UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Inference and Forecasting Using Infectious Disease Surveillance Data

Permalink
https://escholarship.org/uc/item/2r48w6mj

Author
Bayer, Damon

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Inference and Forecasting Using Infectious Disease Surveillance Data

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

Damon Bayer

Dissertation Committee:
Professor Volodymyr Minin, Chair
Professor Veronica Berrocal
Associate Professor Weining Shen
Associate Professor Daniel Parker

2023

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work would not have been possible without the help and guidance of many friends, colleagues, and mentors throughout the last five years.

First, I offer my deepest gratitude to my advisor, Volodymyr Minin, who has been an exceptional role model to me as both a scholar and a person. Each week, I looked forward to our meetings and chatting about research, my career, or whatever else was on our minds. As Volodymyr's first statistics PhD graduate at UCI, I have been delighted, but not surprised, by the extraordinary group he has assembled in Irvine thus far: Yushang, Isaac, Catalina, Thanasi, and Jessalyn. They are every bit as intelligent, kind, and humorous as their advisor. Beyond our juicy gossip, tasteful slack emoji, and culinary innovations, it has been a joy to learn statistics alongside each of them.

I am also fortunate to have met a number of exceptional mentors and colleagues outside of UCI. Among these are the generous and talented statisticians in the Biostatistics Research Branch at the National Institute of Allergy and Infectious Diseases, including Mike Fay and Jon Fintzi. Mike encouraged my ambitions to write what eventually became Chapter 3 of this dissertation during my summer internship at NIAID in 2021 and enabled me to explore new areas of statistics I had not encountered at UCI. Jon was extremely generous with his time in answering so, so many of my questions in the early days of my work modeling COVID-19 and has continued to be a great source of guidance throughout my graduate schooling. I am also indebted to my collaborators at other University of California schools and the California Department of Public Health, especially Tomás León and his team, for their insights and many fruitful conversations about our COVID-19 modeling efforts.

Finally, I have been lucky to be supported by many friends and family members while working on this dissertation. They have made life outside of school so rich.

The text of Chapter 3 of this dissertation is a modified reprint of the material as it appears in *Statistics in Medicine* [Bayer et al., 2023], used with permission from Wiley.

VITA

Damon Bayer

EDUCATION

Doctor of Philosophy in Statistics **2023**
University of California, Irvine *Irvine, CA*

Master of Science in Statistics **2020**
University of California, Irvine *Irvine, CA*

Master of Science in Mathematics **2018**
South Dakota State University *Brookings, SD*

Bachelor of Science in Mathematics **2016**
South Dakota State University *Brookings, SD*

ABSTRACT OF THE DISSERTATION

Inference and Forecasting Using Infectious Disease Surveillance Data

By

Damon Bayer

Doctor of Philosophy in Statistics

University of California, Irvine, 2023

Professor Volodymyr Minin, Chair

Statistical modeling of infectious disease data is among the oldest applications of statistics. Today, it is an increasingly relevant application of research, due to globalization that enables diseases to spread further and faster, as well as the abundance of relevant data from electronic surveillance systems, seroprevalence studies, and genetic sequencing of pathogens. In this work, we develop novel statistical methods to combine varied data sources to improve both inference and forecasting. First, we work with data from assay validation studies and active surveillance studies to develop confidence intervals for prevalence estimates from complex surveys with imperfect assays. In this complicated setting, there are no established competitive methods, and ours exhibits at least nominal coverage. In addition, we apply our model in simplified cases where competitors exist and demonstrate desirable properties. Next, we develop a semi-parametric Bayesian compartmental model that effectively integrates passively collected time series of diagnostic tests and mortality data, as well as actively collected seroprevalence data. We emphasize retrospective inference and evaluate the utility of each data stream in the context of short-term forecasting. Finally, we focus on healthcare demand forecasting during epidemic surges of pathogen variants capable of immune escape. We build upon our Bayesian compartmental model to incorporate time series of cases, hospitalizations, ICU admissions, deaths, and genetic sequence counts. We show that using genetic information leads to superior forecasting performance, compared to

traditional models. Throughout each project, we employ our methods to analyze a variety of COVID-19 data sets at the county, state, and national levels.

Chapter 1

Introduction

The statistical modeling of infectious disease data is among the oldest applications of statistics, beginning with Bernoulli's work on smallpox in the 18th century [Bernoulli and Blower, 2004]. Today, it is an increasingly relevant application of research, due to globalization that enables diseases to spread further and faster, as well as the abundance of relevant data from electronic surveillance systems, social contact and mobility patterns from mobile phones, and genetic sequencing of pathogens. The importance of this pursuit is apparent when reflecting on the recent COVID-19 pandemic, which resulted in millions of deaths and the largest global recession in nearly a century [WHO, 2020, Zumbrun, 2020].

1.1 Epidemic Surveillance Data and Its Applications

Infectious disease data differs from data used in traditional statistical applications because they are highly dependent in time and space, and are almost always partially observed [Held et al., 2019]. People without healthcare access or who do not exhibit disease symptoms may not seek diagnosis, leading to systematic bias and under-reporting in case counts. Among

those who seek diagnosis, their disease status may not be correctly identified, and their results may not be reported to a centralized database. Even among those who receive a correct diagnosis and whose diagnostic test results are reported, the precise timing of infection, transmission, and recovery are generally unknown. These factors make inference and forecasting challenging. For instance, a small outbreak with a high reporting rate may produce similar observed case counts as a large outbreak with a low reporting rate.

Infectious disease surveillance data can be reported at a variety of resolutions and used for a variety of tasks. Some methods are designed to work with detailed data from a specific, relatively small outbreak, while others are more suitable for widespread diseases observed with less detail. This work is primarily concerned with the latter scenario and uses both passively and actively collected surveillance data. These data sources traditionally take the form of aggregated counts of incidence data (e.g., new tests, new cases, and new deaths), counts of prevalence data (e.g., the number of hospitalized patients with a disease or the number of people with seropositive tests from a sampled population) at some coarse demographic, spatial, and temporal resolution (e.g., stratified by age and county and documented weekly). Our work and other recent advances seek to integrate other forms of data into this framework [Rasmussen et al., 2011, Tang et al., 2023].

These data and models are vital tools for both inference and forecasting tasks. One major goal in the inference context is nowcasting, which typically involves estimating the effective reproduction number, R_t, in real time [Cori et al., 2013]. This parameter is defined as the expected number of secondary cases that arise from a primary case and is influenced by factors inherent to the pathogen, as well as the environmental conditions. For example, consider the case of a new virus variant emerging in the summer. The new variant may be inherently more transmissible than the previously circulating strain, driving up the effective reproduction number. Simultaneously, the warmer weather and school closures work to decrease the effective reproduction number. Of most concern is the threshold $R_t = 1$. When $R_t < 1$, the

average primary case produces fewer than one secondary case. If this is sustained, prevalence will decrease. When $R_t > 1$, the average primary case produces more than one secondary case, leading to an exponential increase in cases. Other inference tasks involve quantifying the effects of interventions, whether they be pharmaceutical (e.g., vaccination campaigns) or non-pharmaceutical (e.g., mask mandates). Other inference tasks may be more basic. For example, an infectious disease modeler may simply want to infer the true proportion of the population that has been infected with a disease. Beyond inference, modelers may also be interested in forecasting future disease burden, often with special attention to severe outcomes like hospitalizations and deaths [Pollett et al., 2021]. Relatedly, modelers may consider potential future scenarios to drive public policy by answering questions like "in the presence of a more transmissible variant, how many people do we need to vaccinate to prevent the healthcare system from being overwhelmed?"

In the next section, we provide an overview of surveillance data available to infectious disease modelers to motivate our methodological contributions.

1.2 Motivating Examples

In Chapter 3, we work with data originally analyzed by Kalish et al. [2021]. This data set was collected to estimate the cumulative incidence of SARS-CoV-2 in undiagnosed adults in the United States between May and July 2020. The data consists of 304 seropositive samples among the 8058 total antibody tests. The samples are weighted based on a variety of socioeconomic factors. The assay used is estimated to have perfect sensitivity, based on 56 tests on individuals with confirmed SARS-CoV-2 and perfect specificity based on 300 tests on individuals confirmed to not have SARS-CoV-2.

In Chapter 4, we analyze data from Orange County, California between March 2020 and

February 2021. This data includes case and mortality data provided by Orange County Health Care Agency, including individual-level records of all negative and positive PCR tests during the modeling period. The aggregated counts of cases, tests, deaths, and test-positivity (cases / tests) are presented in Figure 1.1. This figure is labelled with three periods. In each period, there is some tension in the trend portrayed by the various data streams, each of which we believe to be reflective of the underlying number of infections. In the first period, there is a gradual increase in both tests and deaths, a sudden increase in cases, and a gradual decline in test positivity, which is followed by a sudden increase. In the second period, tests rise slightly, while deaths, cases, and test positivity remain nearly constant. In the final period, there is a sharp increase and decrease in deaths and test positivity, while tests and cases have a large dip in the middle of this period. Resolving the conflicting signals between these data streams is the goal of Chapter 4. To achieve this goal, we also use data from Bruckner et al. [2021], a study conducted to estimate the seroprevalence in Orange County from a population-representative sample, which consists of 343 seropositive cases among 2979 tests conducted between July 10 and August 16, 2020.
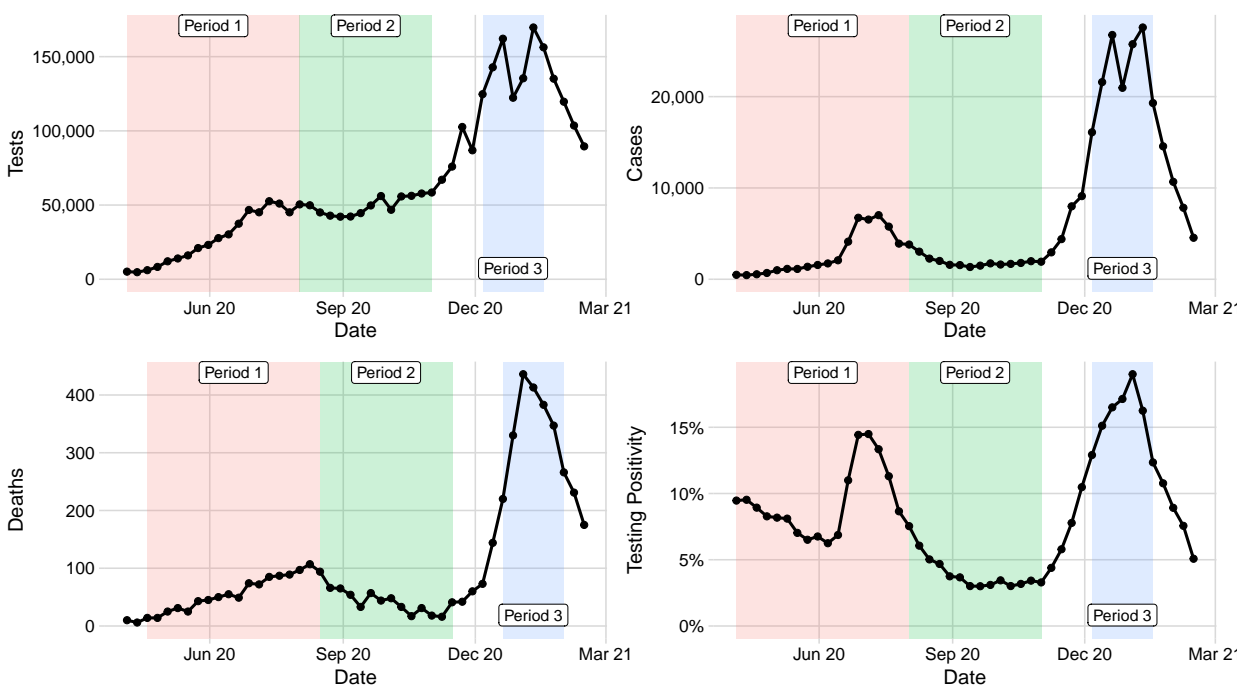
Figure 1.1: COVID-19 surveillance data from Orange County, California used in Chapter 4. The figure shows weekly counts of tests, cases (positive tests), reported deaths due to COVID-19, as well as testing positivity. The distinct periods labelled with different background colors are explained in Section 1.2.

In Chapter 5, we again analyze data from Orange County, as well as statewide data from California, with a focus on forecasting healthcare demand during the Omicron BA.1 wave of winter 2022. Counts of cases, hospital occupancy, ICU occupancy, and deaths are provided to us by the California Department of Public Health. The counts of virus sequence by variant are provided by the Global Initiative on Sharing All Influenza Data (GISAID) [Shu and McCauley, 2017] via Outbreak.info [Gangavarapu et al., 2023]. The aggregated counts of cases, hospital occupancy, ICU occupancy, deaths, and sequence counts of the BA.1 and non-BA.1 variants are presented in Figure 1.2 for Orange County and Figure 1.3 at the statewide level. The gray highlighted regions indicate the times for which we create forecasts. The objective of this chapter is to develop a model that uses the proportion of BA.1 sequences to improve forecasting during this time of rapidly changing dynamics.
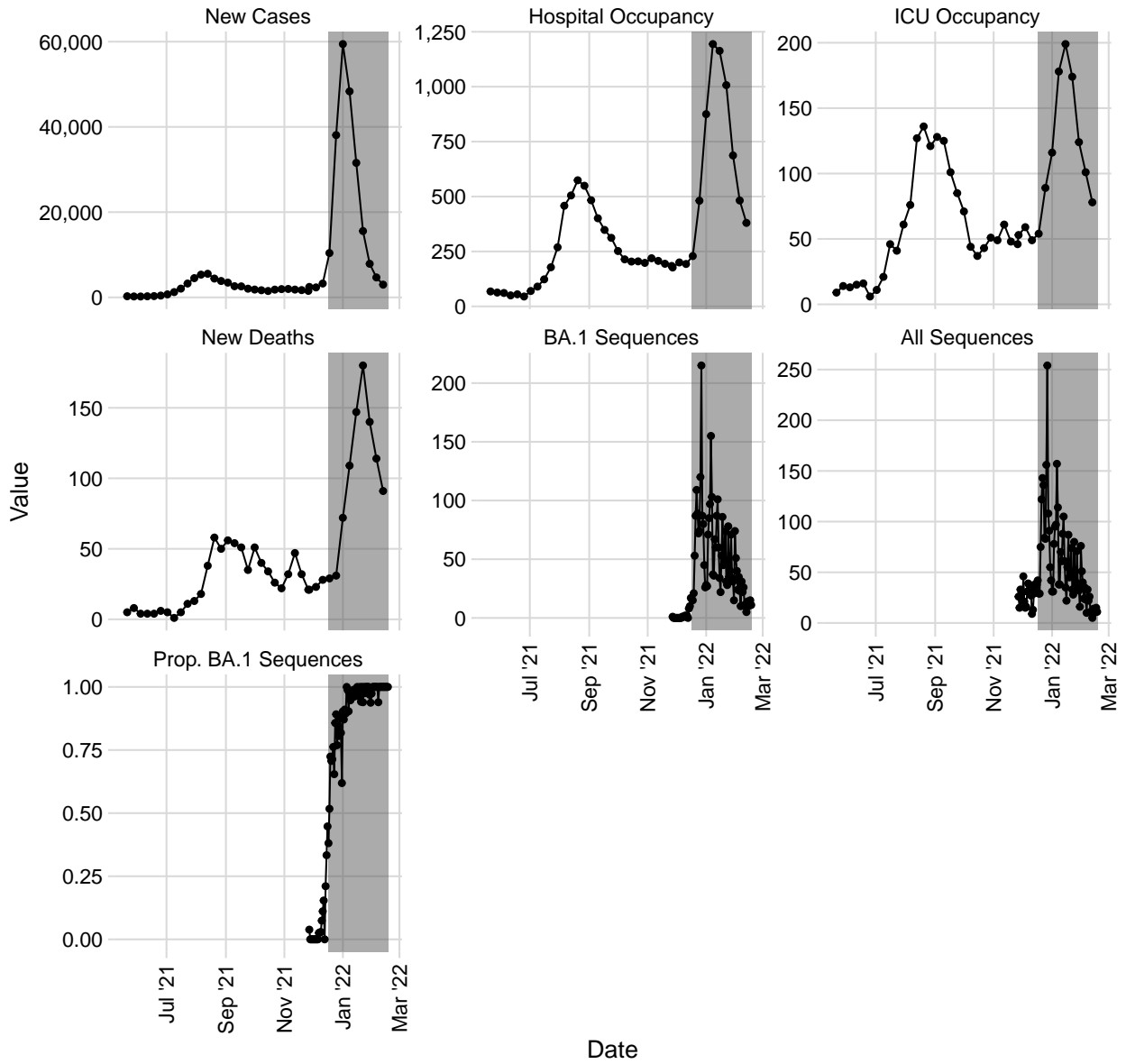
Figure 1.2: COVID-19 surveillance data from Orange County, California used in Chapter 5. The plots show weekly counts of cases, hospital and ICU occupancy of patients with COVID-19, reported deaths due to COVID-19, as well as counts of virus sequences for Omicron BA.1 and all lineages, and the proportion of BA.1 lineages. The gray highlighted regions indicate the times for which we produce forecasts.
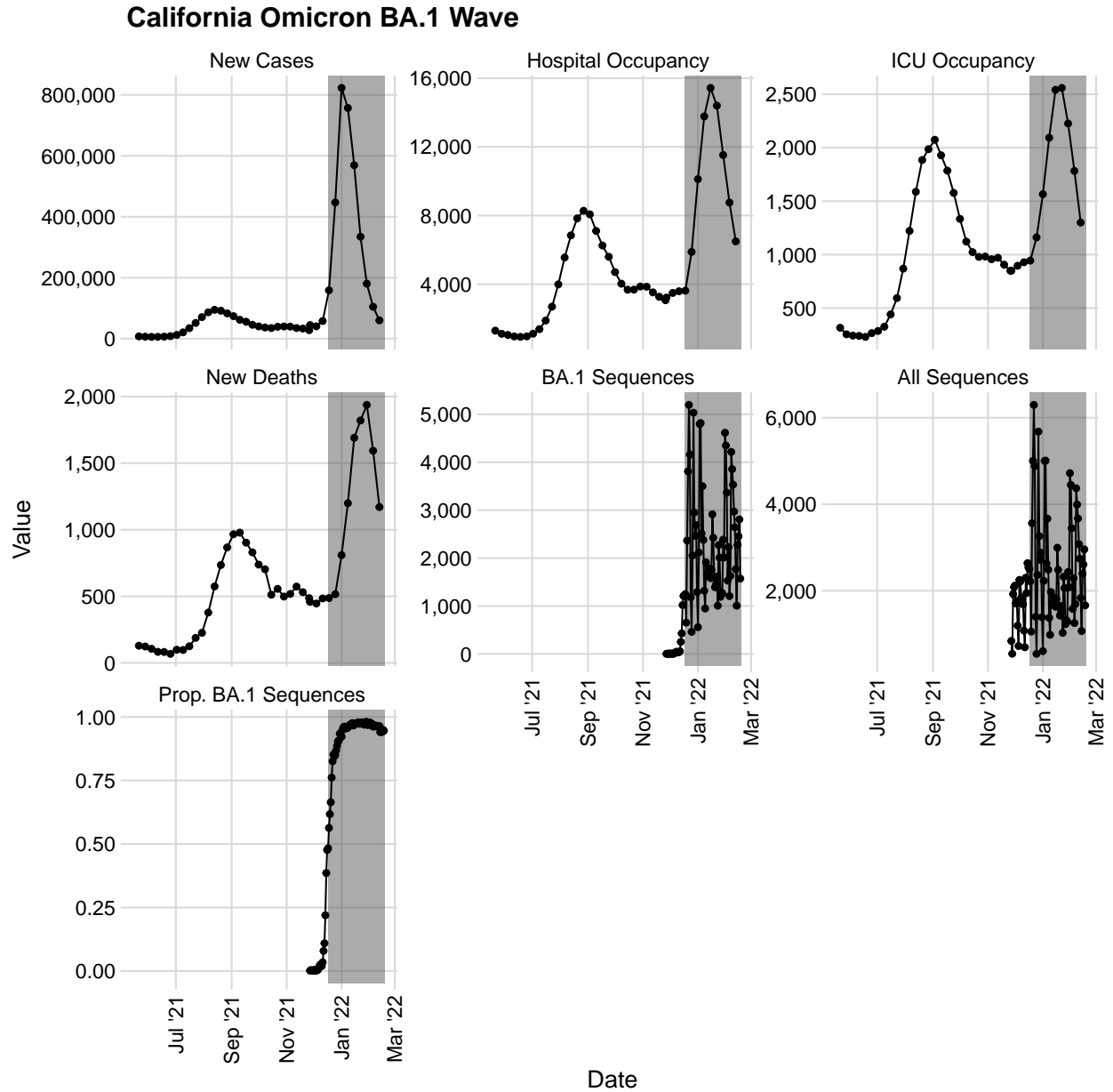
Figure 1.3: Statewide COVID-19 surveillance data from California used in Chapter 5. The plots show weekly counts of cases, hospital and ICU occupancy of patients with COVID-19, reported deaths due to COVID-19, as well as counts of virus sequences for Omicron BA.1 and all lineages, and the proportion of BA.1 lineages. The gray highlighted regions indicate the times for which we produce forecasts.

1.3   Overview of this dissertation

Chapter 2 provides background information on areas of statistics essential to understanding the methods proposed and assessed in this dissertation. The topics covered include mathematical models for the spread of infectious diseases, Bayesian inference and Markov chain Monte Carlo, forecast assessment, and fiducial inference.

In Chapter 3, we eschew the temporal complications of analyzing infectious disease data and focus on issues of sampling schemes and diagnostic test accuracy. While there are established methods for estimating disease prevalence with associated confidence intervals for complex surveys with perfect assays and simple random sample surveys with imperfect assays, the case of complex surveys with imperfect assays remains relatively unexplored. We develop and study new methods for this setting. The new methods use the melding method to combine gamma intervals for directly standardized rates and established adjustments for imperfect assays by estimating sensitivity and specificity. One of the new methods appears to have at least nominal coverage in all simulated scenarios. We compare our new methods to established methods in special cases (complex surveys with perfect assays or simple surveys with imperfect assays). In some simulations, our methods appear to guarantee coverage, while competing methods have much lower than nominal coverage, especially when overall prevalence is very low. In other settings, our methods are shown to have higher than nominal coverage. We apply our method to a seroprevalence survey of SARS-CoV-2 in undiagnosed adults in the United States between May and July 2020 and find that our methods estimate less overall prevalence than a method which does not account for imperfections in the assay.

In Chapter 4, we turn our attention to the temporal dynamics of infectious diseases in the presence of changing policy and behavior. Mechanistic models fit to streaming surveillance data are critical for understanding the transmission dynamics of an outbreak as it unfolds in real-time. However, transmission model parameter estimation can be imprecise, and some-

times even impossible because surveillance data are noisy and not informative about all aspects of the mechanistic model. To partially overcome this obstacle, Bayesian models have been proposed to integrate multiple surveillance data streams. We devise a modeling framework for integrating SARS-CoV-2 diagnostics test and mortality time series data, as well as seroprevalence data from cross-sectional studies, and tested the importance of individual data streams for both inference and forecasting. Importantly, our model for incidence data accounts for changes in the total number of tests performed. We apply our Bayesian data integration method to COVID-19 surveillance data collected in Orange County, California between March 2020 and February 2021 and find that 32–72% of the Orange County residents experienced SARS-CoV-2 infection by mid-January, 2021. Despite this high number of infections, our results suggest that the abrupt end of the winter surge in January 2021 was due to both behavioral changes and a high level of accumulated natural immunity.

In Chapter 5, we work in a similar setting as Chapter 4, but where the changing disease dynamics are due to novel disease variants, rather than changing policy and behavior. In this chapter, we are primarily concerned with forecasting, rather than inference. Accurate forecasting of epidemic surges is an important tool for public health agencies to make informed decisions about interventions and resource allocation. Forecasting is particularly challenging when a new variant arises, which may result in increased transmissibility, severity, or immune evasion, leading to surges in cases, hospitalizations, or deaths. To improve forecasting in these scenarios, we propose a way to integrate raw counts of variant sequences into a Bayesian compartmental model. We assume that the observed count of sequences of a novel variant has a beta-binomial distribution whose mean is a product of the total number of observed genetic sequences and the proportion of the infectious population with the novel variant. We then model the average duration of immunity as a flexible function of this proportion. We evaluate our method in a simulation study wherein novel variants become dominant at varying rates and to data from the Omicron wave in Orange County, California and the state of California as a whole. In these assessments, our model is shown

to have superior forecasting performance and is especially better at forecasting the timing and magnitude of the peak hospital occupancy, a metric crucial for public health officers and hospital managers making staffing decisions.

We conclude with a summary of our work in Chapter 6 and discuss opportunities for future research in inference and forecasting using infectious disease surveillance data.

Chapter 2

Background

2.1 Bayesian Inference

When performing Bayesian statistical inference, we aim to quantify uncertainty about parameters of interest, $\boldsymbol{\theta}$, conditional on some observed data, $\mathbf{X}$. To accomplish this, we employ a prior distribution, which expresses our beliefs about $\boldsymbol{\theta}$ without data, and a sampling distribution for $\mathbf{X}$. Formally, by Bayes' theorem and the law of total probability, we have

$$\pi\left(\boldsymbol{\theta} \mid \mathbf{X}\right) = \frac{\pi\left(\mathbf{X} \mid \boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)}{\pi\left(\mathbf{X}\right)} = \frac{\pi\left(\mathbf{X} \mid \boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)}{\int \pi\left(\mathbf{X} \mid \boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)\mathrm{d}\boldsymbol{\theta}}, \tag{2.1}$$

where $\pi\left(\mathbf{X} \mid \boldsymbol{\theta}\right)$ is the likelihood for $\mathbf{X}$ and $\pi\left(\boldsymbol{\theta}\right)$ is the prior distribution. Because $\pi\left(\mathbf{X}\right)$ is constant, with respect to $\boldsymbol{\theta}$, we typically work with an unnormalized posterior

$$\pi\left(\boldsymbol{\theta}\mid\mathbf{X}\right)\propto\pi\left(\mathbf{X}\mid\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right). \tag{2.2}$$

In practice, the posterior distribution, $\pi\left(\boldsymbol{\theta}\mid\mathbf{X}\right)$, rarely exists in closed form. However, we can approximate the posterior distribution via Markov chain Monte Carlo, wherein we construct a Markov Chain whose stationary distribution is the posterior distribution.

A foundational algorithm for sampling from the posterior distribution is the Metropolis-Hastings algorithm [Hastings, 1970]. In this algorithm, at each step in the Markov chain, a new state, $\boldsymbol{\theta}'$, is proposed by sampling from a proposal distribution q. The proposed value $\boldsymbol{\theta}'$, is accepted with probability

$$\alpha\left(\boldsymbol{\theta},\boldsymbol{\theta}'\right)=\min\left\{1,\frac{\pi\left(\boldsymbol{\theta}'\mid\mathbf{X}\right)}{\pi\left(\boldsymbol{\theta}\mid\mathbf{X}\right)}\frac{q\left(\boldsymbol{\theta}\mid\boldsymbol{\theta}'\right)}{q\left(\boldsymbol{\theta}'\mid\boldsymbol{\theta}\right)}\right\} \tag{2.3}$$

$$=\min\left\{1,\frac{\pi\left(\mathbf{X}\mid\boldsymbol{\theta}'\right)}{\pi\left(\mathbf{X}\mid\boldsymbol{\theta}\right)}\frac{\pi\left(\boldsymbol{\theta}'\right)}{\pi\left(\boldsymbol{\theta}\right)}\frac{q\left(\boldsymbol{\theta}\mid\boldsymbol{\theta}'\right)}{q\left(\boldsymbol{\theta}'\mid\boldsymbol{\theta}\right)}\right\}. \tag{2.4}$$

Efficiently exploring the posterior distribution using Metropolis-Hastings requires selecting a proposal distribution that produces proposals that are far enough away from the current parameters to explore the parameter space quickly, but not so widely distributed that they have a very low probability of acceptance. As $\boldsymbol{\theta}$ increases in dimensionality, striking this balance becomes more difficult, and Markov chains constructed with relatively simple algorithms tend to become "stuck" in a small region, unable to explore the full posterior. Sophisticated methods, such as the Metropolis-adjusted Langevin algorithm [Besag, 1994], elliptical slice sampling [Murray et al., 2010], and Hamiltonian Monte Carlo [Duane et al., 1987] have been developed to improve this efficiency in high dimensional settings.

Throughout this dissertation, we rely on the Hamiltonian Monte Carlo. Briefly, Hamilto-

nian Monte Carlo is a form of the Metropolis–Hastings algorithm which uses Hamiltonian dynamics to generate proposals. To accomplish this, the parameter space is augmented with "momentum" variables which guide the proposals through regions of high probability. For a detailed overview, see [Neal, 2011] and [Betancourt, 2018].

2.2 Confidence Distributions

Confidence distributions are a frequentist estimators of a parameter, although they are historically associated with the fiducial distribution from fiducial statistics [Xie and Singh, 2013]. They are "distribution estimators," similar in spirit to the bootstrap and the posterior distribution from Bayesian inference (see section 2.1). A modern definition and short history of confidence distributions is available in [Xie and Singh, 2013]. For clarity, we present a more classical derivation, which relies on previously derived confidence interval procedures.

Under the classical derivation, we can create random variables associated with the lower and upper confidence distributions from two one-sided, nested, confidence interval procedures. For a nested confidence interval procedure, the $(1 - \alpha_1)$ interval is a subset of the $(1 - \alpha_2)$ interval whenever $(1 - \alpha_1) \leq (1 - \alpha_2)$. Denote those one-sided intervals by $[L(\mathbf{x}, 1 - \alpha), \infty)$ and $(-\infty, U(\mathbf{x}, 1 - \alpha)]$, which are defined for any observed data $\mathbf{x}$, and $(1 - \alpha) \in (0, 1)$. For fixed $\mathbf{x}$, define the lower and upper confidence distribution random variables as $L(\mathbf{x}, A)$ and $U(\mathbf{x}, B)$, where A and B are independent uniform random variables. Then a $100(1 - \alpha)\%$ central confidence interval has the $(\alpha/2)$th quantile of $L(\mathbf{x}, A)$ as the lower limit, and the $(1 - \alpha/2)$th quantile of $U(\mathbf{x}, B)$ as the upper limit.

As an example, let $X_1, \ldots, X_n$ be independently and identically distributed Normal(μ, σ^2).

With σ^2 known, a $(1-\alpha)\%$ confidence interval for μ is

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right], \tag{2.5}$$

where $\bar{x}$ is the sample mean and $z_{\alpha/2}$ satisfies $P(Z \geq z_{\alpha/2}) = \alpha/2$ and $Z \sim \text{Normal}(0,1)$.

Letting $L_\mu(\mathbf{x}, q) = U_\mu(\mathbf{x}, q) = \bar{x} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(q)$, we have $L_\mu(\mathbf{x}, A) \sim \text{Normal}\left(\bar{x}, \frac{\sigma^2}{n}\right)$ and $U_\mu(\mathbf{x}, A) \sim \text{Normal}\left(\bar{x}, \frac{\sigma^2}{n}\right)$.

We now note that (2.5) is equivalent to

$$\left[F_{L_\mu}^{-1}(\alpha/2), F_{U_\mu}^{-1}(1-\alpha/2)\right], \tag{2.6}$$

where F_Y^{-1} indicates the inverse of the cumulative distribution function of a random variable Y. Thus, by definition, L_μ and U_μ are lower and upper confidence distribution random variables for μ.

While the focus of our work is on combining confidence distributions of different variables to produce confidence intervals, the distributions have additional uses. For instance, we can create point estimators by considering the mean, median, or mode of a confidence distribution. Under modest conditions, these can be shown to be consistent estimators. We can also perform hypothesis tests using confidence distributions. For a one-sided hypotheses test at the level α of $H_0 : \theta \in C$ versus $H_0 : \theta \notin C$, where C is an interval of the form $[b, \infty]$ or $-\infty, b$, we reject H_0 if the integral of the confidence distribution under C is less than α.

2.3 Mathematical Models for the Spread of Infectious Diseases

2.3.1 Compartmental Models

The compartmental model is a popular tool used to describe the spread of an infectious disease agent. The most basic and prominent of these models is the Susceptible-Infected-Removed (SIR) model, which we use as an example in this section. In these models, the population of interest is divided into compartments that indicate a disease status (e.g., susceptible, infectious, and recovered), and individuals transition between compartments at a specified rate (e.g., transitions from infectious compartment to recovered compartment happen at a rate proportional to the number of infectious individuals). Overviews of mechanistic compartmental models for disease dynamics can be found in [Anderson and May, 1992, Brauer et al., 2008, Keeling and Rohani, 2011, Lessler and Cummings, 2016].

Compartmental models can be represented deterministically or stochastically. In the stochastic treatment, the model is formulated as a continuous-time Markov chain (CTMC). A CTMC is a stochastic process $\{X_t\}, t \geq 0$, on a state space Ω such that

$$\Pr\left(X_{t+s} = j \mid X_s = i, X_{s_n} = i_n, \ldots, X_{s_0} = i_0\right) = \Pr\left(X_{t+s} = j \mid X_s = i\right), \quad (2.7)$$

for any $0 < s_0 < \cdots < s_n < s$ and $i_0, \ldots, i_n, i, j \in \Omega$. For the SIR model with closed population size N, the state space is $\left\{(S, I, R) \in [0, N]^3 \cap \mathbb{N}^3 \mid S + I + R = N\right\}$, where S, I, and R indicate the number of people from the population in the susceptible, infectious, and recovered compartments, respectively. Because N is fixed, the state space can be fully defined using any two of S, I, and R, with the third implicitly defined as N minus the sum of the other two. We use the triplet representation for clarity.

In this model, individuals may only transition from susceptible to infectious ($S \to I$) or from infectious to recovered ($I \to R$) (see Figure 2.1). Infection events are assumed to happen at the rate $\frac{\beta}{N}SI$, where β is the contact rate. Recovery events occur at a rate γI, where γ is the recovery rate. The diagram in Figure 2.2 represents the system's evolution from one state to the next.

$$S \longrightarrow I \longrightarrow R$$
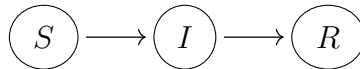
Figure 2.1: The basic SIR compartmental model. Only transitions from susceptible to infectious ($S \to I$) and from infectious to recovered ($I \to R$) are allowed.
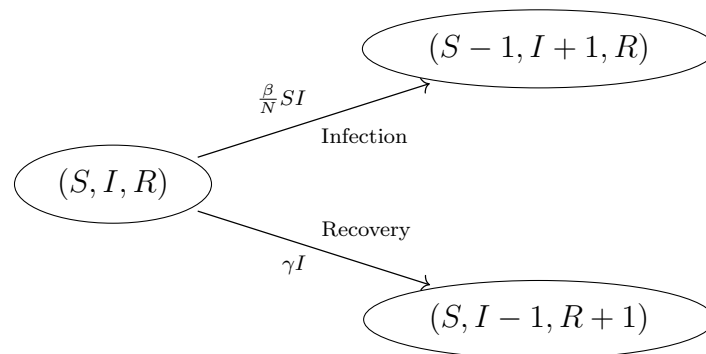
Figure 2.2: State transition graph for SIR model with current state (S, I, R). The population experiences an infection and transitions to the state $(S-1, I+1, R)$ at the rate $\frac{\beta}{N}SI$. The population experiences a recovery and transitions to the state $(S, I-1, R+1)$ at the rate γI.

The deterministic treatment of a stochastic compartmental model is a system of ordinary differential equations (ODEs). For the SIR model, we have

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta \frac{SI}{N}, \qquad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta \frac{SI}{N} - \gamma I, \qquad \text{and} \qquad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I. \tag{2.8}$$

Letting $s(t) = S(t)/N$, $i(t) = I(t)/N$, $r(t) = R(t)/N$, where $A(t)$ is the value of compartment A at time t, the system of ODEs can be viewed as the diffusion limit of the stochastic model as $N \to \infty$, while $s(0), i(0), r(0)$ are held constant [Greenwood and Gordillo, 2009, Fuchs, 2013].

Compared to the deterministic models, the stochastic models are more computationally difficult to work with, but they are particularly useful when there are few infections, as at the beginning or end of an outbreak and in small populations. In these scenarios, the randomness involved at the subject level can have a major impact on the course of the outbreak. In an extreme case, one infectious individual could be introduced to a population, but, by chance, not infect anyone else in the population. Then, no outbreak occurs. In the deterministic treatment, the outbreak is guaranteed to begin and will last indefinitely because $S(t)$, $I(t)$, and $R(t)$ will always be greater than 0, due to the behavior of the differential equations in (2.8). The benefit to using the less realistic deterministic models is that they are more computationally feasible and tend to work well when all compartments are suitably large [King et al., 2015]. This work focuses on the deterministic models, as we typically work in large population settings, where an outbreak is ongoing.

The basic SIR model can be augmented with more compartments, as well as time-varying parameters, to more realistically represent an outbreak. Additional compartments can account for different disease states (e.g., vaccinated or deceased), competing disease variants, or stratification of the population (e.g., by age or location). It is also possible to model an endemic disease by allowing reinfection after recovery (e.g., $R \rightarrow S$ transitions). Time-varying parameters can also model changes in behavior or policy (e.g., stay at home orders and restaurant closures) that cannot be captured by the constant parameters.
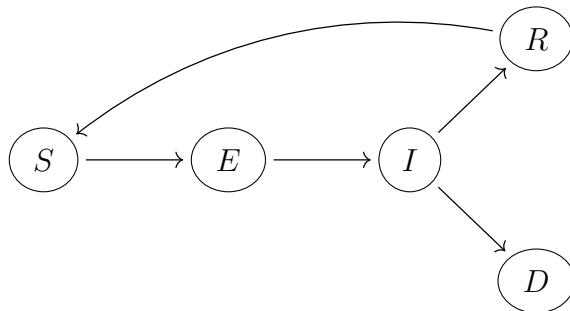
Figure 2.3: Augmented SIR model. The states in this model are susceptible (S), exposed (E), infectious (I), Recovered (R) and Deceased D.

Figure 2.3 presents a model demonstrating some of these features, including additional compartments for individuals who are infected, but not infections (E) and those who are deceased (D), as well as loss of immunity and possibility of reinfections after recovery.

2.3.2 Data Integration

We are interested in estimating parameters of SIR-like models using surveillance data. Broadly, we can categorize data used for this modeling as either actively or passively collected surveillance data. Common examples of passively collected incidence data include daily counts of new diagnostic tests, confirmed cases, and deaths, as well as genetic sequences of viruses or bacteria that cause diseases. We refer to this as incidence data because each count is assumed to correspond to a new event during that time. In contrast, we can work with prevalence data, which is reflective of the frequency of a condition at a given time. Examples of prevalence data include counts of current hospital or intensive care unit (ICU) occupants with a certain disease or the number of seropositive tests from a carefully constructed population sample. We also loosely differentiate between data that is passively or actively collected. Passively collected data like case counts and hospital occupancy is routinely reported to health care agencies and is generally more affordable to collect because it does not require a formal study to be conducted. Actively collected data is often the result of a survey that has been deliberately designed and conducted to measure some aspect of a population, like the number of people previously infected with a disease or the contact rates between different subpopulations. This data is more expensive to collect, but can provide information that is difficult to gather through passive surveillance.

As alluded to in Section 2.3.1, modelers can construct compartmental models which encode arbitrarily complex dynamics. While simulating from these models can be straightforward, connecting them to real-world data to perform statistical inference is a non-trivial task.

18

Broadly, a model uses a compartmental model and basic parameters, like the reproductive number, duration of infectiousness, and infection-fatality ratio, to model the size of a compartment at a series of time points. These compartment sizes can be mapped to prevalence data with some emission distribution or loss function. To integrate incidence data, we can keep track of the cumulative incidence (e.g., by counting the cumulative number of $S \to I$ transitions) at each time point and compute the difference between successive times. Then these differences can be mapped to incidence data with an emission distribution or loss function. For example, let $N_{SI}(t_l)$ be the cumulative number of $S \to I$ transitions by time $t = l$, as tracked by the compartmental model. Then $\Delta N_{SI}(t_l) = N_{SI}(t_l) - N_{SI}(t_{l-1})$ is the number new $S \to I$ transitions between time $t = l$ and time $t = l - 1$. We can link this to C_l, the number of observed cases between time $t = l$ and time $t = l - 1$ by the following distributional assumption:

$$C_l \sim \text{Negative binomial} \left(\mu_l = \rho \times \Delta N_{SI}(t_l), \ \sigma_l^2 = \mu_l(1 + \mu_l/\phi) \right), \tag{2.9}$$

where μ_l and σ_l^2 are the mean and variance of this negative binomial distribution, ρ is a case reporting probability, reflecting that not all people who become infected with a disease are tested, and ϕ is an over-dispersion parameter.

The specifics of the data and compartmental model used can make the mapping from the latent compartments to the observed data challenging. In some cases, the distinction between incidence and prevalence data is not clear. For example, when someone receives a positive test result for an infectious disease with a relatively short period of infectiousness, they may receive several tests in the following weeks to assess whether they are still infectious. In that case, each positive test is not necessarily reflective of a new disease case. The link between positive test results and the underlying cases may also be complicated by imperfect specificity and sensitivity of the tests. Similarly, there is ambiguity about the disease's relevance to an event reported in the data. Someone admitted to the hospital with a broken leg may also

have an infectious disease and be reported as part of the hospital's passive surveillance data. Thus, the hospitalization data may not be truly reflective of the number of people with severe disease cases. Often, there is a significant delay between when data is captured and when it is reported. When performing inference or predictions in real time, it may appear that there are fewer cases in one week than in the previous week, simply due to the most recent week's tests not yet being fully reported. Furthermore, reporting practices for passive surveillance data may be inconsistent. The inconsistency could be due to any number of reasons. For example, many testing locations may not be open on the weekends or holidays, making it appear that new cases are less common on those days. Inconsistencies may also be due to policy changes, changes in funding, shifting demographics among the infected population, or even accidental non-reporting.

While overcoming each of the above individual complications is itself difficult and requires modelers to make many subtle choices, there is additional complexity involved in integrating each of the data streams in a unified model. This additional complexity can be worth pursuing, as integrating more data sources may make more model parameters identifiable. De Angelis et al. [2015] present four important challenges in data integration when modeling infectious diseases. The first is the weighting of evidence. When combining data streams of varying quality, it is desirable to rely more on the high-quality data than the low-quality data. This can be done simply by excluding the low-quality data altogether, or by formally modeling data quality issues. Under the umbrella of model criticism, the authors present identifiability, conflict, and influence as areas of concern. In Section 1.2, we provided an example of different data sources presenting conflicting evidence about underlying disease dynamics. Understanding how each data stream informs modeling results and how their apparent differences are reconciled is a key challenge. In addition, the authors cite issues related to computational complexity when combining many data sources, as well as potential problems of dependencies between the datasets included in the model.

2.4 Forecasting Infectious Disease Outbreaks

While compartmental models are a common framework for forecasting infectious disease outbreaks, other methods are available for the task. A sample of these methods are gathered in the COVID-19 Forecast Hub, which collected and synthesized forecasts from a variety of modelers with diverse approaches throughout the COVID-19 pandemic [Cramer et al., 2022]. Among the methods presented are those using compartmental models, along with approaches common to other areas of statistics, such as time series and machine learning methods. In addition, Cramer et al. [2022] evaluate an ensemble forecast, which combines the models into a single forecast, which is generally shown to have superior performance to any of the individual models. These methods typically use the data described in Section 2.3.2, but some incorporate additional data sources, such as mobility data collected from mobile phones.

2.4.1 Forecast Assessment

Methods used for forecast assessment differ based on the format of the forecasts. We begin with detailed exploration of forecasts in the form of full probability distributions, which we use throughout this work. We end this section with a brief overview of other forecast evaluation methods.

Probabilistic forecasts are probability distributions over future events. In this section, let $F(x)$ denote the cumulative distribution function of a forecast. We assume that new data are realized samples from some underlying probability distribution G, which is only known to nature. Thus, an ideal forecast is $F = G$. Because G is unknown, assessing whether $F = G$ is impossible. The sharpness principle conjectures that issuing a forecast that maximizes sharpness, subject to calibration, is equivalent to making an ideal forecast [Gneiting et al.,

2007].

In plain terms, a forecast is well-calibrated if events that are forecasted to have a $y\%$ chance of happening, are actually observed $y\%$ of the time. Being well-calibrated is insufficient for a forecast to be "good." For example, consider the "climatological" forecaster, who issues forecasts based on the marginal distribution of the predictand. In the context of infectious disease modeling, a climatological forecaster could issue a forecast for daily influenza cases by issuing the same forecast every day, where their forecast is simply a smoothed estimate of daily influenza cases from the last decade. Assuming that influenza case counts in the future are, on average, distributed as they were in the last decade, this forecaster will be well-calibrated. However, competing forecasters should be able to make narrower forecasts while maintaining calibration, perhaps by using data from the preceding week to predict case counts the next week. These forecasts are clearly more useful and are said to be "sharper."

In practice, calibration and sharpness can be assessed by both numerical and graphical summaries. Calibration is often assessed by computing the probability integral transform (PIT), $p_t = F_t(x_t)$ for each forecast, F_t, and its corresponding observed value, x_t, for all forecast times t. If the forecasts are well-calibrated, $p_t \sim \text{Uniform}(0, 1)$ and the p_t's will exhibit no autocorrelation. Visual inspection of a histogram of PIT values for uniformity can, therefore, diagnose issues of miscalibration. Calibration can also be assessed by computing the empirical coverage of prediction intervals. Sharpness can be assessed by visually inspecting the predictive distributions or by computing summaries like interval widths or variance of the distribution.

To numerically assess both sharpness and calibration, a strictly proper scoring rule can also be used. We denote a score for a probabilistic forecast F and observed outcome x by $s(F, x)$. Generally, the forecaster wishes to minimize the score, which can be considered a penalty or cost function. A scoring rule is said to be proper when the expected value of the score is minimized when $F = G$, where $x \sim G$. The scoring rule is strictly proper when this

minimum is unique.

For continuous variables, the continuous ranked probability score is a popular scoring rule with attractive properties:

$$\text{CRPS}(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbb{1}(y \geq x))^2 \, dy. \tag{2.10}$$

In contrast to other scoring rules, it is based on the cumulative distribution function of the predictive distribution, rather than the density function, which may not exist in some contexts. Additionally, the CRPS is sensitive to distance, meaning that it gives credit to forecasts which assign high probability to values near x, even if they do not assign high probability to x itself. For a discussion of scoring rules, see [Gneiting and Raftery, 2007].

Predictive distributions arise naturally from the posterior predictive distribution in the Bayesian setting. The posterior predictive distribution is the distribution of possible unobserved values, $\mathbf{X}^*$, conditional on the observed values, $\mathbf{X}$. Following the notation from Section 2.1, the posterior predictive distribution is $\pi(\mathbf{X}^* \mid \mathbf{X}) = \int \pi(\mathbf{X}^* \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \mathbf{X}) \, d\boldsymbol{\theta}$. When working in the context of MCMC (Section 2.1), the posterior predictive distribution, F, is not available in closed form. Rather, we only have some finite number of samples from the posterior distribution, $(\boldsymbol{\theta}_i, \ldots \boldsymbol{\theta}_n)$. Krüger et al. [2021] present three options for using these samples from MCMC procedures to produce approximated probabilistic forests.

The first is to use the mixture-of-parameters method. Often, the predictive distribution, conditional on $\boldsymbol{\theta}$, is available in closed form. Call this $F_c(x \mid \boldsymbol{\theta})$. We can approximate F by

$$\hat{F}^{\text{MP}}(x) = \frac{1}{n} \sum_{i=1}^{n} F_c(x \mid \boldsymbol{\theta}_i). \tag{2.11}$$

Alternatively, we can work directly with samples from the posterior predictive, $(x_1^*, \ldots, x_n^*)$.

We could estimate F by the empirical cumulative distribution function of these samples:

$$\hat{F}^{\text{ECDF}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left[x_i^* \leq x\right]. \tag{2.12}$$

We could also estimate the predictive density, f by a kernel density estimate of the posterior predictive samples.

The authors recommend using the mixture-of-parameters method when possible, as it tends to produce scores closer to those obtained using the true posterior predictive distribution. When the mixture-of-parameters method cannot be used, the authors advocate for the use of the ECDF method over the kernel density method.

Forecasts may be represented with less detail than a full probabilistic forecast distribution (or samples from such a distribution). If, instead, only quantiles of the forecast distribution are provided, it is common to use the pinball loss function to assess forecast quality:

$$\text{Pinball-Loss}(f_\alpha, x) = (\mathbb{1}\{x \leq f_\alpha\} - \alpha)(f_\alpha - x) = \begin{cases} (1-\alpha)(f_\alpha - x) & x \leq f_\alpha, \\ \alpha(x - f_\alpha) & x \geq f_\alpha, \end{cases}, \tag{2.13}$$

where x is the observed value and f_α is the α quantile of the predictive distribution. When $\alpha = 0.5$, this is the same as $1/2$ of the absolute error. When $\alpha \neq 0.5$, the penalty is weighted based on the probability of $f_\alpha - x$ being positive or negative.

When forecasts are presented in an interval format (or intervals are constructed from quantiles), the Winkler score is commonly used:

$$\text{Winkler}_\alpha(l_\alpha, u_\alpha, x) = (u_\alpha - l_\alpha) + \frac{2}{\alpha}(l_\alpha - x)\mathbb{1}\{x < l_\alpha\} + \frac{2}{\alpha}(x - u_\alpha)\mathbb{1}\{x > u_\alpha\}, \tag{2.14}$$

where $[l_\alpha, u_\alpha]$ is a $100(1-\alpha)\%$ prediction interval. In [Bracher et al., 2021], a weighted interval score is developed for combining Winkler scores at several levels of α. A recent

overview of methods for evaluating forecast quantiles is available in [Gneiting et al., 2023]. Further discussion of all of the above scoring methods is presented by Gneiting and Raftery [2007].

When only point forecasts are available, many standard loss functions are applicable to measure the error between an observed value, x, and a forecasted value, $\hat{x}$. The most common of these are the absolute error, $|x - \hat{x}|$, and the squared error, $(x - \hat{x})^2$. These measures are scale-dependent, making them inappropriate for comparing forecasts on quantities of different units. To overcome this limitation, the absolute percentage error $|100 \, (x - \hat{x}) \, /x|$ can be used. A detailed discussion of error measurements available for point forecasts is presented in [Hyndman and Koehler, 2006].

When assessing any of the measures of forecast quality mentioned in this section, it is useful to aggregate them at a fixed forecast horizon. For example, for a j-week forecast horizon, we create a forecast for week i using all data up to week $i - j$. After repeating this for all weeks of interest, we can compute the relevant score or loss for each forecast. We can then compare multiple forecasting procedures by comparing the average score or loss for each model at the given forecast horizon.

Chapter 3

Confidence intervals for prevalence estimates from complex surveys with imperfect assays

3.1   Introduction

Estimating and quantifying uncertainty for disease prevalence is a standard task in epidemiology. For rare events, these estimates are highly sensitive to misclassification [Hemenway, 1997], making adjustments for sensitivity and specificity critically important. While estimating prevalence (or any event proportion in a population) in complex surveys and adjusting estimates for misclassification have been well studied separately, performing both of these tasks simultaneously remains relatively unexplored. This paper fills that gap by studying estimates and confidence intervals for prevalence from complex surveys with misclassification. We develop a new tractable confidence interval procedure designed to guarantee coverage. Our confidence interval combines three statistical methods: (1) the gamma confidence in-

terval for directly standardized rates is applied to the apparent prevalence [Fay and Feuer, 1997] and (2) standard misclassification adjustments to that apparent prevalence for sensitivity and specificity [Rogan and Gladen, 1978] are used together with (3) the melding method, which allows combination of confidence intervals for different parameters [Fay et al., 2015]. We show by simulation only (there are no proofs of guaranteed coverage) that, with a complex survey with low prevalence and misclassification, our method appears to have at least nominal coverage. Even in special cases (e.g., simple random samples with imperfect assays, or weighted samples with perfect assays) where there are existing methods, our new method has at least nominal simulated coverage. while some of the tractable existing methods may under-cover. The cost for apparently achieving nominal coverage is that our intervals may be quite wide in some situations.

Recent overviews of confidence interval procedures for prevalence in surveys without misclassification are provided by Dean and Pagano [2015] and Franco et al. [2019]. For simple random sample surveys with imperfect sensitivity and specificity, Lang and Reiczigel [2014] proposed an approximate confidence interval that performed well in simulations. Recent works by DiCiccio et al. [2022] and Cai et al. [2022] study both valid (i.e., exact) and approximate intervals. Their valid intervals use test inversion and the adjustment of [Berger and Boos, 1994], while their approximation intervals use the bootstrap with the test inversion approach. Fewer methods are available for constructing frequentist confidence intervals for prevalence estimates from complex surveys while adjusting for sensitivity and specificity. Kalish et al. [2021] developed one such method that is closely related to one of the methods presented here, but that method's properties were not studied. Cai et al. [2022] (see also discussion in [DiCiccio et al., 2022]) modify their approximation approach to allow sample weights for strata or individual specific weights. Another recent advancement is the method developed by Rosin et al. [2023] that makes use of asymptotic normal approximations, which reduce to the Wald interval when sensitivity and specificity are perfect. This problem has also previously been addressed in Bayesian literature, recently by Gelman and Carpenter

[2020].

We work up to our ultimate goal in stages. First, in Section 3.2.2, we propose confidence intervals for simple random samples where prevalence is assessed with an assay with imperfect sensitivity and/or specificity. Next, in Section 3.2.3, we present confidence intervals for weighted samples where prevalence is assessed with an assay without misclassification. In Section 3.2.4, we combine these methods to create confidence intervals for weighted samples where prevalence is assessed with an assay with imperfect sensitivity and specificity. Because the combined method reduces to one of the first two methods as a special case, we can think of the first two stages as testing the combined method in those cases. Finally, in Section 3.2.5 we show how certain complex surveys may fit into the format for our new method.

In simulations, we compare our method to established frequentist competitors and show through simulations that it beats the best of those in each of the three stages with respect to guaranteeing coverage. However, in some simulated settings, our proposed methods are overly conservative, meaning that they demonstrate higher than nominal coverage, while competitor methods maintain closer to nominal coverage. We did not include in our simulations some recent methods that have been developed in response to the COVID-19 pandemic [Cai et al., 2022, DiCiccio et al., 2022, Rosin et al., 2023]. The exact method of [DiCiccio et al., 2022] would guarantee coverage, although applying it to a survey with a large number of strata would be "computationally expensive," and it has not been applied to surveys using post-stratification weighting. In contrast, our new method can very tractable in those situations.

3.2 Confidence Interval Methods

3.2.1 Notation and Problem Set-up

To introduce notation, consider first the stratified simple random sample. Suppose we have a population partitioned into K strata, with $N_1, N_2, \ldots, N_K$ individuals in the K strata of the population. We sample $n_1, n_2, \ldots, n_K$ individuals via a simple random sample from each of the K strata to have an assay performed on each individual to determine who has the disease. Let X_i be the number of positive results from an assay performed on the n_i individuals from stratum i and assume $X_i \sim \text{Binomial}(n_i, \theta_i)$, where θ_i is the population frequency of positive results for assays performed on individuals from stratum i. Similarly, let X_i^* be the unobserved true number of people with the disease among the n_i individuals from stratum i and assume $X_i^* \sim \text{Binomial}(n_i, \theta_i^*)$, where θ_i^* is the population frequency of cases in stratum i. In the case of a perfect assay, $\theta_i = \theta_i^*$.

Therefore, the population prevalence is

$$\beta^* = \frac{\sum_{i=1}^K N_i \theta_i^*}{\sum_{j=1}^K N_j} = \sum_{i=1}^K w_i \theta_i^*, \tag{3.1}$$

and the apparent prevalence is

$$\beta = \frac{\sum_{i=1}^K N_i \theta_i}{\sum_{j=1}^K N_j} = \sum_{i=1}^K w_i \theta_i, \tag{3.2}$$

where $w_i = N_i / \sum_{j=1}^K N_j$ and, therefore, $\sum_{i=1}^K w_i = 1$. This set-up will approximately work

for other complex survey samples, where we can estimate survey weights such that the complex survey sample may be treated as a multinomial sample with probabilities proportional to those weights (see Section 3.2.5).

We can relate θ_i and θ_i^* using the sensitivity (ϕ_p) and specificity (1-ϕ_n) of the assay, where ϕ_p and ϕ_n are the proportion of positive assays from a population of positive controls (i.e., individuals known to have the disease) and negative controls (i.e., individuals known to be without the disease), respectively. Then $\theta_i = \phi_p \theta_i^* + \phi_n(1 - \theta_i^*)$, or equivalently [Rogan and Gladen, 1978],

$$\theta_i^* = \frac{\theta_i - \phi_n}{\phi_p - \phi_n},$$

and we have

$$
\begin{aligned}
\beta^* &= \sum_{i=1}^K w_i \theta_i^* = \sum_{i=1}^K w_i \left(\frac{\theta_i - \phi_n}{\phi_p - \phi_n} \right) \\
&= \frac{\sum_{i=1}^K w_i \theta_i}{\phi_p - \phi_n} - \frac{\phi_n \sum_{i=1}^K w_i}{\phi_p - \phi_n} = \frac{\sum_{i=1}^K w_i \theta_i}{\phi_p - \phi_n} - \frac{\phi_n}{\phi_p - \phi_n}
\end{aligned}
\tag{3.3}
$$

Suppose the assay is measured on m_n individuals known not to have the disease and on m_p individuals known to have the disease. Let C_n and C_p be the number who test positive from the respective samples. Assume that the negative and positive controls act like simple random samples from their respective populations. Thus, $C_n \sim \text{Binomial}(m_n, \phi_n)$ where $1 - \phi_n$ is the specificity of the assay, and $C_p \sim \text{Binomial}(m_p, \phi_p)$, where ϕ_p is the sensitivity of the assay. Let $\hat{\theta}_i = \frac{X_i}{n_i}$, $\hat{\phi}_n = \frac{C_n}{m_n}$, and $\hat{\phi}_p = \frac{C_p}{m_p}$. Then a plug-in estimator for β^* is

$$\hat{\beta}^* = \frac{\sum_{i=1}^K w_i \hat{\theta}_i}{\hat{\phi}_p - \hat{\phi}_n} - \frac{\hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n}. \tag{3.4}$$

This estimator serves as an important basis for developing confidence intervals in this work. Section 3.2.2 is concerned with confidence intervals for β^* in the case where $K = 1$, $\phi_n > 0$, $\phi_p < 1$, i.e., estimating prevalence from a simple random sample with an imperfect assay. Section 3.2.3 is concerned with confidence intervals for β^* in the case where $K > 1$, $\phi_n = 0$, $\phi_p = 1$, i.e., estimating prevalence from a weighted sample with a perfect assay. Section 3.2.4 is concerned with confidence intervals for β^* in the case where $K > 1$, $\phi_n > 0$, $\phi_p < 1$, i.e., estimating prevalence from a weighted sample with an imperfect assay.

3.2.2 Estimating Prevalence from a Simple Random Sample with an Imperfect Assay

First we consider the scenario where $K = 1$, $\phi_n > 0$, and $\phi_p < 1$. We develop a confidence interval for the population prevalence, β^*. When $K = 1$, the estimand in Equation 3.3 becomes $\beta^* = (\theta_1 - \phi_n)/(\phi_p - \phi_n)$. We have $\phi_p > \phi_n$ for any useful assay, and since the sample is a mixture of individuals with and without the disease of interest, $\phi_p \geq \theta_1 \geq \phi_n$. The estimator of β^* is

$$
\hat{\beta}^* \equiv g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) \equiv
\begin{cases}
1 & \text{if } \hat{\phi}_n < \hat{\phi}_p \leq \hat{\theta}_1 \\
\frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} & \text{if } \hat{\phi}_p > \hat{\theta}_1 \geq \hat{\phi}_n \\
0 & \text{otherwise.}
\end{cases}
\tag{3.5}
$$

To create a confidence interval for $\hat{\beta}^*$, we use a generalization of the melding method [Fay et al., 2015], which makes use of lower and upper confidence distributions on functions of independent estimators to account for variability in $\hat{\theta}_1$, $\hat{\phi}_n$, and $\hat{\phi}_p$. Before describing the melding method, we briefly review confidence distributions.

Confidence distributions are frequentist estimators of a parameter, which can be used similarly to the bootstrap or posterior distributions. For continuous responses (and asymptotically for discrete responses), one can formally define confidence distributions without relying on previously derived confidence interval procedures [Xie and Singh, 2013], but for intuition with discrete responses, it is easier to derive confidence distributions the classical way, i.e., directly from confidence interval procedures. Here, we reproduce the definition of confidence distributions from Section 2.2. For discrete responses, there are two confidence distributions (the lower and upper ones), used to ensure validity of the resulting inferences. Under the classical derivation, we can create random variables associated with the lower and upper confidence distributions from two one-sided, nested, confidence interval procedures, where for a nested confidence interval procedure, the $(1 - \alpha_1)$ interval is a subset of the $(1 - \alpha_2)$ interval whenever $(1 - \alpha_1) \leq (1 - \alpha_2)$. Denote those one-sided intervals by $[L(\mathbf{x}, 1 - \alpha), \infty)$ and $(-\infty, U(\mathbf{x}, 1 - \alpha)]$, which are defined for any observed data $\mathbf{x}$, and $(1 - \alpha) \in (0, 1)$. For fixed $\mathbf{x}$, define the lower and upper confidence distribution random variables as $L(\mathbf{x}, A)$ and $U(\mathbf{x}, B)$, where A and B are independent uniform random variables. Then a $100(1 - \alpha)\%$ central confidence interval has the $(\alpha/2)$th quantile of $L(\mathbf{x}, A)$ as the lower limit, and the $(1 - \alpha/2)$th quantile of $U(\mathbf{x}, B)$ as the upper limit.

Because each estimated component in Equation 3.5 is a binomial probability parameter, we now focus on the confidence distributions associated with the exact binomial confidence interval. For a binomial experiment with x successes out of n trials, the lower confidence distribution is $\mathrm{Beta}(x, n - x + 1)$ with associated random variable B^L, and the upper confidence distribution is $\mathrm{Beta}(x + 1, n - x)$ with random variable B^U, where for $a > 0$ we let $\mathrm{Beta}(0, a)$ and $\mathrm{Beta}(a, 0)$ be point masses at 0 and 1, respectively. This result comes from the relationship between the binomial and beta distributions, namely $\Pr[X \leq x] = \Pr[B > \theta]$, where $B \sim \mathrm{Beta}(x + 1, n - x)$ for $x \in \{-1, 0, 1, \ldots, n\}$ (see e.g., [Blyth, 1986, Section 2], [Fay et al., 2022, Section S2]). Additionally, in the binomial case, the lower and upper confidence distributions are equivalent to the posterior distributions that result from using

well-calibrated null preference priors [Fay et al., 2022]. Let $q(a, W)$ be the ath quantile of a random variable W. Then the exact $1 - \alpha\%$ central confidence interval of [Clopper and Pearson, 1934] for the binomial parameter is

$$\left\{ q\left(\frac{\alpha}{2}, B^L\right), q\left(1 - \frac{\alpha}{2}, B^U,\right)\right\}. \tag{3.6}$$

Fay et al. [2015] proposed the melding method for obtaining confidence intervals for functions of two parameters that are monotonic within the allowable range for each parameter, given the other is fixed. The melding method uses quantiles of those functions except replacing lower or upper confidence distribution random variables for their associated parameter in the function. Here, we generalize that method to β^*, which is a function of three parameters. When $1 \geq \phi_p > \theta_1 > \phi_n \geq 0$ then β^* is monotonically increasing in θ_1, monotonically decreasing in ϕ_p, and monotonically decreasing in ϕ_n. For an assessment of monotonicity in other scenarios, see Appendix A.1. Then the $1 - \alpha\%$ confidence interval for β^* is

$$\left\{ q\left(\frac{\alpha}{2}, g\left\{B^L_{\theta_1}, B^U_{\phi_n}, B^U_{\phi_p}\right\}\right), q\left(1 - \frac{\alpha}{2}, g\left\{B^U_{\theta_1}, B^L_{\phi_n}, B^L_{\phi_p}\right\}\right)\right\}. \tag{3.7}$$

where $g(\cdot)$ is defined in equation 3.5. In equation 3.7, the choice of confidence distribution random variable (lower or upper) is determined by the associated the direction of the monotonicity to attempt to ensure coverage. As the sample sizes (e.g., n_i, m_p, and m_n) increase, there is little difference between the lower and upper confidence distributions.

