The Pennsylvania State University
The J. Jeffrey and Ann Marie Fox Graduate School

A STOCHASTIC MODEL STUDYING INTERACTIONS BETWEEN BREASTFEEDING AND VACCINATION IN PROVIDING ROTAVIRUS IMMUNITY

A Thesis in
Statistics
by
John Mattiace

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2024

The thesis of John Mattiace was reviewed and approved by the following:

Murali Haran
Professor of Statistics
Thesis Advisor

Stephen Berg
Assistant Professor of Statistics

Runze Li
Eberly Family Chair Professor in Statistics
Chair of Graduate Studies

Abstract

It is well known that breastfeeding bolsters infant immune systems, but whether breastfeeding impacts long-term immunity acquired via vaccination is less well-studied. Of particular interest is whether breastfeeding actively assists or detracts from vaccine uptake in otherwise susceptible children. We consider data from the National Immunization Survey as well as the Healthcare Cost and Utilization Project and develop a discrete-time, stochastic compartmental model, explicitly accounting for the number of breastfed children, with the end goal of investigating the interaction between rates of breastfeeding, rates of infection, and rates of successful vaccination. Likelihood-based inference for this model is computationally challenging, so we resort to an approach based on Approximate Bayesian Computation Sequential Monte Carlo (ABCSMC) methodology to estimate parameters driving the model dynamics. We present preliminary findings from fitting our new stochastic model to these data.

Table of Contents

List of Figures

List of Tables

List of Symbols

α Interaction term between breastfeeding and successful vaccination, p. 11

β Rate of infection by rotavirus, p. 11

γ Rate of recovery from infection, p. 11

ω Probability of possessing maternal antibodies, p. 11

η Probability of breastfeeding, p. 11

ν Rate of being weaned, p. 11

θ Rate of waning maternal protection, p. 11

δ Rate of waning immunity, p. 11

σ Chance of successful vaccination, p. 11

κ Rate of vaccine coverage, p. 11

Acknowledgments

Thanks to the National Institutes of Health, which funded this and other related projects. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the NIH.

Chapter 1
Introduction

Even in our modern times of genetically-modified organisms and expansive supplement markets, breast milk remains the best, most nutritious nourishment for a human baby. The American Academy of Pediatrics recommends mothers to exclusively breastfeed their children until 6 months old. After that, mothers should continue to breastfeed, but also introduce solid foods, eventually weaning the baby from the breast entirely. Breast milk contains essential macronutrients such as protein, carbohydrate, and fat, but it also contains many micronutrients like vitamin D and amino acids. Researchers still continue to discover the nutritional benefits of breast milk [1]. In terms of preventative health benefits, breastfed infants appear to have a lower incidence of asthma, type 1 diabetes, and sudden infant death syndrome, to name a few. For the mother, breastfeeding seems to lower the risk of high blood pressure and ovarian and breast cancers [2]. Moreover, in the final trimester of a pregnancy, antibodies from the mother pass through the placenta to the fetus. These antibodies, though limited in lifespan, confer some amount of protection to the baby once it is born. Mothers can extend this protection through breastfeeding, as antibodies are transferred from the mother to the child via breast milk [3]. In sum, there are clear benefits to the mother and child; however, breastfeeding rates have differed across space and time. For example, in the United States, the Northeast tends to have high rates (around 80%), while the Midwest and some parts of the South have much lower rates (approximately 50-60%). The introduction of formula provided a paradigm shift, and people in these regions subsequently viewed breastfeeding as uncultured and unclean. This stigma remains extant within some communities to this day. To combat this, organizations such as the CDC have engaged in campaigns to popularize breastfeeding and extend support to struggling mothers [4]. Of key importance is breastfeeding bestowing protection from infection. To examine this phenomenon further, we focus our attention on rotavirus, a pervasive but largely harmless disease (in the U.S. and developed world).

Rotavirus, a double-stranded RNA virus, is one of the most common causes of diarrhea among young children. In fact, rotavirus infects nearly all children at least once by the age of five. We know further that it is possible to be infected multiple times (whether that occurs due to waning immunity or due to exposure to a different strain of the virus, the outcome is the same); in many analyses, three typifies the maximum number of infections. Typical infection periods last about a week , and the typical route of infection is the fecal-oral route [5].

Though it accounts for roughly one third of all diarrhea-related hospitalizations of infants and young children, its role in undermining public health goes virtually unannounced. The global burden of disease includes, a 2007 paper estimates, 450,000 to 700,000 deaths with almost two million more hospitalizations every year for children under five years of age [6]. The overwhelming majority of these cases occur in the developing world; the United States annually, by contrast, had roughly 60,000 hospitalizations and 37 deaths in the pre-vaccination era [6] [7]. In the post-vaccination era, with developed countries having access to the vaccine, this gulf has only widened.

In the West, two vaccines exist for rotavirus: RotaTeq, a three-dose vaccine, emerged first in February 2006, while Rotarix, a two-dose vaccine, entered the US market in April 2008. Both vaccines require oral administration. Countries that adopted rotavirus vaccination as part of a routine childhood vaccination program saw several benefits to public health. Notably, studies in eight of those countries saw a 49–89% reduction in lab-confirmed gastroenteritis hospitalizations and a 17–55% decline in all-cause gastroenteritis hospitalizations among children under five years of age, within 2 years of vaccine introduction [8]. The rotavirus vaccination of young infants in some countries also led to the decline of disease among children who missed vaccination—including older children and even adults who were not vaccinated. Most vitally, studies from Mexico, Panama, and Brazil show a reduction in childhood deaths from all-cause diarrhea following vaccine implementation. While the goal is to distribute the vaccine to populations in Africa and Asia, where it would do the most good at this point, researchers and public health officials still hold some concerns. Since either form of the vaccine is orally administered, its efficacy may be negatively impacted by concurrent enteric infections, higher rates of malnutrition and comorbidities, and greater levels of maternal antibodies, which are risk factors in these populations. [9]

To briefly elaborate, newborns carry maternal antibodies (that is, antibodies for pathogens the mother has been exposed to during her life) as a side effect of sharing a blood supply with the mother while in the womb. These offer some protection for

the lifespan of the antibodies (6-12 months), but this protection wanes until the child is able to produce antibodies for itself. For that to occur, the child must be exposed to the pathogen of interest–in this case, rotavirus. In populations without the vaccine, infection rates are naturally higher, so it is more likely that mothers possess antibodies to rotavirus, and therefore, it is more likely that their children have maternal antibodies for rotavirus.

Therein lies the motivation for this endeavor. While it seems natural to conclude from the preceding information that infants ought to be breastfed and to receive some form of rotavirus vaccine, this might not be the case in developing countries. If the higher levels of maternal antibodies decrease the effectiveness of the vaccine, then the only protection the infants have is short-lived, lying in those still-active maternal antibodies.

There exist some conflicting experimental studies on the potential interactions between the antibodies provided through breastfeeding and the protection conferred from the vaccine [10] [11] [12]. Here, we use observational data. In spite of the numerous benefits to breastfeeding, we should, in good conscience, investigate potential negatives warranting mitigation with respect to maternally-conferred immunity. Of interest are the following two questions: 1) Does breastfeeding have a significant impact on rates of rotavirus infection?, and 2) Is there a detectable interaction between chance of successful vaccination against rotavirus and protection offered by breastfeeding? Since no formal experiments have been conducted with regard to the questions at hand we must use observational data to draw a conclusion of some kind. The remainder of the paper is divided as follows: the next section discusses the considered data, then a discourse on the proposed method of analysis, then an appraisal of preliminary results, then a summary of challenges, concluding with thoughts for future work.

Chapter 2
Data

We have introduced the problem: estimating the effect of breastfeeding on both infection by rotavirus and successful vaccination against said infection. In the following section, we detail the data used to approach these questions. Importantly, we note that these data are observational, and not experimental, in nature.

2.1 HCUP

Principal data come from the Healthcare Cost and Utilization Project (HCUP), a family of healthcare databases in collaboration with the U.S. Department of Health and Human Services. In particular, we consider the National Inpatient Survey (NIS). We examine a dataset of individuals with rotavirus as an admitting diagnosis from January 2002 to December 2016. Moreover, since the population of interest is children under the age of 5, we eschew any observations of those older than 4. A sample of the data is as follows.

