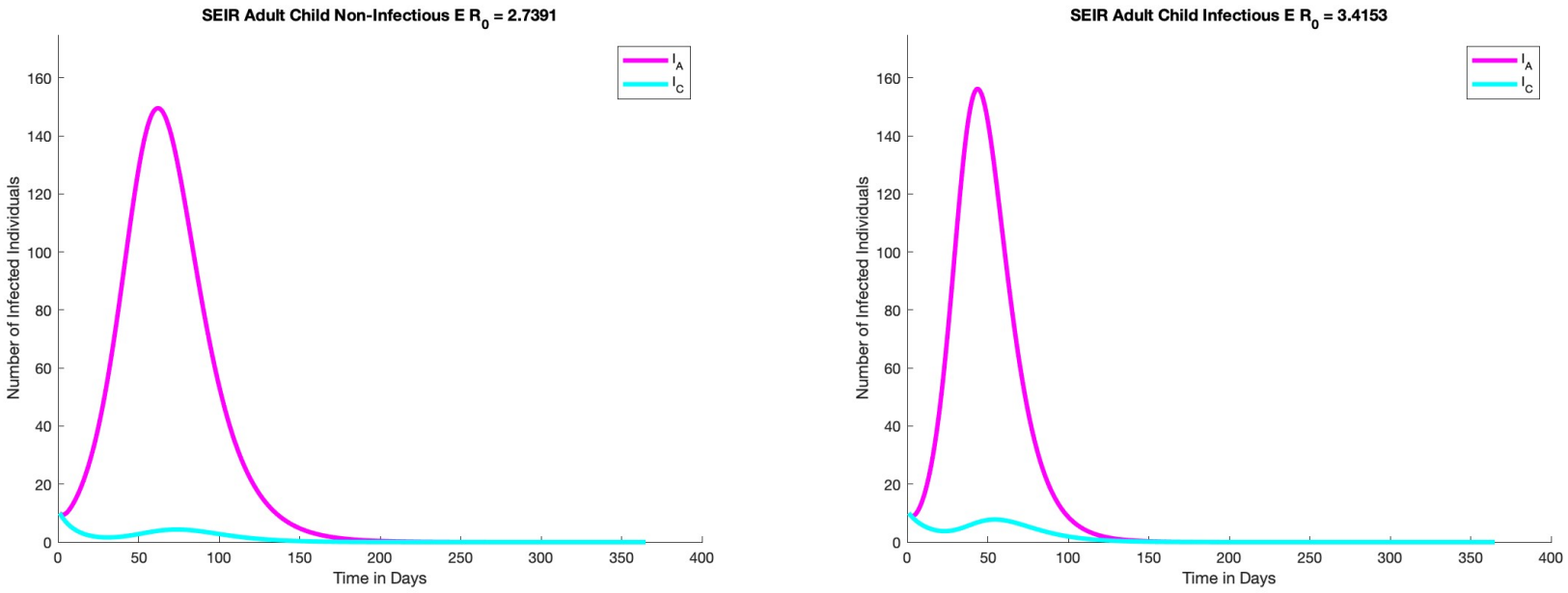


Figure: Infectious and non-infectious E SEIR adult child model dynamics with initial conditions $S_{C0} = 240$, $S_{A0} = 740$, $E_{C0} = 0$, $E_{A0} = 0$, $R_{C0} = 0$, and $R_{A0} = 0$.

Note: We adjust the time scale in our calculations to account for the differences in dynamics.

Computing Practical Identifiability

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**
    $ARE(p_\sigma^{(k)}) = 100\% \times \frac{1}{100} \sum_{i=1}^{100} \frac{p^{(k)} - \hat{p}_{\sigma i}^{(k)}}{p^{(k)}}$
  end for
end for

Def: A parameter set p is practically identifiable if $ARE(p_\sigma^{(k)}) \leq \sigma$ for all σ.

Identifiability of Adult-Child Non-Infectious E SEIR Model

σ	β_{AA}	β_{CC}	β_{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	γ_A	γ_C	ϵ_C	ϵ_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

Figure: For 365 days using (time of peak infection for children is at day 1 and time of peak infection for adults is at day 62)

Identifiability of Adult-Child Non-Infectious E Beta Dependencies

Parameter	Dependency	Parameter	Dependency
β_{AA}	β	$\frac{\beta_{CC}}{N_{C0}}$	0.1434β
β_{CA}	0.1912β	$\frac{\beta_{AC}}{N_{A0}}$	0.1667β

Figure: COVID specific infection rate dependencies

- Scale the contact rate and probability of infection for the other three according to their relationships to β_{AA}.
- β_{AA} is the the largest followed by β_{CA}, $\dfrac{\beta_{AC}}{N_A(0)}$, and $\dfrac{\beta_{CC}}{N_C(0)}$.

Identifiability of Adult-Child Non-Infectious E Beta Dependencies Continued

σ	β	γ_A	γ_C	ϵ_C	ϵ_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

Figure: For 365 days (time of peak infection for children is at day 71 and time of peak infection for adults is at day 62)

Conclusions

- Differentiating between adults and children allows us to determine unique disease dynamics for the respective groups.
- Interactions between populations significantly change infection peaks and can create secondary infection peaks
- Isolated, non-infectious populations may still experience an epidemic after interacting with a more infectious population
- Cumulative outputs are not structurally identifiable.
- Including relevant β and ξ dependencies produce a full set of practically identifiable parameters.
- Age-based model captures additional disease dynamics without losing identifiability properties.

Selected References

- 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.
- Michael T Meehan, Diana P Rojas, Adeshina I Adekunle, Oyelola A Adegboye, Jamie M Caldwell, Evelyn Turek, Brid- get 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.
- Hongyu Miao, Xiaohua Xia, Alan S Perelson, and Hulin Wu. On identifiability of nonlinear ode models and applications in viral dynamics. SIAM review, 53(1):3–39, 2011.
- Edward Goldstein, Marc Lipsitch, and Muge Cevik. On the effect of age on the transmission of sars-cov-2 in house- holds, schools, and the community. The Journal of infectious diseases, 223(3):362–369, 2021.
- Kimberlyn Roosa and Gerardo Chowell. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. Theoretical Biology and Medical Modelling, 16(1):1–15, 2019.
- Harvey Thomas Banks, Shuhua Hu, and W Clayton Thompson. Modeling and inverse problems in the presence of uncertainty. CRC Press, 2014.

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

- Xiuquan Nie, Lieyang Fan, Ge Mu, Qiyou Tan, Mengyi Wang, Yujia Xie, Limin Cao, Min Zhou, Zhuang Zhang, and Wei- hong 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.