The quantiles of these melded distributions are calculated by Monte Carlo sampling from each of the component distributions. We compare this method to one described in [Lang and Reiczigel, 2014] as implemented in prevSeSp function in the `asht` R package [Fay, 2020], which provides approximate confidence intervals for true prevalence when sensitivity and

specificity are estimated from independent samples, as they are in this section. The Lang-Reiczigel interval is given by

$$\beta_1^{*\prime} + d\beta \pm q\left(1 - \frac{\alpha}{2}, Z\right) \cdot \mathrm{Var}(\beta_1^{*\prime})^{1/2}, \tag{3.8}$$

where $q_Z \equiv q\left(1 - \frac{\alpha}{2}, Z\right)$, $Z \sim N(0,1)$,

$$d\beta = 2 \cdot q_Z^2 \cdot \left\{ \beta_1^{*\prime} \cdot \frac{\phi_p'(1 - \phi_p')}{m_p'} - (1 - \beta_1^{*\prime}) \cdot \frac{(1 - \phi_n')\phi_n'}{m_n'} \right\},$$

$$\mathrm{Var}(\beta_1^{*\prime}) = \frac{\frac{\beta_1^{*\prime}(1 - \beta_1^{*\prime})}{n_1} + (\beta_1^{*\prime})^2 \frac{\phi_p'(1 - \phi_p')}{m_p} + (1 + \beta_1^{*\prime})^2 \frac{(1 - \phi_n')\phi_n'}{m_n}}{(\phi_p' - \phi_n')^2},$$

$$m_p' = m_p + 2, \qquad\qquad m_n' = m_n + 2,$$

$$\phi_p' = \frac{m_p \cdot \hat{\phi}_p + 1}{m_p + 2}, \qquad\qquad 1 - \phi_n' = \frac{m_n \cdot (1 - \hat{\phi}_n) + 1}{m_n + 2},$$

$$\beta_1^{*\prime} = \frac{\beta_1' - \phi_n'}{\phi_p' - \phi_n'}, \qquad\qquad \beta_1' = \frac{n_1 \cdot \hat{\theta}_1 + q_Z^2/2}{n_1 + q_Z^2}.$$

3.2.3 Estimating Prevalence from a Weighted Sample with a Perfect Assay

Next, we present a confidence interval for the population prevalence, β^*, in the scenario $K > 1$, $\phi_n = 0$, $\phi_p = 1$. Our method is a straightforward adaptation of the gamma confidence interval presented in [Fay and Feuer, 1997], which was developed to create confidence intervals for a standardized population rate which is assumed to be a weighted sum of Poisson rate

parameters. We note that for sufficiently large sample size n and small rate λ, a Poisson$(n\lambda)$ distribution is approximately equal in distribution to a Binomial(n, λ) distribution. Under this Poisson assumption, we suggest the $100(1 - \alpha)\%$ gamma confidence interval for β^*:

$$\left(q\left(\frac{\alpha}{2}, G_{\beta^*}^L\right), q\left(1 - \frac{\alpha}{2}, G_{\beta^*}^U\right) \right), \tag{3.9}$$

where

$$
\begin{aligned}
G_{\beta^*}^L &\sim \text{Gamma}\left(\frac{y^2}{v}, \frac{v}{y}\right), & G_{\beta^*}^U &\sim \text{Gamma}\left(\frac{y^{*2}}{v^*}, \frac{v^*}{y^*}\right), \\
y &= \sum_{i=1}^K \frac{w_i}{n_i} x_i, & v &= \sum_{i=1}^K \left(\frac{w_i}{n_i}\right)^2 x_i, \\
y^* &= y + \max\left(\frac{w_1}{n_1}, \ldots, \frac{w_K}{n_K}\right), & v^* &= v + \left\{\max\left(\frac{w_1}{n_1}, \ldots, \frac{w_K}{n_K}\right)\right\}^2.
\end{aligned}
$$

We call this the wsPoisson method, since it assumes a weighted sum of Poissons. We compare the wsPoisson confidence interval to two methods presented in [Dean and Pagano, 2015], which were recommended for scenarios with low prevalence. Dean and Pagano showed in simulations that the standard Wald interval had poor coverage with low prevalence (e.g., Fig. 1 of that paper showed 95% confidence intervals with coverage of less than 85% for prevalence values less than 2%). Since the confidence interval of [Rosin et al., 2023] reduces to the Wald interval with perfect assays, we will not include that method in the simulation comparisons.

The first recommended method of [Dean and Pagano, 2015] is an adaptation of the method of [Agresti and Coull, 1998] for the survey setting. The interval for β^* is given by:

$$\tilde{p} \pm q_Z \sqrt{\tilde{p}(1-\tilde{p})/\tilde{n}}, \tag{3.10}$$

where

$$\tilde{x} = \left(\sum_{i=1}^{k} w_i \hat{\theta}_i\right) n_{\text{eff}} + c, \qquad\qquad \tilde{n} = n_{\text{eff}} + 2c,$$

$$\tilde{p} = \tilde{x}/\tilde{n}, \qquad\qquad c = q_Z^2/2,$$

$$n_{\text{eff}} = \frac{\left(\sum_{i=1}^{k} w_i \hat{\theta}_i\right)\left(1 - \sum_{i=1}^{k} w_i \hat{\theta}_i\right)}{\sum_{i=1}^{k} \frac{w_i^2}{n_i} \hat{\theta}_i}. \tag{3.11}$$

In the case where $\sum_{i=1}^{k} \frac{w_i^2}{n_i} \hat{\theta}_i = 0$, we instead let $n_{\text{eff}} = \sum_{i=1}^{k} n_i$.

We also compare our suggested method to Dean and Pagano's modification of the method of [Korn and Graubard, 1998, Dean and Pagano, 2015]. This interval is given by

$$\left(q\left(\frac{\alpha}{2}, B_{KG}^L\right), q\left(1 - \frac{\alpha}{2}, B_{KG}^U\right)\right), \tag{3.12}$$

where, analogously to the Clopper-Pearson interval (see Equation 3.6),

$$B_{KG}^L \sim \text{Beta}\left(x_{\text{eff}}, n_{\text{eff}} - x_{\text{eff}} + 1\right), \qquad\qquad B_{KG}^U \sim \text{Beta}\left(x_{\text{eff}} + 1, n_{\text{eff}} - x_{\text{eff}}\right),$$

with $x_{\text{eff}} = n_{\text{eff}} \sum_{i=1}^{k} w_i \hat{\theta}_i$, and n_{eff} defined in Equation 3.11. Although Dean and Pagano [2015] expressed this in terms of F-distributions, the beta distribution representation is equivalent.

3.2.4   Estimating Prevalence from a Weighted Sample with an Imperfect Assay

Lastly, we develop a confidence interval for the population prevalence, β^*, in the case where $K > 1$, $\phi_n > 0$, $\phi_p < 1$.

The two methods we discuss are closely related to each other and the methods discussed in Sections 3.2.2 and 3.2.3. As in Section 3.2.2, we use the melding method [Fay et al., 2015] to create $1 - \alpha\%$ confidence interval very similar to Equation 3.7. The confidence distributions for ϕ_p and ϕ_n are the same Beta distributions as in Section 3.2.2. The two methods differ in their confidence distributions for the apparent prevalence β.

In the first case, we use the adaptation of the gamma confidence interval [Fay and Feuer, 1997] presented in Section 3.2.3 to derive the $1 - \alpha\%$ confidence interval for β^*:

$$\left\{ q\left(\frac{\alpha}{2}, g\left[G_{\beta^*}^L, B_{\phi_n}^U, B_{\phi_p}^U \right] \right), q\left(1 - \frac{\alpha}{2}, g\left[G_{\beta^*}^U, B_{\phi_n}^L, B_{\phi_p}^L \right] \right) \right\}, \tag{3.13}$$

where $G_{\beta^*}^L$ and $G_{\beta^*}^U$ are defined in Section 3.2.3. We refer to this method as the WprevSeSp Poisson - weighted prevalence with sensitivity and specificity, where the prevalence confidence distribution is based on the weighted sum of Poissons.

The alternative method is very similar to that used in [Kalish et al., 2021]. We use the

modification of the method from [Korn and Graubard, 1998] presented by Dean and Pagano [2015] as in Section 3.2.3, to derive the $100(1 - \alpha)\%$ confidence interval for β^*:

$$\left\{ q\left(\frac{\alpha}{2}, g\left[B_{KG}^L, B_{\phi_n}^U, B_{\phi_p}^U\right]\right), q\left(1 - \frac{\alpha}{2}, g\left[B_{KG}^U, B_{\phi_n}^L, B_{\phi_p}^L\right]\right)\right\}, \tag{3.14}$$

where B_{KG}^L and B_{KG}^U are as defined in Section 3.2.3. Although equivalent, this expression looks different than the one used in [Kalish et al., 2021] because they used a parameter for specificity, rather than ϕ_n, which is 1 minus specificity. We refer to this as WprevSeSp Binomial - weighted prevalence with sensitivity and specificity, where the prevalence confidence distributions are based on a binomial assumption. The WprevSeSp Binomial and WprevSeSp Poisson methods are implemented in the `asht` R package [Fay, 2020].

3.2.5 Applications to More Complex Surveys

When We Can Use These Methods

In Section 3.2.1, we assumed that the apparent prevalence was a weighted sum of binomial random variables, $\beta = \sum_{i=1}^{K} w_i \frac{X_i}{n_i}$, where $X_i \sim \text{Binomial}(n_i, \theta_i)$. Then in Section 3.2.3, we used the fact that for small θ_i and large n_i the binomial can be approximated by the Poisson, giving $X_i \overset{.}{\sim} \text{Poisson}(n_i\theta_i)$. Thus, whenever we can model a complex survey estimator of apparent prevalence as a weighted sum of Poisson variates, then we can apply the methods of this paper.

In the upcoming Section 3.2.5, we give a detailed review relating the multinomial sampling model to a weighted sum of Poisson variates model. The multinomial sampling model treats the survey sample as if it is a sampling with replacement from the entire population of N

individuals, where each of the N individuals has a probability of P_j of being sampled for each of the n samples from the survey, with $\sum_{j=1}^{N} P_j = 1$. Under this model the number of times each of the N individuals is included in the sample is a multinomial with parameters n and $[P_1, \ldots, P_N]$. The multinomial model describes sampling with replacement, but it is nevertheless used to approximate a sampling design where the jth individual is sampled *without* replacement with probability P_j, even though under that design (unlike the multinomial model) no individual is included in the sample more than once. The multinomial model is a common approximation for other complex survey designs.[see e.g., Korn and Graubard, 1999, p. 14]. For example, in the Kalish et al. [2021] analysis of Section 3.4, each individual in the sample is assigned a pseudo-weight approximating one over their sampling probability from a multinomial model. The actual sample was not a probability sample. In fact, it was a quota sample from a very large pool of self-selected volunteers, and the pseudo-weights were calculated using a different large survey that was a probability weighted survey. The pseudo-weights were calculated such that if they were analyzed under the multinomial model, they would adjust for selection bias due to self-selection of the volunteers and the imperfection of the quota sampling.

Multinomial Sampling Model

Let $Y_1, \ldots, Y_N$ be the binary indicators of event in the N individuals in the population of interest, so the prevalence is $\beta = N^{-1} \sum_{j=1}^{N} Y_j$. There are many ways to design a complex survey sample, and it is often useful to analyze them as if individuals were sampled with replacement with the sampling probability of the jth individual equal to P_j, with $\sum_{j=1}^{N} P_j = 1$. In other words, we treat the sample as if it were n independent multinomial samples each with one trial and selection probability vector $[P_1, \ldots, P_N]$. Let $I_{ij} = 1$ if the ith draw for the sample is individual j in the population, and 0 otherwise. Then let $y_i = Y_j$ and $p_i = P_j$ when $I_{ij} = 1$. Here, following the tradition in the survey literature, we use capital letters

for the population of interest (e.g., N, Y, P), and lower case letters for the sample (e.g., n, y, p). In this notation, both y_i and Y_j are fixed, and only the variables representing the sampling (i.e., the I_{ij} variables) are random. Under this independent multinomial model, since $E(I_{ij}) = P_j$, an unbiased estimator of β is

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{Np_i} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\frac{I_{ij}Y_j}{NP_j}, \tag{3.15}$$

and an unbiased estimator of $\text{var}(\hat{\beta})$ under the multinomial model is

$$\widehat{\text{var}}_M(\hat{\beta}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{Np_i} - \hat{\beta}\right)^2 \tag{3.16}$$

(see [Korn and Graubard, 1999] Problem 2.2-10). We can write $\hat{\beta}$ as a weighted sum. Traditional survey weighting defines the weights so that the weight for the ith sampled individual can be interpreted as the number of individuals in the population that the ith sampled individual represents. Following that tradition, let $w_i^{(trad)} = 1/(np_i)$ and $W_j^{trad)} = 1/(nP_j)$, then the expected sum of the sampled weights is N,

$$E\left(\sum_{i=1}^{n}w_i^{(trad)}\right) = E\left(\sum_{i=1}^{n}\sum_{j=1}^{N}I_{ij}W_j^{(trad)}\right) = \sum_{i=1}^{n}\sum_{j=1}^{N}\frac{E(I_{ij})}{nP_j} = \sum_{i=1}^{n}\sum_{j=1}^{N}\frac{1}{n} = N.$$

Sometimes the weights are scaled after selection so that the scaled weights are $w_i^{(strad)} = \frac{Nw_i^{(trad)}}{\sum_{i=1}^{n}w_i^{(trad)}}$ and are forced to sum to N. For example, in [Kalish et al., 2021] rescaling (sometimes called post-stratification) was done in a more complicated manner to ensure that the weights summed to the US census population within age group, sex, race, ethnicity

and region.

For this paper, we define the weights differently because we want to model our estimator as a weighted sum of Poisson random variables. Thus, we use $w_i = 1/(nNp_i)$ and $W_j = 1/(nNP_j)$ so that the sums have expectation 1. In the complex survey case, we start with the independent multinomial model as in equation 3.15, then we use the relationship between the multinomial and Poisson distributions. Using the "multinomial-Poisson transformation", the maximum likelihood estimates (MLE) for a multinomial random variable are equivalent to the MLEs for independent Poisson random variables, and the variances are asymptotically equivalent (see [Baker, 1994]). Even though we model $\hat{\beta}$ using multinomial random variables, where there are many missing values (which occurs in our situation whenever $I_{ij} = 0$), the multinomial-Poisson relationship holds even when there are missing variables (see [Baker, 1994, Section 3]). For both the Poisson and multinomial models, $E(I_{ij}) = P_j$, and $\hat{\beta}$ is unbiased under either model. For the Poisson model, all the I_{ij} are independent, and each mean equals its variance, so that the variance of $\hat{\beta}$ under this model is

$$
\begin{aligned}
\text{var}_P\left(\hat{\beta}\right) &= \text{var}_P\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N}\frac{I_{ij}Y_j}{NP_j}\right) = \sum_{i=1}^{n}\sum_{j=1}^{N}\frac{\text{var}_P(I_{ij})Y_j^2}{n^2N^2P_j^2} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{N}\frac{\text{var}_P(I_{ij})Y_j}{n^2N^2P_j^2} = \sum_{i=1}^{n}\sum_{j=1}^{N}\frac{Y_j}{n^2N^2P_j}.
\end{aligned}
$$

We estimate $\text{var}_P\left(\hat{\beta}\right)$ by multiplying each term in the sum by I_{ij}/P_j, which has an expectation of 1 and eliminates terms of non-selected individuals, giving

$$
\widehat{\text{var}}_P\left(\hat{\beta}\right) = \sum_{i=1}^{n}\sum_{j=1}^{N}\frac{I_{ij}Y_j}{n^2N^2P_j^2} = \sum_{i=1}^{n}\frac{y_i}{n^2N^2p_i^2} \tag{3.17}
$$

Under the Poisson model $\widehat{\mathrm{var}}_P\left(\hat\beta\right)$ is an unbiased estimator of $\mathrm{var}_P\left(\hat\beta\right)$.

3.3   Simulations

We explore simulations in a variety of scenarios; however, despite this variety, we limit the scenarios to cases similar to the application of Section 3.4, which has low prevalence (when the Poisson assumption is adequate), and high sensitivity and specificity (i.e., good assays).

3.3.1   Estimating Prevalence from a Simple Random Sample with an Imperfect Assay

We assess and compare our new method (Melding, i.e., equation 3.7) to that of Lang and Reiczigel (LR) in a variety of simulated settings. In each simulation, 100 subjects are tested to estimate prevalence, 60 are tested to estimate sensitivity, and 300 are tested to estimate specificity. Several combinations of prevalences (0.5%–2%), sensitivities (75%–100%) and specificities (75%–100%) are assessed. Each simulated scenario is replicated 10,000 times. Figure 3.1 compares the two methods based on coverage, while Figures 3.2 and 3.3 present the lower and upper error frequencies for these scenarios, respectively.
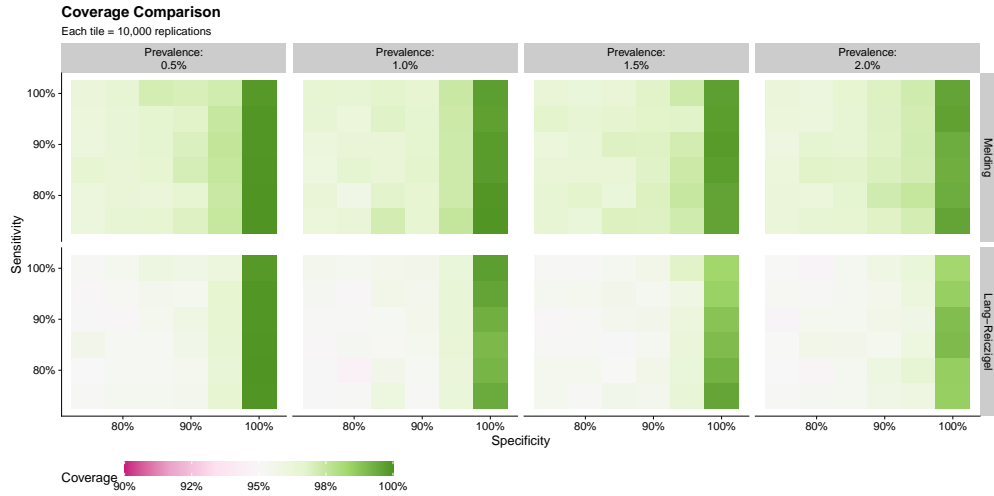
Figure 3.1: Coverage properties of 95% confidence intervals for our new method (Melding) and the Lang-Reiczigel method in a variety of settings, each simulated 10,000 times.

Figure 3.2: Lower error properties of 95% confidence intervals for our new method (Melding) and the Lang-Reiczigel method in a variety of settings, each simulated 10,000 times.
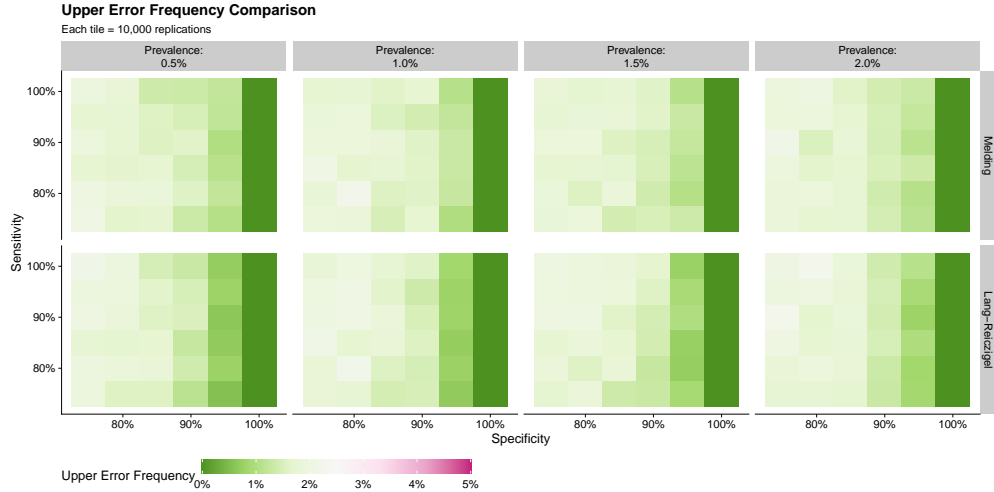
Figure 3.3: Upper error properties of 95% confidence intervals for our new method (Melding) and the Lang-Reiczigel method in a variety of settings, each simulated 10,000 times.

Figure 3.1 shows that, when specificity is less than perfect, the Lang-Reiczigel method achieves approximately nominal coverage, while the melding method is slightly more conservative, generally demonstrating 96%-97% coverage with 95% confidence intervals. When specificity is 100%, both methods are very conservative, achieving nearly 100% coverage. For a more nuanced depiction of each method's properties, we separate the overall coverage into lower and upper errors. Figure 3.3 shows that both methods make upper errors with roughly the same frequency. Figure 3.2 demonstrates that while the melding procedure bounds the lower error frequency below 2.5%, the Lang-Reiczigel method generally has lower error frequency above 2.5%. Each of these behaviors may be undesirable, depending on the context in which the methods are applied. For applications in which there is a need to bound the lower errors, the melding method appears to be superior.

44

3.3.2 Estimating Prevalence from a Weighted Sample with a Perfect Assay

We compare the wsPoisson method to the more traditional Dean-Pagano modification of the Agresti-Coull (DPAC) method and the Korn-Graubard (KG) method for survey proportions in a variety of settings. Our simulations examine varying levels of disease prevalence (0.5% or 5%), different types of survey designs (50 sampling strata with 200 subjects each or 8000 individuals, each with their own weight), distributions of weights among the sampling strata or individuals (coefficient of variation from approximately 0% to nearly 600%), and the number and weights of sampling strata with non-zero prevalence. For each combination of prevalence p, and group type, up to 500 sets of weights are simulated. These 500 sets of weights are designed to span a range of coefficients of variation. For a target coefficient of variation, v, n weights (w_i) are simulated by generating n samples from a Beta $\left(\frac{1}{v^2} - \frac{1}{nv^2} - \frac{1}{n}, \frac{n-1}{v^2} - \frac{n-1}{nv^2} - \frac{n-1}{n} \right)$ distribution and normalizing so that $\sum_{i=1}^n w_i = 1$. This assures that the coefficient of variation among these weights is approximately v. Then, certain weights are chosen to have non-zero prevalence (5%, 25%, or 75% distributed either among the highest weights, lowest weights, or distributed uniformly). These weights with non-zero prevalence are given a prevalence such that $\sum_{i=1}^n w_i \theta_i = p$. For each simulated set of parameters and weights, 10,000 data sets are simulated and assessed.

The coverage properties for these simulations are presented in Figures 3.4–3.7. Additional properties for these simulations are presented in Figures A.1–A.12.

From Figures 3.4–3.7, we note that the two competitor methods generally exhibit lower coverage as the coefficient of variation among the weights increases. In Figure 3.4, this coverage falls below 60% when the prevalence is very low and is concentrated among the highest weights, and the coefficient of variation among the weights exceeds 4. Uniform distribution of prevalence among the weights, increased overall prevalence, and larger sample

sizes among fewer groups all appear to lessen the severity of this problem. In contrast, the wsPoisson method appears to guarantee coverage in all scenarios. The wsPoisson method tends to become more conservative when the coefficient of variation among the weights increases, when the other methods can have problems guaranteeing coverage. In all cases, the wsPoisson method is more conservative than the competitor methods. In scenarios with higher overall prevalence, the Agresti-Coul and Korn-Graubard methods achieve close to nominal coverage, while the wsPoisson method strongly over-covers. This is similar to the behavior observed in [Fay and Feuer, 1997] where, in simulations, the overall error rate for the gamma intervals decreased as the variance of the weights increased. Because our methods appear to be very conservative, with coverage near 100% in some cases, we present the widths of the confidence intervals in Figures A.9–A.12. In scenarios where coefficient of variation among the survey weights is high, the wsPoisson intervals are often two or three times wider than intervals produced by competing methods.

Figure 3.4: Coverage properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 50 groups of 200 people are sampled. The horizontal dashed line indicates the nominal coverage, 95%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure 3.5: Coverage properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence where 8000 individuals are sampled. The horizontal dashed line indicates the nominal coverage, 95%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.
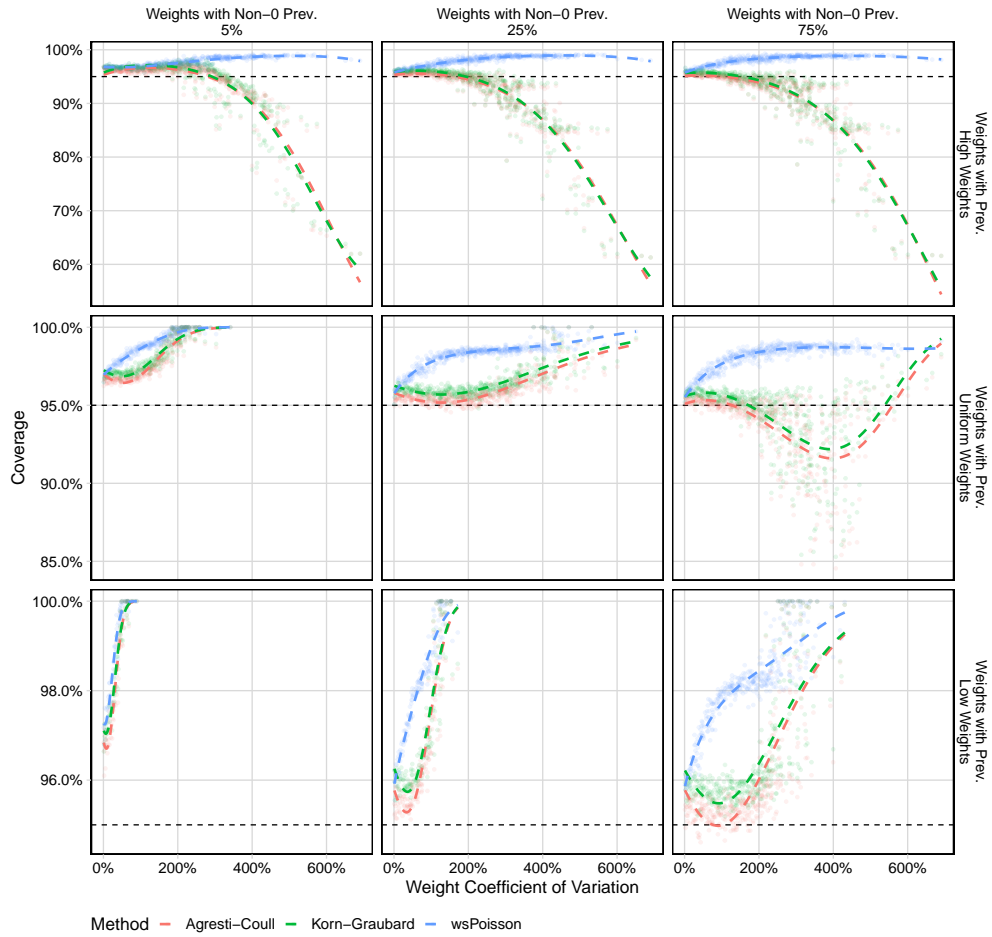
Figure 3.6: Coverage properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 50 groups of 200 people are sampled. The horizontal dashed line indicates the nominal coverage, 95%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.
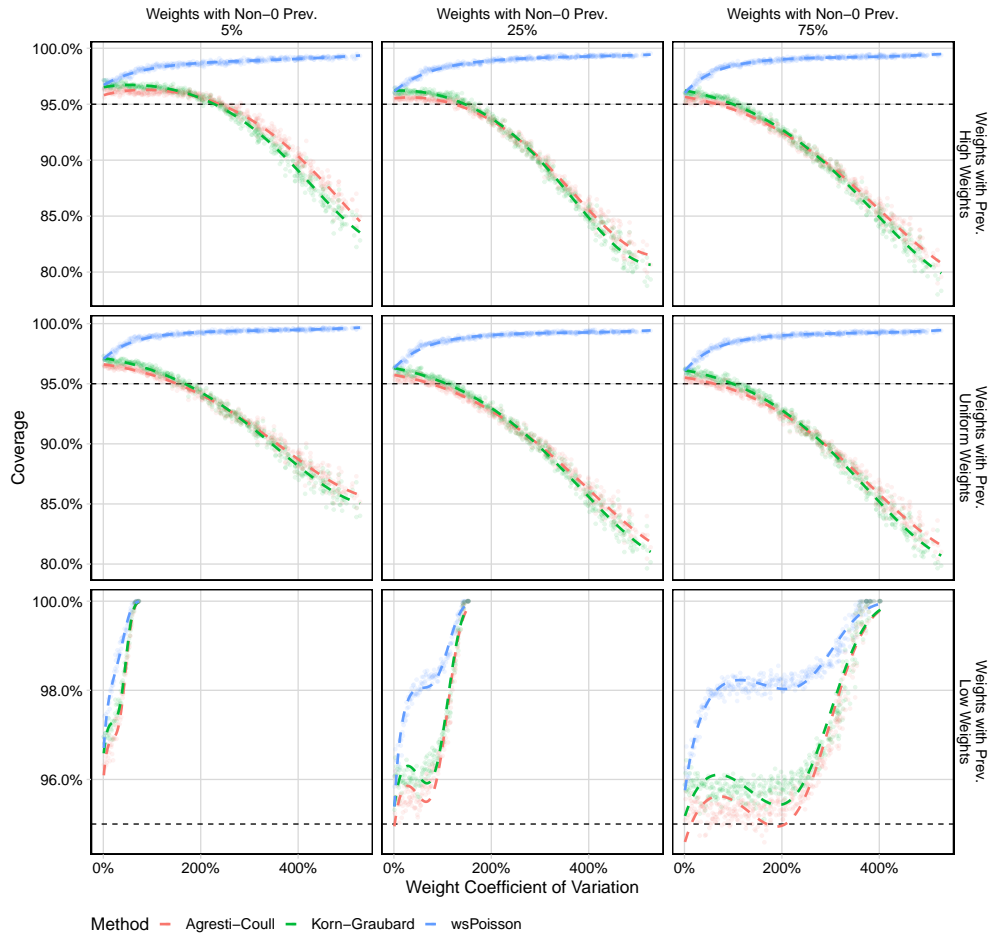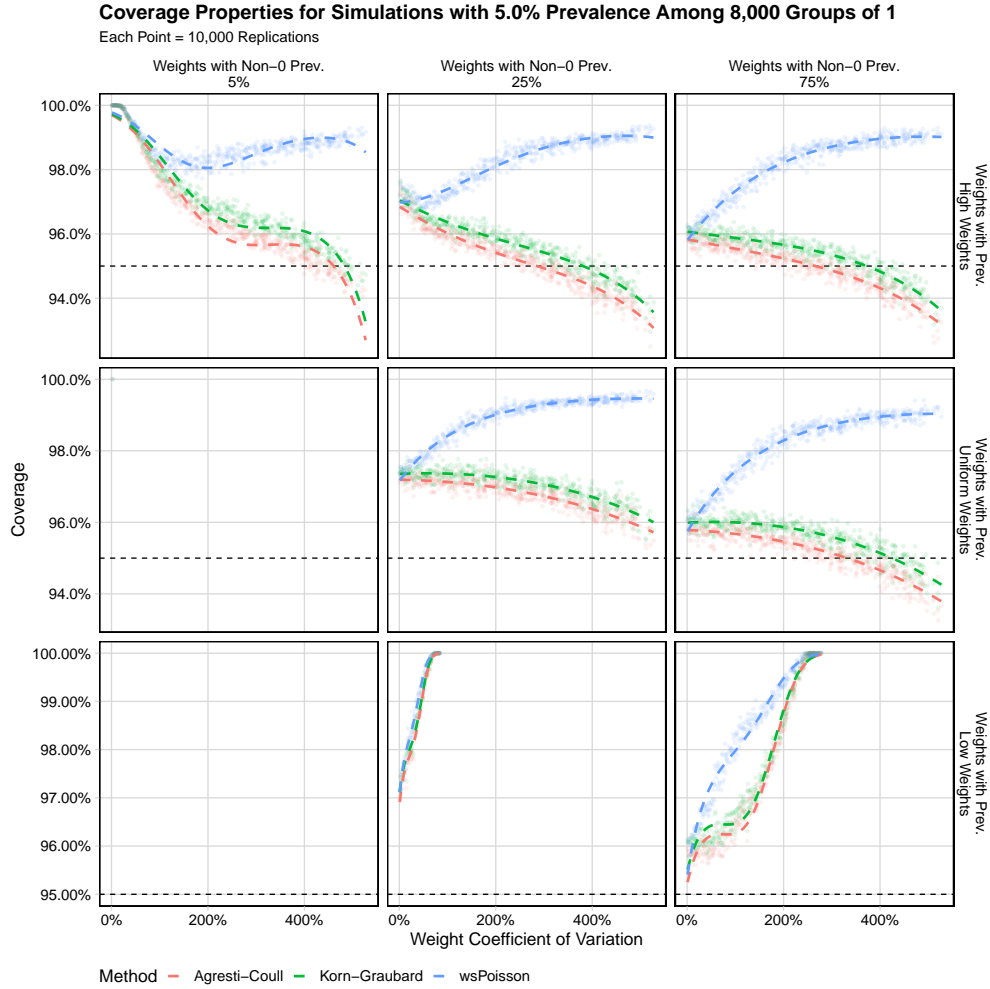
Figure 3.7: Coverage properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of Agresti-Coull and Korn-Graubard. Each point represents 10,000 simulations of datasets from a population with 5% prevalence where 8000 individuals are sampled. The horizontal dashed line indicates the nominal coverage, 95%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

3.3.3 Estimating Prevalence from a Weighted Sample with an Imperfect Assay

We compare properties of our melded confidence interval, WprevSeSp Poisson, to another melded confidence interval method, WprevSeSp Binomial, for cases with weighted samples and imperfect assays. The methods are assessed in several simulated scenarios with varying

levels of disease prevalence (0.5% or 5%), types of groups surveyed (50 groups of 200 subjects each or 8000 groups of 1 subject), distributions of weights among the groups (coefficient of variation from approximately 0% to nearly 600%), the number of groups with non-zero prevalence, and the specificity of the assay (80% - 100%). In each scenario, the assay has 95% sensitivity. Each scenario creates up to 500 new sets of weights and θ_i parameters (as in Section 3.3.2), and each of those is simulated 10,000 times, with new prevalence, sensitivity, and specificity surveys generated and 95% confidence intervals are created. Modeled after the study of Kalish et al. [2021], the simulated sensitivity is assessed based on 60 tests, while specificity is based on 300 tests.

The coverage properties for these simulations are presented in Figures 3.8–3.11. Additional properties for these simulations are presented in Figures A.13–A.24.

Figure 3.8: Coverage properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 50 groups of 200 people are sampled. Each also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal coverage, 95%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure 3.9: Coverage properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence where 8000 individuals are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal coverage, 95%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure 3.10: Coverage properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 50 groups of 200 people are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal coverage, 95%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.
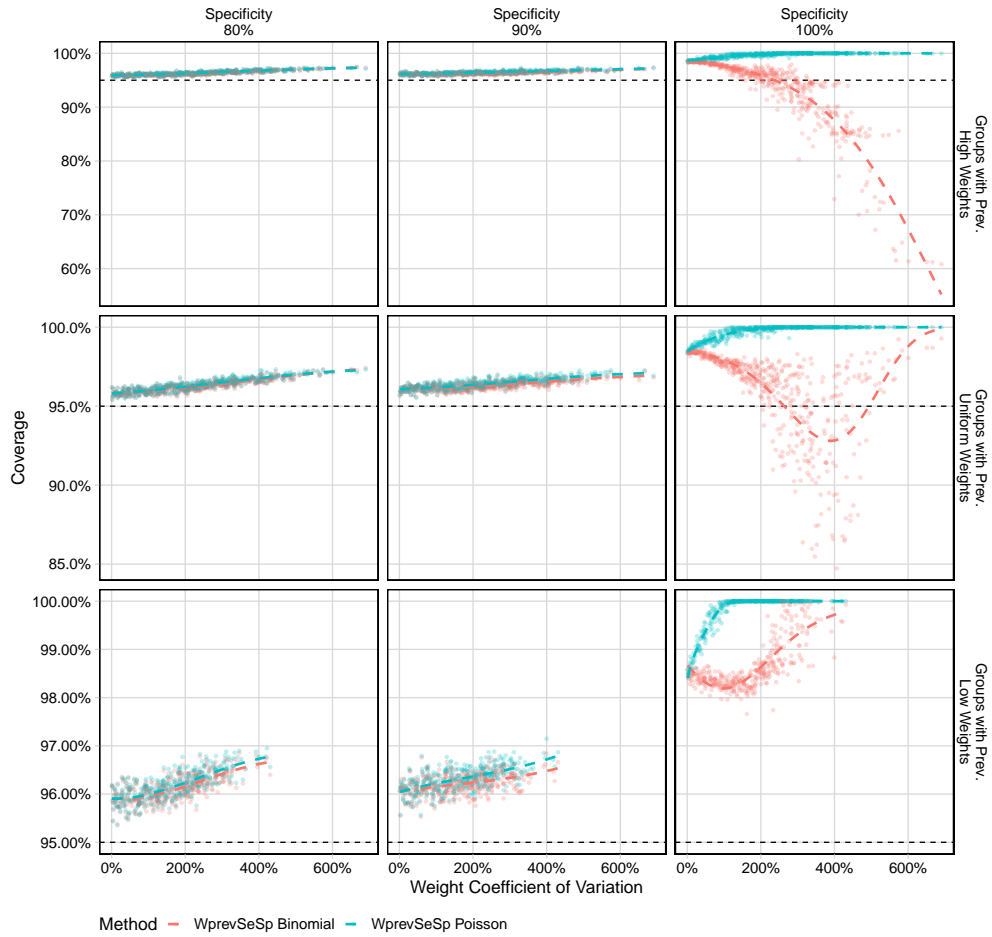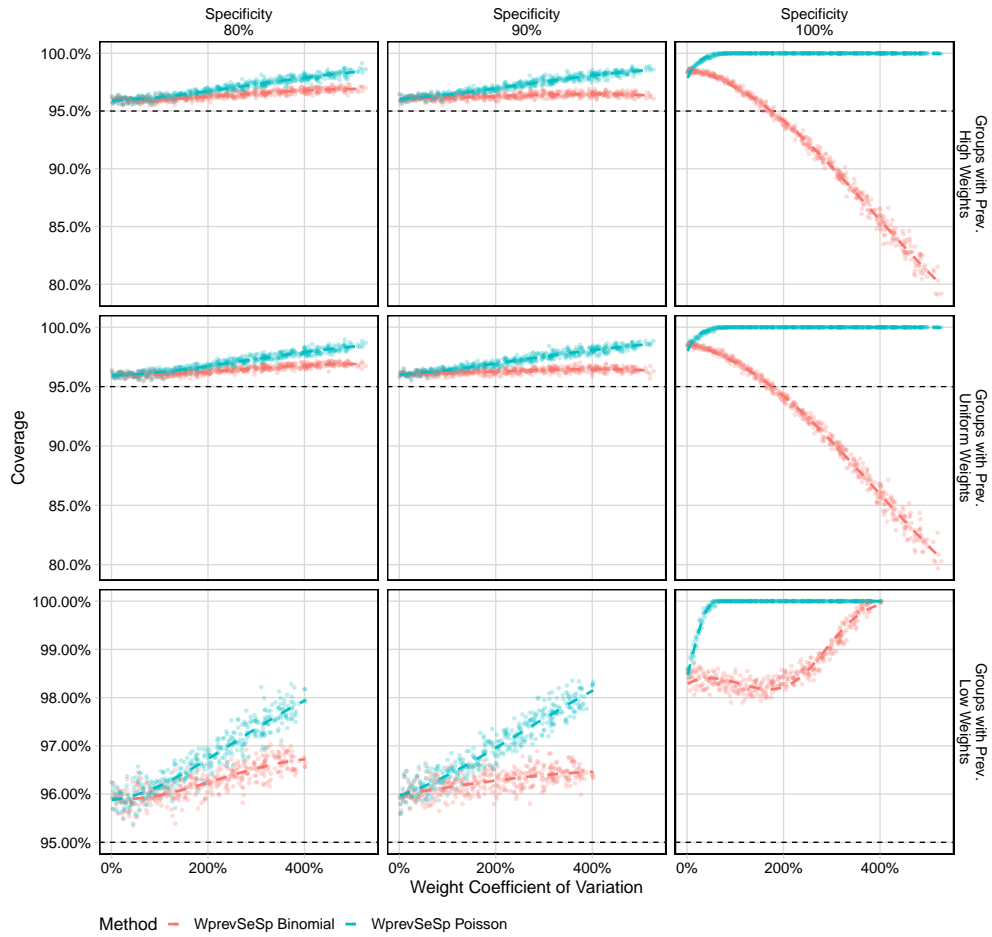
Figure 3.11: Coverage properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 5% prevalence where 8000 individuals are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal coverage, 95%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Based on Figures A.13–A.20, we note that, in most settings, the two melding procedures result in conservative confidence intervals, often nearing 100% coverage. With perfect specificity, the WprevSeSp Binomial method fails to maintain nominal coverage when the coefficient of variation among the weights is high and specificity is perfect. In the other scenarios tested, the WprevSeSp Binomial generally achieves closer to nominal coverage than the

WprevSeSp Poisson, which strongly over-covers. Our proposed WprevSeSp Poisson method maintains or exceeds the desired coverage in all scenarios. In these scenarios, we also assess properties of the wsPoisson procedure, which does not account for the imperfect assay. This method results in approximately 0% coverage in any scenarios where specificity is less than perfect. For this reason, results from that method are omitted in the figures. Because both WprevSeSp methods appear to be very conservative, with coverage near 100% in some cases, we present the widths of the confidence intervals in Figures A.21–A.24. The WprevSeSp Binomial and WprevSeSp Poisson methods typically produce wide intervals of approximately the same size - sometimes as wide as 12%, even when the true prevalence is 0.05%. One notable exception to this is presented in Figure A.23, which shows that for tests with perfect specificity, the WprevSeSP Binomial method produces much narrower confidence intervals than the other method.

3.4 Application

We apply these two methods to a real data set from [Kalish et al., 2021]. This data set was collected to estimate seroprevalence of SARS-CoV-2 in undiagnosed adults in the United States between May and July 2020. The assay used in this data is estimated to have perfect sensitivity, based on 56 tests on individuals with confirmed SARS-CoV-2 and perfect specificity based on 300 tests on individuals confirmed not to have SARS-CoV-2. First, we apply the methods to the full data set ($n = 8058$, weight coefficient of variation $= 252\%$). The seroprevalence in Kalish, *et al* was 4.6% with (95% CI: 2.6% to 6.5%), using a confidence interval method that was nearly the same as the WprevSeSp Binomial method (the method Kalish et al included a calculation of the variability of the weights due to the estimation of the weights, whereas in this paper we treat the weights as fixed constants). The Korn and Graubard type melded confidence interval with imperfect assay adjustments (WprevSeSp Bi-

nomial) studied in this paper produced a 95% confidence interval for population prevalence of (2.53%, 6.68%), very similar to that of Kalish, et al, while the wsPoisson type melded confidence interval with imperfect assay adjustments (WprevSeSp Poisson) produced the 95% confidence interval (2.56%, 7.54%). We also apply the wsPoisson method from Section 3.2.3, which does not account for imperfections in the assay, resulting in a 95% confidence interval of (3.04%, 7.39%). While all three intervals overlap to a large degree, the WprevSeSp Poisson interval is the widest. Our simulations show that in this situation, the WprevSeSp Binomial interval may be the best because, with coefficient of variance about 250% (see Figures B15 and B20, top right panel) the error on both sides of the confidence interval is bounded at 2.5% and the width of the intervals are better (Figure B23).

We also apply the methods to the subset of only Hispanic participants ($n = 1281$, weight coefficient of variation = 306%), where Kalish et al estimated the undiagnosed adult seroprevalence estimate as 6.1% (95% CI: 2.4% to 11.5%). The WprevSeSp Binomial method produces a 95% confidence interval for population prevalence (2.35%, 11.75%), while the WprevSeSp Poisson method produces a 95% confidence interval (2.40%, 20.02%). The wsPoisson method produces a 95% confidence interval of (2.80%, 19.63%). In this case, the two Poisson-based confidence intervals are much wider than the WprevSeSp Binomial interval, which is as expected since the melding method is designed to guarantee coverage, although the simulations show that the WprevSeSp Binomial interval may be reasonable (see e.g., Figure B20). The smaller WprevSeSp Binomial interval is also unsurprising and is similar to the results observed in our simulation study.

3.5 Discussion

We presented several methods for creating confidence intervals to assess disease prevalence in a variety of settings, including simple random samples with imperfect tests, weighted sampling with perfect tests, and weighted sampling with imperfect tests. A main point of this paper was to develop and study methods for the last setting, where there has been little previous work. Two methods were studied for that setting, WprevSeSp Poisson and WprevSeSp Binomial, the latter of which was very similar to the method used by Kalish et al. [2021]. In simulations based on the Kalish et al application, the WprevSeSp Poisson method had greater than nominal coverage in all cases but could be very conservative, while the WprevSeSp Binomial was slightly less conservative but had less than nominal coverage in a few scenarios.

In less complicated scenarios, we could compare the new methods to established methods. These new confidence intervals appear to guarantee coverage in most simulated settings, including some scenarios where competitor methods dramatically under-cover. In general, our methods demonstrate higher coverage than competitor methods in most scenarios, sometimes being very conservative while competitor methods demonstrate closer to nominal coverage. In the case of the simple random sample with an imperfect test, our new methods can bound the lower error rate for a 95% confidence interval at 2.5%, while the Lang and Reiczigel [2014] method maintains 95% coverage by allowing a higher lower error rate.

A big advantage of the new methods is that they may be applied with complex survey methods, where each individual has their own weight, such as in [Kalish et al., 2021]. However, in this paper we have only studied fixed weights and not when the weights are estimated as in Kalish et al. Further, we have not included estimates of the variability of the weights as was done in Kalish, et al; however, recalculating the confidence intervals on the same data shows that ignoring variability of the weights made little difference in that case. Further

58

work is needed to address this issue. In addition, further work is needed to consider other complex sample designs such as multistage cluster designs that are used in household and institutional surveys such as hospital and medical practice surveys.

While the new methods have coverage that is at least nominal in nearly all simulated situations, the cost is that they can be overly conservative. Future work could create less conservative methods by, for example, modifying the new methods by replacing the lower or upper confidence distributions with a mixture of the two (as in [Veronese and Melilli, 2015]), or using a mid-p version of the gamma intervals (as in [Fay and Kim, 2017]).

Although this paper has done extensive simulations, the focus of the simulations was on low prevalence and high sensitivity and specificity. For other situations, especially high prevalence situations, more work is needed to more fully explore the properties of these methods.

Our methods' conservative properties are especially advantageous in settings where the competitor methods exhibit much lower than nominal coverage. For example, when there is high variance among the sample weights and prevalence is concentrated among the highest-weighted samples, competitor coverage of 95% confidence intervals can fall to 60%, while our method exhibits > 95% coverage. Thus, we suggest that our melding methods be employed when working with survey settings which involve high variance among the weights or lower errors are particularly undesirable.

Chapter 4

Semi-parametric modeling of SARS-CoV-2 transmission using tests, cases, deaths, and seroprevalence data

4.1 Introduction

SARS-CoV-2 is a human coronavirus associated with high morbidity and mortality that caused a pandemic in 2020 [Wu and McGoogan, 2020, Song et al., 2020]. Like other human coronaviruses, SARS-CoV-2 is transmitted person to person through close contact and has high transmission potential in crowded indoor settings and around activities that generate aerosols [WHO, 2021]. In the early stages of the COVID-19 pandemic, transmission dynamics modeling played an important role in alerting the public about the potential dangers of unmitigated virus spread [Prem and et. al, 2020, Ferguson and et. al, 2020, Davies et al., 2020]. At later pandemic stages, this modeling helped evaluate intervention effectiveness [Knock et al., 2021] and to quantify transmission advantages of genetic variants [Davies

60

et al., 2021]. We develop models that integrate diagnostics test and mortality time series data with cross-sectional seroprevalence data to estimate underlying transmission dynamics and forecast future case and death counts. These models are flexible in the data sources they incorporate. By comparing the forecasting capabilities of these models, we aim to investigate which data streams should be used in future modeling efforts.

Differences in mitigation strategies, surveillance efforts, and population characteristics across countries and even across different regions within one country prompted development of regional modeling of SARS-CoV-2 transmission [Anderson et al., 2020, Jewell et al., 2021, Morozova et al., 2021, Irons and Raftery, 2021]. However, neither national nor subnational/regional modelers fully integrate all surveillance data available to them, because inclusion of each additional data source leads to an increase in model complexity, which complicates statistical inference and reduces computational efficiency of this inference. In addition, including more data sources necessitates additional modeling assumptions and risks specification in one part of the model influencing inference in other aspects of the model. Modelers were faced with many questions about which data to use, which data to ignore, and how to best integrate them into their models. Incorporating case incidence data into inference proved particularly problematic because a data generating model for cases needs to account for preferential sampling of symptomatic individuals and dependence of case counts on the number of diagnostic tests performed, which significantly varies temporally and spatially. However, even with delayed reporting, positive diagnostic tests (cases) are among the earliest indicators of changing disease dynamics, so we hypothesize that taking advantage of this source of information could be important for producing timely forecasts and for policy decision-making. Similarly, properly incorporating seroprevalence data into models may improve the accuracy of a model's estimations of the underlying cumulative number of infections, which, in part, drives the effective reproduction number and is crucial for forecasting. We investigate these ideas by fitting and comparing the forecasting abilities of multiple models, some of which use these data, while others do not.

We show how to fit a mechanistic model of SARS-CoV-2 spread to incidence and mortality time series, while accounting for the time-varying number of diagnostic tests performed. The mechanistic model is a standard ordinary differential equation (ODE) model that describes changes in the proportions of the population residing in model compartments. Death counts are modeled with a negative binomial distribution that allows for over-dispersion often observed in surveillance data. Our first innovation is the model for cases, where we use a flexible beta-binomial distribution, whose mean is a product of the total number of tests performed and a non-linear function of unobserved infections modeled by the ODE model. This ensures that our estimates are not unduly influenced by large fluctuations of COVID-19 diagnostic test positivity fractions. Our second innovation is nonparametric estimation of time-varying parameters that control both the transmission model and surveillance model. By allowing our model to adapt to temporal changes in transmission and surveillance, we can identify how changes in policy affected the near-term progression of the outbreak. Our third contribution is a careful assessment of the usefulness of various data streams in the context of a case study in Orange County, California, USA.

To benchmark the ability of our model to capture temporal trends in transmission dynamics, we use simulated data to compare our estimation of changes in the effective reproductive number to analogous estimates produced by `epidemia`, a simpler semi-parametric method [Scott et al., 2020]. This comparison shows that failing to account for fluctuations in the number of diagnostic tests performed can result in misleading inferences about effective reproduction numbers, a critical quantity in infectious disease epidemiology. Our data integration approach allows our model to provide a much more detailed picture of the spread of an infectious disease beyond the effective reproduction number, including estimating the time-varying infection fatality ratio, the total number of infected individuals, and changes in testing patterns. Data integration further allows us to produce reasonable short-term forecasts of deaths and testing positivity. We demonstrate these enhanced capabilities by fitting our compartmental model to COVID-19 surveillance data collected in Orange County,

California — the sixth most populous county in the United States of America (U.S.A.), with an estimated 3.2 million inhabitants as of 2019 [United States Census Bureau, 2020]. We analyze Orange County surveillance data collected between March 30, 2020 and February 14, 2021, prior to the start of widespread vaccine availability. We find both basic and effective reproductive numbers varied widely during the first year of the pandemic, which is expected in light of implementation and subsequent relaxation of mitigation measures. We compare several modifications to our primary model which omit negative diagnostic tests or seroprevalence data. We demonstrate that our models produce reasonable short term (up to 4 weeks ahead) probabilistic forecasts of mortality, but that different data streams may be more or less useful during different periods of a pandemic.

4.2   Methods

4.2.1   Data

We start with time series of daily numbers of SARS-CoV-2 diagnostic tests (positive and negative), case counts (positive tests), and deaths observed over some time period of interest. We aggregate the three types of counts in weekly intervals. Figure 4.1 shows such a collection of aggregated time series for Orange County, CA, corresponding to the observation period spanning days between March 30, 2020 and February 14, 2021. We end our modeling period in February because vaccines became more widely available around this time, and our model does not account for vaccine-induced immunity. The data was compiled from anonymized individual test results provided by the Orange County Health Care Agency (OCHCA). We define cases as either confirmed or presumed COVID-19 diagnoses that have been officially reported to the OCHCA. We used specimen collection dates and dates of deaths to tabulate test, case, and death counts. We denote the vector of binned tests by $\mathbf{T} = (T_1, \ldots, T_L)$, the

vector of case counts by $\mathbf{Y} = (Y_1, \ldots, Y_L)$, and the vector of deaths by $\mathbf{M} = (M_1, \ldots, M_L)$, where the weeks are indexed by l and $L = 48$ is the total number of weeks. Additionally, we use data from Bruckner et al. [2021], a study conducted to estimate the seroprevalence in Orange County from a population-representative sample, which consists of 343 seropositive cases (Z_{l^*}) among 2979 (U_{l^*}) tests conducted between July 10 and August 16, 2020. For simplicity, we lump all seroprevalence test dates to a single time point — August 16, 2020 — corresponding to week $l^* = 20$. To formulate the surveillance model for cases $\mathbf{Y}$, deaths $\mathbf{M}$, and seropositive cases Z_{l^*}, we first need a model for latent trajectories of incidence and prevalence of SARS-CoV-2 infections.
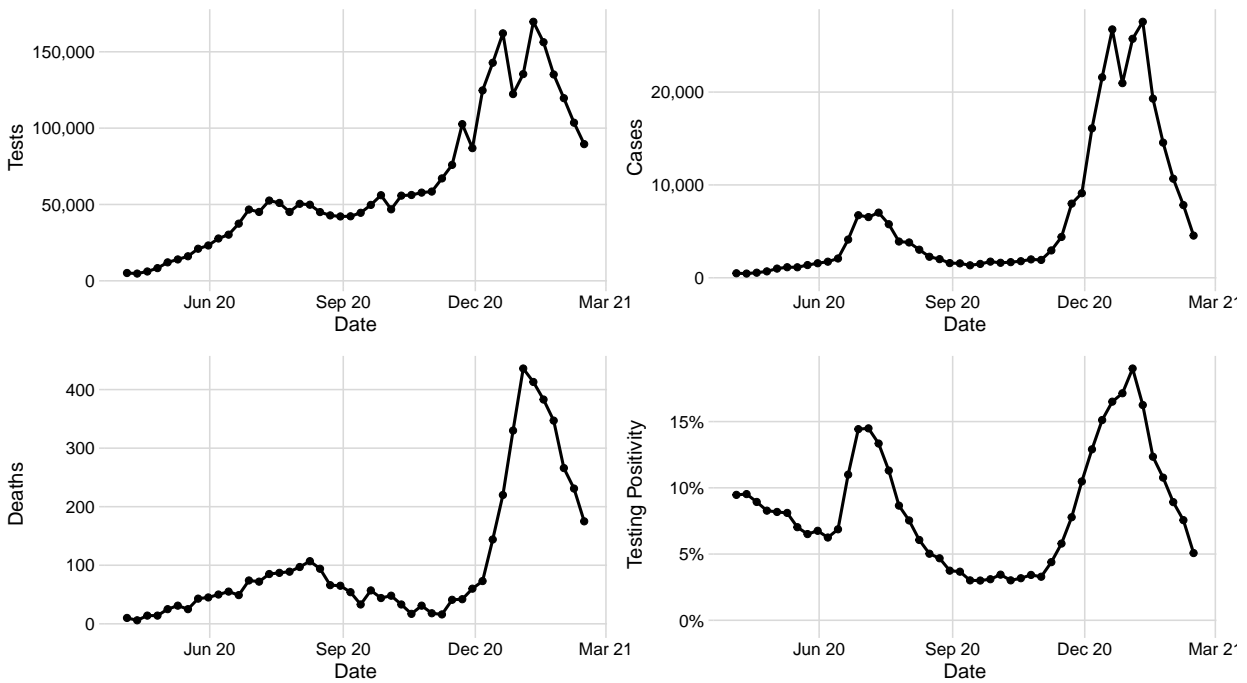

Figure 4.1: COVID-19 surveillance data from Orange County, CA. The figure shows weekly counts of tests, cases (positive tests), reported deaths due to COVID-19, as well as testing positivity.

4.2.2 Transmission model

To model latent incidence and prevalence trajectories, we divide all individuals in the population of Orange County, CA into five compartments: S = susceptible individuals, E = infected, but not yet infectious individuals, I = infectious individuals, R = recovered individuals, D = individuals who died due to COVID-19. Possible progressions of an individual through the above compartments are depicted in Figure 4.2. We model the time-evolution of the proportions of individuals occupying the above compartments with a set of deterministic ordinary differential equations (ODEs). For simplicity, we assume a homogeneously mixing population of fixed size N, although it is possible to relax these assumptions, and we also assume that recovery confers immunity to subsequent infection over the duration of the modeling period. Let $\mathbf{X}(t) = (S(t), E(t), I(t), R(t), D(t))^T$ denote the population proportion in each compartment at time t, and let $\mathbf{X}(t_0) = \mathbf{x}_0$ denote the population proportions at time t_0, the start of the modeling period. By convention, we model the population at risk, i.e., those individuals who may still move throughout the model compartments. Hence, we take $R(t_0) = D(t_0) = 0$ and normalize $\mathbf{X}(t_0)$ so that $\mathbf{X}(t_0)^T \mathbf{1} = 1$, where $\mathbf{1} = (1, 1, 1, 1, 1)^T$. Since we want to fit this model to incidence data, it is convenient to also keep track of the cumulative proportion of the population that experiences transitions between compartments from t_0 to t: $\mathbf{N}(t) = (N_{SE}(t), N_{EI}(t), N_{IR}(t), N_{ID}(t))^T$. To describe mathematically how vectors $\mathbf{X}(t)$ and $\mathbf{N}(t)$ change through time, we first define rates of transitions between compartments, with possible transitions corresponding to the arrows in Figure 4.2:

$$
\begin{aligned}
\lambda_{SE}(S, I, t) &= \frac{\beta(t)}{N} SI, & \lambda_{EI}(E) &= \gamma E, \\
\lambda_{IR}(I, t) &= (1 - \eta(t)) \nu I, & \lambda_{ID}(I, t) &= \eta(t) \nu I,
\end{aligned}
\tag{4.1}
$$

where $\beta(t)$ is the transmission rate, which varies over time, N is the constant population size, $1/\gamma$ is the mean latent period duration, $1/\nu$ is the mean infectious period duration, and $\eta(t)$ is the infection-to-fatality ratio (IFR), which varies over time. The time-varying transmission rate allows our model to capture the effects of interventions and changes in human behavior, while the time-varying IFR should capture changes in age profiles of infected individuals and stress on healthcare providers during surges. We demonstrate this property in Supplementary Section B.3.

Equipped with the population-level transition rates, we define ODEs for our model:

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\lambda_{SE}(S, I, t), & \frac{\mathrm{d}N_{SE}}{\mathrm{d}t} &= \lambda_{SE}(S, I, t), \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \lambda_{SE}(S, I, t) - \lambda_{EI}(E), & \frac{\mathrm{d}N_{EI}}{\mathrm{d}t} &= \lambda_{EI}(E), \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \lambda_{EI}(E) - \lambda_{IR}(I, t) - \lambda_{ID}(I, t), & \frac{\mathrm{d}N_{IR}}{\mathrm{d}t} &= \lambda_{IR}(I, t), \qquad (4.2) \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \lambda_{IR}(I, t), & \frac{\mathrm{d}N_{ID}}{\mathrm{d}t} &= \lambda_{ID}(I, t), \\
\frac{\mathrm{d}D}{\mathrm{d}t} &= \lambda_{ID}(I, t),
\end{aligned}
$$

subject to initial conditions $\mathbf{X}(t_0) = \mathbf{x}_0$ and $\mathbf{N}(t_0) = \mathbf{0}$, where $\mathbf{x}_0 = (S_0, E_0, I_0, R_0, D_0)$ are initial compartment proportions. We set $R_0 = 0$ and $D_0 = 0$, because these proportions do not play a role in future dynamics of the epidemic, leaving S_0, E_0, and I_0 as free model parameters.

The above equations are redundant, and typically only the prevalence ODEs in the left column are used in mathematical modeling. However, the cumulative incidence/transition representation of the model, shown by the ODEs in the right column, is useful for statistical modeling of infectious disease dynamics [Bretó and Ionides, 2011]. In practice, we solve the subset of the above ODEs that are needed to track $\mathbf{X}(t)$ and the parts of $\mathbf{N}(t)$ that "connect" our transmission model to data. We proceed to make this connection in the next subsection.
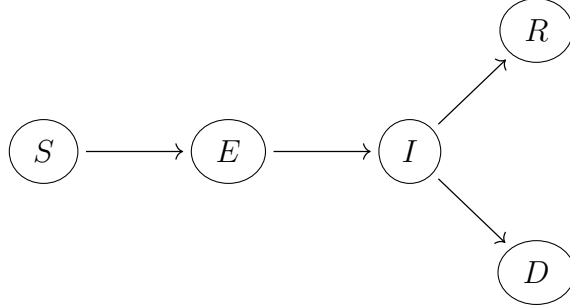
Figure 4.2: Model diagram depicting possible progressions between infection states. The model compartments are as follows: susceptible (S), infected, but not yet infectious (E), infectious (I), recovered (R), and deceased (D).

4.2.3 Surveillance model

We fit our transmission model to seroprevalence data and two time series: numbers of new cases and deaths reported during some pre-specified time periods (e.g., weeks). We do not model changes in the numbers of diagnostic tests performed. Rather, we condition on test counts in the specification of the sampling model for the vector of case counts, which describes the probability of the observed case count given the observed number of tests and unobserved/latent incidence of cases over each time interval. First, we assume that, conditional on $\mathbf{N}(t)$, case and death counts are independent of each other and across time intervals, because they are just noisy realizations of information encoded by $\mathbf{N}(t)$. This leaves us with formulating models for cases and deaths in each individual observation interval.