Table 2.1. HCUP Data

	Age	Census_Division	Hospitalizations	Race	Date
1	0	9	14	1	2002-01-01
2	0	9	6	2	2002-01-01
3	0	9	53	3	2002-01-01
4	1	9	15	1	2002-01-01
5	1	9	4	2	2002-01-01
6	1	9	25	3	2002-01-01
7	1	9	2	4	2002-01-01
8	1	9	3	6	2002-01-01
9	2	9	4	1	2002-01-01
10	2	9	2	2	2002-01-01

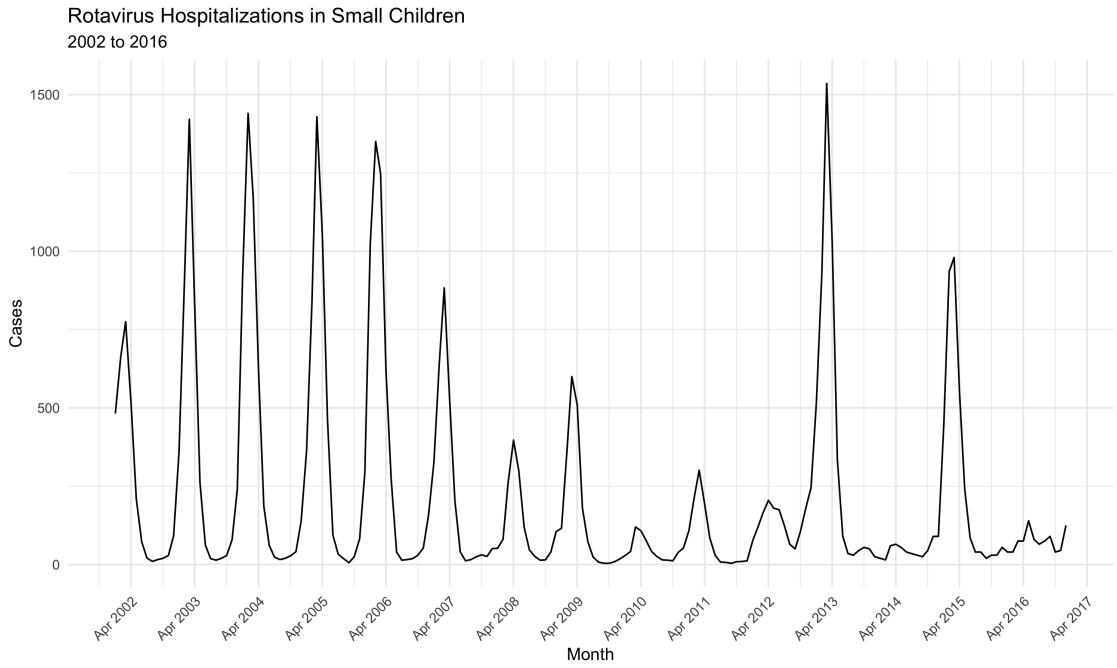

Figure 2.1. Cases of Rotavirus in the United States from 2002 to 2016

Here we see that hospitalizations is an integer-valued variable that counts the total number of admissions due to diarrhea during a given month. The variables race and census division are both categorical, meant to code a specific race and a specific geographic region of the United States. There are 6 levels of race (e.g. 1 represents White) and 9 levels of census division (e.g. 9 corresponds to California, Oregon, Washington, Alaska, and Hawai'i). In sum, we have monthly counts of rotavirus hospitalizations aggregated by census region. Naturally, the aggregation to region-level results from efforts to protect healthcare information privacy.

An additional wrinkle to the HCUP data lies in a change to the sampling scheme over the years: prior to 2012, the NIS took all discharges from a sample of hospitals across the different census divisions; whereas from 2012 on, the NIS takes a sample of all discharges from all hospitals across the different census divisions. For clarification, all hospitals means all hospitals participating in the NIS; which hospitals do and do not participate is not available to us. This move ultimately provides better estimates, since different hospitals tend to have different populations of patients in the long run (depending on specialty and regional demographics). We can see the tangible results of this shift in sampling policy in Figure 2.1.

We see regular, annual rotavirus epidemics before 2007, which, in lieu of the release of the vaccine, makes sense. The consistency of the peaks does capture one's attention.

After 2007, the vaccine had widely become available and administered in the United States, so the peaks become much smaller, though still temporally in line with the previous epidemics. Then, after a handful of years, the time series shows a supposed marked increase in hospitalizations. Epidemiologically, an explanation might go like this: several years went by without a large outbreak, which means most children went unexposed to the disease and thus were susceptible to infection. This population of susceptibles grew so large that, when the next outbreak occurs, that outbreak would be larger than any "ordinary" outbreak in recent history. As an analogy, a forest that has not undergone even small fires for years can accumulate vast quantities of dead wood such that, when a fire does eventually happen, it will be much larger than usual. Though this offers a logically valid explanation, epidemiologists have no reason to believe anything like this happened, as vaccination coverage only increased over the years. Rather, it remains much more likely this is a quirk resulting from the switch in sampling methods. One final feature worth noting is the apparent triannual epidemics with very small outbreaks occurring between them. This demonstrates the efficacy of the vaccine.

2.2 CDC Breastfeeding Survey

As a second data component, we include results from the CDC's National Immunization Survey, conducted yearly, which attempts to ascertain vaccination coverage among children and teenagers. For mothers of newborn or young children, an ancillary component of the questionnaire addresses breastfeeding behavior. Some of the questions asked address, for example, whether the child was ever breastfed and the age at which it occurred. A snapshot of these data looks like the following.

Table 2.2. CDC Breastfeeding/National Immunization Survey Results

	Bf_Init	Bf_6mo	E_bf_6mo	R_3doses	R_2doses	State	Date
1	1	0	0	0.00	0.00	Texas	2003-01-01
2	1.00	0	0			Louisiana	2003-01-01
3	0	0	0			Arizona	2003-01-01
4	1	1	0			California	2003-01-01
5	1	0	0	0.00	0.00	Alabama	2003-01-01
6	1	1	1	0.00	0.00	Arizona	2003-01-01
7	0	0	0	0.00	0.00	Tennessee	2003-01-01
8	1	1	0	0.00	0.00	Tennessee	2003-01-01
9	1	1	0	0.00	0.00	Washington	2003-01-01
10	0	0	0	0.00	0.00	Maryland	2003-01-01

As neither vaccine was available to the public before 2006, there are no observed vaccinations during this time period. In fact, only a handful of vaccination observations exist before 2008. All variables of interest except state and year are binary. *Breastfeeding_initiation* refers to whether the child was ever breastfed; *breastfeeding _6months* refers to whether the child was breastfed at least until it was 6 months old; and *exclusive_breastfeeding_6months* refers to whether the child was given food other than breast milk or formula during that first 6 months. One positive feature of these data, though there are many missing values related to vaccination, is that we have individual responses at the state level. An unfortunate yet not fatal negative of these data concerns the yearly time scale in comparison to the monthly time scale of the hospitalization counts. This means we must assume constant (or close to it) breastfeeding rates throughout each given year, though they may, in reality, change during the year.

Data available upon request.

Chapter 3
Model

Now that we have explored the available data, we proceed with first an overview of compartmental models and second a modification in the form of "subcompartmental" models, meant to assist in estimating the effect that breastfeeding has on rotavirus infection and vaccination.

3.1 Compartmental Models

Perhaps the most ubiquitous model in epidemiology, thanks to its power and flexibility, is the Susceptible-Infected-Recovered/Removed (SIR) model. Essentially, the SIR model partitions a population into three different states: first, there are those who do not have sufficient immunity and are thus susceptible to being infected by a pathogen; second, there are those who just were susceptible, but are now infected by said pathogen for a period of time; and third, there are those who have recovered from their illness, either acquiring immunity (in which case they cannot be infected again, and so do not become susceptible) or not (in which case they return to the susceptible state), depending on the disease. We can represent this relationship in the following figure.

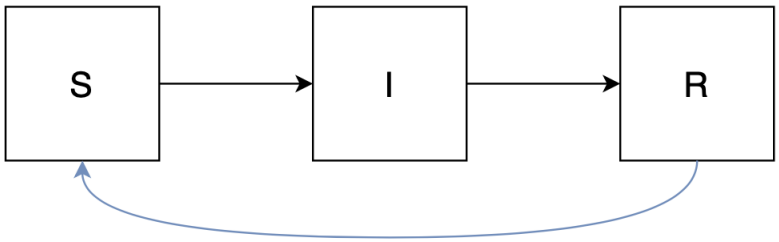

Figure 3.1. A Basic Compartmental Model for Infectious Disease

While the move between R and S is optional, something we will consider in future work, for all intents and purposes, the moves between S and I, I and R are not optional. That said, this being a stochastic model, an individual may simply by chance evade infection for the length of the study or observation period, but this would only belie the true state of affairs. Given its widespread use and recognized utility, it is natural to proceed with this as our modeling framework.

3.2 "Subcompartmental" Models

Returning to the problem of interest, we are interested in investigating infection differences between those children who are breastfed and those who are not. Moreover, we want to get an idea of any interaction between successful vaccination and breastfeeding. We therefore must add a "vaccinated" state, and further, we need some way to delineate subpopulations of breastfed/non-breastfed children. Therefore, we propose the following model.