Consider the number of deaths M_l observed in time interval $(t_{l-1}, t_l]$, where $l = 1, \ldots, L$. Since our ODEs track the latent cumulative fraction of deaths $N_{ID}(t_l)$, we can compute $\Delta N_{ID}(t_l) = N_{ID}(t_l) - N_{ID}(t_{l-1})$ — the latent fraction of the population that died in the interval $(t_{l-1}, t_l]$. We model the observed death count M_l as a realization from the following negative binomial distribution:

$$M_l \sim \text{Negative binomial} \left(\mu_l^D = \rho_D \times N \times \Delta N_{ID}(t_l),\ \sigma_l^{2D} = \mu_l^D(1 + \mu_l^D/\phi_D) \right), \quad (4.3)$$

where N is the population size, μ_l^D and σ_l^{2D} are the mean and variance of the negative binomial distribution, $\rho_D \in [0, 1]$ is the mean overall death detection probability, and $\phi_D > 0$ is an over-dispersion parameter. Informally, our mortality model says that, on average, the observed number of deaths, M_l, is a fraction of the true death count estimated by the model, $N \times \Delta N_{ID}(t_l)$, with some noise due to underreporting, delayed reporting, and sampling variability.

Next, we develop a model for the number of positive tests (cases), Y_l, observed in the time interval $(t_{l-1}, t_l]$. We start with a simple binomial model with per-test positivity probability ψ_l:

$$Y_l \mid \psi_l \sim \text{Binomial}(T_l, \psi_l),$$

where T_l is the number of COVID-19 diagnostic tests administered during the time interval $(t_{l-1}, t_l]$. We use another layer of randomness to account for unobserved factors affecting positivity probabilities (e.g., variable testing guidelines and test shortages) and assume that the positivity probability in interval $(t_{l-1}, t_l]$ follows the beta distribution:

$$\psi_l \sim \text{Beta}\left(\phi_C \mu_l^C, \ \phi_C \left(1 - \mu_l^C\right)\right), \tag{4.4}$$

where ϕ_C is an over-dispersion parameter and μ_l^C is the mean test positivity probability. We assume that mean test positivity odds is proportional to the unobserved odds of transitioning from exposed to infectious, $\Delta N_{EI}(t_l) = N_{EI}(t_l) - N_{EI}(t_{l-1})$ in interval $(t_{l-1}, t_l]$:

$$\frac{\mu_l^C}{1 - \mu_l^C} = e^{\alpha_l} \cdot \left(\frac{\Delta N_{EI}(t_l)}{1 - \Delta N_{EI}(t_l)}\right), \tag{4.5}$$

where $\alpha_l > 0$. This functional form ensures that, on average, the probability of detecting a

SARS-CoV-2 infection grows with the population incidence. Parameter α_l can be thought of as an effect of testing guidelines and practices. A model with $\alpha_l \approx 0$ (i.e., $\mu_l^C \approx \Delta N_{EI}(t_l)$) says that in interval $(t_{l-1}, t_l]$ testing is done approximately by sampling individuals uniformly at random, so that the positivity probability over a time interval l is equal to the fraction of the population that transitions from the latent to infectious state. As we increase α_l above 0, the model mimics preferential testing of individuals who are more likely to have severe infection (e.g., testing only individuals with certain symptoms).

We can streamline our surveillance model for case counts by integrating over positivity probabilities and arriving at the following beta-binomial distribution:

$$Y_l \mid \mu_l^C, \phi_C \sim \text{Beta-binomial}\left(T_l, \ \phi_C \mu_l^C, \ \phi_C\left(1 - \mu_l^C\right)\right). \tag{4.6}$$

Properties of the beta-binomial distribution imply that $\mathrm{E}(Y_l) = T_l \times \mu_l^C$. This means that our model predicts that, on average, cases grow linearly with the number of diagnostic tests administered. Keeping in mind our assumed relationship between μ_l^C and $\Delta N_{EI}(t_l)$, the average number of cases also grows with the accumulation of new infections. Furthermore, the variance of the fraction of tests that are positive under the beta-binomial distribution is

$$\mathrm{Var}(Y_l/T_l \mid T_l, \mu_l^C, \phi_C) = \frac{\mu_l^C(1 - \mu_l^C)}{T_l}\left(1 + \frac{T_l - 1}{\phi_C + 1}\right),$$

where the variance under an analogous pure binomial model would be $\mu_l^C(1 - \mu_l^C)/T_l$. Hence, the over-dispersion parameter, ϕ_C, can be interpreted in terms of the excess variance of the beta-binomial model relative to a pure binomial distribution. In summary, our beta-binomial distribution for observed case counts ensures that we do not confuse increase in testing for increase in SARS-CoV-2 incidence and implicitly allows for heterogeneity in the mean test positivity probability.

Finally, we model the number of observed seropositive cases Z_{l^*} among U_{l^*} tests with a

binomial distribution:

$$Z_{l^*} \sim \text{Binomial}\left(U_{l^*}, \frac{R_{l^*}}{S_{l^*} + E_{l^*} + I_{l^*} + R_{l^*}}\right).$$ (4.7)

This simple model assumes the seroprevalence data comes from a high-quality study based on random sampling, which does not exhibit the problems observed in the testing data.

In addition, we also consider a more typical approach to modeling observed cases, which is not conditional on tests and is similar to (4.3).

$$Y_l \sim \text{Negative binomial}\left(\mu_l^Y = \rho_l^Y \times N \times \Delta N_{EI}(t_l),\ \sigma_l^{2^Y} = \mu_l^Y(1 + \mu_l^Y/\phi_Y)\right),$$ (4.8)

where N is the population size, μ_l^Y and $\sigma_l^{2^Y}$ are the mean and variance of the negative binomial distribution, $\rho_l^Y \in [0,1]$ is the mean case detection probability, which varies over time, and $\phi_Y > 0$ is an over-dispersion parameter.

4.2.4   Putting all the pieces into a Bayesian model

We now describe our inferential Bayesian procedure. First, we re-parameterize our model by replacing $\beta(t)$ with a basic reproductive number $R_0(t) = \beta(t)/\nu$. We parameterize each of our time-varying parameters, $R_0(t)$, $\eta(t)$, $\alpha(t)$, $\rho_l^Y(t)$ as piecewise constant functions, where each vector defining the constants *a priori* follows a Gaussian Markov random field (GMRF).

More precisely, we define the auxiliary vectors:

$$\tilde{\boldsymbol{R_0}} = (\tilde{R}_{0,1}, \tilde{R}_{0,2}, \ldots, \tilde{R}_{0,L}),\ \tilde{\boldsymbol{\eta}} = (\tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_L),$$

$$\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_L),\ \text{and}\ \tilde{\boldsymbol{\rho}}^{\boldsymbol{Y}} = (\tilde{\boldsymbol{\rho}}^{\boldsymbol{Y}}{}_1, \tilde{\boldsymbol{\rho}}^{\boldsymbol{Y}}{}_2, \ldots, \tilde{\boldsymbol{\rho}}^{\boldsymbol{Y}}{}_L),$$

which follow the Gaussian Markov random field priors:

$$
\begin{aligned}
\tilde{R}_{0,l} &\sim N\left(\tilde{R}_{0,l-1}, \sigma_{R_0}^2\right), \text{ where } l = 2, \ldots, L \text{ and } \tilde{R}_{0,1} \sim N\left(\mu_{R_{01}}, \sigma_{R_{01}}^2\right), \\
\tilde{\eta}_l &\sim N\left(\tilde{\eta}_{l-1}, \sigma_\eta^2\right), \text{ where } l = 2, \ldots, L \text{ and } \tilde{\eta}_1 \sim N\left(\mu_{\eta_1}, \sigma_{\eta_1}^2\right), \\
\tilde{\alpha}_l &\sim N\left(\tilde{\alpha}_{l-1}, \sigma_\alpha^2\right), \text{ where } l = 2, \ldots, L \text{ and } \tilde{\alpha}_1 \sim N\left(\mu_{\alpha_1}, \sigma_{\alpha_1}^2\right), \\
\tilde{\rho}_l^Y &\sim N\left(\tilde{\rho}_{l-1}^Y, \sigma_{\rho^Y}^2\right), \text{ where } l = 2, \ldots, L \text{ and } \tilde{\rho}_1^Y \sim N\left(\mu_{\rho_1^Y}, \sigma_{\rho_1^Y}^2\right),
\end{aligned}
\tag{4.9}
$$

and define the piecewise constant functions:

$$R_0(t) = \sum_{l=1}^L \exp\left\{\left(\tilde{R}_{0,l}\right) \mathbb{1}\left[t \in (t_{l-1}, t_l]\right]\right\}, \tag{4.10}$$

$$\eta(t) = \sum_{l=1}^L \frac{\exp\{(\tilde{\eta}_l) \mathbb{1}\left[t \in (t_{l-1}, t_l]\right]\}}{\exp\{(\tilde{\eta}_l) \mathbb{1}\left[t \in (t_{l-1}, t_l]\right]\} + 1}, \tag{4.11}$$

$$\alpha(t) = \sum_{l=1}^L \exp\{(\tilde{\alpha}_l)\ \mathbb{1}\left[t \in (t_{l-1}, t_l]\right]\}, \tag{4.12}$$

$$\rho^Y(t) = \sum_{l=1}^L \frac{\exp\{\left(\tilde{\rho}_l^Y\right) \mathbb{1}\left[t \in (t_{l-1}, t_l]\right]\}}{\exp\{(\tilde{\rho}_l^Y) \mathbb{1}\left[t \in (t_{l-1}, t_l]\right]\} + 1}. \tag{4.13}$$

In addition, we parameterize initial compartment fractions as S_0, $I_0 = \widetilde{I}_0(1 - S_0)$, $E_0 = (1 - \widetilde{I}_0)(1 - S_0)$, $R_0 = 0$, $D_0 = 0$, where $\widetilde{I}_0 = I_0/(E_0 + I_0)$. This construction allows us to specify independent prior distributions for S_0 and $\widetilde{I}_0$ while preserving the sum-to-one constraint on the original initial compartmental fractions. Next, we collect all our model

parameters into a vector $\boldsymbol{\theta} = (S_0, \tilde{I}_0, \tilde{\boldsymbol{R_0}}, \gamma, \nu, \tilde{\boldsymbol{\eta}}, \rho^D, \phi_D, \tilde{\boldsymbol{\alpha}}, \phi_C, \sigma_{R_0}, \sigma_\eta, \sigma_\alpha)$. When using the traditional case-emission model (4.8), $\tilde{\boldsymbol{\rho}}^{\boldsymbol{Y}}$, σ_{ρ^Y}, and ϕ_Y are substituted for $\tilde{\boldsymbol{\alpha}}$, ϕ_C, and σ_α. Our probabilistic construction described above implies that the likelihood function — probability of observing incidence, mortality, and seroprevalence data — can be written in the following way:

$$\Pr(\mathbf{M}, \mathbf{Y}, Z_{l^*} \mid \boldsymbol{\theta}) = \Pr(\mathbf{M} \mid \boldsymbol{\theta})\Pr(\mathbf{Y} \mid \boldsymbol{\theta})\Pr(Z_{l^*} \mid \boldsymbol{\theta}) = \prod_{l=1}^{L} \Pr(M_l \mid \boldsymbol{\theta}) \Pr(Y_l \mid \boldsymbol{\theta}) \Pr(Z_{l^*} \mid \boldsymbol{\theta}),$$

where $\Pr(M_l \mid \boldsymbol{\theta})$, $\Pr(Y_l \mid \boldsymbol{\theta})$, and $\Pr(Z_{l^*} \mid \boldsymbol{\theta})$ are the probability mass functions given by (4.3), (4.6) or (4.8), and (4.7) respectively.

We encode available information about our model parameters in a prior distribution with density $\pi(\boldsymbol{\theta})$. We assume that all univariate non-GMRF distributed parameters are *a priori* independent and list our prior assumptions in Table B.1. Since our model is highly parametric, we rely on informative prior distributions that we parameterize using existing scientific studies. We base all our inferences and predictions on the posterior distribution of all model parameters:

$$\pi(\boldsymbol{\theta} \mid \mathbf{M}, \mathbf{Y}, \mathbf{Z}) \propto \Pr(\mathbf{M}, \mathbf{Y}, \mathbf{Z} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{4.14}$$

We sample from this posterior using the No-U-Turn Sampler [Hoffman and Gelman, 2014] as implemented in the Turing Julia package [Ge et al., 2018]. Model code and data are available at the following GitHub repository: `https://github.com/damonbayer/semi_parametric_` `COVID_19_OC_model`.

4.3 Results

4.3.1 Simulation Study

To validate our model, we simulated 200 datasets with parameters given in Supplementary Table B.2. An example of one of these datasets is presented in Figure B.1. The number of tests at each time point is the same as in the Orange County data set, and the parameters were deliberately chosen to produce data similar to the Orange County data. Priors used for these model fits are the same as those used in the Orange County model (see Table B.1). The prior and posterior distribution of the model fit to the single simulated dataset from Figure B.1 are presented in Supplementary Figures B.2–B.3. For each simulated dataset, we used four Markov chains run in parallel to draw a total of 1000 posterior samples. In this single dataset example, most of the scalar parameter posteriors shift slightly toward the true parameters compared to the priors, without much posterior variance contraction relative to the prior. We define posterior variance contraction as one minus the ratio of standard deviation of the posterior and the prior, where negative contraction indicates that the posterior is wider than the prior, and 100% contraction indicates that the posterior is a degenerate distribution. For the time-varying parameters and compartments, the variance contraction and shift are much more apparent. We further explore our simulation study results with summary measurements, presented in Supplementary Figures B.5–B.9. Figure B.5 shows the coverage of the posterior 80% credible intervals constructed for the scalar parameters in the 200 simulated datasets. Most parameters demonstrate nearly 100% coverage, except for ϕ_C, which shows approximately 90% coverage, which is still above the nominal 80%. Similarly, Figure B.7 displays coverage of the posterior 80% credible intervals constructed for the time-varying parameters, with only one time point for α demonstrating less than nominal coverage. We observe slightly less conservative coverage when examining the latent compartments in Figure B.9, with the D compartment falling to around 60% coverage at

some points. In addition to coverage, we also consider posterior contraction. Figure B.6 shows contraction of the scalar parameters, with most parameters demonstrating nearly no contraction. Notable exceptions to this are σ_{R_0}, and σ_α, which exhibit positive contraction. Figure B.8 shows contraction of the time-varying parameters, with all parameters exhibiting a high amount of positive contraction. Similarly, the latent compartments also demonstrate a large degree of positive contraction in Figure B.10.

We used the `epidemia` R package [Scott et al., 2020] to fit a state-of-the-art method for effective reproduction number estimation (R_t), to the same 200 simulated datasets. This semi-mechanistic method does not attempt to estimate the unobserved number of susceptible and recovered individuals and does not account for changes in testing volume. See Supplementary Section B.2 for a brief description of the `epidemia` statistical model. Metrics based on estimates of R_t from the true model and `epidemia` for the simulation study are presented in Figure B.11. We assess the envelope, which is the proportion of time points which the 80% posterior credible interval contains the true R_t value specified in the simulation. Mean credible interval width (MCIW) is calculated as the mean of credible interval widths across time points within a simulation replication. Absolute deviation is a measure of bias, and is calculated as the mean of the absolute difference between the posterior median and the true R_t value at each time point. The mean absolute sequential variation (MASV) is calculated as the mean of the absolute difference between the posterior median at a time point and the posterior median at the previous time point. In all metrics, the true model fit achieves the superior result. The envelope for `epidemia` is typically around 50%, while it is near 100% for the true model. The mean credible interval width for `epidemia` (around 0.50) is larger than for the true model (around 0.35). The `epidemia` results also indicate more bias compared to the true model, as measured by the absolute deviation (0.20 for epidemia vs 0.05 for the true model). Additionally, the true model has a mean absolute sequential variation close to that of the simulated parameters (around 0.085), while the MASV reported by `epidemia` is larger (around 0.10).

4.3.2 Application to Orange County, California Data

Next, we apply our Bayesian inferential procedure to COVID-19 surveillance data collected in Orange County, California between March 30, 2020 and January 17, 2021. We again used four Markov chains run in parallel to draw a total of 1000 posterior samples. By the end of the modeling period, approximately 4% of Orange County residents were at least partially vaccinated. Because our model does not incorporate vaccination directly, it doesn't make sense to use our model beyond January 2021. Throughout the modeling period, a variety of non-pharmaceutical interventions were enacted and sometimes lifted at the state, county, and city level. Notably, in-person school closures, indoor dining bans, and mask mandates were in effect for most or all of this period. Fitting the model took approximately 7.5 hours to generate 1,000 posterior samples from 4 chains, totaling around 29 CPU hours. These run times show that our model can be fit frequently enough to be useful for a real-time policy response. This number of chains and posterior samples resulted in a satisfactory effective sample size for posterior inference. Convergence and mixing were assessed using potential scale reduction factors, effective posterior sample sizes, and traceplots of model parameters, which are presented in Appendix B.5. Our main interest is in understanding differences in transmission dynamics and surveillance efforts throughout this period. Since our model is highly parametric, we used existing knowledge of SARS-CoV-2 transmission dynamics to formulate informative priors for all model parameters that we list in Table B.1. We briefly highlight some of our assumptions. Our priors for the initial compartment sizes reflect our belief that the number of infections was small, but potentially underreported by a factor of 10, at the beginning of the pandemic. Since our observation period starts close to the stay-at-home order taking effect in California, we assume that the March 2020 basic reproduction number should be around 1.0 to reflect reduced contacts during this period. Lengths of latent and infectious periods *a priori* assumed to be 0.8 and 1.2 weeks respectively, with substantial variance. Based on Orange County, CA seroprevalence study, we assume initial

infection-to-fatality ratio to be around 0.4% [Bruckner et al., 2021]. We compare reported deaths in Orange County, CA with estimates of U.S. county-specific excess mortality to set the prior for death reporting probability to be around 0.9 [Stokes et al., 2021]. Prior and posterior distributional summaries of all model parameters are available in Supplementary Figures B.17 and B.18. These figures also contain the results of our sensitivity analysis, where we examined the effects of our prior assumptions on our inference. Our main conclusion is our results are not sensitive to reasonable prior perturbations.

The upper-left plot of Figure 4.3 presents the posterior distribution of the basic reproductive number (R_0) for Orange County. Throughout the late spring and summer, the basic reproductive number is estimated to be slightly above 1.0, with some probability of being below 1.0 in the early fall. Beginning in October, the basic reproductive number begins to rise and surpasses 2.0 at the peak of the winter wave. This rise in the fall may be associated with the school re-openings that occurred around this time. Despite the high basic reproductive number throughout the modeling period, the upper-right plot of Figure 4.3 shows that the effective reproductive fell below 1.0 for much of the summer and again in January, following the winter surge, allowing us to separate the effects of reducing the average community contact rate and accumulated infection-induced immunity. We also apply the `epidemia` method to the Orange County data and plot the results in Figure 4.4. From this, we observe that the two methods lead to similar conclusions about the posterior distribution of R_t. At all but one of the time points, the 80% credible intervals of the posteriors from both methods overlap with one another, but the full model appears to generally produce smaller credible intervals, especially near the beginning of the fitting period. Additionally, the full model produces a smoother posterior than the `epidemia` model.

We proceed with describing inference results for the other two time-varying parameters in our model: infection-to-fatality ratio η and the parameter α that governs the relationship between testing positivity and the true proportion of newly infected individuals in the popu-
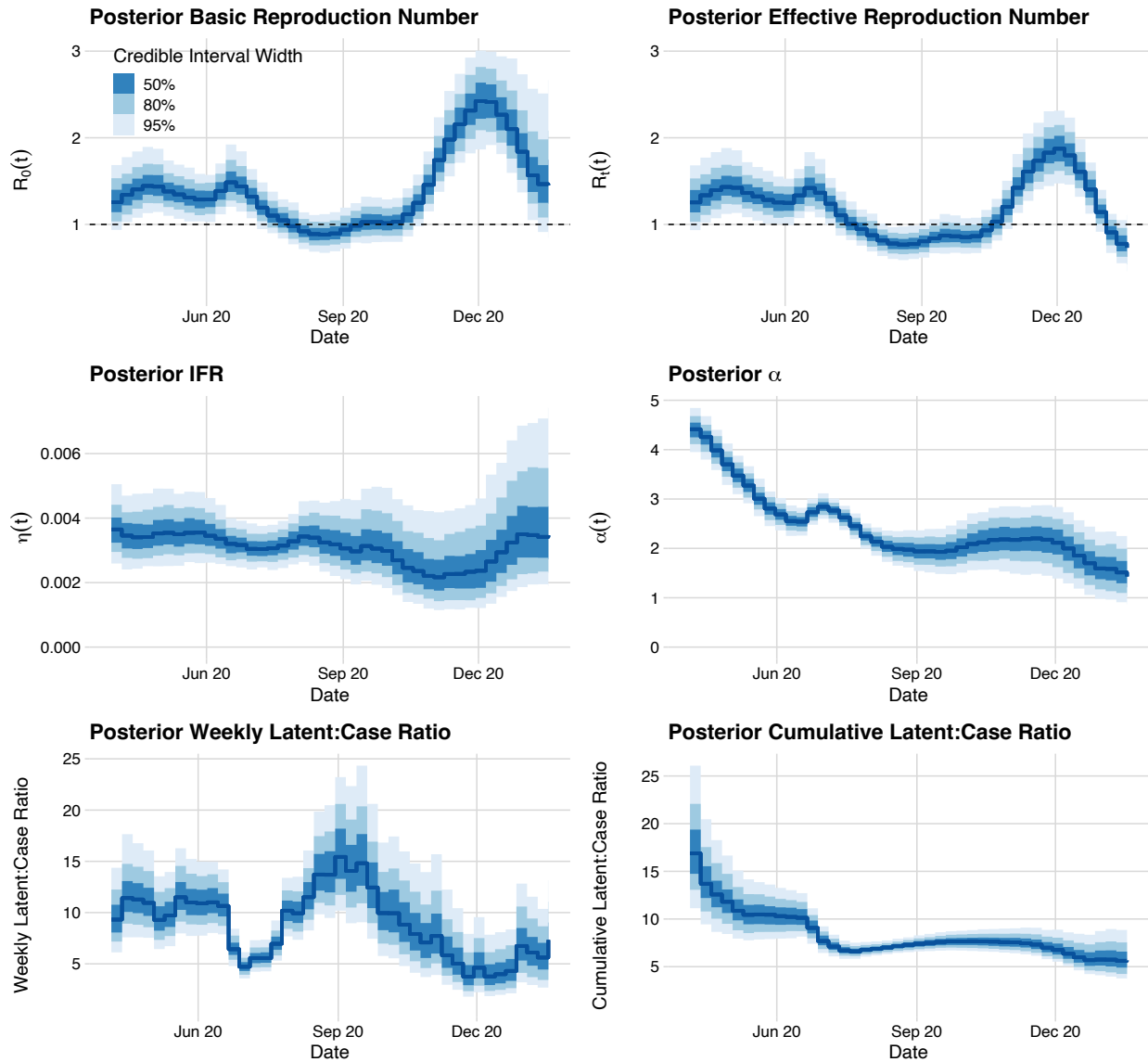
Figure 4.3: Posterior distributions of the time-varying basic reproductive number R_0, effective reproductive number R_e, infection-to-fatality ratio (IFR), proportion in the proportional log-odds model of the beta-binomial observational model for cases α, weekly latent:case ratio, and cumulative latent:case ratio. Solid blue lines show point-wise posterior medians, while shaded areas denote 50%, 80%, and 95% Bayesian credible intervals.
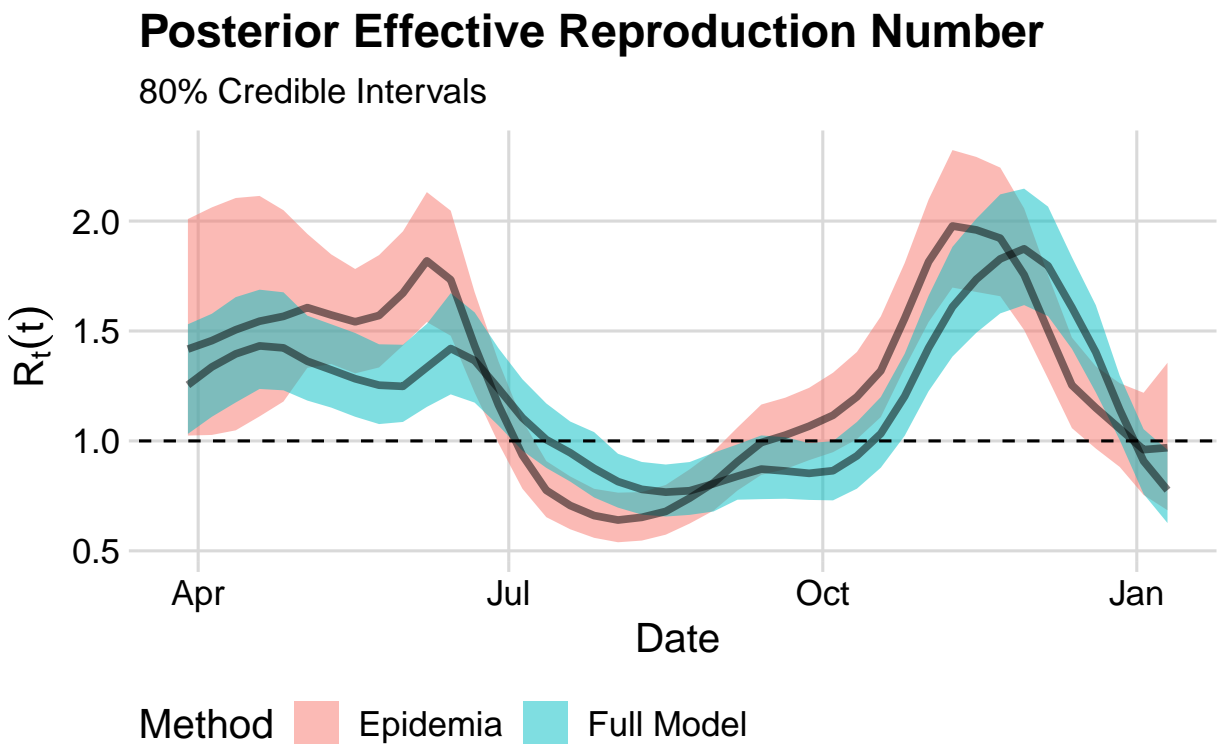
Figure 4.4: Posterior inference for the effective reproduction number from the full model and `epidemia` fit to the Orange County data.

lation. The posterior distribution of the infection-to-fatality ratio is presented in the middle-left plot of Figure 4.3. The IFR is estimated to be consistent over time, hovering around 0.3%, but our estimates are less certain near the end of the modeling period. This potential rise in IFR could have been caused by a combination of the overwhelmed healthcare system and the increasing prevalence of the Alpha variant at this time, which has been tied to more severe outcomes [Grint et al., 2021]. The middle-right plot and bottom plots of Figure 4.3 present three perspectives on testing policy and case detection: the posterior α, the weekly latent:case ratio, and the cumulative latent:case ratio. Generally, α, drifts lower over time, indicating that testing policy became less preferential toward selecting infected individuals as testing became more accessible. This trend is reversed slightly during the summer and winter waves, which is reflected in the decreasing weekly latent:case ratio during these times. The cumulative latent:case ratio also drifts lower over time, eventually arriving at a final cumulative latent:case ratio of $4:1 - 9:1$.
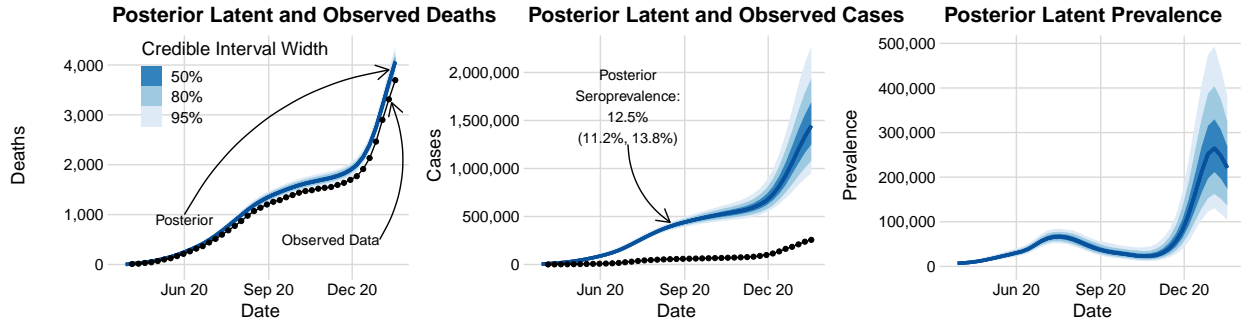

Figure 4.5: Latent and observed cumulative death (left) and incidence (center) trajectories and latent prevalence trajectories (right) in Orange County, CA (population 3.2 million). Solid blue lines show point-wise posterior medians, while shaded areas denote 50%, 80%, and 95% Bayesian credible intervals. Black circles denote observed data. Note that the posterior predictive distributions are of latent deaths and cases are not forecasts of their observed counterparts. Forecasts are plotted in Figure 4.6.

We plot posterior medians and Bayesian credible intervals of the latent cumulative death counts ($N_{ID}(t)$) between March 2020 and January 2021, using three credibility levels shown in the left plot of Figure 4.5. Reported death counts are shown as black circles in the same plot. The plot reflects an overall death reporting rate of 87% - 94%. The center plot of

Figure 4.5 shows the posterior distributions of the cumulative number of infections $(N_{SE}(t))$ occurred in Orange County, with the cumulative observed cases displayed as black circles. We estimate that 32–72% of Orange County residents experienced SARS-CoV-2 infection by mid-January 2021. As in Figure 4.3, this shows a cumulative latent:case ratio of 4:1 – 9:1, with 1/3 – 2/3 of all Orange County residents having been infected by the end of January 2021. From this plot, we also note that our posterior estimate of seroprevalence in mid-August of 2020 (11.2%-13.7%) closely matches the 11.5% estimate from [Bruckner et al., 2021]. We explicitly used this seroprevalence data in our inference, so this is unsurprising. The right plot of Figure 4.5 shows the prevalence of SARS-CoV-2 infected individuals at a particular time, $(E(t) + I(t))$. At the peak of the winter wave, we estimate that 7.8% – 14.9% of Orange County residents had an active infection at the same time.
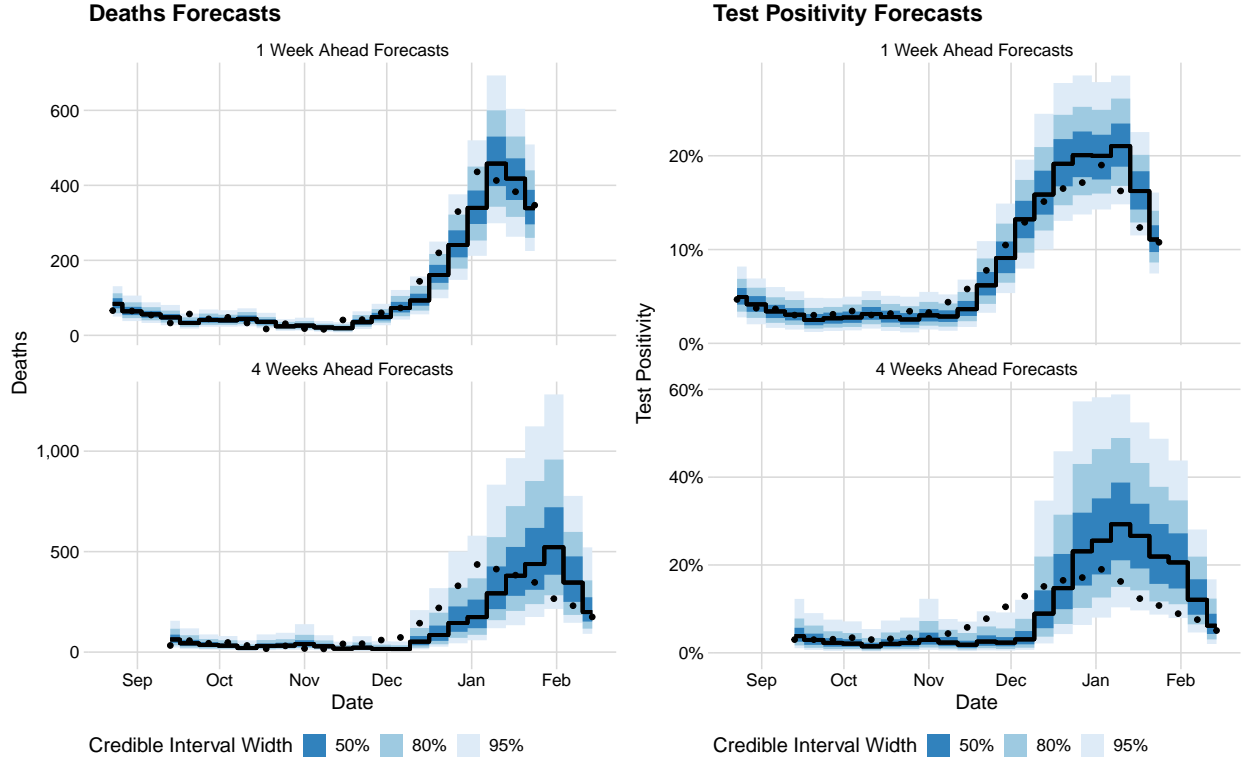

Figure 4.6: Forecast distributions for observed deaths (left column) and testing positivity (right column). Solid blue lines show point-wise posterior medians, while shaded areas denote 50%, 80%, and 95% Bayesian credible intervals. Observed values are presented as black circles.

Finally, we turn to model-based forecasting of observable quantities. In Figure 4.6, we present one week and four week ahead forecasts of observed deaths and test positivity. The credible intervals shown for a given date are generated from a model that is fit to data from March 30, 2020 up to 1 week or 4 weeks prior to the given date. The forecasts are produced by augmenting the posterior time-varying parameters by carrying forward the previous mean values in (4.9) and solving the ODEs from (4.2) into the future. Since this model makes use of the seroprevalence data, we only produce forecasts for times after this data is available, beginning in late August 2020. Because forecasting cases with this model is impossible without knowing how many tests will be conducted in the future, we focus on the positivity fraction (cases divided by the total number of tests) instead. As in the other figures, we use three credibility levels in Figure 4.6, and observed values are displayed as black circles. Our one-week ahead probabilistic forecasts for both observed deaths in the upper half of Figure 4.6 generally capture the observed values, indicating our method to be precise and well calibrated. The four-week ahead forecasts predict the data well in times of relative stability, but exhibit poor performance when time-varying parameters are changing rapidly, such as during the winter wave. In this case, when forecasting four weeks out, we tend to underestimate the rise at the beginning of the wave and overestimate the fall at the end of the wave. This is not of major concern because the four-week time horizon is long enough that interventions and behavioral changes may take effect that are not foreseen by the model. Scenario-based modeling, where some values are specified for the future time-varying parameters, rather than simulating from the prior, may be more appropriate for this task.

We now compare our forecasting results to three variants of our model, which make use of different data streams. Each model can either be conditioned on tests or not conditioned on tests. When conditioning on tests, we use the case emission model given by (4.6). When not conditioning on tests, we use the case emission model given by (4.8). Each model can also make use of the seroprevalence data or not. When using the seroprevalence data,

we use the emission distribution given by (4.7). When not using the seroprevalence data, no emission distribution is used. The model results discussed above are for the model which is conditioned on tests and uses the seroprevalence data. We compare these models by calculating the Continuous Ranked Probability Score (CRPS) [Matheson and Winkler, 1976], as implemented in the `scoringRules` package [Jordan et al., 2019], facilitated by the `scoringutils` package [Bosse et al., 2022]. We present comparisons of these scores in Figure 4.7. We only show scores based on deaths because comparisons based on cases would require developing a method to forecast future test counts, which we have not considered in this work.
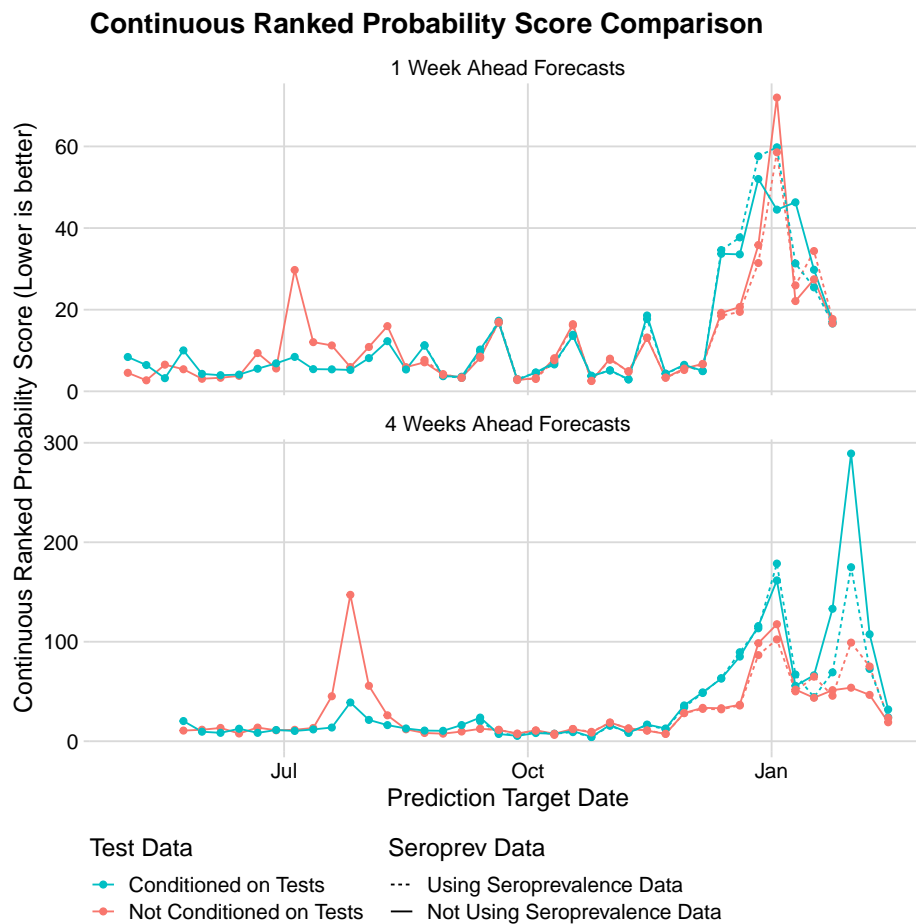
Figure 4.7: Comparison of Continuous Rank Probability Score for models fit to the Orange County data. Lower is better.

From Figure 4.7, we observe that all models appear to perform similarly throughout much

of the assessed time period. However, in both one week and four week ahead forecasts, the models which are conditioned on tests tend to score slightly better in the late summer period, when testing policy was rapidly changing. During the winter surge, the differences in the model forecasting abilities are more pronounced, with the models not conditioned on tests appearing to be consistently superior to those which are conditioned on tests. There is no clear pattern differentiating the models which use the seroprevalence data from those that do not.

4.4 Discussion

We developed a Bayesian SARS-CoV-2 transmission model that integrates information from incidence, mortality, and seroprevalence data. Our approach combines an ODE-based SEIR compartmental model of SARS-CoV-2 transmission dynamics and a carefully constructed surveillance model for cases, deaths, and seroprevalence. Importantly, our method accounts for variability in the number of SARS-CoV-2 diagnostics tests across time, thus ensuring that we do not confuse increases in testing with increases in incidence. Another distinguishing feature of our approach is nonparametric modeling of changes in key transmission and surveillance model parameters. Since we are integrating multiple sources of information, we can afford to be fairly ambitious and to include three such parameters into our model. Changes in one of these parameters accounts for changing the strength of preferentially testing SARS-CoV-2 infected individuals, which helps us avoid an important source of potential bias when inferring transmission model parameters. We reconstruct latent dynamics of the two pre-delta COVID-19 waves in Orange County, CA and estimate that 32–72% of Orange County residents experienced SARS-CoV-2 infection by mid-January 2021. Retrospective analysis shows that our model produces accurate and well calibrated one week ahead and reasonable four week ahead forecasts, but the latter lack accuracy during periods when SARS-CoV-2

transmission dynamics and mitigation policies change rapidly. Additionally, we evaluated our forecasting performance when including or excluding certain data streams (test counts and seroprevalence study data) and found that incorporating negative test counts into our model was useful near the beginning of the modeling period, when testing policy was changing rapidly, but less useful when the policy became consistent. We also found that excluding seroprevalence data did not negatively affect our forecasting ability.

Our primary focus in this work was on developing a framework for integrating multiple data streams into a transmission model. However, there are a number of extensions we could pursue to improve the realism of the assumed transmission dynamics and strengthen the model's forecasting skill. Our model assumes that the population of interest is well mixed and that all individuals in the population infect others and get infected at the same per capita rate. In fact, the actual SARS-CoV-2 transmission process is much more complex because individuals come into contact with each other based on their geographical and social network proximity. Furthermore, it is well established that COVID-19 disease progression process depends on the individual's age and other characteristics [Kim et al., 2021, Bhargava et al., 2020, Petrilli et al., 2020]. Similarly, all transmission model parameters may depend on the vaccination status of an individual. Fortunately, compartmental models can be extended to account for these complexities. For example, we can stratify each model compartment by age, vaccination status, and geographical location, as is commonly done in epidemiological modeling [Li and Brauer, 2008, Van den Driessche, 2008].

We have addressed changes in control/mitigation measures and in human behavior by non-parametrically modeling variability of some of the SARS-CoV-2 transmission model parameters across time. Anderson et al. [2020] and Jewell et al. [2021] use parametric approaches to model the effects of mitigation measures on R_0. It would be interesting to try a semi-parametric approach that combines parametric and non-parametric components, which would allow us to include indicators of human behavior (e.g., mobility data as in Jewell et al.

[2021]) into our inference and forecasting.

In this paper, we have sidestepped the thorny issue of reporting delays by restricting our analyses to time periods in which the data have stabilized. Hence, our analyses should be robust to reporting delays so long as we have either "run out the clock" on the extent of the delays or reporting delays do not differ between positive and negative COVID-19 diagnostic tests. A useful set of extensions that would make our model more useful for real-time surveillance involve estimating the reporting delay distribution [Höhle and an der Heiden, 2014, Stoner and Economou, 2020] and using this distribution in our surveillance model.

Finally, we would like to point out that our deterministic representation of the latent epidemic process could be substituted for a fully stochastic model where the latent epidemic is represented as a Markov jump process, albeit with some loss of computational efficiency. In our large population setting, this could be achieved via simulation-based methods [Bretó et al., 2009, Andrieu et al., 2010, Dukic et al., 2012], data augmentation [Pooley et al., 2015, Nguyen-Van-Yen et al., 2021], or a variety of approximations of the latent stochastic epidemic process [Lekone and Finkenstädt, 2006, Cauchemez and Ferguson, 2008, Fintzi et al., 2022]. Scaling our model to the state or national level could be done by analyzing multiple counties independently, or by building a Bayesian hierarchical model that would allow borrowing information among counties. An even more ambitious undertaking would be allowing importation/exportation events across county lines, as was done by Pei et al. [2021]. We hope that our methodology and other works in this spirit, along with better quality of surveillance data, will provide us with better predictive analytics tools when the next pandemic strikes.

Chapter 5

Forecasting during epidemic surges of pathogen variants capable of immune escape

5.1 Introduction

Accurate forecasting of epidemic surges is an important tool for public health agencies to make informed decisions about interventions and resource allocation. Forecasting is particularly challenging when conditions in the environment change. These changes could be due to a variety of factors, including non-pharmaceutical interventions, vaccination campaigns, treatment developments, or phenotypic changes in the pathogen due to evolution. These phenotypic changes may result in increased transmissibility, disease severity, or immune evasion, leading to surges in cases, hospitalizations, or deaths. This has been particularly relevant throughout the COVID-19 pandemic, where the emergence of novel virus strains such as the Alpha, Delta, and Omicron variants was repeatedly accompanied by these surges. In

this work, we focus on developing forecasting methodology for times when a new variant becomes dominant due to its ability to escape immunity induced by infections from previously circulating variants.

Forecasting epidemic surges in changing environments has previously been addressed by including time-varying parameters in infectious disease dynamics models. These time-varying parameters can be treated as fixed inputs or estimated via parametric or non-parametric methods. Many models focus on a time-varying transmission rate. For example, Zhou and Ji [2020] and Xu et al. [2018] modeled the time-varying transmission rate via a Gaussian process. Iyaniwura et al. [2022] use survey data to derive separate time-varying contact rates for different subpopulations and incorporate these derived rates into their transmission rate model. Other models include additional time-varying parameters. In Chapter 4, we developed a model which included time-varying transmission rate, testing policy, and case detection rate. In [O'Dea and Drake, 2022], time-varying transmission rate, reporting probability, and observation variance are modeled with random walks. Similarly, Gibson et al. [2020] also model time-varying transmission rate and case detection rate with random walks. Additionally, Morozova et al. [2021] incorporate time-varying contact rate, transmission rate, recovery rate, severe fraction, hospital discharge rate, and hospital case fatality ratio into a Bayesian model. While none of these works allow the average immunity duration to vary in time, a topic of interest in this chapter, their methods could be adapted to model this dynamic behavior.

Despite the apparent relevance of the novel variants, a recent comprehensive review of 136 infectious disease dynamics models noted that not one of the papers analyzed used variant prevalence data [Nixon et al., 2022]. Since that review's publication, we are aware of two papers which have attempted to use this type of data [Du et al., 2023, Rashed et al., 2022]. These two works used a deep learning framework and developed long short-term memory (LSTM) models which incorporate many data sources, including pre-processed estimates of

variant proportions.

In contrast to these methods, we propose a way to integrate raw counts of variant sequences into a Bayesian compartmental model. Rather than adapting one of the traditional models for multiple strains (see Kucharski et al. [2016] for an overview), which would necessitate many additional compartments and parameters, we propose a modification to a single strain model. We take inspiration from Figgins and Bedford [2022], who model variant proportions over time using a Dirichlet-multinomial likelihood to estimate variant-specific effective reproduction numbers. Analogously, we assume that the observed count of sequences of a novel variant has a beta-binomial distribution whose mean is a product of the total number of observed genetic sequences and the proportion of the infectious population infected with the novel variant. We then model the average duration of immunity as a flexible function of this proportion. The form of our function ensures that the immunity duration decreases when the new variant is introduced to the population, before increasing again once the new variant becomes dominant.

We evaluate our method in a simulation study wherein novel variants become dominant at varying rates and compare our forecasts to those from more conventional Bayesian compartmental models with time-varying parameters. We demonstrate competitive or superior ability to forecast cases, hospital occupancy, ICU occupancy, and deaths, as judged by the continuous ranked probability score — a popular proper scoring rule for probabilistic forecasts [Gneiting and Raftery, 2007]. We also apply our model to real data from the Omicron wave in Orange County, California, and the state of California as a whole. In particular, our proposed model is much better than others at forecasting the timing and size of the peak hospital occupancy, a metric crucial for public health officers making resource allocation recommendations to medical providers.

5.2 Methods

5.2.1 Data

We seek to integrate the following data sources into our modeling: time series of daily numbers of cases, infected hospital occupants, infected ICU occupants, deaths due to the infection, genetic sequences from a novel pathogen variant, and genetic sequences from all variants. These time series need not be observed at the same temporal resolution, but, for clear exposition in this section, we assume they are each reported at times $t_1, \ldots, t_L$. We denote the vector of cases, $\mathbf{W} = (W_1, \ldots, W_L)$; infected hospital occupants, $\mathbf{X} = (X_1, \ldots, X_L)$; infected ICU occupants, $\mathbf{Y} = (Y_1, \ldots, Y_L)$; deaths due to the infection, $\mathbf{Z} = (Z_1, \ldots, Z_L)$; number of genetic sequences from the novel variant, $\mathbf{V}^N = (V_1^N, \ldots, V_L^N)$; total number of genetic sequences from all variants, $\mathbf{V}^A = (V_1^A, \ldots, V_L^A)$. In each vector, the subscript indicates that the observation is made at the corresponding time, $t_1, l = 1, \ldots, L$. We work our way up to a surveillance model for these time series by first describing a transmission model for the underlying population dynamics.

5.2.2 Transmission model

We model latent incidence and prevalence trajectories by dividing the population of interest into seven compartments: susceptible individuals (S), infected, but not yet infectious, individuals (E), infectious individuals (I), individuals hospitalized with infection (H), individuals in the ICU with infection (U), recovered individuals (R), and individuals who died due to infection (D). The potential progression of an individual through these states is presented in Figure 5.1. Note that we assume all deaths are transitions from the ICU compartment, not from the hospitalized or infectious compartments.
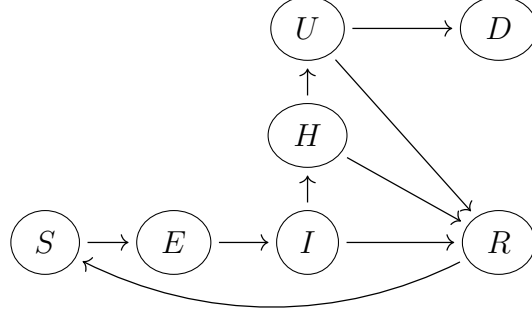
Figure 5.1: Model diagram for potential progression between infection states. Model compartments are susceptible individuals (S), infected, but not yet infectious, individuals (E), infectious individuals (I), individuals hospitalized with infection (H), individuals in the ICU with infection (U), recovered individuals (R), and individuals who died due to infection (D).

We model the time-evolution of the proportions of individuals occupying these compartments with a set of ordinary differential equations (ODEs). For simplicity, we assume a homogeneously mixing population of fixed size N. Let $\mathbf{A}(t) = (S(t),\ E(t),\ I(t),\ H(t),\ U(t),$ $R(t),\ D(t))^T$ denote the population proportions of all compartments at time t. We normalize $\mathbf{A}(t_0)$ so that $\mathbf{A}(t_0)^T \mathbf{1} = 1$, where $\mathbf{1} = (1, 1, 1, 1, 1, 1, 1)^T$. Because we fit this model to incidence data, it is convenient to also keep track of the cumulative proportion of the population that experiences transitions between compartments from t_0 to t: $\mathbf{N}(t) = (N_{SE}(t),\ N_{EI}(t),$ $N_{IH}(t),\ N_{HU}(t),\ N_{UD}(t),\ N_{IR}(t),\ N_{HR}(t),\ N_{UR}(t),\ N_{RS}(t))^T$. To describe mathematically how vectors $\mathbf{A}(t)$ and $\mathbf{N}(t)$ change through time, we first define rates of transitions between compartments, with possible transitions corresponding to the arrows in Figure 5.1:

$$
\begin{aligned}
\lambda_{SE}(S, I) &= \frac{\beta}{N} SI, \\
\lambda_{EI}(E) &= \gamma E, \\
\lambda_{IH}(I) &= \nu\tau I, \\
\lambda_{HU}(H) &= \eta\upsilon H, \\
\lambda_{UD}(U) &= \omega\chi U,
\end{aligned}
\qquad
\begin{aligned}
\lambda_{IR}(I) &= \nu\left(1 - \tau\right)I, \\
\lambda_{HR}(H) &= \eta\left(1 - \upsilon\right)H, \\
\lambda_{UR}(U) &= \omega\left(1 - \chi\right)U, \\
\lambda_{RS}(R) &= \kappa R,
\end{aligned}
\tag{5.1}
$$

90

where β is the transmission rate, N is the population size, $1/\gamma$ is the mean latent period duration, $1/\nu$ is the mean infectious period duration, $1/\eta$ is the mean hospitalization duration, $1/\omega$ is the mean ICU stay duration, $1/\kappa$ is the mean immunity duration, τ is the infection-hospitalization ratio, υ is the hospitalization-ICU admission ratio, and χ is the ICU-fatality ratio. In some cases, we allow certain parameters to vary in time to capture changing dynamics without the need for additional compartments in the model. These changing dynamics could be due to policy or behavioral changes, or the presence of a new variant with increased transmissibility or immune escape. In particular, in the models we consider in our simulation study and application, we replace β and κ with time-varying versions, $\beta(t)$, and $\kappa(t)$. We discuss the details of this implementation in Section 5.2.4.

Using the transition rates in (5.1), we define the ODEs in the model:

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= \lambda_{RS}(R) - \lambda_{SE}(S, I), & \frac{\mathrm{d}N_{SE}}{\mathrm{d}t} &= \lambda_{SE}(S, I), \\
& & \frac{\mathrm{d}N_{EI}}{\mathrm{d}t} &= \lambda_{EI}(E), \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \lambda_{SE}(S, I) - \lambda_{EI}(E), & \frac{\mathrm{d}N_{IH}}{\mathrm{d}t} &= \lambda_{IH}(I), \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \lambda_{EI}(E) - (\lambda_{IH}(I) + \lambda_{IR}(I)), & \frac{\mathrm{d}N_{HU}}{\mathrm{d}t} &= \lambda_{HU}(H), \\
\frac{\mathrm{d}H}{\mathrm{d}t} &= \lambda_{IH}(I) - (\lambda_{HU}(H) + \lambda_{HR}(H)), & \frac{\mathrm{d}N_{UD}}{\mathrm{d}t} &= \lambda_{UD}(U), \\
\frac{\mathrm{d}U}{\mathrm{d}t} &= \lambda_{HU}(H) - (\lambda_{UD}(U) + \lambda_{UR}(U)), & \frac{\mathrm{d}N_{IR}}{\mathrm{d}t} &= \lambda_{IR}(I), \\
\frac{\mathrm{d}D}{\mathrm{d}t} &= \lambda_{UD}(U) & \frac{\mathrm{d}N_{HR}}{\mathrm{d}t} &= \lambda_{HR}(H), \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \lambda_{IR}(I) + \lambda_{HR}(H) + \lambda_{UR}(U) - \lambda_{RS}(R), & \frac{\mathrm{d}N_{UR}}{\mathrm{d}t} &= \lambda_{UR}(U), \\
& & \frac{\mathrm{d}N_{RS}}{\mathrm{d}t} &= \lambda_{RS}(R),
\end{aligned}
\tag{5.2}
$$

subject to initial conditions $\mathbf{A}(t_0) = (S_0, E_0, I_0, H_0, U_0, D_0, R_0)$. Because the total population size is fixed at size N, and we take H_0, U_0, and D_0 to match the reported values of the

hospitalizations, ICU occupancy, and deaths in the data at t_0, we are left with E_0, I_0, and R_0 as free parameters (because $S_0 = N - (E_0 + I_0 + H_0 + U_0 + D_0 + R_0)$). The equations in (5.2) are redundant, but tracking both the prevalence (left column) and the cumulative incidence (right column) is useful for linking the transmission model to data via a surveillance model.

5.2.3   Surveillance model

We fit the transmission model to six time series: number of new cases $(\mathbf{W})$, number of infected hospital occupants $(\mathbf{X})$, number of infected ICU occupants $(\mathbf{Y})$, number of new deaths due to the infection $(\mathbf{Z})$, number of genetic sequences from the novel variant $(\mathbf{V}^N)$, and number of genetic sequences from all variants $(\mathbf{V}^A)$ observed at time points $t_1, \dots t_l$. The count for each non-genetic data stream at each time (l) is modelled by a negative binomial distribution, which we parameterize in terms of its mean (μ) and variance (σ^2). The degree of over-dispersion in each of these negative binomial distributions is controlled by a ϕ parameter. The number of sequences from the novel variant, is modeled with a beta-binomial distribution which is conditional on the total number of observed sequences. We parameterize this distribution in terms of δ, its mean success probability and ϕ, an over-dispersion parameter.

We model the observed cases, W_l, in the interval $(t_{l-1}, t_l]$, for $l = 1 \dots L$, by

$$W_l \sim \text{NegativeBinomial}\left(\mu_l^W = \rho^W \cdot N \cdot \Delta N_{EI}(t_l), \left(\sigma_l^W\right)^2 = \mu_l^W\left(1 + \mu_l^W/\phi_W\right)\right), \quad (5.3)$$

where $\rho^W \in [0, 1]$ is the overall case detection rate and $\Delta N_{EI}(t_l) = N_{EI}(t_l) - N_{EI}(t_{l-1})$, meaning that the number of observed cases is assumed to be some noisy realization of the fraction of the true new cases in the population, as estimated by the ODE model. In some models, we allow the case detection rate to vary over time, in which case we replace ρ^W with

$\rho^W(t_l)$ in (5.3). We discuss modeling choices for this and $\delta(t_l)$ in the next section.

The observed number of hospital occupants with an infection at time t_l is modeled by

$$X_l \sim \text{NegativeBinomial}\left(\mu_l^X = N \cdot H(t_l), \left(\sigma_l^X\right)^2 = \mu_l^X\left(1 + \mu_l^X/\phi_X\right)\right), \qquad (5.4)$$

which means that the observed number of hospital occupants is a noisy realization of the hospitalized population, as estimated by the ODE model.

We assume that the number of ICU occupants with an infection has the distribution

$$Y_l \sim \text{NegativeBinomial}\left(\mu_l^Y = N \cdot U(t_l), \left(\sigma_l^Y\right)^2 = \mu_l^Y\left(1 + \mu_l^Y/\phi_Y\right)\right), \qquad (5.5)$$

indicating that the observed number of ICU occupants is a noisy realization of the ICU population, as estimated by the ODE model.

The observed deaths due to the infection are modeled by

$$Z_l \sim \text{NegativeBinomial}\left(\mu_l^Z = \rho^Z \cdot N \cdot \Delta N_{UD}(t_l), \left(\sigma_l^Z\right)^2 = \mu_l^Z\left(1 + \mu_l^Z/\phi_Z\right)\right), \qquad (5.6)$$

where $\rho^Z \in [0, 1]$ is the overall death detection rate and $\Delta N_{UD}(t_l) = N_{UD}(t_l) - N_{UD}(t_{l-1})$, meaning that the number of observed deaths is assumed to be some noisy realization of the fraction of the true new deaths in the population, as estimated by the ODE model.

We model the observed counts of genetic sequences of the novel variant, conditional on the observed number of all genetic sequences by

$$V_l^N \sim \text{Beta-Binomial}\left(V_l^A, \alpha = \phi_V \delta(t_l), \beta = \phi_V(1 - \delta(t_l))\right), \qquad (5.7)$$

where $\delta(t_l)$ is the proportion of currently infectious individuals who are infected with the

novel variant. Thus, the observed proportion of genetic sequences of the novel variant is a noisy realization of the true proportion in the population.

5.2.4 Bayesian inference

In Sections 5.2.2 and 5.2.3, we noted that, for some models considered in this work, β, κ, and ρ^W could be substituted for time-varying counterparts, $\beta(t)$, $\kappa(t)$, and $\rho^W(t)$. We now present some options for modeling how these parameters may change over time. We proceed by reparameterizing $\beta(t)$ as basic reproduction number, $R_0(t) = \beta(t)/\nu$, and $\kappa(t)$ as the average immunity duration, $1/\kappa(t)$.

In some models, we parameterize $R_0(t)$, $1/\kappa(t)$, or $\rho(t)$ as piecewise constant functions, where each vector defining the constants a prior follows a Gaussian Markov random field (GMRF). More precisely, we define the auxiliary vectors:

$$\tilde{\boldsymbol{R}_0} = (\tilde{R}_{0,1}, \tilde{R}_{0,2}, \ldots, \tilde{R}_{0,L}), \ \tilde{\boldsymbol{\kappa}} = (\tilde{\kappa}_1, \tilde{\kappa}_2, \ldots, \tilde{\kappa}_L), \text{ and } \tilde{\boldsymbol{\rho}^W} = (\tilde{\rho}_1^W, \tilde{\rho}_2^W, \ldots, \tilde{\rho}_L^W),$$

which follow the Gaussian Markov random field priors

$$
\begin{aligned}
\tilde{R}_{0,l} &\sim N\left(\tilde{R}_{0,l-1}, \sigma_{R_0}^2\right), \text{ where } l = 2, \ldots, L \text{ and } \tilde{R}_{0,1} \sim N\left(\mu_{R_{01}}, \sigma_{R_{01}}^2\right), \\
\frac{1}{\tilde{\kappa}_l} &\sim N\left(\frac{1}{\tilde{\kappa}_{l-1}}, \sigma_\kappa^2\right), \text{ where } l = 2, \ldots, L \text{ and } \frac{1}{\tilde{\kappa}_1} \sim N\left(\mu_{\kappa_1}, \sigma_{\kappa_1}^2\right), \\
\tilde{\rho}_l^W &\sim N\left(\tilde{\rho}_{l-1}^W, \sigma_{\rho^Y}^2\right), \text{ where } l = 2, \ldots, L \text{ and } \tilde{\rho}_1^W \sim N\left(\mu_{\rho_1^W}, \sigma_{\rho_1^W}^2\right),
\end{aligned}
\tag{5.8}
$$

and define the piecewise constant functions:

$$R_0(t) = \sum_{l=1}^{L} \exp\left\{ \left(\tilde{R}_{0,l} \right) \mathbb{1}\left[t \in (t_{l-1}, t_l] \right] \right\}, \tag{5.9}$$

$$1/\kappa(t) = \sum_{l=1}^{L} \exp\{ (1/\tilde{\kappa}_l) \ \mathbb{1}\left[t \in (t_{l-1}, t_l] \right] \}, \tag{5.10}$$

$$\rho^W(t) = \sum_{l=1}^{L} \frac{\exp\left\{ \left(\tilde{\rho}_l^W \right) \mathbb{1}\left[t \in (t_{l-1}, t_l] \right] \right\}}{\exp\{ (\tilde{\rho}_l^W) \ \mathbb{1}\left[t \in (t_{l-1}, t_l] \right] \} + 1}. \tag{5.11}$$

In our simulation study and application, we assume that these time-varying dynamics are due to the emergence of a novel variant to which the population has reduced immunity. Thus, we hypothesize that defining time-varying parameters as a function of $\delta(t)$, the proportion of currently infectious individuals who are infected with the novel variant, may lead to improvements in forecasting capabilities. In principle, we can do this for any of the parameters in the model that might differ by variant, but we only consider this parameterization for the average immunity duration, $1/\kappa$, in our simulation study and application. We model $1/\kappa(t)$ as dependent on $\delta(t)$ by

$$1/\kappa(t) = 1/\hat{\kappa} \left(\delta(t - \zeta) \right) = \exp\left\{ \alpha_0 + \alpha_1 \left[\delta(t - \zeta) \left(1 - \delta(t + \zeta) \right) \right]^{\alpha_2} \right\}, \tag{5.12}$$

where $\alpha_2 > 0$. This functional form ensures that the average immunity duration is maximized when the entire infectious population is infected with the same variant (i.e., $\delta = 0$ or $\delta = 1$) and minimized when $\delta = 0.5$. Thus, when a new variant is introduced, but only constitutes a small proportion of the infectious population, the average immunity duration begins to decrease, moving more of the population to the susceptible compartments. The immunity duration continues to decrease until the new variant is half of the infectious population, at which point there is a large infectious population carrying the new variant and a large proportion of the population with reduced immunity to the new variant. As the new variant reaches dominance, the immunity duration returns to its original level. We allow

this relationship to be flexible by including an offset in time, ζ.

We note that the maximum immunity duration is $\exp(\alpha_0)$, and the minimum immunity duration is $\exp\{\alpha_0 + \alpha_1 (1/4)^{\alpha_2}\}$. To use a more interpretable prior, we reparameterize α_1 in terms of α_1^*:

$$\alpha_1 = 4^{\alpha_2} \ln(\alpha_1^*). \tag{5.13}$$

Then the minimum immunity duration is $\exp\{\alpha_0 + 4^{\alpha_2} \ln(\alpha_1^*)(1/4)^{\alpha_2}\} = \alpha_1^* \exp\{\alpha_0\}$. Recalling that $\exp\{\alpha_0\}$ is the maximum immunity duration, we see that $\alpha_1^* \in [0, 1]$ is simply a scaling factor, which determines the minimum duration, based on the maximum duration. We show several example curves for (5.12) in Figure 5.2.
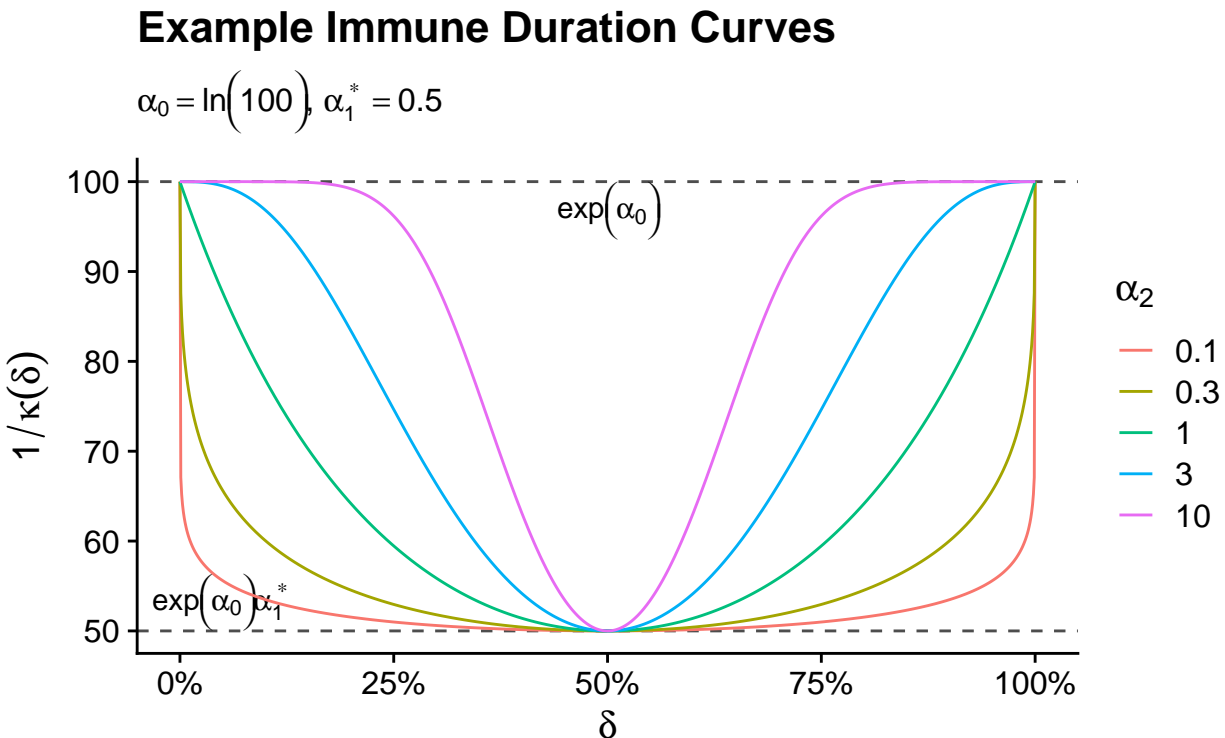
Figure 5.2: Example immunity duration curves for several values of α_2, with $\alpha_0 = \ln(100)$, $\alpha_1^* = 0.5$. The full formula for these curves is given by (5.12) and (5.13).

Now, we discuss our model for $\delta(t)$, the proportion of currently infectious individuals who

are infected with the novel variant. We assume that, after the novel variant is introduced, its proportion in the population evolves according to logistic growth:

$$\delta(t) = \frac{\exp\left\{\iota_0 + \iota_1(t)\right\}}{\exp\left\{\iota_0 + \iota_1(t)\right\} + 1}. \tag{5.14}$$

This parametric approach is borrowed from population genetics literature devoted to studying the dynamics of the novel variant proportion in the population of interest [Lacerda and Seoighe, 2014, Zhao et al., 2023]. To set interpretable priors on ι_0 and ι_1, we parameterize the logistic growth as

$$\delta(t) = \frac{\exp\left\{\iota_0 + \iota_1(t - t^*)\right\}}{\exp\left\{\iota_0 + \iota_1(t - t^*)\right\} + 1}, \tag{5.15}$$

where t^* is the time point for which $\delta(t^*) = \frac{\exp\{\iota_0\}}{\exp\{\iota_0\}+1}$. Then $\iota_1 = \frac{\ln\left(\frac{0.99}{1-0.99}\right) - \ln\left(\frac{0.01}{1-0.01}\right)}{\iota_1^*}$, where ι_1^* is the amount of time the novel variant takes to grow from 1% of the infectious population to 99% of the infectious population.

With these parameters all defined, we describe our Bayesian procedure. We collect all our model parameters into a vector $\boldsymbol{\theta}$ and write the likelihood function as

$$\Pr\left(\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{V}^N \mid \boldsymbol{\theta}\right)$$
$$= \Pr\left(\mathbf{W} \mid \boldsymbol{\theta}\right) \Pr\left(\mathbf{X} \mid \boldsymbol{\theta}\right) \Pr\left(\mathbf{Y} \mid \boldsymbol{\theta}\right) \Pr\left(\mathbf{Z} \mid \boldsymbol{\theta}\right) \Pr\left(\mathbf{V}^N \mid \boldsymbol{\theta}\right)$$
$$= \prod_{l=1}^{L} \Pr\left(W_l \mid \boldsymbol{\theta}\right) \Pr\left(X_l \mid \boldsymbol{\theta}\right) \Pr\left(Y_l \mid \boldsymbol{\theta}\right) \Pr\left(Z_l \mid \boldsymbol{\theta}\right) \Pr\left(V_l^N \mid \boldsymbol{\theta}\right),$$

where $\Pr\left(W_l \mid \boldsymbol{\theta}\right)$, $\Pr\left(X_l \mid \boldsymbol{\theta}\right)$, $\Pr\left(Y_l \mid \boldsymbol{\theta}\right)$, $\Pr\left(Z_l \mid \boldsymbol{\theta}\right)$, and $\Pr\left(V_l^N \mid \boldsymbol{\theta}\right)$, are the probability mass functions given by (5.3), (5.4), (5.5), (5.6), and (5.7), respectively. We encode available information about our model parameters in a prior distribution with density $\pi(\boldsymbol{\theta})$. We assume that all univariate non-GMRF distributed parameters are *a priori* independent and list our prior assumptions for our simulation study in Table C.2 and for our application in

Table C.4. We base all our inferences and predictions on the posterior distribution of all model parameters:

$$\Pr\left(\boldsymbol{\theta} \mid \mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{V}^N\right) \propto \Pr\left(\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{V}^N \mid \boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}\right). \tag{5.16}$$

We sample from this posterior using the No-U-Turn Sampler, [Hoffman and Gelman, 2014] as implemented in the Turing Julia package [Ge et al., 2018]. Model code and data are available at the following GitHub repository: `https://github.com/damonbayer/immunity_semi_parametric_model`.

In total, the form of the model without any GMRF components has 24 parameters. Adding a GMRF component requires one additional parameter for the variance and one more parameter per observation time. In our simulation study and real data analyses, the number of model parameters ranges between 24 and 91.

5.3   Results

5.3.1   Simulation study

We simulate three data sets for this study, each where the novel variant becomes dominant at a different speed: slow (24 weeks to go from 1% to 99% of sequences), medium (13 weeks), and fast (7 weeks). The data sets are simulated from a two strain model, which is depicted in Figure 5.3 and explained in full detail in Appendix C.1. Briefly, the model is similar to the one depicted in Figure 5.1 but is modified to account for two distinct disease variants. The variants associated with each compartment are indicated by the subscripts: "N" for the novel variant, "O" for other variants, and "A" for all types for variants. When all compartments associated with one variant type are empty, the model is equivalent to the

one from Figure 5.1.

These otherwise independent models are linked by allowing transitions from S_A and R_O to E_N. That is, people who are susceptible to all variant types (S_A) and those who are only recovered from other variants (R_O) can become infected with the novel variant (E_N). The rates for these transitions are given by

$$\lambda_{S_A E_N}(S_A, I_N) = \frac{\beta_0}{N} S_A I_N \quad \text{and} \quad \lambda_{R_O E_N}(R_O, I_N) = \epsilon \frac{\beta_N}{N} R_O I_N, \tag{5.17}$$

where β and N are defined as in (5.1), and $\epsilon \in [0, 1]$ is a factor that represents the effect of partial immunity to the novel variant conferred by a recent infection from another variant. After some time, all the individuals in the O subscript compartments will have transitioned to the N subscript compartments, and the model again behaves like the one in Figure 5.1.

Table 5.1: Simulation parameters that differ by scenario.

Scenario	β_N	ϵ
Slow	2.1	0.75
Medium	2.1	1.0
Fast	3.5	1.0
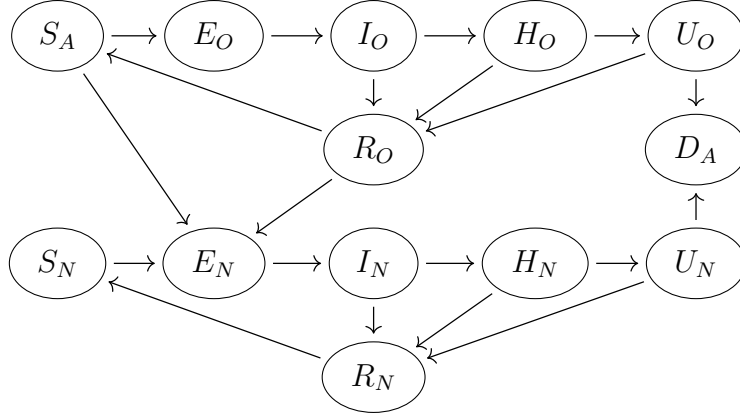
Figure 5.3: Compartmental diagram for simulation model. Model compartments are susceptible individuals (S), infected, but not yet infectious, individuals (E), infectious individuals (I), individuals hospitalized with infection (H), individuals in the ICU with infection (U), recovered individuals (R), and individuals who died due to infection (D). The variants associated with each compartment are indicated by the subscripts: "N" for the novel variant, "O" for other variants, and "A" for all types for variants.

For our simulations, the compartments are initiated with the entire population (3 million) in S_0, except a small number who are in E_O. Because all the N subscript compartments are empty, this behaves like the model in Figure 5.1. We simulate the compartment trajectories until the initial surge begins to subside. While the overall prevalence is beginning to decrease, we simulate an importation of 1000 people who are infectious with the novel variant into the I_N compartment. Then, the individuals begin to transition into the N subscript compartments, resulting in a second wave. Forecasting this second wave is the objective of our modeling efforts.

We manipulate the size of this wave and the speed with which the novel variant takes over by altering β and ϵ in the different simulation settings. Table 5.1 presents the exact β and ϵ values used. The other simulation parameters are given in Table C.1. Changes of the resulting proportion of individuals infected with the novel variant over time for each simulated scenario is presented in Figure 5.4. The simulated data for the medium takeover speed scenario is plotted in Figure 5.5, with the gray shaded areas indicating the time points for which we produce forecasts. Similar figures for the other data sets are presented in

Figures C.1 and C.2. In each scenario, non-genetic data is reported at a weekly resolution, starting 20 weeks into the simulation, which we call $t = 0$. Genetic data is reported at a daily resolution, which is crucial to capturing the early exponential-like growth of the novel variant, relative to the other variants. Additionally, the genetic data is only reported beginning one week before the novel variant importation event, which is necessary to meet the assumption of logistic growth in (5.14).
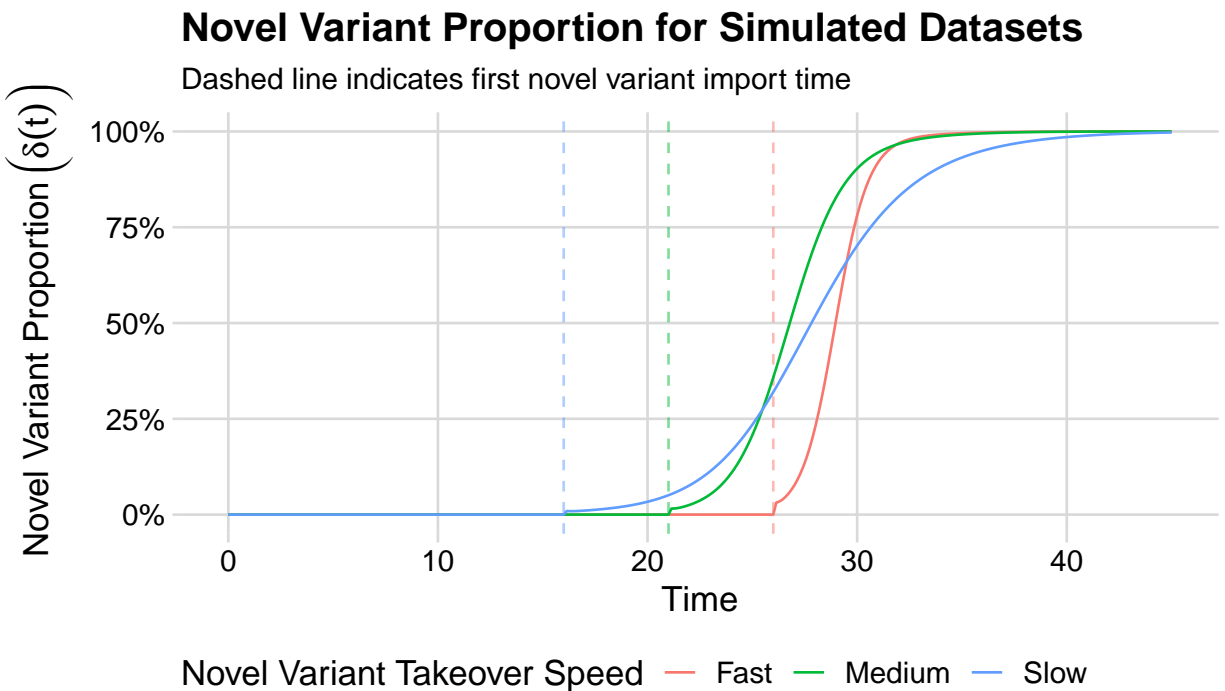
Figure 5.4: Proportion of infectious individuals infected with the novel variant over time for three simulated scenarios. The dashed line indicates the time that the initial importation event of the novel variant occurs.

Simulated Data

Medium Novel Variant Takeover Speed
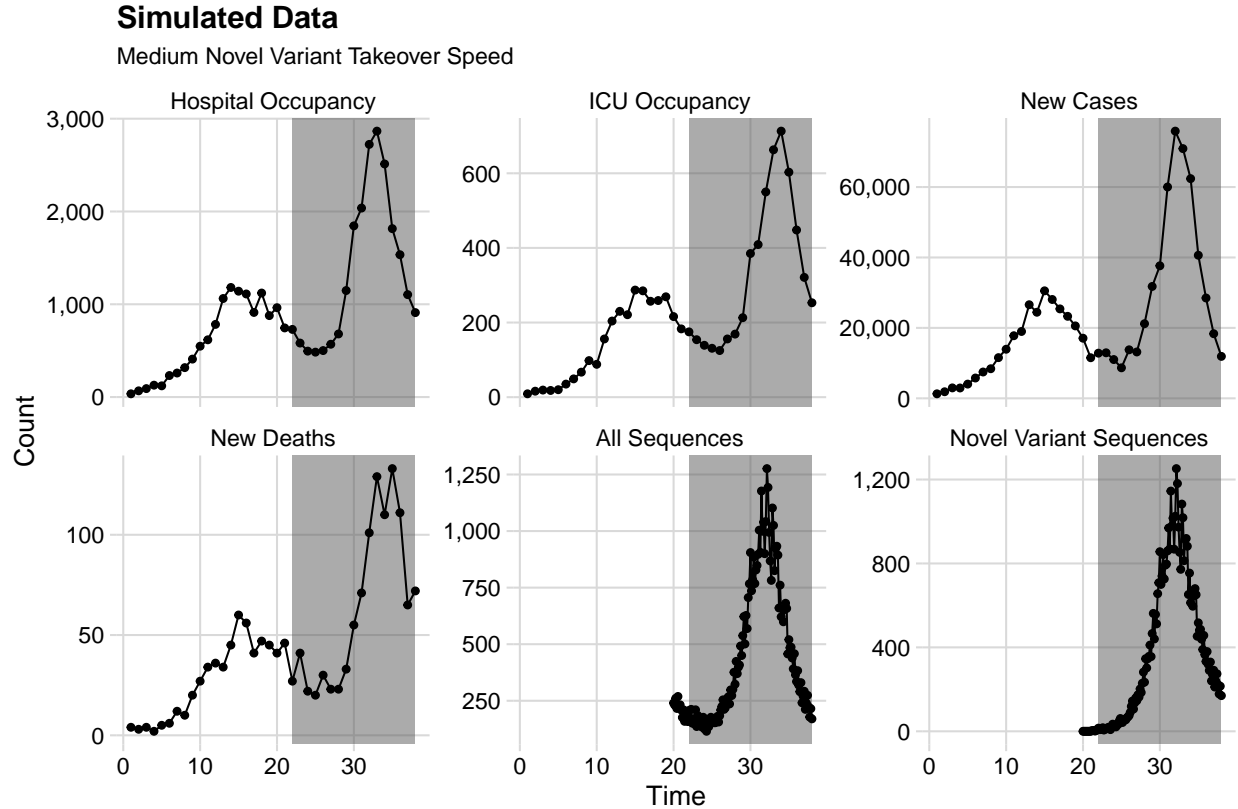
Figure 5.5: Simulated data set for the medium takeover speed scenario. The gray shaded areas indicate the time points for which we produce forecasts

We fit three models to these simulated data sets: (1) a model where $R_0(t)$ is a priori modeled as a GMRF and $1/\kappa(t)$ is constant, (2) a model where $R_0(t)$ is constant and $1/\kappa(t)$ is a priori modeled as a GMRF, and (3) a model where $R_0(t)$ is constant and $1/\kappa(t)$ is a function of the proportion of infectious individuals infected with the novel variant shown in (5.12). The first model is reflective of typical practice in infectious disease modeling (see, e.g., [Gibson et al., 2020, O'Dea and Drake, 2022]). The second is a slight variation on the first, where we make the immunity duration, rather than the basic reproductive number, vary over time. The third model is our main innovation, which models the mean immunity duration based on genetic data. The priors used in these models are presented in Table C.2. For each simulated data set, we used four Markov chains run in parallel to draw a total of 1000 posterior samples.

While these models forecast several data streams, we present only forecasts related to hospi-

talization in the main text, as they are the most relevant for public health policymakers. We present additional results for cases, ICU occupancy, and deaths in Section C.2. In general, these results exhibit the same patterns as observed in the hospitalization results. Additionally, we conducted a sensitivity analysis wherein we modified the prior for the anticipated speed of the novel variant takeover and found our model to be robust to these changes. These priors are presented in Table C.3. The results of this analysis are presented in Section C.3.

We first show 1, 2, and 4-week ahead forecasts for the simulated medium takeover speed data in Figure 5.6. Analogous plots for the slow and fast takeover data are presented in Figures C.3 and C.4, respectively.

From these figures, we note that, for all models and all data sets, one-week ahead forecasts match the data quite closely. As may be expected, performance degrades as the forecast horizon increases. For the slow takeover data, the forecasts where $1/\kappa(t)$ is informed by the genetic data closely match the data up to the four-week ahead forecasts. The forecasts where $1/\kappa(t)$ is a priori modeled as a GMRF also match the data quite well, while the model where $R_0(t)$ is a priori modeled as a GMRF is much less sharp than the other models. For the medium data, the forecasts with $1/\kappa(t)$ informed by the genetic data again predict the data well up to the four-week forecast horizon. The $1/\kappa(t)$ GMRF model misjudges the height and the timing of the peak, and the forecasts from the $R_0(t)$ GMRF model are again much wider than the other models. When forecasting the fast variant takeover data, the $R_0(t)$ GMRF model is competitive with the $1/\kappa(t)$ genetic model, while the $1/\kappa(t)$ GMRF model again misjudges the peak.

Forecast Comparison for Hospital Occupancy

Medium Takeover Speed Data
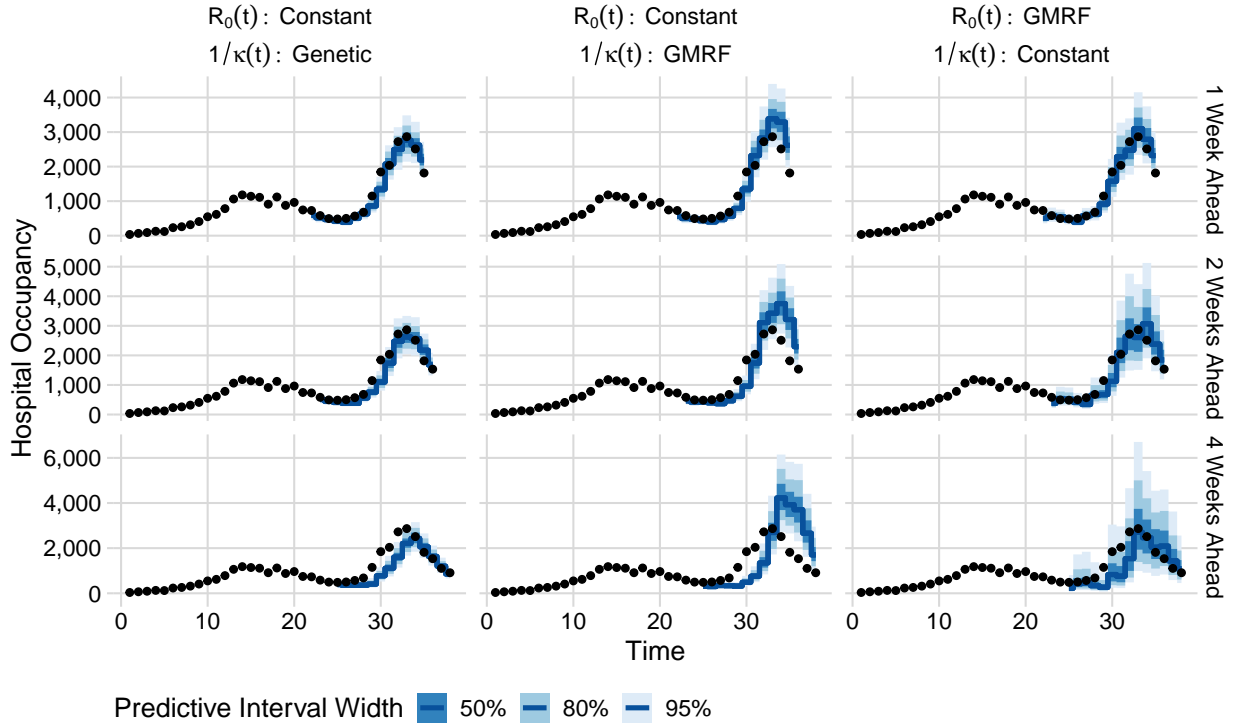
Figure 5.6: Hospital occupancy forecasts from three models at 1, 2, and 4-week forecast horizons for the simulated medium takeover speed data.

To numerically assess the quality of these forecasts, we compute the continuous ranked probability score (CRPS) for each forecast at 1, 2, and 4-week horizons. Plots summarizing the average scores of the hospitalization forecasts are displayed in Figure 5.7. Analogous plots for forecasts of cases, ICU occupancy, and deaths are presented in Figures C.5–C.7. Additionally, we show the scores for each individual forecast for each data stream in Figures C.8–C.11. We observe that for the slow and medium takeover speed scenarios, the model where $1/\kappa(t)$ is informed by the genetic data achieves the lowest (best) average scores. For the slow takeover data, the model where $1/\kappa(t)$ is modeled by a GMRF performs better than the model where $R_0(t)$ is a priori modeled as a GMRF. For the medium takeover data, the two GMRF models are more competitive with each other. With respect to the fast takeover scenario, the model where $R_0(t)$ is modeled as a GMRF achieves slightly lower scores than the model that uses

genetic data, and both of these models outperform the model where $1/\kappa(t)$ is modeled as a GMRF.
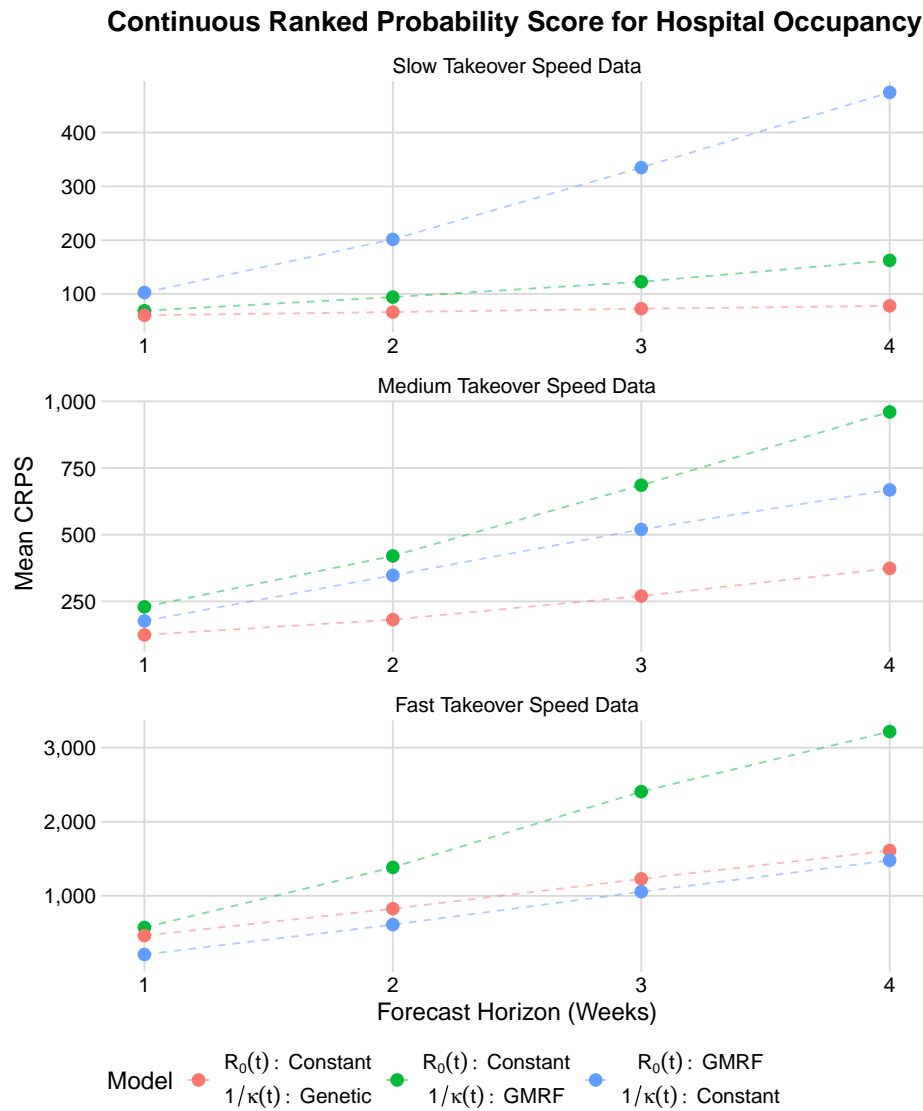
Figure 5.7: CRPS summaries for hospital occupancy forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.

We also specifically assess each model's ability to estimate the timing and size of the peak hospital occupancy. Posterior predictive intervals for the peak timing and peak value are presented in Figures 5.8 and 5.9. In general, the model where $1/\kappa(t)$ is informed by the genetic data tends to have the smallest predictive interval widths and is overconfident about

peak timing and value early on, but can accurately forecast these quantities when subsequent data is included in the model. The model where $R_0(t)$ follows a GMRF has the widest intervals and is, in general, under-confident about the peak timing and value, even when the peak has already passed. The model where $1/\kappa(t)$ follows a GMRF falls somewhere in the middle.
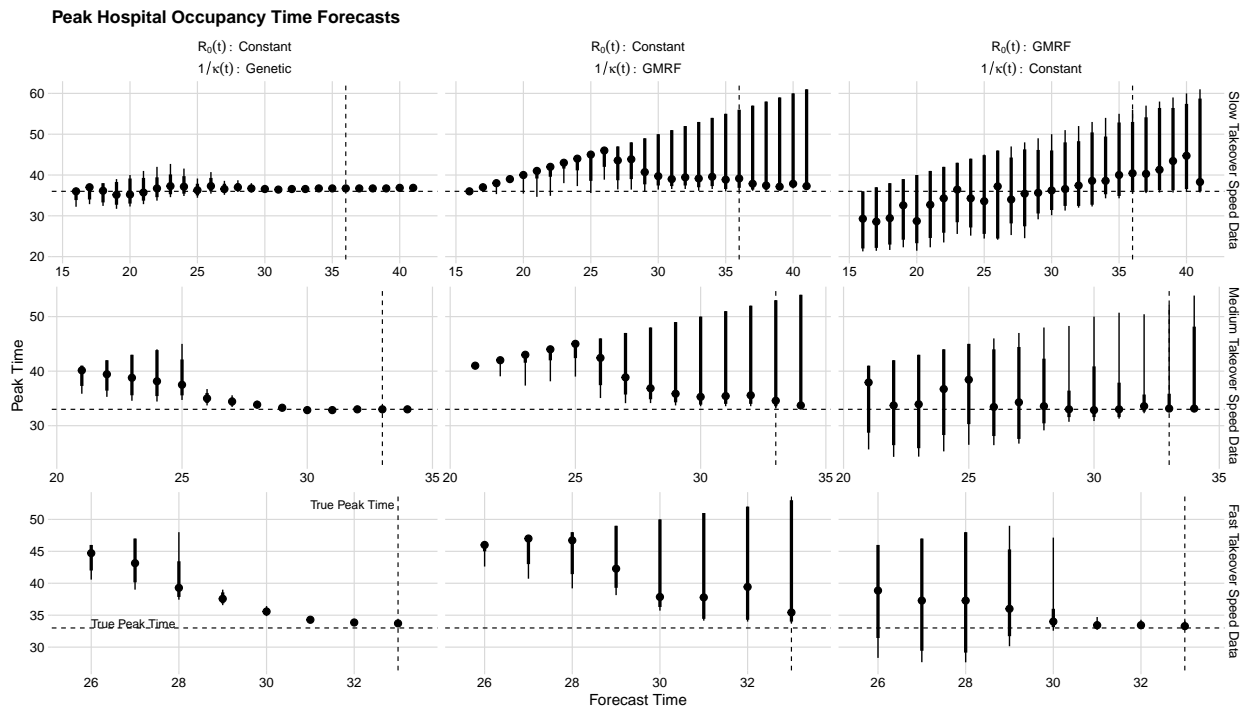
Figure 5.8: Posterior predictive intervals for the time at which hospital occupancy reaches its maximum in three simulated data sets. Dots indicate the median of the predictive distribution, while the thick and thin lines represent central 80% and 95% intervals, respectively. Horizontal and vertical dashed lines indicate the true peak hospitalization time.

Peak Hospital Occupancy Value Forecasts
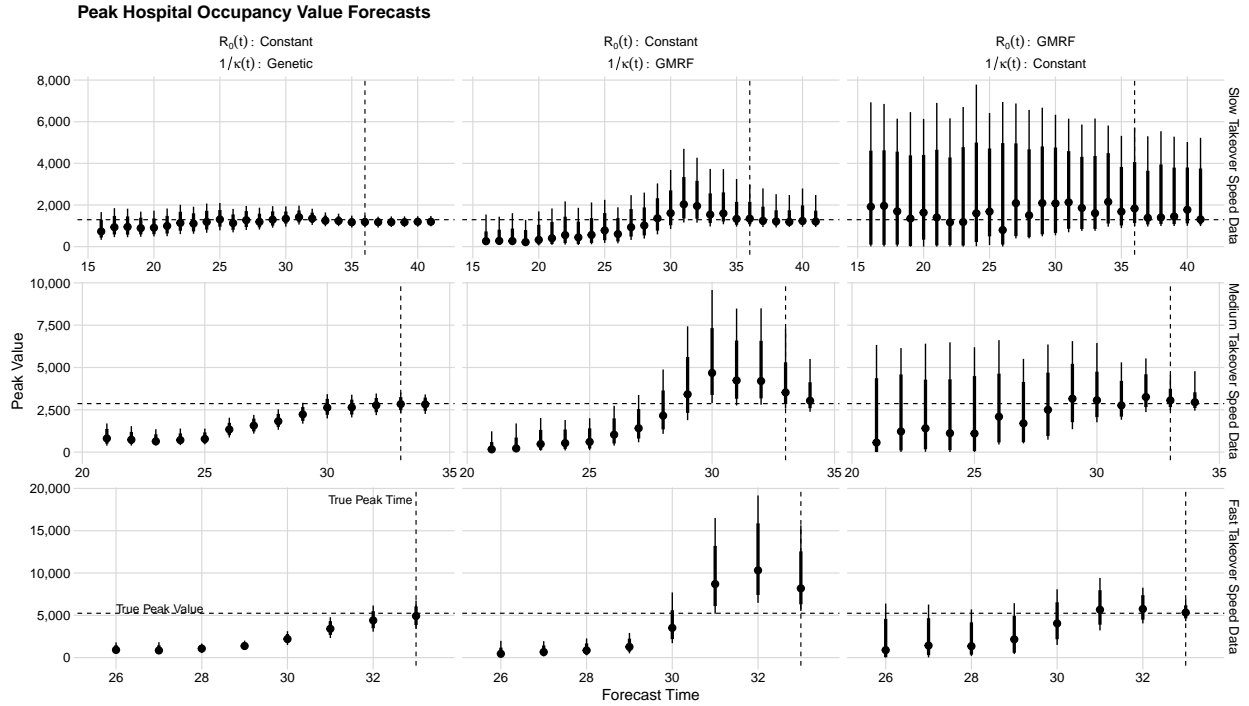
Figure 5.9: Posterior predictive intervals for the maximum hospital occupancy in three simulated data sets. Dots indicate the median of the predictive distribution, while the thick and thin lines represent central 80% and 95% intervals, respectively. Horizontal dashed lines indicate the true peak hospital occupancy, while the vertical dashed lines indicate the true peak hospital occupancy time.

Summaries of the average scores for these forecasts are depicted in Figure 5.10. The score at each forecast time is shown in Figure C.12. For the slow takeover data, the model where $1/\kappa(t)$ is informed by the genetic data achieves the best scores for both peak timing and value. When the models are fit to the medium and fast takeover data, the model where $R_0(t)$ is a priori modeled as a GMRF performs the best.

Continuous Ranked Probability Score for Peak Hospitalization
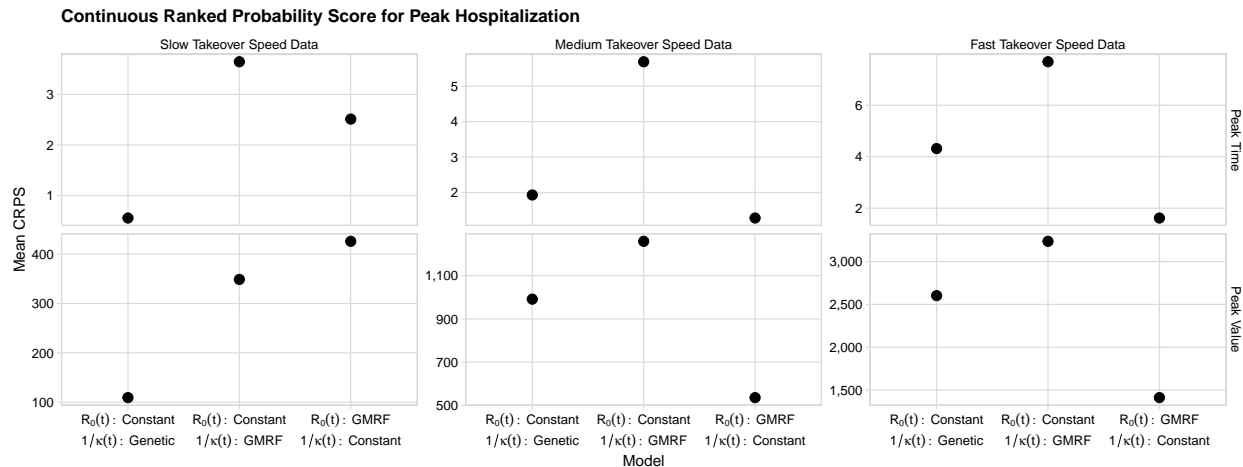
Figure 5.10: CRPS summaries for peak hospital occupancy timing and size for three simulated data sets. Lower CRPS is better.

Beyond the forecasting advantages, the model that uses genetic data is more computationally efficient and uses the same number of parameters, regardless of the number of observation times. In contrast, the number of parameters in the models with GMRFs scales with the number of observation times. A summary of the computation time required to fit models in the simulation study is presented in Table 5.2.

Table 5.2: Computation time summary for all models in our simulation study. Each fit consists of four Markov chains run in parallel to draw a total of 1000 posterior samples.

Model	Avg. CPU Hrs	Min. CPU Hrs	Max. CPU Hrs
$R_0(t)$: Constant, $1/\kappa(t)$: Genetic	2.54	1.53	3.61
$R_0(t)$: Constant, $1/\kappa(t)$: GMRF	5.47	2.93	8.63
$R_0(t)$: GMRF, $1/\kappa(t)$: Constant	11.05	4.15	21.70

5.3.2 Application to California data

In our real data application, we focus on modeling the wave of cases, hospitalizations, ICU admissions, and deaths associated with the first Omicron variant of SARS-CoV-2 in Cal-

ifornia. This wave lasted from roughly December 2021 to March 2022, but we use data beginning in May 2021 to forecast the wave, as this time period includes the previous wave of cases. We fit models to both Orange County data and statewide California data. The time series of daily counts of cases, hospital occupancy, ICU occupancy, and deaths due to COVID-19 are provided by the California Department of Public of Health, and published on the California Open Data Portal (`https://data.ca.gov`). Additionally, the daily counts of sequenced viruses, aggregated by pango lineage [Rambaut et al., 2020], are provided by the Global Initiative on Sharing All Influenza Data (GISAID) [Shu and McCauley, 2017] and made available via Outbreak.info [Gangavarapu et al., 2023]. To match the format described in Section 5.2, the lineages are further aggregated into those that begin with "BA.1" (e.g., BA.1, BA.1.1, BA.1.17.2) and those that do not. The BA.1 sequences are the "novel" variant, and the BA.1 and non-BA.1 sequences are summed together as "all" variants. The non-genetic time series are aggregated at a weekly level, while the genetic data is used at a daily resolution, beginning about one week before the first novel variant is observed. Because the BA.1 variant becomes dominant so quickly, the daily observation of the genetic data is crucial to capture this change, while the weekly observation of other data streams prevents us from needing to account for data inconsistencies, like the "weekend effect" where fewer cases are reported on weekends.

Figure 5.11 shows the binned data at the statewide level, while Figure 5.12 displays the binned data from Orange County. The gray highlighted regions indicate the times for which we produce forecasts. We note that, while the peak weekly case count in the BA.1 wave is about six times higher than the first wave depicted in the figure, the peak hospitalizations are only twice as high in the second wave, and the peak ICU occupancy is only about 20% higher in the second wave. This is a clear departure from the models we fit in our simulation study, where the second peak's height compared to the first peak is the same for all data streams. Because of this, we chose to fit the same models as in the simulation study, but modified them to also include modeling the case detection rate, $\rho^W(t)$, with a

109

GMRF. In reality, the reasoning for the changing relationship between the peaks is complex and difficult to untangle, even in retrospect. It is likely that some combination of partial protection from previous infections or vaccinations and possible less inherent propensity for severe outcomes in the Omicron variant are at play. Anticipating and modeling this scenario in a real forecasting scenario with a novel variant would be extremely difficult, so we chose to incorporate some additional flexibility into the model in a place that could be realistic and useful.

Figure 5.11: COVID-19 surveillance data from California. The plots show weekly counts of cases, hospital and ICU occupancy of patients with COVID-19, reported deaths due to COVID-19, as well as counts of virus sequences for Omicron BA.1 and all lineages, and the proportion of BA.1 lineages. The gray highlighted regions indicate the times for which we produce forecasts.
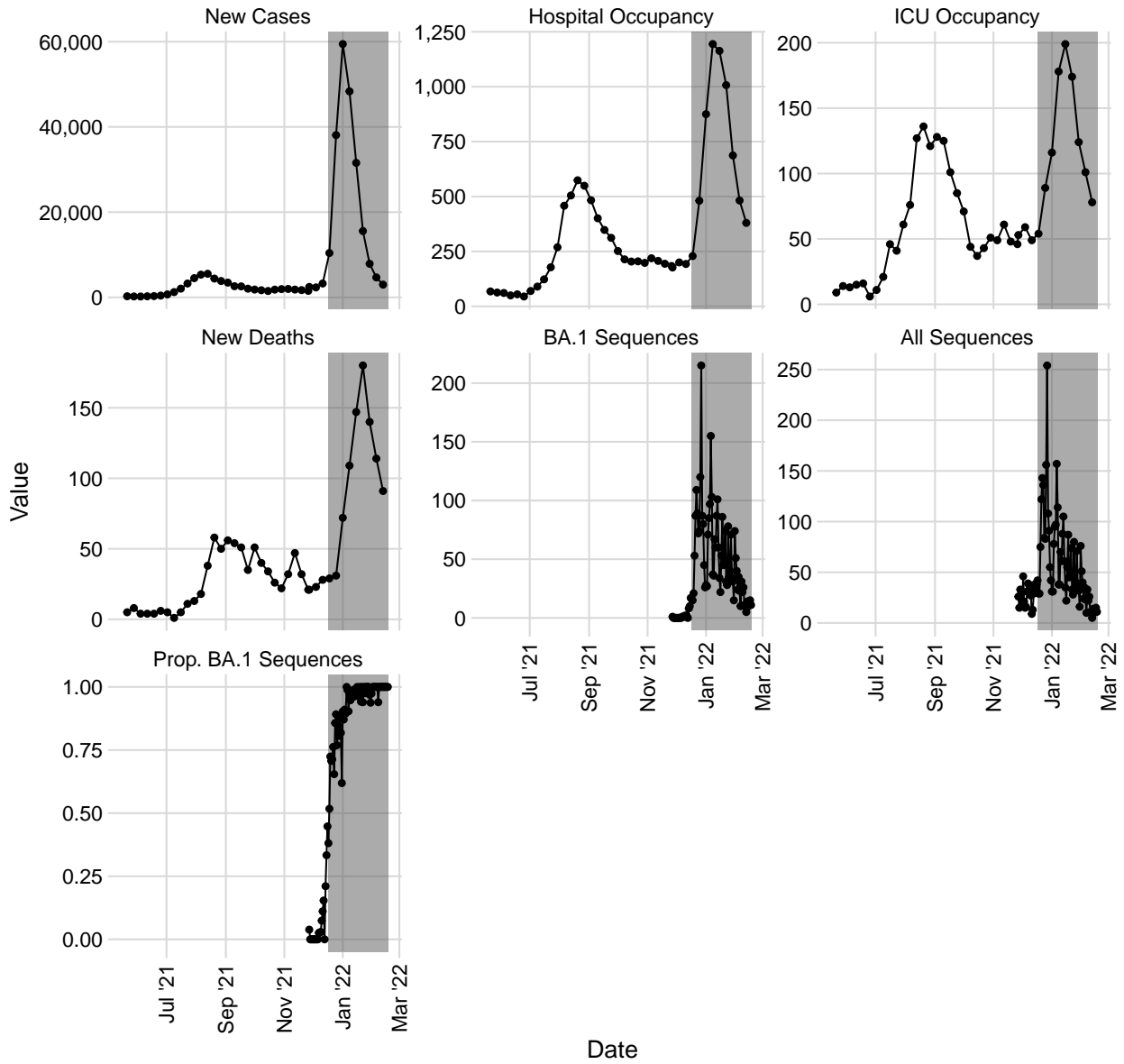
Figure 5.12: COVID-19 surveillance data from Orange County, CA. The plots show weekly counts of cases, hospital and ICU occupancy of patients with COVID-19, reported deaths due to COVID-19, as well as counts of virus sequences for Omicron BA.1 and all lineages, and the proportion of BA.1 lineages. The gray highlighted regions indicate the times for which we produce forecasts.

The priors used for our models are presented in Table C.4. For each data set, we used four Markov chains run in parallel to draw a total of 1000 posterior samples. As in the simulation study, in the main text, we focus on results related to hospital occupancy, but present results

for cases, ICU occupancy, and deaths in Section C.4. In general, these results exhibit the similar patterns as observed in the hospitalization results.

Figures 5.13 and 5.14 show the 1, 2, and 4-week ahead forecasts for hospital occupancy in Orange County and California, respectively. For both data sets, we note that one-week ahead forecasts match the data quite closely for the models that do not use the genetic data, but the model where $1/\kappa(t)$ is informed by the genetic data exhibits comparatively wider credible intervals. As the forecast horizon increases, the two models where $1/\kappa(t)$ varies in time can correctly identify the peak of the hospital occupancy, while the model where $1/\kappa(t)$ is constant greatly overestimates the peak. These observations are reflected in Figure 5.15, where we observe that the model where $1/\kappa(t)$ is informed by genetic data achieves the lowest average CRPS values at forecast horizons greater than one week, and the model where $1/\kappa(t)$ is constant achieves the highest CRPS values. Analogous figures for the other data streams are presented in Figures C.17–C.19. We show the individual scores for each forecast in Figures C.20–C.23. In Figures 5.16 and 5.17, we present posterior predictive intervals for the peak timing and peak value for hospital occupancy. One again, the models where $1/\kappa(t)$ varies in time are shown to be superior, with narrower credible intervals near the observed value in the data. In particular, the model where $1/\kappa(t)$ is informed by genetic data appears to be best at forecasting the time of the hospital occupancy peak. This is demonstrated in Figure 5.18, which shows a summary of the average CRPS values for forecasts of peak hospital occupancy time and amount. Individual CRPS values for each forecast of a peak hospital occupancy time and amount are presented in Figure C.24.
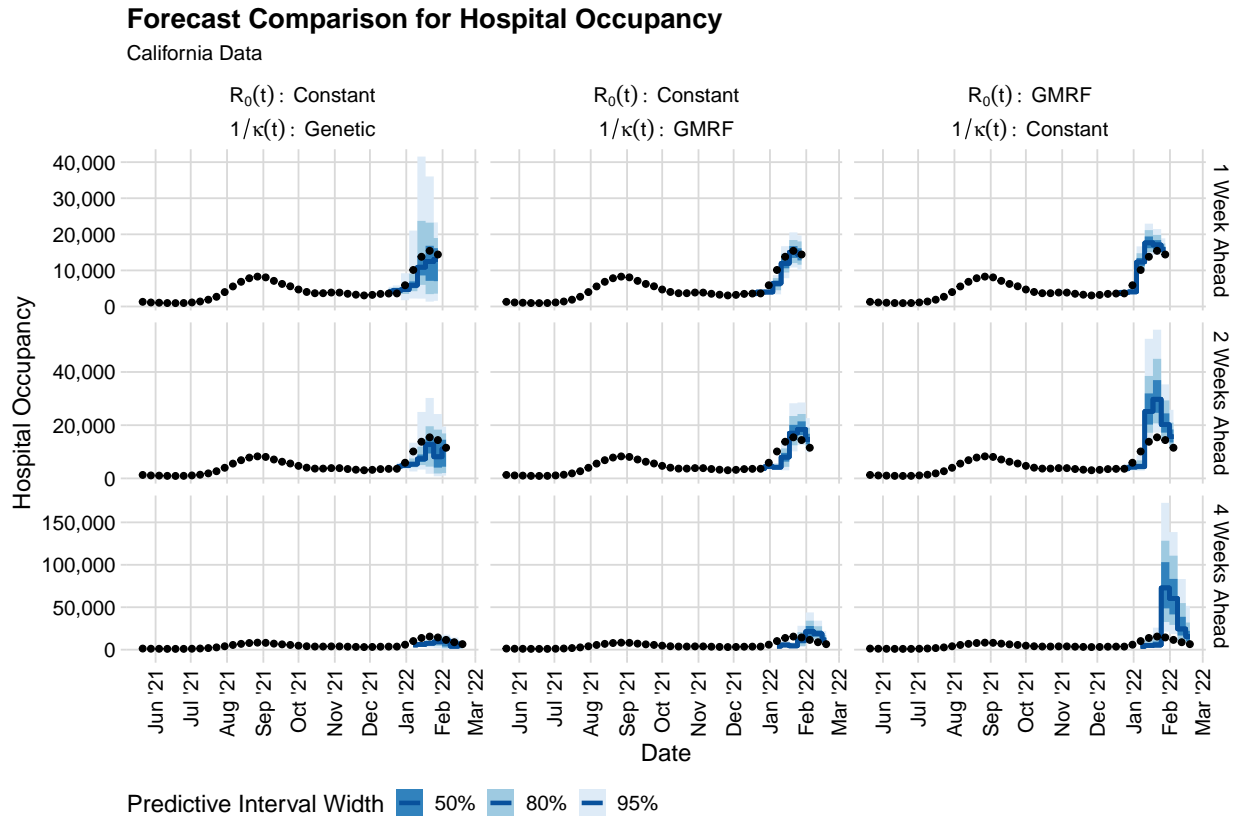
Figure 5.13: Hospital occupancy forecasts from three models at 1, 2, and 4-week forecast horizons for the California data.
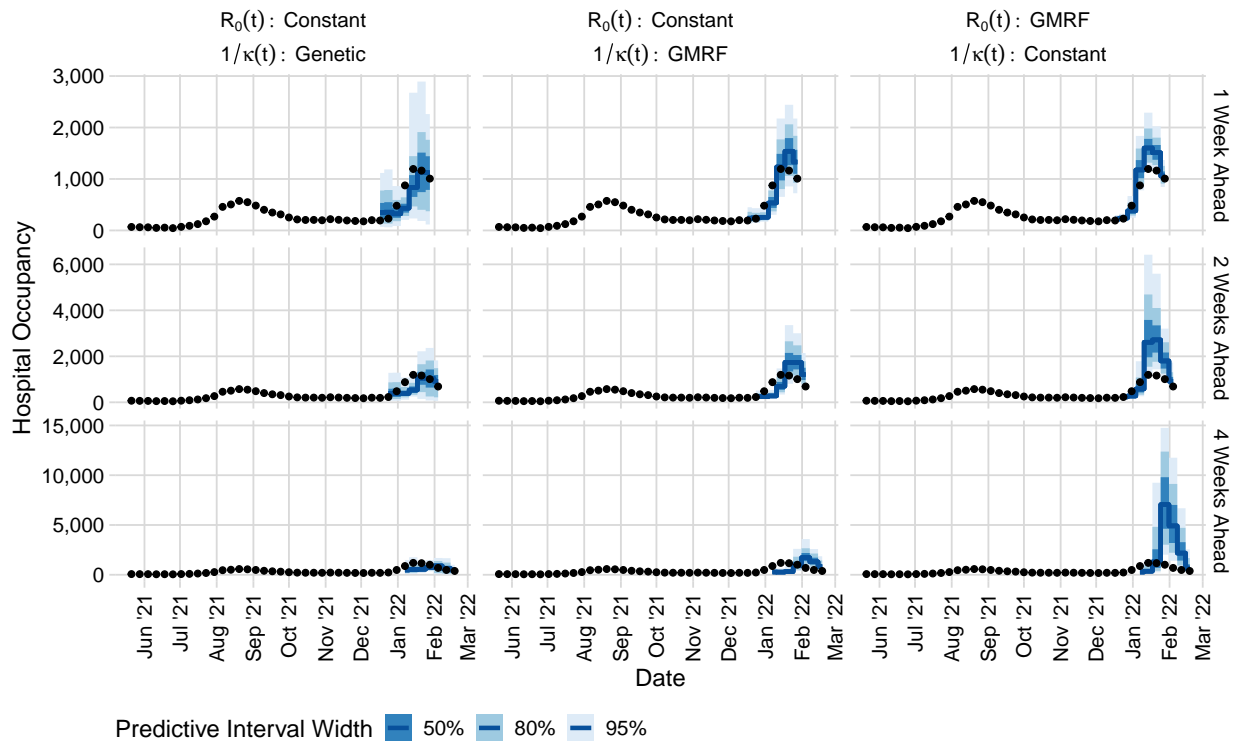
Figure 5.14: Hospital occupancy forecasts from three models at 1, 2, and 4-week forecast horizons for the Orange County data.
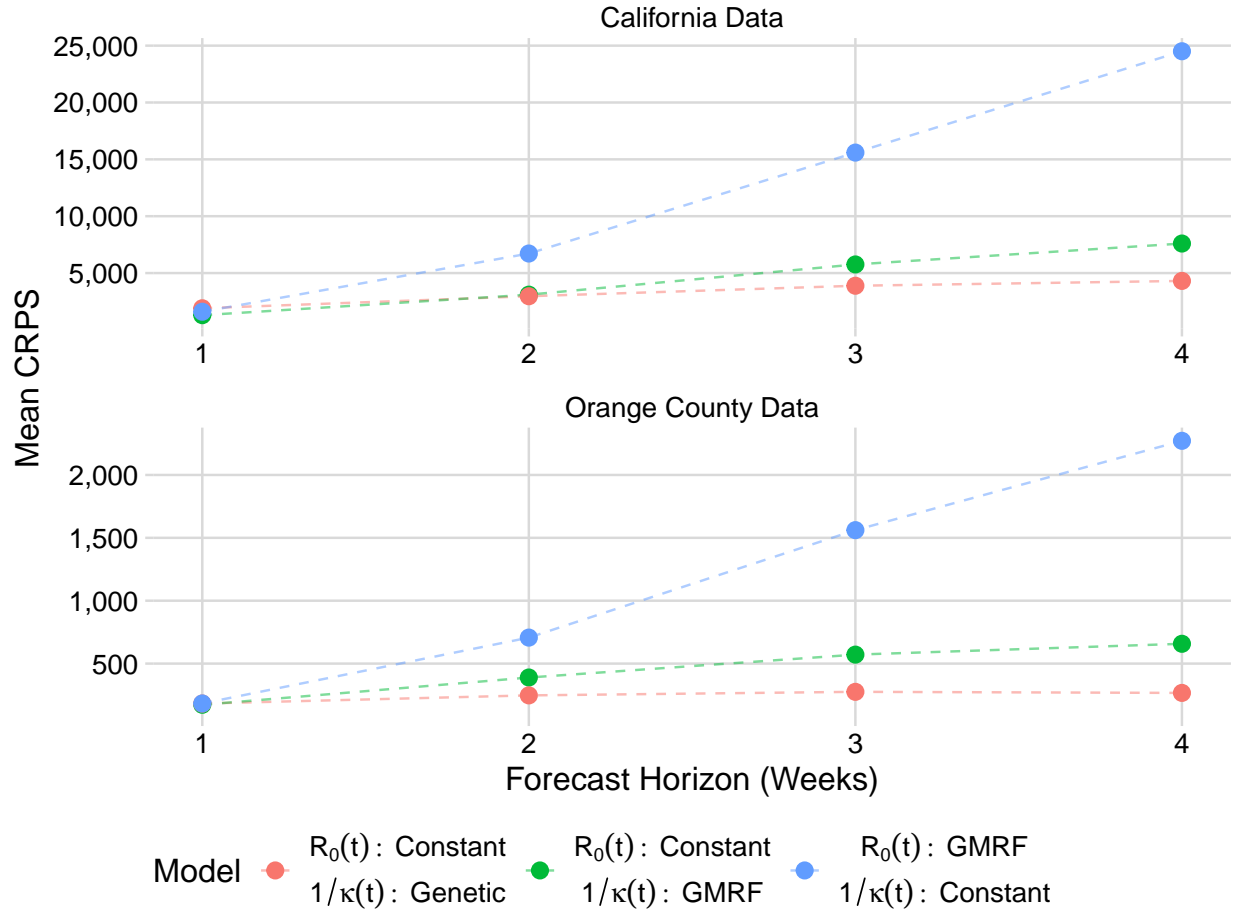
Figure 5.15: CRPS summaries for hospital occupancy forecasts at 1, 2, and 4-week horizons for California and Orange County data sets. Lower CRPS is better.
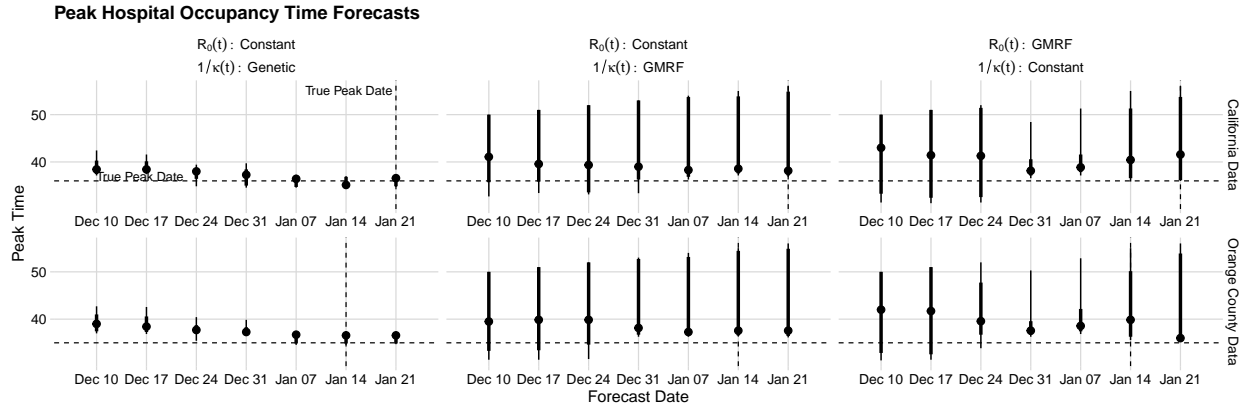
Figure 5.16: Posterior predictive intervals for the time at which hospital occupancy reaches its maximum in California and Orange County data sets. Dots indicate the median of the predictive distribution, while the thick and thin lines represent central 80% and 95% intervals, respectively. Horizontal and vertical dashed lines indicate the true peak hospitalization time.
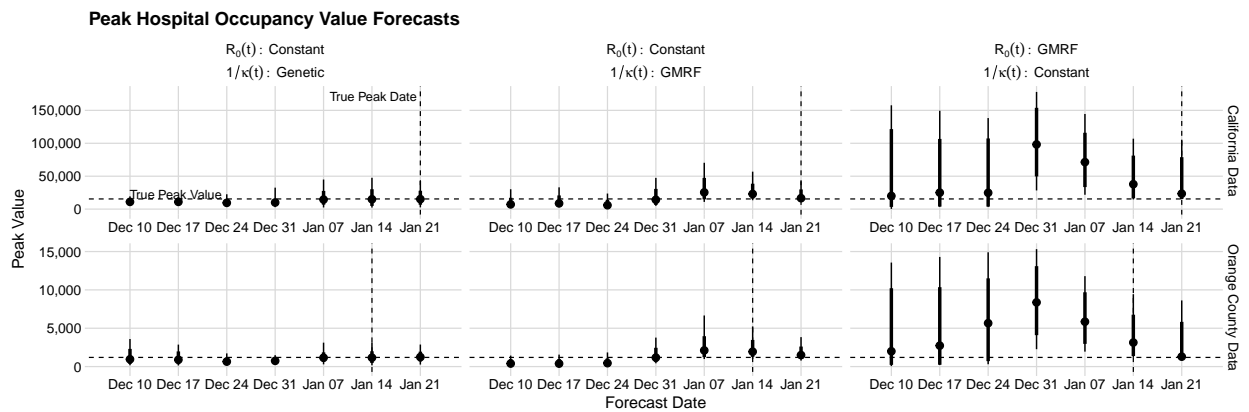
Figure 5.17: Posterior predictive intervals for the maximum hospital occupancy in Orange County and California data sets. Dots indicate the median of the predictive distribution, while the thick and thin lines represent central 80% and 95% intervals, respectively. Horizontal dashed lines indicate the true peak hospital occupancy, while the vertical dashed lines indicate the true peak hospital occupancy time.
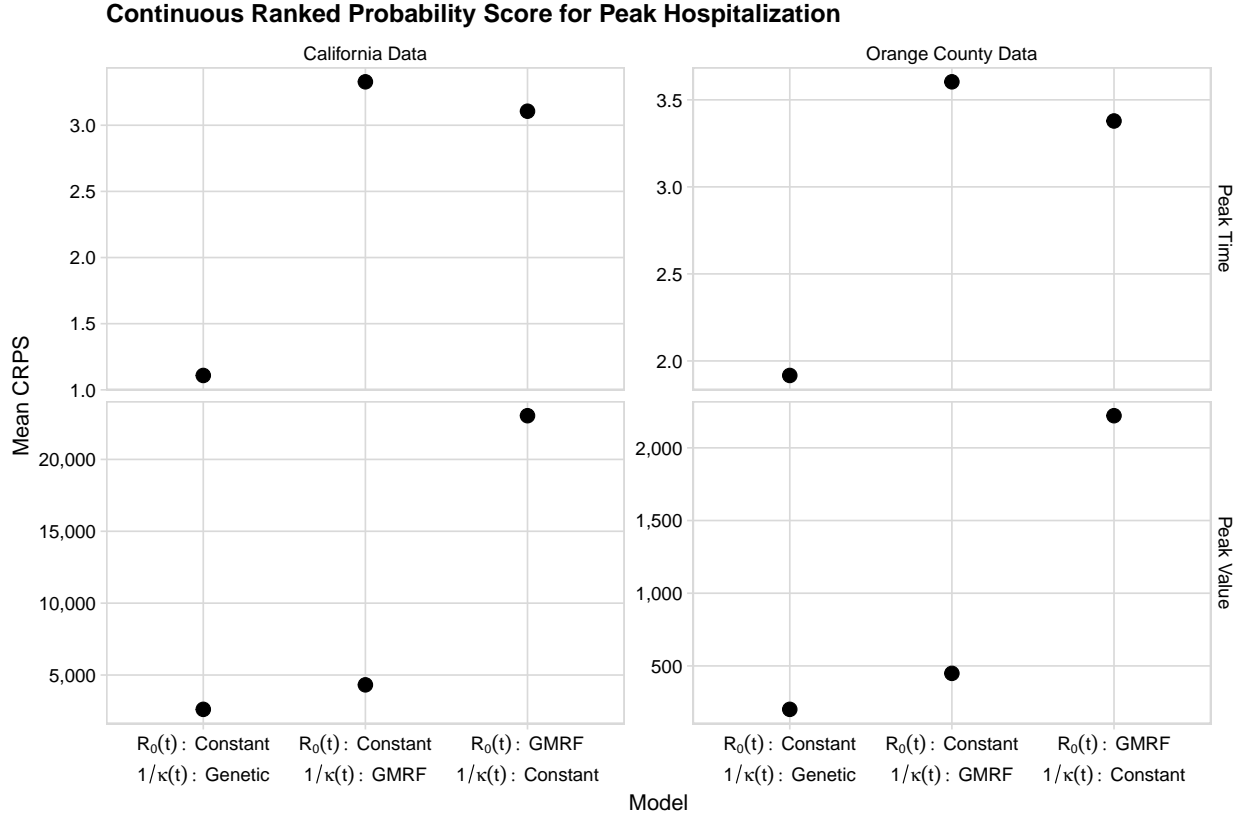
Figure 5.18: CRPS summaries for peak hospital occupancy timing and size in California and Orange County data sets. Lower CRPS is better.

As in the simulation study, the model that uses genetic data is more computationally efficient than the models that use more GMRFs. A summary of the computation time required to fit models in the simulation study is presented in Table 5.3

Table 5.3: Computation time summary for all models in our real data application. Each fit consists of four Markov chains run in parallel to draw a total of 1000 posterior samples.