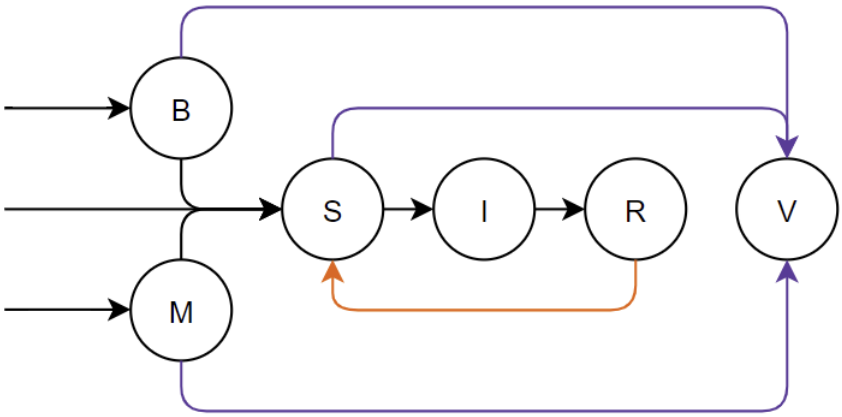

Figure 3.2. An Initial "Subcompartmental" Model

Infants, according to this schema, are born and have some probability of starting in B, M, or S. B denotes the "breastfed" state while M denotes the "maternal antibodies" state. From any of the pre-infected states, you can enter the "successfully vaccinated" state V.

As aforementioned, newborns carry maternal antibodies (that is, antibodies for pathogens to which the mother has been exposed) as a side effect of sharing a blood supply with the mother while in the womb. These offer some protection for the first

6-12 months of life, but wane eventually. Until the child has been exposed to the pathogens themselves, the child will be unable to manufacture their own antibodies, so these maternal antibodies offer limited protection [13]. If the mother has antibodies against rotavirus and breastfeeds her infant, the antibodies are transmitted through the breast milk to extend that protection [3]. The mother eventually weans the child from breast milk and any maternal protection will subsequently cease, at which point the infant enters the susceptible state. Otherwise, the mother does not possess antibodies for rotavirus, so whether she breastfeeds does not matter, and the infant resides in the susceptible state from the beginning.
In summary, the modifications to the traditional compartmental model are as follows:

- if the mother was never infected with rotavirus, whether or not she breastfeeds, then the infant goes immediately to S upon birth;
- if the mother was infected with rotavirus, but does not breastfeed, then the infant goes to M and eventually S, at the rate of those antibodies fading;
- if the mother was infected with rotavirus and does breastfeed, then the infant goes to B and eventually S, at a slower rate than M to S;
- if at any point the child is vaccinated before infection, then the infant goes to V and stays there (in other words, vaccination prevents infection in this model).

Finally, the black lines denote what would happen if no intervention were present, whereas the purple represents that intervention (through vaccination). The orange shows a potentiality of reinfection. Note that it is possible to be vaccinated in any of the pre-infection states, and once an individual is vaccinated, they cannot leave state V. We additionally make the assumption that children do not die from infection, as the number of deaths per year are negligible to the broader population.

Since disease transmission occurs on a continuous time scale, SIR models favor mathematical description through Ordinary Differential Equations (ODEs), where the quantities of interest are described as rates (in the simplest models, constant throughout time). For example, in a closed population,

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

where the left-hand side denotes a change in the states S, I, and R over time. Even though these equations appear simple, they must be solved numerically, not analytically [14]. By extension, the differential equations for the full breastfeeding model look like

$$\begin{aligned}
\frac{dB}{dt} &= \omega\eta G - \nu B - \alpha\sigma\kappa B - \mu B \\
\frac{dM}{dt} &= \omega(1-\eta)G - \theta M - \sigma\kappa M - \mu M \\
\frac{dS}{dt} &= (1-\omega)G + \nu B + \theta M + \delta R - \beta\frac{SI}{N} - \sigma\kappa S - \mu S \\
\frac{dI}{dt} &= \beta\frac{SI}{N} - \gamma I - \mu I \\
\frac{dR}{dt} &= \gamma I - \delta R - \mu R \\
\frac{dV}{dt} &= \alpha\sigma\kappa B + \sigma\kappa M + \sigma\kappa S - \mu V,
\end{aligned}$$

where G denotes the births during some time interval, μ acts as a constant rate of removal from the population to balance the incoming births across all states (since eventually children age outside our purview), α serves as an interaction effect between B and V, and $N = B + M + S + I + R + V$, where N is the population of children under five. In particular, α is nonnegative–breastfeeding may completely hinder the vaccine from taking hold, or it may bolster its efficacy. This parameter essentially would modify the uptake of the vaccinated from state B versus M or S. That being said, the "typical" uptake rate of $\sigma\kappa$ is not identifiable on its own, and would likely require domain knowledge with respect to vaccine coverage (κ) to infer σ. Other parameters' definitions are located in the front matter. The purple text highlights the components of the model that relate to a vaccinated state, the orange to a reinfection rate, and the black to a more straightforward SIR process.

Naturally, the purpose behind any model lies in its ability to fit to data. Immediately an issue arises: we have monthly counts for a disease whose infectious periods (sometimes called "generations") last only one week on average. In other words, by the time we see a single observation, four generations have come and gone. The counts, therefore, are

aggregated not only by census division, but also with respect to time. In order for us to use this model, we would have to upscale the data from a monthly timeframe to a continuous one, which poses its own unique challenge. With this in mind, we consider changing our approach from one rooted in continuous-time assumptions to one rooted more in discrete-time ones.

We therefore turn to the Poisson point process, a counting process that in this case records the number of transitions from one state to another, dependent on the rate of transmission, over an interval of time. That is, the Poisson process would count the number of transmissions from S to I as the number of infections that occurred during that time period, and likewise, the number of transmissions from I to R as the number of recoveries that occurred during that time period. The Poisson process possesses a number of pleasing attributes, such as 1) independent increments and 2) the number of events in any interval of length t being distributed Poisson with mean λt [15]. This means we can shift our interpretation of time to a day to day, week to week, or better yet, month to month (as this matches our data), and the Poisson random variable emulates the stochasticity of the disease transmission over the interval length of interest.

In essence, what we hope to achieve is an approximation of this continuous-time phenomenon as follows (in slightly abusive notation).

$$\begin{aligned}\frac{dB}{dt} &= \omega\eta G - \nu B - \alpha\sigma\kappa B - \mu B \\ &\approx \text{Binom}(G, \omega\eta) - \text{Pois}(\nu B\Delta t) - \text{Pois}(\alpha\sigma\kappa B\Delta t) - \text{Pois}(\mu B\Delta t) \\ &= \text{Binom}(G, \omega\eta) - \text{Pois}((\nu + \alpha\sigma\kappa + \mu)B\Delta t) \qquad (1)\end{aligned}$$

for small Δt, where now G represents the number of births over an interval of length Δt (and we consider a binomial random variable to ensure that G is split according to the rates of breastfeeding and maternal antibody presence). As a result of the Poisson process assumption, we may take each component of this overall counting process as independent Poisson random variables. We can combine these three as the negated sum of three independent Poisson random variables, given a Δt, and can simply regard this change in B as a single binomial random variable minus a single Poisson one.

One may take issue with the assumption of independent increments: since the number of susceptibles must necessarily decrease in accordance with the number of infected from generation to generation (though it may grow overall with new births), if, say, 90% of

the susceptibles were infected the previous generation, clearly, it will not be possible for the disease to infect a large number of susceptibles in the coming generation. While true, we assume the rates of infection, recovery, and so on themselves to be constant, and they are the quantities of interest. The states' numbers will fluctuate over time because the states "downstream" necessarily rely on the states "upstream," seen through several states over many generations.

Since we are interested in performing inference on chiefly α, it follows that we must obtain the distributions of counts on states that involve it as a parameter. That leaves states B and V as our primary targets; unfortunately, however, these are latent variables. We simply do not have access to even useful estimates of these states. Again, from our infection data, we only have access to a sample of the hospitalizations that result from rotavirus, aggregated over space and time. Given these complications, we can simplify in the short term by analyzing synthetic data consistent with the pre-vaccination period (thus, V is not a state and α is not a parameter). Any model that can tackle the real data should be able to do so with higher-resolution synthetic data.