Model	Avg. CPU Hrs	Min. CPU Hrs	Max. CPU Hrs
$R_0(t)$: Constant, $1/\kappa(t)$: Genetic	4.47	2.00	9.12
$R_0(t)$: Constant, $1/\kappa(t)$: GMRF	16.17	9.10	28.01
$R_0(t)$: GMRF, $1/\kappa(t)$: Constant	30.98	22.44	41.15

5.4   Discussion

We developed a Bayesian model that integrates time series of cases, hospitalizations, ICU admissions, and deaths, as well as genetic data to produce forecasts during time periods where a novel virus variant becomes dominant. Our approach is based on the idea that the introduction of a novel variant leads to increased infections, primarily due to immune escape. As such, we allow the parameter in our model reflecting the average immunity duration to change over time and consider modeling this time-varying parameter both parametrically using genetic data, and non-parametrically. We applied our model to simulated data sets as well as real data from the Omicron wave in Orange County and the state of California, with a particular focus on forecasting hospital occupancy. We compare our models with time-varying immunity duration to a standard model where the basic reproduction number varies in time and is modeled non-parametrically. In simulated data sets, the model where the time-varying immunity duration is informed by genetic data produces superior forecasts when the new variant becomes dominant at a "slow" or "medium" pace (24 or 13 weeks to go from 1% to 99% of sequences, respectively). When the new variant becomes dominant quickly (6 weeks), this model is competitive with the model where the basic reproduction number is modeled non-parametrically, which performs the best. In the real data scenarios, the model where the time-varying immunity duration is informed by genetic data is shown to be especially useful for forecasting the timing and size of peak hospital occupancy, a metric that is of great interest to public health agencies.

Despite these advancements, some potential issues remain unaddressed by our model. First, our model asserts that a novel variant must lead to reduced immunity in the population, and therefore cause a second wave. In reality, a variant may become dominant simply due to genetic drift and not be accompanied by an increase in cases, hospitalizations or deaths. Importantly, we assess our model only on retrospective data and do not account for reporting delays. In a real-time forecasting, it should be possible to obtain up-to-date counts of hospital

and ICU occupancy, but case, death, and sequence counts may be reported with significant delays. Since our likelihood for the new variant sequences is conditional on the total number of sequences, this data does not directly impact our estimates of latent incidence, and it may not be necessary to model these delays for the genetic data. In contrast, the case and death counts do directly influence estimates of latent incidence, so reporting delays of these data streams would likely need to be modeled in a real-world use case. Our model also assumes that genetic sequences from each variant are reported at the same rate. In practice, genetic sequences reporting is likely to be biased. For example, samples for sequencing may come disproportionately from an outbreak in a certain location or for patients who have severe symptoms. Certain lineages may also be able to be identified more easily than others (e.g., the Alpha variant of SARS-CoV-2 [McMillen et al., 2022]). Some effort could be made to correct for this in either the model or the data collection. Finally, it should be possible to use the estimated proportion of the infectious population infected with the novel variant to inform model parameters which might differ between variants (e.g., the infection hospitalization ratio or duration of infectiousness). However, ascertaining which of these parameters may differ between variants may be difficult to do in real-time.

Chapter 6

Discussion and Future Directions

In this thesis, we developed novel statistical methods to combine varied data sources to improve both inference and forecasting using infectious disease surveillance data.

In Chapter 3, we focused on issues of sampling schemes and diagnostic test accuracy when estimating disease prevalence. While there are established methods for estimating disease prevalence with associated confidence intervals for complex surveys with perfect assays and simple random sample surveys with imperfect assays, the case of complex surveys with imperfect assays was relatively unexplored before the development of our methods. Opportunities for future scholarship related to this work could focus on assessment of our methods in settings beyond those considered here. In particular, we would be interested to understand how our methods perform in situations with high prevalence. In addition, our work relied on having fixed weights for our survey response. Extending our methodology to scenarios where weights are estimated and accounting for this uncertainty could be useful. Finally, our methods are shown to be overly conservative in some scenarios. Future work could attempt to address this by replacing our lower or upper confidence distributions with a mixture of the two (as in [Veronese and Melilli, 2015]), or using a mid-p version of the gamma intervals

(as in [Fay and Kim, 2017]).

In Chapter 4, we turned our attention to the temporal dynamics of infectious diseases in the presence of changing policy and behavior. We devised a modeling framework for integrating SARS-CoV-2 diagnostics test and mortality time series data, as well as seroprevalence data from cross-sectional studies, and tested the importance of individual data streams for both inference and forecasting. Importantly, our model for incidence data accounts for changes in the total number of tests performed. This chapter largely functions as an extensive case study, with some efforts to produce generalizable results via various sensitivity analyses and model comparisons. Extensions to this work could focus on establishing further evidence for some of the observations made in our work. For example, we noted that conditioning our case observation model on the number of performed diagnostic tests appeared to be more useful at some points in the outbreak than others. This observation could be more formally assessed in an extensive simulation study or by applying our methodology to similar data from other locations.

In Chapter 5, we worked in a similar setting as Chapter 4, but where the changing disease dynamics were due to novel disease variants, rather than changing policy and behavior. To improve forecasting in these scenarios, we proposed a way to integrate raw counts of variant sequences into a Bayesian compartmental model. We then modeled the average duration of immunity as a flexible function of the proportion of the infectious population infected with the novel variant. A major drawback of our approach is that our model asserts that the dominance of a novel variant must lead to reduced immunity in the population, and therefore cause a new wave of infections. In reality, a variant may become dominant simply due to genetic drift and not be accompanied by an increase in cases, hospitalizations or deaths. Future work could develop a model which allows for this flexibility. Our model also assumes that genetic sequences from each variant are reported at the same rate. In practice, genetic sequences reporting is likely to be biased. Extensions to our work could correct

122

for this in the model. Finally, it should be possible to use the estimated proportion of the infectious population infected with the novel variant to inform model parameters that might differ between variants (e.g., the infection hospitalization ratio or duration of infectiousness). However, ascertaining which of these parameters may differ between variants may be difficult to do in real-time.

Considering Chapters 4 and 5 together, there are common opportunities for extensions. First, the forecasting assessment for these models was performed using retrospective data, not real-time data. This enabled us to sidestep the issue of modeling reporting delays, which could negatively affect the models' forecasting capabilities. The models proposed in these chapters both included a term in their likelihood which was conditioned on some observed data (diagnostic test counts in Chapter 4 and total counts of genetic sequences in Chapter 5). Future work could consider modeling these data streams as well, rather than conditioning on them. Beyond considering alternative ways to model data already used in our work, additional data sources, such as wastewater, mobility, or survey data, could be integrated in these models. The model dynamics could also be expanded to stratify compartments by age or geographical location. Chapter 4 developed a model suitable for forecasting the COVID-19 pandemic from its onset until vaccination and reinfection became common, and Chapter 5 proposed a model for forecasting in the period of novel variants capable of immune escape. The timespan in between, where the majority of vaccines were administered, remains unaddressed and presents opportunities to future infectious disease modelers.

Bibliography

A. Agresti and B. A. Coull. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126, 1998.

R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1992.

S. C. Anderson, A. M. Edwards, M. Yerlanov, N. Mulberry, J. E. Stockdale, S. A. Iyaniwura, R. C. Falcao, M. C. Otterstatter, M. A. Irvine, N. Z. Janjua, D. Coombs, and C. Colijn. Quantifying the impact of COVID-19 control measures using a Bayesian model of physical distancing. *PLOS Computational Biology*, 16(12):1–15, 2020.

C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72:269–342, 2010.

S. G. Baker. The multinomial-Poisson transformation. *Journal of the Royal Statistical Society. Series D, Statistician*, 43(4):495–504, 1994.

D. M. Bayer, M. P. Fay, and B. I. Graubard. Confidence intervals for prevalence estimates from complex surveys with imperfect assays. *Statistics in Medicine*, 42(11):1822–1867, 2023.

R. L. Berger and D. D. Boos. P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*, 89(427):1012–1016, 1994.

D. Bernoulli and S. Blower. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. *Reviews in Medical Virology*, 14(5):275–288, 2004.

J. Besag. Discussion of the paper by Grenander and Miller. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(4):591–592, 1994.

M. Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv*, 2018.

A. Bhargava, E. A. Fukushima, M. Levine, W. Zhao, F. Tanveer, S. M. Szpunar, and L. Saravolatz. Predictors for severe COVID-19 infection. *Clinical Infectious Diseases*, 71 (8):1962–1968, 2020.

C. R. Blyth. Approximate binomial confidence limits. *Journal of the American Statistical Association*, 81(395):843–855, 1986.

N. I. Bosse, H. Gruson, A. Cori, E. van Leeuwen, S. Funk, and S. Abbott. Evaluating forecasts with scoringutils in R. *arXiv*, 2022.

J. Bracher, E. L. Ray, T. Gneiting, and N. G. Reich. Evaluating epidemic forecasts in an interval format. *PLOS Computational Biology*, 17(2):1–15, Feb. 2021.

F. Brauer, P. van den Driessche, and J. Wu. *Compartmental Models in Epidemiology*, pages 19–79. Springer, Berlin, Heidelberg, 2008. ISBN 978-3-540-78911-6.

C. Bretó and E. Ionides. Compound Markov counting processes and their applications to modeling infinitesimally over–dispersed systems. *Stochastic Processes and their Applications*, 121:2571–2591, 2011.

C. Bretó, D. He, E. Ionides, and A. King. Time series analysis via mechanistic models. *The Annals of Applied Statistics*, 3:319–348, 2009.

T. A. Bruckner, D. M. Parker, S. M. Bartell, V. M. Vieira, S. Khan, A. Noymer, E. Drum, B. Albala, M. Zahn, and B. Boden-Albala. Estimated seroprevalence of SARS-CoV-2 antibodies among adults in Orange County, California. *Scientific Reports*, 11(1), 2021.

A. W. Byrne, D. McEvoy, A. B. Collins, K. Hunt, M. Casey, A. Barber, F. Butler, J. Griffin, E. A. Lane, C. McAloon, K. O'Brien, P. Wall, K. A. Walsh, and S. J. More. 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), 2020.

P.-C. Bürkner, J. Gabry, M. Kay, and A. Vehtari. posterior: Tools for working with posterior distributions, 2022. URL `https://mc-stan.org/posterior/`. R package version 1.3.1.

B. Cai, J. P. A. Ioannidis, E. Bendavid, and L. Tian. Exact inference for disease prevalence based on a test with unknown specificity and sensitivity. *Journal of Applied Statistics*, pages 1–25, 2022.

S. Cauchemez and N. Ferguson. Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. *Journal of the Royal Society Interface*, 5:885–897, 2008.

C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, Dec. 1934.

A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, Sept. 2013.

E. Y. Cramer, E. L. Ray, V. K. Lopez, J. Bracher, A. Brennen, A. J. Castro Rivadeneira, A. Gerding, T. Gneiting, K. H. House, Y. Huang, D. Jayawardena, A. H. Kanji, A. Khandelwal, K. Le, A. Mühlemann, J. Niemi, A. Shah, A. Stark, Y. Wang, N. Wattanachit, M. W. Zorn, Y. Gu, S. Jain, N. Bannur, A. Deva, M. Kulkarni, S. Merugu, A. Raval, S. Shingi, A. Tiwari, J. White, N. F. Abernethy, S. Woody, M. Dahan, S. Fox, K. Gaither, M. Lachmann, L. A. Meyers, J. G. Scott, M. Tec, A. Srivastava, G. E. George, J. C. Cegan, I. D. Dettwiller, W. P. England, M. W. Farthing, R. H. Hunter, B. Lafferty, I. Linkov, M. L. Mayo, M. D. Parno, M. A. Rowland, B. D. Trump, Y. Zhang-James, S. Chen, S. V. Faraone, J. Hess, C. P. Morley, A. Salekin, D. Wang, S. M. Corsetti, T. M. Baer, M. C. Eisenberg, K. Falb, Y. Huang, E. T. Martin, E. McCauley, R. L. Myers, T. Schwarz, D. Sheldon, G. C. Gibson, R. Yu, L. Gao, Y. Ma, D. Wu, X. Yan, X. Jin, Y.-X. Wang, Y. Chen, L. Guo, Y. Zhao, Q. Gu, J. Chen, L. Wang, P. Xu, W. Zhang, D. Zou, H. Biegel, J. Lega, S. McConnell, V. P. Nagraj, S. L. Guertin, C. Hulme-Lowe, S. D. Turner, Y. Shi, X. Ban, R. Walraven, Q.-J. Hong, S. Kong, A. van de Walle, J. A. Turtle, M. Ben-Nun, S. Riley, P. Riley, U. Koyluoglu, D. DesRoches, P. Forli, B. Hamory, C. Kyriakides, H. Leis, J. Milliken, M. Moloney, J. Morgan, N. Nirgudkar, G. Ozcan, N. Piwonka, M. Ravi, C. Schrader, E. Shakhnovich, D. Siegel, R. Spatz, C. Stiefeling, B. Wilkinson, A. Wong, S. Cavany, G. España, S. Moore, R. Oidtman, A. Perkins, D. Kraus, A. Kraus, Z. Gao, J. Bian, W. Cao, J. Lavista Ferres, C. Li, T.-Y. Liu, X. Xie, S. Zhang, S. Zheng, A. Vespignani, M. Chinazzi, J. T. Davis, K. Mu, A. Pastore Y Piontti, X. Xiong, A. Zheng, J. Baek, V. Farias, A. Georgescu, R. Levi, D. Sinha, J. Wilde, G. Perakis, M. A. Bennouna, D. Nze-Ndong, D. Singhvi, I. Spantidakis, L. Thayaparan, A. Tsiourvas, A. Sarker, A. Jadbabaie, D. Shah, N. Della Penna, L. A. Celi, S. Sundar, R. Wolfinger, D. Osthus, L. Castro, G. Fairchild, I. Michaud, D. Karlen, M. Kinsey, L. C. Mullany, K. Rainwater-Lovett, L. Shin, K. Tallaksen, S. Wilson, E. C. Lee, J. Dent, K. H. Grantz, A. L. Hill, J. Kaminsky, K. Kaminsky, L. T. Keegan, S. A. Lauer, J. C. Lemaitre, J. Lessler, H. R. Meredith, J. Perez-Saez, S. Shah, C. P. Smith, S. A. Truelove, J. Wills, M. Marshall, L. Gardner, K. Nixon, J. C. Burant, L. Wang, L. Gao, Z. Gu, M. Kim, X. Li, G. Wang, Y. Wang, S. Yu, R. C. Reiner, R. Barber, E. Gakidou, S. I. Hay, S. Lim, C. Murray, D. Pigott, H. L. Gurung, P. Baccam, S. A. Stage, B. T. Suchoski, B. A. Prakash, B. Adhikari, J. Cui, A. Rodríguez, A. Tabassum, J. Xie, P. Keskinocak, J. Asplund, A. Baxter, B. E. Oruc, N. Serban, S. O. Arik, M. Dusenberry, A. Epshteyn, E. Kanal, L. T. Le, C.-L. Li, T. Pfister, D. Sava, R. Sinha, T. Tsai, N. Yoder, J. Yoon, L. Zhang, S. Abbott, N. I. Bosse, S. Funk, J. Hellewell, S. R. Meakin, K. Sherratt, M. Zhou, R. Kalantari, T. K. Yamana, S. Pei, J. Shaman, M. L. Li, D. Bertsimas, O. Skali Lami, S. Soni, H. Tazi Bouardi, T. Ayer, M. Adee, J. Chhatwal, O. O. Dalgic, M. A. Ladd, B. P. Linas, P. Mueller, J. Xiao, Y. Wang, Q. Wang, S. Xie, D. Zeng, A. Green, J. Bien, L. Brooks, A. J. Hu, M. Jahja, D. McDonald, B. Narasimhan, C. Politsch, S. Rajanala, A. Rumack, N. Simon, R. J. Tibshirani, R. Tibshirani, V. Ventura, L. Wasserman, E. B. O'Dea, J. M. Drake, R. Pagano, Q. T. Tran, L. S. T. Ho, H. Huynh, J. W. Walker, R. B. Slayton, M. A. Johansson, M. Biggerstaff, and N. G. Reich. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proc. Natl. Acad. Sci. U. S. A.*, 119(15):e2113561119, Apr. 2022.

N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, W. J. Edmunds, T. Jombart, K. O'Reilly, A. Endo, J. Hellewell, E. S. Nightingale, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, 5(7):e375–e385, 2020.

N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, A. J. Kucharski, J. D. Munday, C. A. B. Pearson, T. W. Russell, D. C. Tully, A. D. Washburne, T. Wenseleers, A. Gimma, W. Waites, K. L. M. Wong, K. van Zandvoort, J. D. Silverman, C. C.-. W. Group, C.-. G. U. C.-U. Consortium, K. Diaz-Ordaz, R. Keogh, R. M. Eggo, S. Funk, M. Jit, K. E. Atkins, and W. J. Edmunds. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538), 2021.

D. De Angelis, A. M. Presanis, P. J. Birrell, G. S. Tomba, and T. House. Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics*, 10:83–87, 2015.

N. Dean and M. Pagano. Evaluating confidence interval methods for binomial proportions in clustered surveys. *Journal of Survey Statistics and Methodology*, 3(4):484–503, 2015.

T. J. DiCiccio, D. M. Ritzwoller, J. P. Romano, and A. M. Shaikh. Confidence intervals for seroprevalence. *Statistical Science*, 37:306–321, 2022.

H. Du, E. Dong, H. S. Badr, M. E. Petrone, N. D. Grubaugh, and L. M. Gardner. Incorporating variant frequencies data into short-term forecasting for COVID-19 cases and deaths in the USA: a deep learning approach. *eBioMedicine*, 89, 2023.

S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.

V. Dukic, H. Lopes, and N. Polson. Tracking epidemics with Google flu trends data and a state-space SEIR model. *Journal of the American Statistical Association*, 107:1410–1426, 2012.

M. P. Fay. *asht: Applied Statistical Hypothesis Tests*, 2020. URL `https://CRAN.R-project.org/package=asht`. R package version 0.9.6.

M. P. Fay and E. J. Feuer. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Statistics in Medicine*, 16(7):791–801, 1997.

M. P. Fay and S. Kim. Confidence intervals for directly standardized rates using mid-p gamma intervals. *Biometrical Journal*, 59(2):377–387, 2017.

M. P. Fay, M. A. Proschan, and E. Brittain. Combining one-sample confidence procedures for inference in the two-sample case. *Biometrics*, 71(1):146–156, 2015.

M. P. Fay, M. A. Proschan, E. Brittain, and R. Tiwari. Interpreting p-values and confidence intervals using well-calibrated null preference priors. *Statistical Science*, 37: 455–472, 2022.

N. Ferguson and et. al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *MRC Centre for Global Infectious Disease Analysis Reports*, 2020. Accessed: 2020-06-19.

M. D. Figgins and T. Bedford. SARS-CoV-2 variant dynamics across US states show consistent differences in effective reproduction numbers. *medRxiv*, 2022.

J. Fintzi, J. Wakefield, and V. N. Minin. A linear noise approximation for stochastic epidemic models fit to partially observed incidence counts. *Biometrics*, 78(4):1530–1541, 2022.

C. Franco, R. J. Little, T. A. Louis, and E. V. Slud. Comparative study of confidence intervals for proportions in complex sample surveys. *Journal of Survey Statistics and Methodology*, 7(3):334–364, 2019.

C. Fuchs. *Inference for Diffusion Processes: with Applications in Life Sciences.* Springer Berlin, Heidelberg, 2013. ISBN 978-3-642-25969-2.

K. Gangavarapu, A. A. Latif, J. L. Mullen, M. Alkuzweny, E. Hufbauer, G. Tsueng, E. Haag, M. Zeller, C. M. Aceves, K. Zaiets, M. Cano, X. Zhou, Z. Qian, R. Sattler, N. L. Matteson, J. I. Levy, R. T. C. Lee, L. Freitas, S. Maurer-Stroh, M. A. Suchard, C. Wu, A. I. Su, K. G. Andersen, L. D. Hughes, G. I. S. A. I. D. Core, and C. Team. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Nature Methods*, 20(4):512–522, Apr. 2023.

H. Ge, K. Xu, and Z. Ghahramani. Turing: A language for flexible probabilistic inference. In A. Storkey and F. Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1682–1690. PMLR, Apr. 2018.

A. Gelman and B. Carpenter. Bayesian analysis of tests with unknown specificity and sensitivity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(5): 1269–1283, 2020.

G. C. Gibson, N. G. Reich, and D. Sheldon. Real-time mechanistic Bayesian forecasts of COVID-19 mortality. *medRxiv*, 2020.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69 (2):243–268, 2007.

T. Gneiting, D. Wolffram, J. Resin, K. Kraus, J. Bracher, T. Dimitriadis, V. Hagenmeyer, A. I. Jordan, S. Lerch, K. Phipps, and M. Schienle. Model diagnostics and forecast evaluation for quantiles. *Annual Review of Statistics and Its Application*, 10(1):597–621, 2023.

P. E. Greenwood and L. F. Gordillo. *Stochastic Epidemic Modeling*, pages 31–52. Springer Netherlands, Dordrecht, 2009. ISBN 978-90-481-2313-1.

D. J. Grint, K. Wing, C. Houlihan, H. P. Gibbs, S. J. W. Evans, E. Williamson, H. I. McDonald, K. Bhaskaran, D. Evans, A. J. Walker, G. Hickman, E. Nightingale, A. Schultze, C. T. Rentsch, C. Bates, J. Cockburn, H. J. Curtis, C. E. Morton, S. Bacon, S. Davy, A. Y. S. Wong, A. Mehrkar, L. Tomlinson, I. J. Douglas, R. Mathur, B. MacKenna, P. Ingelsby, R. Croker, J. Parry, F. Hester, S. Harper, N. J. DeVito, W. Hulme, J. Tazare, L. Smeeth, B. Goldacre, and R. M. Eggo. Severity of Severe Acute Respiratory System Coronavirus 2 (SARS-CoV-2) Alpha Variant (B.1.1.7) in England. *Clinical Infectious Diseases*, 75(1):e1120–e1127, Sept. 2021.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, Apr. 1970.

L. Held, N. Hens, P. D O'Neill, and J. Wallinga. *Handbook of Infectious Disease Data Analysis*. CRC Press, 2019.

D. Hemenway. The myth of millions of annual self-defense gun uses: a case study of survey overestimates of rare events. *CHANCE*, 10(3):6–10, 1997.

M. D. Hoffman and A. Gelman. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014.

M. Höhle and M. an der Heiden. Bayesian nowcasting during the STEC O104: H4 outbreak in Germany, 2011. *Biometrics*, 70:993–1002, 2014.

R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

N. J. Irons and A. E. Raftery. Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proceedings of the National Academy of Sciences*, 118 (31):e2103272118, 2021.

S. A. Iyaniwura, R. C. Falcão, N. Ringa, P. A. Adu, M. Spencer, M. Taylor, C. Colijn, D. Coombs, N. Z. Janjua, M. A. Irvine, and M. Otterstatter. Mathematical modeling of COVID-19 in British Columbia: An age-structured model with time-dependent contact rates. *Epidemics*, 39, 2022.

S. Jewell, J. Futoma, L. Hannah, A. C. Miller, N. J. Foti, and E. B. Fox. It's complicated: characterizing the time-varying relationship between cell phone mobility and COVID-19 spread in the US. *npj Digital Medicine*, 4(1):152, 2021.

A. Jordan, F. Krüger, and S. Lerch. Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software*, 90(12):1–37, 2019.

H. Kalish, C. Klumpp-Thomas, S. Hunsberger, H. A. Baus, M. P. Fay, N. Siripong, J. Wang, J. Hicks, J. Mehalko, J. Travers, M. Drew, K. Pauly, J. Spathies, T. Ngo, K. M. Adusei, M. Karkanitsa, J. A. Croker, Y. Li, B. I. Graubard, L. Czajkowski, O. Belliveau, C. Chairez, K. R. Snead, P. Frank, A. Shunmugavel, A. Han, L. T. Giurgea, L. A. Rosas, R. Bean, R. Athota, A. Cervantes-Medina, M. Gouzoulis, B. Heffelfinger, S. Valenti, R. Caldararo, M. M. Kolberg, A. Kelly, R. Simon, S. Shafiq, V. Wall, S. Reed, E. W. Ford, R. Lokwani, J.-P. Denson, S. Messing, S. G. Michael, W. Gillette, R. P. Kimberly, S. E. Reis, M. D. Hall, D. Esposito, M. J. Memoli, and K. Sadtler. Undiagnosed SARS-CoV-2 seropositivity during the first 6 months of the COVID-19 pandemic in the united states. *Science Translational Medicine*, 13(601), 2021.

M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals.* Princeton University Press, 2011.

L. Kim, S. Garg, A. O?Halloran, M. Whitaker, H. Pham, E. J. Anderson, I. Armistead, N. M. Bennett, L. Billing, K. Como-Sabetti, et al. Risk factors for intensive care unit admission and in-hospital mortality among hospitalized adults identified through the US coronavirus disease 2019 (COVID-19)-associated hospitalization surveillance network (COVID-NET). *Clinical Infectious Diseases*, 72(9):e206–e214, 2021.

A. A. King, M. Domenech de Cellès, F. M. G. Magpantay, and P. Rohani. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society B: Biological Sciences*, 282(1806), 2015.

E. S. Knock, L. K. Whittles, J. A. Lees, P. N. Perez-Guzman, R. Verity, R. G. FitzJohn, K. A. M. Gaythorpe, N. Imai, W. Hinsley, L. C. Okell, A. Rosello, N. Kantas, C. E. Walters, S. Bhatia, O. J. Watson, C. Whittaker, L. Cattarino, A. Boonyasiri, B. A. Djaafara, K. Fraser, H. Fu, H. Wang, X. Xi, C. A. Donnelly, E. Jauneikaite, D. J. Laydon, P. J. White, A. C. Ghani, N. M. Ferguson, A. Cori, and M. Baguelin. Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Science Translational Medicine*, 13(602):eabg4262, 2021.

E. L. Korn and B. I. Graubard. Confidence intervals for proportions with small expected number of positive counts estimated from survey data. *Survey Methodology*, 24:193–201, 1998.

E. L. Korn and B. I. Graubard. *Analysis of Health Surveys.* John Wiley & Sons, 1999.

F. Krüger, S. Lerch, T. Thorarinsdottir, and T. Gneiting. Predictive inference based on Markov chain Monte Carlo output. *International Statistical Review*, 89(2):274–301, 2021.

A. J. Kucharski, V. Andreasen, and J. R. Gog. Capturing the dynamics of pathogens with many strains. *Journal of Mathematical Biology*, 72(1):1–24, 2016.

M. Lacerda and C. Seoighe. Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics*, 198(3):1237–1250, Sept. 2014.

Z. Lang and J. Reiczigel. Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. *Preventive Veterinary Medicine*, 113(1):13–22, 2014.

P. Lekone and B. Finkenstädt. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 62:1170–1177, 2006.

J. Lessler and D. A. T. Cummings. Mechanistic models of infectious disease and their impact on public health. *American Journal of Epidemiology*, 183(5):415–422, Feb. 2016.

J. Li and F. Brauer. Continuous-time age-structured models in population dynamics and epidemiology. In F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, chapter 9, pages 205–227. Springer, 2008.

J. E. Matheson and R. L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, 1976.

T. McMillen, K. Jani, E. V. Robilotti, M. Kamboj, and N. E. Babady. The spike gene target failure (SGTF) genomic signature is highly accurate for the identification of Alpha and Omicron SARS-CoV-2 variants. *Sci. Rep.*, 12(1):18968, Nov. 2022.

O. Morozova, Z. R. Li, and F. W. Crawford. One year of modeling and forecasting COVID-19 transmission to support policymakers in Connecticut. *Scientific Reports*, 13: 20271, 2021.

I. Murray, R. Adams, and D. MacKay. Elliptical slice sampling. In Y. W. Teh and M. Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 541–548, Chia Laguna Resort, Sardinia, Italy, May 2010. PMLR.

R. M. Neal. *MCMC Using Hamiltonian Dynamics*, pages 113–162. Chapman and Hall/CRC, 2011. ISBN 9780429138508.

B. Nguyen-Van-Yen, P. Del Moral, and B. Cazelles. Stochastic epidemic models inference and diagnosis with Poisson random measure data augmentation. *Mathematical Biosciences*, 335:108583, 2021.

K. Nixon, S. Jindal, F. Parker, N. G. Reich, K. Ghobadi, E. C. Lee, S. Truelove, and L. Gardner. An evaluation of prospective COVID-19 modelling studies in the USA: from data to science translation. *The Lancet Digital Health*, 4(10):e738–e747, 2022.

E. B. O'Dea and J. M. Drake. A semi-parametric, state-space compartmental model with time-dependent parameters for forecasting COVID-19 cases, hospitalizations and deaths. *Journal of The Royal Society Interface*, 19(187):20210702, 2022.

S. Pei, T. K. Yamana, S. Kandula, M. Galanti, and J. Shaman. Burden and characteristics of COVID-19 in the United States during 2020. *Nature*, 598(7880):338–341, 2021.

C. M. Petrilli, S. A. Jones, J. Yang, H. Rajagopalan, L. O'Donnell, Y. Chernyak, K. A. Tobin, R. J. Cerfolio, F. Francois, and L. I. Horwitz. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *British Medical Journal*, 369, 2020.

S. Pollett, M. A. Johansson, N. G. Reich, D. Brett-Major, S. Y. Del Valle, S. Venkatramanan, R. Lowe, T. Porco, I. M. Berry, A. Deshpande, M. U. G. Kraemer, D. L. Blazes, W. Pan-ngum, A. Vespigiani, S. E. Mate, S. P. Silal, S. Kandula, R. Sippy, T. M. Quandelacy, J. J. Morgan, J. Ball, L. C. Morton, B. M. Althouse, J. Pavlin, W. van Panhuis, S. Riley, M. Biggerstaff, C. Viboud, O. Brady, and C. Rivers. Recommended reporting items for epidemic forecasting and prediction research: The epiforge 2020 guidelines. *PLOS Medicine*, 18(10):1–12, Oct. 2021.

C. Pooley, S. Bishop, and G. Marion. Using model-based proposals for fast parameter inference on discrete state space, continuous-time Markov processes. *Journal of The Royal Society Interface*, 12:20150225, 2015.

K. Prem and 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:e261–e270, 2020.

A. Rambaut, E. C. Holmes, Á. O'Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, and O. G. Pybus. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5(11):1403–1407, 2020.

E. A. Rashed, S. Kodera, and A. Hirata. COVID-19 forecasting using new viral variants and vaccination effectiveness models. *Computers in Biology and Medicine*, 149:105986, 2022.

D. A. Rasmussen, O. Ratmann, and K. Koelle. Inference for nonlinear epidemiological models using genealogies and time series. *PLOS Computational Biology*, 7:1–11, Aug. 2011.

W. J. Rogan and B. Gladen. Estimating prevalence from the results of a screening test. *American Journal of Epidemiology*, 107(1):71–76, 1978.

S. P. Rosin, B. E. Shook-Sa, S. R. Cole, and M. G. Hudgens. Estimating SARS-CoV-2 seroprevalence. *Journal of the Royal Statistical Society Series A: Statistics in Society*, page qnad068, May 2023.

J. A. Scott, A. Gandy, S. Mishra, J. Unwin, S. Flaxman, and S. Bhatt. epidemia: Modeling of epidemics using hierarchical Bayesian models, 2020. URL `https://imperialcollegelondon.github.io/epidemia/`. R package version 1.0.0.

Y. Shu and J. McCauley. Gisaid: Global initiative on sharing all influenza data–from vision to reality. *Eurosurveillance*, 22(13):30494, 2017.

J.-W. Song, C. Zhang, X. Fan, F.-P. Meng, Z. Xu, P. Xia, W.-J. Cao, T. Yang, X.-P. Dai, S.-Y. Wang, et al. Immunological and inflammatory profiles in mild and severe cases of COVID-19. *Nature Communications*, 11(1):1–10, 2020.

Stan Development Team. RStan: the R interface to Stan, 2020. URL `http://mc-stan.org/`. R package version 2.21.2.

A. C. Stokes, D. J. Lundberg, I. T. Elo, K. Hempstead, J. Bor, and S. H. Preston. COVID-19 and excess mortality in the United States: A county-level analysis. *PLOS Medicine*, 18(5):1–18, May 2021.

O. Stoner and T. Economou. Multivariate hierarchical frameworks for modeling delayed reporting in count data. *Biometrics*, 76(3):789–798, 2020.

M. Tang, G. Dudas, T. Bedford, and V. N. Minin. Fitting stochastic epidemic models to gene genealogies using linear noise approximation. *The Annals of Applied Statistics*, 17 (1):1 – 22, 2023.

United States Census Bureau. Quick facts: Orange County, California. `https://www.census.gov/quickfacts/orangecountycalifornia`, 2020. Accessed: 2020-09-05.

P. Van den Driessche. Spatial structure: Patch models. In F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, chapter 7, pages 179–189. Springer, 2008.

A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2):667 – 718, 2021.

P. Veronese and E. Melilli. Fiducial and confidence distributions for real exponential families. *Scandinavian Journal of Statistics*, 42(2):471–484, 2015.

WHO. Word Health Organization coronavirus disease situation report. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019`, 2020. Accessed: 2020-06-17.

WHO. Word Health Organization q&a: Coronavirus disease (COVID-19): How is it transmitted? `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted`, 2021. Accessed: 2021-09-12.

Z. Wu and J. M. McGoogan. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *Journal of the American Medical Association*, 323(13):1239–1242, 2020.

M.-g. Xie and K. Singh. Confidence distribution, the frequentist distribution estimator of a parameter: a review. *International Statistical Review*, 81(1):3–39, 2013.

H. Xin, Y. Li, P. Wu, Z. Li, E. H. Y. Lau, Y. Qin, L. Wang, B. J. Cowling, T. K. Tsang, and Z. Li. Estimating the Latent Period of Coronavirus Disease 2019 (COVID-19). *Clinical Infectious Diseases*, 74(9):1678–1681, Sept. 2021.

X. Xu, T. Kypraios, and P. D. O'Neill. Bayesian non-parametric inference for stochastic epidemic models using Gaussian Processes. *Biostatistics*, 17(4):619–633, Mar. 2018.

S. Zhao, I. Hu, J. Lou, M. K. Chong, L. Cao, D. He, B. C. Zee, and M. H. Wang. The mechanism shaping the logistic growth of mutation proportion in epidemics at population scale. *Infectious Disease Modelling*, 8(1):107–121, 2023.

T. Zhou and Y. Ji. Semiparametric Bayesian inference for the transmission dynamics of COVID-19 with a state-space model. *Contemporary Clinical Trials*, 97:106146, 2020.

J. Zumbrun. Coronavirus slump is worst since great depression. will it be as painful?, May 2020. URL https://www.wsj.com/articles/coronavirus-slump-is-worst-since-great-depression-will-it-be-as-painful-11589115601. Accessed: 2020-09-05.

Appendix A

Additional Material for Chapter 3

A.1 Monotonicity of g

It is clear that g is monotonic within each of its piecewise-defined functions. In the following sections, we consider if monotonicity holds at the change points between piecewise functions.

A.1.1 Monotonicity in $\hat{\phi}_p$

- Case 1: $\hat{\phi}_n < \hat{\theta}_1$. Not monotonic.

$$
g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases}
0 & \hat{\phi}_p < \hat{\phi}_n < \hat{\theta}_1 \\
0 & \hat{\phi}_n = \hat{\phi}_p < \hat{\theta}_1 \\
1 & \hat{\phi}_n < \hat{\phi}_p < \hat{\theta}_1 \\
\frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = 1 & \hat{\phi}_n < \hat{\phi}_p = \hat{\theta}_1 \\
\frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} & \hat{\phi}_n < \hat{\theta}_1 < \hat{\phi}_p
\end{cases}
\tag{A.1}
$$

- Case 2: $\hat{\theta}_1 < \hat{\phi}_n$. Monotonic.

$$
g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases} 0 & \hat{\phi}_p < \hat{\theta}_1 < \hat{\phi}_n \\ 0 & \hat{\theta}_1 = \hat{\phi}_p < \hat{\phi}_n \\ 0 & \hat{\theta}_1 < \hat{\phi}_p < \hat{\phi}_n \\ 0 & \hat{\theta}_1 < \hat{\phi}_p = \hat{\phi}_n \\ 0 & \hat{\theta}_1 < \hat{\phi}_n < \hat{\phi}_p \end{cases} \tag{A.2}
$$

- Case 3: $\hat{\theta}_1 = \hat{\phi}_n$. Monotonic.

$$
g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases} 0 & \hat{\phi}_p < \hat{\theta}_1 = \hat{\phi}_n \\ \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = \frac{0}{0} \equiv 0 & \hat{\phi}_p = \hat{\theta}_1 = \hat{\phi}_n \\ \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = \frac{0}{\hat{\phi}_p - \hat{\phi}_n} = 0 & \hat{\theta}_1 = \hat{\phi}_n < \hat{\phi}_p \end{cases} \tag{A.3}
$$

A.1.2 Monotonicity in $\hat{\theta}_1$

- Case 1: $\hat{\phi}_n < \hat{\phi}_p$. Monotonic.

$$
g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases} 0 & \hat{\theta}_1 < \hat{\phi}_n < \hat{\phi}_p \\ \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = 0 & \hat{\phi}_n = \hat{\theta}_1 < \hat{\phi}_p \\ \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} & \hat{\phi}_n < \hat{\theta}_1 < \hat{\phi}_p \\ \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = 1 & \hat{\phi}_n < \hat{\theta}_1 = \hat{\phi}_p \\ 1 & \hat{\phi}_n < \hat{\phi}_p < \hat{\theta}_1 \end{cases} \tag{A.4}
$$

- Case 2: $\hat{\phi}_p < \hat{\phi}_n$. Monotonic.

$$g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases} 0 & \hat{\theta}_1 < \hat{\phi}_p < \hat{\phi}_n \\ 0 & \hat{\phi}_p = \hat{\theta}_1 < \hat{\phi}_n \\ 0 & \hat{\phi}_p < \hat{\theta}_1 < \hat{\phi}_n \\ 0 & \hat{\phi}_p < \hat{\theta}_1 = \hat{\phi}_n \\ 0 & \hat{\phi}_p < \hat{\phi}_n < \hat{\theta}_1 \end{cases} \tag{A.5}$$

- Case 3: $\hat{\phi}_p = \hat{\phi}_n$. Monotonic.

$$g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases} 0 & \hat{\theta}_1 < \hat{\phi}_p = \hat{\phi}_n \\ \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = \frac{0}{0} \equiv 0 & \hat{\theta}_1 = \hat{\phi}_p = \hat{\phi}_n \\ 0 & \hat{\phi}_p = \hat{\phi}_n < \hat{\theta}_1 \end{cases} \tag{A.6}$$

A.1.3 Monotonicity in $\hat{\phi}_n$

- Case 1: $\hat{\phi}_p < \hat{\theta}_1$. Monotonic.

$$g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases} 1 & \hat{\phi}_n < \hat{\phi}_p < \hat{\theta}_1 \\ 0 & \hat{\phi}_p = \hat{\phi}_n < \hat{\theta}_1 \\ 0 & \hat{\phi}_p < \hat{\phi}_n < \hat{\theta}_1 \\ 0 & \hat{\phi}_p < \hat{\phi}_n = \hat{\theta}_1 \\ 0 & \hat{\phi}_p < \hat{\theta}_1 < \hat{\phi}_n \end{cases} \tag{A.7}$$

- Case 2: $\hat{\theta}_1 < \hat{\phi}_p$. Monotonic.

$$g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases} \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} & \hat{\phi}_n < \hat{\theta}_1 < \hat{\phi}_p \\ \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = \frac{0}{\hat{\phi}_p - \hat{\phi}_n} = 0 & \hat{\theta}_1 = \hat{\phi}_n < \hat{\phi}_p \\ 0 & \hat{\theta}_1 < \hat{\phi}_n < \hat{\phi}_p \\ 0 & \hat{\theta}_1 < \hat{\phi}_n = \hat{\phi}_p \\ 0 & \hat{\theta}_1 < \hat{\phi}_p < \hat{\phi}_n \end{cases}$$

(A.8)

- Case 3: $\hat{\theta}_1 = \hat{\phi}_p$. Monotonic.

$$g(\hat{\theta}_1, \hat{\phi}_n, \hat{\phi}_p) = \begin{cases} \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = 1 & \hat{\phi}_n < \hat{\theta}_1 = \hat{\phi}_p \\ \frac{\hat{\theta}_1 - \hat{\phi}_n}{\hat{\phi}_p - \hat{\phi}_n} = \frac{0}{0} \equiv 0 & \hat{\phi}_n = \hat{\theta}_1 = \hat{\phi}_p \\ 0 & \hat{\theta}_1 = \hat{\phi}_p < \hat{\phi}_n \end{cases}$$

(A.9)

A.2 Additional Figures
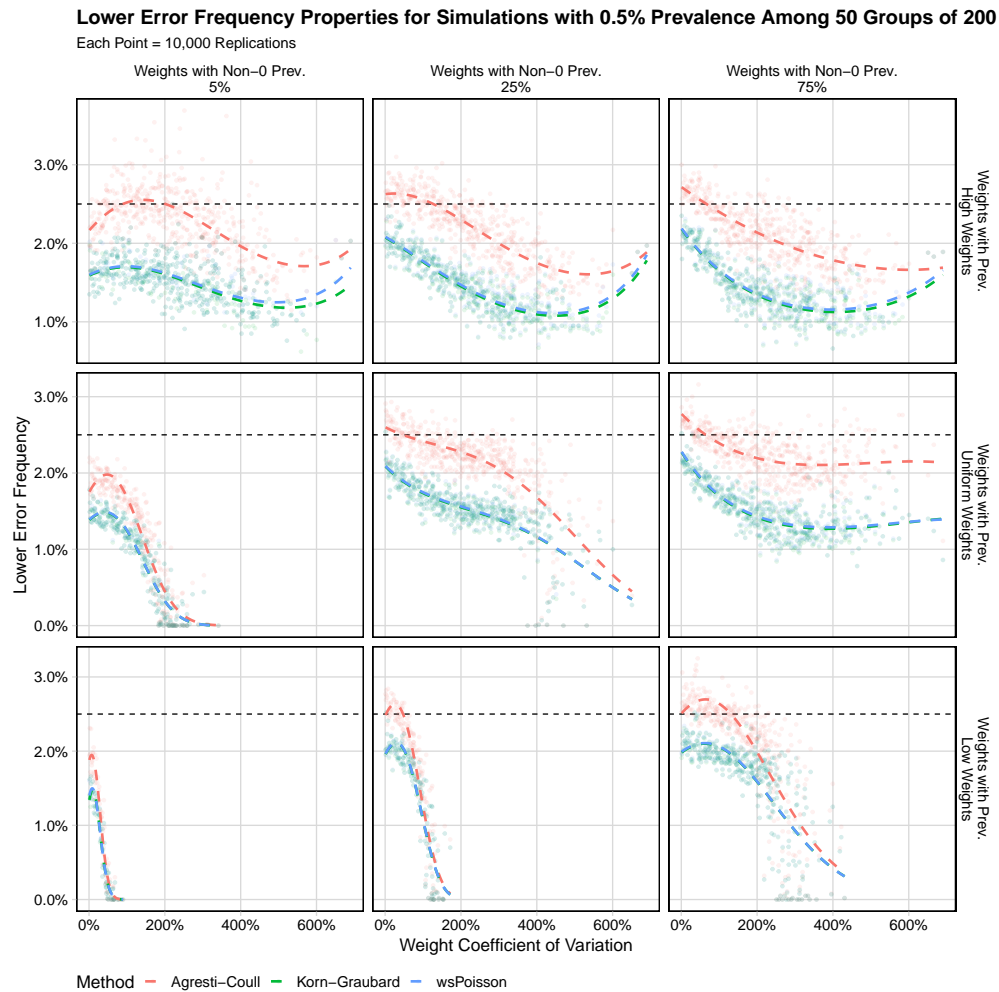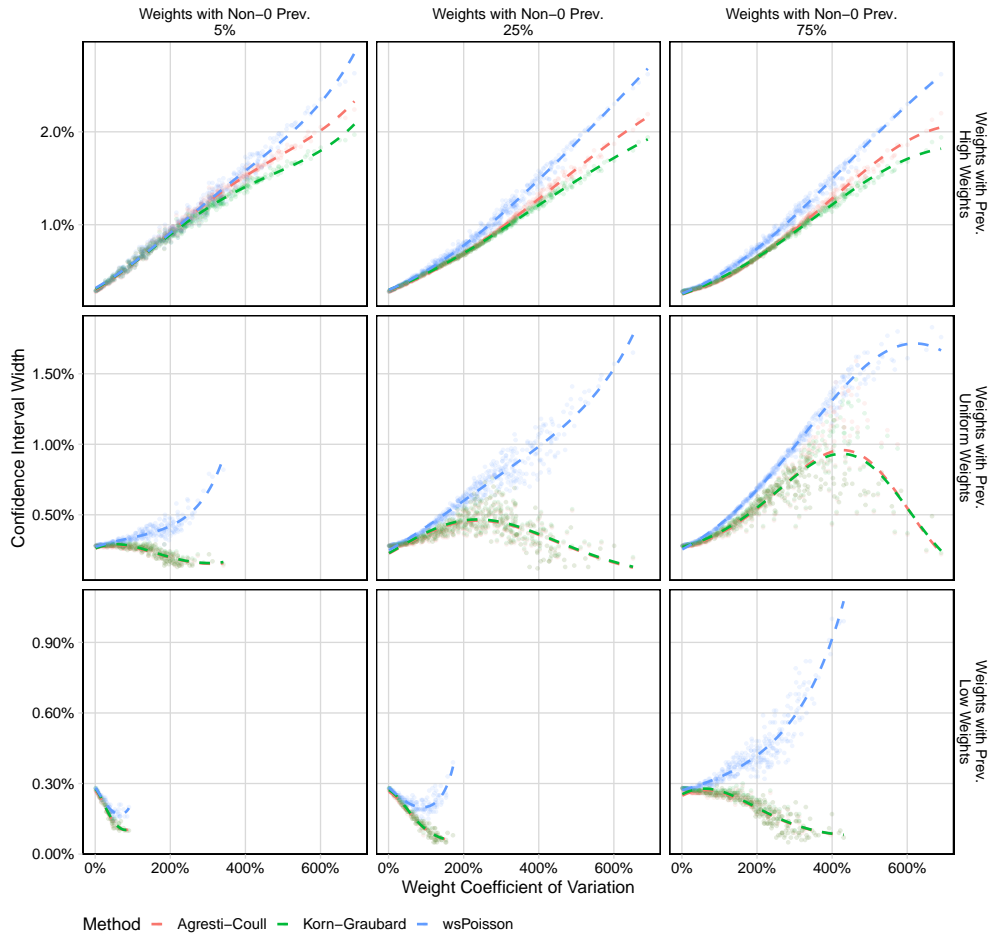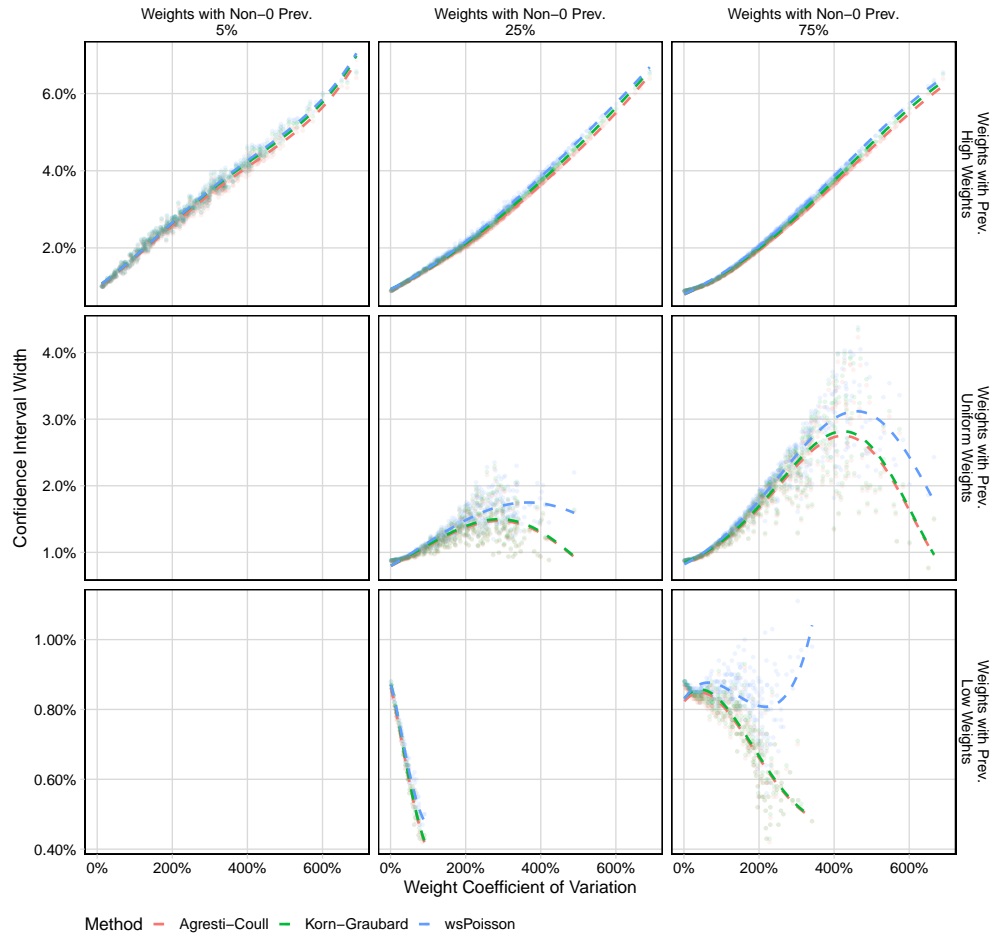

Figure A.1: Lower error properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 50 groups of 200 people are sampled. The horizontal dashed line indicates the nominal lower error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Lower Error Frequency Properties for Simulations with 0.5% Prevalence Among 8,000 Groups of 1

Each Point = 10,000 Replications
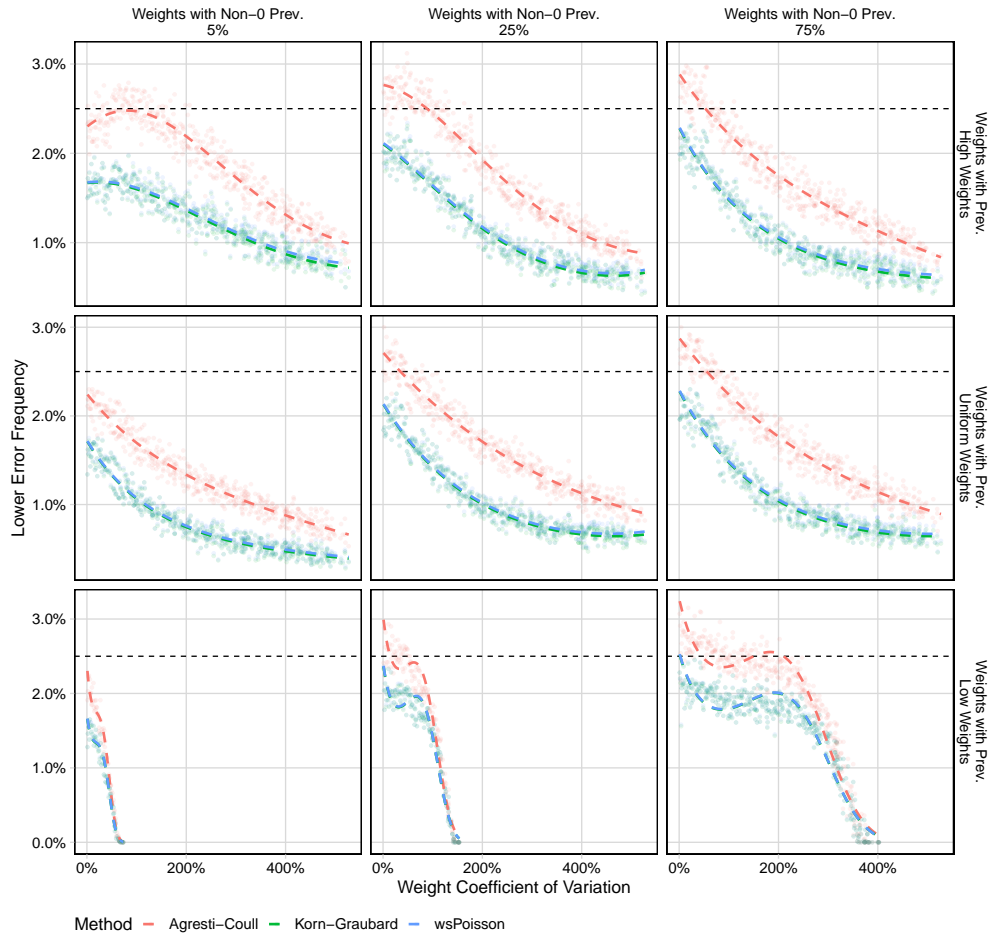
Figure A.2: Lower error properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 8000 individuals are sampled. The horizontal dashed line indicates the nominal lower error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.
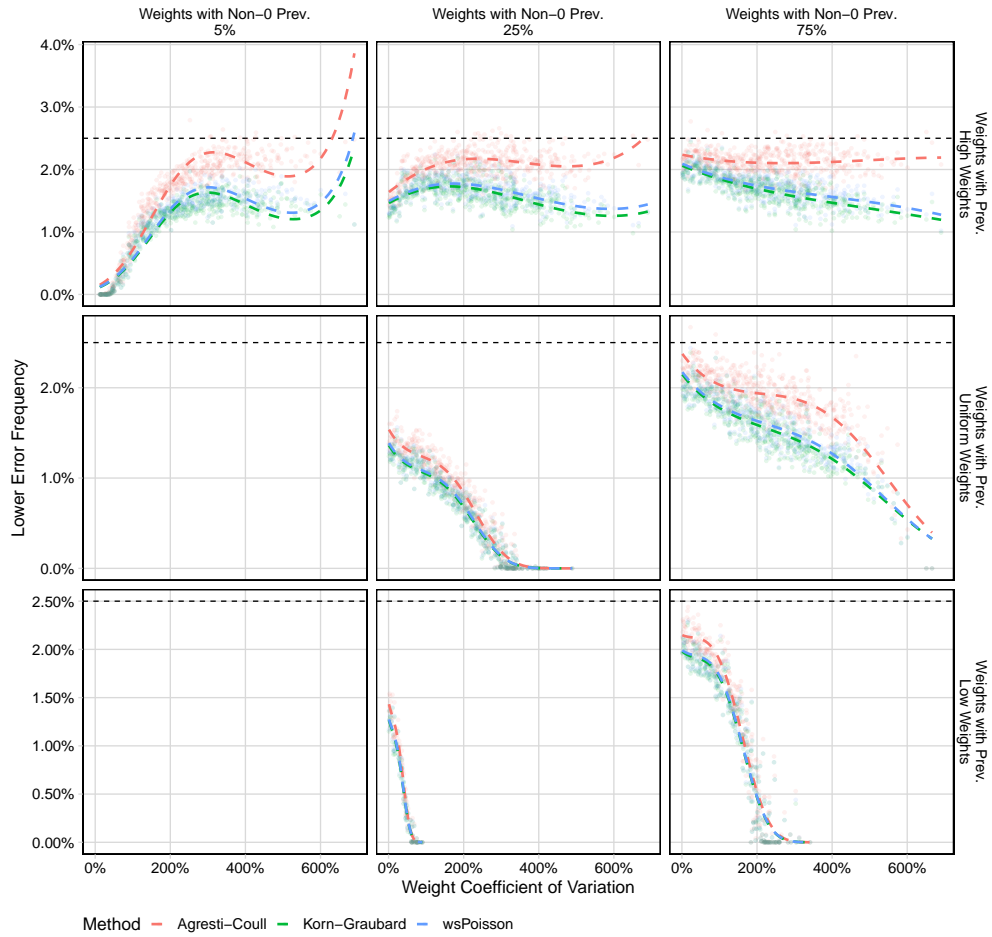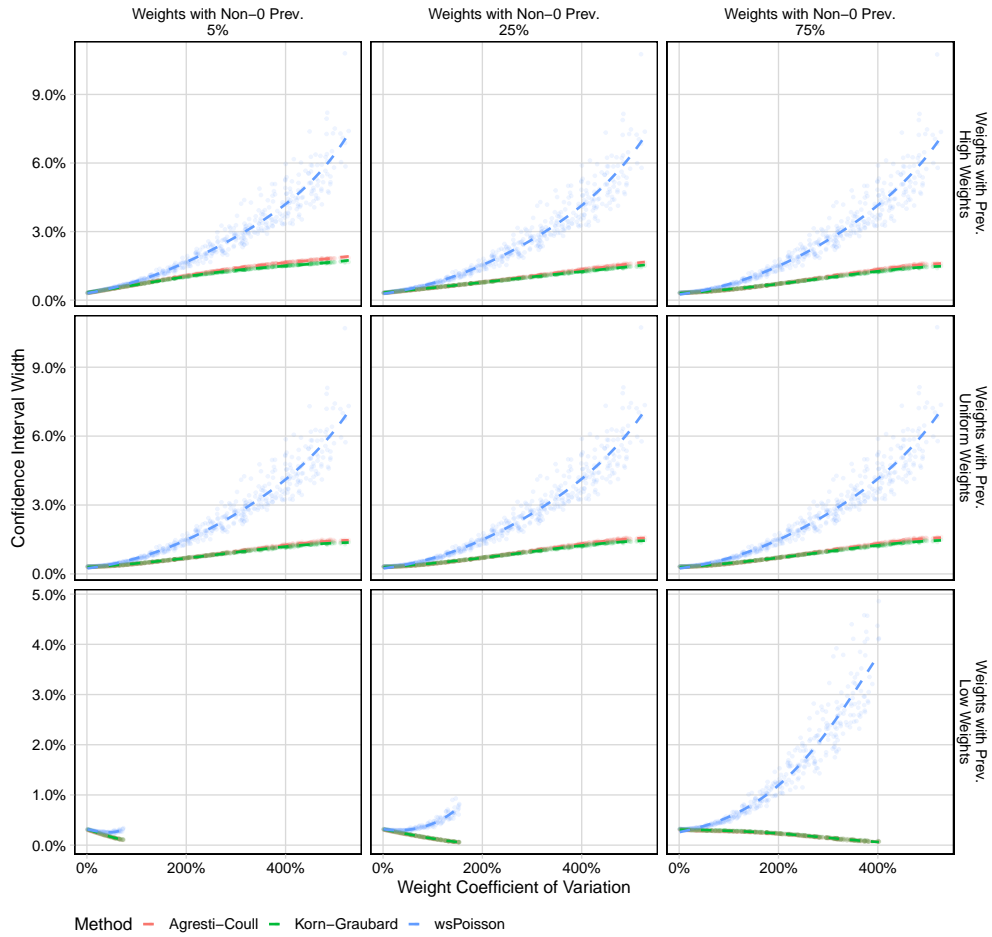
Figure A.3: Lower error properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 50 groups of 200 people are sampled. The horizontal dashed line indicates the nominal lower error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.
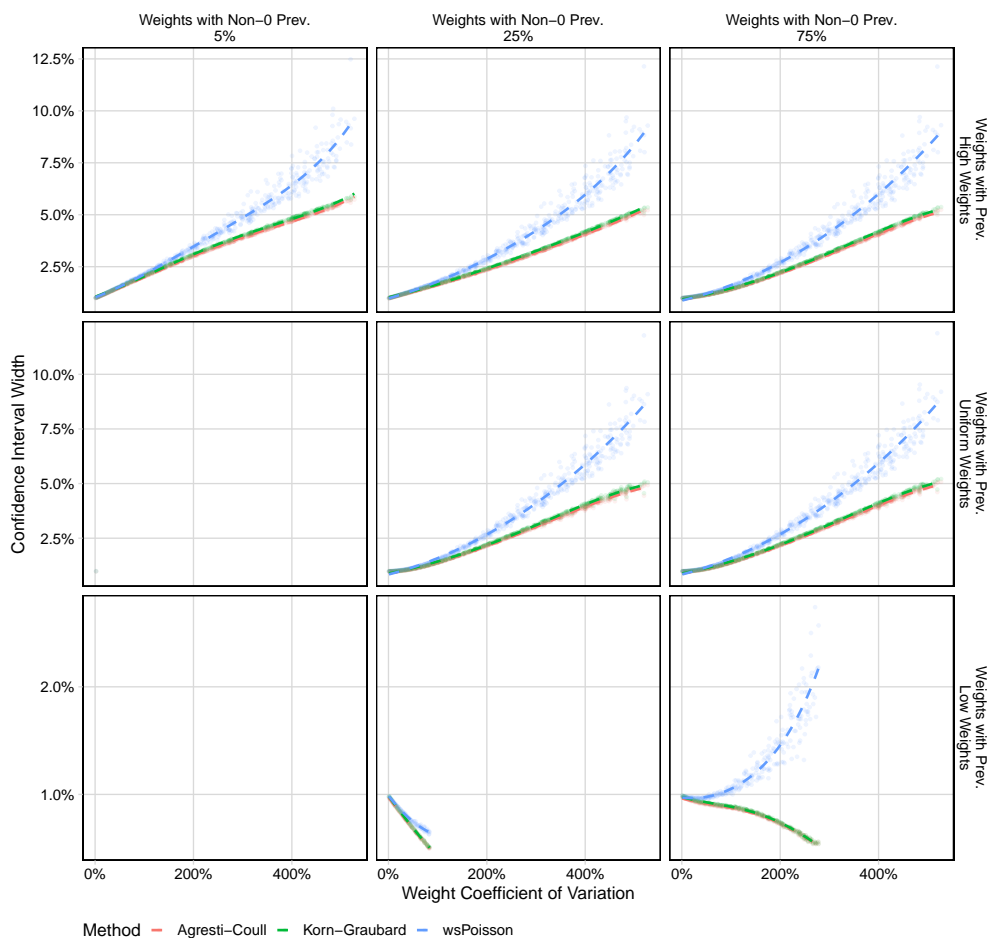
Figure A.4: Lower error properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 8000 individuals are sampled. The horizontal dashed line indicates the nominal lower error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.5: Upper error properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 50 groups of 200 people are sampled. The horizontal dashed line indicates the nominal upper error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.6: Upper error properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 8000 individuals are sampled. The horizontal dashed line indicates the nominal upper error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.7: Upper error properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 50 groups of 200 people are sampled. The horizontal dashed line indicates the nominal upper error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.8: Upper error properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 8000 individuals are sampled. The horizontal dashed line indicates the nominal upper error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.9: Confidence interval width properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 50 groups of 200 people are sampled. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.10: Confidence interval width properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 50 groups of 200 people are sampled. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.11: Confidence interval width properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 8000 individuals are sampled. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.12: Confidence interval width properties for the wsPoisson model and two standard methods, the Dean-Pagano modification of the Agresti-Coull method and of the Korn-Graubard method. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 8000 individuals are sampled. Colored dashed lines are estimates from a logistic regression model using quadratic splines.
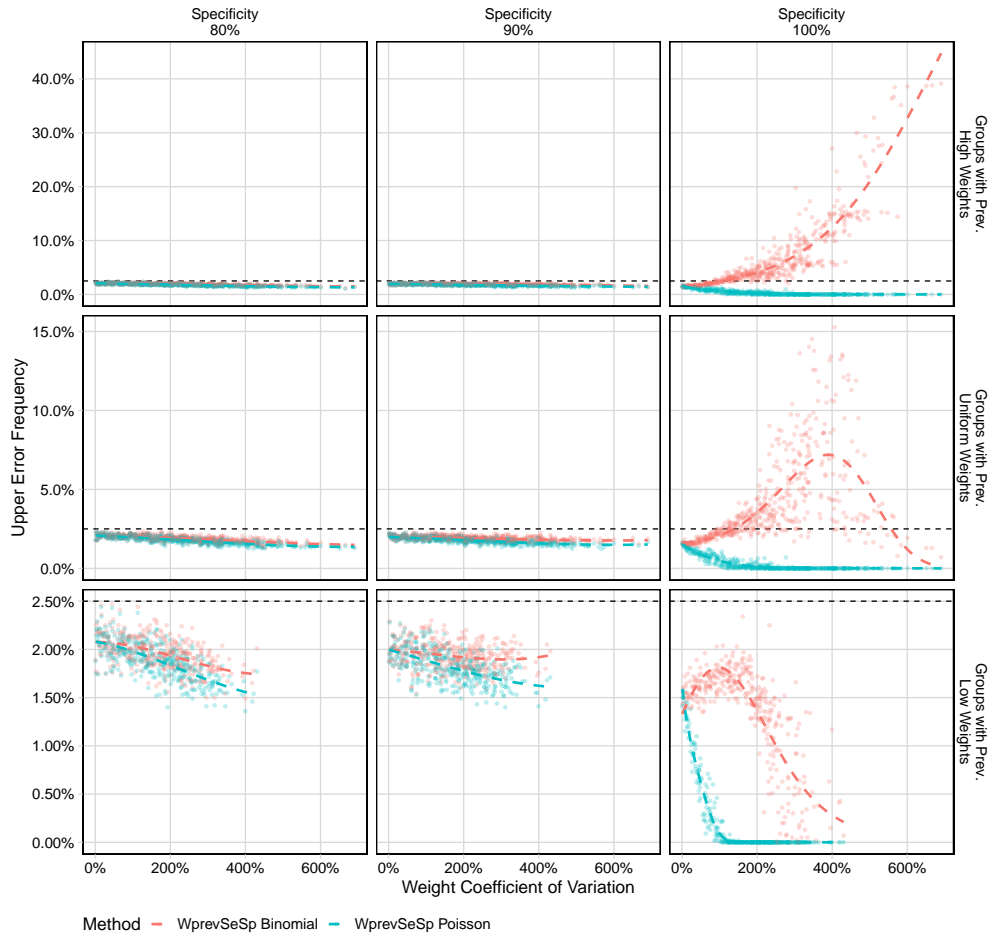
Figure A.13: Lower error properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 50 groups of 200 people are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal lower error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines. If the WprevSeSp Binomial line is not visible, then it is covered by the WprevSeSp Poisson line.
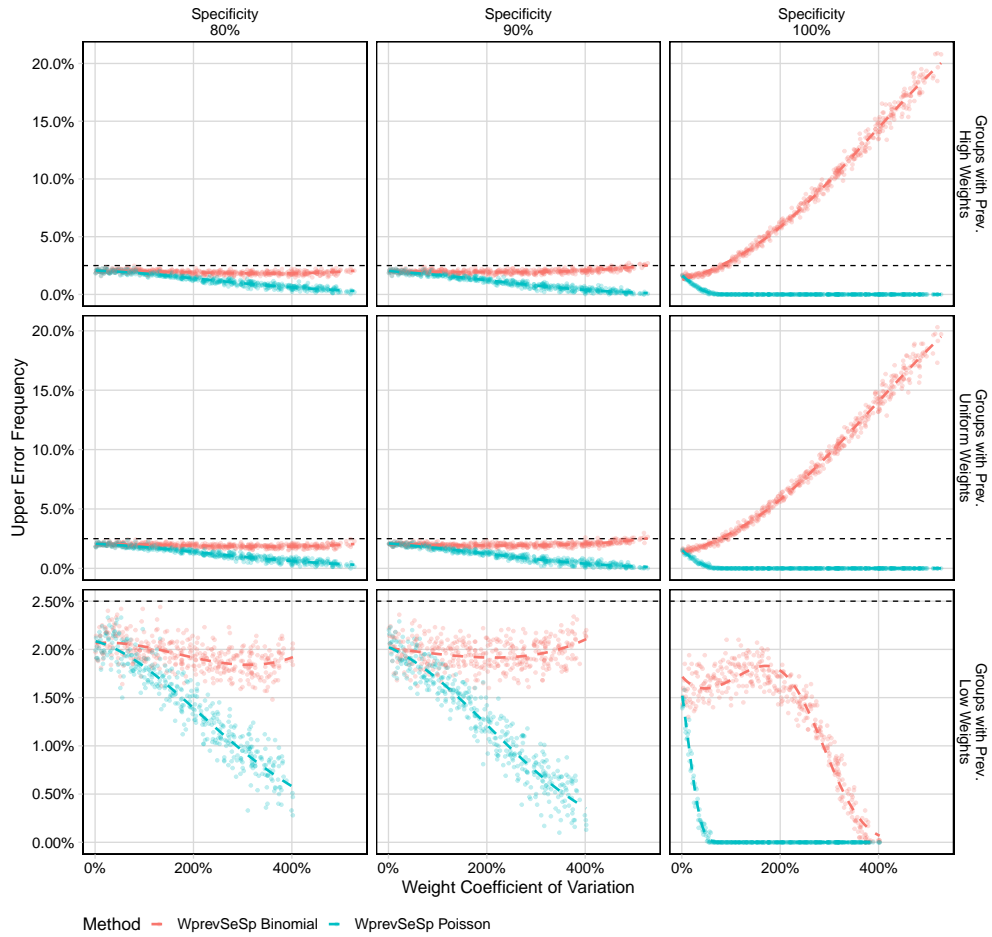
Figure A.14: Lower error properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 8000 individuals are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal lower error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.15: Lower error properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 50 groups of 200 people are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal lower error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.
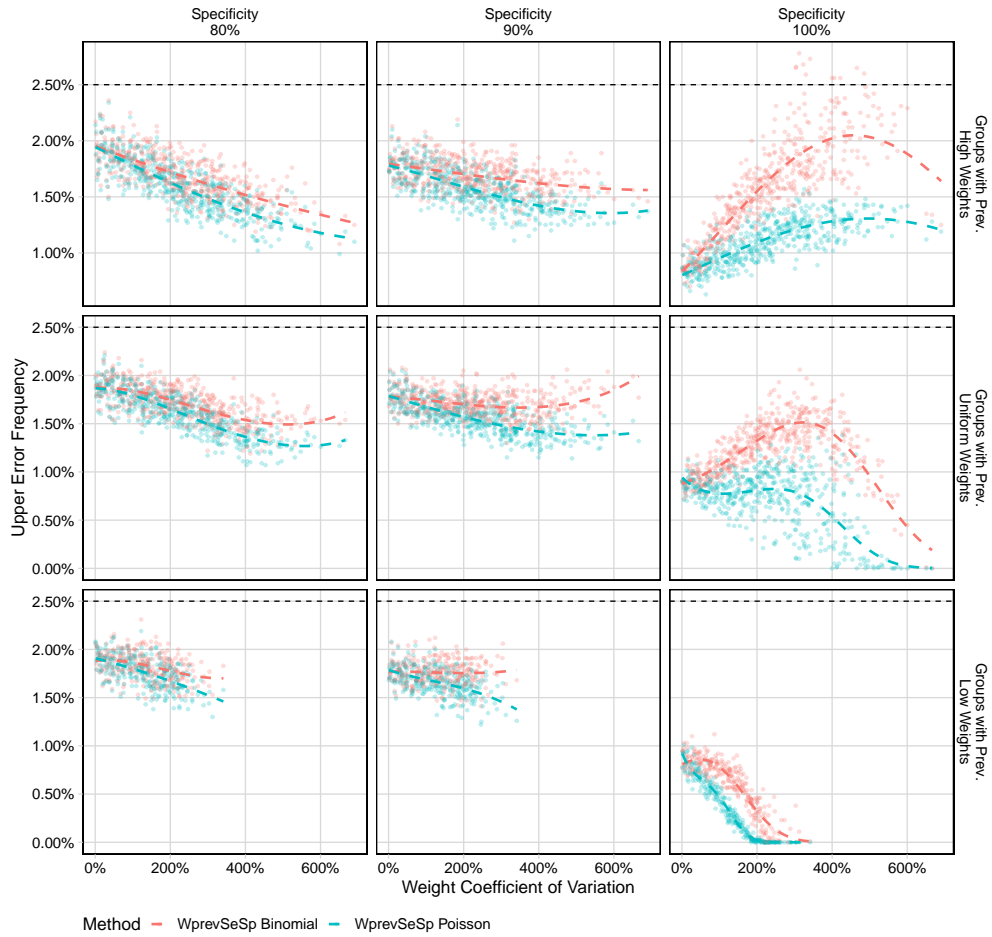
Figure A.16: Lower error properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 8000 individuals are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal lower error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Upper Error Frequency Properties for Simulations with 0.5% Prevalence Among 50 Groups of 200