3.3 Ordinary Differential Equation-Simulated Data

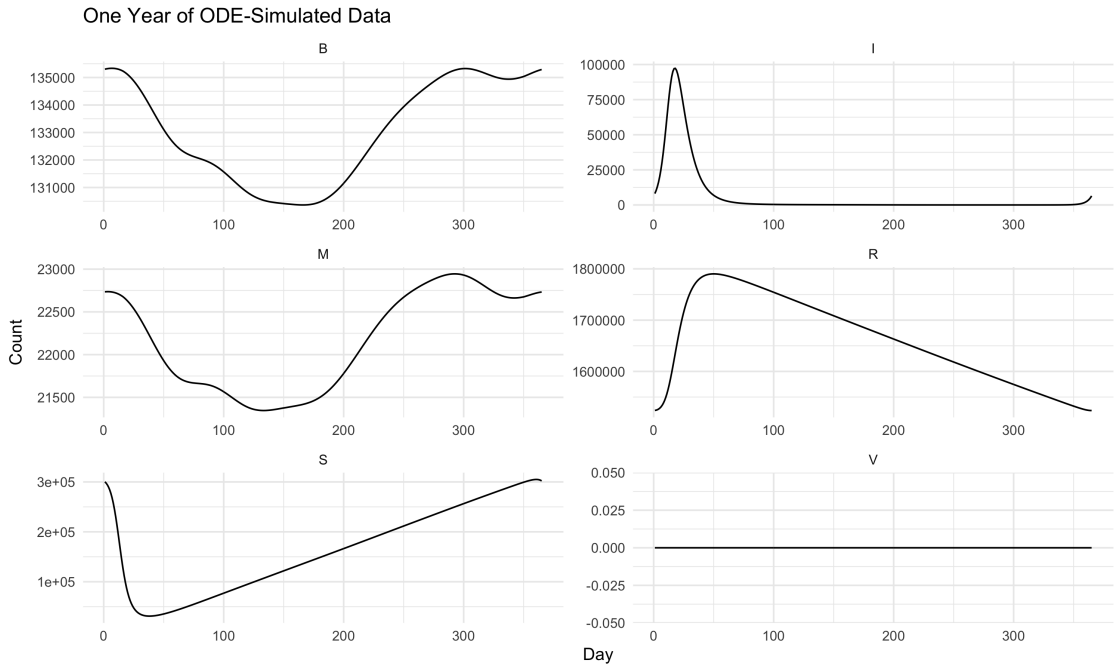

Figure 3.3. Simulated State Data

Because of the complexity in the hospitalization data (spatial and temporal aggregation, accounting for pre- and post-vaccine effects, accounting for a change in survey methodology), our first step, and the remainder of the paper, will center on some simplified simulated data with much higher granularity. In particular, these data arise from a series of ordinary differential equations, discussed earlier. These are notably derived from a single state's (Texas, 2002) parameters, and they provide daily, rather than monthly, counts of all states of interest.

This leaves us with the other parameters, and if we can estimate them, then that will make it easier to estimate α in considering the full model. To reiterate, we know γ and μ through domain knowledge and assumption respectively; what we want to know is ω, ν, η, θ and β. This leaves us with a *BMSIR* disease model.

Let us then formally define the sub-compartmental model:

- We have the following random variables
  - Transfers $T_{B,t}$, $T_{M,t}$, $T_{S,t}$, $T_{I,t}$
  - Births $G_{B,t}$, $G_{M,t}$, $G_{S,t}$
  - Removals $O_{B,t}$, $O_{M,t}$, $O_{S,t}$, $O_{I,t}$, $O_{R,t}$
- with respective distributions
  - $T_{B,t} \sim \text{Pois}(\nu B_s \Delta t)$
  - $T_{M,t} \sim \text{Pois}(\theta M_s \Delta t)$
  - $T_{S,t} \sim \text{Pois}(\beta \frac{S_s I_s}{N} \Delta t)$
  - $T_{I,t} \sim \text{Pois}(\gamma I_s \Delta t)$
  - $[G_B, G_M, G_S]_t \sim \text{Multinom}(G_{\Delta t}, \{\omega\eta, \omega(1-\eta), 1-\omega\})$
  - $O_{B,t} \sim \text{Pois}(\mu B_s \Delta t)$
  - $O_{M,t} \sim \text{Pois}(\mu M_s \Delta t)$
  - $O_{S,t} \sim \text{Pois}(\mu S_s \Delta t)$
  - $O_{I,t} \sim \text{Pois}(\mu I_s \Delta t)$
  - $O_{R,t} \sim \text{Pois}(\mu R_s \Delta t)$

where s and t denote times, and $s = t - \Delta t$.

Putting this all together, we have

$$\frac{\Delta B}{\Delta t} = G_{B,t} - T_{B,t} - O_{B,t}$$

$$\frac{\Delta M}{\Delta t} = G_{M,t} - T_{M,t} - O_{M,t}$$

$$\frac{\Delta S}{\Delta t} = G_{S,t} + T_{B,t} + T_{M,t} - T_{S,t} - O_{S,t}$$

$$\frac{\Delta I}{\Delta t} = T_{S,t} - T_{I,t} - O_{I,t}$$

$$\frac{\Delta R}{\Delta t} = T_{I,t} - O_{R,t}$$

all that remains to do inference is finding the joint likelihood function of these five variables.

Chapter 4
Methods and Results

To this point, we have discussed the problem, the data, and the model of choice to answer the research questions at hand–namely, estimating the extent to which breastfeeding impacts infection by and successful vaccination against rotavirus in infants. We aim to use a subcompartmental model as the foundation, while we translate a typical ODE disease model to a Poisson point process. In the following section, we delve into its use to perform inference via such a model.

4.1 Likelihood-Based Inference

In the realm of likelihood-based inference—that is, an approach in which one optimizes the titular likelihood equation with respect to some parameters—there persist two main schools of thought (frequentist vs. Bayesian). We elect a Bayesian approach for its straightforwardness, summarized by

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta})$$

where the left-hand side denotes the posterior distribution of some vector of parameters, the right-hand side, the product of the likelihood and prior distributions, and the proportionality stems from an unknown constant coefficient. The ultimate goal is to sample from the posterior distribution. This is commonly done through Markov Chain Monte Carlo (MCMC) in conjunction with the Metropolis-Hastings (MH) algorithm. The MH algorithm functions in the following way: we first randomly generate a proposed parameter value (or values if many parameters are under consideration), then evaluate the likelihood and prior with that value, and if this current proposal leads to a more optimized product, then we accept it with varying probability, proportional to how much

better the proposal is than the current step in the chain. The algorithm is included in the appendix [16].

Because of this need to evaluate the likelihood function to determine the validity of the proposal, much of the computational viability of using MCMC in this way depends on the complexity of that likelihood function. If the dimensionality markedly increases, or if there are sufficiently taxing terms in the function from the outset, then the efficiency of this algorithm plummets quickly [17]. Considering only the first two states (that is, B and M), we have

$$\pi(\omega, \eta, \nu, \theta | \mathbf{B}, \mathbf{M}) \propto \mathcal{L}(\mathbf{B}, \mathbf{M} | \omega, \eta, \nu, \theta) \pi(\omega, \eta, \nu, \theta)$$

$$\mathcal{L}(\mathbf{B}, \mathbf{M} | \omega, \eta, \nu, \theta) = \prod_{t=\{\Delta t, 2\Delta t, \ldots\}} [f((\Delta \mathrm{B})_t | \mathrm{B}_s)] \times [f((\Delta \mathrm{M})_t | \mathrm{M}_s)]$$

$$f((\Delta \mathrm{B})_t | \mathrm{B}_s) = \sum_{\{G_{B,t} - T_{B,t} - O_{B,t} = (\Delta \mathrm{B})_t\}} f(G_{B,t}, T_{B,t}, O_{B,t} | \mathrm{B}_s, \mathrm{G}_t, \omega, \eta, \nu)$$

$$f((\Delta \mathrm{M})_t | \mathrm{M}_s) = \sum_{\{G_{M,t} - T_{M,t} - O_{M,t} = (\Delta \mathrm{M})_t\}} f(G_{M,t}, T_{M,t}, O_{M,t} | \mathrm{M}_s, \mathrm{G}_t, \omega, \eta, \theta)$$

In other words, we would have to sum over the combinations of possible events to calculate the likelihood for each time step. This may be feasible when the numbers are small, but the U.S. sees over 3 million births annually, and many of those babies will belong to either B or M.

Further, in our fully realized model, we ideally want to infer ten parameters that drive the dynamics of the states over time (though it is more likely we lean on domain knowledge or extraneous estimates). We therefore need those states at our disposal. Since they do not appear in the dataset, we must create them as latent variables while running the algorithm, and with sufficiently small time intervals, that means we must have thousands upon thousands of guesses for each given state over time. Hence, the dimensionality will increase dramatically over the course of tweaking and updating the algorithm to match the model. Therefore, a straightforward application of Metropolis-Hastings Markov Chain Monte Carlo (MCMC) will not work here, thanks to the complexity of the likelihood function; however, alternatives to the standard exist.

Ignoring the myriad latent variables in the full model for the moment and considering

only the hospitalization time-series data, it stands to reason that these data represent measurements of a fundamentally hidden process. The true number of infectives each generation lies under the surface, and what we actually observe—that is, the number of cases reported by either doctors or insurance companies—represents a biased sample of that true number. Hence, some process generates the true number of cases, and then some other process dictates the number of cases we observe, and this happens for each generation. If we can assume that the underlying process generating the true number of infectives is Markov, then what we obtain is a model concerning a partially-observed Markov process (POMP), or hidden Markov model (HMM).

We can, viewing the problem this way, recognize the merit behind sequential Monte Carlo (SMC). Essentially, sequential Monte Carlo attempts to sidestep direct evaluation of the likelihood function, which incorporates all the data simultaneously (this might be very difficult or even impossible), through building a numerical approximation. We establish this approximation through many particles' trajectories over the sample space, one time-step at a time (hence, sequential). Put simply, it is much easier to write and evaluate the probability model for a nonlinear process considering only one step to the next rather than the entire chain at once. The particles represent trajectories of the process under the probability model, and if the probability model is correct, then with favorable odds, it should happen that the particles tend to stick closely to the trajectory seen in the real data. In our pursuit, we clearly have a well-defined model when looking only one time step ahead, but a much more complicated model when considering multiple or many time steps ahead. Sequential Monte Carlo, therefore, might serve well.