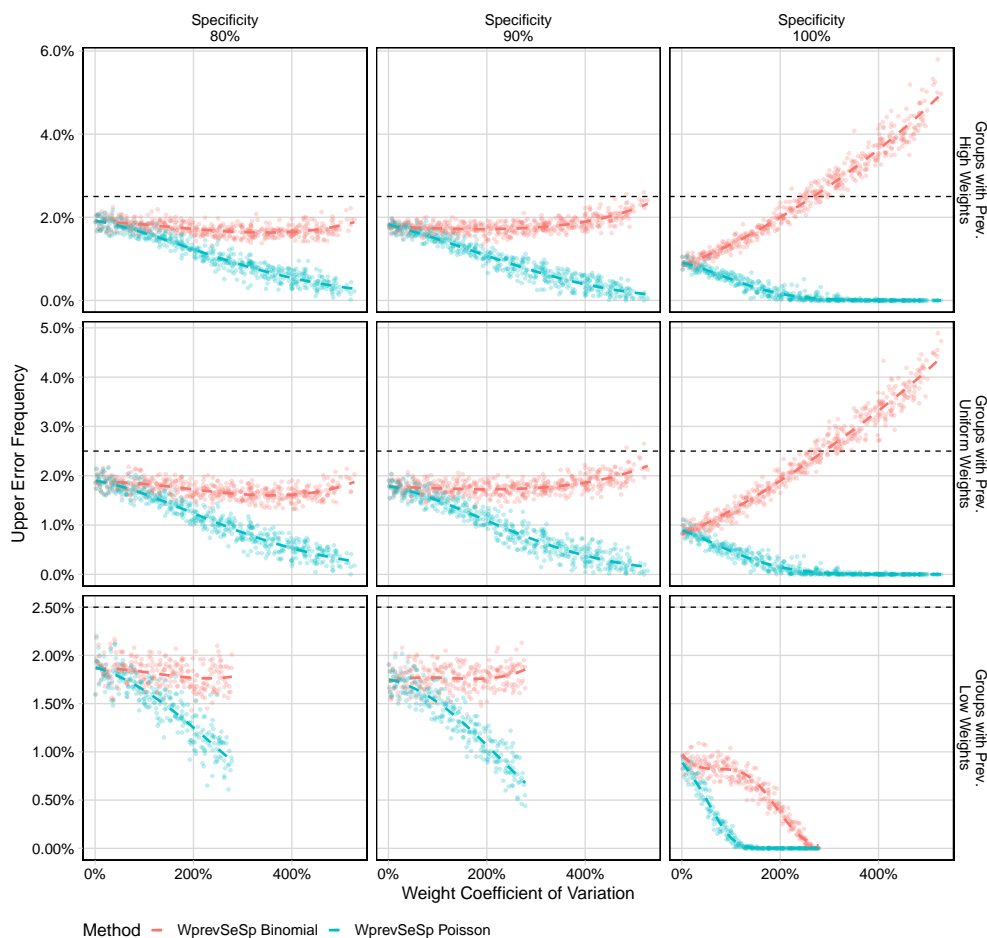Each Facet = 95% Sensitivity, Each Point = 10,000 Replications

Figure A.17: Upper error properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 50 groups of 200 people are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal upper error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.18: Upper error properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 8000 individuals are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal upper error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Upper Error Frequency Properties for Simulations with 5.0% Prevalence Among 50 Groups of 200

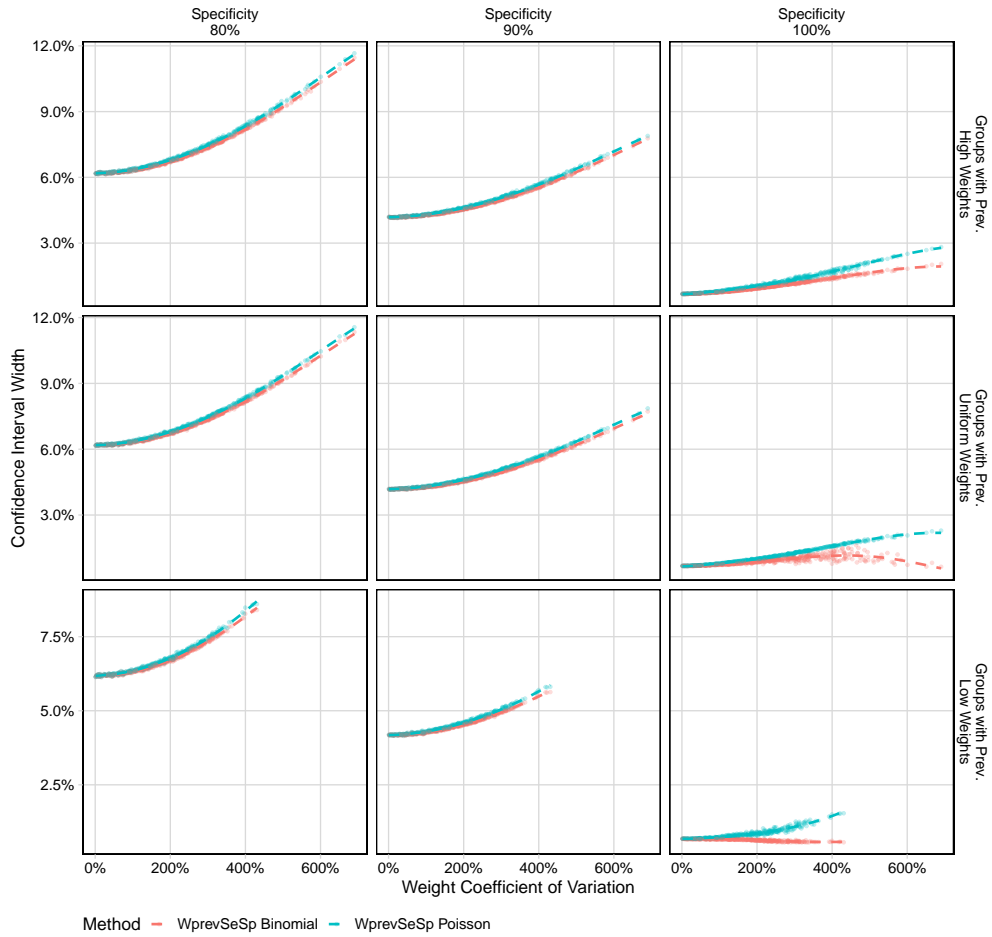Each Facet = 95% Sensitivity, Each Point = 10,000 Replications

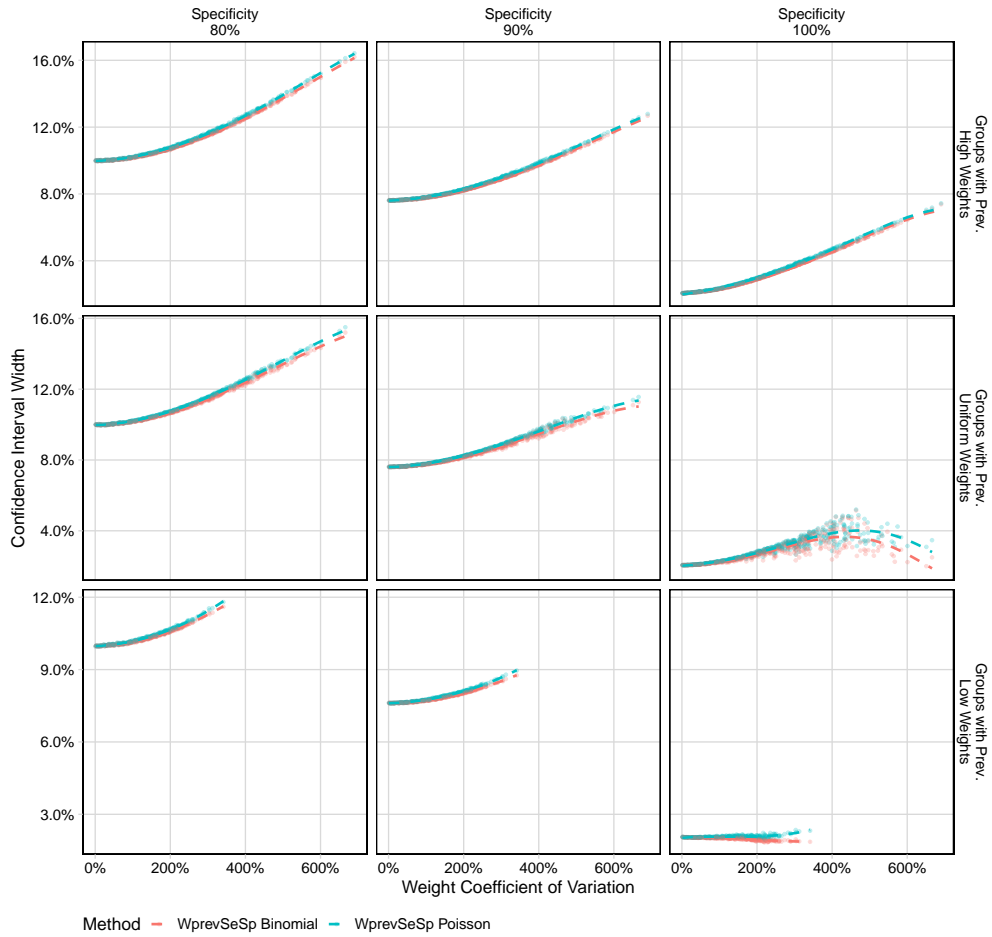Figure A.19: Upper error properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 50 groups of 200 people are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal upper error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Figure A.20: Upper error properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 8000 individuals are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. The horizontal dashed line indicates the nominal upper error rate, 2.5%. Colored dashed lines are estimates from a logistic regression model using quadratic splines.
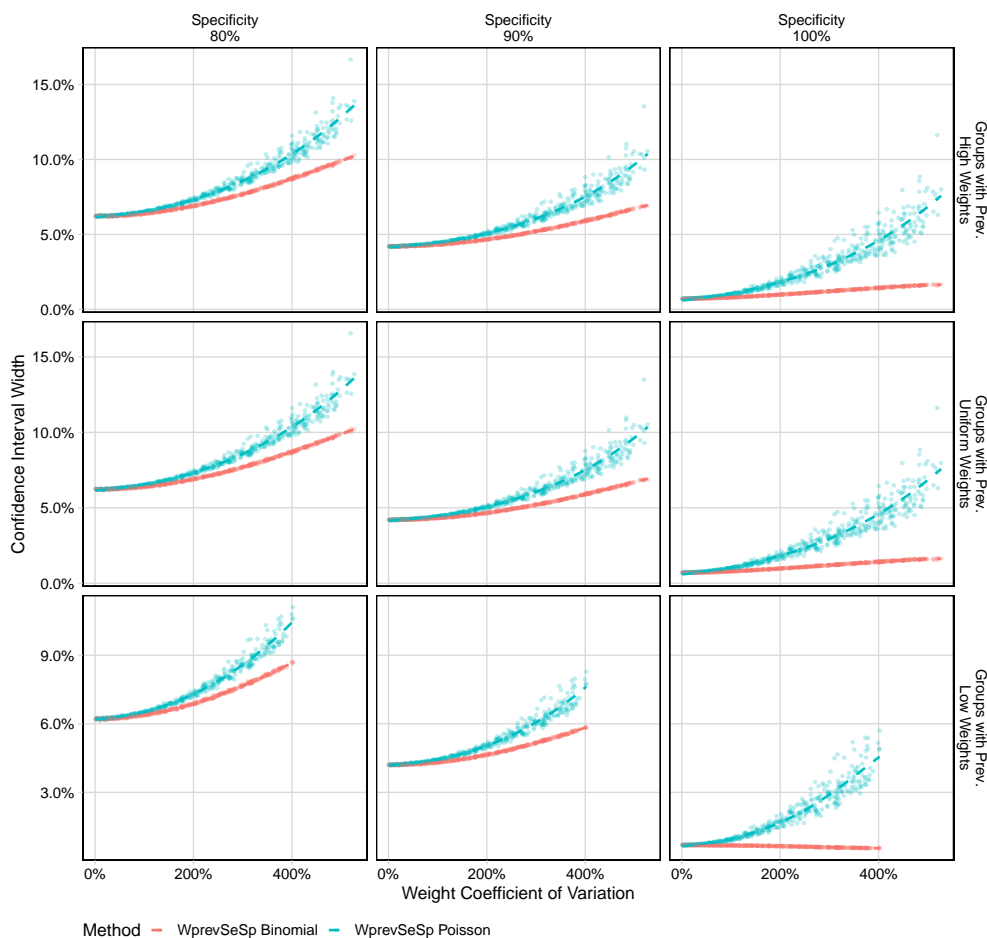
Figure A.21: Confidence interval width properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 50 groups of 200 people are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Confidence Interval Width Properties for Simulations with 5.0% Prevalence Among 50 Groups of 2

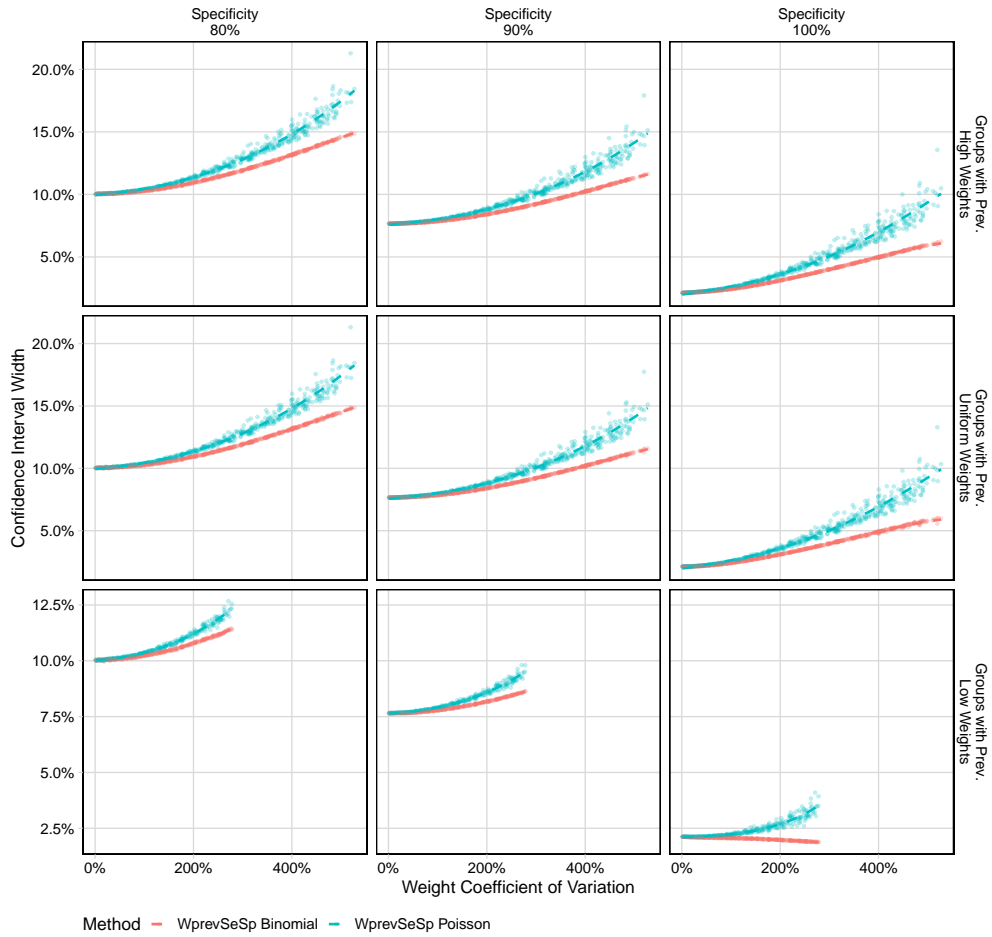Each Facet = 95% Sensitivity, Each Point = 10,000 Replications

Figure A.22: Confidence interval width properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 50 groups of 200 people are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Confidence Interval Width Properties for Simulations with 0.5% Prevalence Among 8,000 Groups

Each Facet = 95% Sensitivity, Each Point = 10,000 Replications

Figure A.23: Confidence interval width properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 0.5% prevalence, where 8000 individuals are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Confidence Interval Width Properties for Simulations with 5.0% Prevalence Among 8,000 Groups

Each Facet = 95% Sensitivity, Each Point = 10,000 Replications

Figure A.24: Confidence interval width properties for the confidence interval procedures, WprevSeSp Binomial and WprevSeSp Poisson. Each point represents 10,000 simulations of datasets from a population with 5% prevalence, where 8000 individuals are sampled. Each dataset also includes simulated results of tests to evaluate the sensitivity and specificity of the assay performed on 60 and 300 individuals, respectively. Colored dashed lines are estimates from a logistic regression model using quadratic splines.

Appendix B

Additional Material for Chapter 4

B.1 Simulation study

We performed a simulation study on 200 datasets to validate our models. We use the same prior distributions for the parameters as in the main text. These distributions are presented in Table B.2. We purposely chose parameter values that resulted in data similar to the Orange Country data used in the main text. Exact values for these parameters are presented in Table B.2. One of the 200 simulated datasets is presented in Figure B.1. Figures B.2–B.4 present the prior and posterior distribution for this single dataset. Figures B.5–B.8 show coverage and contraction properties for the whole simulation study. Contraction is calculated as one minus the ratio of standard deviation of the posterior and the prior. Commentary on these results is presented in Section 4.3 of the main text.

Table B.1: Model parameters and their prior distributions.

Parameter	Interpretation	Prior	Prior Median (95% Interval)	Source
S_0	Initial susceptible proportion	Logit-Normal(6, 0.25)	0.998 (0.993, 0.999)	
$\tilde{I}_0$	Initial proportion of non-susceptibles who are infectious	Logit-Normal(0.6, 0.0009)	0.646 (0.632, 0.659)	
$\exp\left(\tilde{R}_{0,1}\right)$	Initial basic reproduction number	Log-Normal(0, 0.0625)	1.000 (0.613, 1.630)	
$1/\gamma$	Mean latent period (weeks)	Log-Normal(-0.25, 0.01)	0.779 (0.640, 0.947)	Xin et al. [2021]
$1/\nu$	Mean infectious period (weeks)	Log-Normal(0.15, 0.01)	1.160 (0.955, 1.410)	Byrne et al. [2020]
$\mathrm{expit}\,(\tilde{\eta}_1)$	Initial infection fatality ratio	Logit-Normal(-5.3, 0.04)	0.00497 (0.00336, 0.00733)	Bruckner et al. [2021]
ρ^D	Mean death detection rate	Logit-Normal(2.3, 0.04)	0.909 (0.871, 0.937)	Bruckner et al. [2021]
ϕ_D	over-dispersion in observed deaths negative-binomial model	Log-Normal(4.16, 0.293)	63.9 (22.1, 185.0)	
$\exp\,(\tilde{\alpha}_1)$	Initial proportion in prop. odds test positivity model	Log-Normal(1.35, 0.0121)	3.86 (3.11, 4.79)	
ϕ_C	over-dispersion in observed cases beta-binomial model	Log-Normal(6.5, 0.0673)	665 (400, 1110)	
σ_{R_0}	Standard deviation of log-GMRF for time-varying R_0	Log-Normal(-1.9, 0.09)	0.1500 (0.0831, 0.2690)	
σ_η	Standard deviation of logit-GMRF for time-varying η	Log-Normal(-2.4, 0.0144)	0.0907 (0.0717, 0.1150)	
σ_α	Standard deviation of log-GMRF for time-varying α	Log-Normal(-2.7, 0.0225)	0.0672 (0.0501, 0.0902)	
$\tilde{\rho}_1^Y$	Initial case detection rate	Logit-Normal(-2.5, 0.01)	0.0759 (0.0632, 0.0908)	Bruckner et al. [2021]
ϕ_Y	over-dispersion in observed cases negative-binomial model	Log-Normal(3.93, 0.0684)	51.1 (30.6, 85.3)	
σ_{ρ^Y}	Standard deviation of logit-GMRF for time-varying ρ^Y	Log-Normal(-2.2, 0.04)	0.1110 (0.0749, 0.1640)	

Table B.2: Simulation parameters.

Parameter	Interpretation	Value
S_0	Initial susceptible proportion	0.9979
$\tilde{I}_0$	Initial proportion of non-susceptibles who are infectious	0.6455
$\exp\left(\tilde{R}_{0,1}\right)$	Initial basic reproduction number	1.2602
$1/\gamma$	Mean latent period (weeks)	0.7697
$1/\nu$	Mean infectious period (weeks)	1.1997
$\text{expit}\left(\tilde{\eta}_1\right)$	Initial infection fatality ratio	0.0005
ρ^D	Mean death detection rate	0.9061
ϕ_D	over-dispersion in observed deaths negative-binomial model	87.2776
$\exp\left(\tilde{\alpha}_1\right)$	Initial proportion in proportional odds test positivity model	4.3958
ϕ_C	over-dispersion in observed cases Beta-Binomial model	1,026.6765
σ_{R_0}	Standard deviation of log-Guassian Markov random field for time-varying R_0	0.1481
σ_η	Standard deviation of logit-Guassian Markov random field for time-varying η	0.0944
σ_α	Standard deviation of log-Guassian Markov random field for time-varying α	0.0696

Simulated data

Counts binned into weekly periods
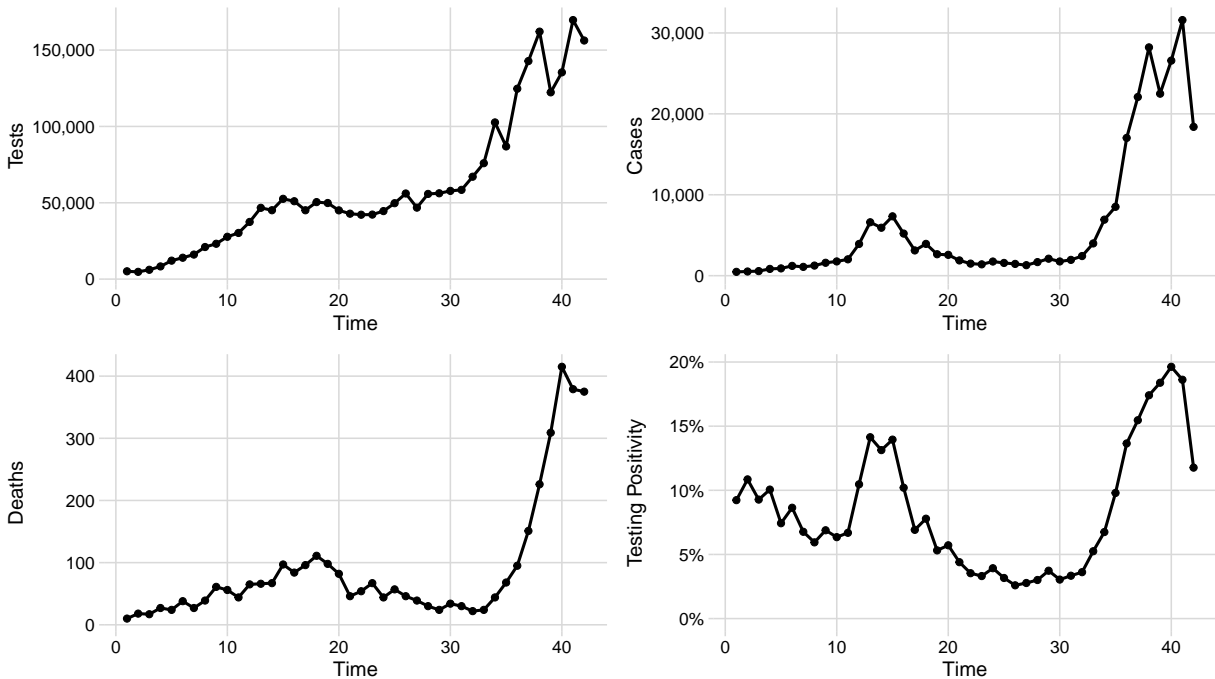
Figure B.1: Simulated data. The figure shows weekly counts of tests, cases (positive tests), reported deaths due to COVID-19, as well testing positivity.
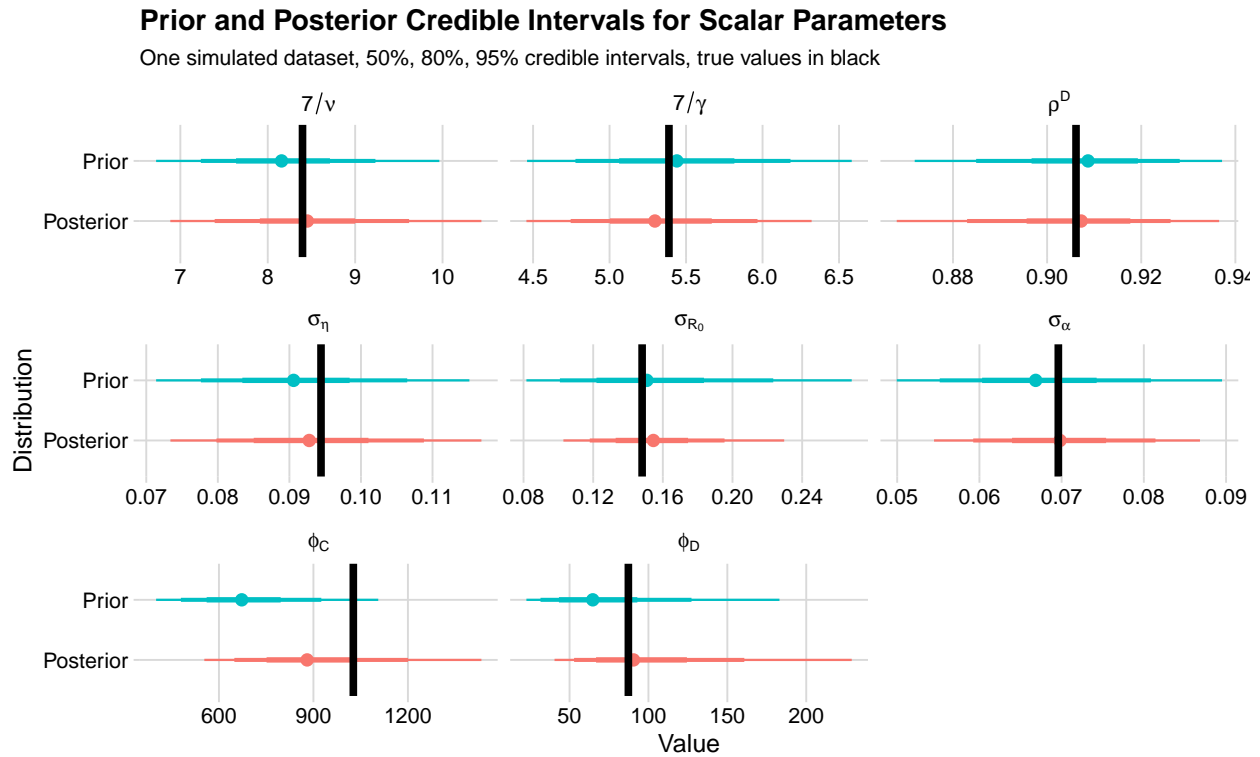
Figure B.2: Prior and posterior credible intervals for scalar parameters for a model fit to the dataset presented in Figure B.1. True values for the simulated parameters are indicated by solid black lines.
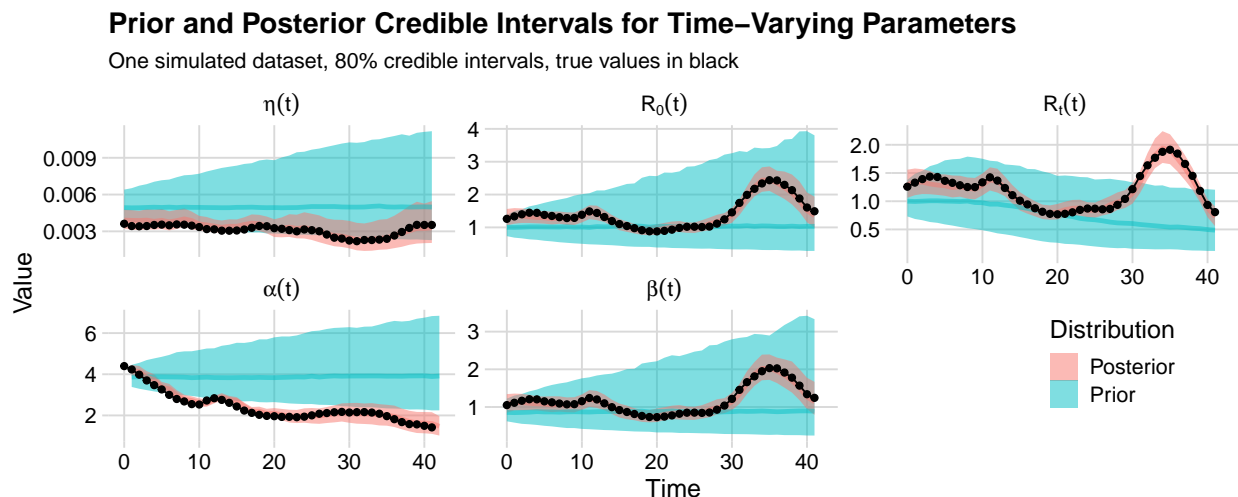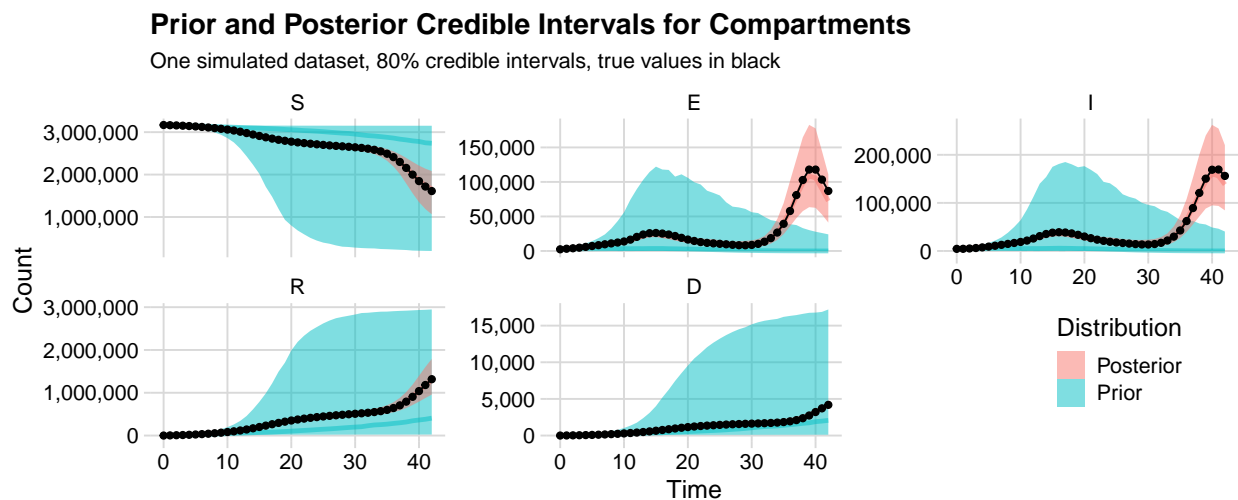

Figure B.3: Prior and posterior 80% credible intervals for time-varying parameters for a model fit to the dataset presented in Figure B.1. True values for the simulated parameters are indicated by solid black lines.

Prior and Posterior Credible Intervals for Compartments

One simulated dataset, 80% credible intervals, true values in black

Figure B.4: Prior and posterior 80% credible intervals for latent compartments for a model fit to the dataset presented in Figure B.1. True values for the simulated compartment sizes are indicated by solid black lines.

Posterior Coverage Properties of Scalar Parameters

Mean and 95% Confidence Interval from 200 Simulations
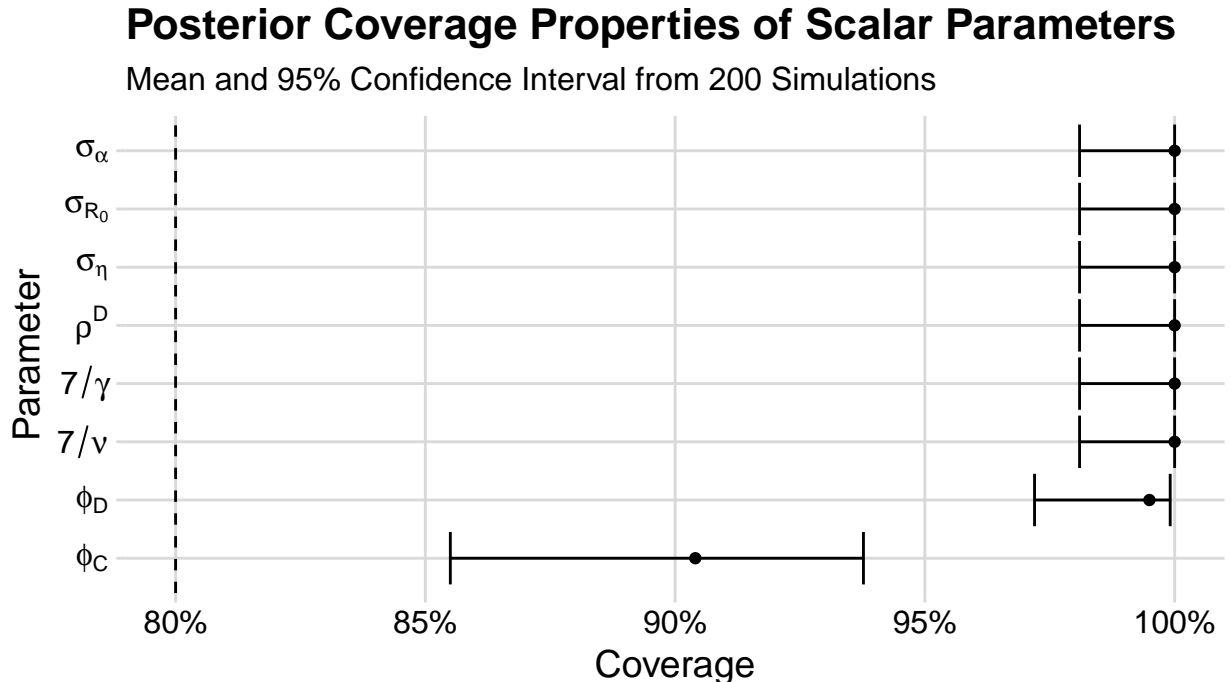
Figure B.5: Coverage properties of 80% posterior credible intervals for scalar parameters from models fit to 200 simulated datasets. Nominal coverage is indicated by the dashed line.

Figure B.6: Contraction properties of scalar parameters from models fit to 200 simulated datasets. Contraction is calculated as one minus the ratio of standard deviation of the posterior and the prior.
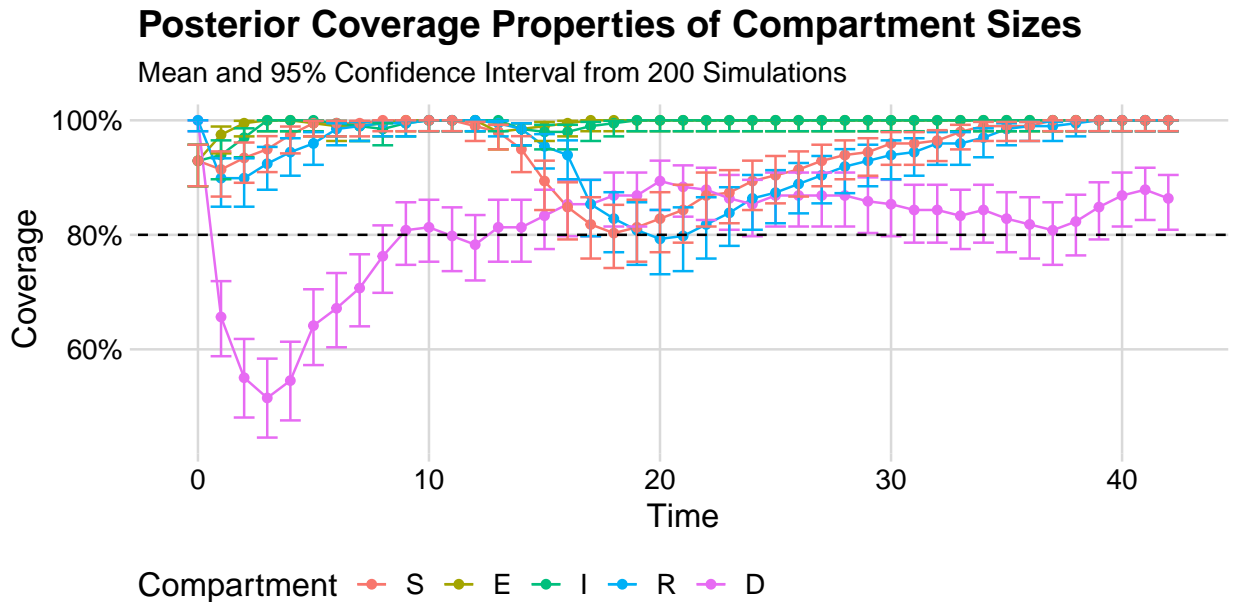

Figure B.7: Coverage properties of 80% posterior credible intervals for time-varying parameters from models fit to 200 simulated datasets. Nominal coverage is indicated by the dashed line.
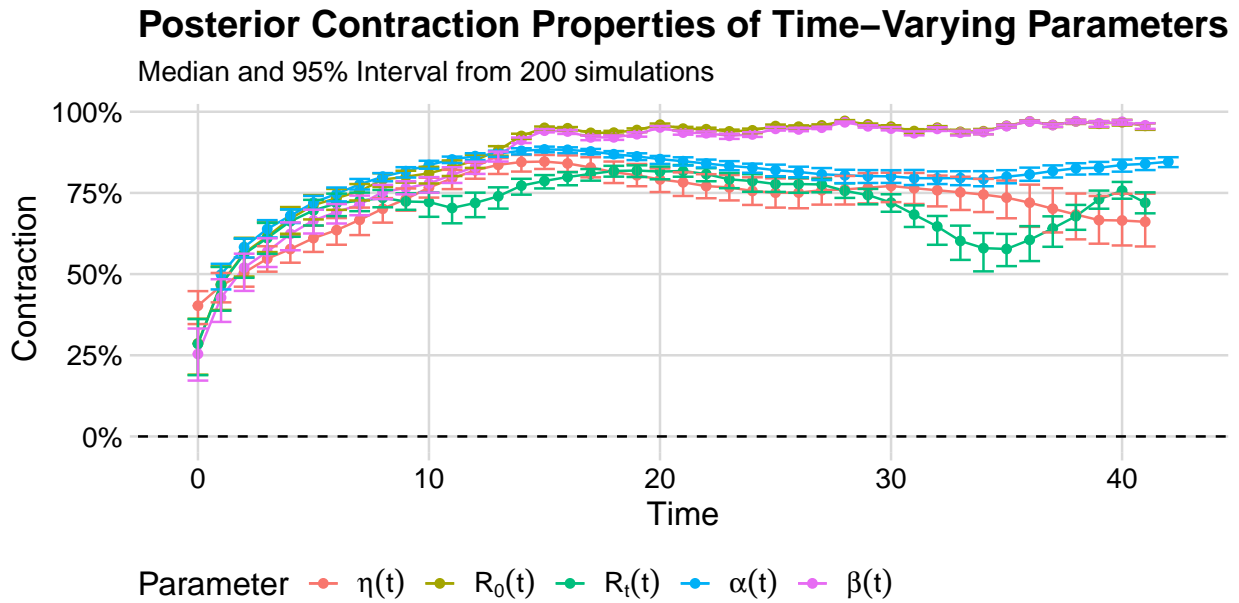
Figure B.8: Contraction properties of time-varying parameters from models fit to 200 simulated datasets. Contraction is calculated as one minus the ratio of standard deviation of the posterior and the prior.

Figure B.9: Coverage properties of 80% posterior credible intervals for latent compartments from models fit to 200 simulated datasets. Nominal coverage is indicated by the dashed line.
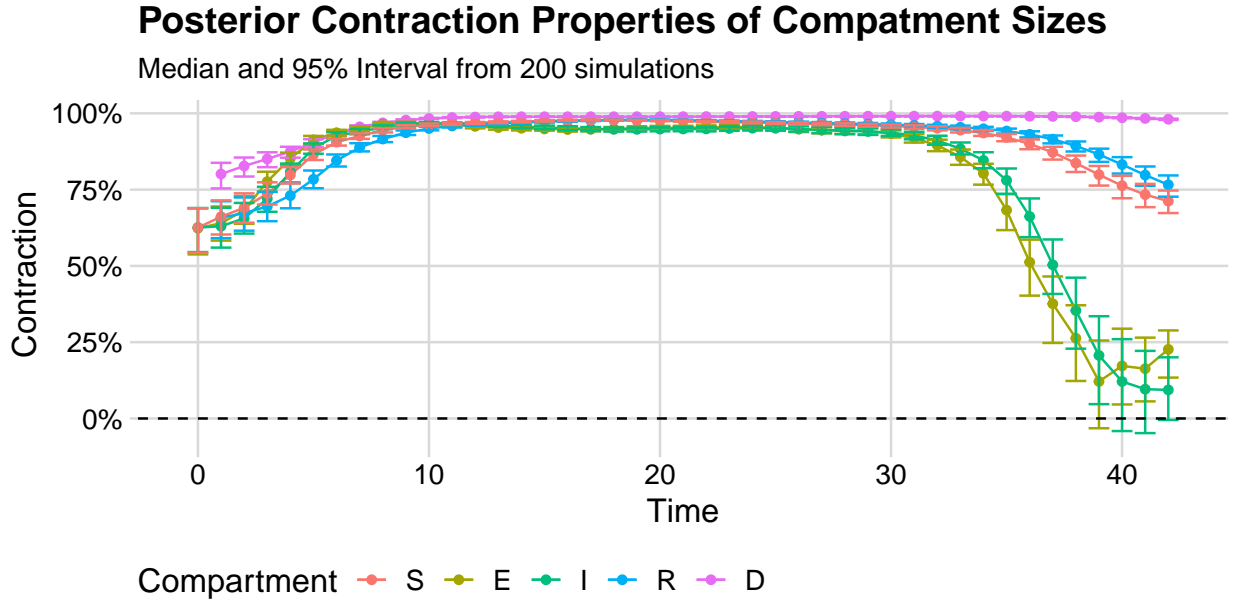
Posterior Contraction Properties of Compatment Sizes

Median and 95% Interval from 200 simulations

Figure B.10: Contraction properties of latent compartments from models fit to 200 simulated datasets. Contraction is calculated as one minus the ratio of standard deviation of the posterior and the prior.

B.2 Comparison with `epidemia`

We used the `epidemia` R package to infer R_t in the same 200 simulated datasets, as well as the Orange County data. Statistical details of these methods are presented below. Commentary on these results is presented in Section 4.3 of the main text.

The `epidemia` package can be used to create different branching process inspired models to estimate the effective reproduction number. In contrast to the compartmental model used in this paper, branching process inspired models have related the mean of current incidence to a weighted sum of previous incidence and the effective reproduction number R_t. Let I_t be the incidence at time t, R_t be the effective reproduction number at time t, and $g(t)$ be the probability density function of the generation time distribution (the time between an individual becoming infected and infecting another individual; under the compartmental model framework this is usually taken to be equivalent to the sum of the latent period and

the infectious period). Then the mean relationship used is:

$$E[I_t|I_1,\ldots,I_{t-1}] = R_t \sum_{s=1}^{t-1} I_s g(t-s).$$

For the model we used in this study, we then added an observation model for new cases, modeled the effective reproduction number as a random walk, and modeled unobserved incidence as an autoregressive normal random variable with variance equal to the mean multiplied by an over-dispersion parameter.

$$\tau \sim \exp(\lambda) \quad \text{Hyperprior for unobserved incidence}$$

$$I_\nu \sim \exp(\tau) \quad \text{Prior on unobserved incidence } \nu \text{ days before observation}$$

$$I_{\nu+1},\ldots,I_0 = I_\nu \quad \text{Unobserved incidence}$$

$$\sigma \sim \text{Truncated-Normal}(0, 0.1^2)$$

$$\log R_0 \sim \text{Normal}(\log 2, 0.2^2) \quad \text{Prior on } R_0$$

$$\log R_t|\log R_{t-1} \sim \text{Normal}(\log R_{t-1}, \sigma) \quad \text{Random Walk prior on } R_t$$

$$\psi \sim \text{Normal}(10, 2) \quad \text{Prior on variance parameter for incidence}$$

$$I_t|I_\nu,\ldots,I_{t-1} \sim \text{Normal}(R_t \sum_{s<t} I_s g_{t-s}, \psi) \quad \text{Model for incidence}$$

$$\alpha \sim \text{Normal}(0.13, 0.7^2) \quad \text{Prior on case detection rate}$$

$$y_t = \alpha_t \sum_{s<t} I_s \pi_{t-s} \quad \text{Mean of observed data model}$$

$$\phi \sim P(\phi) \quad \text{Prior on dispersion parameter for observed data}$$

$$Y_t \sim \text{Neg-Binom}(y_t, \phi) \quad \text{Observed data model}$$

Here π_t are the values of the probability density function for the delay distribution, the time between an individual being infected and being observed. This distribution is assumed to

be a gamma distribution with shape parameter one and mean equal to the true mean latent period. To sample from the posterior distribution, `epidemia` uses Hamiltonian Monte Carlo via the `Stan` simulation software [Stan Development Team, 2020]. We draw 2000 posterior samples and discard the first 1000 for this analysis.
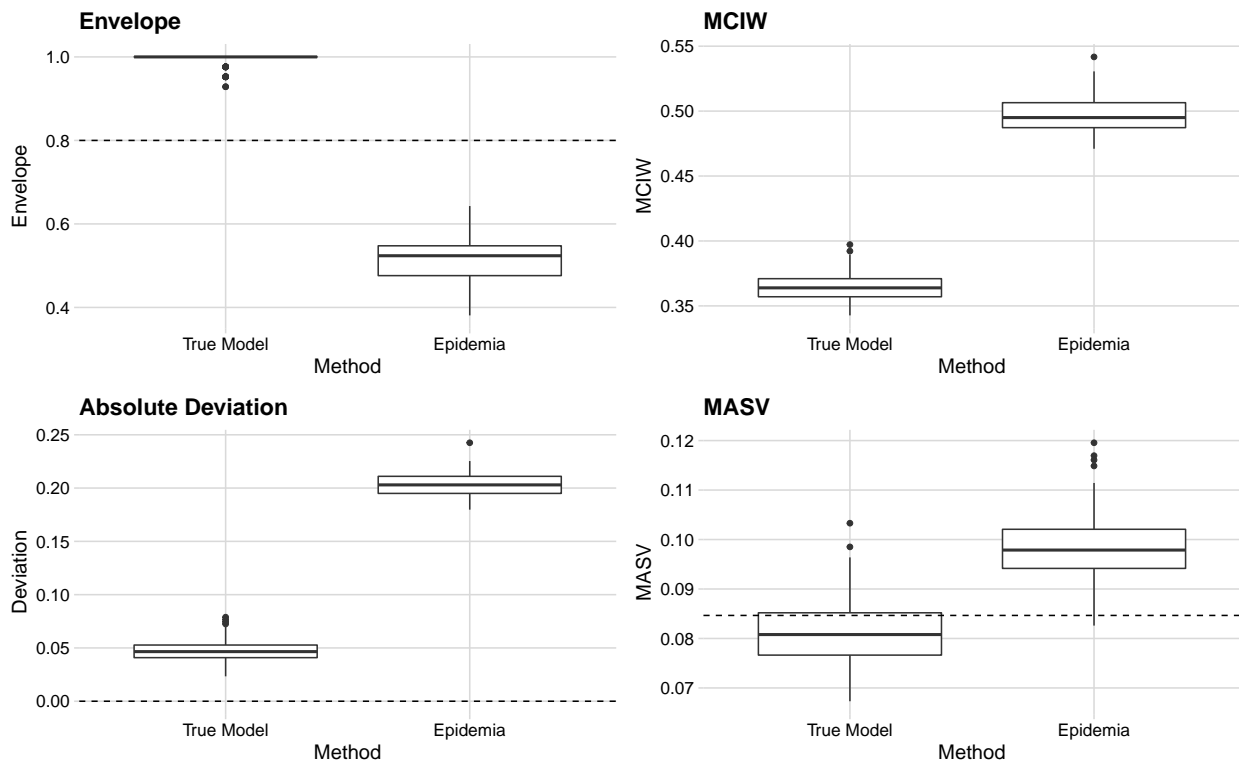

Figure B.11: Properties of R_t estimation from 200 simulated data sets. The envelope is the proportion of time points which the 80% posterior credible interval contains the true R_t value specified in the simulation. Mean credible interval width (MCIW) is the mean of credible interval widths across time points within a simulation replication. Absolute deviation is calculated as the mean of the absolute difference between the posterior median and the true R_t value at each time point. The mean absolute sequential variation (MASV) is the mean of the absolute difference between the posterior median at a time point and the posterior median at the previous time point.

B.3 Comparison with structured populations model

Here, we demonstrate that semi-parametric modeling of key parameters can obviate the need for modeling heterogeneous populations with separate compartments. We construct a model

wherein a disease spreads among two subpopulations: the "general" population and the "vulnerable" population, which interact with each other. Progression through compartments is governed by the following system of differential equations, with "g" subscripts denoting the general subpopulation and "v" subscripts denoting the vulnerable subpopulation and parameters having the same interpretations as in the main text. The differential equations used for this model presented in (B.1).

$$
\begin{aligned}
\frac{\mathrm{d}S_v}{\mathrm{d}t} &= -\left(\beta_{vv}I_v + \beta_{vg}I_g\right)\frac{S_v}{N} & \frac{\mathrm{d}S_g}{\mathrm{d}t} &= -\left(\beta_{gg}I_g + \beta_{gv}I_v\right)\frac{S_g}{N} \\
\frac{\mathrm{d}E_v}{\mathrm{d}t} &= \left(\beta_{vv}I_v + \beta_{vg}I_g\right)\frac{S_v}{N} - \gamma E_v & \frac{\mathrm{d}E_g}{\mathrm{d}t} &= \left(\beta_{gg}I_g + \beta_{gv}I_v\right)\frac{S_g}{N} - \gamma E_g \\
\frac{\mathrm{d}I_v}{\mathrm{d}t} &= \gamma E_v - \nu I_v & \frac{\mathrm{d}I_g}{\mathrm{d}t} &= \gamma E_g - \nu I_g \\
\frac{\mathrm{d}R_v}{\mathrm{d}t} &= \nu(1 - \eta_v)I_v & \frac{\mathrm{d}R_g}{\mathrm{d}t} &= \nu(1 - \eta_g)I_g \\
\frac{\mathrm{d}D_v}{\mathrm{d}t} &= \nu\eta_v I_v & \frac{\mathrm{d}D_g}{\mathrm{d}t} &= \nu\eta_g I_g
\end{aligned}
\tag{B.1}
$$

subject to initial conditions $\mathbf{X}(t_0) = \mathbf{x}_0$ and $\mathbf{N}(t_0) = \mathbf{0}$, where $\mathbf{x}_0 = (S_{v0}, E_{v0}, I_{v0}, R_{v0}, D_{v0}, S_{g0}, E_{g0}, I_{g0}, R_{g0}, D_{g0})$ are initial compartment counts.

We only observed the unstratified case and death counts. Observed cases and deaths are Poisson distributed with the rate parameter equal to the number of latent cases and deaths, respectively.

$$
Y_l \sim \mathrm{Poisson}(\Delta N_{E_v I_v}(t_l) + \Delta N_{E_g I_g}(t_l))
$$

$$
M_l \sim \mathrm{Poisson}(\Delta N_{I_v D_v}(t_l) + \Delta N_{I_g D_g}(t_l))
$$

We construct a scenario where a disease outbreak occurs in a small vulnerable population

173

with a true infection-fatality ratio of 10% before spreading to a larger general population with a true infection-fatality ratio of 1%. Because the outbreak spreads through the different populations at different times, the true population infection-fatality ratio varies in time. Figure B.12 shows the latent new cases and new deaths for each subpopulation, as well as the combined latent new cases and new deaths, and the observed new cases and new deaths for this constructed scenario.
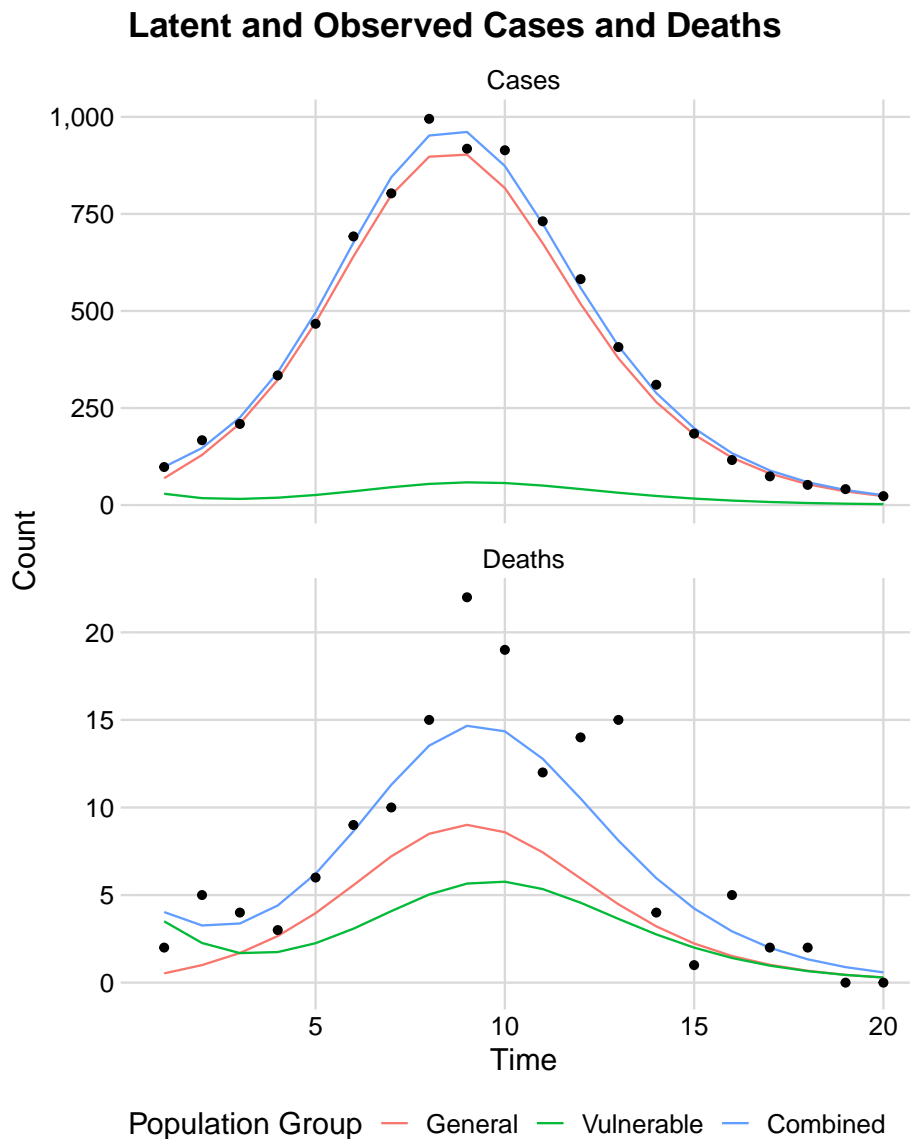

Figure B.12: Latent new cases and deaths for vulnerable and general subpopulations, along with combined latent new cases and new deaths and observed (combined) new cases and new deaths for a simulated scenario.