Put formally, SMC will help us compute $f(\mathbf{B}, \mathbf{M}|\omega, \eta, \nu, \theta)$, but through handling the log-likelihood, or $\log f(\mathbf{B}, \mathbf{M}|\omega, \eta, \nu, \theta) \equiv \ell(\mathbf{Y}|\omega, \eta, \nu, \theta) \equiv \ell(\omega, \eta, \nu, \theta)$. Consider

$$\ell(\omega, \eta, \nu, \theta) = \sum_{t=1}^{T} \ell_{t|1:t-1}(\omega, \eta, \nu, \theta) = \sum_{t=1}^{T} \log f(Y_t|Y_{1:(t-1)}, \omega, \eta, \nu, \theta) = \log f(\mathbf{Y}|\omega, \eta, \nu, \theta).$$

where t represents time increments. We can further rewrite this as

$$\begin{aligned}\ell_{t|1:t-1}(\omega, \eta, \nu, \theta) &= \log f(Y_t|Y_{1:(t-1)}, \omega, \eta, \nu, \theta) \\ &= \log \int f_{Y_t|X_t}(Y_t|X_t; \omega, \eta, \nu, \theta)) f_{X_t|Y_{1:(t-1)}}(X_t|Y_{1:(t-1)}; \omega, \eta, \nu, \theta) dX_t\end{aligned}$$

where $Y_t = (B_t, M_t)$ and X_t is the true, unobserved number of infectives. The entire

enterprise, then, centers around sampling from the (target) distribution $f_{X_t|Y_{1:(t-1)}}(X_t|Y_{1:(t-1)};\omega,\eta,\nu,\theta)$ in order to approximate this integral. The particles, tracing high-plausibility areas, are the vehicle through which this occurs effectively [18] [19].

While this solves our issue with an intractable likelihood, we still wish to sample from the posterior distribution if we are to continue in Bayesian inference. SMC provides a numerical approximation to the likelihood for a given parameter or set of parameters, but does not include a mechanism for proposing and accepting/rejecting those proposals [18] [19]. For that, we turn to back to MCMC and, putting the two together, we obtain Particle MCMC. Particle MCMC simply takes the MH algorithm, and, rather than using the analytical solution of the likelihood function, uses the representation born of SMC to assess those proposals. Hence, we eventually obtain a posterior distribution. The algorithm for this approach is listed in the index [17].

Naturally, dynamic, nonlinear systems can effortlessly slide into unpredictability, especially as the number of time steps increases. As such, it is unsurprising that a misspecification in the model can cause the particles' trajectories to veer far from that of the real data. This leads to the offending particles being discarded as "bad" proposals and only a very small few accepted as "good" samples. This phenomenon is especially common when the hidden state is high dimensional. This happened to be the case here, as each particle accounted for, in the base model, multiple hidden states with multiple hidden generations from one observation to the next. In short, a vanishingly small number of particles survived even one year's worth of trajectory simulation, leading to the likelihood approximation rapidly deteriorating, leading to an inability to effectively assess the proposals, finally leading to a degenerate posterior distribution. One must finely tune such an algorithm in high dimensions and employ a well-fitting proposal distribution to offset the high computation cost.

4.2 Simulation-Based Inference

Therefore, we choose to abandon the likelihood-based approach in favor of a likelihood-free one, which becomes ever more feasible with every increase in available computing power. In order to remain in the Bayesian paradigm while switching to a likelihood-free method, we turn to Approximate Bayesian Computation (ABC), a technique currently growing in popularity in applied sciences [20] [21] [22]. The value proposition of ABC is the following: while the likelihood function characterizing the data-generating process may be

challenging, computationally inefficient, or even impossible to evaluate directly, we ought to have a conception of what kind of random process might simulate data that "look like" the real data [23]. Whereas this notion of "looking alike" has a precise meaning in likelihood-based approaches (e.g. comparable sufficient statistics), it becomes looser when the likelihood is not available to us. The type of statistic desired will vary depending on the problem [24]. For example, one may want each observation of the pseudo-data to match an observation of the real data: this is more feasible in some contexts, like low-dimensional datasets, than others. We could, if faced with data of higher dimension, attempt to reduce that high-dimensionality through the use of summary statistics and then compare the summary statistics of the pseudo-data to the summary statistics of the real data. One must take caution in doing so, as the efficacy of this approach hinges on the capacity of those elected summary statistics to characterize the distribution of the high-dimensional data.

Once a random process and statistic(s) have been chosen to simulate and evaluate the data respectively, eventually we should obtain a simulated sample that resembles the real data close enough that we are satisfied, where "enough" is controlled by some explicit threshold. We count the parameters that generated that reminiscent simulation as "good," and add them to our approximate posterior distribution. The accuracy of this approximation to the posterior distribution depends explicitly upon the chosen threshold/tolerance level. This is commonly referred to as the ABC rejection algorithm [25].

More formally, we have

- our real data $\mathbf{x}$,
- some probability model,
- some statistic $T(\cdot)$,
- some tolerance level ϵ,
- and some distance function ρ.

Using the probability model, we simulate a candidate dataset $\mathbf{y}$ with the goal of achieving

$$\rho(T(\mathbf{x}), T(\mathbf{y})) \leq \epsilon$$

This means that the specified feature(s) of the simulated data correspond well enough to the real data, as dictated by the threshold, ϵ. A good parameter proposal $\tilde{\theta}$ generates a

pseudo-dataset whose T-statistic is at most ϵ away from the T-statistic of the real dataset. Note, as $\epsilon \to 0$, we sample from the true posterior. This fundamentally changes the way in which we evaluate proposals within the Monte Carlo framework. The algorithm for ABC rejection sampling is included in the appendix [25].

Turning to the problem at hand, we have already mentioned the real data and the probability model (with which we shall simulate pseudo-data). Our T-statistic of choice changed with time: initially, we considered use of the identity function, so that the simulated time-series would have to come sufficiently close, again, controlled by the threshold, to the trajectory of the real time-series data to be considered acceptable. Considering that the dimensionality of the pseudo-data is 36,500 (if $\Delta t = 1/100$) by 2, 5, or 6 states (simulating BM versus $BMSIR$ versus $BMSIRV$), it stands as unsurprising that the dimensionality of the data was high enough that simulating an acceptable trajectory, especially as we lowered the threshold, became akin to throwing darts at a dartboard, blindfolded, at a distance of 30 meters or more. Therefore, we chose to consider some summary statistics to reduce that burden. To choose the summary statistics, we considered a slew of basic descriptive statistics such as mean, variance, minimum, maximum, and even some functions of the mean and variance in kurtosis, skew, and mean-to-variance ratio to try and better match the pseudo-data to the real data. While at some point we considered principal components of the series, instead we opted for summary statistics thanks to their interpretability. Now, since we simulate 2, 5, or 6 time series per iteration (1 for each state); that means we have 2, 5, or 6 sets of said descriptive statistics to contemplate. Likely only a small subset of these would provide useful information; we use regression to determine which of them deserve the most attention.

We conduct this regression through, in essence, a manual grid search (the best summary statistics being regarded as the optimizable quantity): we effectively grid sample over the space of the joint parameters by combining individual grid samples. For example, if a parameter has support from 1 to 10, and choose to take 19 values over that interval, our parameter value set will be {1, 1.5, 2, ..., 10}. We do this for every parameter involved in the probability model, and then we obtain every possible combination of joint parameter values. Then, we simulate pseudo-data based on these different combinations, compute the relevant summary statistics, and run an inverse regression where we use the values of the summary statistics to predict the values of the parameters. As a result, we find that some of the summary statistics explain much of the variation in the parameters, and we use these as our T-statistics when running the

ABC algorithm to provide flexibility. Which statistics reach significance in the regression depends on a few factors because changes in the prior distribution, which states are included, and the number of grid samples, to name a few, affect the distribution of the pseudo-data and therefore the distribution of the summary statistics as well. These significant summary statistics also provide us with the opportunity of fitting the real data to the regression model and using the fitted values of the parameters as sensible starting points for the ABC algorithm.

Now that we have chosen our summary statistics, we may proceed to discussing the tolerance level, ϵ. Note that for a run of the ABC rejection algorithm, the value of ϵ remains fixed since it dictates the approximation to the posterior, and hence, it is not dynamic. Therefore, we must decide a value of ϵ before running the algorithm. If our proposal distribution is overtuned and ϵ is very small, then our algorithm will reject the overwhelming majority of proposals, but because our threshold is so strict, we will obtain a high fidelity approximation to the posterior, given enough samples. On the other hand, if our threshold is too lax, then the opposite will occur. Additionally, a useful value for ϵ depends on the T-statistic(s) and its (their) scale(s). We can easily parallelize this operation at differing tolerance levels to see the breakpoints the efficiency/fidelity tradeoff for later increases in the sample size, but there is no guarantee that the (in)efficiency will be worth the fidelity or vice-versa.

Our final reagent, the distance function, also presents a burden of choice. Note first that this distance function need not be a metric; some instances of the ABC rejection algorithm utilize Kullback-Leibler divergence, for example [26]. In this matter, we opt for something simple and intuitive, with reasonable theoretical properties—k-dimensional Euclidean distance. While some have modified this with the use of weights in application, with weights meant to normalize the scales of the summary statistics, the sheer number of moving parts already present in this algorithm invite one to keep it simple [27]. We forgo any special weighting scheme in the distance function. One final wrinkle worth noting concerns, for multi-dimensional distance functions, whether to use an "intersection" or a "union" criterion when computing this distance (this was felt in particular during the use of principal components). In the former, we would require that each pair of respective artificial and real statistics measure less than ϵ apart; in the latter, we would require that the distance from all of the artificial statistics—that is, jointly—to the real ones measures less than ϵ. The intersection rules plainly provide a stricter basis of evaluation for the same tolerance level.

Finalized, our first attempt at ABC rejection followed this contour: first, we grid

sample a pre-determined number of joint parameter combinations, dependent on the specified prior distributions' domains for each parameter; second, we run a regression and extract those terms that achieve statistical significance; third, implement the ABC rejection algorithm at some tolerance level using the k-dimensional Euclidean distance to represent discrepancies between the significant summary statistics of the pseudo and real data.

Unfortunately, results did not fare well with this approach either. A small acceptance rate stood as the primary issue, with only 1.6% of candidates being accepted at a certain tolerance level; moreover, two of the variables, ω and η had roughly uniform posterior distributions, indicating that this tolerance level was still not small enough to recover useful information about them from the simulations. Lowering the tolerance level to better approximate the true posterior would only exacerbate the bad acceptance rate issues. Finally, we turned to Adaptive Approximate Bayesian Computation Sequential Monte Carlo, or ABCSMC.

The main draws to ABCSMC include increased efficiency and better handling of complex models. First, ABCSMC allows for the use of a proposal distribution, conditioned on already accepted candidates; whereas plain ABC simply samples from the prior distribution. Second, particles (sets of proposed parameter values) undergo a form of importance sampling, where particles that generate pseudo-data closer to the real data are given more weight, leading to a higher likelihood of resampling. These, in theory, should reduce the number of simulations required to achieve convergence to the posterior. Finally, the sequential proposal distributions allow for a diversity of samples while still having some form of memory of already evaluated "good" proposals in addition to refining the search over the parameter space [23] [28] .

In addition to its other features, the adaptive portion of the Adaptive ABCSMC algorithm operates using a decreasing sequence of automatically-selected tolerances, a process crucial for refining the approximation of the posterior distribution over time. Initially, we set a relatively high tolerance level, which we determine based on domain expertise or as a quantile of the distances observed when comparing pseudo-data generated from the prior distribution with the actual data. As the algorithm progresses, this tolerance level is automatically reduced. The reduction can follow a fixed schedule, decreasing the tolerance at each iteration by a fixed factor, or it can be adjusted dynamically based on the performance of the simulation, often setting the next tolerance level to a specific quantile of the distances from the last iteration. This strategy ensures that as the tolerance decreases, the approximation of the posterior distribution converges

more closely to the true posterior, enhancing the accuracy and reliability of the inference process [28].

All this to say Adaptive ABCSMC reduces computation time by saving information across runs of the algorithm and by eliminating areas of prior space from which we should not propose since we know those areas give "bad" results. The dynamically adaptive tolerances also should reduce user error in selecting the tolerance levels across many iterations.

We run this algorithm on the first two states, B and M. After more than 96 hours to procure 1,000 samples at 7 different tolerance levels, we have the following.

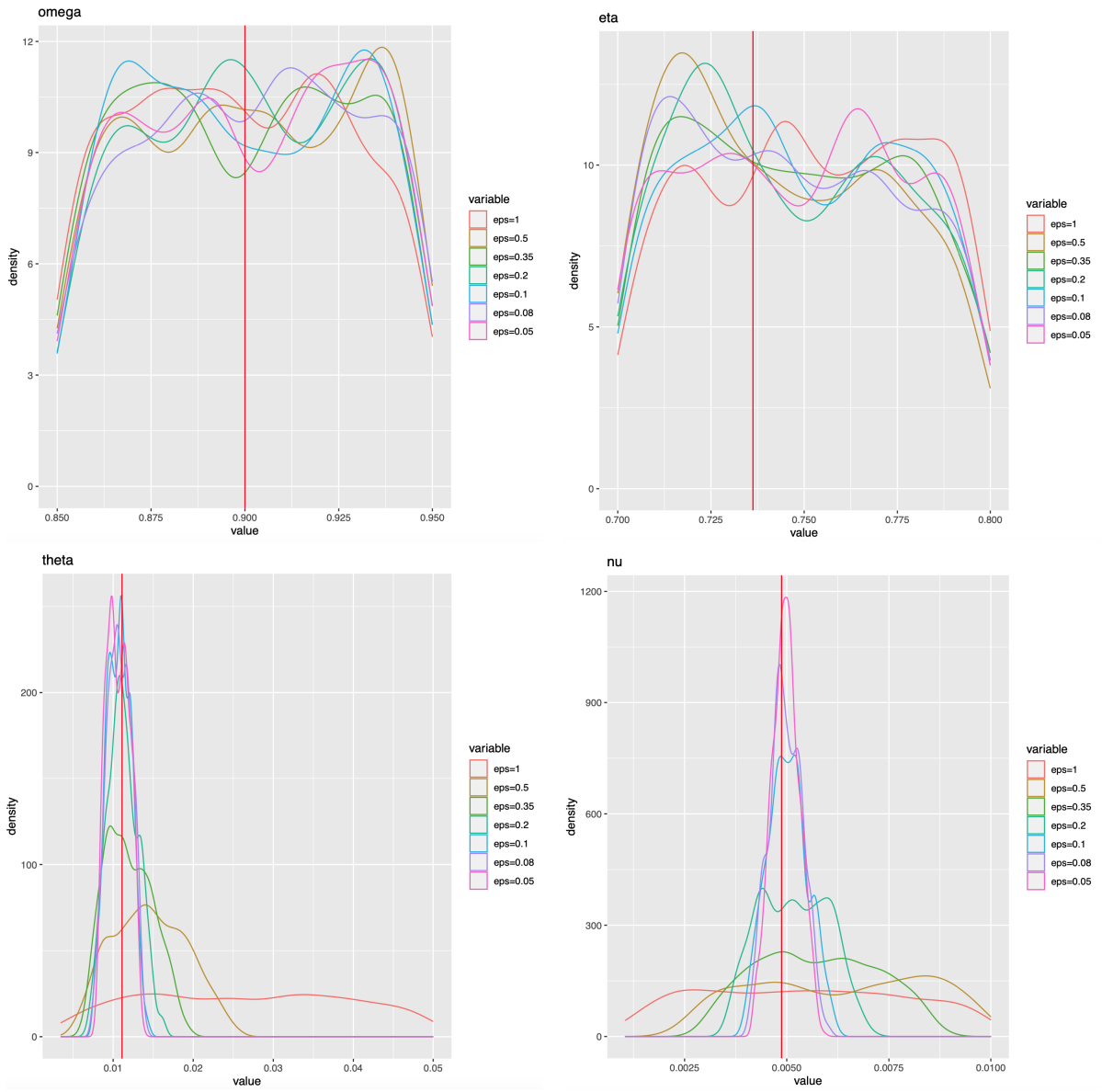

Figure 4.1. Parameter Posteriors via Adaptive ABCSMC

In Figure 4.1, we see the same problem mentioned with respect to the simple ABC rejection algorithm, where ω's and η 's posterior estimates do not converge while θ's and ν's do. The colored lines show an approximate posterior for each respective parameter, where the violet and magenta lines should outline a shapely distribution, since they correspond to the strictest tolerance level. The vertical red line denotes the true parameter value, so ideally those distributions would be centered around it. Finally, note that each approximate posterior here is comprised of 1,000 samples. We can therefore estimate ν and θ, but not ω and η. Inspection of the correlation plot shows that the relationship