Now, we fit a semi-parametric model, similar to the one in the main text, to this data. We model $\eta(t)$ with a logit-Guassian Markov random field, as in the main text. The differential equations used for this model presented in (B.2).

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\beta I \frac{S}{N} \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \beta I \frac{S}{N} - \gamma E \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \gamma E - \nu I \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \nu(1 - \eta(t))I \\
\frac{\mathrm{d}D}{\mathrm{d}t} &= \nu\eta(t)I
\end{aligned}
\tag{B.2}
$$

subject to initial conditions $\mathbf{X}(t_0) = \mathbf{x}_0$ and $\mathbf{N}(t_0) = \mathbf{0}$, where $\mathbf{x}_0 = (S_0, E_0, I_0, R_0, D_0)$ are initial compartment counts.

Figures B.13–B.16, demonstrate, that when we fit our semi-parametric model, we can generally fit the data well and recover the true values of the parameters without modeling the two heterogeneous populations.

B.4 Sensitivity analysis

We conducted four sensitivity analyses to see how our results change depending on the specified priors. In each additional analysis, we change only one aspect of the model priors. We perform one analysis where, *a priori*, twice the number of people are initially infected (denoted Half S_0), one with a lower initial basic reproduction number prior (denoted Half $\exp\left(\tilde{R}_{0,1}\right)$), one with a higher initial infection fatality ratio prior (denoted Double expit $(\tilde{\eta}_1)$, and one with a lower initial α prior (denoted Half $\exp(\tilde{\alpha}_1)$. Precise descriptions of the priors used in the sensitivity analyses are presented in Table B.3. Graphical results of the
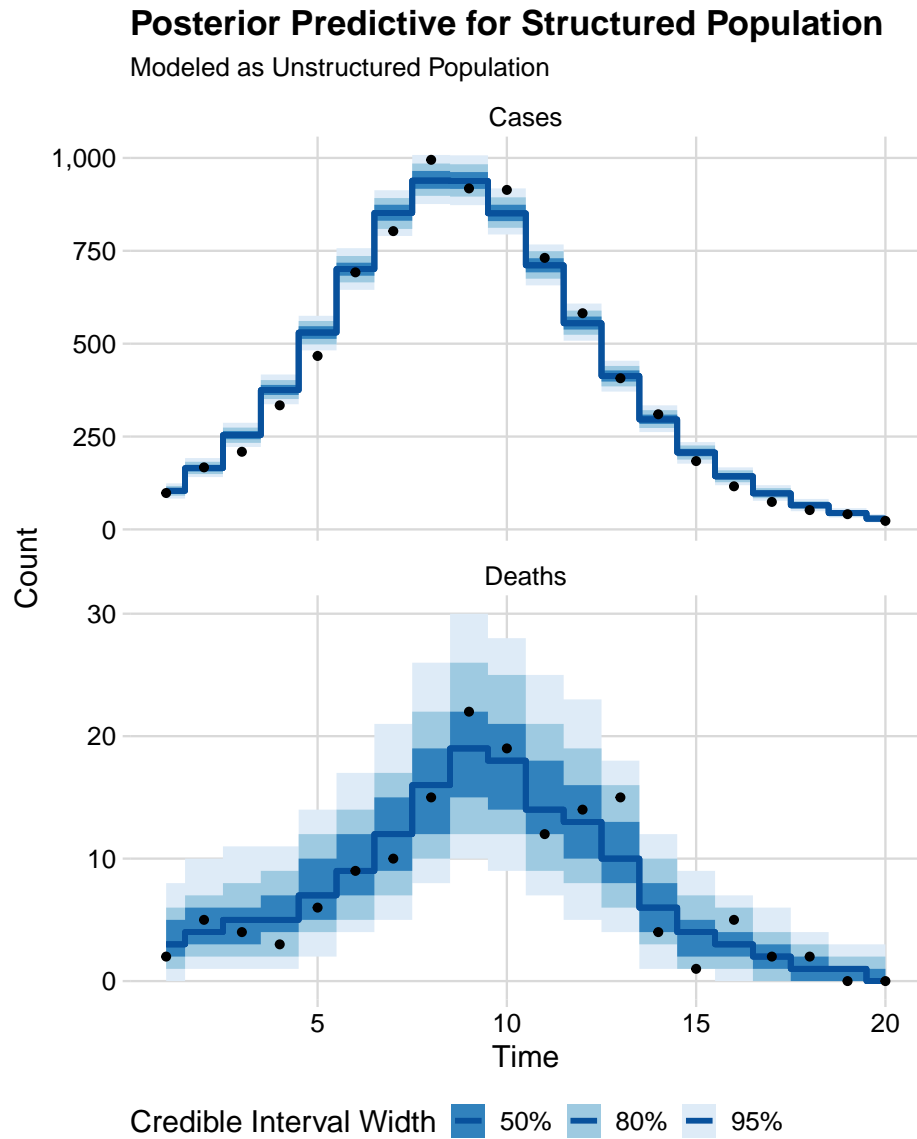
Figure B.13: Posterior predictive distributions for a model with non-parametric IFR fit to a simulated dataset with a heterogeneous population. The case and death data used are shown as black dots.
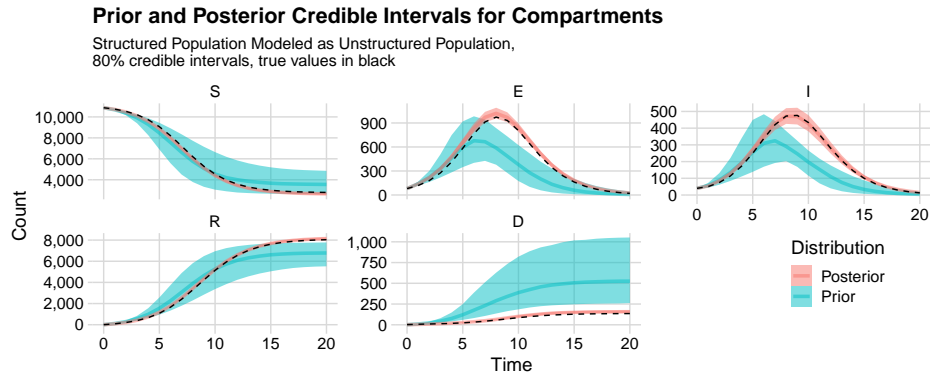
Figure B.14: Prior and Posterior distributions for latent compartments for a model with non-parametric IFR fit to a simulated dataset with a heterogeneous population. The true time-varying parameters are indicated by the dashed line.
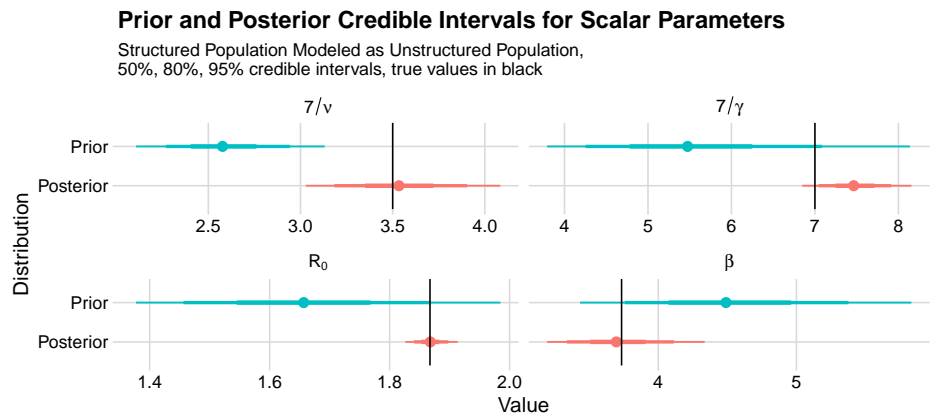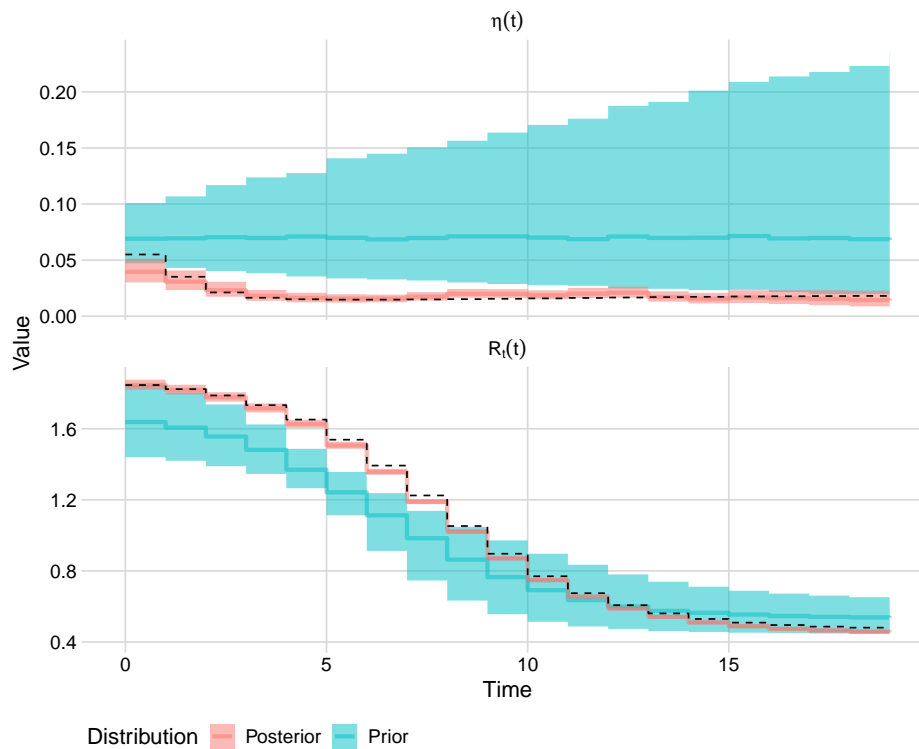

Figure B.15: Prior and posterior distributions for scalar parameters for a model with non-parametric IFR fit to a simulated dataset with a heterogeneous population. The true time-varying parameters are indicated by the vertical line.

Figure B.16: Prior and Posterior distributions for time-varying parameters for a model with non-parametric IFR fit to a simulated dataset with a heterogeneous population. The true time-varying parameters are indicated by the dashed line.

sensitivity analyses are presented in Figures B.17–B.19. We find that our model is typically robust to these alternative priors, and no alternative model leads to substantively different conclusions.

Table B.3: Model parameters and their prior distributions.

Analysis	Parameter	Original Prior Median (95% Interval)	Sensitivity Prior Median (95% Interval)
Half S_0	S_0	0.998 (0.993, 0.999)	0.995 (0.987, 0.998)
Half $\exp\left(\tilde{R}_{0,1}\right)$	$\exp\left(\tilde{R}_{0,1}\right)$	1.000 (0.613, 1.630)	0.500 (0.306, 0.816)
Double $\operatorname{expit}(\tilde{\eta}_1)$	$\operatorname{expit}(\tilde{\eta}_1)$	0.00497 (0.00336, 0.00733)	0.00988 (0.00670, 0.01460)
Half $\exp(\tilde{\alpha}_1)$	Half $\exp(\tilde{\alpha}_1)$	3.86 (3.11, 4.79)	1.93 (1.55, 2.39)

Additionally, we perform analyses where we modify the main model to, one at a time, fix each of the time-varying parameters, R_0, α, and η. As demonstrated in Figure B.20, fixing these parameters has no negative impact on the model's ability to properly fit the test positivity and death data, with each of the models exhibiting nearly identical posterior predictive distributions. However, these modified models do lead to substantially different inferences about the time-varying parameters themselves. This is shown in Figure B.21, where it appears that when one parameter is fixed, the others can become more flexible to still precisely match the observed data. The most dramatic effect is seen when fixing R_0, which leads to drastically different inferences about η and α. In contrast, there appears to be little impact from fixing the infection-fatality ratio, η as constant through time.
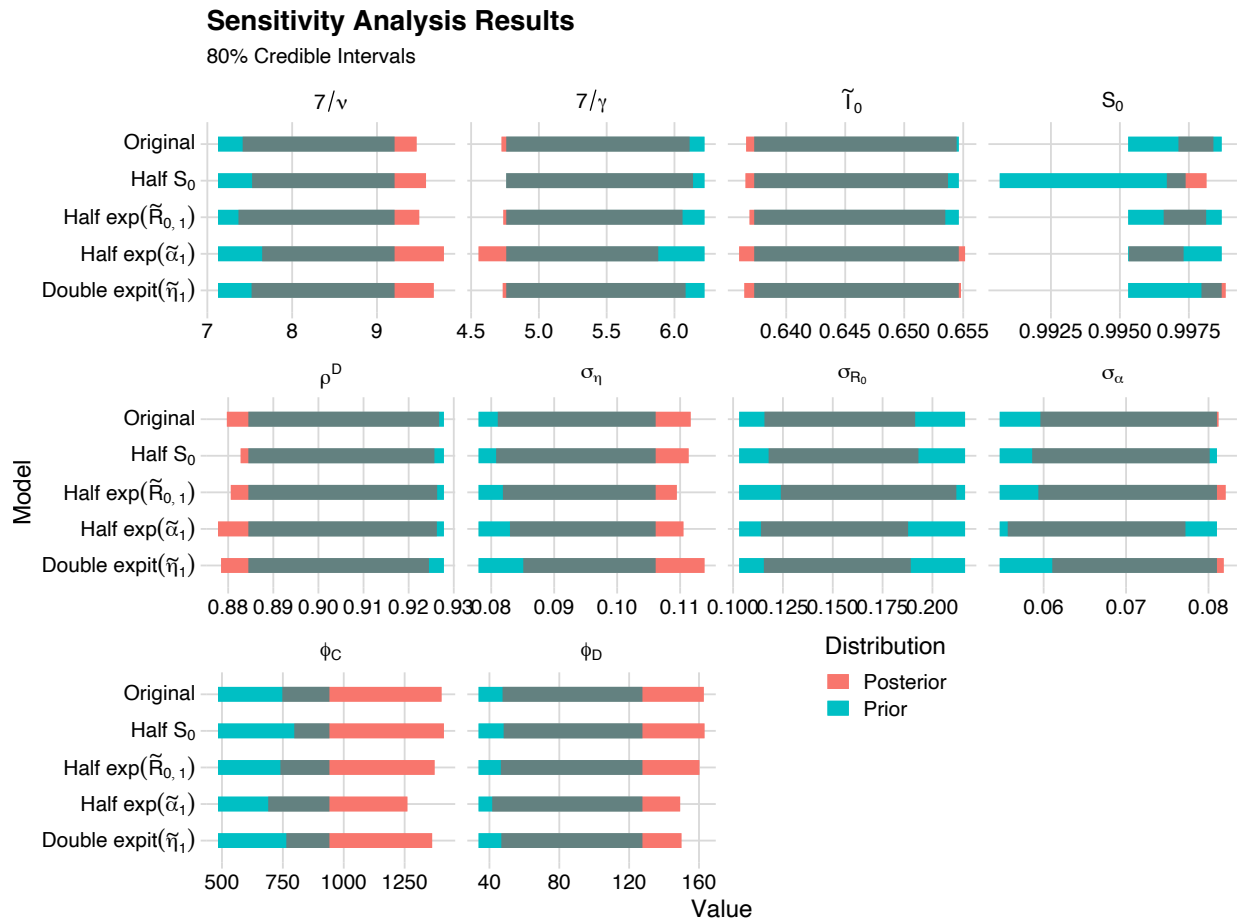
Figure B.17: Prior and posterior 80% credible intervals for scalar parameters from four sensitivity analyses and the original analysis.
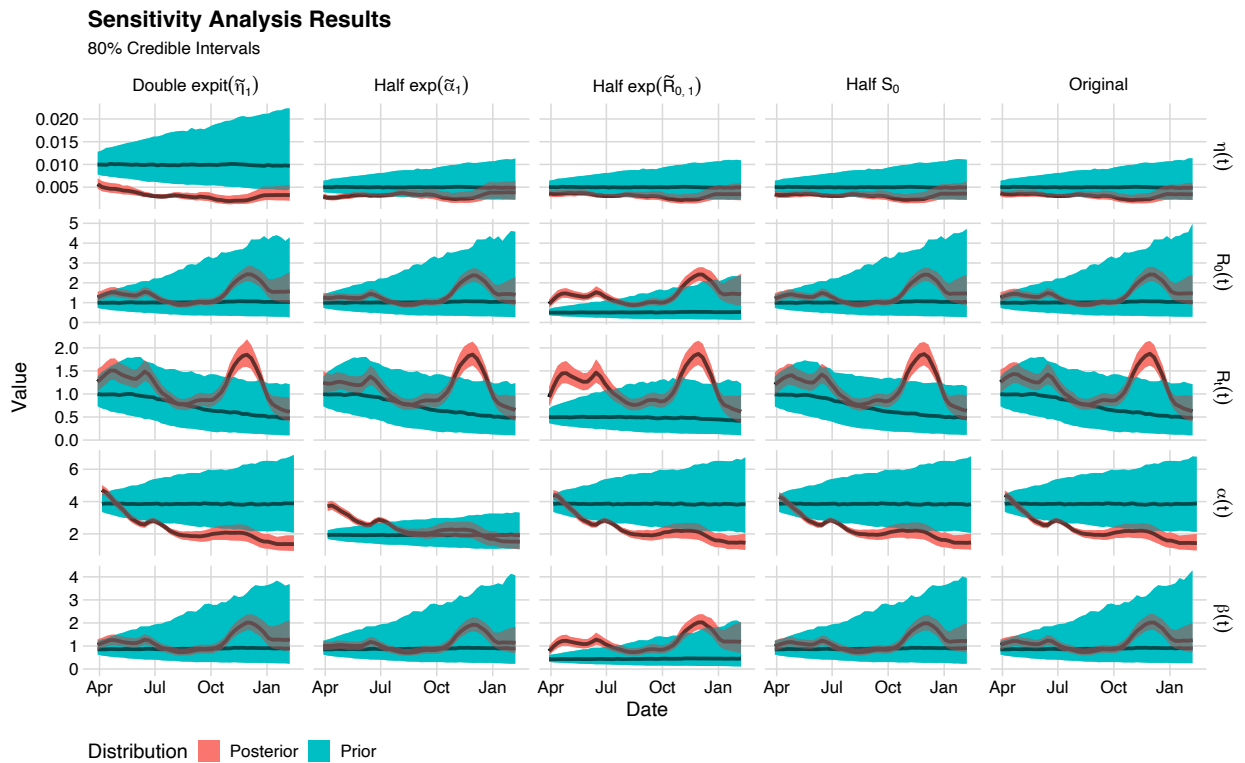
Figure B.18: Prior and posterior 80% credible intervals for time-varying parameters from four sensitivity analyses and the original analysis.
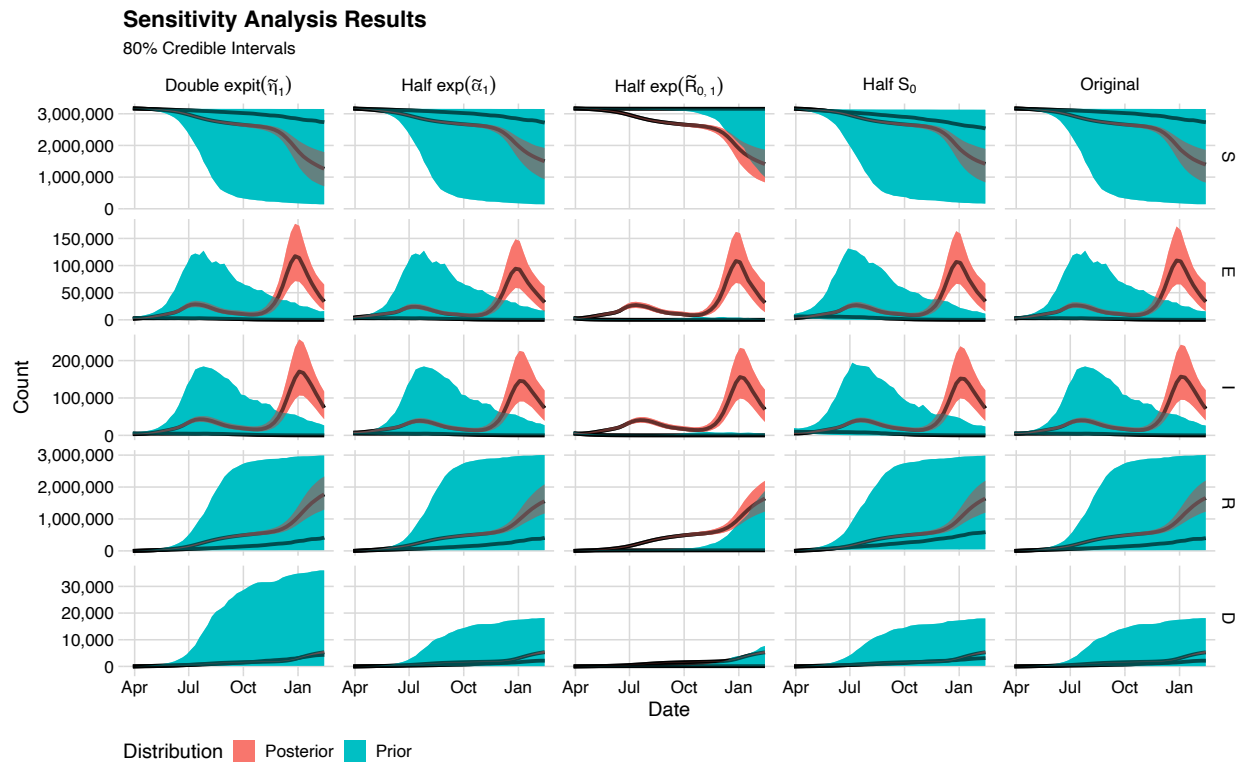
Figure B.19: Prior and posterior 80% credible intervals for time-varying parameters from four sensitivity analyses and the original analysis.
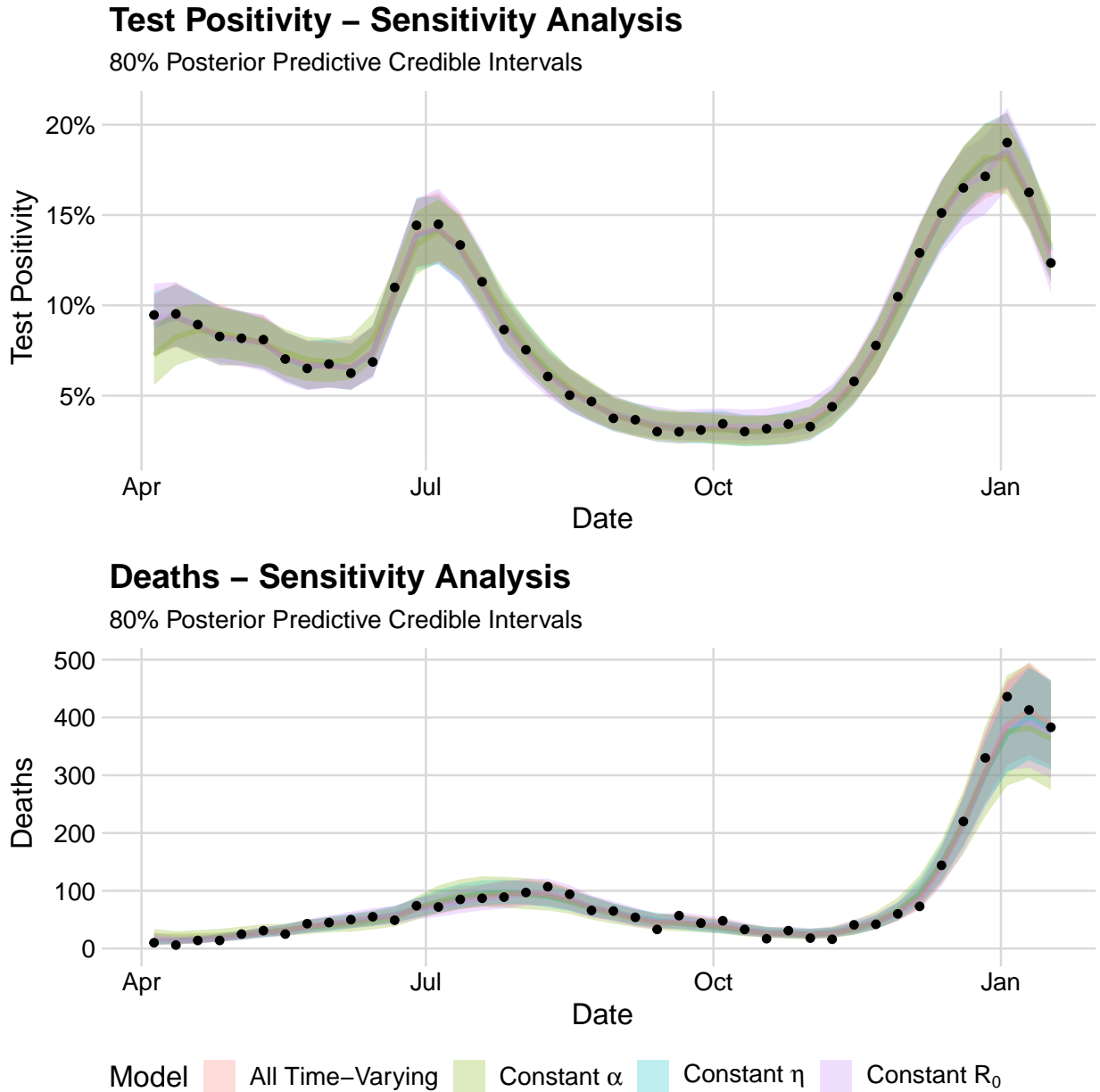
Figure B.20: Posterior predictive distributions when one of the typically time-varying parameters is made to be fixed through time.
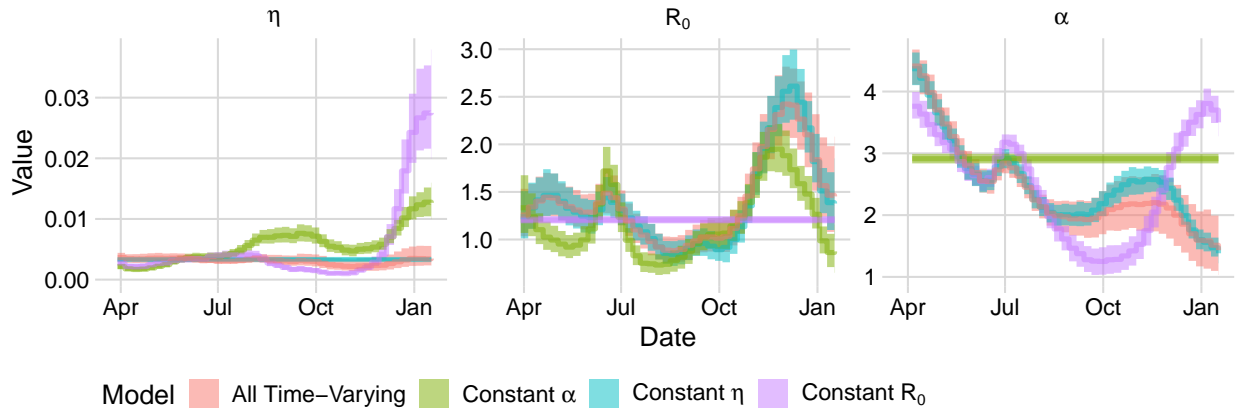
Figure B.21: Posterior inference for time-varying parameters when one of the typically time-varying parameters is made to be fixed through time.

B.5   MCMC Diagnostics

Convergence diagnostics are presented in Tables B.4 and B.5, where $\hat{R}$ is the potential scale reduction factor [Vehtari et al., 2021], and ESS is the effective sample size, both as computed in the posterior R package [Bürkner et al., 2022]. All parameters show potential scale reduction factors between 1 and 1.02, providing no evidence of lack of convergence. Additionally, all model parameters have effective sample sizes of multiple hundreds, which is sufficient for our inferences.

We also produce a trace plot of the log-posterior probability for each chain in Figure B.22, which indicates that each chain explores a region of similar probability.

Table B.4: Convergence diagnostics for scalar parameters for the main model fit to the Orange County data set.

Parameter	$\hat{R}$	ESS
S_0	1.00	992.14
$\tilde{I}_0$	1.01	1128.48
$1/\gamma$	1.01	712.86
$1/\nu$	1.02	907.22
ϕ_D	1.00	876.28
ρ^D	1.00	812.16
ϕ_C	1.00	841.67
σ_{R_0}	1.00	544.21
σ_η	1.00	788.91
σ_α	1.01	567.30

Table B.5: Convergence diagnostics for scalar parameters for the main model fit to the Orange County data set.

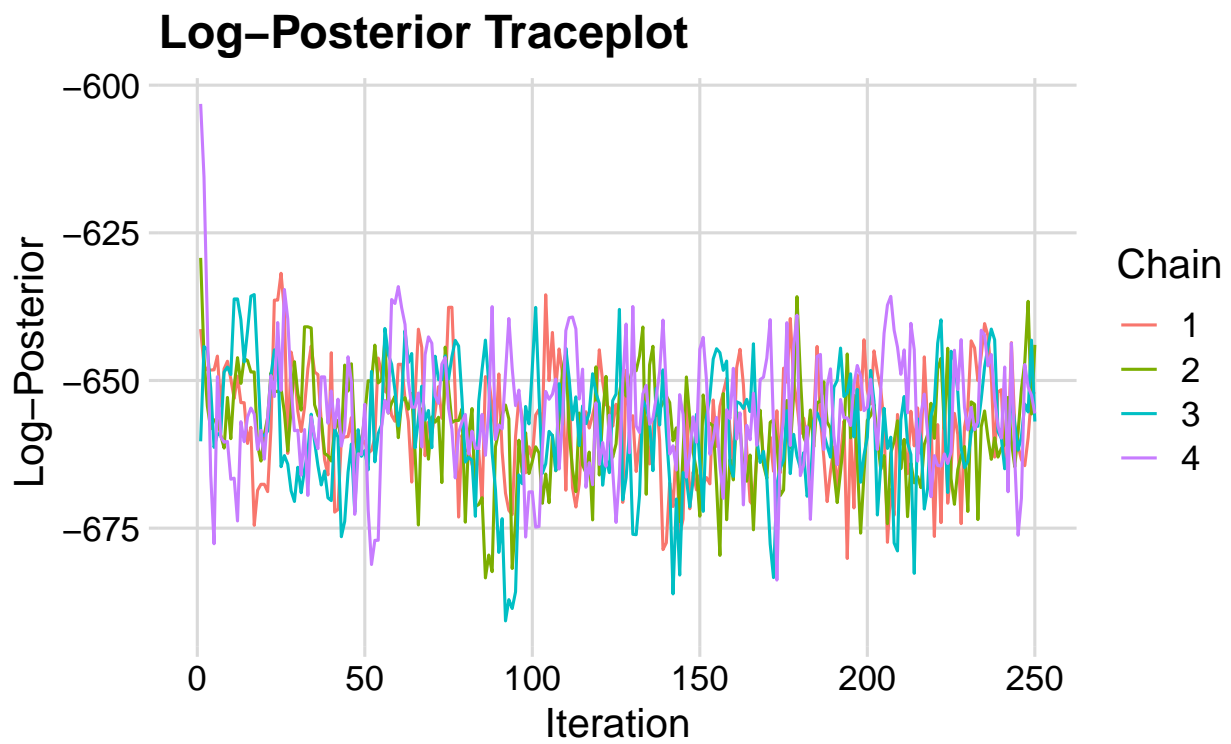| Parameter	Min. $\hat{R}$	Avg. $\hat{R}$	Max. $\hat{R}$	Min. ESS	Avg. ESS	Max. ESS
$\exp(\tilde{\alpha}_t)$	1	1	1.01	630.78	827.20	1088.73
$\exp\left(\tilde{R}_{t,t}\right)$	1	1	1.02	693.98	977.45	1368.80
$\mathrm{expit}(\tilde{\eta}_t)$	1	1	1.02	563.38	793.53	1259.37

Figure B.22: Trace plot of log-posterior probability for the main model fit.

Appendix C

Additional Material for Chapter 5

C.1 Simulation Model

Data for the simulation study in Section 5.3.1 are generated from a two strain model, which is described in the main text. Here, we present the mathematical details of the model presented in Figure 5.3. We first present the rates of transitions between compartments in the model in (C.1), where the interpretation and values of the parameters used in the simulation are given in Table C.1. Using the transition rates in (C.1), we define the ODEs in the model in (C.2), subject to the initial conditions given in Table C.1.

We use the solutions to the system of differential equations to simulate the six time series referenced in Section 5.2.3 according to (C.3)–(C.8). The time series of counts of new sequences from the old variant $\left(\mathbf{V}^O\right)$ and the novel variant $\left(\mathbf{V}^N\right)$ are summed to create the time series of counts of all variants $\left(\mathbf{V}^A\right)$.

The simulated data for the medium takeover speed scenario is plotted in Figure 5.5, with the gray shaded areas indicating the time points for which we produce forecasts. Similar

figures for the other data sets are presented in Figures C.1 and C.2.

$$
\begin{aligned}
\lambda_{S_A E_O}(S_A, I_O) &= \frac{\beta_O}{N} S_A I_O, & \lambda_{R_O E_N}(R_O, I_N) &= \epsilon \frac{\beta_N}{N} R_O I_N, \\
\lambda_{S_A E_N}(S_A, I_N) &= \frac{\beta_N}{N} S_A I_N, & \lambda_{S_N E_N}(S_N, I_N) &= \frac{\beta}{N} S_N I_N, \\
\lambda_{E_O I_O}(E_0) &= \gamma E_O, & \lambda_{E_N I_N}(E_N) &= \gamma E_N, \\
\lambda_{I_O H_O}(I_O) &= \nu\tau I_O, & \lambda_{I_N H_N}(I_N) &= \nu\tau I_N, \\
\lambda_{I_O R_O}(I_O) &= \nu(1-\tau) I_O, & \lambda_{I_N R_N}(I_N) &= \nu(1-\tau) I_N, \\
\lambda_{H_O U_O}(H_O) &= \eta\upsilon H_O, & \lambda_{H_N U_N}(H_N) &= \eta\upsilon H_N, \\
\lambda_{H_O R_O}(H_O) &= \eta(1-\upsilon) H_O, & \lambda_{H_N R_N}(H_N) &= \eta(1-\upsilon) H_N, \\
\lambda_{U_O D_A}(U_O) &= \omega\chi U_O, & \lambda_{U_N D_A}(U_N) &= \omega\chi U_N, \\
\lambda_{U_O R_O}(U_O) &= \omega(1-\chi) U_O, & \lambda_{U_N R_N}(U_N) &= \omega(1-\chi) U_N, \\
\lambda_{R_O S_A}(R_O) &= \kappa R_O, & \lambda_{R_N S_N}(R_N) &= \kappa R_N,
\end{aligned}
\tag{C.1}
$$

Table C.1: Parameters used to simulate data in Section 5.3.1.

Parameter	Interpretation	Value
β_O	Contact rate for old variant	1.5
β_N	Contact rate for novel variant	see Table 5.1
ϵ	Parital immunity to novel variant conferred by old variant	see Table 5.1
$1/\nu$	Mean infectious period duration (weeks)	5 / 7
$1/\gamma$	Mean latent period duration (weeks)	2 / 7
$1/\eta$	Mean hospitalization duration (weeks)	3 / 7
$1/\omega$	Mean ICU stay duration (weeks)	5 / 7
$1/\kappa$	Mean immunity duration (weeks)	20

τ	Infection-hospitalization ratio	0.02
υ	Hospitalization-ICU admission ratio	0.15
χ	ICU-fatality ratio	0.15
ρ^W	Case detection rate	0.2
ρ^Z	Death detection rate	0.9
ρ^V	Sequence detection rate (relative to cases)	0.1
ϕ_W	Over-dispersion for observed new cases	100
ϕ_X	Over-dispersion for observed hospitalizations	100
ϕ_Y	Over-dispersion for observed ICU	100
ϕ_Z	Over-dispersion for observed new deaths	100
ϕ_V	Over-dispersion for observed new sequences	100
$S_{A,0}$	Initial size of population susceptible to all variants	1,899,600
$E_{O,0}$	Initial size of population infected, but not yet infectious with the old variant	100
$I_{O,0}$	Initial size of population infectious with the old variant	300
$R_{O,0}$	Initial size of population recovered from the old variant	300
$S_{N,0}$	Initial size of population susceptible to only the novel variant	0
$E_{N,0}$	Initial size of population infected, but not yet infectious with the novel variant	0
$I_{N,0}$	Initial size of population infectious with the novel variant	0
$R_{N,0}$	Initial size of population recovered from the novel variant	0
$H_{O,0}$	Initial size of population hospitalized with the old variant	0
$H_{N,0}$	Initial size of population hospitalized with the novel variant	0
$U_{O,0}$	Initial size of population in the ICU with the old variant	0
$U_{N,0}$	Initial size of population in the ICU with the novel variant	0
$D_{A,0}$	Initial size of population deceased due to infection from all variants	0

$$\frac{\mathrm{d}S_A}{\mathrm{d}t} = \lambda_{R_O S_A}(R_O) - \left(\lambda_{S_A E_O}(S_A, I_O) + \lambda_{S_A E_N}(S_A, I_N)\right),$$

$$\frac{\mathrm{d}E_O}{\mathrm{d}t} = \lambda_{S_A E_O}(S_A, I_O) - \lambda_{E_O I_O}(E_0),$$

$$\frac{\mathrm{d}I_O}{\mathrm{d}t} = \lambda_{E_O I_O}(E_0) - \left(\lambda_{I_O H_O}(I_O) + \lambda_{I_O R_O}(I_O)\right),$$

$$\frac{\mathrm{d}H_O}{\mathrm{d}t} = \lambda_{I_O H_O}(I_O) - \left(\lambda_{H_O U_O}(H_O) + \lambda_{H_O R_O}(H_O)\right),$$

$$\frac{\mathrm{d}U_O}{\mathrm{d}t} = \lambda_{H_O U_O}(H_O) - \left(\lambda_{U_O D_A}(U_O) + \lambda_{U_O R_O}(U_O)\right),$$

$$\frac{\mathrm{d}R_O}{\mathrm{d}t} = \lambda_{I_O R_O}(I_O) + \lambda_{H_O R_O}(H_O) + \lambda_{U_O R_O}(U_O) - \left(\lambda_{R_O S_A}(R_O) + \lambda_{R_O E_N}(R_O, I_N)\right),$$

$$\frac{\mathrm{d}S_N}{\mathrm{d}t} = \lambda_{R_N S_N}(R_N) - \lambda_{S_N E_N}(S_N, I_N), \tag{C.2}$$

$$\frac{\mathrm{d}E_N}{\mathrm{d}t} = \lambda_{S_N E_N}(S_N, I_N) + \lambda_{R_O E_N}(R_O, I_N) + \lambda_{S_N E_N}(S_N, I_N) - \lambda_{E_N I_N}(E_N)$$

$$\frac{\mathrm{d}I_N}{\mathrm{d}t} = \lambda_{E_N I_N}(E_0) - \left(\lambda_{I_N H_N}(I_N) + \lambda_{I_N R_N}(I_N)\right),$$

$$\frac{\mathrm{d}H_N}{\mathrm{d}t} = \lambda_{I_N H_N}(I_N) - \left(\lambda_{H_N U_N}(H_N) + \lambda_{H_N R_N}(H_N)\right),$$

$$\frac{\mathrm{d}U_N}{\mathrm{d}t} = \lambda_{H_N U_N}(H_N) - \left(\lambda_{U_N D_A}(U_N) + \lambda_{U_N R_N}(U_N)\right),$$

$$\frac{\mathrm{d}R_N}{\mathrm{d}t} = \lambda_{I_N R_N}(I_N) + \lambda_{H_N R_N}(H_N) + \lambda_{U_N R_N}(U_N) - \lambda_{R_N S_N}(R_N),$$

$$\frac{\mathrm{d}D_A}{\mathrm{d}t} = \lambda_{U_O D_A}(U_O) + \lambda_{U_N D_A}(U_N)$$

$$W_l \sim \text{NegativeBinomial}\left(\mu_l^W = \rho^W \cdot N \cdot \left(\Delta N_{E_O I_O}(t_l) + \Delta N_{E_N I_N}(t_l)\right),\right.$$
$$\left. \left(\sigma_l^W\right)^2 = \mu_l^W\left(1 + \mu_l^W / \phi_W\right)\right) \tag{C.3}$$

$$X_l \sim \text{NegativeBinomial}\left(\mu_l^X = N \cdot \left(H_O(t_l) + H_N(t_l)\right), \left(\sigma_l^X\right)^2 = \mu_l^X\left(1 + \mu_l^X / \phi_X\right)\right) \tag{C.4}$$

$$Y_l \sim \text{NegativeBinomial}\left(\mu_l^Y = N \cdot (U_O(t_l) + U_N(t_l)), \left(\sigma_l^Y\right)^2 = \mu_l^Y\left(1 + \mu_l^Y/\phi_Y\right)\right) \quad \text{(C.5)}$$

$$Z_l \sim \text{NegativeBinomial}\left(\mu_l^Z = \rho^Z \cdot N \cdot \left(\Delta N_{U_O D_O}(t_l) + \Delta N_{U_N D_N}(t_l)\right),\right.$$
$$\left.\left(\sigma_l^Z\right)^2 = \mu_l^Z\left(1 + \mu_l^Z/\phi_Z\right)\right) \quad \text{(C.6)}$$

$$V_l^O \sim \text{NegativeBinomial}\left(\mu_l^{V^O} = \mu_l^W \cdot (1 - \delta(t_l)) \cdot \rho^V, \left(\sigma_l^Z\right)^2 = \mu_l^{V^O}\left(1 + \mu_l^{V^O}/\phi_V\right)\right) \quad \text{(C.7)}$$

$$V_l^N \sim \text{NegativeBinomial}\left(\mu_l^{V^N} = \mu_l^W \cdot \delta(t_l) \cdot \rho^V, \left(\sigma_l^Z\right)^2 = \mu_l^{V^O}\left(1 + \mu_l^{V^N}/\phi_V\right)\right) \quad \text{(C.8)}$$
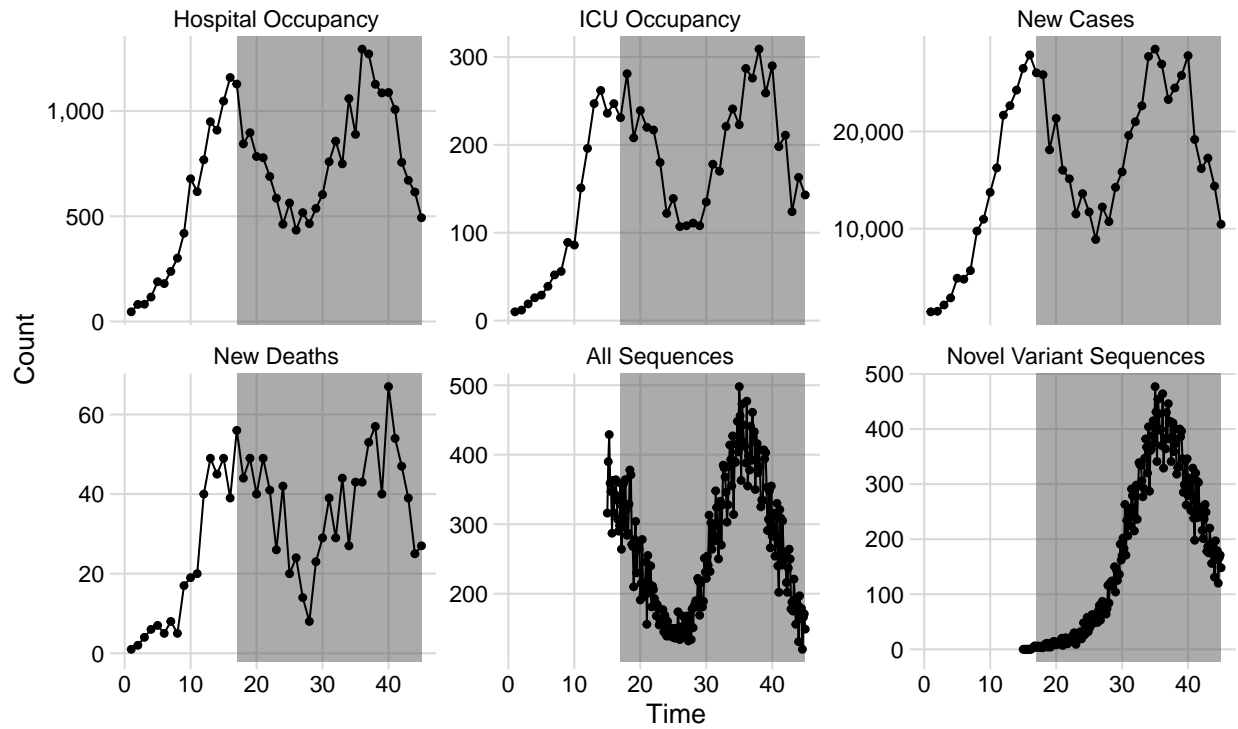
Figure C.1: Simulated data set for the slow takeover speed scenario. The gray shaded areas indicate the time points for which we create forecasts
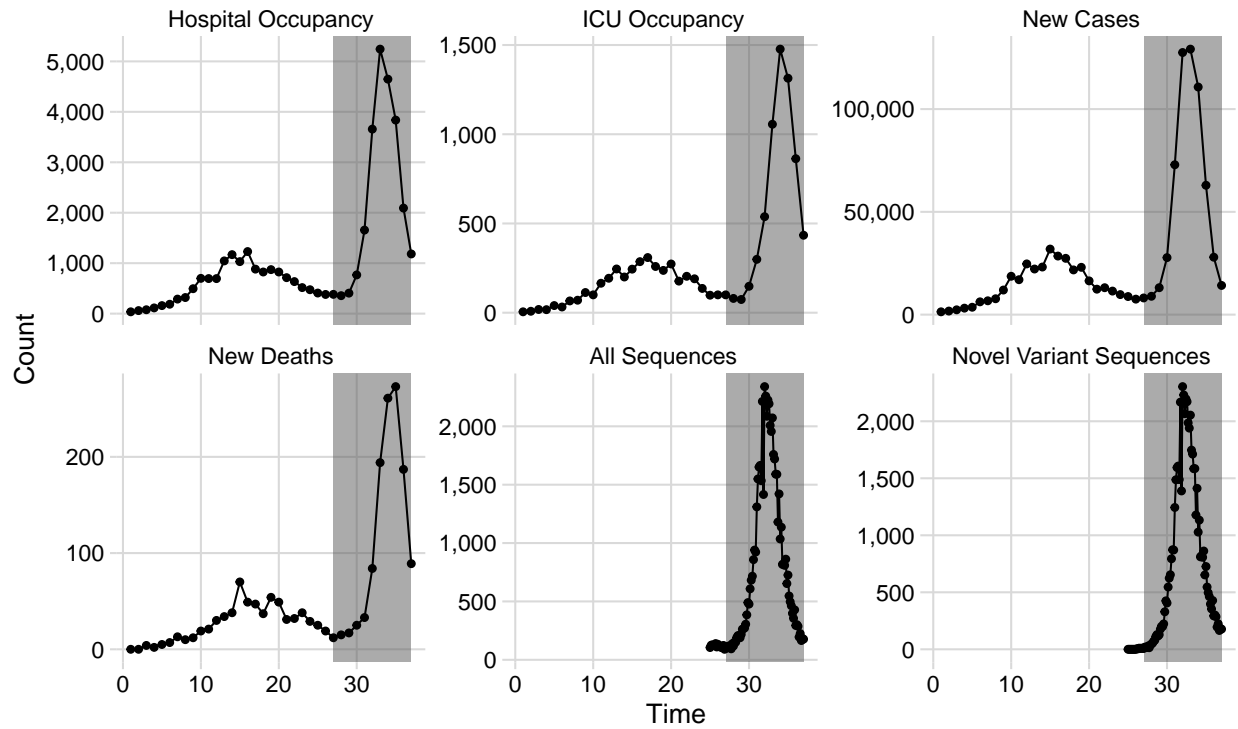
Figure C.2: Simulated data set for the fast takeover speed scenario. The gray shaded areas indicate the time points for which we create forecasts

C.2 Additional simulation study results

Table C.2: Priors used in simulation study in Section 5.3.1. In model 1, $R_0(t)$ is a priori modeled as a GMRF and $1/\kappa(t)$ is constant. In model 2, $R_0(t)$ is constant and $1/\kappa(t)$ is a priori modeled as a GMRF. In model 3, $R_0(t)$ is constant and $1/\kappa(t)$ is a function of the proportion of infectious individuals infected with the novel variant shown in (5.12).

Parameter	Interpretation	Prior	Prior Median (95% Interval)	Models
$1/\gamma$	mean latent period duration (weeks)	Log-Normal(-1.25, 0.04)	0.286 (0.193, 0.423)	1,2,3
$1/\nu$	Mean infectious period duration (weeks)	Log-Normal(-0.336, 0.04)	0.714 (0.483, 1.060)	1,2,3
$1/\eta$	mean hospital-ization duration (weeks)	Log-Normal(-0.847, 0.04)	0.429 (0.290, 0.634)	1,2,3
$1/\omega$	mean ICU stay duration (weeks)	Log-Normal(-0.336, 0.04)	0.714 (0.483, 1.060)	1,2,3
τ	Infection-hospitalization ratio	Logit-Normal(-3.89, 0.04)	0.0200 (0.0136, 0.0293)	1,2,3

v	Hospitalization-ICU admission ratio	Logit-Normal(-1.73, 0.04)	0.150 (0.107, 0.207)	1,2,3
χ	ICU-fatality ratio	Logit-Normal(-1.73, 0.04)	0.150 (0.107, 0.207)	1,2,3
$1/\sqrt{\phi_W}$	over-dispersion parameter for observed new cases	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	1,2,3
$1/\sqrt{\phi_X}$	over-dispersion parameter for observed hospital occupancy	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	1,2,3
$1/\sqrt{\phi_Y}$	over-dispersion parameter for observed ICU occupancy	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	1,2,3
$1/\sqrt{\phi_Z}$	over-dispersion parameter for observed new deaths	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	1,2,3
ρ^W	case detection rate	Logit-Normal(-1.39, 0.04)	0.200 (0.145, 0.270)	1,2,3

ρ^Z	death detection rate	Logit-Normal(2.2, 0.04)	0.900 (0.859, 0.930)	1,2,3
$E(0)$	Initial proportion of infected, but not yet infectious population	Logit-Normal(-8.52, 0.04)	0.000200 (0.000135, 0.000296)	1,2,3
$I(0)$	Initial proportion of infectious population	Logit-Normal(-7.82, 0.04)	0.000400 (0.000270, 0.000592)	1,2,3
$R(0)$	Initial proportion of immune population	Logit-Normal(-1.39, 0.16)	0.200 (0.102, 0.354)	1,2,3
$1/\kappa$	immunity duration	Log-Normal(3, 0.04)	20.0 (13.5, 29.6)	1
$\exp\left(\tilde{R}_{0,1}\right)$	initial basic reproduction number	Log-Normal(0.405, 0.04)	1.50 (1.01, 2.22)	1
σ_{R_0}	standard deviation for basic reproduction number GMRF	Log-Normal(-2, 0.01)	0.135 (0.111, 0.165)	1

$\exp\left(1/\tilde{\kappa}_1\right)$	initial immunity duration	Log-Normal(3, 0.04)	20.0 (13.5, 29.6)	2
σ_κ	standard deviation for immunity duration GMRF	Log-Normal(-2, 0.01)	0.135 (0.111, 0.165)	2
R_0	basic reproduction number	Log-Normal(0.405, 0.04)	1.50 (1.01, 2.22)	2,3
$\exp\left(\alpha_0\right)$	maximum immunity duration	Log-Normal(3, 0.0025)	20.0 (18.1, 22.1)	3
α_1^*	proportion of maximum immunity duration that is minimum immunity duration	Logit-Normal(-1.1, 0.09)	0.250 (0.156, 0.375)	3

α_2	shape of transition from maximum immunity duration to minimum immunity duration	Log-Normal(0, 0.25)	1.000 (0.375, 2.660)	3
ζ	time offset (weeks) for relationship between immunity duration $1/\kappa(t)$ and proportion of novel variant $\delta(t)$	Log-Normal(0, 0.64)	1.000 (0.208, 4.800)	3
ι_1^*	time novel variant takes to grow from 1% to 99% of the infectious population (weeks)	Log-Normal(2.56, 0.16)	13.00 (5.94, 28.50)	3
ι_0	proportion of novel variant at t^*	Logit-Normal(-2.94, 0.25)	0.0500 (0.0194, 0.1230)	3

$1/\sqrt{\phi_V}$	over-dispersion parameter for observed new sequences of new variant	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	3