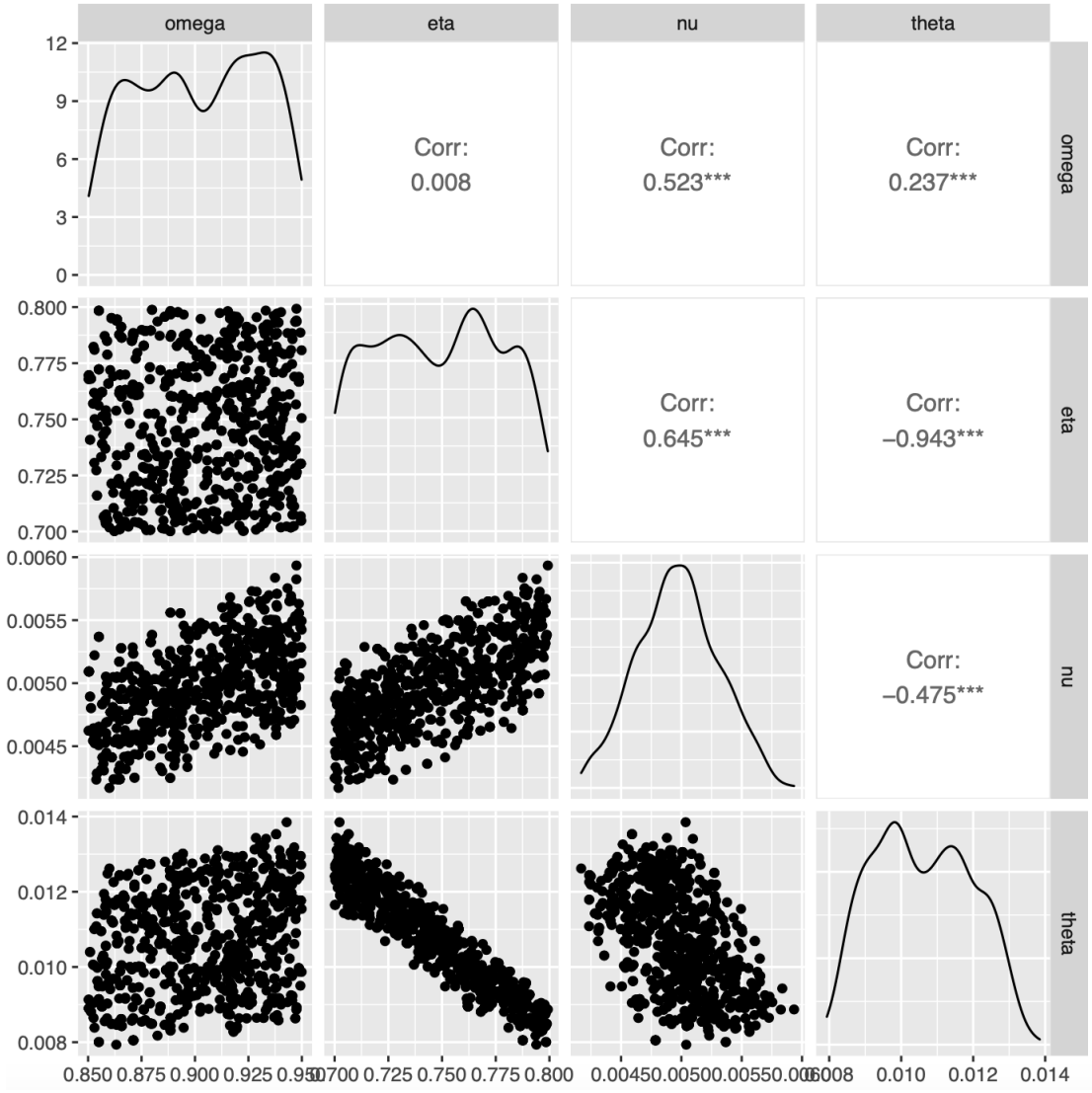

Figure 4.2. Correlation Plots

between η and ω in this model is either unimportant or nonexistent, which is puzzling considering the definition of the model, where the product of ω and η directly influences the number of newborns entering states B and M. We believe that the main culprit behind the poor performance is an identifiability issue between these two parameters: the model likely requires a revision to resolve this.

Chapter 5
Discussion

To this point, we have introduced the model of choice, the subcompartmental model with Poisson point processes; we have covered the inferential strategy, the simulation-based Approximate Bayesian Computation; and we have recounted the lacking results. Here, we discuss the challenges to answering the research questions through our approach, some factors that would make answering those questions easier, and some future directions to explore to better refine our given approach.

5.1 Challenges

In testing these various approaches, we have given the algorithms much more information than would be feasible with real data: that is, knowledge of the would be latent variables in addition to more granularity on the timescale. It is therefore clear that there remain several challenges to be surmounted to attain even a rough inference.

First and foremost, even in simple *SIR* modeling, we do not typically know the population of susceptibles over time. Some rare exceptions include a novel epidemic (like COVID-19), where ostensibly everyone is susceptible to the disease at the beginning of the outbreak; however, with rotavirus, only a fraction of the total population is susceptible at any given time. Estimating this fraction stands as a tricky task, for one of the few direct ways to do accomplish this lies in collecting a large sample of serological data and surveying the number of people who have antibodies to the virus. Otherwise, we are left with estimation.

Second, the data with the highest availability, hospitalization/case data, suffer from a number of bias-inducing factors. Consider one such observation in the data set. In order to record this observation, this person must first suffer an illness severe enough to seek medical help. Once seen by a doctor, the doctor must then make the correct diagnosis.

This chain of events may be further complicated by the biology of the particular disease. Some diseases, for example, have very long latent/incubation periods where the host appears unaffected, thus appearing susceptible but actually belonging to the infective class. Therefore, these hospitalization data do not offer us even an unbiased random sample, let alone the entirety, of the infectives.

Third, and compounding with the previous point, we turn to this specific problem and the singular dataset examined from HCUP. We noted that the time-series data are very irregular. While there is a persistent trend of annual epidemics in the beginning few years, after the introduction of the vaccine, sheer irregularity surmounts that pattern. The typical epidemiological model focuses on nonlinear processes with clear trends (e.g., sinusoidal as in measles data from the mid-20th century or exponential as in a novel epidemic), but these data do not offer a clear trend after the introduction of the vaccine [29]. Certainly, this complicates any inferential approach aiming to incorporate data both pre- and post-vaccine introduction. Moreover, as noted previously, the data themselves arrive in poor spatial and temporal resolution, aggregated by month and by census division. Given that our attempt with these methods yielded mediocre results with a much higher time resolution and foregoing the spatial component, it bears careful consideration how one may overcome this deficiency. Intuition tells us that any hope of spatial analysis must wait until the temporal factor is captured.

Fourth, as aforementioned, there remain some burdensome computational costs. While the likelihood-based approach fared well in the preliminary tests, we must mention a few caveats: one is that we were forced to use highly informative priors for a pair of parameters due to identifiability issues. Another is being able to write out the probability model. Yet another is that the preliminary tests were low-dimensional. On the other hand, the likelihood-free approach does not require that we are able to write the probability model, only that we write a way to simulate from it. Even though the computational cost may be higher in the preliminary tests, that cost scales much better for the likelihood-free approach than the likelihood-based approach. As more terms are added to the probability model and as the dimensionality of the parameter space increases, so too does the complexity of the calculations necessary in order to achieve a valid, functional Monte Carlo sampler. This means that for our ultimate goal of including a vaccination state and a re-infection process, the likelihood-based approach would, theoretically, cost much more computationally than the likelihood-free approach. That said, some runs of the algorithm for 6 tolerance levels took upwards of 43 hours, with the final level taking 10 hours, even with just the basic two-state model, and only to obtain 500 samples from

the posterior.

Fifth, whereas the typical *SIR* model holds only three unique states, our subcompartmental model has five, two of which divide S. Since we do not know S, discerning what its subcompartments are becomes even more ambitious. That said, some researchers have concocted methods for retrieving the unknown susceptible class using only the infection data. One approach to consider, for example, is set forth by Bobashev et. al in a 2000 paper entitled "Reconstructing Susceptible and Recruitment Dynamics from Measles Epidemic Data." The authors claim they can recover, under "general theoretical assumptions" and "up to linear scaling parameters," the dynamics of the susceptibles and the recruitment to the susceptibles based only on case report data [30]. This would immensely aid the effort using the HCUP data, though it remains unclear if the assumptions used in the paper, specifically tailored to mid-20th century measles outbreaks in the developed world (prior to immunizations), can effectively translate to turn-of-the-century rotavirus outbreaks in the developed world both before and after immunizations. Even if these assumptions are valid, unfortunately, their approach concerned the traditional *SIR* model. Moreover, more of the preceding challenges must be solved before this would even begin to help.

5.2 Conclusion

Here we have presented a novel approach to inferring the parameters driving the disease dynamics of a discrete-time, stochastic, subcompartmental model. The paramount parameter of our study measures an interaction between breastfeeding and successful vaccination against rotavirus. Examining its effect remains crucial because no considerable backlog of credible experimental data exist: thus the data at our disposal originate from hospitalizations, not laboratories. When considering analysis of disease data, the common approach is one of continuous-time modeling; however, the spatio-temporal resolution of the data here is a far cry from continuous, so we instead opt for a discrete-time method. In doing so we translate the ordinary differential equations from classical epidemiology into a plethora of Poisson point processes. While this provides flexibility (e.g. choice of time interval length), it also comes at a computational cost. To better exploit this inherent cost, we consider a likelihood-free method of inference, which already has a high cost at baseline (and thus does not suffer greatly from the translation), and whose upside allows easier inclusion of latent variables in the model, of which this model has several.

Unfortunately, the likelihood-free approach's efficiency also deteriorates rapidly with

increased dimensionality unless we assign strict priors; however, this brings the challenge of knowing which priors make sense to use. A different modeling approach, on the other hand, may stand as a wise modification. For example, one may maintain the continuous-time components of ODEs but add a discrete, monthly reporting/measurement process to incorporate the data. Moreover, the research goal might be tweaked in the future, applying to a different illness of more interest to both the public and epidemiological communities, or at such a time that more data become available, leading to the marginalization of the troublesome, myriad latent variables in the model.

Appendix
MCMC Algorithms

1 The Metropolis-Hastings Algorithm

1. Initialize x_0.
2. For $t = 1, 2, ..., N$:
   (a) Generate $y \sim q(\cdot|\boldsymbol{\theta})$, where q is a specified proposal distribution relying on the parameters $\boldsymbol{\theta}$.
   (b) Generate $u \sim \text{Unif}(0, 1)$.
   (c) If $u < \alpha(x_t, y) = \min\left(\frac{\pi(y)q(y,x_t)}{\pi(x_t)q(x_t,y)},\ 1\right)$, set $x_{t+1} = y$.
   (d) Otherwise, set $x_{t+1} = x_t$.

2 Particle MCMC Algorithm

1. Initialize $\boldsymbol{\theta_0}$, $\hat{\ell}(\theta_0)$
2. Draw $\boldsymbol{\theta}^* \sim q(\cdot|\boldsymbol{\theta})$, where q is a specified proposal function
3. Run SMC algorithm using $f_{X_t|X_{(t-1)}}(\cdot|X_{(t-1)}; \boldsymbol{\theta^*})$
   (a) Initialize P particles: $X_{p,0} \sim \pi_{X_0}(\cdot)$
   (b) Simulate from each particle: $X^*_{p,t} \sim f_{X_t|X_{(t-1)}}(\cdot|X_{p,(t-1)}; \boldsymbol{\theta^*})$
   (c) Weight each simulated value: $w_{p,t} = f_{Y_t|X_t}(Y_t|X^*_{p,t})$
   (d) Sum weights: $\hat{\ell}_t(\boldsymbol{\theta^*}) = \sum_{p=1}^{P} w_{p,t}$
   (e) Normalize weights: $\tilde{\mathbf{w}}_\mathbf{t} = \frac{w_{p,t}}{\hat{\ell}_t(\boldsymbol{\theta^*})}$

(f) Resample particles with replacement with prob $\tilde{\mathbf{w}}_\mathbf{t}$

(g) Repeat ii-vi for T total iterations

(h) Obtain $\hat{\ell}(\boldsymbol{\theta}^*) = \sum_{t=1}^{T} \hat{\ell}_t(\boldsymbol{\theta}^*)$

4. Calculate acceptance probability $\alpha = \frac{\pi(\boldsymbol{\theta}^*)\exp(\hat{\ell}(\boldsymbol{\theta}^*))}{\pi(\boldsymbol{\theta_{m-1}})\exp(\hat{\ell}(\boldsymbol{\theta}_{m-1}))}$

5. Draw $U \sim \text{Unif}(0, 1)$

6. If $U < \alpha$, set $\hat{\ell}(\boldsymbol{\theta}_m) = \hat{\ell}(\boldsymbol{\theta}^*)$ and $\boldsymbol{\theta}_m = \boldsymbol{\theta}^*$. Otherwise, set $\hat{\ell}(\boldsymbol{\theta}_m) = \hat{\ell}(\boldsymbol{\theta}_{m-1})$, $\boldsymbol{\theta}_m = \boldsymbol{\theta}_{m-1}$

7. Repeat 2-7 for M total iterations

Bibliography

[1] Clinic, M. (2020), "Breast-feeding vs. formula-feeding: What's best?" .
URL https://www.mayoclinic.org/healthy-lifestyle/infant-and-toddler-health/in-depth/breast-feeding/art-20047898

[2] CDC (2021), "Breastfeeding Is an Investment in Health, Not Just a Lifestyle Decision," .
URL https://www.cdc.gov/breastfeeding/about-breastfeeding/why-it-matters.html

[3] NHS (2021), "How long do babies carry their mother's immunity?" .
URL https://www.nhs.uk/common-health-questions/childrens-health/how-long-do-babies-carry-their-mothers-immunity/

[4] CDC (2020), "Breastfeeding Report Card," .
URL https://www.cdc.gov/breastfeeding/data/reportcard.htm

[5] ——— (2021), "Rotavirus," .
URL https://www.cdc.gov/rotavirus/index.html

[6] Simpson, E., S. Wittet, J. Bonilla, K. Gamazina, L. Cooley, and J. L. Winkler (2007) "Use of formative research in developing a knowledge translation approach to rotavirus vaccine introduction in developing countries," *BMC Public Health*, **7**(281).

[7] Pitzer, V. E., C. Viboud, L. Simonsen, C. Steiner, C. A. Panozzo, W. J. Alonso, M. A. Miller, R. I. Glass, J. W. Glasser, U. D. Parashar, et al. (2009) "Demographic variability, vaccination, and the spatiotemporal dynamics of rotavirus epidemics," *Science*, **325**(5938), pp. 290–294.

[8] Patel, M. M., R. Glass, R. Desai, J. E. Tate, and U. D. Parashar (2012) "Fulfilling the promise of rotavirus vaccines: how far have we come since licensure?" *The Lancet infectious diseases*, **12**(7), pp. 561–570.

[9] Parashar, U. D., H. Johnson, A. D. Steele, and J. E. Tate (2016) "Health Impact of Rotavirus Vaccination in Developing Countries: Progress and Way Forward," *Clinical Infectious Diseases*, **62**(2), pp. S91–S95.

[10] Moon, S.-S., Y. Wang, A. L. Shane, T. Nguyen, P. Ray, P. Dennehy, L. J. Baek, U. Parashar, R. I. Glass, and B. Jiang (2010) "Inhibitory effect of breast milk on infectivity of live oral rotavirus vaccines," *The Pediatric infectious disease journal*, **29**(10), pp. 919–923.

[11] (2021) "Immunization of Mothers and Infants," in *Red Book: 2021–2024 Report of the Committee on Infectious Diseases*, American Academy of Pediatrics, `https://publications.aap.org/redbook/book/chapter-pdf/1602638/rbo2021\_s2\_001\_004\_en.pdf`.
URL `https://doi.org/10.1542/9781610025782-S2\_001\_004`

[12] Ali, A., A. M. Kazi, M. M. Cortese, J. A. Fleming, S. Moon, U. D. Parashar, B. Jiang, M. M. McNeal, D. Steele, Z. Bhutta, et al. (2015) "Impact of withholding breastfeeding at the time of vaccination on the immunogenicity of oral rotavirus vaccine—a randomized trial," *PloS one*, **10**(6), p. e0127622.

[13] Niewiesk, S. (2014) "Maternal antibodies: clinical significance, mechanism of interference with immune responses, and possible vaccination strategies," *Frontiers in immunology*, **5**, p. 103203.

[14] Keeling, M. J. and P. Rohani (2008) *Modeling Infectious Diseases in Humans and Animals*, 1st ed., Princeton University Press.

[15] Streit, R. L. and R. L. Streit (2010) *The poisson point process*, Springer.

[16] Chib, S. and E. Greenberg (1995) "Understanding the metropolis-hastings algorithm," *The american statistician*, **49**(4), pp. 327–335.

[17] Endo, A., E. Van Leeuwen, and M. Baguelin (2019) "Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers," *Epidemics*, **29**, p. 100363.

[18] King, A. A., D. Nguyen, and E. L. Ionides (2015) "Statistical inference for partially observed Markov processes via the R package pomp," *arXiv preprint arXiv:1509.00503*.

[19] Doucet, A., S. Godsill, and C. Andrieu (2000) "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing*, **10**, pp. 197–208.

[20] Cornuet, J.-M., F. Santos, M. A. Beaumont, C. P. Robert, J.-M. Marin, D. J. Balding, T. Guillemaud, and A. Estoup (2008) "Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation," *Bioinformatics*, **24**(23), pp. 2713–2719.

[21] Beaumont, M. A. (2010) "Approximate Bayesian computation in evolution and ecology," *Annual review of ecology, evolution, and systematics*, **41**, pp. 379–406.

[22] Hazra, I., M. D. Pandey, and N. Manzana (2020) "Approximate Bayesian computation (ABC) method for estimating parameters of the gamma process using noisy data," *Reliability Engineering & System Safety*, **198**, p. 106780.

[23] Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf (2009) "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *Journal of the Royal Society Interface*, **6**(31), pp. 187–202.

[24] Aeschbacher, S., M. A. Beaumont, and A. Futschik (2012) "A novel approach for choosing summary statistics in approximate Bayesian computation," *Genetics*, **192**(3), pp. 1027–1047.

[25] Toni, T. and M. P. Stumpf (2009) "Tutorial on ABC rejection and ABC SMC for parameter estimation and model selection," *arXiv preprint arXiv:0910.4472*.

[26] Jiang, B. (2018), "Approximate Bayesian computation with Kullback-Leibler divergence as data discrepancy," .

[27] Prangle, D. (2017) "Adapting the ABC distance function," .

[28] Del Moral, P., A. Doucet, and A. Jasra (2012) "An adaptive sequential Monte Carlo method for approximate Bayesian computation," *Statistics and computing*, **22**, pp. 1009–1020.

[29] Bjørnstad, O. N. (2022) *Epidemics: models and data using R*, Springer Nature.

[30] Bobashev, G. V., S. P. Ellner, D. W. Nychka, and B. T. Grenfell (2000) "Reconstructing susceptible and recruitment dynamics from measles epidemic data," *Mathematical Population Studies*, **8**(1), pp. 1–29.