Forecast Comparison for Hospital Occupancy

Slow Takeover Speed Data
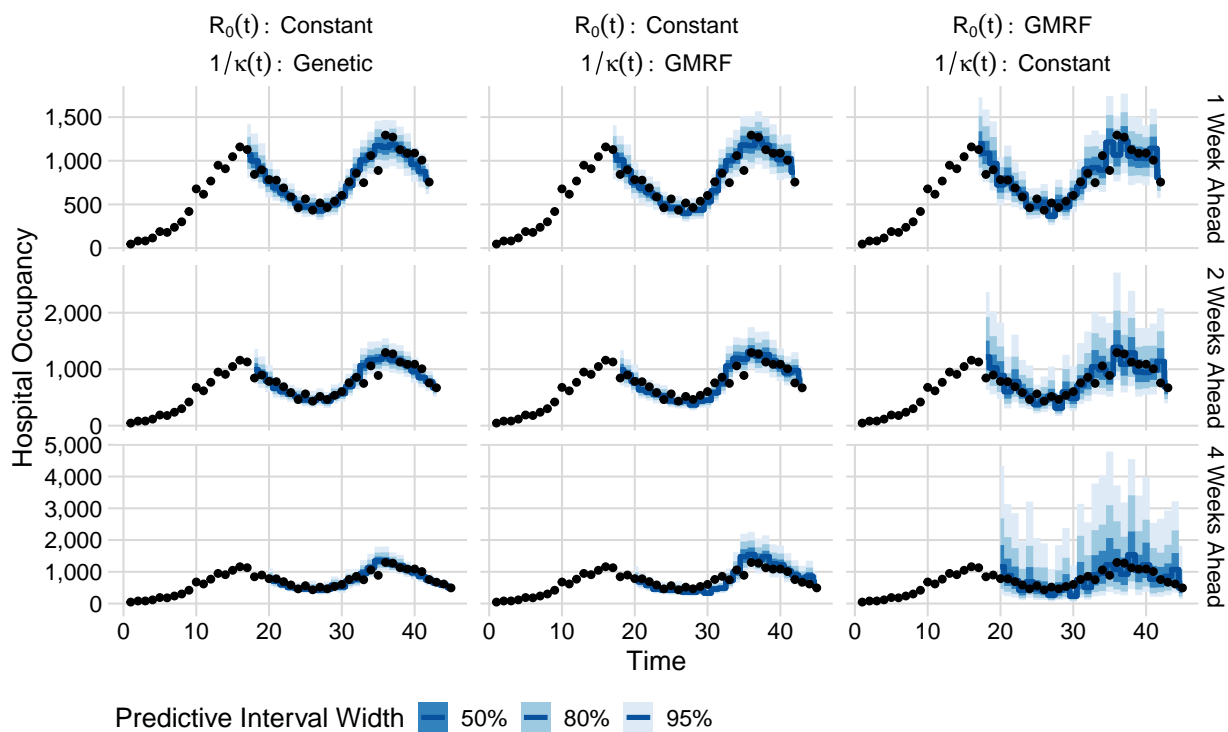
Figure C.3: Hospital occupancy forecasts from three models at 1, 2, and 4-week forecast horizons for the simulated slow takeover speed data.
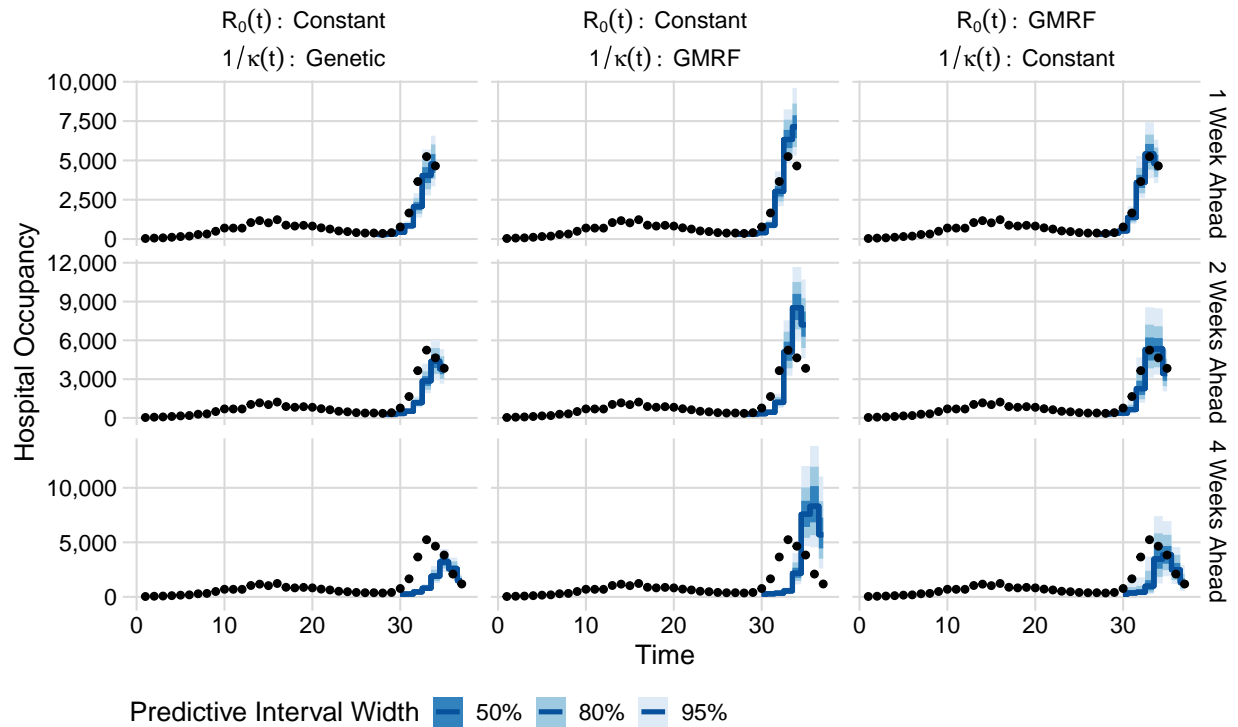
Figure C.4: Hospital occupancy forecasts from three models at 1, 2, and 4-week forecast horizons for the simulated fast takeover speed data.
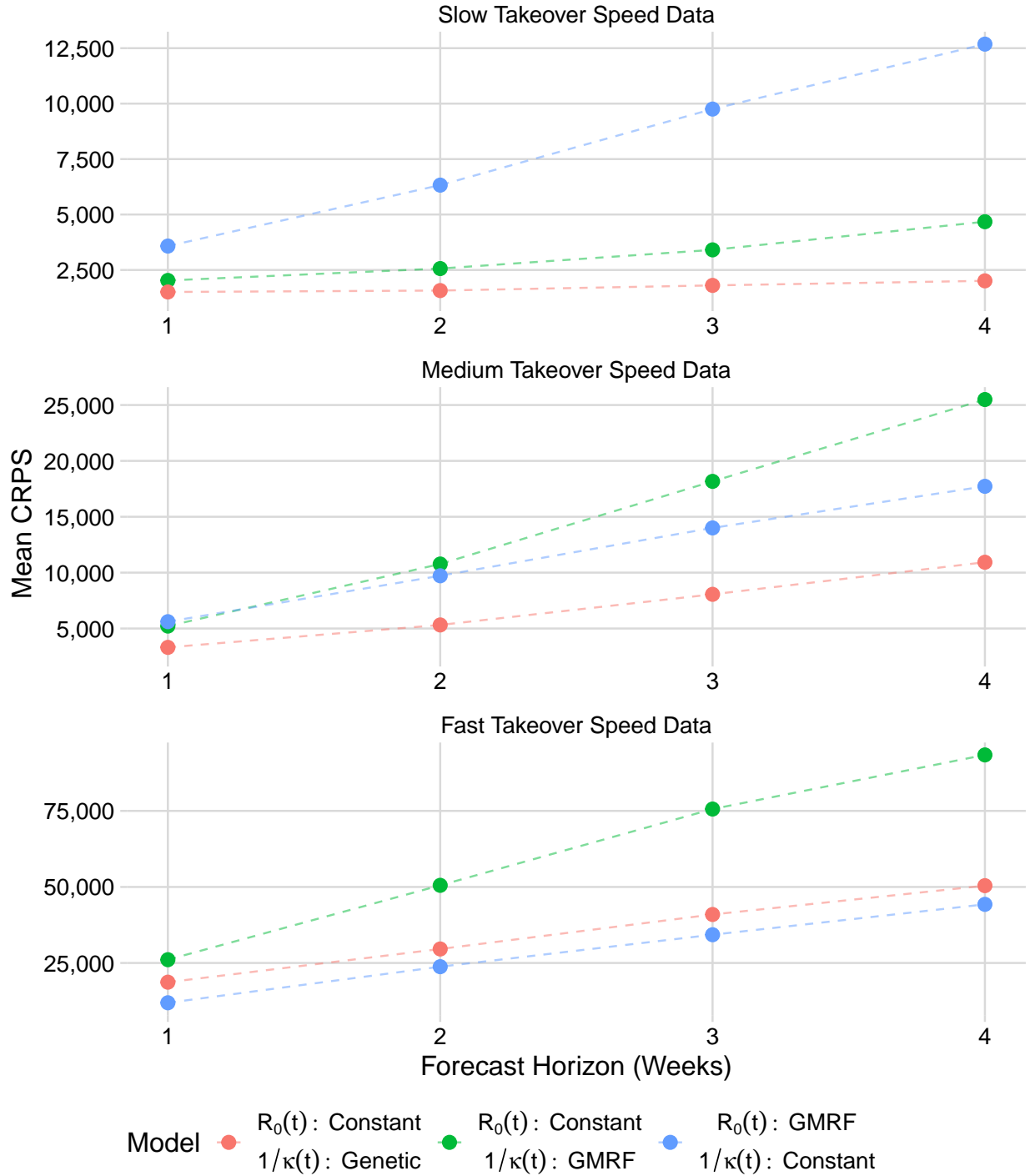
Figure C.5: CRPS summaries for new cases forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.
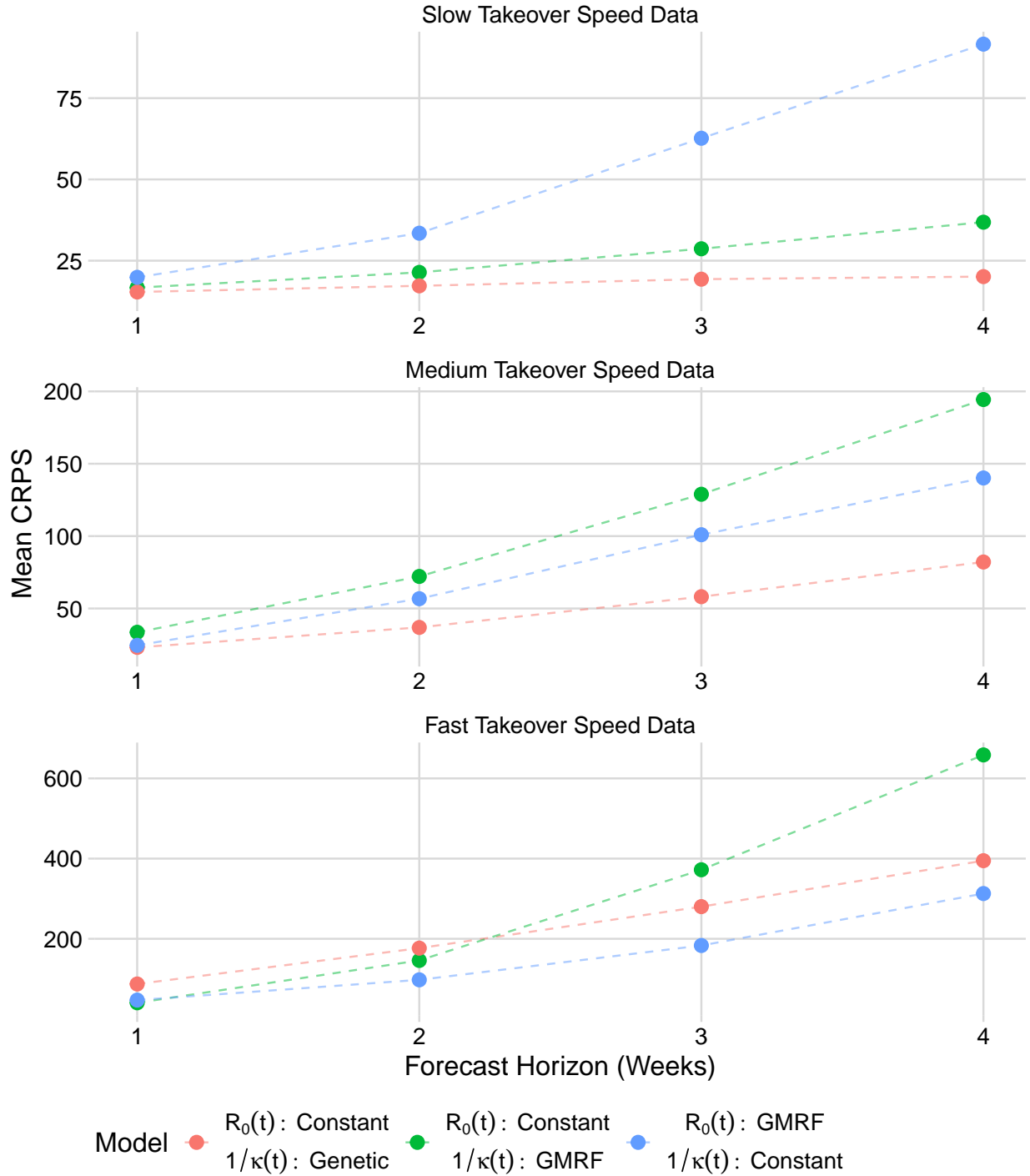
Figure C.6: CRPS summaries for ICU occupancy forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.
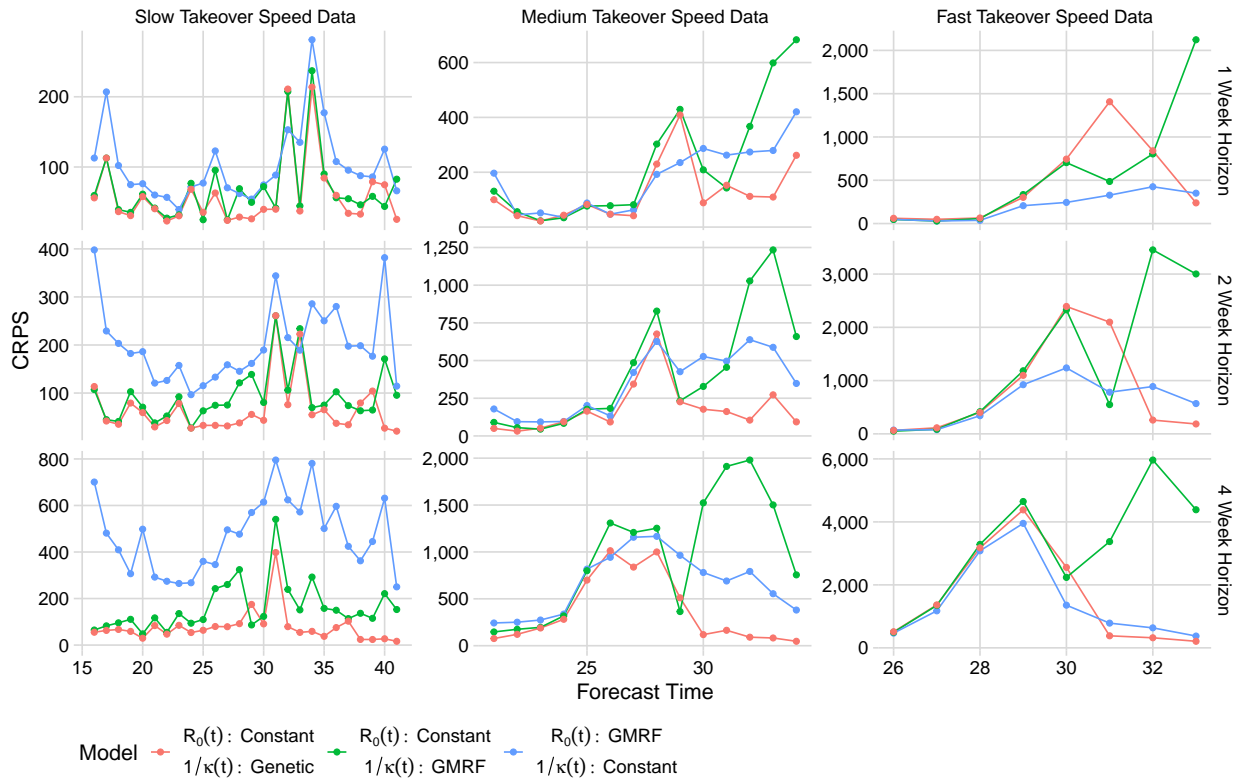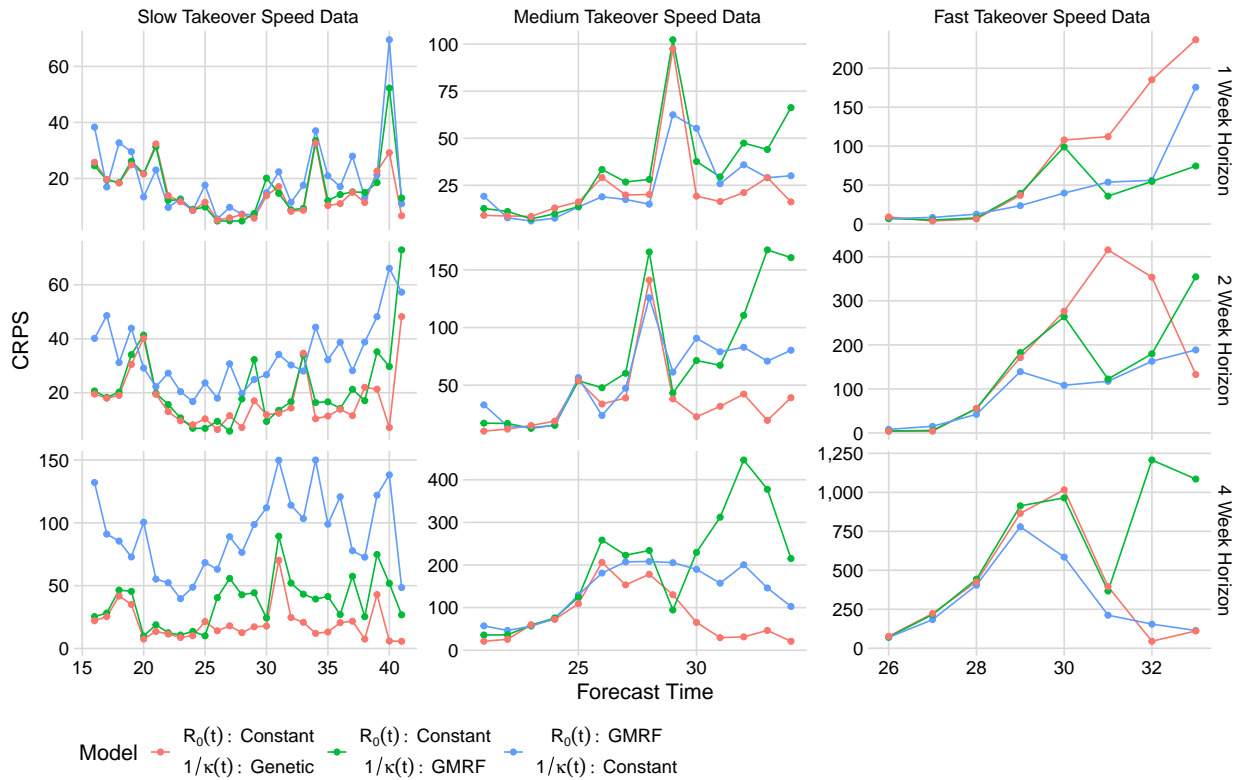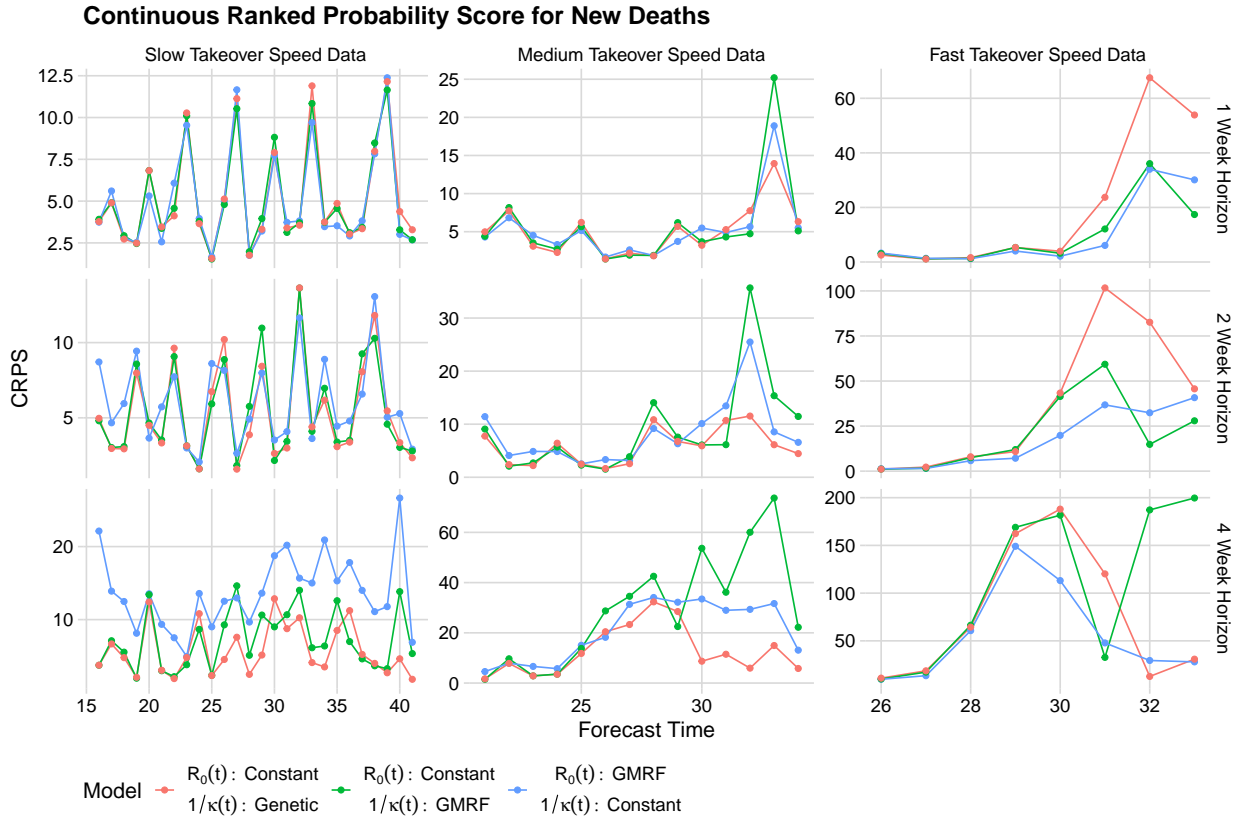
Continuous Ranked Probability Score for New Deaths
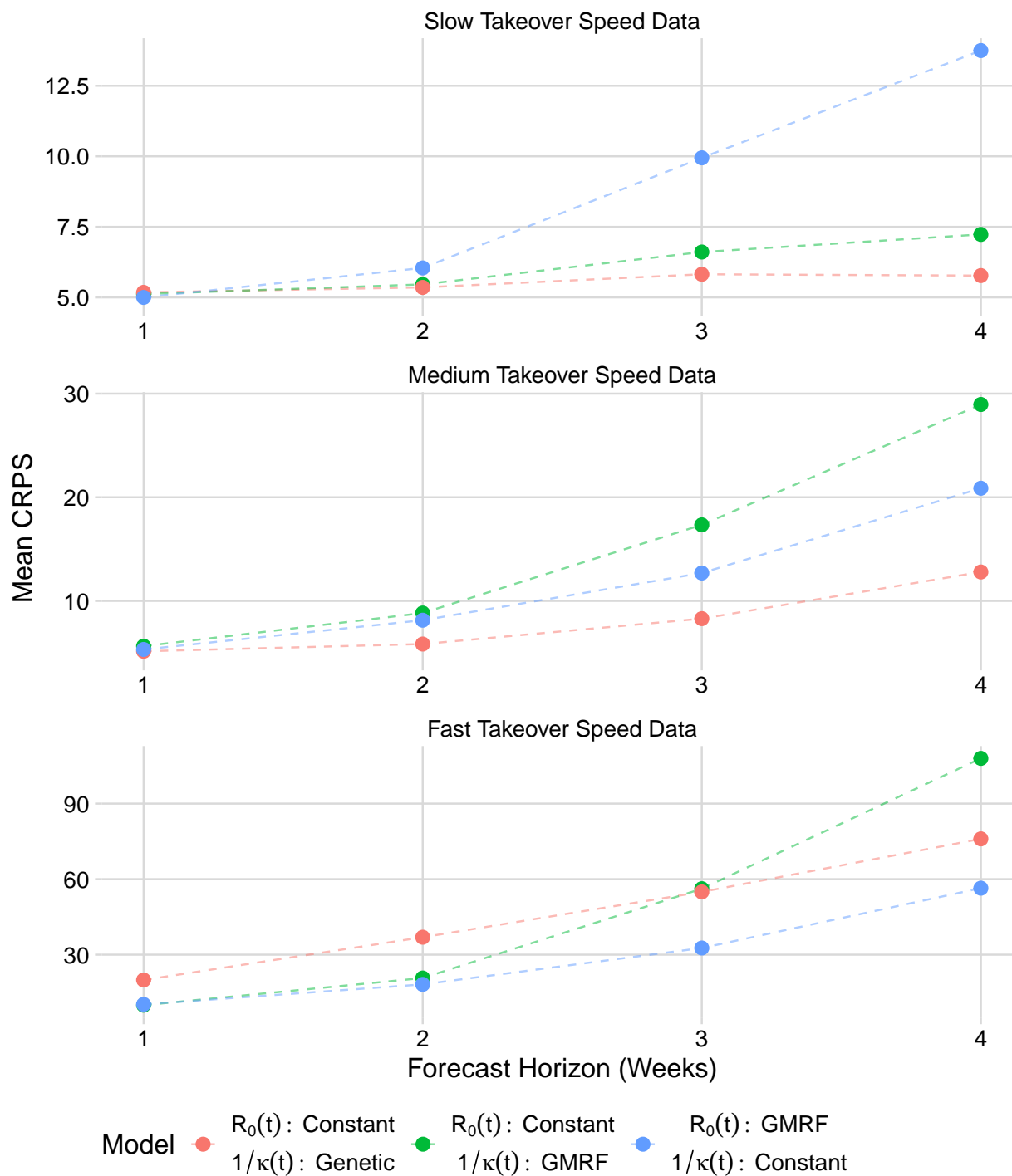
Figure C.7: CRPS summaries for new deaths forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.
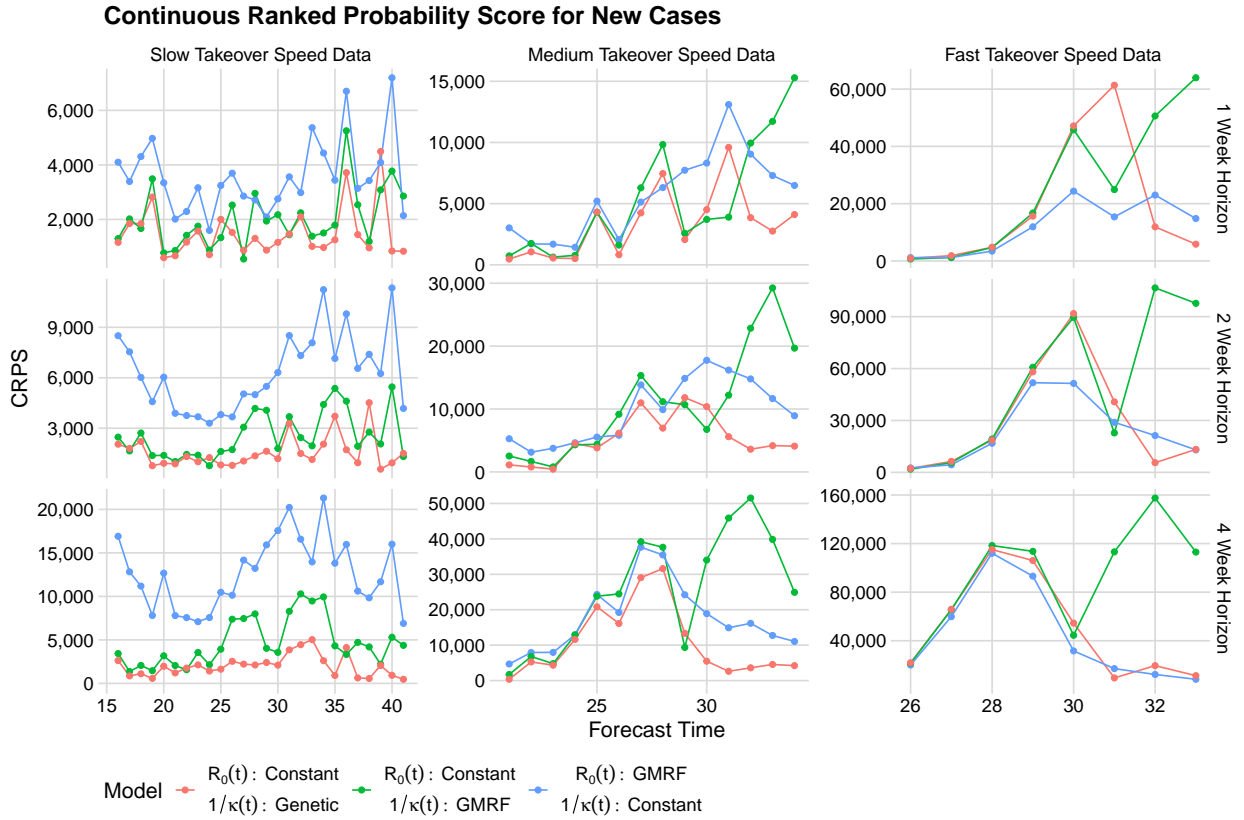
Figure C.8: Individual CRPS for new cases forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.

Figure C.9: Individual CRPS for hospital occupancy forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.

Figure C.10: Individual CRPS for ICU occupancy forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.

Figure C.11: Individual CRPS for new deaths forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.
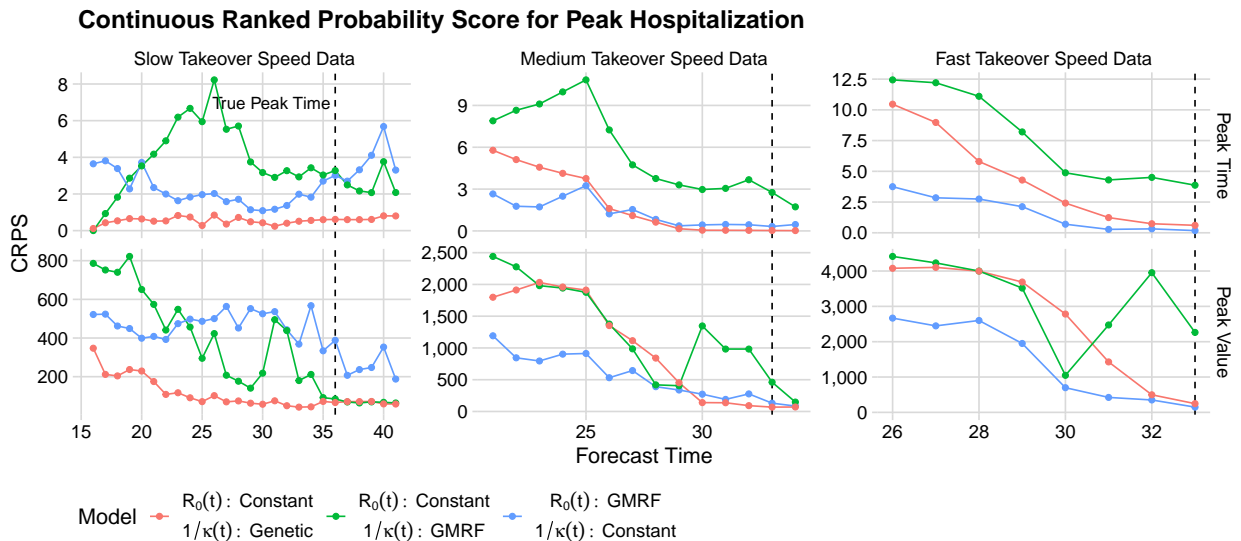
Figure C.12: Caption

C.3 Simulation study sensitivity analysis

We conducted a sensitivity analysis wherein we modified the prior for the anticipated speed of the novel variant takeover. We present the different priors used in Table C.3. All other priors are those relevant to model 3 in Table C.2. Based on the figure in this section, we found our model to be robust to these prior modifications.

Table C.3: Priors that differ from those in Table C.2 in simulation study sensitivity analysis.

Parameter	Interpretation	Prior	Prior Median (95% Interval)	Prior
ι_1^*	time novel variant takes to grow from 1% to 99% of the infectious population (weeks)	Log-Normal(1.95, 0.04)	7.00 (4.73, 10.40)	Fast
ι_1^*	time novel variant takes to grow from 1% to 99% of the infectious population (weeks)	Log-Normal(2.56, 0.04)	13.00 (8.78, 19.20)	Medium
ι_1^*	time novel variant takes to grow from 1% to 99% of the infectious population (weeks)	Log-Normal(3.18, 0.04)	24.0 (16.2, 35.5)	Slow

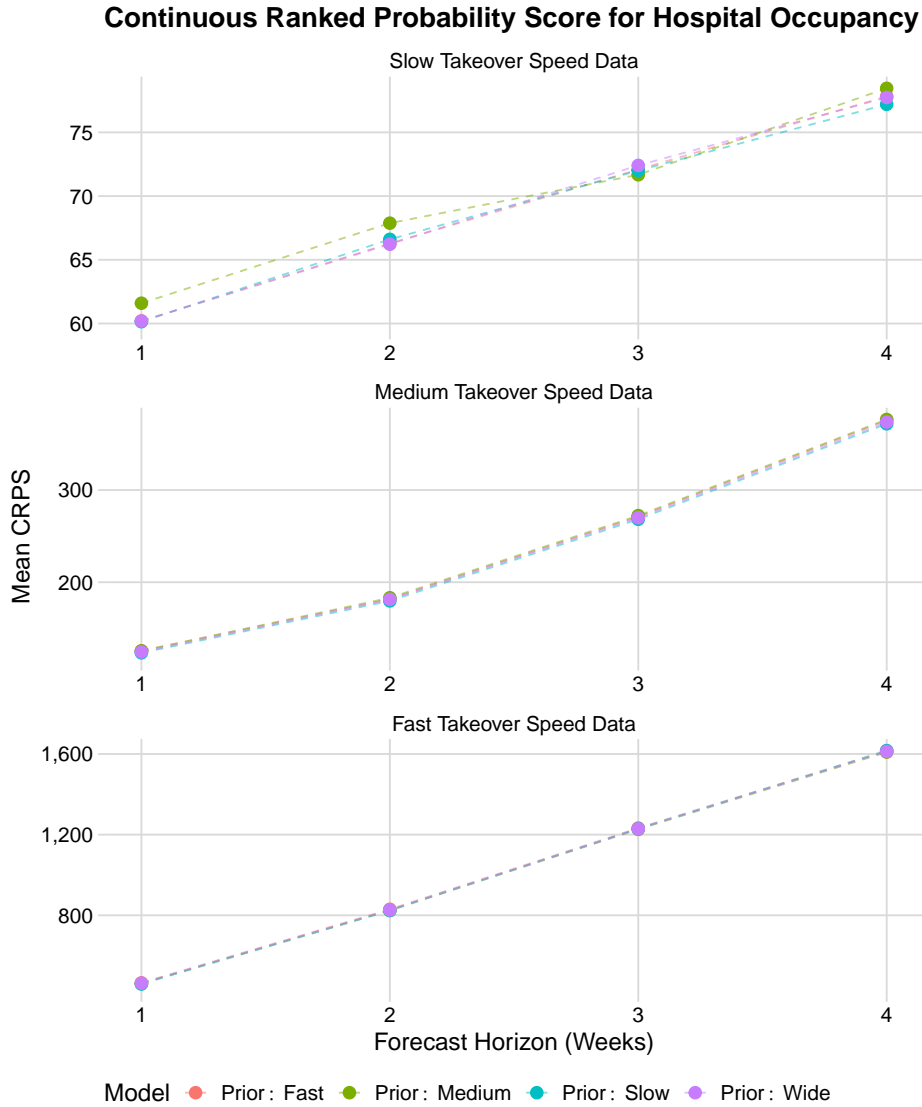| ι_1^* | time novel variant takes to grow from 1% to 99% of the infectious population (weeks) | Log-Normal(2.56, 0.16) | 13.00 (5.94, 28.50) | Wide |

Figure C.13: CRPS summaries for hospital occupancy forecasts at 1, 2, and 4-week horizons for three simulated data sets. Lower CRPS is better.
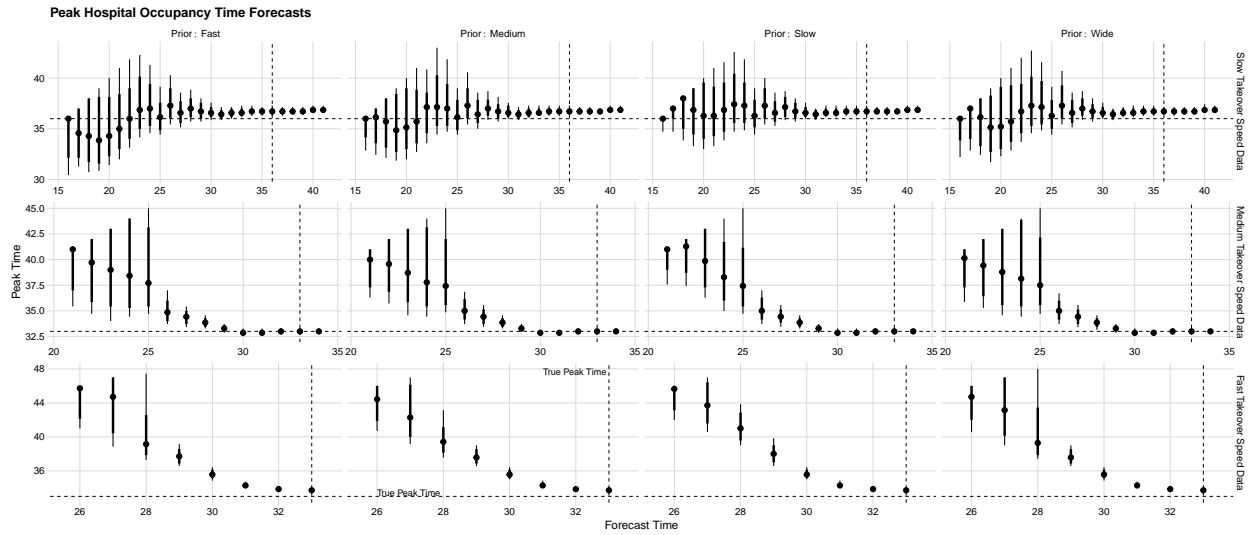
Figure C.14: Posterior predictive intervals for the time at which hospital occupancy reaches its maximum in three simulated data sets. Dots indicate the median of the predictive distribution, while the thick and thin lines represent central 80% and 95% intervals, respectively. Horizontal and vertical dashed lines indicate the true peak hospitalization time.
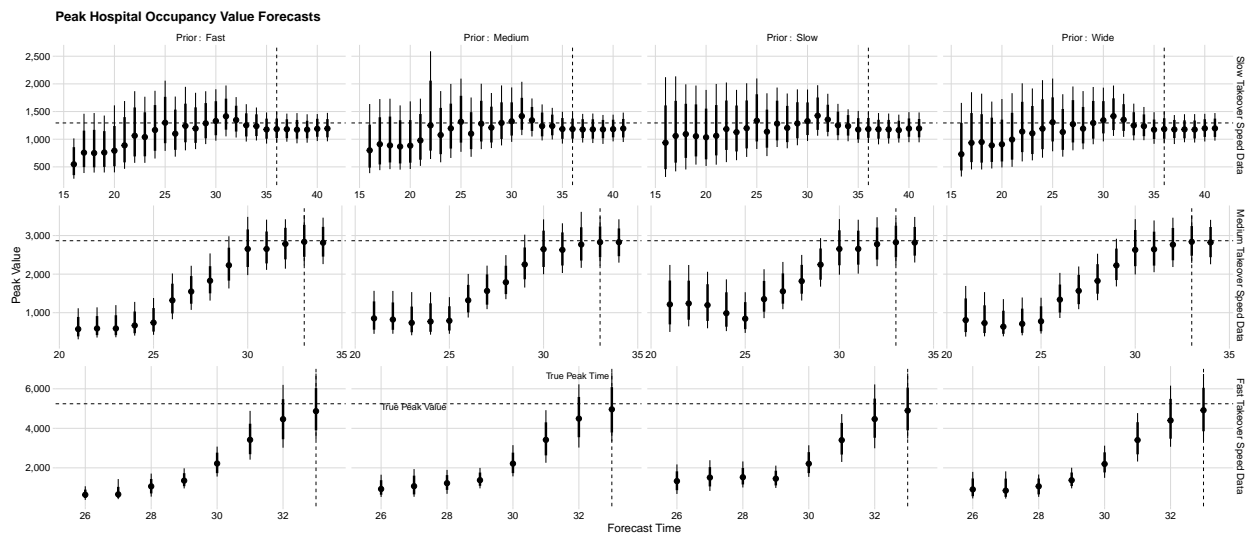

Figure C.15: Posterior predictive intervals for the maximum hospital occupancy in three simulated data sets. Dots indicate the median of the predictive distribution, while the thick and thin lines represent central 80% and 95% intervals, respectively. Horizontal dashed lines indicate the true peak hospital occupancy, while the vertical dashed lines indicate the true peak hospital occupancy time.
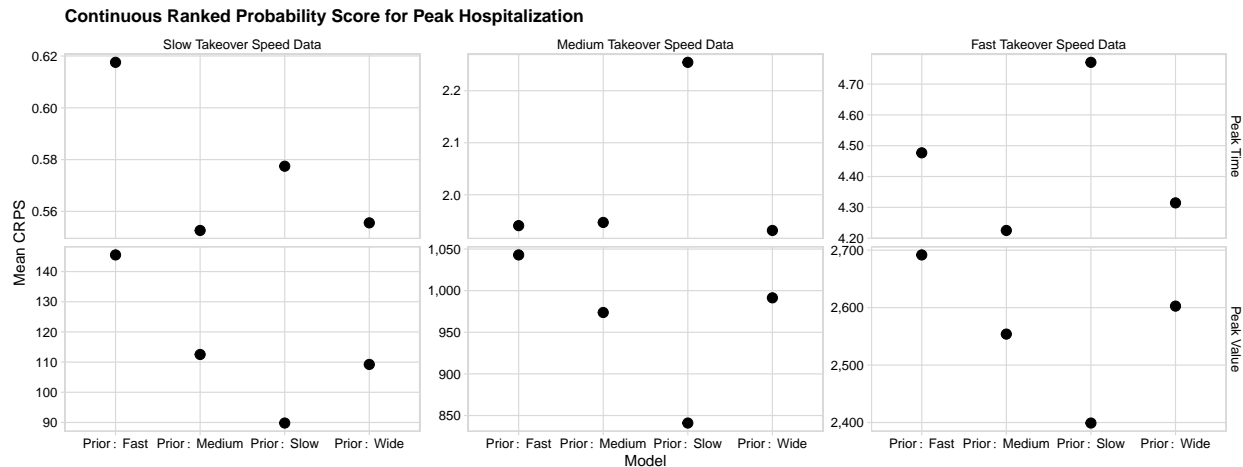
Continuous Ranked Probability Score for Peak Hospitalization

Figure C.16: CRPS summaries for peak hospital occupancy timing and size for three simulated data sets. Lower CRPS is better.

C.4 Additional California data results

Table C.4: Priors used in real data application in Section 5.3.2. In model 1, $R_0(t)$ is a priori modeled as a GMRF and $1/\kappa(t)$ is constant. In model 2, $R_0(t)$ is constant and $1/\kappa(t)$ is a priori modeled as a GMRF. In model 3, $R_0(t)$ is constant and $1/\kappa(t)$ is a function of the proportion of infectious individuals infected with the novel variant shown in (5.12). "CA" indicates the prior is only used when fitting the statewide California data. "OC" indicates the prior is only used when fitting the Orange County data.

Parameter	Interpretation	Prior	Prior Median (95% Interval)	Models
$1/\gamma$	mean latent period duration (weeks)	Log-Normal(-1.25, 0.04)	0.286 (0.193, 0.423)	1,2,3

$1/\nu$	Mean infectious period duration (weeks)	Log-Normal(-0.154, 0.04)	0.857 (0.579, 1.270)	1,2,3
$1/\eta$	mean hospitalization duration (weeks)	Log-Normal(-1.03, 0.04)	0.357 (0.241, 0.529)	1,2,3
$1/\omega$	mean ICU stay duration (weeks)	Log-Normal(-0.56, 0.04)	0.571 (0.386, 0.846)	1,2,3
τ	Infection-hospitalization ratio	Logit-Normal(-4.6, 0.04)	0.01000 (0.00678, 0.01470)	1,2,3
υ	Hospitalization-ICU admission ratio	Logit-Normal(-1.59, 0.04)	0.170 (0.122, 0.233)	1,2,3
χ	ICU-fatality ratio	Logit-Normal(-1.73, 0.04)	0.150 (0.107, 0.207)	1,2,3

$1/\sqrt{\phi_W}$	over-dispersion parameter for observed new cases	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	1,2,3
$1/\sqrt{\phi_X}$	over-dispersion parameter for observed hospital occupancy	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	1,2,3
$1/\sqrt{\phi_Y}$	over-dispersion parameter for observed ICU occupancy	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	1,2,3
$1/\sqrt{\phi_Z}$	over-dispersion parameter for observed new deaths	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	1,2,3
ρ^Z	death detection rate	Logit-Normal(2.2, 0.04)	0.900 (0.859, 0.930)	1,2,3

$E(0)$	Initial proportion of infected, but not yet infectious population	Logit-Normal(-7.91, 0.04)	0.000366 (0.000247, 0.000542)	1,2,3 CA
$E(0)$	Initial proportion of infected, but not yet infectious population	Logit-Normal(-8.64, 0.04)	0.000177 (0.000119, 0.000261)	1,2,3 OC
$I(0)$	Initial proportion of infectious population	Logit-Normal(-7.22, 0.04)	0.000732 (0.000495, 0.001080)	1,2,3 CA
$I(0)$	Initial proportion of infectious population	Logit-Normal(-7.95, 0.04)	0.000353 (0.000239, 0.000522)	1,2,3 OC
$R(0)$	Initial proportion of immune population	Logit-Normal(1.1, 0.16)	0.750 (0.578, 0.868)	1,2,3

logistic $\left(\tilde{\rho}_1^W\right)$	initial case detection rate	Logit-Normal(-1.39, 0.04)	0.200 (0.145, 0.270)	1,2,3
σ_{ρ^W}	standard deviation for case detection rate GMRF	Log-Normal(-2, 0.01)	0.135 (0.111, 0.165)	1,2,3
$1/\kappa$	immunity duration	Log-Normal(2.48, 0.04)	12.00 (8.11, 17.80)	1
$\exp\left(\tilde{R}_{0,1}\right)$	initial basic reproduction number	Log-Normal(1.61, 0.04)	5.00 (3.38, 7.40)	1
σ_{R_0}	standard deviation for basic reproduction number GMRF	Log-Normal(-2, 0.01)	0.135 (0.111, 0.165)	1
$\exp\left(1/\tilde{\kappa}_1\right)$	initial immunity duration	Log-Normal(2.48, 0.04)	12.00 (8.11, 17.80)	2
σ_κ	standard deviation for immunity duration GMRF	Log-Normal(-2, 0.01)	0.135 (0.111, 0.165)	2

			5.00	2,3
R_0	basic reproduction number	Log-Normal(1.61, 0.04)	(3.38, 7.40)	
$\exp(\alpha_0)$	maximum immunity duration	Log-Normal(2.48, 0.0025)	12.0 (10.9, 13.2)	3
α_1^*	proportion of maximum immunity duration that is minimum immunity duration	Logit-Normal(-1.1, 0.09)	0.250 (0.156, 0.375)	3
α_2	shape of transition from maximum immunity duration to minimum immunity duration	Log-Normal(0, 0.25)	1.000 (0.375, 2.660)	3

ζ	time offset (weeks) for relationship between immunity duration $1/\kappa(t)$ and proportion of novel variant $\delta(t)$	Log-Normal(0, 0.64)	1.000 (0.208, 4.800)	3
ι_1^*	time novel variant takes to grow from 1% to 99% of the infectious population (weeks)	Log-Normal(1.95, 0.16)	7.0 (3.2, 15.3)	3
ι_0	proportion of novel variant at t^*	Logit-Normal(-2.94, 0.25)	0.0500 (0.0194, 0.1230)	3
$1/\sqrt{\phi_V}$	over-dispersion parameter for observed new sequences of new variant	Half-Normal(0,1)	0.6740 (0.0313, 2.2400)	3

Continuous Ranked Probability Score for New Cases
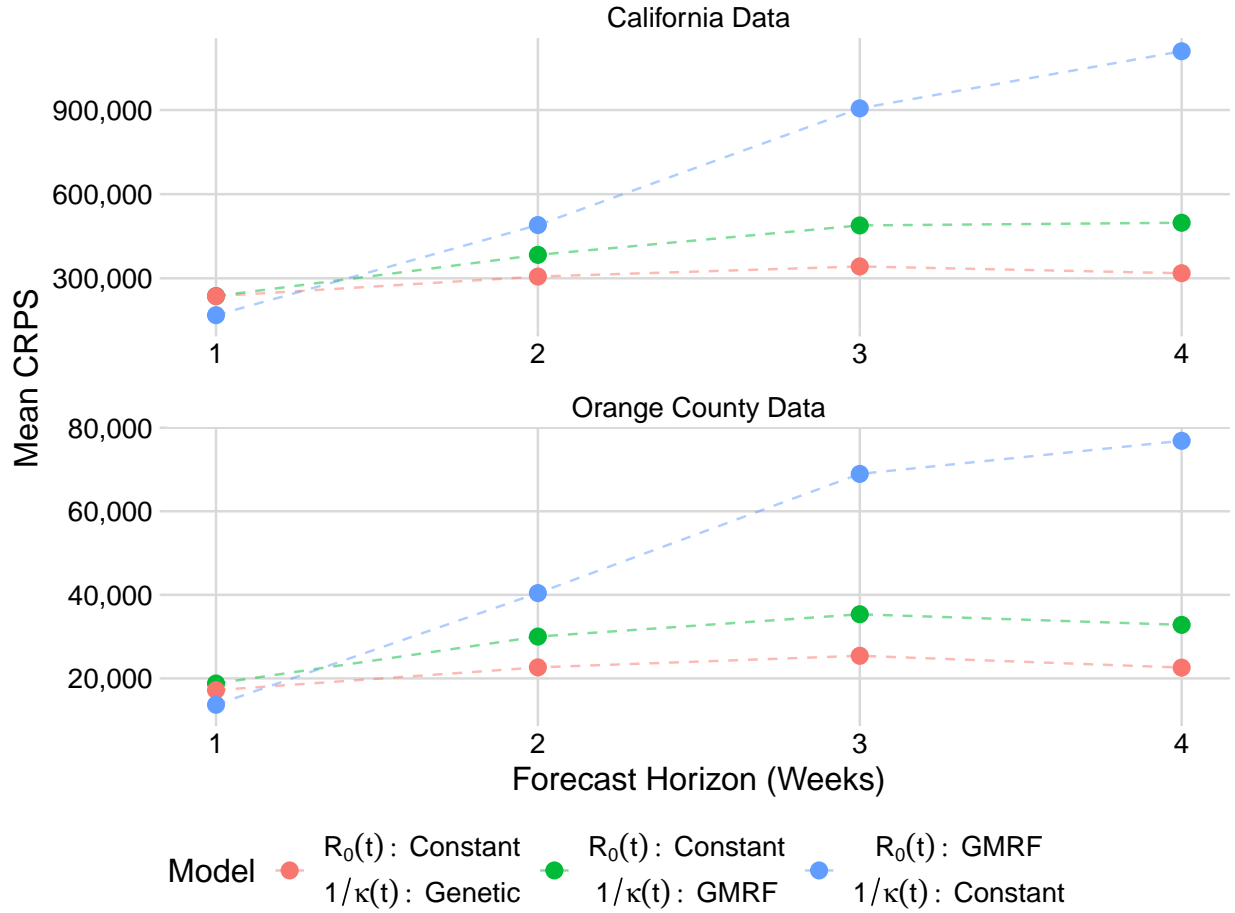
Figure C.17: CRPS summaries for new cases forecasts at 1, 2, and 4-week horizons for California and Orange County data sets. Lower CRPS is better.
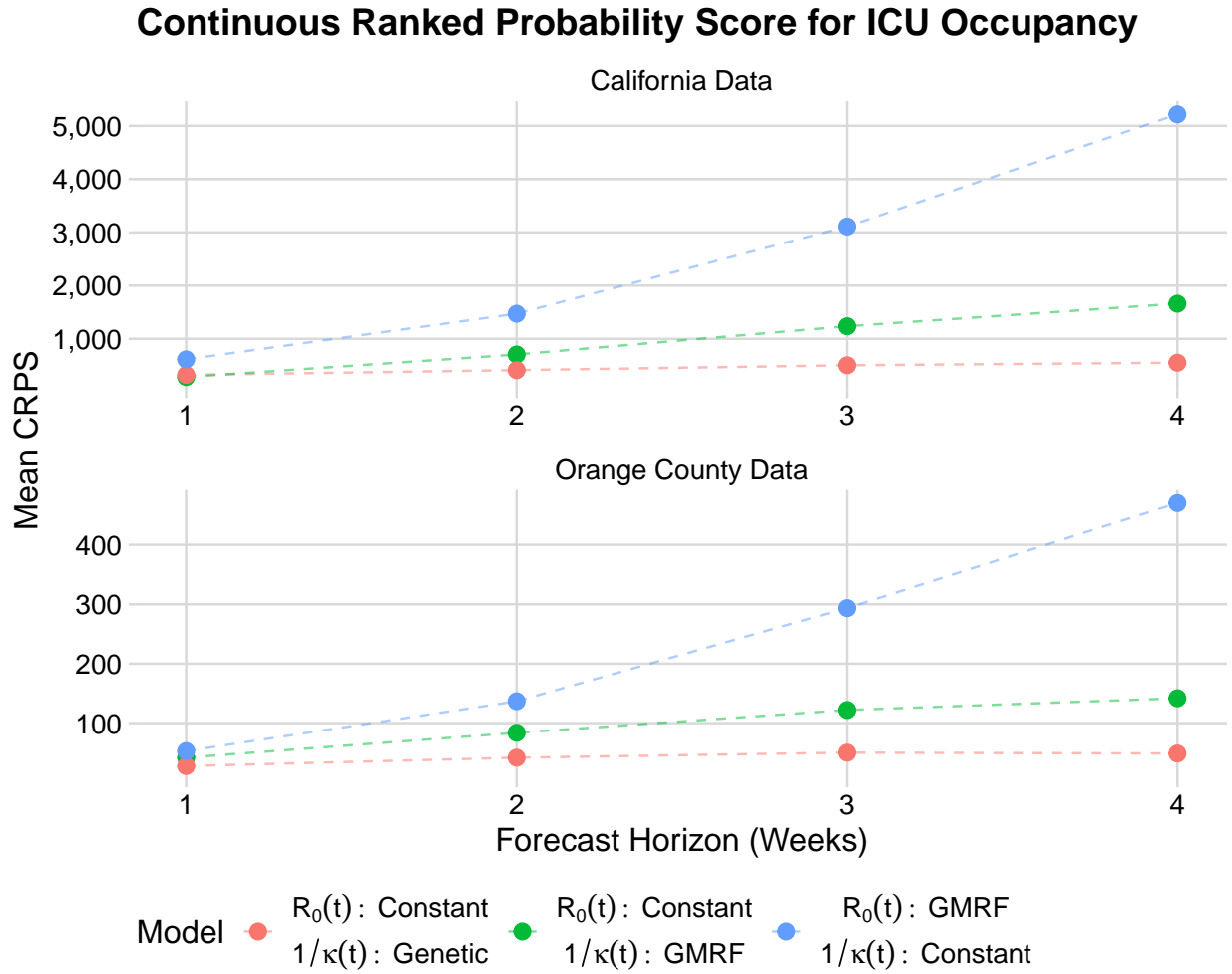
Figure C.18: CRPS summaries for ICU occupancy forecasts at 1, 2, and 4-week horizons for California and Orange County data sets. Lower CRPS is better.
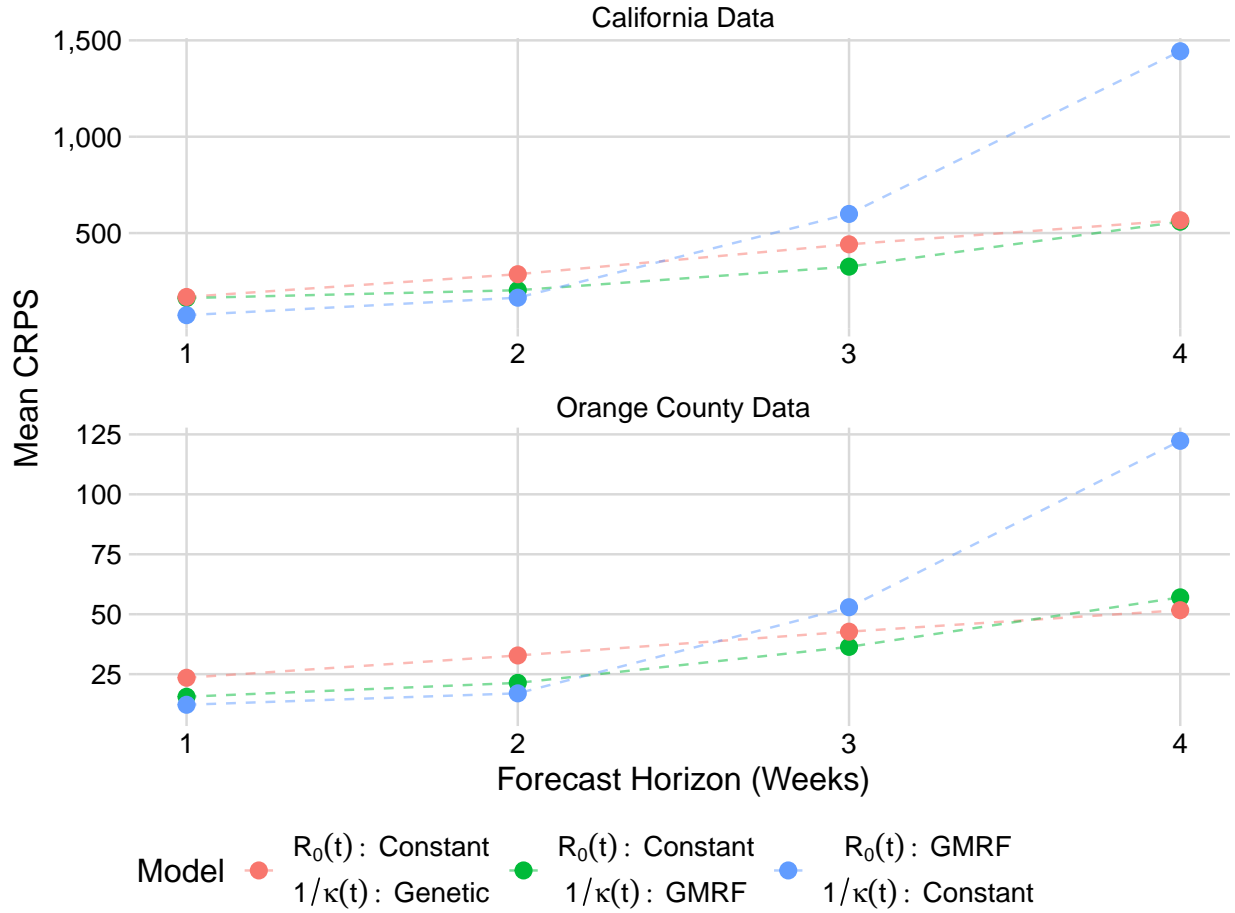
Continuous Ranked Probability Score for New Deaths

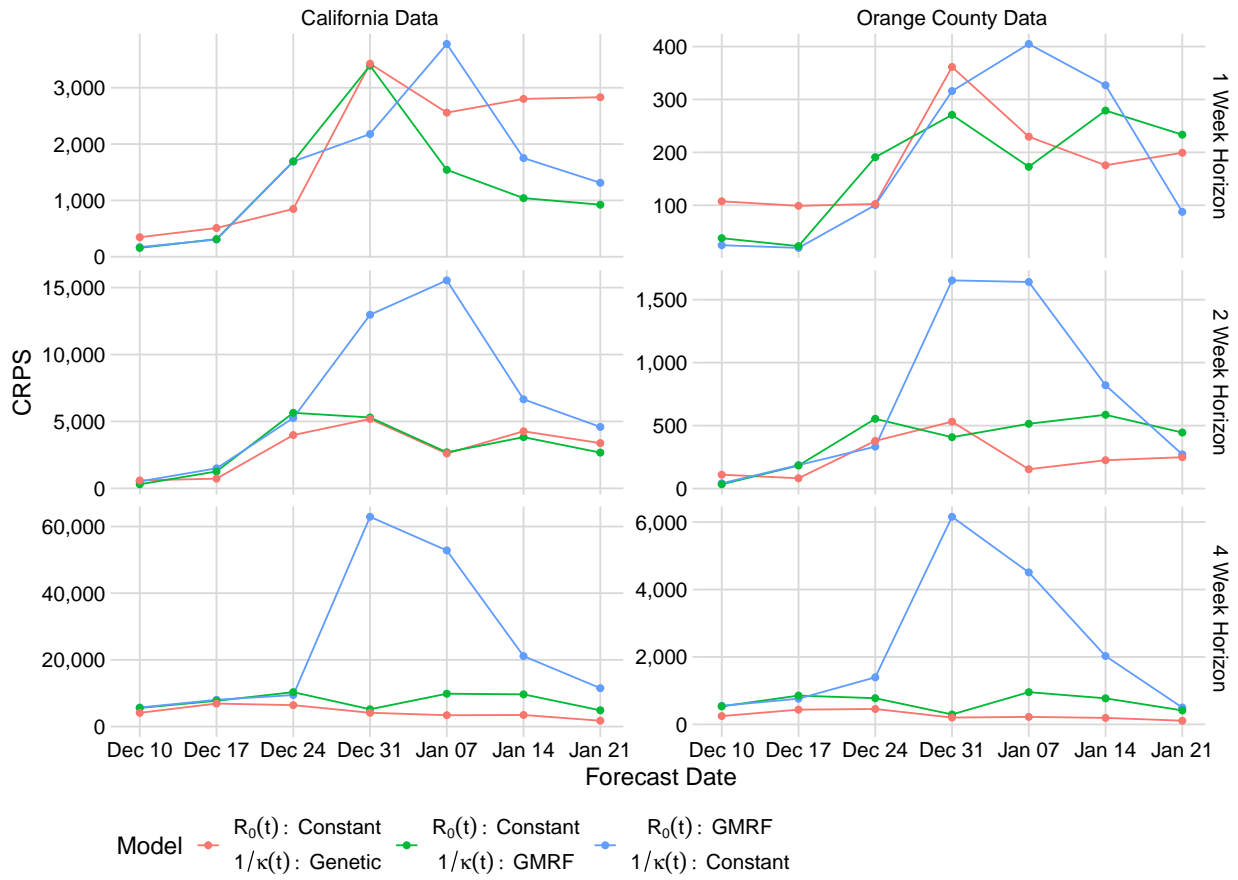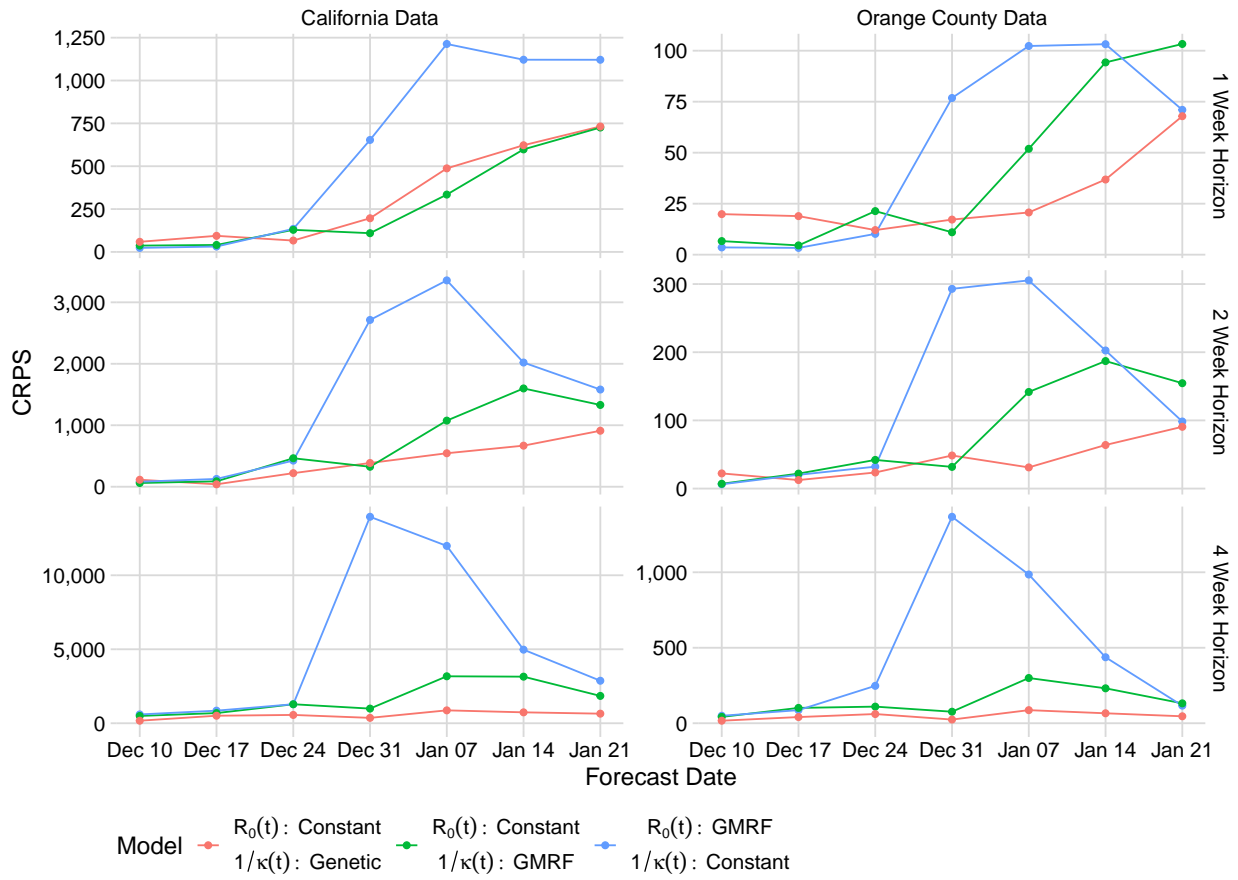Figure C.19: CRPS summaries for new deaths occupancy forecasts at 1, 2, and 4-week horizons for California and Orange County data sets. Lower CRPS is better.
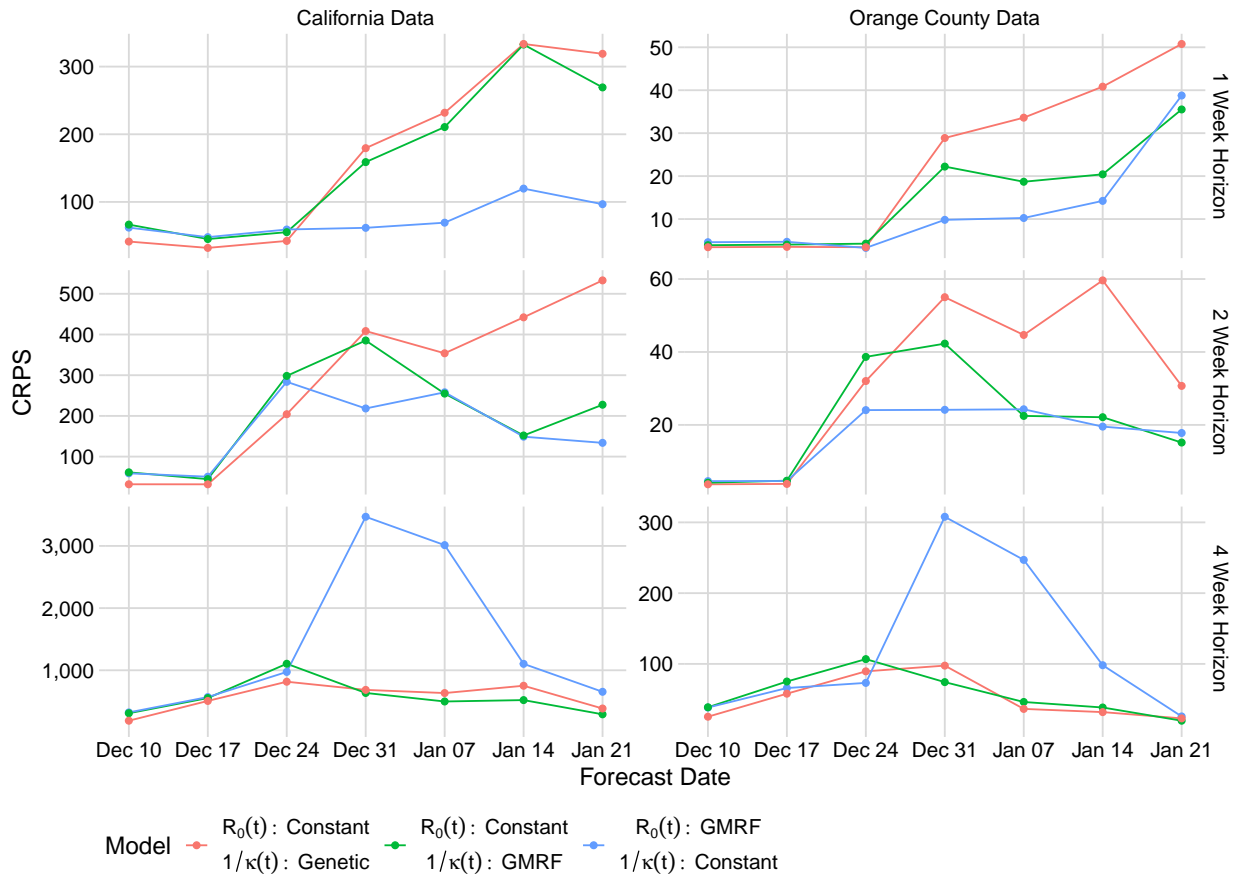
Figure C.20: Individual CRPS for new cases forecasts at 1, 2, and 4-week horizons for California and Orange County data sets. Lower CRPS is better.

Figure C.21: Individual CRPS for hospital occupancy forecasts at 1, 2, and 4-week horizons for California and Orange County data sets. Lower CRPS is better.

Figure C.22: Individual CRPS for ICU occupancy forecasts at 1, 2, and 4-week horizons for California and Orange County data sets. Lower CRPS is better.

Figure C.23: Individual CRPS for new deaths forecasts at 1, 2, and 4-week horizons for California and Orange County data sets. Lower CRPS is better.
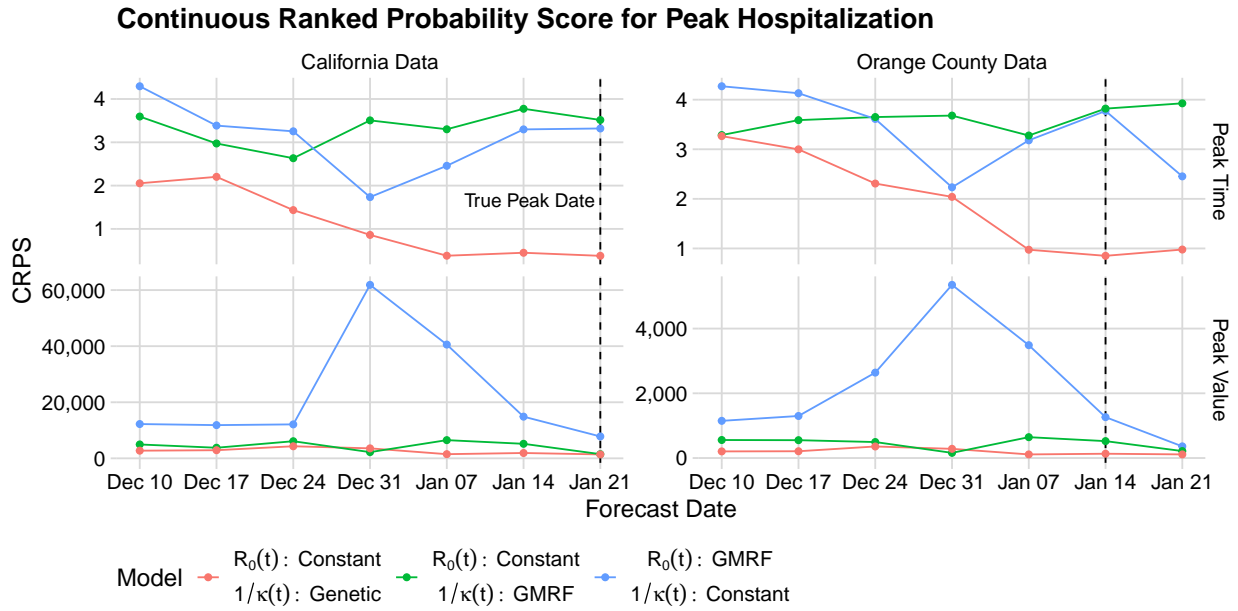
Figure C.24: Individual CRPS for peak hospital occupancy timing and size in California and Orange County data sets. Lower CRPS is better.